Journal for ImmunoTherapy of Cancer

Comprehensive molecular characterization of cfDNA as predictive and monitoring biomarkers in advanced gastric cancer receiving immunotherapy

Jingshuai Fang,[1,2,3] Yuhui Yu,[2] Yan Sun,[2,4] Mengmeng Ji,[5,6] Yue Jiang,[2] Xiaofeng Chen,[7,8] Caiwang Yan,[1,2,3] Guangfu Jin [1,2,4]

To cite: Fang J, Yu Y, Sun Y, *et al*. Comprehensive molecular characterization of cfDNA as predictive and monitoring biomarkers in advanced gastric cancer receiving immunotherapy. *Journal for ImmunoTherapy of Cancer* 2026;**14**:e013792. doi:10.1136/jitc-2025-013792

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jitc-2025-013792).

JF, YY and YS contributed equally.

Accepted 01 April 2026

For numbered affiliations see end of article.

Correspondence to
Professor Guangfu Jin;
guangfujin@njmu.edu.cn

Professor Caiwang Yan;
caiwangyan@njmu.edu.cn

ABSTRACT

Background Advanced gastric cancer (aGC) exhibits substantial heterogeneity in response to combination immunotherapy. Circulating cell-free DNA (cfDNA) enables non-invasive profiling of tumor dynamics and may provide biomarkers for response prediction.

Methods We enrolled 94 patients with aGC undergoing combination immunotherapy and assigned them to a discovery set (n=49) and an internal validation set (n=45). Plasma cfDNA was collected pre-treatment and post-treatment and profiled by low-pass whole-genome sequencing and whole-genome bisulfite sequencing in the discovery set, with targeted bisulfite sequencing used in the validation set.

Results The discovery set included 34 responders and 15 non-responders, and the validation set included 30 responders and 15 non-responders. We found that responders showed longer cfDNA, lower cfDNA tumor fraction (median: 0.06 vs 0.01, p<0.001), reduced chromosomal instability (median genomic instability index: 0.026 vs 0.007, p<0.001), and higher global methylation (median: 0.715 vs 0.724, p=0.007). Additionally, the differentially methylated region (DMR) at chr20:25849353-25849490 consistently showed higher methylation in responders in both the discovery (adjusted p=0.006) and validation sets (adjusted p=0.049). A predictor based on this DMR outperformed programmed death-ligand 1 (PD-L1) combined positive score (CPS) with area under the curve (AUCs) of 0.79 (discovery: 95% CI 0.65 to 0.93) and 0.72 (validation: 95% CI 0.54 to 0.91). When integrating PD-L1 CPS with this DMR, the AUCs were 0.81 (discovery: 95% CI 0.67 to 0.95) and 0.75 (validation: 95% CI 0.56 to 0.94), respectively. For on-treatment monitoring, three DMRs increased specifically in responders; among them, increased methylation of chr8:110479193-110480324 and chr8:50891437-50892120 was associated with improved progression-free survival (median: 4.90 vs 11.57 months, p<0.001; median: 5.20 vs 10.20 months, p=0.007).

Conclusion Integrated cfDNA profiling captures immunotherapy-associated molecular dynamics in aGC. A single pretreatment cfDNA methylation marker (chr20:25849353-25849490) improves response prediction beyond PD-L1 CPS and represents a potential predictive biomarker for combination immunotherapy.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Currently, programmed death-ligand 1 (PD-L1) combined positive score, tumor mutation burden, and microsatellite instability are regarded as the most reliable predictive biomarkers for response to immunotherapy. However, significant intrapatient heterogeneity and the lack of adequate tumor tissue may limit their effectiveness in clinical practice.

WHAT THIS STUDY ADDS

⇒ This study establishes a novel multidimensional circulating cell-free DNA (cfDNA) profiling framework that integrates fragment length, chromosomal instability, global methylation, and specific differentially methylated regions (DMRs).

⇒ It identifies baseline methylation at chr20:25849353-25849490 as a predictive biomarker (area under the curve (AUC)=0.79), which demonstrates enhanced predictive power when combined with PD-L1 combined positive score (CPS) (AUC=0.81).

⇒ Longitudinal analysis reveals dynamic shifts in methylation at chr8:110479193-110480324 post-treatment, which correlate with survival benefits. These findings provide mechanistic insights into the epigenetic regulation of immunotherapy response.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The cfDNA-based integrative model challenges current single-biomarker paradigms and advocates for the adoption of multimodal liquid biopsy in clinical decision-making. The validation of the synergy between chr20 DMR and PD-L1 CPS could refine patient stratification in clinical trials, while the dynamics of chr8 methylation may guide adaptive treatment monitoring. These findings position cfDNA methylation as a guideline-relevant tool aimed at accelerating the development of precision oncology frameworks for advanced gastric cancer.

INTRODUCTION

The advent of immune checkpoint blockade targeting the programmed cell death protein

1 (PD-1)/programmed death-ligand 1 (PD-L1) axis has revolutionized oncology, ushering in a new era of cancer therapeutics characterized by durable responses and distinct biological mechanisms of action.[1] Immunotherapy, particularly immune checkpoint inhibitors (ICIs), has dramatically transformed the treatment landscape for various malignancies, including advanced gastric cancer (aGC),[2 3] a disease historically associated with dismal survival outcomes and limited therapeutic options.[4 5] Unlike molecularly targeted therapies that focus on defined oncogenic drivers, ICIs frequently induce long-lasting remissions and lower recurrence rates, underscoring their unique and multifaceted mechanisms of antitumor activity.[6]

Following regulatory approval of anti-PD-1/PD-L1 antibodies for a broad spectrum of solid tumors, recent phase III clinical trials have demonstrated their clinical efficacy in aGC.[1] Biomarkers such as tumor mutational burden (TMB) and PD-L1 expression have been endorsed by the US Food and Drug Administration for predicting ICI benefit.[7 8] Additionally, hematologic indices, including platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio, and Systemic Immune-Inflammation Index, alongside conventional tumor markers like carcinoembryonic antigen, carbohydrate antigens (CA72-4, CA125, CA19-9), and alpha-fetoprotein, are frequently used for disease monitoring and therapeutic assessment.[9–11] However, these approaches face significant limitations, including inherent biological variability, the necessity for invasive tissue sampling for molecular analyses, and interpretive challenges inherent to immunohistochemical assays.[11 12] Additionally, pronounced spatial heterogeneity between primary and metastatic lesions erodes biomarker accuracy, accentuating the critical need for minimally invasive, temporally dynamic predictive biomarkers to guide ICI therapy.[13]

Emerging liquid biopsy technologies, particularly the interrogation of circulating cell-free DNA (cfDNA), have emerged as promising tools for tumor detection and monitoring.[14 15] Tumor-derived cfDNA (circulating tumor DNA (ctDNA)) harbors unique genetic and epigenetic alterations, enabling serial, real-time molecular profiling of cancer dynamics.[16 17] Notably, among the characteristics of ctDNA, methylation patterns demonstrate particular clinical potential: these epigenetic modifications manifest earlier than somatic mutations during tumorigenesis, exhibit tissue-specific patterns, maintain exceptional stability in biofluids, and are minimally affected by clonal hematopoiesis.[18 19] In the context of metastatic castration-resistant prostate cancer, methylation-sensitive restriction enzyme quantitative PCR analyses of liquid biopsies have identified three genes (*AKR1B1*, *LDAH*, *KLF8*) that effectively differentiate responders from non-responders following treatment.[20] However, current methodologies for methylation detection remain constrained by technical limitations; locus-specific assays,[21] methylated DNA immunoprecipitation sequencing,[22] and reduced representation bisulfite sequencing[23] lack the resolution required for comprehensive epigenomic profiling. Previous literature has reported multiple studies using whole-genome bisulfite sequencing (WGBS) to investigate the landscape of cfDNA at single-base resolution, identifying differentially methylated regions (DMRs) by integrating information from multiple individual cytosine-phosphate-guanine (CpG) sites.[24] Collectively, these studies have established a growing consensus in the field that, due to the inherently noisy nature of cfDNA, combining data from neighboring CpG sites yields more reliable and robust outcomes.[24 25] Nevertheless, the genome-wide methylation landscape of cfDNA at single-base resolution remains largely underexplored in the context of gastric cancer.

In this study, we performed low-pass whole-genome sequencing (LP-WGS) and WGBS of cfDNA in a discovery set comprising 49 patients with aGC with 98 plasma samples. For validation, targeted bisulfite sequencing was conducted on 90 plasma samples from 45 immunotherapy-treated patients with aGC (figure 1). This study aimed to characterize genome-wide methylation signatures in cfDNA, establish their predictive value for immunotherapy response, and develop dynamic monitoring frameworks for precision immuno-oncology applications.

METHODS

Study design and patient

Between December 2020 and October 2023, a total of 94 patients with aGC who met eligibility criteria were enrolled from The First Affiliated Hospital of Nanjing Medical University (ChiCTR2000034109[26]) and The Affiliated Cancer Hospital of Nanjing Medical University. The patients were divided into a discovery set (n=49) and an internal validation set (n=45). The inclusion criteria were as follows: (1) confirmed metastatic or recurrent gastric adenocarcinoma; (2) at least one measurable lesion as defined by the Response Evaluation Criteria in Solid Tumors (RECIST) V.1.1[27]; (3) age of 18 years or older; (4) adequate organ function as per-protocol; and (5) an Eastern Cooperative Oncology Group performance status of 0 or 1; (6) none of the patients had previously received treatment with ICIs.

All patients, including those in the discovery and internal validation sets, received combination therapy consisting of PD-1 inhibitors (nivolumab, tislelizumab, or camrelizumab) along with platinum-based chemotherapy (oxaliplatin, fluorouracil, or capecitabine), with or without angiogenesis inhibitors (anlotinib or apatinib), until disease progression or intolerable toxicity occurred.

This prospective biomarker study analyzed 188 plasma samples from: (1) a discovery set comprising 49 patients with aGC and 98 longitudinal samples subjected to multiomics profiling; (2) an internal validation set consisting of 45 patients with aGC with 90 longitudinal samples undergoing validation. Comprehensive molecular characterization included LP-WGS and WGBS for the discovery set,

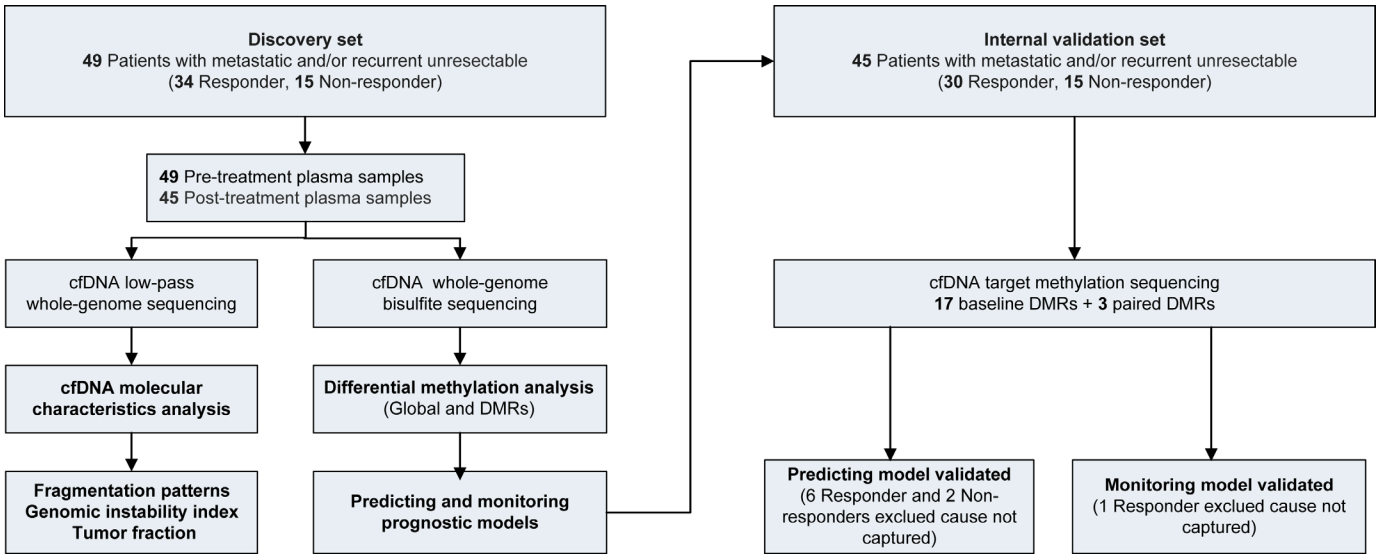

Figure 1 Study design. cfDNA, cell-free DNA.

followed by targeted bisulfite sequencing in the internal validation set (figure 1).

Response assessment and patient stratification

Tumor responses were assessed by radiographic evaluation using computed tomography (CT) every 6 weeks in accordance with RECIST V.1.1 guidelines. At baseline, all lesions were reviewed and classified as either measurable or non-measurable. Measurable lesions were defined as non-nodal lesions with a longest diameter of ≥10 mm or malignant lymph nodes with a short-axis diameter of ≥15 mm. Lymph nodes with short-axis diameters ranging from ≥10 mm to <15 mm, as well as all lesions measuring <10 mm, were classified as non-measurable.

For patients with measurable disease, up to five target lesions in total (with a maximum of two per organ) were selected at baseline to represent the overall tumor size. For each target lesion, the longest diameter was measured for non-nodal lesions, while the short-axis diameter was measured for lymph nodes. The baseline sum of diameters was calculated as the total of these measurements and served as the reference for all subsequent comparisons. Follow-up tumor assessments adhered to the same measurement criteria, and changes in the sum of diameters were used to determine objective response or disease progression.f

The best overall response (BOR) is the best response recorded from the start of the study treatment until the end of treatment. Patients classified as responders achieved either a complete response (CR), which is characterized by the complete disappearance of all target lesions, or a partial response (PR), defined as a minimum 30% reduction in the sum of the diameters of target lesions compared with the baseline sum of diameters. Additionally, stable disease (SD), defined as neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for progressive disease (PD), with a progression-free survival (PFS) of 6 months or longer, also categorizes patients as responders. Conversely, patients exhibiting SD with a PFS of less than 6 months or those with PD, defined as at least a 20% increase in the sum of the diameters of target lesions compared with the smallest sum observed during the study or the emergence of one or more new lesions, are categorized as non-responders.[28]

PD-L1 status classification

PD-L1 protein expression was evaluated in formalin-fixed paraffin-embedded (FFPE) tumor specimens using the clinically validated 22C3 pharmDx immunohistochemistry assay (Agilent Technologies, Santa Clara, California, USA). Stained sections were systematically analyzed using the combined positive score (CPS) quantification system, calculated as follows: CPS = (PD-L1$^+$tumor cells+PD-L1$^+$lymphocytes+PD-L1$^+$macrophages)/total viable tumor cells×100. Specimens demonstrating a CPS ≥1 were considered PD-L1 positive.[29 30]

Microsatellite instability profiling

The microsatellite instability (MSI) status of tumor tissue was determined through two approaches: immunohistochemical (IHC) evaluation and next-generation sequencing (NGS). For the IHC evaluation, sequential 4 µm FFPE sections were stained with monoclonal antibodies against mismatch repair proteins, specifically MLH1, MSH2, MSH6, and PMS2.[31] In the NGS approach, DNA from tumor-normal pairs underwent hybrid capture, targeting 706 cancer-related genes and 39 intronic regions, achieving a mean coverage of ≥1000×. MSI scores were calculated using the formula: MSI score=(unstable loci/total evaluable loci)×100. Tumors exhibiting more than 15% locus instability were classified as microsatellite instability-high (MSI-H) in comparison to corresponding peripheral blood samples.[32]

Plasma biospecimen processing

Peripheral venous blood (8 mL) was collected from the study patients (N=94) using standardized phlebotomy protocols at both baseline and post-treatment. In our study, the pretreatment sample was defined as plasma collected within 7 days prior to the initiation of immunotherapy (from day −7 to day 0). The post-treatment sample was collected on completion of two standard 21-day treatment cycles, specifically within 7 days before the start of the third cycle (from day −7 to day 0 relative to cycle 3). The samples were immediately aliquoted into 10 mL K2 EDTA Vacutainer tubes (BD Biosciences, Franklin Lakes, New Jersey, USA) and maintained vertically at 4°C during transport. Plasma isolation was initiated within 2 hours post-collection through a sequential centrifugation process: primary separation involved centrifugation at 3,000 rpm for 10 min at 4°C, followed by clarification at 16,000×g for 10 min at 4°C to effectively remove residual cellular components.[33]

cfDNA isolation

cfDNA was isolated from 500 µL of plasma using the MiniMax High Efficiency cfDNA Isolation Kit (Apostle, San Jose, California, USA), following the manufacturer's protocol with a modification for protease digestion. The concentrations of DNA were quantified using a Qubit V.2.0 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The purified cfDNA was then resuspended in a final volume of 20 µL and stored at −80°C until further analysis.

Plasma cfDNA was extracted and quantified prior to library preparation. To minimize variability caused by differences in cfDNA concentration across samples, cfDNA input was normalized before downstream processing. For each sample, approximately 5 ng of cfDNA was used for library construction.

cfDNA LP-WGS library preparation and data analysis

The sequencing library preparation was performed following standard low-input cfDNA library construction protocols. Briefly, 12.5 µL of cfDNA samples was subjected to end repair and A-tailing using the KAPA HyperPlus kit (Roche Diagnostics, Basel, Switzerland) following the manufacturer's instructions. Subsequently, adapter ligation and amplification were performed using the KAPA HIFI HotStart ReadyMix (Roche Diagnostics, Basel, Switzerland). The libraries were then purified, quantified, and pooled for sequencing on the Illumina NovaSeq 6000 with paired-end 150 bp mode. For individual samples, LP-WGS yielded a median sequence coverage of approximately 3.9×.

The cfDNA fragment size profiles were generated from paired-end LP-WGS data. After the removal of adapter sequences, the reads were aligned to the hg19 reference genome using BWA-MEM, followed by the elimination of PCR duplicates using Picard Tools "Markduplicates". The fragment size distributions were quantified from the deduplicated BAM files using Picard's "CollectInsertSizeMetrics" tool. This analysis produced genome-wide cfDNA fragment length profiles without stratification by chromosome or region. The median size of cfDNA fragments[34] and the proportions of fragments ranging from 20 to 150 bp,[35] as well as the ratio of fragments measuring 100–150 bp to those measuring 151–220 bp,[36] were subjected to statistical analysis.

CNA analysis and quantification of tumor-derived cfDNA based on LP-WGS

Somatic copy number alterations (CNAs) and tumor fraction (TF) were inferred from cfDNA LP-WGS data using "ichorCNA" (https://github.com/broadinstitute/ichorCNA, V.0.2.0), a tool specifically developed for the detection of CNAs and estimation of TFs in cfDNA without prior knowledge of tumor-specific mutations.[37] Briefly, paired-end reads with a mapping quality of ≥20 were counted in non-overlapping 1 Mb genomic bins using the "readCounter" tool from the "HMMcopy" Suite.[38] Bins overlapping centromeric regions and genomic gaps were excluded. To correct for systematic technical biases, read counts were normalized for guanine-cytosine (GC) content and mappability using locally estimated scatterplot smoothing (LOESS) regression, and further normalized against a reference panel generated from LP-WGS data of 22 healthy donors processed using the same experimental and computational pipeline. Log2 copy -ratio values were then computed for each bin relative to the healthy reference.

"ichorCNA" employs a hidden Markov model-based Bayesian framework to segment the genome into discrete copy-number states, including deletions, copy-neutral regions, gains, and amplifications. The cfDNA signal is modeled as a mixture of tumor-derived and non-tumor-derived DNA fragments, allowing for the estimation of the TF as the inferred proportion of ctDNA in each sample. The model jointly estimates CNA states, TF, and tumor ploidy. Only large-scale CNAs were considered in this analysis; allele-specific imbalance and copy-neutral loss of heterozygosity were not assessed due to the limited allelic resolution of LP-WGS data.

Chromosomal instability (CIN) was quantified using the Genomic Instability Index (GII), which summarizes the overall burden of somatic CNAs across the genome. For each sample, the GII was calculated from segmented CNA profiles generated by "ichorCNA", adhering to previously published methodologies.[39] Briefly, for each genomic segment, we evaluated the absolute log2 copy-ratio relative to the diploid baseline and identified segments whose deviations exceeded a predefined threshold. The GII was subsequently defined as the proportion of all assessed genomic segments that surpassed this deviation threshold, thereby reflecting the overall extent of genome-wide copy number gains and losses and providing a quantitative measure of chromosomal instability.

Chromosome arm-level alterations, along with statistically significant amplifications and deletions, were identified and visualized using "GISTIC" V.2.0 (https://

genepattern.broadinstitute.org/) with the following parameters: 0.95 confidence level, 0.25 amplification and deletion thresholds, and a 0.25 significance threshold for q values.[40]

cfDNA WGBS library preparation and data analysis

Library preparation included end repair, A-tailing, and adapter ligation, performed according to previously established methods. The sequencing library preparation followed standard procedures, including end repair, A-tailing, and adapter ligation as previously described. Subsequently, plasma-derived cfDNA underwent bisulfite conversion using the EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, California, USA). Amplification steps were executed using the KAPA HIFI HotStart ReadyMix (Roche Diagnostics, Basel, Switzerland). Then, the libraries were purified, quantified, and pooled for sequencing on Illumina NovaSeq 6000 with paired-end 150 bp mode. For individual samples, WGBS produced a median sequence coverage of approximately 38.3×.

The sequencing quality and raw reads trimming were performed as previously described. Subsequently, we adhered to the standard "Bismark" workflow to perform mapping to the hg19 reference genome, PCR deduplication, and methylation extraction (https://github.com/FelixKrueger/Bismark, V.0.24.2) for all WGBS samples.[41] Only CpG sites were used for subsequent analysis and visualization.

cfDNA targeted bisulfite sequencing library preparation and data analysis

End repair, A-tailing, adapter ligation, bisulfite conversion, PCR amplification, purification, and quantification were performed according to established protocols. According to the manual, pre-library samples are mixed in equal amounts to create a pool for every 12 samples, with each pool containing a total cfDNA input of approximately 1,500 ng. Hybridization capture is subsequently performed using the QuarHy One Reagent Kit (Dynegene, Guangzhou, China). The libraries are then purified, quantified, and pooled for sequencing on the Illumina NovaSeq 6000 in paired-end 150 bp mode. For individual samples, targeted bisulfite sequencing produced a median of approximately 9.64 million reads.

The sequencing quality, raw read trimming, and mapping were conducted as previously described. The methylation levels of DMRs were calculated for all CpG dinucleotides within the regions captured by the probes. This was achieved by dividing the number of unique reads containing a cytosine at the interrogated position by the total number of reads covering that position.

Computational identification of DMRs in cfDNA

DMRs were detected using the R package "DSS (Dispersion Shrinkage for Sequencing data)".[42] Briefly, DSS employs a Bayesian hierarchical model with local smoothing to borrow information across neighboring CpG sites, which increases sensitivity under uneven sequencing depth and reduces noise at individual CpGs. DMRs are subsequently constructed by aggregating contiguous differentially methylated loci based on user-defined criteria, including the minimum number of CpGs within a region (minCG), the minimum genomic length (minlen), a methylation difference threshold (delta), and a statistical significance threshold at the single CpG level (p.threshold). Thus, regions are not defined by a fixed distance cut-off alone, but rather by the density, consistency, and magnitude of CpG-level methylation differences within genomic neighborhoods.

For the baseline analysis (responders vs non-responders), we applied used criteria (minCG=3, minlen=50 bp, delta=0.1, p.threshold=0.01) to maintain sensitivity to shorter regulatory regions, such as promoters and enhancers. For the longitudinal analysis (pretreatment vs post-treatment), to avoid false positives and ensure that only robust treatment-associated changes were captured, we adopted more stringent criteria (minCG=5, minlen=500 bp, delta=0.1, p.threshold=1×10^{-5}) to prioritize longer regions supported by multiple CpGs and strong statistical evidence. We then annotated the identified DMRs with specific functional regions (CpG island, CpG shore, CpG shelve, CpG inter, enhancer, CDS, 3′UTR, 5′UTR, intron, exon, intergenic, and promoter) using the "annotatr" package.[43]

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was conducted to evaluate whether the genes associated with baseline DMRs and those associated with therapy-induced DMRs were enriched in the transcriptional changes linked to immunotherapy responses in human gastric cancer. RNA sequencing data from both immunotherapy responders and non-responders were sourced from the public dataset PRJEB25780.[44] A differential expression analysis was performed to compare responders and non-responders, ranking all genes based on their log2 fold change. A total of 172 DMR-associated genes from 325 baseline DMRs were used to construct gene sets. GSEA was executed using the "clusterProfiler" package,[45] and the results were visualized with the "GseaVis" package.[46]

Machine learning model

The R packages "caret" and "glmnet" were employed for data preprocessing and feature selection in the discovery set. First, recursive feature elimination (RFE) was conducted using a Random Forest (RF) classifier, where feature importance was quantified based on each feature's contribution to predicting treatment response. To address the stochastic variability inherent to RF models, the complete RFE procedure was repeated 500 times with different random seeds. DMRs that were selected in more than 250 of the 500 repetitions were retained as stable and reproducible candidates.

In the second step, the stability-selected DMRs were further refined using least absolute shrinkage and selection operator (LASSO) regression with 10-fold

cross-validation. Cross-validation was employed to determine the optimal regularization parameter (λ.min), after which LASSO was refitted on the full dataset using this λ value. The DMRs with non-zero coefficients in the refitted model were defined as the final candidate features. These features were subsequently assessed for their ability to discriminate between responders and non-responders in both the discovery and internal validation sets.

To evaluate the combined predictive value of PD-L1 CPS and the validated DMR, we constructed a generalized linear model (GLM) using both variables as continuous predictors. Specifically, PD-L1 CPS and the methylation level of the selected DMR were jointly included in a binomial GLM to estimate the probability of response. Model training was performed with leave-one-out cross-validation (LOOCV) implemented in the "caret" package in R. The performance of these models was evaluated in both the discovery and internal validation sets using receiver operating characteristic (ROC) curves.[47]

Outcomes

This study was designed as a prospective translational analysis to investigate cfDNA-based biomarkers associated with response to ICIs in patients with aGC. Radiographic response evaluated according to RECIST V.1.1 was used to categorize patients into responders and non-responders and served as the clinical reference for biomarker discovery and validation.

PFS was defined as the time from treatment initiation to disease progression or death, and overall survival (OS) as the time from treatment initiation to death from any cause. Objective response rate (ORR), PFS, and OS were summarized descriptively to support interpretation of cfDNA-based biomarker analyses.

Statistical analysis

ORR along with 95% CIs were calculated using the " binom.test" function. Survival analysis was performed using the R packages "survival" and "survminer". PFS and OS curves were generated using the Kaplan-Meier method, with survival probability plotted against survival time (in months). The patients were divided into methylation high and low groups with the median methylation level of each DMR as an unbiased, data-driven cut-off. Additionally, in the longitudinal analysis of cfDNA DMRs before and after treatment, patients were assigned to the decreased group when the post-treatment methylation level of a given DMR was lower than baseline, and to the increased group when the post-treatment level was higher than baseline. Univariable survival differences between groups were assessed using the log-rank test, with a p value of <0.05 considered statistically significant.

The Fisher's exact test was employed to analyze and compare categorical variables. The Wilcoxon test was used for comparisons of independent continuous variables, while the paired Wilcoxon test was applied for paired samples. Additionally, heatmap analyses of DMRs were conducted using the R package "pheatmap". We also calculated the statistical power of the tests using the " power.roc.test" function from the "pROC" package, establishing a target area under the curve (AUC) of 0.75 and a significance level of 0.05. In the discovery set, which included 34 responders and 15 non-responders, the determined power was 0.85. In the internal validation set, comprising 30 responders and 15 non-responders, the power was 0.83. All statistical analyses were executed using R V.4.2.3.

RESULTS

Patient characteristics and treatment response

A total of 94 patients with aGC who underwent combination immunotherapy were enrolled and assigned to a discovery set (n=49) and an internal validation set (n=45) (online supplemental table 1). According to the BOR following the criteria of RECIST V.1.1[27] throughout the treatment course, the discovery set included 29 patients with PR, 16 with SD, and 4 with PD (ORR, 59.2%), whereas the validation set comprised 23 patients with PR, 16 with SD, and 6 with PD (ORR, 51.1%) (online supplemental table 1). Based on the commonly used criteria,[28] responders were defined as patients with CR, PR, or SD with PFS >6 months, whereas non-responders were defined as patients with SD with PFS <6 months or PD. Accordingly, 34 responders and 15 non-responders were identified in the discovery set, while 30 responders and 15 non-responders were identified in the validation set.

No significant differences were observed in demographic characteristics (including age and sex) or clinicopathological features (including Lauren classification and primary tumor site) between responders and non-responders in either the discovery set or the internal validation set (online supplemental table 1). Notably, PD-L1 status[48] and MSI status,[44] which are potential predictors of immunotherapy response, were also comparable between the two groups (online supplemental table 1).

Genomic profiling of cfDNA in patients with aGC stratified by treatment response

In the discovery set, we extracted cfDNA from plasma samples collected from responders and non-responders at both pretreatment (R-pre, NR-pre) and post-treatment (R-post, NR-post) time points. The cfDNA concentrations did not differ significantly between responders and non-responders at baseline (median: 0.48 vs 0.43 ng/µL, p=0.474) or after treatment (median: 0.51 vs 0.47 ng/µL, p=0.965; online supplemental figure 1A and table 2). Additionally, no correlation was detected between baseline cfDNA concentration and tumor size across all pretreatment samples (R=0.157; p=0.280; online supplemental figure 1B).

We then assessed genome-wide cfDNA features, including fragmentation patterns and copy-number derived metrics (figure 1). At baseline, cfDNA median fragment size was negatively correlated with tumor size (R=−0.469, p<0.001; Online supplemental figure 1C). However, no significant

differences were observed between responders and non-responders in median fragment length (figure 2A, online supplemental figure 1D), the proportion of fragments between 20–150 bp (online supplemental figure 1E), or the fragment short/long ratio (100–150 bp/151-220bp) (online supplemental figure 1F) at baseline. After treatment, responders showed a shift toward longer cfDNA fragments (median fragment size: 166.85 vs 168.49 bp, p=0.016; figure 2A, online supplemental figure 1D), accompanied by a significant decrease in the proportion of fragments between 20–150 bp (p=0.011; online supplemental figure 1E) and in the fragment short/long ratio (p=0.004; online supplemental figure 1F), whereas no significant changes were observed among non-responders, indicating a treatment-associated decrease in tumor-derived cfDNA in responders, with concomitant enrichment of longer cfDNA fragments.[49 50]

We next compared cfDNA TF and GII, estimated from cfDNA CNA profiles (figure 2B), between the responders and non-responders at baseline and post-treatment (online supplemental table 2). As expected, both baseline GII (R=0.510, p<0.001, online supplemental figure 1G) and cfDNA TF (R=0.606, p<0.001, online supplemental figure H) were positively correlated with tumor size. Additionally, the baseline GII (figure 2C) and cfDNA TF (figure 2D) did not differ significantly between responders and non-responders. After treatment, both the GII (median: 0.026 vs 0.007, p<0.001, figure 2C) and cfDNA TF (median: 0.057 vs 0.010, p<0.001; figure 2D) significantly decreased in responders, indicating a restoration of chromosomal instability. In contrast, non-responders largely retained their baseline CNA profiles (GII, median: 0.061 vs 0.010, p=0.330; figure 2C; cfDNA TF, median: 0.123 vs 0.023, p=0.252; figure 2D).

For example, in patient EP05 (responder; PD-L1 CPS=2), a 65.9% reduction in tumor size and a concomitant decrease in cfDNA TF were observed after two cycles of immunotherapy (figure 2E,F). Conversely, patient EP24 (non-responder; PD-L1 CPS=10) exhibited only a 2.6% decrease in tumor size but an increase in cfDNA TF post-treatment (figure 2G,H), highlighting the high quality of the extracted cfDNA and the accuracy of cfDNA-derived signals in reflecting patient tumor dynamics.

cfDNA methylomic profiling reveals response-associated signatures in patients with aGC

Similarly, based on WGBS data from cfDNA in the discovery set (figure 1), we observed a strong inverse correlation between baseline global cfDNA methylation levels and tumor size (R=−0.612, p<0.001; online supplemental figure 2A and table 2). In addition, the global cfDNA methylation levels did not differ between responders and non-responders at baseline (p=0.889; online supplemental figure 2B). After treatment initiation, responders showed a significant increase in global methylation levels (median: 0.715 vs 0.724, p=0.007; online supplemental figure 2B), whereas non-responders exhibited no discernible change (median: 0.717 vs 0.714, p=0.639; online supplemental figure 2B). Post-treatment methylation levels also tended to be higher in responders (p=0.022; online supplemental figure 2B).

To further assess whether methylation signatures could serve as predictive biomarkers for combination immunotherapy in aGC, we integrated information across individual CpG sites in pretreatment cfDNA and identified 325 DMRs (online supplemental table 3). Among these, 64.3% showed higher methylation levels in responders compared with non-responders, whereas 35.7% were hypomethylated in responders (figure 3A). These DMRs were mapped to 172 genes (online supplemental table 3), including *TNFSF9*,[51] *HLA-DPA1*,[52] and *LTA*,[53] which have been implicated in antitumor immunity and response to immune checkpoint blockade. Pathway enrichment analysis of these DMR-associated genes revealed significant enrichment in immune-related pathways, including regulation of mononuclear cell proliferation, positive regulation of lymphocyte proliferation, positive regulation of T-cell proliferation (online supplemental figure 2C). To further support the functional relevance of these genes, we analyzed an external transcriptomic dataset of tumor tissues from patients with gastric cancers treated with immunotherapy (PRJEB25780). Notably, 8.1% of the DMR-associated genes showed differential expression between responders and non-responders, and GSEA revealed significant enrichment of these DMR-associated genes in responders (online supplemental figure 2D), supporting a potential role in modulating antitumor immunity.

To distinguish immunotherapy responders from non-responders, we applied Random Forest-based variable importance scoring and LASSO regression (online supplemental figure 2E) to the 325 preselected DMRs, yielding 17 candidate DMRs (online supplemental table 4), which were subsequently validated in the internal validation set among individuals with successful capture of all 17 DMRs by targeted bisulfite sequencing (figure 1). Among these, the DMR at chr20:25849353-25849490 exhibited consistently elevated methylation levels in responders compared with non-responders across both the discovery set (adjusted p=0.006; figure 3B, online supplemental table 4) and internal validation set (adjusted p=0.049; figure 3C, online supplemental table 4). To exclude the potential confounding by MSI-H status, sensitivity analyses restricted to MSS individuals confirmed that chr20:25849353-25849490 remained different between responders and non-responders (online supplemental table 5).

We next evaluated the predictive performance of DMR chr20:25849353-25849490, both alone and in combination with PD-L1 CPS. In the discovery set, the DMR-only model achieved an AUC of 0.79 (95% CI 0.65 to 0.93; figure 3D), outperforming PD-L1 CPS alone (AUC=0.65, 95% CI 0.49 to 0.82; figure 3D). The combined model further improved performance (AUC=0.81, 95% CI 0.67 to 0.95; figure 3D). These findings were replicated in the validation set, with corresponding AUCs of 0.61 (PD-L1 CPS alone: 95% CI

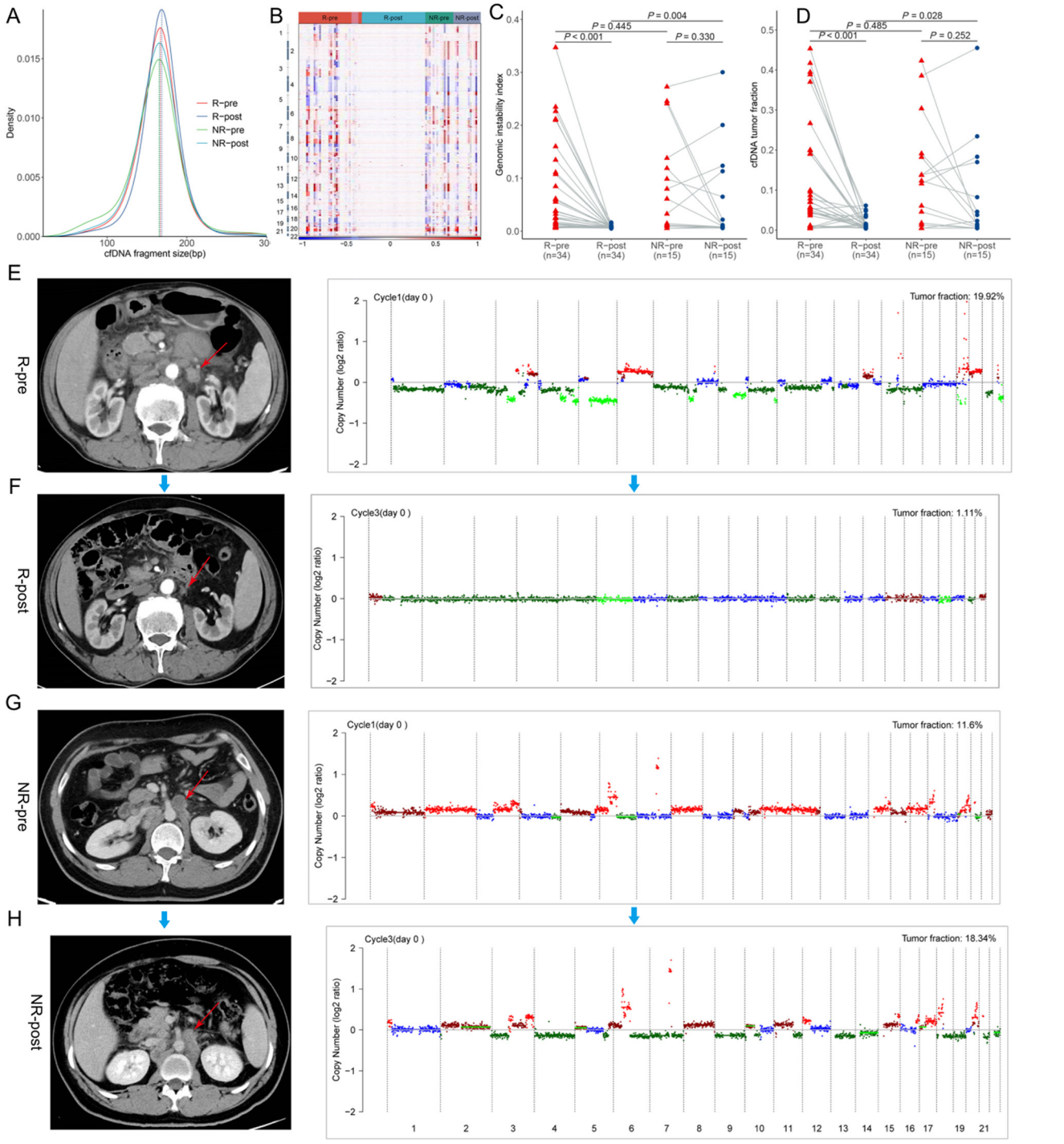

Figure 2 Molecular profiling of cfDNA dynamics pre-immunotherapy and post-immunotherapy. (A) Fragment size distribution of human-mapped cfDNA across treatment groups: responders (R-pre, R-post) and non-responders (NR-pre, NR-post). Vertical dashed lines denote group-specific median fragment lengths. (B) Copy-number alteration heatmaps for R-pre, R-post, NR-pre, and NR-post are displayed, with the x-axis representing individual samples arranged from left to right. Each row corresponds to a different chromosome, organized vertically from top to bottom. The color gradient in the heatmap indicates the degree of copy-number alteration, with red and blue representing copy gains and losses, respectively. The comparative analysis of cfDNA reveals Genomic Instability Index (C) and TFx (D) in responders (n=34) versus non-responders (n=15) both pre-treatment and post-treatment. A Wilcoxon test was used for comparisons of unpaired samples, while a paired Wilcoxon test was employed for paired samples. (E–H) Representative imaging-genomic correlations: responder baseline (E) and follow-up CT (F) after two immunotherapy cycles (left panels), with paired genome-wide copy-number profiles and TFx (right panels). Non-responder baseline (G) and follow-up CT (H) (left panels), with corresponding genomic profiles (right panels). cfDNA, cell-free DNA; NR-pre, non-responders at pre-treatment; NR-post, non-responders at post-treatment; R-pre, responders at pre-treatment; R-post, responders at post-treatment; TFx, tumor fraction.

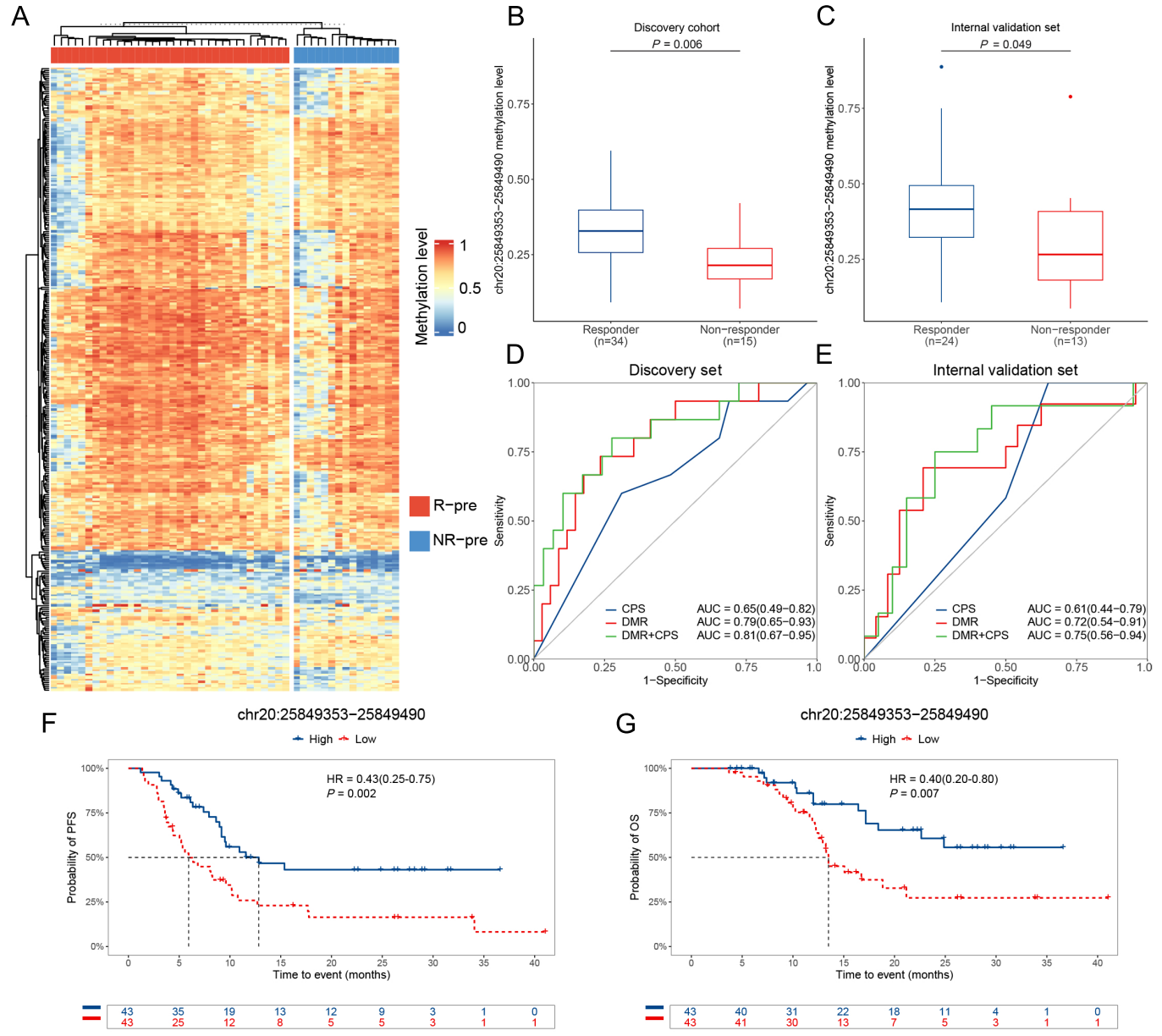

Figure 3 Differential baseline cfDNA methylation signatures are predictive of immunotherapy response in advanced gastric cancer. (A) Unsupervised clustering of whole-genome bisulfite sequencing data from pretreatment plasma cfDNA of patients with advanced gastric cancer (discovery set: 34 responders (R-pre), 15 non-responders (NR-pre)). The heatmap displays methylation levels at DMRs. (B–C) Methylation levels at DMR chr20:25849353-25849490 in the discovery (B) and internal validation sets (C) (validation: 24 R-pre, 13 NR-pre). The boxes illustrate the IQR, with the midline representing the median. Statistical significance was determined using the Wilcoxon test. (D–E) ROC curves evaluating predictive performance in the discovery (D), and internal validation sets (E). The models compared include: PD-L1 CPS (combined positive score), DMR chr20:25849353-25849490, PD-L1 CPS+DMR. (F–G) Kaplan-Meier survival analysis stratified by DMR chr20 methylation (high vs low): the dashed lines indicate median survival. AUC, area under the curve; cfDNA, cell-free DNA; DMRs, differentially methylated regions; NR-pre, non-responders at pre-treatment; OS, overall survival; PD-L1, programmed death-ligand 1; PFS, progression-free survival; ROC, receiver operating characteristic; R-pre, responders at pre-treatment.

0.44 to 0.79), 0.72 (DMR alone: 95% CI 0.54 to 0.91), and 0.75 (DMR+CPS: 95% CI 0.56 to 0.94; figure 3E), respectively. The performance metrics for each predictive model in both the discovery and validation sets are detailed in online supplemental table 6. Furthermore, we also found that the fragments between 20–150 bp and the fragment short/long ratio exhibited only modest discrimination between responders and non-responders (AUC=0.61; online supplemental figure 2F). Moreover, adding these fragmentation features to DMR chr20:25849353-25849490 did not improve performance (online supplemental figure 2F). Together, these results suggest that DMR chr20:25849353-25849490 may

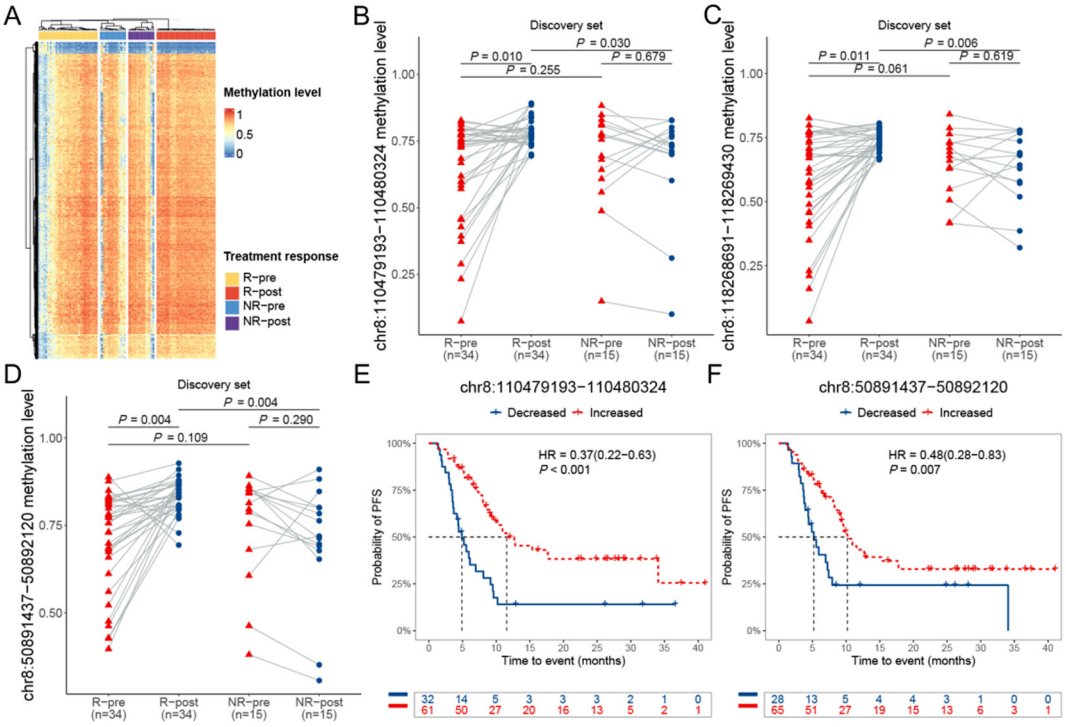

Figure 4 Dynamic cfDNA methylation changes and survival associations in immunotherapy-treated advanced gastric cancer. (A) Unsupervised clustering of pretreatment (pre) and post-treatment (post) plasma cfDNA methylation profiles from patients with advanced gastric cancer treated with combination immunotherapy (discovery set: 34 responders (R-pre/R-post), 15 non-responders (NR-pre/NR-post)). Heatmap displays whole-genome bisulfite sequencing data for DMRs. (B–D) Box plots comparing methylation levels of three chromosome 8 DMRs across treatment phases: (B) DMR chr8:110479193-110480324, (C) DMR chr8:118268691-118269430, (D) DMR chr8:50891437-50892120. Adjusted p value was generated from a logistic model including age, sex, tumor fraction, tumor size and DMR methylation level for unpaired samples, while a logistic model including tumor size and DMR methylation level for paired samples. Kaplan-Meier survival curve compares PFS in patients stratified by decreased versus increased methylation levels of DMR chr8:110479193-110480324 (E) and chr8:50891437-50892120 (F), with the p value calculated using the log-rank test. Patients were assigned to the decreased group when the post-treatment methylation level of a given DMR was lower than baseline, and to the increased group when the post-treatment level was higher than baseline. Horizontal dashed lines indicate median survival. cfDNA, cell-free DNA; DMRs, differentially methylated regions; NR-pre, non-responders at pre-treatment; NR-post, non-responders at post-treatment; PFS, progression-free survival; R-pre, responders at pre-treatment; R-post, responders at post-treatment.

serve as a potential biomarker for predicting immunotherapy response.

Finally, in all patients with follow-up data (including responders and non-responders), we assessed the association between chr20:25849353-25849490 methylation and clinical outcomes. We found that patients with high methylation levels of this DMR had significantly prolonged PFS (median: 5.93 vs 12.83 months; HR=0.43, 95% CI 0.25 to 0.75, p=0.002; figure 3F) and OS (median: 13.5 months vs not reached; HR=0.40, 95% CI 0.20 to 0.80, p=0.007; figure 3G).

cfDNA methylation changes in aGC across immunotherapy treatment

A comparative analysis of pretreatment and post-treatment plasma cfDNA methylation profiles in the discovery set revealed 5,199 treatment-associated DMRs (online supplemental table 7). Notably, 96.3% of these DMRs exhibited increased methylation levels following treatment compared with baseline, while only 3.7% showed decreased methylation (figure 4A). Among these DMRs, we annotated 995 genes (online supplemental table 7), which were enriched in immune response regulating signaling pathway and B-cell receptor signaling pathway (online supplemental figure 3A).

A dual feature selection strategy that combined Random Forest-based variable ranking with LASSO regression (online supplemental figure 3B) identified three DMRs with potential utility for immunotherapy monitoring: chr8:110479193-110480324, chr8:118268691-118269430, and chr8:50891437-50892120 (online supplemental table 8). Notably, these three DMRs showed a significant trend of increased methylation levels in responders (all adjusted p<0.05; figure 4B–D, online supplemental table 8), but not in non-responders (figure 4B–D, online supplemental table 8), and this differing trend was not influenced by MSI status (online supplemental table 9).

Similarly, we then stratified all patients by high versus low methylation levels for each of the three DMRs and conducted survival analyses. We found that patients with increased methylation levels of

chr8:110479193-110480324 and chr8:50891437-50892120 demonstrated significantly improved PFS (median: 4.90 vs 11.57 months; HR=0.37, 95% CI 0.22 to 0.63, p<0.001; figure 4E; median: 5.20 vs 10.20 months, HR=0.48, 95% CI 0.28 to 0.83, p=0.007; figure 4F), but not OS (median: 15.07 vs 24.87 months; HR=0.56, 95% CI 0.29 to 1.08, p=0.079; online supplemental figure 3C; median: 15.07 vs 21.17 months, HR=0.76, 95% CI 0.38 to 1.54, p=0.443; online supplemental figure 3D) In contrast, methylation of chr8:118268691-118269430 was not significantly associated with either PFS or OS (online supplemental figure 3E,F).

DISCUSSION

The development of universally accessible biomarkers to predict patient responses to combination immunotherapy represents a critical unmet clinical need, particularly given the limited global accessibility of genomic and immunological assays. In this study, we integrated preplanned clinical outcomes with cfDNA profiling to identify determinants of therapeutic response in aGC.

While cfDNA has been extensively studied in cancer research, comprehensive pan-cancer analyses of its concentration across various tumor types and stages remain limited. In our study, baseline and post-treatment cfDNA concentrations did not differ significantly between responders and non-responders to combination immunotherapy. This finding may reflect the confounding effects of demographic and physiological variables, such as age, sex, body mass index, and ethnicity, which are known to influence cfDNA levels.[54] Similarly, Spearman correlation analysis indicated no significant association between cfDNA concentration and tumor size.

Notably, responders demonstrated a lower abundance of shorter cfDNA fragments (<167 bp), with fragment size inversely correlated with tumor size. Previous studies suggest that cfDNA fragments centered around 167 bp (and its multiples) are indicative of caspase-dependent apoptosis,[55] whereas ctDNA fragments are enriched at approximately 145 bp.[35] This distinction underscores the potential utility of fragment size analysis in differentiating tumor-derived DNA from non-malignant sources.

Another observation is that responders demonstrated reduced copy number variation (CNVs), indicating a restoration of chromosomal stability following treatment. Using LP-WGS and ichorCNA-based CNV analysis, we observed a significant decrease in cfDNA TF among responders, which positively correlated with tumor size. Genome-doubling events, a hallmark of CIN, are strongly associated with poor prognosis.[56 57] The CIN-driven acquisition of CNAs during tumor proliferation fosters intratumoral heterogeneity,[58] while LP-WGS of cfDNA enables non-invasive tracking of somatic copy number alteration evolution and CIN dynamics.[59] This approach proves particularly valuable in low TF samples, offering a scalable strategy for CIN biomarker development and patient stratification.[60]

We developed a cfDNA methylation-based model for predicting and monitoring immunotherapy efficacy, which outperforms established biomarkers such as PD-L1 IHC, TMB, and MSI-H. Current biomarkers face significant limitations: (1) PD-L1 testing lacks standardization across platforms and antibodies[61]; (2) TMB quantification requires resource-intensive genomic profiling with variability across panels[62]; (3) MSI-H is rare and confined to specific subtypes.[63 64] Rather than relying on genome-wide WGBS, the DMRs identified in our study can be translated into focused, small-panel methylation assays that equire lower sequencing depth and reduced assay complexit. When combined with conventional assessments, such targeted methylation markers may provide a scalable and clinically practical complement to existing biomarkers.

Among the 17 baseline DMRs, one DMR located at chr20:25849353–25849490, which maps to *FAM182B*, was validated in the internal validation set. *FAM182B* is a long non-coding RNA, and although its exact biological function remains to be fully characterized, low expression of *FAM182B* has been associated with a trend towards improved prognosis in hepatocellular carcinoma.[65] Of the three longitudinal DMRs identified, two are located within the introns of *PKHD1L1* (chr8:110479193–110480324) and syntrophin gamma 1 (*SNTG1*) (chr8:50891437–50892120), while the third DMR (chr8:118268691–118269430) does not map to any known gene. *PKHD1L1* has been identified as a potential tumor suppressor in thyroid cancer, with reduced expression correlating with larger tumor size, distant metastasis, and advanced disease stage.[66 67] It has also been implicated in immune cell infiltration, suggesting a role in shaping the tumor microenvironment.[68 69] *SNTG1* has been associated with paclitaxel response, with higher expression levels observed in sensitive cell lines.[70] The SNP rs318885 in *SNTG1* may contribute to suboptimal tumor response by impairing breast epithelial cell adhesion, thereby facilitating malignant transformation,[71] or by decreasing mitotic microtubule stability, leading to reduced paclitaxel binding and diminished therapeutic efficacy.[72] Supporting these hypotheses is the expression of the dystrophin-associated protein complex, which includes *SNTG1* and sarcoglycan delta (*SGCD*) and links the cytoskeleton to the extracellular matrix in muscle. Specifically, β-sarcoglycan, another component of this complex, colocalizes with α-tubulin (the site of action of paclitaxel),[73] corresponding to the cell cycle stage when paclitaxel interacts with tubulin. The association of these genes with microtubules suggests they may influence microtubule dynamics and thereby modulate paclitaxel response, particularly given that the *SNTG1* SNP lies within an intronic region bound by multiple transcription factors, many of which are implicated in transcriptional misregulation in cancer.[74]

Several limitations should be acknowledged. First, the discovery and internal validation sets were relatively small, particularly for paired longitudinal analyses, which increases the risk of overfitting when applying multistep feature-selection procedures to high-dimensional methylation

data. Although we employed a stability-selection framework combined 500 iterations of LOOCV-based RFE and 10-fold cross-validated LASSO and further examined all candidate DMRs in an internal validation set, the results should still be regarded as hypothesis-generating. The modest decline in AUC from discovery to validation likely reflects both intercohort heterogeneity and the inherent constraints of small-sample modeling, including potential optimism bias from using the same dataset for feature selection and performance estimation. Second, one baseline DMR and three longitudinal DMRs were consistently reproduced, underscoring the necessity for larger, multicenter cohorts to confirm the robustness and clinical relevance of these markers. Additionally, the discovery set was profiled using WGBS, whereas the internal validation set was assessed using targeted bisulfite sequencing of the candidate DMRs in an independent patient cohort. Differences in assay platforms, technical workflows, and underlying patient characteristics between these two sets may also contribute to the observed reduction in AUC. Third, the absence of intermediate sampling time points limits our ability to determine whether cfDNA methylation changes precede radiographic responses and to fully characterize their temporal dynamics. Finally, Epstein-Barr virus (EBV) status was not available in this study due to insufficient tumor tissue. Overall, despite the reproducibility of several DMRs in the internal validation set, the limited sample size remains a significant constraint and may contribute to residual overfitting. Larger, independent cohorts with more granular longitudinal sampling will be essential to validate these preliminary findings and improve the generalizability of the proposed methylation-based predictors.

In conclusion, using LP-WGS and WGBS, we identified distinct cfDNA features in aGC responders, including reduced short-fragment abundance, attenuated CIN, lower TF, and elevated global methylation levels. Our methylation-based predictive model represents a clinically accessible tool for identifying patients likely to benefit from immunotherapy. Nevertheless, multicenter validation and mechanistic investigations are essential before clinical translation.

Author affiliations
[1]School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu, China
[2]Department of Epidemiology, School of Public Health, State Key Laboratory Cultivation Base of Biomarkers for Cancer Precision Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu, China
[3]The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi People's Hospital, Wuxi Medical Center, Wuxi, Shanghai, China
[4]Department of Medical Oncology, Jiangsu Key Laboratory of Innovative Cancer Diagnosis & Therapeutics, Jiangsu Cancer Hospital and Jiangsu Institute of Cancer Research, The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, Jiangsu, China
[5]Innovation Center of Suzhou, Nanjing Medical University, Suzhou, Jiangsu, China
[6]National Center of Technology Innovation for Biopharmaceuticals, Suzhou, Jiangsu, China
[7]Department of Oncology, The First Affiliated Hospital With Nanjing Medical University, Nanjing, Jiangsu, China
[8]Gastric Cancer Center, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu, China

Acknowledgements The research team expresses profound appreciation to all clinical participants and collaborating investigators for their invaluable contributions to this study.

Contributors Conceptualization: GJ, CY, and XC. Methodology: JF, YY, MJ, and YJ Investigation: YS, XC, JF, and YY. Visualization: JF and YY. Funding acquisition: GJ, CY, and XC. Project administration: GJ. Supervision: GJ. Writing—original draft: JF. Writing—review and editing: GJ, CY, JF, and YY. GJ is the study guarantor, had full access to the data, and made the final decision to submit for publication.

Funding This study was supported by the National Natural Science Foundation of China (82125033, 82230110), Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501500), the Clinical Capacity Enhancement Project of the First Affiliated Hospital of Nanjing Medical University, the Pukou District Social Cause Science and Technology Development Project in 2020 (S2020-21), and the Major Project of Science and Technology Development Fund of the Pukou Branch Hospital of Jiangsu Province Hospital (Nanjing Pukou Central Hospital) in 2021 (KJ2021-1).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This research protocol received formal ethical approval from the Institutional Review Boards of Nanjing Medical University (No. NJMUIRB (2021) 589) and The First Affiliated Hospital of Nanjing Medical University (No. 2020-SR-002 and 2020-SR-558). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon reasonable request. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

ORCID iD
Guangfu Jin https://orcid.org/0000-0003-0249-5337

REFERENCES

1 Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science* 2018;359:1350–5.
2 Ribas A, Hamid O, Daud A, *et al*. Association of Pembrolizumab With Tumor Response and Survival Among Patients With Advanced Melanoma. *JAMA* 2016;315:1600–9.
3 Wang F, Wei XL, Wang FH, *et al*. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. *Ann Oncol* 2019;30:1479–86.
4 Smyth EC, Nilsson M, Grabsch HI, *et al*. Gastric cancer. *Lancet* 2020;396:635–48.
5 Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74:229–63.
6 Sharma P, Hu-Lieskovan S, Wargo JA, *et al*. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* 2017;168:707–23.

7 Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019;19:133–50.
8 Subbiah V, Solit DB, Chan TA, *et al*. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) ≥10: a decision centered on empowering patients and their physicians. *Ann Oncol* 2020;31:1115–8.
9 Chen Y, Zhang C, Peng Z, *et al*. Association of Lymphocyte-to-Monocyte Ratio With Survival in Advanced Gastric Cancer Patients Treated With Immune Checkpoint Inhibitor. *Front Oncol* 2021;11:589022.
10 Nakano M, Kuromatsu R, Niizeki T, *et al*. Immunological inflammatory biomarkers as prognostic predictors for advanced hepatocellular carcinoma. *ESMO Open* 2021;6:100020.
11 Valero C, Lee M, Hoen D, *et al*. Pretreatment neutrophil-to-lymphocyte ratio and mutational burden as biomarkers of tumor response to immune checkpoint inhibitors. *Nat Commun* 2021;12:729.
12 Valero C, Lee M, Hoen D, *et al*. Response Rates to Anti–PD-1 Immunotherapy in Microsatellite-Stable Solid Tumors With 10 or More Mutations per Megabase. *JAMA Oncol* 2021;7:739.
13 McLaughlin J, Han G, Schalper KA, *et al*. Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer. *JAMA Oncol* 2016;2:46–54.
14 Klein EA, Richards D, Cohn A, *et al*. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* 2021;32:1167–77.
15 Gao Q, Lin YP, Li BS, *et al*. Unintrusive multi-cancer detection by circulating cell-free DNA methylation sequencing (THUNDER): development and independent validation studies. *Ann Oncol* 2023;34:486–95.
16 Rothwell DG, Ayub M, Cook N, *et al*. Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med* 2019;25:738–43.
17 Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol* 2021;18:297–312.
18 Soto J, Rodriguez-Antolin C, Vallespín E, *et al*. The impact of next-generation sequencing on the DNA methylation-based translational cancer research. *Transl Res* 2016;169:1–18.
19 Widschwendter M, Jones A, Evans I, *et al*. Epigenome-based cancer risk prediction: rationale, opportunities and challenges. *Nat Rev Clin Oncol* 2018;15:292–309.
20 Dillinger T, Sheibani-Tezerji R, Pulverer W, *et al*. Identification of tumor tissue-derived DNA methylation biomarkers for the detection and therapy response evaluation of metastatic castration resistant prostate cancer in liquid biopsies. *Mol Cancer* 2022;21:7.
21 Xu R-H, Wei W, Krawczyk M, *et al*. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater* 2017;16:1155–61.
22 Shen SY, Singhania R, Fehringer G, *et al*. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature New Biol* 2018;563:579–83.
23 Guo S, Diep D, Plongthongkum N, *et al*. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet* 2017;49:635–42.
24 Loyfer N, Magenheim J, Peretz A, *et al*. A DNA methylation atlas of normal human cell types. *Nature New Biol* 2023;613:355–64.
25 Li W, Li Q, Kang S, *et al*. CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res* 2018;46:e89.
26 Chen X, Xu H, Chen X, *et al*. First-line camrelizumab (a PD-1 inhibitor) plus apatinib (an VEGFR-2 inhibitor) and chemotherapy for advanced gastric cancer (SPACE): a phase 1 study. *Sig Transduct Target Ther* 2024;9:73.
27 Eisenhauer EA, Therasse P, Bogaerts J, *et al*. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
28 Gogleva A, Polychronopoulos D, Pfeifer M, *et al*. Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. *Nat Commun* 2022;13:1667.
29 Kang Y-K, Boku N, Satoh T, *et al*. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet* 2017;390:2461–71.
30 Doi T, Piha-Paul SA, Jalal SI, *et al*. Safety and Antitumor Activity of the Anti–Programmed Death-1 Antibody Pembrolizumab in Patients With Advanced Esophageal Carcinoma. *JCO* 2018;36:61–7.
31 Samowitz WS. Evaluation of colorectal cancers for Lynch syndrome: practical molecular diagnostics for surgical pathologists. *Mod Pathol* 2015;28 Suppl 1:S109–13.
32 Roock WD, Vriendt VD, Normanno N, *et al*. KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol* 2011;12:594–603.
33 Meddeb R, Pisareva E, Thierry AR. Guidelines for the Preanalytical Conditions for Analyzing Circulating Cell-Free DNA. *Clin Chem* 2019;65:623–33.
34 Zhang H, Dong P, Guo S, *et al*. Hypomethylation in HBV integration regions aids non-invasive surveillance to hepatocellular carcinoma by low-pass genome-wide bisulfite sequencing. *BMC Med* 2020;18:200.
35 Mouliere F, Chandrananda D, Piskorz AM, *et al*. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 2018;10:eaat4921.
36 Cristiano S, Leal A, Phallen J, *et al*. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature New Biol* 2019;570:385–9.
37 Adalsteinsson VA, Ha G, Freeman SS, *et al*. n.d. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*8.
38 Ha G, Roth A, Lai D, *et al*. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res* 2012;22:1995–2007.
39 Silberman A, Goldman O, Boukobza Assayag O, *et al*. Acid-Induced Downregulation of ASS1 Contributes to the Maintenance of Intracellular pH in Cancer. *Cancer Res* 2019;79:518–33.
40 Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12:R41.
41 Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 2011;27:1571–2.
42 Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 2016;32:1446–53.
43 Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics* 2017;33:2381–3.
44 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.
45 Xu S, Hu E, Cai Y, *et al*. Using clusterProfiler to characterize multiomics data. *Nat Protoc* 2024;19:3292–320.
46 Zhang J, Li H, Tao W, *et al*. GseaVis: An R Package for Enhanced Visualization of Gene Set Enrichment Analysis in Biomedicine. *Med Research* 2025;1:131–5.
47 Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
48 Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *j immunotherapy cancer* 2019;7:278.
49 Manier S, Park J, Capelletti M, *et al*. Whole-exome sequencing of cell-free DNA and circulating tumor cells in multiple myeloma. *Nat Commun* 2018;9:1691.
50 Peneder P, Stütz AM, Surdez D, *et al*. Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. *Nat Commun* 2021;12:3230.
51 Pennycuick A, Teixeira VH, AbdulJabbar K, *et al*. Immune Surveillance in Clinical Regression of Preinvasive Squamous Cell Lung Cancer. *Cancer Discov* 2020;10:1489–99.
52 Liu H, Zhao R, Qin R, *et al*. Panoramic comparison between NK cells in healthy and cancerous liver through single-cell RNA sequencing. *Cancer Biol Med* 2022;19:1–18.
53 Höckendorf U, Dutta S, Kloos A, *et al*. Lymphotoxin alpha eradicates acute myeloid leukemia and simultaneously promotes healthy hematopoiesis in mice. *Sci Transl Med* 2025;17:eadu3313.
54 Hestand MS, Bessem M, van Rijn P, *et al*. Fetal fraction evaluation in non-invasive prenatal screening (NIPS). *Eur J Hum Genet* 2019;27:198–202.
55 Snyder MW, Kircher M, Hill AJ, *et al*. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* 2016;164:57–68.
56 Dewhurst SM, McGranahan N, Burrell RA, *et al*. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov* 2014;4:175–85.

57 Jamal-Hanjani M, Wilson GA, McGranahan N, *et al*. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* 2017;376:2109–21.
58 Schvartzman J-M, Sotillo R, Benezra R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat Rev Cancer* 2010;10:102–15.
59 Van Roy N, Van Der Linden M, Menten B, *et al*. Shallow Whole Genome Sequencing on Circulating Cell-Free DNA Allows Reliable Noninvasive Copy-Number Profiling in Neuroblastoma Patients. *Clin Cancer Res* 2017;23:6305–14.
60 Sansregret L, Vanhaesebroeck B, Swanton C. Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol* 2018;15:139–50.
61 Mino-Kenudson M, Le Stang N, Daigneault JB, *et al*. The International Association for the Study of Lung Cancer Global Survey on Programmed Death-Ligand 1 Testing for NSCLC. *J Thorac Oncol* 2021;16:686–96.
62 Nassar AH, Adib E, Abou Alaiwi S, *et al*. Ancestry-driven recalibration of tumor mutational burden and disparate clinical outcomes in response to immune checkpoint inhibitors. *Cancer Cell* 2022;40:1161–72.
63 Smyth EC, Wotherspoon A, Peckitt C, *et al*. Mismatch Repair Deficiency, Microsatellite Instability, and Survival: An Exploratory Analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) Trial. *JAMA Oncol* 2017;3:1197–203.
64 Sundar R, Huang KK, Qamra A, *et al*. Epigenomic promoter alterations predict for benefit from immune checkpoint inhibition in metastatic gastric cancer. *Ann Oncol* 2019;30:424–30.
65 Liu J, Li W, Zhang J, *et al*. Identification of key genes and long non-coding RNA associated ceRNA networks in hepatocellular carcinoma. *PeerJ* 2019;7:e8021.
66 Wang Q-X, Chen E-D, Cai Y-F, *et al*. A panel of four genes accurately differentiates benign from malignant thyroid nodules. *J Exp Clin Cancer Res* 2016;35:169.
67 Liu W, Zheng J, Han L, *et al*. Clinical performance of a machine learning-based model for detecting lymph node metastasis in papillary thyroid carcinoma: A multicenter study. *Int J Surg* 2025;111:4062–7.
68 Hogan MC, Griffin MD, Rossetti S, *et al*. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. *Hum Mol Genet* 2003;12:685–98.
69 Liao Z, Cheng Y, Zhang H, *et al*. A novel prognostic signature and immune microenvironment characteristics associated with disulfidptosis in papillary thyroid carcinoma based on single-cell RNA sequencing. *Front Cell Dev Biol* 2023;11:1308352.
70 Eng L, Ibrahim-zada I, Jarjanazi H, *et al*. Bioinformatic analyses identifies novel protein-coding pharmacogenomic markers associated with paclitaxel sensitivity in NCI60 cancer cell lines. *BMC Med Genomics* 2011;4:18.
71 Arco A, Favaloro A, Gioffrè M, *et al*. Sarcoglycans in the Normal and Pathological Breast Tissue of Humans: An Immunohistochemical and Molecular Study. *Cells Tissues Organs (Print)* 2012;195:550–62.
72 Yin S, Bhattacharya R, Cabral F. Human mutations that confer paclitaxel resistance. *Mol Cancer Ther* 2010;9:327–35.
73 Hashimoto R, Yamaguchi M. Dynamic changes in the subcellular localization of Drosophila beta-sarcoglycan during the cell cycle. *Cell Struct Funct* 2006;31:173–80.
74 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4.