The Journal of Pathology: Clinical Research
J Pathol Clin Res 2023
Published online 3 July 2023 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cjp2.335

ORIGINAL ARTICLE

The epigenetic modifier *lysine methyltransferase 2C* is frequently mutated in gastric remnant carcinoma

Bo Sun[1,2†], Haojie Chen[3†], Jiawen Lao[4†], Cong Tan[5], Yue Zhang[5], Zhen Shao[3*] and Dazhi Xu[1,2*]

[1]*Department of Gastric Surgery, Fudan University Shanghai Cancer Center, Shanghai, PR China*
[2]*Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, PR China*
[3]*CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, PR China*
[4]*Department of Gastric Surgery, Sun Yat-sen University Cancer Center, Guangzhou, PR China*
[5]*Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, PR China*

**Correspondence to: Zhen Shao, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, PR China. E-mail: shaozhen@picb.ac.cn; Dazhi Xu, Department of Gastric Surgery, Fudan University Shanghai Cancer Center, Shanghai 200030, PR China. E-mail: xudzh@shca.org.cn*

†These authors contributed equally to this work.

Abstract

Gastric remnant carcinoma (GRC), which occurs in the stomach after partial gastrectomy, is a rare and aggressive form of gastric adenocarcinoma (GAC). Comprehensive profiling of genomic mutations in GRC could provide the basis for elucidating the origin and characteristics of this cancer. Herein, whole-exome sequencing (WES) was performed on 36 matched tumor–normal samples from patients with GRC and identified recurrent mutations in epigenetic modifiers, notably *KMT2C*, *ARID1A*, *NSD1*, and *KMT2D*, in 61.11% of cases. Mutational signature analysis revealed a low frequency of microsatellite instability (MSI) in GRC, which was further identified by MSIsensor, MSI-polymerase chain reaction, and immunohistochemistry analysis. Comparative analysis demonstrated that GRC had a distinct mutation spectrum compared to that of GAC in The Cancer Genome Atlas samples, with a significantly higher mutation rate of *KMT2C*. Targeted deep sequencing (Target-seq) of an additional 25 paired tumor–normal samples verified the high mutation frequency (48%) of *KMT2C* in GRC. *KMT2C* mutations correlated with poor overall survival in both WES and Target-seq cohorts and were independent prognosticators in GRC. In addition, *KMT2C* mutations were positively correlated with favorable outcomes in immune checkpoint inhibitor-treated pan-cancer patients and associated with higher intratumoral $CD3^+$, $CD8^+$ tumor-infiltrating lymphocyte counts, and PD-L1 expression in GRC samples ($p = 0.018$, 0.092, 0.047, 0.010, and 0.034, respectively). Our dataset provides a platform for information and knowledge mining of the genomic characteristics of GRC and helps to frame new therapeutic approaches for this disease.

Keywords: gastric carcinoma; whole-exome sequencing; microsatellite instability; tumor microenvironment

Received 27 July 2022; Revised 20 May 2023; Accepted 25 May 2023

No conflicts of interest were declared.

Introduction

Gastric remnant carcinoma (GRC) is a relatively rare malignancy that develops in the remnant stomach after partial gastrectomy, regardless of the initial pathology or reconstruction technique [1]. GRC has an incidence of 1–5% after gastrectomy. Due to adjacent organ invasion and extensive lymph node metastasis, GRC has an unfavorable prognosis [2], and improving its clinical outcome remains a challenge.

Recently, omics analyses have been performed to determine the genomic characteristics of gastric adenocarcinoma (GAC). The Cancer Genome Atlas (TCGA) has conducted comprehensive genomic and transcriptomic analyses of GAC and categorized it into four molecular subtypes [3]. Subsequently, the Asian Cancer Research Group classified GAC into four subtypes with different outcomes [4]. Recently, several studies have identified the molecular features of some GAC subtypes, such as diffuse-type GAC, early onset

GAC, and Epstein–Barr virus $(EBV)^+$ GAC [5–7]. These clinicopathological subtypes and associated omics signatures allow better patient stratification and personalized treatment. However, the genomic characteristics of GRC, a rare GAC subtype, remain largely unexplored.

In this study, we aimed to uncover the underlying genomic landscape of GRC by performing whole-exome sequencing (WES) on a cohort of 36 paired tumor–normal samples and targeted deep sequencing (Target-seq) on an additional set of 25 paired tumor–normal samples. We provide an integrated analysis of somatic sequence mutations, tumor-infiltrating immune cells, and programmed death-ligand 1 (PD–L1) expression in GRC. These results provide valuable biological and clinical insights into the underlying genomic features of this disease and possible new therapeutic targets.

Materials and methods

Patients and samples

We obtained 36 formalin-fixed, paraffin-embedded paired tumor–normal samples from GRC patients at Sun Yat-sen University Cancer Center for WES. An expanded cohort of 25 GRC samples was obtained from Fudan University Shanghai Cancer Center as the validation cohort for Target-seq. All patients were independently diagnosed with GRC after surgery by at least two pathologists. Patient demographics, tumor characteristics, and survival information were obtained. This study was approved by the ethics committees of both participating institutions.

WES and data processing

DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) according to standard protocols. Genomic DNA was extracted from tumor-adjacent normal tissues using the DNeasy Blood & Tissue Kit (Qiagen) and sequenced as a germline reference. Exon capture was performed using Clinical Research Exome V2 (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA, USA), with 150 bp paired-end reads.

Adapters and low-quality bases in FASTQ reads were removed, and the clean reads were aligned to the Genome Reference Consortium Human Build 37 (GRCh37) using BWA [8]. QualiMap was used for further quality control of the alignment result [9]. The BAM files were then subjected to realignment, duplicate marking, and recalibration using the Picard and GATK software tools [10]. The resulting BAM files were used for further downstream analyses. We ran MuTect for single-nucleotide variants (SNVs) [11] and Pindel for small insertions and deletions (indels) [12]. Only MuTect calls marked as 'KEEP' were selected for SNVs analysis. For indels, mutations with low coverage (<20 reads for tumor samples; <10 reads for normal samples) or identified as germline mutations in other normal samples were removed. ANNOVAR was used to annotate all somatic mutations after filtering [13]. Somatic mutations annotated as nonsynonymous or unknown were excluded from downstream analysis.

Target-seq analysis

Library preparation was performed using a total amount of 40 ng genomic DNA per sample with adaptor and barcode ligation. Sequencing adaptors and barcodes were ligated to the amplicons using ligase. After the barcoded library construction and purification, Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Q33216, Waltham, MA, USA) and Qsep100 (BiOptic, New Taipei City, Taiwan) were used to quantify concentrations and determine the length of library fragments (from 320 to 420 bp), respectively. Sequencing libraries were generated using MultipSeq® Custom Panel (iGeneTech, Beijing, PR China) following the manufacturer's recommendations and index codes were added to each sample. An Illumina Amplicon library was sequenced on a NovaSeq 6000 instrument (Illumina) on rapid run mode to a read depth of 5,000× (~0.2 GB of total data with 2 × 150 nt reads). The reads with low quality were deleted from the raw data using Trimmomatic-0.38. After removing primer sequences, the reads were mapped to the hg19 reference sequence with BWA-0.7.12. Variants were then called using Varscan-2.4.3 according to best practice (min MQ ≥20, min coverage ≥4); indel realigned by GATK.3.8.1. The final variant sets were annotated using Annovar-201707. The quality of Target-seq was then adjusted in the downstream analysis. In the bioinformatic analysis, we defined variants with variant allele frequency (VAF) > 2% as somatic mutations; most mutated samples had at least one variant with VAF > 5% in this study (i.e. 9/12).

Mutational signature and microsatellite instability analysis

Mutational signature decomposition was performed using the Mutalisk toolkit, based on 30 characterized

COSMIC mutational signatures [14,15]. The SNVs from each individual were aggregated to achieve mutational signature decomposition. Mutalisk was further run-on aggregated mutations from GRC cohort and TCGA-STAD. The proportion of the top signatures is presented for comparison between the GRC cohort and TCGA-STAD. Microsatellite instability (MSI) was detected using MSIsensor, with a score >3.5 indicating MSI, according to the original publication [16].

Somatic copy number variation

Somatic copy number variation (SCNV) analysis was performed using VarScan2 [17]. After correcting for sequencing depth and GC bias, log2 copy number changes were subjected to the DNAcopy R package for SCNV calling. Log2 copy ratios >1 or <−1 were defined as copy number gains or losses, respectively. The SCNV burden was defined as the proportion of the genome with copy number gain or loss against the total length of the profiled genome.

Pathway enrichment analysis

Genes with mutation frequency ≥10% were subjected to gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses using the gseapyPython package [18]. Only the GO terms for molecular functions were used in the GO enrichment analysis. Top-ranked significantly enriched pathways were selected to demonstrate recurrently altered pathways.

Comparison of gene mutation rates

We downloaded TCGA-STAD data from the GDC database using the function GDCquery of the TCGAbiolinks R package and molecular classification from the GDC Data Portal (https://portal.gdc.cancer.gov/) [19]. We then compared gene mutation rates and mutational signatures between GRC and TCGA-STAD samples (including four subtypes of GAC).

Immunohistochemistry

Immunohistochemistry (IHC) was performed using standard procedures [20]. Tissue sections were incubated with the following antibodies: anti-KMT2C (1:1,000), anti-H3K4me1 (1:1,000), anti-H3K4me3 (1:1,000), anti-CD163 (1:500), anti-CD3 (1:200), anti-CD4 (1:1,000), anti-CD8 (1:2,000), and anti-PD-L1 (1:500) antibodies (all from Abcam, Cambridge, UK). Details are given in Supplementary materials and methods.

Helicobacter pylori and EBV testing

The presence of EBV infection was determined using an *in situ* hybridization kit (Zsbio, Beijing, PR China) according to the manufacturer's instructions on formalin-fixed, paraffin-embedded samples. Histologic diagnosis of *Helicobacter pylori* was based on Wright-Giemsa staining for each specimen and was made by two experienced pathologists (CT and YZ) without discrepancy.

Statistical analysis

Statistical significance was determined using the two-tailed *t* test, rank-sum test (for mutations associated with the SCNV burden), or Fisher's exact test (for differences in mutation rates). Survival rates were estimated and compared using the Kaplan–Meier method and log-rank test. Statistical significance was set at $p < 0.05$.

Results

Patient characteristics

The experimental flowchart is presented in supplementary material, Figure S1. Patient demographic and clinical characteristics are shown in Table 1. The mean age at diagnosis in the WES cohort was 64.5 years. Among the 36 cases, 21 (58.3%) had a history of benign diseases, while 15 (41.7%) had malignant diseases. The time interval between the previous gastrectomy and the current diagnosis was 22.1 ± 14.3 years. The reconstruction methods included B-I (16.7%), B-II (72.2%), and others (11.1%). Seventeen patients had tumors at the anastomotic site, whereas the remaining had tumors in a nonanastomotic location. Tumors were well or moderately differentiated in 17 cases, while the rest were poorly differentiated. Based on the eighth edition of the TNM classification, there were 8 stage I, 10 stage II, and 18 stage III patients. The rate of *H. pylori* infection was 41.7% and the frequency of EBV infection was 22.2%. Four patients received neoadjuvant therapy, and 24 received postoperative chemotherapy or radiation therapy. After a median follow-up of 30.50 months, 12 patients died and 24 were alive.

Genomic landscape of GRC

The WES analysis pipeline is illustrated in supplementary material, Figure S2. WES was performed on the

Table 1. Clinical charateristics

	WES	Target-seq
Total number of tissues	36	25
Gender, *n* (%)		
Female	5	3
Male	31	22
Age (years), mean ± SD	64.50 ± 7.78	61.56 ± 9.30
Initial disease		
Benign	21	11
Cancer	15	14
Interval (median, range) (years)	22.14 ± 14.33	20.12 ± 14.83
Initial gastrectomy (%)		
Billroth I	6	7
Billroth II	26	15
Others	4	3
Tumor location		
Anastomotic site	17	10
Nonanastomotic site	19	15
Grade		
Well or moderate	17	12
Poor	19	13
T stage		
1–2	10	8
3–4	26	17
N stage		
0	15	10
1–3	21	15
TNM stage		
1–2	18	10
3–4	18	15
CEA		
≦5 ng/ml	31	20
>5 ng/ml	5	5
Epstein–Barr Encoding Region (EBER)		
Negative	28	20
Positive	8	5
Helicobacter pylori infection		
Negative	21	15
Positive	15	10
MSI status		
MSS	35	25
MSI	1	0
Adjuvant treatment		
None	8	3
Neoadjuvant ± adjuvant	4	3
Chemotherapy and/or radiation therapy	24	19

genomic DNA of 36 GRC and matched normal samples, with an average coverage of 81× and 31×, respectively (supplementary material, Figure S3 and Table S1). Sixty-one nonsynonymous SNVs and 126.5 indels per sample were identified upon variant filtering, equivalent to 3.25 nonsynonymous mutations per megabase of targeted DNA (supplementary material, Table S2). The predominant single-nucleotide mutations were C > T transitions and C > A transversions, accounting for nearly 60% of SNVs (Figure 1A). The most frequently mutated genes in GRC were *KMT2C* (41.7%), *TP53* (30.6%), *MUC16* (25.0%), *MUC4* (19.4%), *ARID1A* (19.4%), and *PIK3CA* (19.4%) (Figure 1A). A comparison was made between the gene mutation rates of GRC and TCGA-STAD (supplementary material, Table S3). The mutation rates of *KMT2C* (36.1% versus 8.8%, $p < 0.0001$) and *GNAS* (13.9% versus 4.2%, $p = 0.02$) were significantly higher in GRC than in TCGA-STAD (Figure 1B). Additionally, the *KMT2C* mutation rate in GRC significantly surpassed that in TCGA-GS, TCGA-EBV, and TCGA-CIN, but not in TCGA-MSI, which is characterized by mismatch repair (MMR) deficiency (Figure 1C). The mutation distributions of *TP53* and *PIK3CA* are shown in supplementary material, Figure S4, and the association between mutated genes and GRC clinicopathological characteristics in supplementary material, Figure S5 and Table S4.

To decipher the underlying biological process operative in generating the mutational profile of GRC, we analyzed the mutation signature compositions in GRC and TCGA-STAD. We found six base substitution signatures in GRC specimens (Figure 1D), according to the COSMIC nomenclature: signatures #1 (34.8%), #6 (21.0%), #3 (19.5%), #5 (17.4%), #17 (4.3%), and #7 (3.1%). Signature #6 correlated with defective DNA MMR and MSI status. The proportion of signature #6 was lower in GRC than in TCGA-STAD (21.0% versus 35.0%). In addition, signatures #15 (22.8% in TCGA-STAD), #26 (1.7% in TCGA-STAD), #13 (1.0% in TCGA-STAD), and #10 (1.0% in TCGA-STAD) were all not observed in GRC. These signatures were associated with defective MMR or somatic hotspot mutations in *POLE*, suggesting a low MSI frequency in GRC. To further assess the MSI status in GRC, MSIsensor was used to evaluate the somatic MSI index in these samples [21]; most GRC samples (35/36) were identified as microsatellite stable (MSS) (supplementary material, Figure S6A and Table S5). We further determined the MSI status in an expanded cohort of 25 GRC samples using MSI-polymerase chain reaction and IHC analysis. Both assays indicated that all samples were MSS.

We then identified SCNVs in GRC and correlated the mutated genes with the corresponding SCNV burden. *ARID1A* mutation was associated with decreased copy-number instability ($p < 0.05$), whereas *TP53* mutation was associated with increased copy-number instability ($p < 0.05$) (supplementary material, Figure S6B).

GO and KEGG enrichment analyses

GO and KEGG enrichment analyses are instrumental in determining the functionality and interactions of

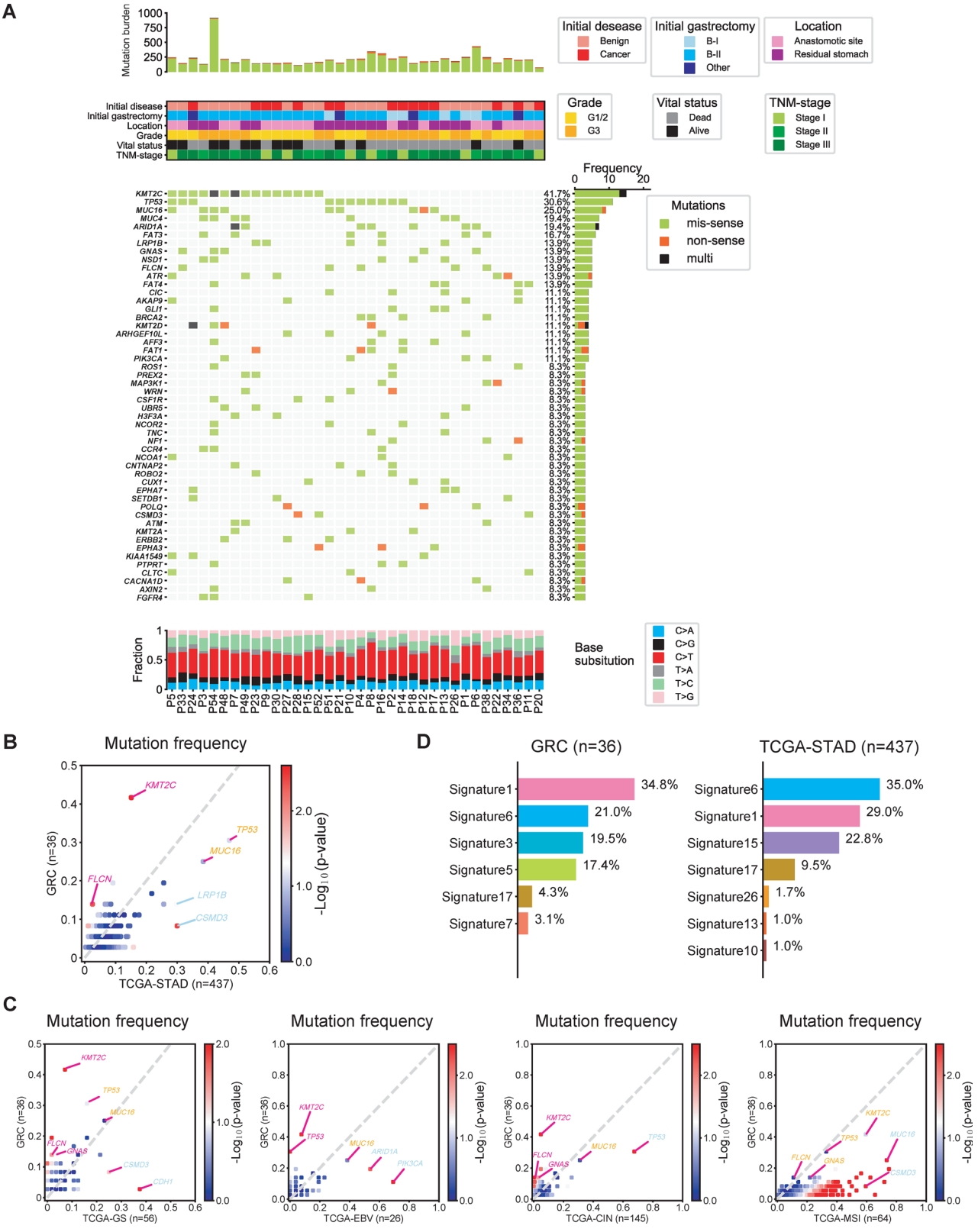

Figure 1. Legend on next page.

genes and their products, contributing to the understanding of biological processes and pathways. In this study, the top five most significantly enriched KEGG pathways were 'Lysine degradation', 'Pathways in cancer', 'Cellular senescence', 'Fanconi anemia pathway', and 'p53 signaling pathway' (Figure 2A). The GO enrichment analysis indicated that significantly mutated genes were predominantly enriched in epigenetic regulation, such as 'histone-lysine N-methyltransferase activity', 'histone methyltransferase activity', 'histone methyltransferase activity (H3-K4 specific)', 'DNA binding', and 'histone methyltransferase activity (H3-K36 specific)' (Figure 2B and supplementary material, Table S6).

Frequent mutations in epigenetic modifiers

Since most of the highly mutated genes were enriched in epigenetic regulation, we further analyzed the alteration frequency in epigenetic modifiers with a mutation rate ≥10% in the database of epigenetic modifiers. We found that 22/36 GRC samples (61.1%) carried at least one mutated epigenetic modifier [22]. *KMT2C* was mutated in 41.7% (15/36) of cases, followed by *ARID1A* (19.4%), *NSD1* (13.9%), and *KMT2D* (11.1%) (Figure 2C). Moreover, the relationship between mutations in epigenetic modifiers and overall survival (OS) was assessed. Patients with mutations in epigenetic modifiers had worse OS than those without such mutations (Hazard Ratio (HR) = 3.78, 95% CI = 0.81–16.65, $p = 0.070$) (Figure 2D). These results suggest that the mutation status of epigenetic modifiers has a prognosis value in GRC.

A linear depiction of the protein domains of the frequently mutated epigenetic modifiers in GRC is shown in Figure 3A–C. Eighteen out of 25 (72.0%) mutations in *KMT2C* were enriched in the plant homeodomain (PHD) ($p = 3.1e{-}15$) in GRC samples (Figure 3A), while *KMT2C* mutations in TCGA-STAD were dispersed throughout the entire gene (Figure 3B). Recent findings suggest that missense mutations in *PHD* diminish KMT2C recruitment to gene enhancers [23], implying an oncogenic effect of *KMT2C* mutations in this domain in GRC.

Target-seq validation of the *KMT2C* mutation spectrum and its clinical significance

In the cohort subjected to WES, 24 *KMT2C* mutations were identified in 15/36 (41.7%) patients with GRC, consisting of 20 missense mutations and 4 truncation mutations (supplementary material, Table S7). Then, Target-seq was performed to verify *KMT2C* mutation in an extended cohort of 25 GRC patients. This cohort showed a high rate (12/25) of *KMT2C* mutations (Figure 4A). In this extended cohort, 21 mutations were detected in 12 patients, comprising 19 missense mutations and 2 splicing mutations (supplementary material, Table S8).

KMT2C mutations in different datasets collected from various GAC studies were analyzed using cBioPortal, an open-access tool for large-scale cancer genomic dataset analysis. The mutation frequency of *KMT2C* ranged from 2.56% to 13.00% in different GAC datasets [3,24–27] (Figure 4A and supplementary material, Table S9). The mutation frequency of *KMT2C* ranged from 10.13 to 14.81% in different GAC subtypes, including signet ring carcinoma of the stomach (14.81%), intestinal-type GAC (12.35%), diffuse-type GAC (11.81%), mucinous GAC (11.36%), and tubular GAC (10.13%) (Figure 4A and supplementary material, Table S10). The mutation frequency of *KMT2C* in GRC was significantly higher than that in TCGA-STAD.

The relationship between *KMT2C* mutation and OS in the Target-seq cohort is shown in Figure 4B. *KMT2C* mutation was associated with poorer OS (HaR = 8.02, 95% CI = 1.69–38.09, $p = 0.009$). Similarly, *KMT2C* mutation corresponded with reduced OS in the WES cohort (HR = 3.35, 95% CI = 1.02–11.02, $p = 0.035$) (Figure 4C). We combined patient data from both WES and Target-seq cohorts and conducted a Cox regression analysis

Figure 1. Landscape of somatic mutations and mutational signatures in GRC. (A) Somatic genomic mutations identified in GRC specimens. The top-ranked frequently mutated genes in GRC are displayed. The top-most histogram shows the frequency of nonsynonymous mutations in each patient, followed by tracks showing the clinical and histological characteristics of each patient and their tumor, respectively. The matrix in the middle shows somatic mutations by patient (column) and gene (row), and the histogram on its right-hand side shows the frequency of different mutations for each gene. The bar plot at the bottom shows the mutational spectrum for each patient. (B) Comparison of gene mutation rates between GRC specimens from this study and primary GAC from The Cancer Genome Atlas GAC cohort (TCGA-STAD). (C) Comparison of gene mutation rates between GRC specimens from this study and four subtypes of GAC from TCGA-STAD. (D) Distribution of somatic mutational signatures in GRC and TCGA-STAD. CIN, chromosomal instability; GS, genomic stable.

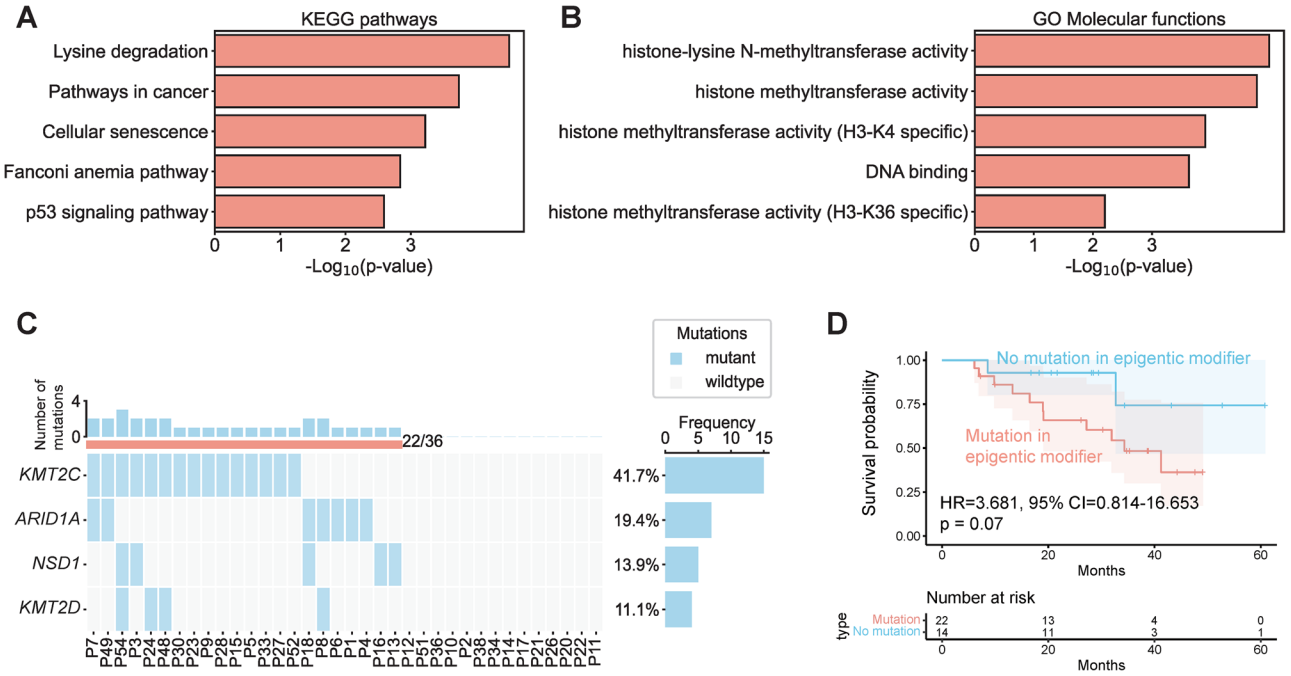

Figure 2. Recurrently mutated genes involved in epigenetic modification. (A, B) KEGG and GO enrichment analysis of genes with recurrent somatic mutations (>10%). (C) Somatic genomic alterations in epigenetic modifiers with mutation frequency >10%. (D) Kaplan–Meier survival curve of GRC patients stratified by the mutation status of epigenetic modifier genes.

(Figure 4D). *KMT2C* mutation (HR = 5.04, 95% CI = 1.82–14.01, $p = 0.002$) and tumor-associated factors, such as lymph node invasion (HR = 2.78, 95% CI = 1.01–7.69, $p = 0.049$), carcinoembryonic antigen (CEA) > 5 ng/ml (HR = 2.81, 95% CI = 1.07–7.39, $p = 0.037$), and carbohydrate antigen 199 (CA199) > 37 U/ml (HR = 3.42, 95% CI = 1.20–9.78, $p = 0.022$), were correlated significantly with decreased OS. Upon multivariate analysis, the independent prognosticators of OS encompassed lymph node invasion (HR = 2.99, 95% CI = 1.03–8.73, $p = 0.045$) and *KMT2C* mutation (HR = 5.00, 95% CI = 1.67–14.98, $p = 0.004$).

KMT2C mutation in primary and GRC patient samples

Primary surgical samples were collected from six GRC patients, including two primary benign samples and four primary malignant samples, among which two benign samples (P26 and P48) and one malignant sample (P27) failed to yield quality DNA. *KMT2C* mutations were detected in the remaining three malignant samples. The *KMT2C* mutation detected in the primary malignant sample from P28 (p.G315S) was consistent with one of the mutations (p.K2797fs, p.D348N, and p.G315S) in the GRC sample. The *KMT2C* mutations in the primary malignant sample and GRC sample were different in P23 (p.R866Q versus p.R284Q) and P24 (p.R3398W versus p.R380L and p.G315S).

Relationship between *KMT2C* mutations and the tumor immune microenvironment

Evidence has suggested *KMT2C* mutations as potential predictors for immunotherapy response in solid malignancies [28]. We obtained pan-cancer MSKCC data of patients undergoing immune checkpoint inhibitor (ICI) therapy from cBioPortal [29]. The relationship between the top-ranked frequently mutated epigenetic modifiers and the efficacy of ICI treatment was then analyzed. Mutated *KMT2C* was significantly and positively correlated with a better prognosis in patients receiving ICI treatment ($p = 0.001$), while *NSD1*, *ARID1A*, and *KMT2D* mutations were not (Figure 5A,B).

Hence, we investigated the impact of mutated *KMT2C* on the tumor immune microenvironment by immunostaining GRC tissue samples for CD3, CD4, CD8, CD163, and PD-L1 (Figure 5C). *KMT2C* mutations were associated with higher intratumoral $CD3^+$ and $CD8^+$ tumor-

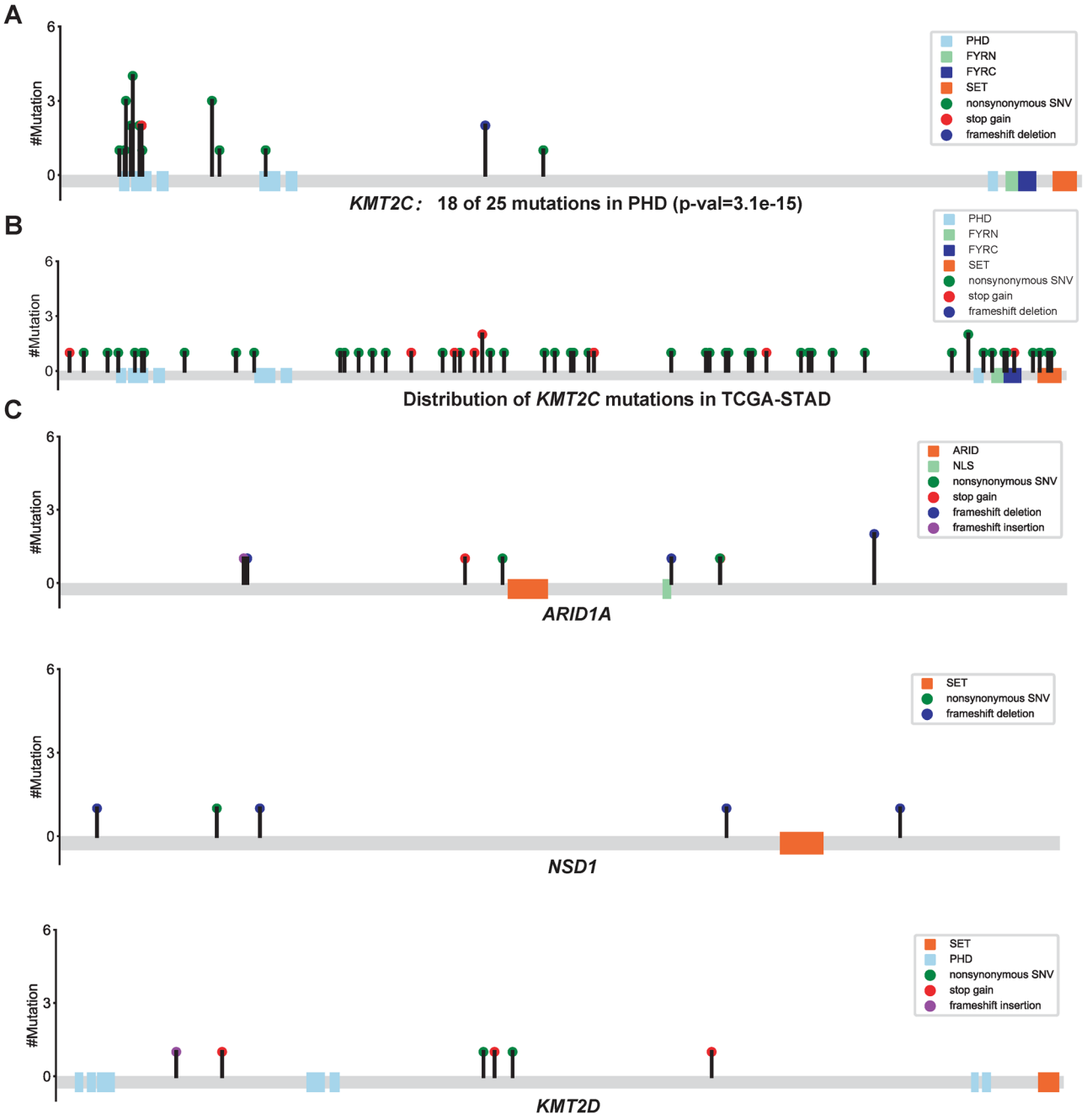

Figure 3. Mutation distribution in highly recurrent mutated epigenetic modifiers. Location and type of mutations in three highly recurrent mutated epigenetic modifiers. Protein domains of these gene products were annotated based on the Pfam database. Each variant is represented by a colored circle, with missense mutations in green, nonsense mutations in red, frameshift insertion mutations in blue and frameshift deletion mutations in purple. (A) Distribution of somatic *KMT2C* mutations identified in GRC. (B) Distribution of somatic *KMT2C* mutations identified in TCGA-STAD. (C) Distribution of somatic *ARID1A*, *NSD1*, and *KMT2D* mutations identified in GRC.

infiltrating lymphocyte (TIL) counts ($p = 0.018$ and 0.047, respectively). There was no significant relationship between the number of $CD4^+$ T cells and *KMT2C* mutations ($p = 0.092$). *KMT2C* mutations were also positively correlated with PD-L1 expression ($p = 0.034$). Furthermore, a significant correlation was found between $CD163^+$ tumor-associated macrophage infiltration and *KMT2C* mutations ($p = 0.010$).

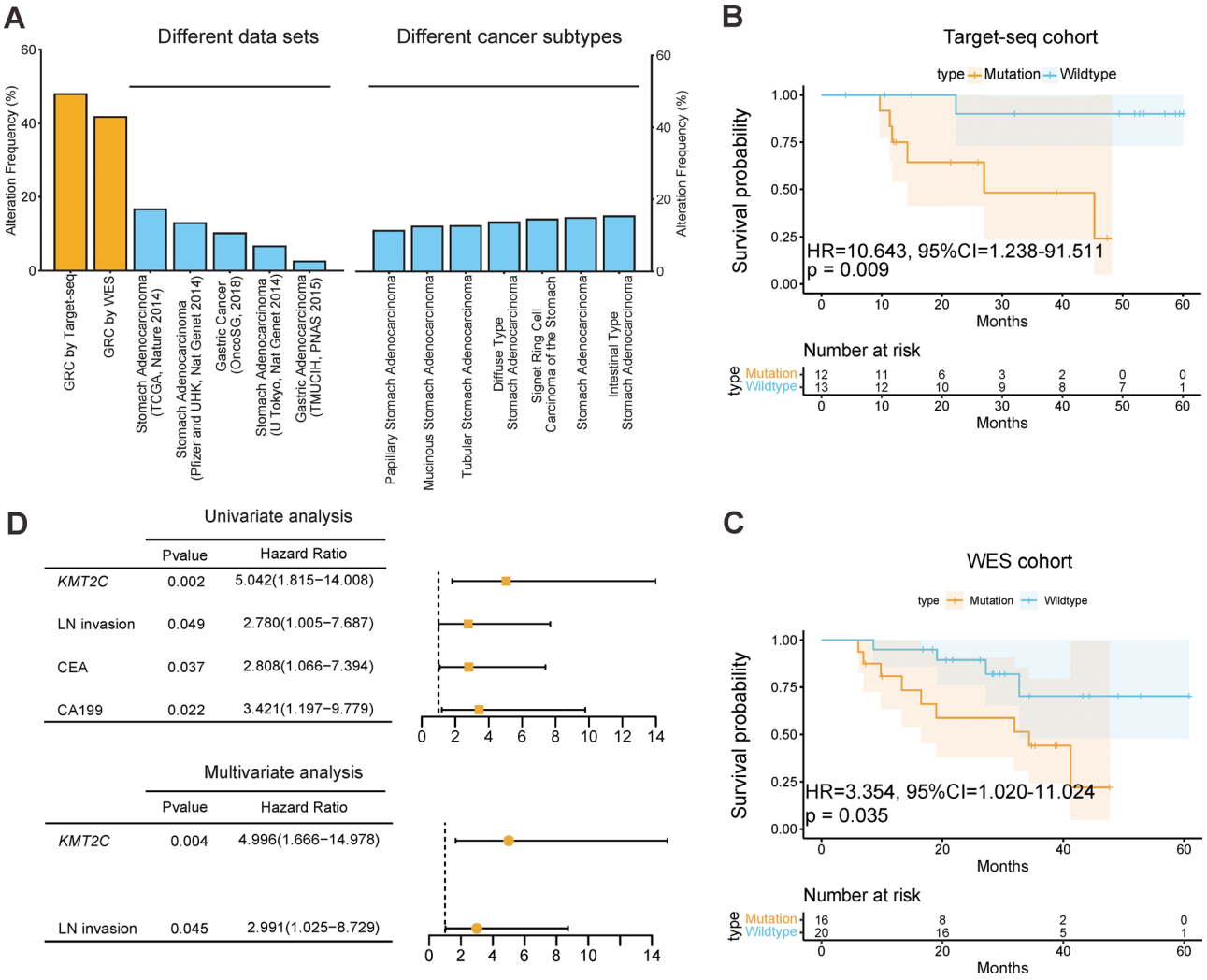

Figure 4. Verification and clinical significance of KMT2C mutations in GRC. (A) Frequency of KMT2C mutations identified by WES and Target-seq in GRC compared with different GAC subtypes and different GAC datasets. (B) Kaplan–Meier OS analysis for patients with and without KMT2C mutations in the Target-seq cohort. (C) Kaplan–Meier OS analysis for patients with and without KMT2C mutations in the WES cohort. (D) Cox regression analysis of KMT2C mutation and tumor-associated factors influencing OS in the combined WES and Target-seq patient cohorts.

Discussion

In this study, WES data from 36 GRC patients revealed recurrent mutations in epigenetic modifiers, which were investigated as predictors for clinical outcomes. The high frequency of *KMT2C* mutation by WES was re-identified in 25 GRC patients by Target-seq. GRC with *KMT2C* mutations presented higher densities of $CD3^+$ TILs, $CD8^+$ TILs, and $CD163^+$ tumor-associated macrophages, and higher expression of PD-L1.

Epigenetic mutations have been reported to affect numerous facets of cancer, including chromatin packaging, programs of distinct cellular gene expression, and signal transduction, incurring enhanced cancer growth, invasion, and metastasis [30,31]. Recent evidence suggests that epigenetic regulators can alter the immunological characteristics of tumors and make them susceptible to immune treatments [32]. In GAC, epigenetic changes are increasingly being recognized as crucial modulators of cancer development [33,34]. This study revealed the widespread mutational landscape of epigenetic modifiers in GRC, which provides a novel scenario for GRC pathogenesis and recommends potential candidates for therapeutic intervention.

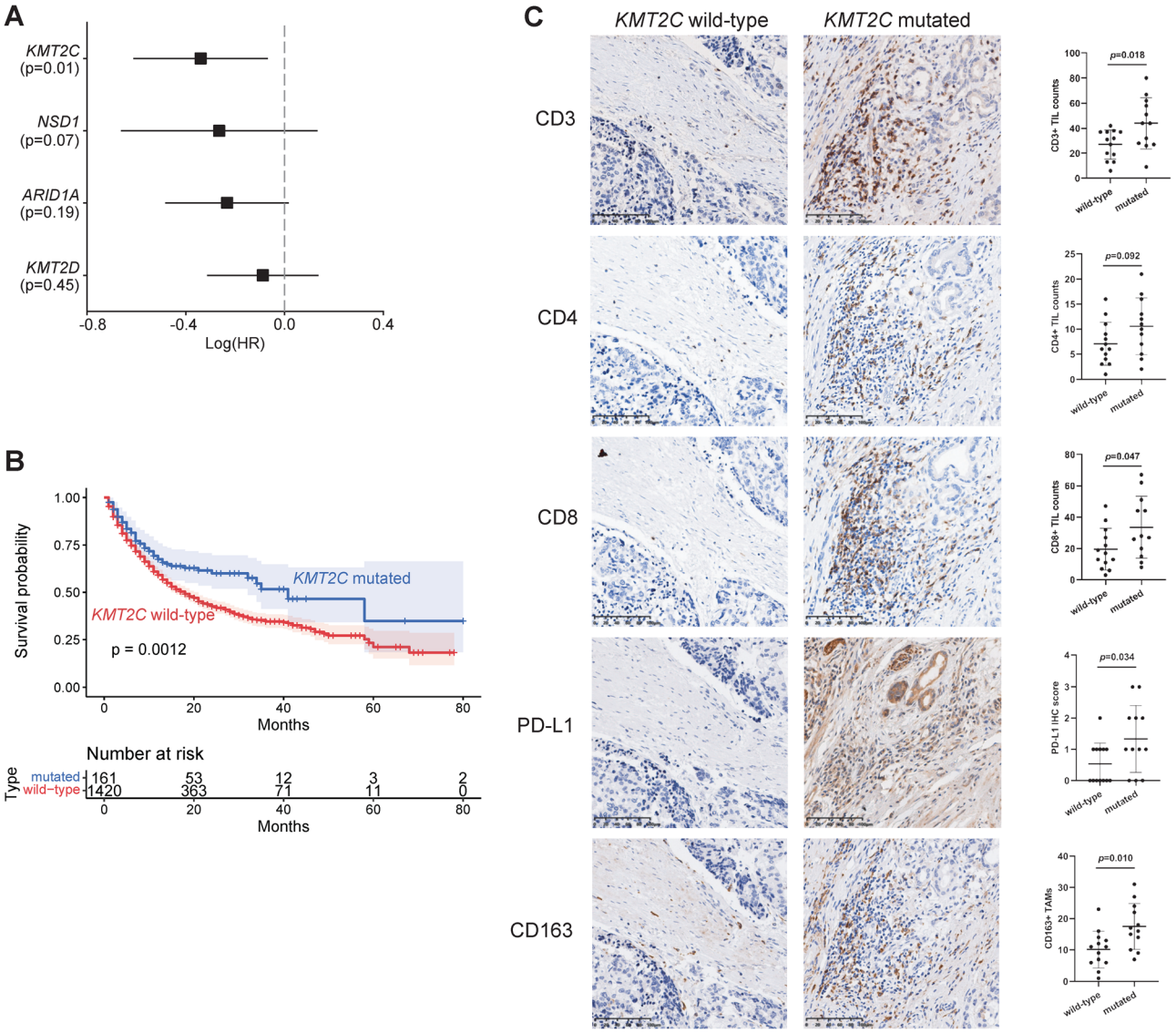

Figure 5. Association between *KMT2C* mutation status and the tumor immune microenvironment. (A) Forest plot depicting the OS benefit of mutations in different epigenetic modifiers for pan-cancer patients subjected to ICI treatment. (B) Kaplan–Meier OS curves of pan-cancer patients subjected to ICI treatment stratified according to the *KMT2C* mutation status. (C) Representative images of IHC analysis of GRC specimens and quantification of $CD3^+$, $CD4^+$, $CD8^+$, and $CD163^+$ cells, and PD-L1) immunostaining score according to the *KMT2C* mutation status. Data were analyzed using the *t* test.

Our results are consistent with previous research on cancer genomics, identifying *TP53* as a well-known tumor suppressor and frequently mutated gene in human cancer [35]. The observed similarities in *TP53* and *MUC16* mutation frequencies between GRC and other TCGA-STAD subtypes further support the relevance of these genes in GAC (supplementary material, Table S3). *PIK3CA* is a frequently mutated oncogene in various cancers [36]. Two (50%) of the four *PIK3CA* mutations in this study are classified as pathogenic in the ClinVar database (supplementary material, Figure S4). Increasing data suggest that *GNAS* mutation promotes carcinogenesis by activating the ERK1/2 MAPK pathway or the Wnt/−catenin pathway [37]. Previous studies have shown that *GNAS* mutations are frequent in the fundic gland type of GAC, but seldom occur in gastric antrum cancer [38]. Given that GRC occurs mainly in the proximal stomach, the distribution characteristics of *GNAS* mutations may heighten the frequency of *GNAS* mutations in GRC.

KMT2C mutations present different mutation ratios in different GAC subtypes, among which GRC shows a significantly higher mutation frequency. The high frequency of *KMT2C* mutations in GRC suggests that these variants act as molecular drivers. KMT2C possesses PHD fingers that serve as histone recognition and binding domains [39]. Missense mutations often disrupt the PHD finger clusters during carcinogenesis, modulating cell growth and differentiation pathways [40,41]. Our study found that *KMT2C* mutations in GRC mainly occurred in the PHD region, whereas *KMT2C* mutations in GAC are scattered throughout the gene, further suggesting a role of *KMT2C* in remnant stomach carcinogenesis. Moreover, *KMT2C* mutations could function as a potential predictor for OS in GRC and serve as a novel and promising predictive biomarker for ICI treatment in multiple solid tumors [28].

Environmental factors, particularly duodenogastric reflux, including bile reflux, play a major role in GRC development [42,43]. Duodenogastric reflux leads to the development of chronic inflammation in the remnant stomach, which may cause epigenetic mutations in genes such as *KMT2C*. *KMT2C* disruption might act as a founder or gatekeeper mutation in early tumor cells, resulting in changes in the epigenomic landscape that enable additional oncogenic modifications [44]. Mechanistically, it has been shown that KMT2C alteration promotes the epithelial–mesenchymal transition of GAC [45]. The pathogenic relationship between duodenogastric reflux and stump cancer would require more mechanistic studies and the exploration of animal models.

MSI refers to a hypermutable pattern of genomic instability, which results from defects and changes in the DNA MMR system [21]. As previously reported, the MSI percentage varies widely in patients with GAC (6.74–33.82%). Our data indicate that MSI may be nonessential to GRC due to the consistent MSS status for GRC patients. MSI was found to be an independent prognostic factor for disease–free survival and OS in GAC. The MSS status of GRC can provide valuable information regarding optimal treatment regimens.

SCNVs can cause changes in mRNA and protein expression [46]. Increasing evidence has supported that SCNVs are prevalent in GAC and can also have implications for the pathogenesis, prognosis, and therapeutic options [47]. This study identified the role of SCNVs in GRC and their association with *ARID1A* and *TP53* mutations. *TP53*-mutated GACs frequently display significant numbers of SCNVs that concern both focal gene regions and wide chromosomal regions [48], suggesting the essential role of *TP53* in maintaining genomic integrity. Additionally, we found that GRC samples with *ARID1A* mutations exhibited fewer SCNVs compared to those without *ARID1A* mutations (supplementary material, Figure S6C), consistent with recent reports suggesting that the loss of ARID1A contributes to chromosome stability [49]. This discovery provides new insights into the molecular mechanisms underlying GRC pathogenesis.

In the past few years, the relationship between epigenetic modifiers and the tumor immune microenvironment has received growing attention [50]. Generally, the $CD3^+$ TIL population is a favorable prognostic factor in cancer, while $CD163^+$ tumor-associated macrophages share similarities with M2 macrophages, which mediate immune suppression [51]. In most tumor types, the tumor-associated macrophage frequency and PD-L1 expression are associated with unfavorable survival [52]. In this study, frequent alterations in *KMT2C* were associated with high TIL numbers. This result suggests that the mechanism by which mutations in *KMT2C* induce GRC tumorigenesis and progression may involve alterations in the tumor immune microenvironment.

Our preliminary findings indicate an unusually high prevalence of *KMT2C* alterations in previous malignancies. Potential explanations encompass gene expression regulation since *KMT2C* mutations may result in dysregulated gene expression, rendering GAC cells more susceptible to growth in the residual gastric tissue [20]. Furthermore, *KMT2C* mutations might influence cell cycle regulation, DNA repair, and apoptosis pathways, further accelerating GRC development. *KMT2C* mutations could potentially heighten tumor heterogeneity, which refers to the genetic and phenotypic variations among tumor cells [53]. This heterogeneity may enable some tumor cells to develop resistance to surgical intervention and chemotherapy, leading to the survival and progression of GRC in the residual gastric tissue. Finally, *KMT2C* mutations might induce abnormalities in pro-inflammatory factors, growth factors, and other signaling molecules within the tumor microenvironment, fostering tumor development in the residual gastric tissue [53]. It is crucial to note that these explanations are speculative, grounded on the potential impact of *KMT2C* mutations on GAC progression to GRC. To substantiate these claims, further research and clinical samples are required.

This study has some limitations. GRC is a typical model of carcinogenesis; thus, this will help in understanding the development of carcinoma in the remnant stomach by comparing the sequencing information of the initial specimens with that of the current

specimens. However, since the initial surgery was done more than 5 years ago, it is difficult to obtain pathological tissues from the initial gastrectomy. Furthermore, the DNA quality of paraffin-embedded tissues may be compromised by long-term storage.

In summary, we performed comprehensive exome analysis and revealed the genetic uniqueness of GRC. To the best of our knowledge, this is the first study to identify and validate new recurrent mutations of epigenetic modifiers in GRC. Moreover, we propose that KMT2C contributes to GRC pathogenesis by affecting multiple factors including epigenetic modifiers and the tumor microenvironment. To conclude, these findings shall facilitate the design of a more accurate tumor profiling and therapy for GRC.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (81902436).

Author contributions statement

DX and ZS conceived and directed the study. JWL and BS provided the samples for this study. HJC performed DNA sequencing, analyzed the data and prepared the figures. BS performed the experiments and wrote the manuscript. CT and YZ carried out pathological assessment and interpretation. All authors have read and approved the final manuscript.

Data availability statement

The datasets generated during the current study are available from the corresponding author on reasonable request.

References

1. Takahashi K, Yoshikawa T, Morita S, *et al*. Different risks of nodal metastasis by tumor location in remnant gastric cancer after curative gastrectomy for gastric cancer. *Gastric Cancer* 2020; **23:** 195–201.
2. Shimada H, Fukagawa T, Haga Y, *et al*. Does remnant gastric cancer really differ from primary gastric cancer? A systematic review of the literature by the Task Force of Japanese Gastric Cancer Association. *Gastric Cancer* 2016; **19:** 339–349.
3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014; **513:** 202–209.
4. Cristescu R, Lee J, Nebozhyn M, *et al*. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 2015; **21:** 449–456.
5. Ge S, Xia X, Ding C, *et al*. A proteomic landscape of diffuse-type gastric cancer. *Nat Commun* 2018; **9:** 1012.
6. Mun DG, Bhin J, Kim S, *et al*. Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell* 2019; **35:** 111–124.e10.
7. Gullo I, Carvalho J, Martins D, *et al*. The transcriptomic landscape of gastric cancer: insights into Epstein-Barr virus infected and microsatellite unstable tumors. *Int J Mol Sci* 2018; **19:** 2079.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25:** 1754–1760.
9. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 2016; **32:** 292–294.
10. McKenna A, Hanna M, Banks E, *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20:** 1297–1303.
11. Cibulskis K, Lawrence MS, Carter SL, *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31:** 213–219.
12. Ye K, Schulz MH, Long Q, *et al*. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; **25:** 2865–2871.
13. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.
14. Rosenthal R, McGranahan N, Herrero J, *et al*. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016; **17:** 31.
15. Lee J, Lee AJ, Lee JK, *et al*. Mutalisk: a web-based somatic MUTation AnaLyIS toolKit for genomic, transcriptional and epigenomic signatures. *Nucleic Acids Res* 2018; **46:** W102–W108.
16. Niu B, Ye K, Zhang Q, *et al*. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 2014; **30:** 1015–1016.
17. Koboldt DC, Zhang Q, Larson DE, *et al*. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22:** 568–576.
18. Kuleshov MV, Jones MR, Rouillard AD, *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016; **44:** W90–W97.
19. Colaprico A, Silva TC, Olsen C, *et al*. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016; **44:** e71.
20. Cho SJ, Yoon C, Lee JH, *et al*. KMT2C mutations in diffuse-type gastric adenocarcinoma promote epithelial-to-mesenchymal transition. *Clin Cancer Res* 2018; **24:** 6556–6569.
21. Choi YY, Kim H, Shin SJ, *et al*. Microsatellite instability and programmed cell death-ligand 1 expression in stage II/III gastric

cancer: post hoc analysis of the CLASSIC randomized controlled study. *Ann Surg* 2019; **270:** 309–316.
22. Singh Nanda J, Kumar R, Raghava GP. dbEM: a database of epigenetic modifiers curated from cancerous and normal genomes. *Sci Rep* 2016; **6:** 19340.
23. Rampias T, Karagiannis D, Avgeris M, *et al*. The lysine-specific methyltransferase KMT2C/MLL3 regulates DNA repair components in cancer. *EMBO Rep* 2019; **20:** e46821.
24. Wang K, Yuen ST, Xu J, *et al*. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* 2014; **46:** 573–582.
25. Kakiuchi M, Nishizawa T, Ueda H, *et al*. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* 2014; **46:** 583–587.
26. Guo YA, Chang MM, Huang W, *et al*. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. *Nat Commun* 2018; **9:** 1520.
27. Chen K, Yang D, Li X, *et al*. Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci U S A* 2015; **112:** 1107–1112.
28. Zhang R, Wu HX, Xu M, *et al*. KMT2A/C mutations function as a potential predictive biomarker for immunotherapy in solid tumors. *Biomark Res* 2020; **8:** 71.
29. Cerami E, Gao J, Dogrusoz U, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; **2:** 401–404.
30. Cheng Y, He C, Wang M, *et al*. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct Target Ther* 2019; **4:** 62.
31. Nervi C, De Marinis E, Codacci-Pisanelli G. Epigenetic treatment of solid tumours: a review of clinical trials. *Clin Epigenetics* 2015; **7:** 127.
32. Bezu L, Chuang AW, Liu P, *et al*. Immunological effects of epigenetic modifiers. *Cancers* 2019; **11:** 1911.
33. Lee SW, Park DY, Kim MY, *et al*. Synergistic triad epistasis of epigenetic H3K27me modifier genes, EZH2, KDM6A, and KDM6B, in gastric cancer susceptibility. *Gastric Cancer* 2019; **22:** 640–644.
34. Zhang Y, Guo D. Epigenetic variation analysis leads to biomarker discovery in gastric adenocarcinoma. *Front Genet* 2020; **11:** 551787.
35. Guimaraes DP, Hainaut P. TP53: a key gene in human cancer. *Biochimie* 2002; **84:** 83–93.
36. Hao Y, Samuels Y, Li Q, *et al*. Oncogenic PIK3CA mutations reprogram glutamine metabolism in colorectal cancer. *Nat Commun* 2016; **7:** 11971.
37. Nomura R, Saito T, Mitomi H, *et al*. GNAS mutation as an alternative mechanism of activation of the Wnt/beta-catenin signaling pathway in gastric adenocarcinoma of the fundic gland type. *Hum Pathol* 2014; **45:** 2488–2496.
38. Kushima R, Sekine S, Matsubara A, *et al*. Gastric adenocarcinoma of the fundic gland type shares common genetic and phenotypic features with pyloric gland adenoma. *Pathol Int* 2013; **63:** 318–325.
39. Sze CC, Shilatifard A. MLL3/MLL4/COMPASS family on epigenetic regulation of enhancer function and cancer. *Cold Spring Harb Perspect Med* 2016; **6:** a026427.
40. Wang L, Zhao Z, Ozark PA, *et al*. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat Med* 2018; **24:** 758–769.
41. Rhee JK, Yoo J, Kim KR, *et al*. Identification of local clusters of mutation hotspots in cancer-related genes and their biological relevance. *IEEE/ACM Trans Comput Biol Bioinform* 2019; **16:** 1656–1662.
42. Kondo K. Duodenogastric reflux and gastric stump carcinoma. *Gastric Cancer* 2002; **5:** 16–22.
43. Mukaisho K, Miwa K, Kumagai H, *et al*. Gastric carcinogenesis by duodenal reflux through gut regenerative cell lineage. *Dig Dis Sci* 2003; **48:** 2153–2158.
44. Fagan RJ, Dingwall AK. COMPASS ascending: emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Lett* 2019; **458:** 56–65.
45. Lu J, Ding Y, Chen Y, *et al*. Whole-exome sequencing of alpha-fetoprotein producing gastric carcinoma reveals genomic profile and therapeutic targets. *Nat Commun* 2021; **12:** 3946.
46. Tan ES, Knepper TC, Wang X, *et al*. Copy number alterations as novel biomarkers and therapeutic targets in colorectal cancer. *Cancers* 2022; **14:** 2223.
47. Onoyama T, Ishikawa S, Isomoto H. Gastric cancer and genomics: review of literature. *J Gastroenterol* 2022; **57:** 505–516.
48. Wang X, Sun Q. TP53 mutations, expression and interaction networks in human cancers. *Oncotarget* 2017; **8:** 624–643.
49. Zhao B, Lin J, Rong L, *et al*. ARID1A promotes genomic stability through protecting telomere cohesion. *Nat Commun* 2019; **10:** 4067.
50. Aspeslagh S, Morel D, Soria JC, *et al*. Epigenetic modifiers as new immunomodulatory therapies in solid tumours. *Ann Oncol* 2018; **29:** 812–824.
51. Motomura T, Shirabe K, Mano Y, *et al*. Neutrophil-lymphocyte ratio reflects hepatocellular carcinoma recurrence after liver transplantation via inflammatory microenvironment. *J Hepatol* 2013; **58:** 58–64.
52. Sun L, Huang C, Zhu M, *et al*. Gastric cancer mesenchymal stem cells regulate PD-L1-CTCF enhancing cancer stem cell-like properties and tumorigenesis. *Theranostics* 2020; **10:** 11950–11962.
53. Xie K, Peng Y, Zhong W, *et al*. KMT2C is a potential biomarker of anti-PD-1 treatment response in metastatic melanoma. *Front Biosci* 2022; **27:** 103.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Flowchart of patient selection and experimental procedures

Figure S2. Analysis pipeline for identifying somatic variants within WES data

Figure S3. Quality control of mapped reads from whole-exome sequencing data

Figure S4. Mutation distribution and frequency in highly recurrent mutated genes

Figure S5. Association between mutated genes and GRC clinicopathological characteristics

Figure S6. Genomic stability of GRC samples

Table S1. Statistics of WES on genomic DNA of 36 paired samples in GRC

Table S2. Statistical table for genomic mutations identified from each GRC patient

Table S3. Comparison of gene mutation frequencies between GRC and TCGA-STAD

Table S4. Correlation of mutated genes with clinicopathological characteristics of GRC patients

Table S5. Using MSIsensor to derive the MSI status of each patient from the corresponding tumor-normal paired WES

Table S6. Significantly enriched pathways based on recurrently mutated genes

Table S7. Location and type of *KMT2*C mutations

Table S8. Targeted sequencing of *KMT2C* in another GRC cohort

Table S9. *KMT2C* mutation frequency in different GAC studies

Table S10. *KMT2C* mutation frequency in different GAC types