Original Article

Development of a streamlined NGS-based TCGA classification scheme for gastric cancer and its implications for personalized therapy

Pengda Guo[1#], Yang Yang[2#], Lu Wang[1], Yu Zhang[1], Bei Zhang[3], Jinping Cai[3], Fabrício Freire de Melo[4], Matthew R. Strickland[5], Min Huang[1,6], Biao Liu[1]

[1]Department of Pathology, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China; [2]Gusu School, Nanjing Medical University, Suzhou, China; [3]The Medical Department, 3D Medicines Inc., Shanghai, China; [4]Multidisciplinary Institute of Health, Federal University of Bahia, Vitória da Conquista, Brazil; [5]Department of Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, MA, USA; [6]General Practice Section, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China

Contributions: (I) Conception and design: P Guo, Y Yang, L Wang, Y Zhang; (II) Administrative support: M Huang, B Liu; (III) Provision of study materials or patients: P Guo, Y Yang, L Wang, Y Zhang; (IV) Collection and assembly of data: Y Zhang, B Zhang; (V) Data analysis and interpretation: B Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Min Huang, MD. Department of Pathology, The Affiliated Suzhou Hospital of Nanjing Medical University, 16 Baita West Road, Gusu District, Suzhou 215006, China; General Practice Section, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou 215006, China. Email: 13962177700@163.com; Biao Liu, MD. Department of Pathology, The Affiliated Suzhou Hospital of Nanjing Medical University, 16 Baita West Road, Gusu District, Suzhou 215006, China. Email: vivi_liu914@163.com.

Background: The Cancer Genome Atlas (TCGA) has identified four distinct molecular subtypes of gastric cancer (GC) with prognostic significance: Epstein-Barr virus (EBV)-positive, microsatellite instability (MSI)-high, genomically stable (GS), and chromosomal instability (CIN). Unfortunately, the complex analysis required for TCGA classification limits its practical use in clinical settings. Our study sought to devise a next-generation sequencing (NGS)-based method to classify GC more efficiently, serving as a promising biomarker for prognosis and immunotherapy efficacy.

Methods: This study was a retrospective observation study, and we employed 2 independent GC cohorts. The 3DMed cohort (n=765), comprising data on 733 cancer-related genes along with 4 EBV-encoded genes, was utilized to develop the new NGS classification. Additionally, the secondary Korean cohort (n=55), which includes both genomic data and information on immune checkpoint inhibitor (ICI) treatment, was employed to establish a correlation between NGS subtypes and ICI responsiveness.

Results: In the 3DMed cohort, we identified 5.2% EBV, 4.6% MSI, 30.6% GS, and 59.6% CIN subtypes. The MSI subtype exhibited the highest number of mutation events, along with the highest tumor mutational burden (TMB) and strong programmed cell death ligand 1 (PD-L1) expression. CIN tumors showed extensive copy number variations (CNVs) and genomic heterogeneity. The EBV subtype presented recurrent *ARID1A* and *PIK3CA* mutations and fewer *TP53* mutations. GS tumors exhibited specific mutations in *CDH1* and *ARID1A*. In the Korean cohort, ICIs were most effective in MSI and EBV cases, showing disease control rates of 100%, compared to 62.9% in GS and 12.5% in CIN subtypes.

Conclusions: The NGS method successfully maps the mutational landscape of GC, providing a practical TCGA classification surrogate to optimize patient-specific treatment strategies.

Keywords: Gastric cancer (GC); next-generation sequencing-based The Cancer Genome Atlas classification (NGS-based TCGA classification); immune checkpoint inhibitor treatment (ICI treatment)

Submitted May 11, 2024. Accepted for publication Aug 16, 2024. Published online Sep 13, 2024.
doi: 10.21037/jgo-24-345
View this article at: https://dx.doi.org/10.21037/jgo-24-345

Introduction

Gastric cancer (GC) is the fifth most common malignant tumor and the third leading cause of cancer mortality, particularly prevalent in Asia (1). The pathogenesis of GC is highly complex and influenced by genetic and epigenetic alterations of oncogenes, tumor suppressor genes and growth factors (2,3). As a highly heterogeneous disease, intratumor genetic heterogeneity (ITH) represents a large hurdle to GC patient treatment, impeding the uniform application of specific molecularly targeted agents. The presence of considerable heterogeneity has been shown through the categorization of subtypes based on histopathology and comprehensive molecular classification methods. The conventional classification systems of GC based on histopathology by the World Health Organization (WHO) and Laurén are, for the most part, not clinically actionable (4,5). Among various molecular classification methods, those proposed by The Cancer Genome Atlas (TCGA) and Asian Cancer Research Group (ACRG) have been the most important and recognized ones in GC (6,7). Based on large-scale integration of DNA, RNA, and proteins, TCGA groups GC into four different subtypes—Epstein-Barr virus (EBV), microsatellite instability (MSI), genomically stable (GS), and chromosomal instability (CIN)—each with distinct salient genomic features (6). The ACRG subtyping is largely complementary to the TCGA system for molecular classification of GC (7). There have been significant molecular and etiologic differences among GC subtypes revealed by the TCGA and ACRG data, as well as many potentially targetable genomic changes and promising biomarkers for prognosis and treatment effectiveness.

The TCGA cohort study reported that the patients with the 4 subtypes had different prognoses, where those with the EBV subtype had the best prognosis, and those with the GS subtype had the worst prognosis (8). Of note, this study also showed that, depending on GC subtype, patients benefited differently from adjuvant chemotherapy, whereas patients with the CIN subtype experienced the greatest benefit from adjuvant chemotherapy while, those with the GS subtype had the least benefit. In recent years, immunotherapy has achieved revolutionary progress in the treatment of malignant tumors, including GC. Response rates have been significantly associated with the TCGA subtypes, exemplified by a nationwide Korean cohort (n=55), where the objective response rates of 100%, 100%, 12%, and 5% were observed in EBV-positive, MSI, GS, and CIN tumors, respectively (9).

These reports presented the clinical utility of the TCGA classification scheme. However, since the TCGA classification scheme for GC was based on a highly complicated integrative analysis of multiple genomic, transcriptomic, epigenetic and proteomic data, it may not be feasible for clinical use due to the complexity of generating these data. Comprehensive genomic profiling data using next-generation sequencing (NGS) technologies offer an alternative solution by prioritizing distinct subtype features and clinically relevant biomarkers, which would be more advantageous for precise target enrichment, enhanced depth of coverage, and conserves samples.

In this study, we employed a 737-gene NGS panel to establish a streamlined, NGS-based classification system for GC in a large Chinese cohort and elucidate the molecular characteristics of each subtype. Furthermore, we evaluated the clinical benefits for patients across identified subtypes using the NGS-based TCGA classification method for immunotherapy in another independent cohort. We present this study in accordance with the STROBE reporting checklist (available at https://jgo.amegroups.com/article/view/10.21037/jgo-24-345/rc).

Highlight box

Key findings

- We mapped the mutational landscape of gastric cancer (GC) with the next-generation sequencing (NGS)-based method, providing a practical The Cancer Genome Atlas (TCGA) classification surrogate to optimize patient-specific treatment strategies.

What is known and what is new?

- TCGA has delineated four molecular subtypes of GC with significant prognostic implications. These classifications offer insights into the heterogeneity of GC and have been instrumental in guiding treatment strategies. However, the application of TCGA classification in clinical practice is limited due to the complexity.
- We introduce a novel NGS-based subtyping system that employs a 737-gene panel to replicate the TCGA molecular classifications in a more streamlined and effective manner. Furthermore, we observed varying responses to immune checkpoint inhibitors among the subtypes, highlighting the potential for personalized therapeutic approaches based on GC molecular subtypes.

What is the implication, and what should change now?

- This research provides a streamlined method for TCGA subtype classification that aids in the development of rational treatment options for GC patients, leading to more meaningful outcomes.

Methods

Samples and study design

The NGS-based TCGA classification scheme was developed in the 3DMed cohort, comprising 735 Chinese patients with GC whose tumor tissues underwent NGS using a 737-gene panel consisting of 733 cancer-related genes and 4 EBV-encoded genes between 6 January 2017 and 14 April 2020, at 3D Medicines (Shanghai, China). All the patients provided written informed consent for examination of their sample and use of their clinical data. Then, the efficacy of immune checkpoint inhibitors (ICIs) was assessed in the subtypes established by the NGS-based TCGA classification scheme. This evaluation was conducted using data from a previously published Korean cohort (PRJEB25780) comprising 55 patients diagnosed with metastatic GC (9). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Nanjing Medical University (No. 2023/279).

In the 3DMed cohort, the pathological diagnosis of the specimens was confirmed by two independent pathologists via review of hematoxylin and eosin (H&E)-stained tumor sections. A specimen was considered suitable for NGS analysis when the size ≥1 mm^3 and the percentage of tumor cells ≥20%. Demographic and clinicopathologic data were collected from medical records.

Molecular subtyping in 735 patients with GC

Following the TCGA classification method for GC, the EBV-positive subgroup identified by NGS was first separated from the 3DMed cohort, followed by high-frequency MSI cases. The remaining patients were classified into GS and CIN subtypes by clustering analysis on the basis of thresholded somatic copy number variation (CNVs) reoccurring analysis from segmented data using GISTIC 2.0 (6). Clustering was done in R based on Euclidean distance using Ward's method. CNVs were plotted by chromosomal location.

DNA preparation and NGS

DNA extraction and NGS analysis were conducted using the previously reported method in a Clinical Laboratory Improvement Amendments/College of American Pathologists (CLIA/CAP)-accredited laboratory (3D Medicines) (10-13). Briefly, genomic DNA was extracted and quantified using the ReliaPrep™ FFPE gDNA Miniprep System (Promega, Madison, WI, USA) and the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA), respectively. DNA fragment libraries with an average size of 250 bp were prepared using the KAPA Hyper Prep Kit (KAPA Biosystems, Wilmington, MA, USA) according to the manufacturer's protocol. Indexed libraries were subjected to hybridization with probes targeting four EBV-encoded genes (*EBNA-1*, *EBNA-2*, *EBNA-3*, and *BZLF1*) and 733 cancer-related genes (3D Medicines). The captured libraries were subsequently loaded onto a NovaSeq 6000 platform (Illumina, San Diego, CA, USA) for 100 bp paired-end sequencing with a mean sequencing depth of 1,000×. Using the Burrows-Wheeler Aligner (v0.7.12), raw data were aligned to hg19. Picard (v1.130) and SAMtools (v1.1.19) were used to filter out PCR duplicate reads and gather sequence metrics. Variant calling was performed only in the targeted regions. Somatic single nucleotide variants (SNVs) were detected using an in-house developed R package to execute a variant detection model based on binomial test, whereas indels were detected by local realignment. Variants were then filtered by their unique supporting read depth, strand bias, and base quality as previously described (11). A false positive filtering pipeline was then used to filter all variants based on an allele frequency of more than 1.0%, ensuring sensitivity and specificity. ANNOVAR was used to annotate single nucleotide polymorphisms (SNPs) and indels against dbSNP (v138), 1000Genome, and ESP6500 (population frequency greater than 0.015). Only missense, stopgain, frameshift, and non-frameshift indel mutations were kept. CNVs and gene rearrangements were detected as described previously (11).

Programmed cell death ligand 1 (PD-L1) expression detected by immunohistochemistry

The detection of PD-L1 expression via immunohistochemistry (IHC) followed a standard protocol (14). Briefly, formalin-fixed paraffin-embedded (FFPE) tissue sections were subjected to assessment of PD-L1 expression using the PD-L1 IHC 22C3 pharmDx assay (Agilent Technologies, Santa Clara, CA, USA). PD-L1 expression levels were ascertained through a Composite Positive Score (CPS), calculated by the ratio of cells positive for PD-L1—encompassing tumor cells, lymphocytes, and macrophages—to the total viable tumor cell count, followed by a scaling factor of 100. A CPS threshold of 1 delineated the samples deemed positive for PD-L1 expression.

MSI, tumor mutational burden (TMB), ITH, and homologous recombination deficiency (HRD) detection by NGS

EBV score was calculated as described previously (15), and the sample was considered EBV-positive if the EBV score was ≥0.05695. For MSI determination, the stability of 100 microsatellites was evaluated using an in-house developed R script. Any sample with at least 40% of unstable MSI loci was classified as MSI-high (also called MSI), and otherwise microsatellite stable (MSS) (16). TMB was defined as the number of somatic SNVs and insertions and deletions (indels) per megabase (Mb) in examined coding regions. SNVs included synonymous and non-synonymous mutations, stop gain/loss, as well as splicing variants. Indels can be both frameshift and non-frameshift. The ITH level was estimated by mutant-allele tumor heterogeneity (MATH), which is derived from the distribution of mutant-allele frequencies across loci unique to each tumor's mutations. MATH value is formulated by taking the median absolute deviation of these frequencies and dividing it by the median value of the mutant-allele fractions, then multiplying by 100 to scale the result: MATH = (100 × MAD) / median (17). HRD score was calculated as the sum of loss of heterozygosity (LOH), telomeric allelic imbalance (TAI), and large-scale state transition (LST) using an in-house algorithm (18).

Statistical analysis

Quantitative variables were expressed as mean ± standard deviation (SD) (quantitative variables with normal distribution) or median and interquartile range (IQR) (quantitative variables with non-normal distribution), and categorical variables were presented as frequency and percentage. The Fisher's exact test or the chi-square test was adopted to compare categorical variables, whereas either the Mann-Whitney *U*-test or Student's *t*-test was used for the comparison of two continuous variables. Differences between multiple groups were tested with Kruskal-Wallis nonparametric analysis of variance with multiple group comparisons. All statistical analyses were performed using R (version 4.3; The R Foundation for Statistical Computing, Vienna, Austria), and differences with P<0.05 were considered statistically significant. Because this was a descriptive study, no formal sample size calculation was performed and the sample size was based on the availability of samples. We handled missing data as missing without data imputation.

Results

Molecular classification of Chinese GC patients

A total of 735 Chinese GC patients were clustered into four TCGA subtypes based on the somatic genomic profiling from the 737-gene panel (*Figure 1*). Distribution and baseline characteristics of the four molecular subtypes are summarized in *Table 1*. Among the 735 patients, 5.2% (38/735) of the patients were identified as EBV subtype, 4.6% (34/735) as MSI subtype, 30.6% (225/735) as GS subtype, and 59.6% (438/735) as CIN subtype. Consistent with TCGA data (6), CIN was the largest among the four subtypes. GS tumors were diagnosed at an earlier age (median age 57.0 years) relative to patients with the other subtypes, whereas the MSI tumors were diagnosed at relatively older ages (median age 66.0 years) (*Figure 2A*), which is highly consistent with reports from TCGA and previous studies (6,8). Patients of all four groups showed male predominance (94.7%, 79.4%, 71.1%, and 82.6%; P<0.001) (*Figure 2B*).

Characteristics of somatic genomic alterations in the four subtypes

Based on the targeted gene sequencing and copy number profiles generated from the 3DMed cohort, we identified the somatic changes in the overall GC population and each GC subtype. In the overall GC population (n=735), altered genes were ranked based on the alteration frequency, and genes with an alteration frequency of >5% were listed (*Figure 1*). The top five most frequently altered genes were *TP53* (61.5%), *ARID1A* (18.2%), *LRP1B* (17.0%), *ERBB2* (16.2%), and *CDH1* (15.1%).

In terms of SNV/indel mutations, MSI subtype was hypermutated, with recurrently mutated genes including *KMT2C* (29/34, 85.3%), *KMT2D* (29/34, 85.3%), *TGFBR2* (27/34, 79.4%), *ACVR2A* (27/34, 79.4%), and *ARID1A* (25/34, 73.5%), *MSH3* (23/34, 67.6%), *RNF43* (22/34, 64.7%), *ZFHX3* (19/34, 55.9%), and *LRP1B* (18/34, 52.9%) (Table S1). The mutation frequency of the genes detected in the 701 non-MSI (i.e., MSS) tumors, was relatively low, less than 20%, except for that of *TP53* (62.6%). Among the MSS groups, the CIN subtype had the highest number of mutation events. In addition to the highest prevalence of *TP53* mutations (322/438, 73.5%), CIN tumors had frequent mutations in *LRP1B* (79/438, 18%), *FAT4* (51/438, 11.6%), *SPTA1* (46/438, 10.5%), *APC* (36/438, 8.2%), *GLI3* (34/438, 7.8%), *PTPRT* (29/438, 6.6%), *CTNND2* (28/438,

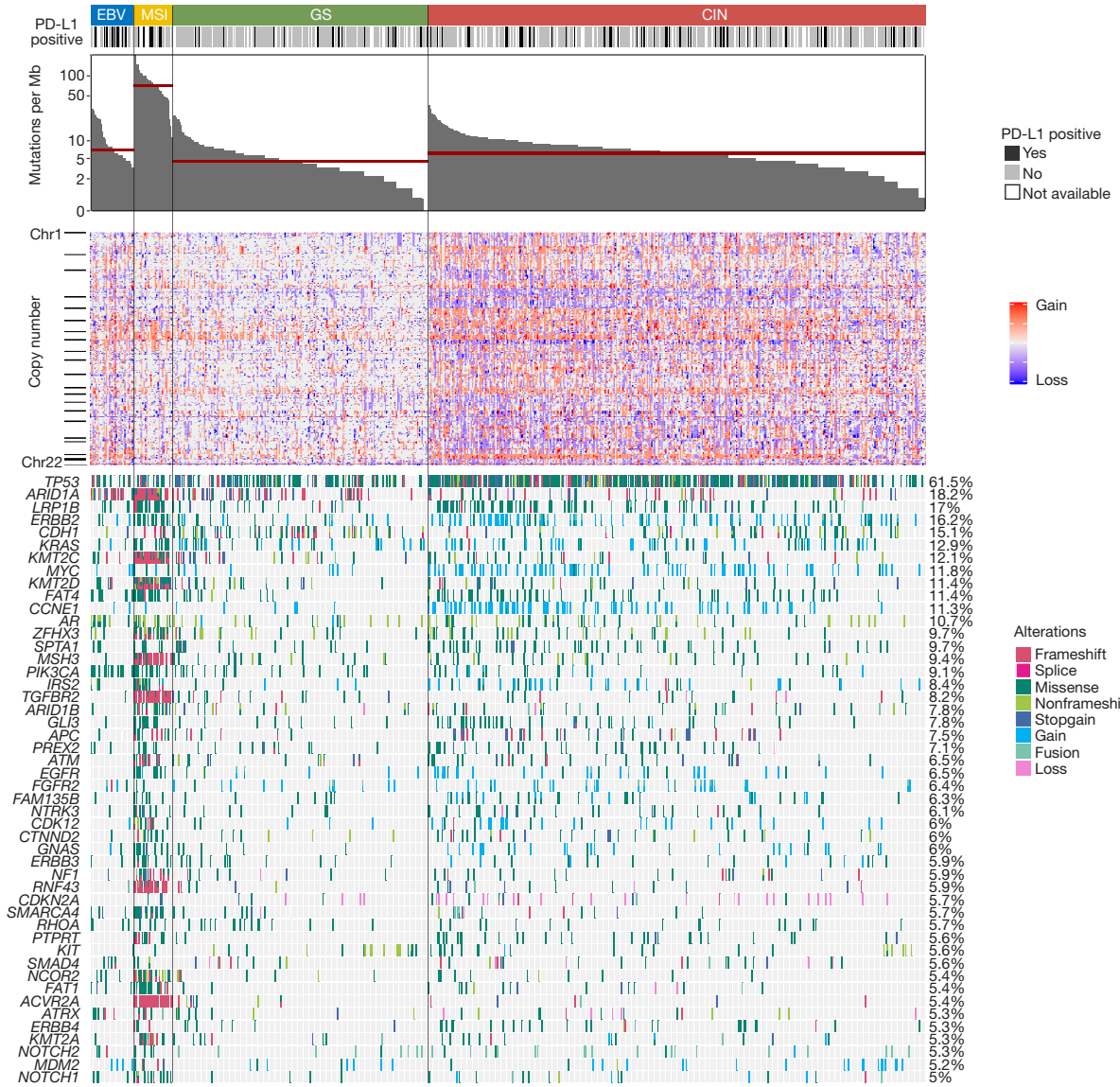


Figure 1 Gastric cancer molecular subtyping. Gastric cancer cases are divided into subtypes: EBV-positive, MSI, GS, and CIN. The molecular data from 735 tumors profiled with an NGS 737-gene panel are depicted. The red line represents the median TMB of each subgroup. EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability; NGS, next-generation sequencing; TMB, tumor mutational burden; PD-L1, programmed cell death ligand 1.

6.4%), *ERBB4* (27/438, 6.2%), and *NTRK3* (27/438, 6.2%). Within the EBV-positive tumors, we observed recurrent *ARID1A* (16/38, 42.1%) and *PIK3CA* mutations (15/38, 39.5%) and fewer *TP53* mutations (12/38, 31.6%). Consistent with TCGA data, the GS subtype exhibited the strong predilection of *CDH1* (53/225, 23.6%) and *ARID1A* mutations (45/225, 20%).

Subsequently, we conducted an analysis of the copy number profiles which revealed a significantly higher frequency of CNV events in the CIN subtype compared to the GS subtype (*Figure 3*). Specifically, the CIN subtype exhibited recurrent amplifications in cell cycle pathway genes, including *CCNE1* (73/438, 16.7%), *MYC* (69/438, 15.8%), and RTKS pathway genes, such as *ERBB2* (66/438, 15.1%), *KRAS* (32/438, 7.3%), *FGFR2* (26/438, 5.9%), and recurrent deletions in *CDKN2B* (5.5%) as TCGA

Table 1 The baseline characteristics of patients with gastric cancer in the 3DMed cohort (n=735)

Characteristic	EBV (N=38)	MSI (N=34)	GS (N=225)	CIN (N=438)	P value
Sex, n (%)					<0.001
Female	2 (5.3)	7 (20.6)	65 (28.9)	76 (17.4)	
Male	36 (94.7)	27 (79.4)	160 (71.1)	362 (82.6)	
Age at diagnosis, n (%)					<0.001
≤60 years	12 (52.2)	7 (35.0)	102 (66.2)	92 (41.3)	
>60 years	11 (47.8)	13 (65.0)	52 (33.8)	131 (58.7)	
Not available	15	14	71	215	
TMB, n (%)					<0.001
Median [min, max]	6.98 [3.35, 31.8]	71.8 [3.91, 207]	4.47 [0, 25.1]	6.15 [0, 35.8]	
PD-L1, n (%)					<0.001
Positive	10 (37.0)	9 (36.0)	18 (9.7)	53 (15.3)	
Negative	17 (63.0)	16 (64.0)	168 (90.3)	294 (84.7)	
Not available	11	9	39	91	

EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability; TMB, tumor mutational burden; PD-L1, programmed cell death ligand 1.

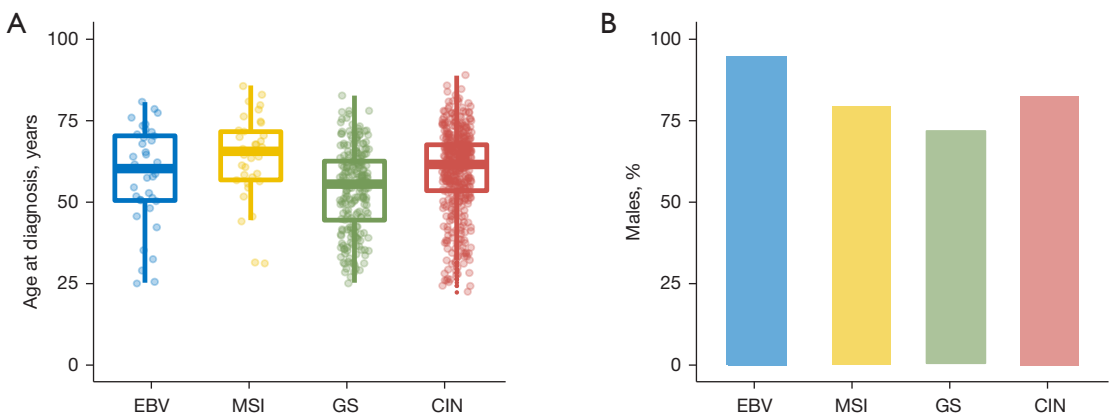
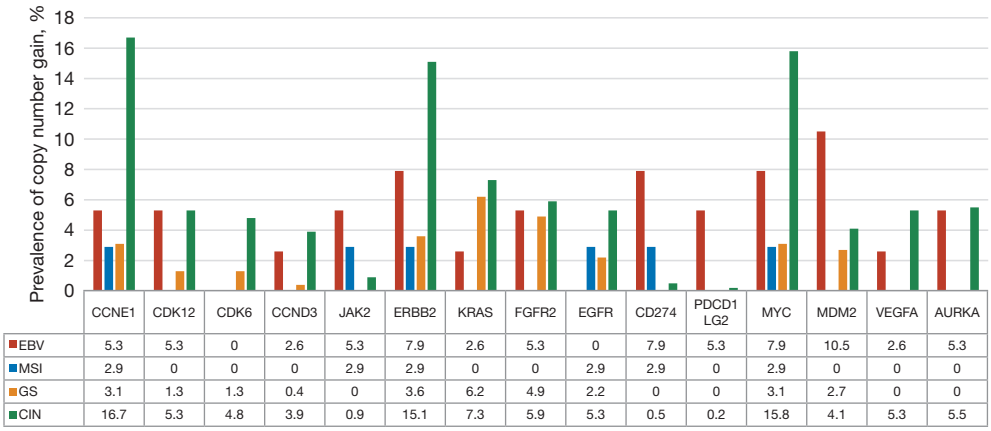


Figure 2 Distribution of sex and age at initial diagnosis in the 4 subtypes of the 3DMed cohort (n=735). The distribution of patient age at initial diagnosis (A) and the proportion of male patients (B) in the four subtypes of the 3DMed cohort. EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability.

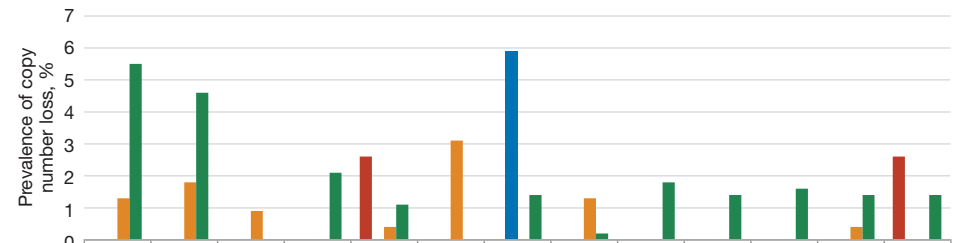
data. Conversely, in the GS subtype, only the amplification frequency of *KRAS* (14/225, 6.2%) exceeded 5%. Additionally, the EBV subtype also demonstrated a high prevalence of amplifications in *MDM2* (10.5%), *ERBB2* (7.9%), *CD274* (7.9%), *MYC* (7.9%), and *PDCD1LG2* (5.3%), *JAK2* (5.3%) and deletion in *PTEN* (2.6%), which is highly consistent with TCGA data. whereas the MSI type displayed a low-level gene amplification event but a recurrent *PMS2* deletion with a high prevalence of (5.9%).

Integrated pathway analysis

We integrated somatic CNVs and mutation data to comprehensively characterize genomic alterations in 10



	CCNE1	CDK12	CDK6	CCND3	JAK2	ERBB2	KRAS	FGFR2	EGFR	CD274	PDCD1 LG2	MYC	MDM2	VEGFA	AURKA
EBV	5.3	5.3	0	2.6	5.3	7.9	2.6	5.3	0	7.9	5.3	7.9	10.5	2.6	5.3
MSI	2.9	0	0	0	2.9	2.9	0	0	2.9	2.9	0	2.9	0	0	0
GS	3.1	1.3	1.3	0.4	0	3.6	6.2	4.9	2.2	0	0	3.1	2.7	0	0
CIN	16.7	5.3	4.8	3.9	0.9	15.1	7.3	5.9	5.3	0.5	0.2	15.8	4.1	5.3	5.5



	CDKN2B	CDKN2A	CDKN2C	AMER1	PTEN	NF2	PMS2	BRCA2	KDM6A	KDM5C	BCOR	GATA1	STAG2
EBV	0	0	0	0	2.6	0	0	0	0	0	0	0	2.6
MSI	0	0	0	0	0	0	5.9	0	0	0	0	0	0
GS	1.3	1.8	0.9	0	0.4	3.1	0	1.3	0	0	0	0.4	0
CIN	5.5	4.6	0	2.1	1.1	0	1.4	0.2	1.8	1.4	1.6	1.4	1.4

Figure 3 The frequently amplified and deleted genes in the 4 subtypes of the 3DMed cohort (n=735). EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability.

well-known signaling pathways (*Figure 4*). Among the four subtypes, the MSI subtype exhibited the highest alteration incidence in 9 out of 10 signaling pathways, including cell cycle, Hippo, MYC, NOTCH, NRF2, PI3K, RTK-RAS, TGF-β, and WNT signaling pathways. The CIN subtype displayed the highest alteration frequency in the TP53 pathway.

Exclusivity and co-occurrence of alterations

Additionally, we performed mutual exclusivity and co-occurrence analysis in the CIN and GS subtypes but not the EBV or MSI subtype due to the limited sample size. Details of alteration interactions in the top 25 highly altered genes are shown in *Figure 5*. In the CIN subtype, co-alterations were observed in *CCNE1* and *CDK12/PTK6/CEBPA/ZFHX3/MYC*, whereas mutual exclusivity was noted in *CDH1* and *PREX2/CDKN2A/GLI3*. In the GS subtype, *TP53* alterations were found to be mutually exclusive with *BAP1/ATM/TGFBR2/KIT*, and *ARID1A* alterations were mutually exclusive with *ERBB2/AR/KRAS* alterations.

Levels of TMB, PD-L1 expression, MATH, and HRD in the four subtypes of Chinese GCs

In addition to well-documented indicators of an active response to ICIs, such as PD-L1 expression, DNA mismatch repair deficiency or MSI, and TMB, ITH and HRD are now being extensively explored as potential biomarkers of ICI efficacy. We characterized these biomarkers in four molecular subtypes to predict the potential benefits from ICI efficacy. The MSI subtype exhibited the highest median TMB (71.8 muts/Mb), followed by the EBV subtype (6.25 muts/Mb), the CIN group (5.59 muts/Mb), and the GS group (4.47 muts/Mb) (P<0.001) (*Figure 6A*). A substantial proportion of the MSI (36.0%) and EBV-positive (37.0%) types showed positive PD-L1 expression (P<0.001) (*Figure 6B*). The CIN subtype exhibited the highest MATH level and HRD score among the four molecular subclasses (P<0.001) (*Figure 6C*,*6D*).

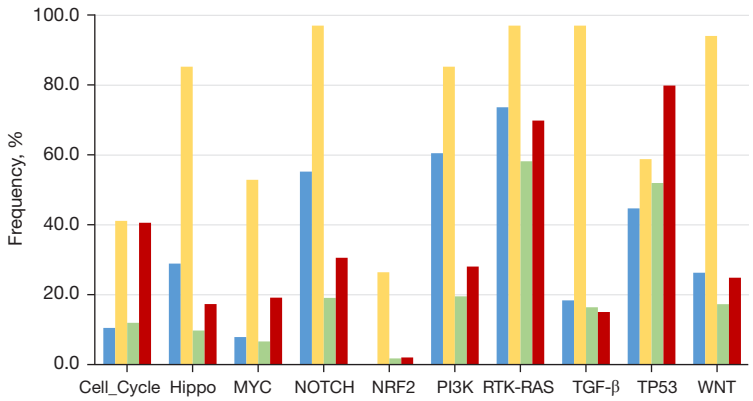


Figure 4 The alteration frequency of 10 well-known signaling pathways in the 4 subtypes of the 3DMed cohort (n=735). The colors represent different molecular subtypes: blue for EBV, yellow for MSI, green for GS, and red for CIN. EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability.

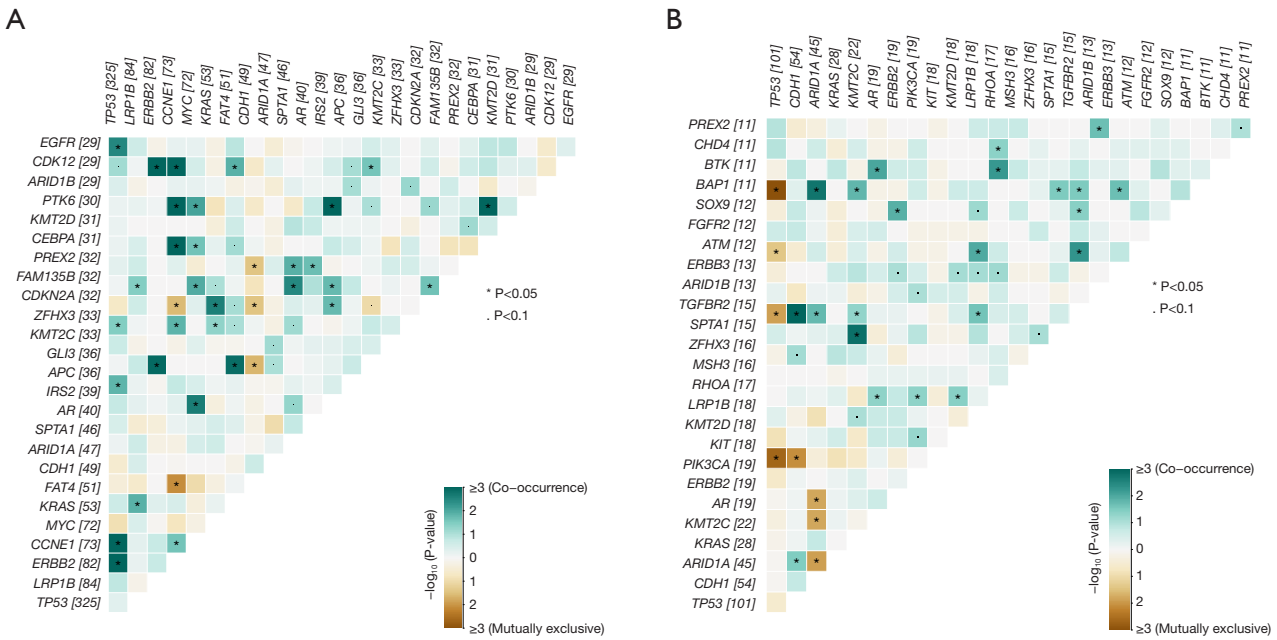


Figure 5 The mutually co-occurring and exclusive mutations of the top 25 frequently mutated genes. The color and symbol in each cell represent the statistical significance of the exclusivity or co-occurrence for each pair of genes. (A) The CIN subtype of the 3DMed cohort. (B) The GS subtype of the 3DMed cohort. CIN, chromosomal instability; GS, genomically stable.

ICI efficacy of the four subtypes of the Korean GCs

Next, we studied another independent ICI treatment cohort (n=55) derived from a published study (9), which evaluated the efficacy of pembrolizumab as a second-line and third-line treatment in the four subtypes of metastatic GC classified according to the whole-exome sequencing-based TCGA classification scheme. Among these patients, the disease control rate (DCR; stable disease + partial response + complete response, per the RECIST 1.1 guidelines) was 100% for the MSI (n=6) subtype, 100% for the EBV (n=4) subtype, 55.0% for the CIN (n=20) subtype, and 48.0% for the GS (n=25) subtype. While the patient population is small, combining the previous data, it is more likely

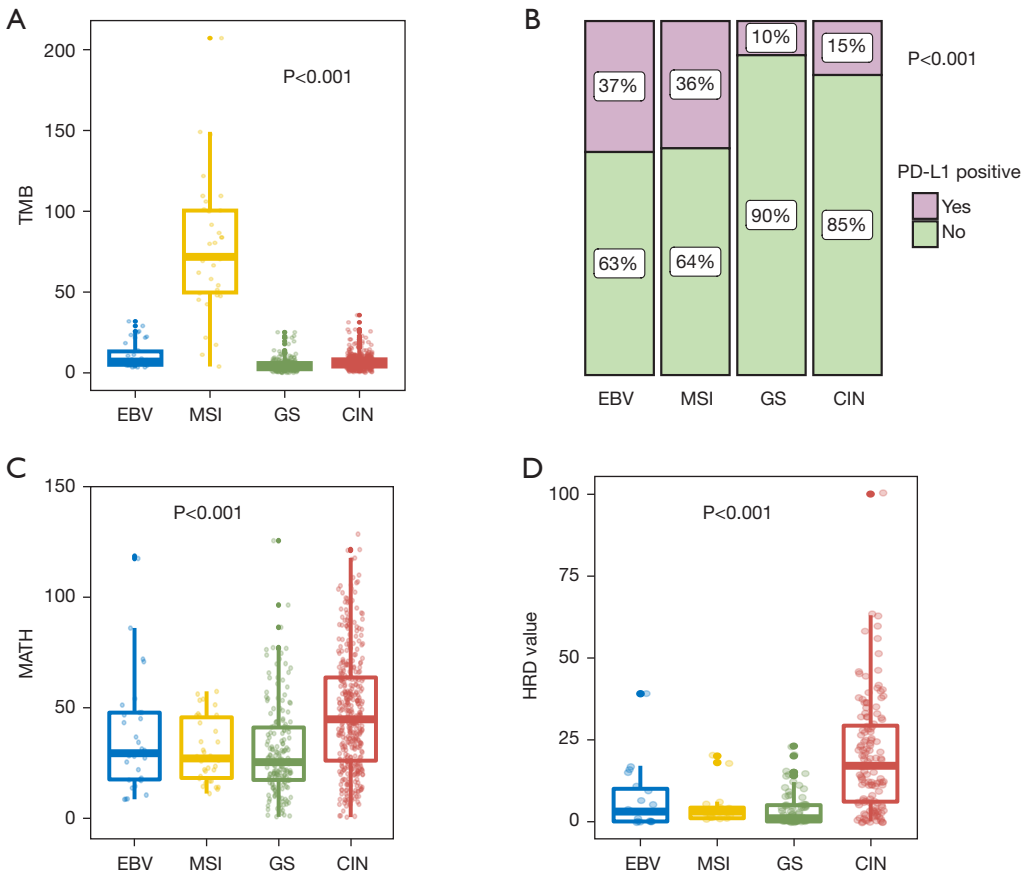


Figure 6 Levels of TMB (A), PD-L1 expression (B), MATH (C), and HRD (D) in the 4 subtypes of the 3DMed cohort (n=735). TMB, tumor mutational burden; PD-L1, programmed cell death ligand 1; MATH, mutant-allele tumor heterogeneity; HRD, homologous recombination deficiency; EBV, Epstein-Barr virus; MSI, microsatellite instability; GS, genomically stable; CIN, chromosomal instability.

that patients with the MSI and EBV subtypes will benefit from immunotherapy. However, the question remains whether immunotherapy can be considered for GS and CIN subtypes, as the immunotherapy benefits are similar (P=0.864), at approximately half (Table S2). Here, we used the NGS-based TCGA classification scheme to subtype these patients: four patients were classified as the EBV subtype, six as the MSI subtype, 35 as the GS subtype, and 8 as the CIN subtype, and their DCR were 100%, 100%, 62.9%, and 12.5%, respectively (*Figure* 7). There was a significant difference in DCR between the GS and CIN subtypes (P=0.01). These results indicate that the GS subtype identified through our method might potentially benefit from immunotherapy. In contrast, the CIN subtype is unlikely to derive benefit from immunotherapy, and therefore may be considered with caution when formulating treatment plans.

Discussion

In this study, our subtype classification drew from the TCGA classification scheme and was simplified, first achieving the molecular subtyping which could be performed in a single NGS test. This NGS-based classification method was rapid, easy-to-use, and also conserved tissue samples. As this NGS-based classification method accurately reflected the distinct genomic characteristics associated with TCGA each subtype, it holds promising potential for guiding rational therapy recommendations. Furthermore, we examined the correlation between ICI efficacy and molecular subtypes, finding that patients with the GS subtype derived less benefit from ICIs compared to those with the EBV and MSI subtypes, but more benefit than those with the CIN subtype.

GC is a highly heterogeneous disease, and ITH poses

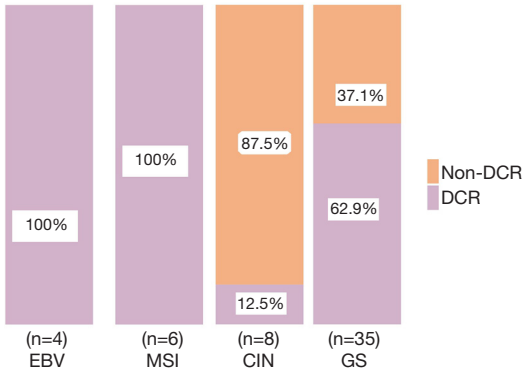


Figure 7 The DCR of the 4 subtypes identified by the NGS-based method from immune checkpoint inhibitors in the Korean cohort (n=55). DCR, disease control rate; EBV, Epstein-Barr virus; MSI, microsatellite instability; CIN, chromosomal instability; GS, genomically stable; NGS, next-generation sequencing.

a significant challenge to the treatment of GC patients. Recent genomic and molecular characterization of GC, especially the findings reported by TCGA through a comprehensive analysis of six distinct data types [somatic mutations, CNVs, CpG methylation, messenger RNA (mRNA), microRNA (miRNA), and protein expression], have shed light on the ITH and potential targeted therapeutics for four different subtypes. These four genomic subtypes identified by TCGA present distinct salient genomic features, which brings the hope of developing personalized treatment strategies. In practical application, EBER-ISH may produce false-negative outcomes as a result of RNA degradation, and false-positive outcomes stemming from background hybridization (19). Nevertheless, the application of the TCGA classification method in clinical practice is hindered by the substantial complexity associated with data generation. In this study, we developed an NGS-based subtyping system utilizing four EBV-encoded genes and tumor genomic alterations in the 3DMed cohort, which presents a simplified approach for replicating the established TCGA molecular classifications, improving diagnostic accuracy for EBV-related GC and streamlining analysis for TCGA classification with less tissue sample required, making it more feasible for regular use in clinical settings.

The clinical characteristics of each subtype, as classified by our NGS-based subtyping system aligned well with the specific tumor subtypes defined by TCGA. CIN emerged as the most prevalent subtype, GS tumors were diagnosed at a younger age, and MSI tumors at an older age, with a consistent male predominance across all subtypes, particularly in the EBV subtype. These findings demonstrate the reproducibility of TCGA tumor classification and the feasibility of our classification method. Certainly, due to the differences in the patient population (China in this study *vs.* USA and Western Europe in TCGA), tumor stage (predominantly advanced stage *vs.* predominantly early stage), and technological platforms (NGS *vs.* six distinct molecular platforms), the distribution and genomic alterations of the four subtypes between Chinese and TCGA cohorts were not completely consistent. It was noted that the occurrence of EBV and MSI subtypes was notably lower than in the TCGA cohort (5.2% *vs.* 9% and 4.6% *vs.* 22%, respectively), aligning with previous studies indicating that the incidence of EBV infection was approximately 5.1% among Chinese GC patients, and MSI occurred more frequently in early stage cancers (20,21).

Similar to TCGA data, the MSI subtype exhibited the highest mutation rates and the CIN subtype displayed the highest copy number alterations (CNAs) among the four subtypes (6). In line with previous studies, patients with MSI or EBV subtypes in the Korean cohort demonstrated the greatest benefits from ICIs (9,15). The underlying mechanisms through which the MSI subtype benefits from ICIs have been extensively documented, primarily involving a high TMB that generates numerous neoantigens, eliciting robust anti-tumor immune responses. Consistently, we also observed that the MSI subtype had the highest TMB among these four subtypes. The mechanisms by which the EBV subtype benefits from ICIs is currently an ongoing subject of debate. Consistent with TCGA data, we found that the EBV subtype was characterized by a high prevalence of *ARID1A* and *PIK3CA* mutations, as well as recurrent *MDM2*, *ERBB2*, *CD274*, *MYC*, *JAK2* and *PDCD1LG2* amplifications and *PTEN* loss. These altered genes were predominantly enriched in the PI3K-AKT-mTOR and chromatin remodeling pathways, all of which are associated with enhanced antitumor immunity or sensitivity to ICIs in solid tumors (22-25). Besides, *CD274* and *PDCD1LG2* encode PD-L1 and PD-L2, immunosuppressive proteins having been well-established targets to augment anti-tumor immune response. These findings might provide a mechanistic explanation for the improved benefit of EBV subtypes from ICIs. Additionally, several preclinical and clinical studies indicated that the inhibitors for MDM2 and MYC can work synergistically with programmed cell death protein 1 (PD-1)/PD-L1 blockade to enhance anti-tumor

efficacy in multiple tumors (26-32), suggesting the potential for testing PD-1/PD-L1 antibodies in combination with MDM2 and MYC inhibitors in this subtype.

The GS subtype was enriched with mutations in *CDH1* (23.6%) as TCGA data, encoding the cell adhesion molecule E-cadherin. *CDH1* mutations have been well-documented as independent prognostic markers for poor prognosis in diffuse-type GC. Given that the most patients with the GS subtype correspond to the diffuse type (33), the prognostic stratification based on the status of *CDH1* mutations could be considered in the GS subtype. Additionally, a recent study reported that *CDH1* mutations were significantly associated with primary resistance to immunotherapy, which might account for the limited benefit of the GS subtype from ICIs (34).

The CIN subtype exhibited the highest frequency of loss or increase of genes in critical areas, and thus has a significant impact on tumor growth and progression (35). A high prevalence of *TP53* mutations and strong ITH in CIN are well-documented as being associated with poor ICI efficacy (36-39), possibly accounting for the limited benefit of this subtype from ICIs. Additionally, the CIN subtype had the highest HRD score. Many studies investigating the correlation between HRD score and the efficacy of ICIs have yielded inconsistent conclusions (40-42). Conversely, HRD has been established as a positive biomarker of platinum therapy in breast and ovarian cancer (43-45). The REAL3 trial (n=74) also showed that esophagogastric cancer patients with the high HRD score experienced prolonged survival when treated with platinum chemotherapy (46). Furthermore, the CIN subtype has been reported to obtain the greatest benefit from adjuvant chemotherapy among these four subtypes (8). These findings suggest that compared with the immunotherapy, the CIN subtype is more likely to benefit from chemotherapy, especially platinum-based chemotherapy. Notably, *LRP1B* and *FAT4* mutations were found for the first time to be enriched and co-occurred in the CIN subtype. *LRP1B* encodes a putative tumor suppressor, the loss of which has been associated with aggressive clinicopathologic features and a poor prognosis in diverse solid cancers, including GC (47-49). *FAT4*, a member of the Fat family, functions as a tumor suppressor in GC by modulating the Wnt/β-catenin signaling pathway (50). A recent study reported comutant *LRP1B* and *FAT* as a positive biomarker for immune response in non-small cell lung cancer. It might be of interest to investigate whether the CIN subtype with *LRP1B* and *FAT* comutation could benefit from ICIs. Additionally, treatment targeting key driver genes copy number gain such as *EGFR*, *HER2*, and *VEGFA* in TRK-RAS pathway and *CCNE1*, *CDK6*, and *AURKA* in cell cycle pathway could been proposed for CIN subtype (51).

Immunotherapies and targeted therapies for GC with HER2 overexpression or amplification, PD-L1 expression, MSI or dMMR status, TMB, *NTRK* fusion, *RET* fusion, and *BRAF* V600E mutations, etc. were recommended by the National Comprehensive Cancer Network (NCCN) Guidelines for GC (52). Additionally, therapies targeting FGFR2b, c-MET, and claudin18.2 are emerging (53-55). Our NGS panel allows for comprehensive and simultaneous examination of various types of alterations in 733 cancer-related genes, covering key clinically relevant biomarkers. By utilizing automated clinical interpretations from genetic annotation databases to shorten the detection period, it could guide clinical treatment decisions and prognosis evaluation.

This study was limited by its retrospective design and small sample size. The relatively modest sample size of the Korean treatment cohort represents the main limitation of our study, which may have decreased the statistical power in ICI efficacy differences among the four subtypes. Nevertheless, our finding that the CIN subtype was associated with the least benefit from ICI, the GS subtype with moderate benefit, and the EBV and MSI subtypes with the greatest benefits is in line with recent reports (9,15). The mechanisms of resistance to ICI in CIN and GS groups are not fully investigated, it is essential to explore them further to improve therapy outcomes in GC. Furthermore, the small number of patients included in the subtypes limited the biomarker analysis of ICI efficacy for each subtype. Further studies with a large sample size are warranted to fully assess the ICI efficacy of these four subtypes grouped by the NGS-based TCGA classification scheme and identify the promising biomarkers of prognosis and treatment efficacy for each subtype.

Conclusions

In summary, we provide a streamlined approach for TCGA subtype classification that could guide rational treatment options for patients living with GC, leading to meaningful outcomes.

Acknowledgments

The authors would like to acknowledge the patients who participated in this study.
Funding: This study was supported by Suzhou Gusu Health Talent Plan Project Fund (No. GSWS2021037).

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://jgo.amegroups.com/article/view/10.21037/jgo-24-345/rc

Data Sharing Statement: Available at https://jgo.amegroups.com/article/view/10.21037/jgo-24-345/dss

Peer Review File: Available at https://jgo.amegroups.com/article/view/10.21037/jgo-24-345/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jgo.amegroups.com/article/view/10.21037/jgo-24-345/coif). B.Z. and J.C. are from 3D Medicines Inc. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Nanjing Medical University (No. 2023/279). All the patients provided written informed consent for examination of their sample and use of their clinical data.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.
2. Otani K, Li X, Arakawa T, et al. Epigenetic-mediated tumor suppressor genes as diagnostic or prognostic biomarkers in gastric cancer. Expert Rev Mol Diagn 2013;13:445-55.
3. Tan P, Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology 2015;149:1153-1162.e3.
4. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand 1965;64:31-49.
5. Bosman FT, Carneiro F, Hruban RH, et al. WHO classification of tumours of the digestive system: World Health Organization 2010.
6. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202-9.
7. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449-56.
8. Sohn BH, Hwang JE, Jang HJ, et al. Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. Clin Cancer Res 2017;23:4441-9.
9. Kim ST, Cristescu R, Bass AJ, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat Med 2018;24:1449-58.
10. Wang Z, Duan J, Cai S, et al. Assessment of Blood Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Patients With Non-Small Cell Lung Cancer With Use of a Next-Generation Sequencing Cancer Gene Panel. JAMA Oncol 2019;5:696-702.
11. Su D, Zhang D, Chen K, et al. High performance of targeted next generation sequencing on variance detection in clinical tumor specimens in comparison with current conventional methods. J Exp Clin Cancer Res 2017;36:121.
12. Yang N, Li Y, Liu Z, et al. The characteristics of ctDNA reveal the high complexity in matching the corresponding tumor tissues. BMC Cancer 2018;18:319.
13. Peng H, Lu L, Zhou Z, et al. CNV Detection from Circulating Tumor DNA in Late Stage Non-Small Cell Lung Cancer Patients. Genes (Basel) 2019;10:926.
14. Xie T, Liu Y, Zhang Z, et al. Positive Status of Epstein-Barr Virus as a Biomarker for Gastric Cancer Immunotherapy: A Prospective Observational Study. J

Immunother 2020;43:139-44.
15. Bai Y, Xie T, Wang Z, et al. Efficacy and predictive biomarkers of immunotherapy in Epstein-Barr virus-associated gastric cancer. J Immunother Cancer 2022;10:e004080.
16. Li J, Wang Z, Yang B, et al. Abstract 5477: A novel loci-selected method for microsatellite instability detection validated on a large Chinese cohort. Cancer Res 2020;80:5477.
17. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol 2013;49:211-5.
18. Chen Y, Wang X, Du F, et al. Association between homologous recombination deficiency and outcomes with platinum and platinum-free chemotherapy in patients with triple-negative breast cancer. Cancer Biol Med 2023;20:155-68.
19. Camargo MC, Bowlby R, Chu A, et al. Validation and calibration of next-generation sequencing to identify Epstein-Barr virus-positive gastric cancer in The Cancer Genome Atlas. Gastric Cancer 2016;19:676-81.
20. Qiu MZ, He CY, Lu SX, et al. Prospective observation: Clinical utility of plasma Epstein-Barr virus DNA load in EBV-associated gastric carcinoma patients. Int J Cancer 2020;146:272-80.
21. Ribic CM, Sargent DJ, Moore MJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med 2003;349:247-57.
22. De P, Dey N. Mutation-Driven Signals of ARID1A and PI3K Pathways in Ovarian Carcinomas: Alteration Is An Opportunity. Int J Mol Sci 2019;20:5732.
23. O'Donnell JS, Massi D, Teng MWL, et al. PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux. Semin Cancer Biol 2018;48:91-103.
24. Krishnamurthy N, Kato S, Lippman S, et al. Chromatin remodeling (SWI/SNF) complexes, cancer, and response to immunotherapy. J Immunother Cancer 2022. doi: 10.1136/jitc-2022-004669.
25. Shi Y, Shin DS. Dysregulation of SWI/SNF Chromatin Remodelers in NSCLC: Its Influence on Cancer Therapies including Immunotherapy. Biomolecules 2023;13:984.
26. Fang DD, Tang Q, Kong Y, et al. MDM2 inhibitor APG-115 synergizes with PD-1 blockade through enhancing antitumor immunity in the tumor microenvironment. J Immunother Cancer 2019;7:327.
27. Tolcher AW, Reeves JA, McKean M, et al. Preliminary results of a phase II study of alrizomadlin (APG-115), a novel, small-molecule MDM2 inhibitor, in combination with pembrolizumab in patients (pts) with unresectable or metastatic melanoma or advanced solid tumors that have failed immuno-oncologic (IO) drugs. Wolters Kluwer Health 2021.
28. Tolcher AW, Karim R, Tang Y, et al. Phase Ib study of a novel, small-molecule MDM2 inhibitor APG-115 combined with pembrolizumab in US patients with metastatic solid tumors. Am Soc Clin Oncol 2020.
29. Han H, Jain AD, Truica MI, et al. Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy. Cancer Cell 2019;36:483-497.e15.
30. Haikala HM, Anttila JM, Marques E, et al. Pharmacological reactivation of MYC-dependent apoptosis induces susceptibility to anti-PD-1 immunotherapy. Nat Commun 2019;10:620.
31. Hashimoto A, Handa H, Hata S, et al. Inhibition of mutant KRAS-driven overexpression of ARF6 and MYC by an eIF4A inhibitor drug improves the effects of anti-PD-1 immunotherapy for pancreatic cancer. Cell Commun Signal 2021;19:54.
32. Pan Y, Fei Q, Xiong P, et al. Synergistic inhibition of pancreatic cancer with anti-PD-L1 and c-Myc inhibitor JQ1. Oncoimmunology 2019;8:e1581529.
33. Kim KW, Kim N, Choi Y, et al. Different effects of p53 protein overexpression on the survival of gastric cancer patients according to Lauren histologic classification: a retrospective study. Gastric Cancer 2021;24:844-57.
34. Wang Z, Zhang Q, Qi C, et al. Combination of AKT1 and CDH1 mutations predicts primary resistance to immunotherapy in dMMR/MSI-H gastrointestinal cancer. J Immunother Cancer 2022;10:e004703.
35. Castellanos G, Valbuena DS, Pérez E, et al. Chromosomal Instability as Enabling Feature and Central Hallmark of Breast Cancer. Breast Cancer (Dove Med Press) 2023;15:189-211.
36. Vitale I, Shema E, Loi S, et al. Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nat Med 2021;27:212-24.
37. Wu W, Liu Y, Zeng S, et al. Intratumor heterogeneity: the hidden barrier to immunotherapy against MSI tumors from the perspective of IFN-γ signaling and tumor-infiltrating lymphocytes. J Hematol Oncol 2021;14:160.
38. Zhao L, Qu X, Wu Z, et al. TP53 somatic mutations are associated with poor survival in non-small cell lung cancer patients who undergo immunotherapy. Aging (Albany NY) 2020;12:14556-68.

39. Jiang Z, Liu Z, Li M, et al. Immunogenomics Analysis Reveals that TP53 Mutations Inhibit Tumor Immunity in Gastric Cancer. Transl Oncol 2018;11:1171-87.
40. Yang C, Zhang Z, Tang X, et al. Pan-cancer analysis reveals homologous recombination deficiency score as a predictive marker for immunotherapy responders. Hum Cell 2022;35:199-213.
41. Kang K, Wu Y, Han C, et al. Homologous recombination deficiency in triple-negative breast cancer: Multi-scale transcriptomics reveals distinct tumor microenvironments and limitations in predicting immunotherapy response. Comput Biol Med 2023;158:106836.
42. Landen C, Molinero L, Sehouli J, et al. Association of BRCA1/2, homologous recombination deficiency, and PD-L1 with clinical outcomes in patients receiving atezolizumab versus placebo combined with carboplatin, paclitaxel, and bevacizumab for newly diagnosed ovarian cancer: exploratory analyses of IMagyn050/GOG3015/ENGOT-ov39. Gynecol Oncol 2021;162:S37-S8.
43. Ter Brugge P, Moser SC, Bièche I, et al. Homologous recombination deficiency derived from whole-genome sequencing predicts platinum response in triple-negative breast cancers. Nat Commun 2023;14:1958.
44. Wen H, Feng Z, Ma Y, et al. Homologous recombination deficiency in diverse cancer types and its correlation with platinum chemotherapy efficiency in ovarian cancer. BMC Cancer 2022;22:550.
45. Zhao EY, Shen Y, Pleasance E, et al. Homologous Recombination Deficiency and Platinum-Based Therapy Outcomes in Advanced Breast Cancer. Clin Cancer Res 2017;23:7521-30.
46. Smyth EC, Cafferkey C, Loehr A, et al. Genomic loss of heterozygosity and survival in the REAL3 trial. Oncotarget 2018;9:36654-65.
47. Langbein S, Szakacs O, Wilhelm M, et al. Alteration of the LRP1B gene region is associated with high grade of urothelial cancer. Lab Invest 2002;82:639-43.
48. Chen H, Chong W, Wu Q, et al. Association of LRP1B Mutation With Tumor Mutation Burden and Outcomes in Melanoma and Non-small Cell Lung Cancer Patients Treated With Immune Check-Point Blockades. Front Immunol 2019;10:1113.
49. Yasufuku I, Saigo C, Kito Y, et al. Prognostic significance of LDL receptor-related protein 1B in patients with gastric cancer. J Mol Histol 2021;52:165-72.
50. Cai J, Feng D, Hu L, et al. FAT4 functions as a tumour suppressor in gastric cancer by modulating Wnt/β-catenin signalling. Br J Cancer 2015;113:1720-9.
51. Nemtsova MV, Kuznetsova EB, Bure IV. Chromosomal Instability in Gastric Cancer: Role in Tumor Development, Progression, and Therapy. Int J Mol Sci 2023;24:16961.
52. NCCN. NCCN Clinical Practice Guideline in Oncology. Gastric Cancer Version 2. 2024.
53. Wainberg ZA, Enzinger PC, Kang YK, et al. Bemarituzumab in patients with FGFR2b-selected gastric or gastro-oesophageal junction adenocarcinoma (FIGHT): a randomised, double-blind, placebo-controlled, phase 2 study. Lancet Oncol 2022;23:1430-40.
54. Lee J, Kim ST, Kim K, et al. Tumor Genomic Profiling Guides Patients with Metastatic Gastric Cancer to Targeted Treatment: The VIKTORY Umbrella Trial. Cancer Discov 2019;9:1388-405.
55. van Laarhoven HWM, Derks S. Claudin-18.2 targeting by zolbetuximab: results of SPOTLIGHT in perspective. Lancet 2023;401:1630-1.

Cite this article as: Guo P, Yang Y, Wang L, Zhang Y, Zhang B, Cai J, de Melo FF, Strickland MR, Huang M, Liu B. Development of a streamlined NGS-based TCGA classification scheme for gastric cancer and its implications for personalized therapy. J Gastrointest Oncol 2024;15(5):2053-2066. doi: 10.21037/jgo-24-345

