Uncovering neutrophil extracellular traps-driven gastric cancer heterogeneity and C5AR1 as a therapeutic target

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Acta Biochimica et Biophysica Sinica</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>ABBS-2023-482.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>07-Dec-2023</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Zhao, Jing; Henan Cancer Hospital Affiliated Cancer Hospital of Zhengzhou University, Department of Medical Oncology  
Li, Xiang-Yu ; Department of Radiation Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China  
Li, Li-Ming ; The First Affiliated Hospital of Zhengzhou University, Department of Radiology  
Chen, Bei-Bei ; Henan Cancer Hospital Affiliated Cancer Hospital of Zhengzhou University, Department of Medical Oncology  
Xu, Wei-Feng ; Henan Cancer Hospital Affiliated Cancer Hospital of Zhengzhou University, Department of Medical Oncology  
He, Yun-Duan ; Henan Cancer Hospital Affiliated Cancer Hospital of Zhengzhou University, Department of Medical Oncology  
Chen, Xiao-Bing ; Department of Medical Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China |
| Keywords: | neutrophil extracellular traps, gastric cancer, heterogeneity, personalized therapy, C5AR1, treatment target |
160x171mm (600 x 600 DPI)
160x185mm (300 x 300 DPI)
160x219mm (300 x 300 DPI)
Figure 4

160x235mm (400 x 400 DPI)
Figure 5

160x93mm (400 x 400 DPI)
Figure 6

160x122mm (400 x 400 DPI)
Figure 7

160x125mm (400 x 400 DPI)
Figure 8

160x80mm (400 x 400 DPI)
Figure 9

160x120mm (400 x 400 DPI)
Original Article

Identification of neutrophil extracellular trap-driven gastric cancer heterogeneity and C5AR1 as a therapeutic target

Jing Zhao¹, Xiangyu Li², Liming Li³, Beibei Chen¹, Weifeng Xu¹, Yunduan He¹, and Xiaobing Chen¹,4,*

¹Department of Medical Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou 450003, China
²Department of Radiation Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou 450003, China
³Department of Radiology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China
⁴State Key Laboratory of Esophageal Cancer Prevention and Treatment, Zhengzhou University, Zhengzhou 450052, China

*Correspondence address. Tel: +86-13937100233 E-mail: zlyychenxb0807@zzu.edu.cn

Received: October 15, 2023
Accepted: November 28, 2023

Abstract
Neutrophil extracellular traps (NETs) are implicated in gastric cancer (GC) growth, metastatic dissemination, cancer-associated thrombosis, etc. This work is conducted to elucidate the heterogeneity of NETs in GC. The transcriptome heterogeneity of NETs is investigated in TCGA-STAD via a consensus clustering algorithm, with subsequent external verification in the GSE88433 and GSE88437 cohorts. Clinical and molecular traits, the immune microenvironment, and drug response are characterized in the identified NET-based clusters. Based upon the feature genes of NETs, a classifier is built for estimating NET-based clusters via machine learning. Multiple experiments are utilized to verify the expressions and implications of the feature genes in GC. A novel NET-based classification system is proposed for reflecting the heterogeneity of NETs in GC. Two NET-based clusters have unique and heterogeneous clinical and molecular features, immune microenvironments, and responses to targeted therapy and immunotherapy. A logistic regression model reliably differentiates the NET-based clusters. The feature genes C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDA, MPEG1, PLEK, SRGN, and STAB1 are proven to be aberrantly expressed in GC cells. Specific knockdown of C5AR1
effectively hinders GC cell growth and elicits intracellular ROS accumulation. In addition, its suppression suppresses the aggressiveness and EMT phenotype of GC cells. In all, NETs are the main contributors to intratumoral heterogeneity and differential drug sensitivity in GC, and C5AR1 has been shown to trigger GC growth and metastatic spread. These findings collectively provide a theoretical basis for the use of anti-NETs in GC treatment.

**Keywords:** neutrophil extracellular trap, gastric cancer, heterogeneity, personalized therapy, C5AR1, treatment target

**Introduction**

Gastric cancer (GC) represents a global health-care challenge [1]. The therapeutic landscape of GC has markedly evolved. Although the effectiveness of chemotherapy and surgical removal has improved, patient prognosis remains unsatisfactory [2]. Chemicalotherapeutic agents with low toxicity, along with molecular-driven targeted treatment, are valuable in sequential therapeutic regimens for the optimization of patient survival. In advanced disease, only trastuzumab and immune checkpoint blockade (ICB), *e.g.*, nivolumab and pembrolizumab in combination with chemotherapy, have shown durable and superior effects in treating HER2-positive and PD-L1-positive patients, respectively [3‒5]. Biomarkers currently utilized for therapeutic choices include HER2 overexpression and amplification, combined positive PD-L1 score, and microsatellite instability. [6]. Intra- and intertumour heterogeneity are notable features of GC and are partially responsible for unfavorable survival [7,8]. Nonetheless, only histological classification is not sufficient to potently stratify patients for personalized therapy or to prolong survival time. Large-scale molecular characterization of GC patients is important for identifying candidate treatment options [9].

Neutrophils are the dominating leukocytes in peripheral blood and are the initial defense against invading pathogens [10–12]. Neutrophil extracellular traps (NETs) are network structures comprising decondensed DNA strands coated with granule proteins that were first discovered in 1996 [13], and these NETs were subsequently named NETosis [14,15]. Based upon mounting evidence, circulating NET levels are aberrantly increased in GC patients, and NETs have indispensable implications for GC growth [16], the metastatic cascade [17‒19] and cancer-associated thrombosis [20]. However, additional studies on NET formation and compounds that hinder or destroy NETs are needed, providing a theoretical basis for the use of NETs in GC treatment.

In the present study, a novel NET-based molecular classification system for GC was proposed, revealing the heterogeneity of NET formation. In addition, a classifier was built for the differentiation of NET-based clusters. Among the feature genes in the classifier, C5AR1 was experimentally proven to mediate GC growth and metastatic spread, indicating the potential of C5AR1 as a treatment target in GC.

**Materials and Methods**
Human GC data acquisition
Using the TCGA database (https://portal.gdc.cancer.gov/), GC RNA sequencing and clinical data from the TCGA-STAD cohort were curated, and the GC expression matrix and clinical traits of the GSE88433 and GSE88437 cohorts were subsequently acquired from the GEO database (https://www.ncbi.nlm.nih.gov/gds/) [21]. Utilizing edgeR [22], the raw data were normalized.

Consensus clustering analysis
Thirty-five NET-associated genes were identified from prior research [23] and are summarized in Supplementary Table S1. Based upon their expression profiles, consensus clustering analysis was performed across TCGA-STAD samples utilizing ConsensusClusterPlus [24]. The K values of the clusters were set as 2 to 5, with subsequent optimal selection based upon the consensus matrix. Such clustering was proven in GSE88433 and GSE88437. Differential overall survival (OS) and disease-free survival (DFS) outcomes were evaluated in distinct clusters.

Somatic variant analysis
The mutation annotation format of the TCGA-STAD cohort was adopted from the TCGA-STAD cohort [25]. The mutation frequency of genes was estimated and compared between clusters by using Fisher’s exact test.

Characterizing cell compositions
CIBERSORT, a deconvolution approach [26], was adopted for estimating immune compositions within GC tumours. The fractions of stromal and immune cell admixture were inferred via the ESTIMATE method [27].

Functional enrichment analysis
GSEA software was used for enrichment analyses of GO and KEGG terms [28]. In addition, enrichment levels of classical tumorigenic signaling pathways were scored via the GSVA package [29].

Treatment response estimation
Based on the GDSC2 database [30], the IC50 of small-molecule agents was inferred via oncoPredict [31]. The ICB response was estimated in accordance with the tumor mutational burden (TMB), T-cell-inflamed score [32], and tumor immune dysfunction and exclusion (TIDE) score [33].

Development of a classifier
First, DEGs with a $|\log_2$ fold change$|>1$ and $q<0.05$ were selected by comparing two NET-based clusters. Gene expression levels were utilized as the prediction value, and the area under the curve (AUC) was estimated. In accordance with the AUC values, the top twelve genes were chosen as feature genes. Through the Python machine learning framework sklearn, linear function support vector machine (linearSVM),
radial basis function SVM (RBFsvm), Gaussian naive Bayes (GaussianNB), and logistic regression (logistic) models were trained. Receiver operating characteristic (ROC) analysis was also conducted to evaluate the prediction efficacy of the pROC package.

**Patients and clinical specimens**

In total, 106 GC patients were recruited from Henan Cancer Hospital. Tumor tissue specimens were collected following institutional protocols. Prognostic information was also obtained. Each participant provided written informed consent. This study was approved by the Ethics Committee of Henan Cancer Hospital (No. 2020012).

**Cell culture**

Normal human gastric mucosal epithelial cells (GES-1) and GC cell lines (AGS, SGC7901 and MKN28) were purchased from the Typical Culture Preservation Committee Cell Bank of the Chinese Academy of Sciences (Shanghai, China). All the cells were cultivated in RPMI-1640 medium (12633012; Gibco, Carlsbad, USA) supplemented with 10% FBS (SH30396.03; HyClone, Carlsbad, USA). The cells were placed in a 5% CO₂ cell incubator (SW-CJ-1FD; Thermo Scientific, Waltham, USA) at 37°C. The medium was changed every 1–2 days.

**RT-qPCR analysis**

Cells or tissues were lysed with Trizol (H10318; TransGen, Beijing, China), after which RNA was extracted. The RNA concentration was measured via a Nanodrop 2000 spectrophotometer (Thermo Scientific). Before quantification, the extracted RNA was stored at −80°C. Complementary DNA was subsequently synthesized. A real-time PCR kit (AQ131-01; TransGen) and SYBR Green Master Mix (Rainbio, Shanghai, China), and RT-qPCR was performed. Table 1 summarizes the primer sequences utilized for RT-qPCR. The 2⁻ΔΔCt method was used to quantify mRNA expression.

Three siRNAs against C5AR1 (si-C5AR1) and a negative control (si-NC) (Table 1) purchase by Sangon Biotech (Shanghai, China) were transfected into GC cells via Lipofectamine 2000 (11668030; Invitrogen, Carlsbad, USA) based on the manufacturer’s instructions. The transfection efficiency was verified via western blot analysis. The optimal si-C5AR1 was subsequently chosen.

**Western blot analysis**

Cells were lysed in RIPA cell lysis buffer supplemented with a protease inhibitor mixture, followed by centrifugation and supernatant collection. Following protein electrophoresis, the proteins were transferred onto PVDF membranes (Millipore, Billerica, USA), which were subsequently blocked. Next, the cells were incubated with primary anti-C5AR1 (1:500; MA5-16937; Invitrogen) or anti-GAPDH (1:3000; AF7021; Affinity, Changzhou, China) antibodies at 4°C overnight. HRP-labelled secondary antibody (1:3000; S0001; Affinity) was added, and the samples were incubated for 1 h at room temperature. The proteins were developed through an
enhanced chemiluminescence (ECL) luminous substrate (P90719; Millipore) and photographed via a chemiluminescence imaging system (ChemiScop Series 3600; Enco Biology, Shenzhen, China).

**Cell viability test**
Following the manufacturer’s instructions for the CCK8 detection kit (C0038; Beyotime, Shanghai, China), 10 μL of CCK8 reagent was added to each well. After 1 h of culture, the absorbance at 450 nm was measured utilizing a microplate reader (Multiskan MK3; Thermo Scientific).

**Reactive oxygen species (ROS) measurement**
In accordance with the manufacturer’s instructions, ROS levels were measured via a DCFH-DA kit (E-BC-K138-F; Elabscience, Wuhan, China). A single-cell suspension was added to the working solution. The cells were incubated at 37°C for 1 h in the dark and centrifuged at 1000 g for 10 min. The gathered cell precipitate was resuspended in the working fluid. Finally, ROS levels were tested through a flow cytometer (FACSCalibur; BD, Franklin Lakes, USA).

**Wound healing assay**
A marker was used to mark the area evenly behind the 6-well plate. Approximately 5×10^5 cells were added to each well. The next day, scratches were made vertically to the horizontal line behind the back. Following removal of the scratched cells, serum-free medium was added. At 0 and 48 h, the cells were photographed.

**Immunofluorescence**
Following fixation with 4% polyformaldehyde, the cells were permeated with PBS containing 0.1% Triton X-100 (Sigma-Aldrich, St Louis, USA), and the cells were subsequently blocked with 5% BSA sealing liquid. Next, the cells were incubated with an anti-E-cadherin (1:500; AF0131; Affinity) or anti-Vimentin (1:500; AF7013; Affinity) antibody for 1 h at room temperature as well as a secondary antibody (1:500; ab6717 or ab150083; Abcam) for 1 h in the dark. The cells were then stained with DAPI (C1002; Beyotime) for 5 min. Under a fluorescence microscope (DMI3000B; Leica), images were acquired.

**Statistical analysis**
All the analyses were conducted with the R platform (version 3.6.1) or GraphPad Prism software (version 9.0.1). Pearson’s or Spearman’s test was adopted for correlation analysis. Differences between two groups were estimated with Student’s t test, the Wilcoxon test, or the chi-square test, with one-way ANOVA for comparisons between ≥3 groups. P<0.05 was considered as statistical significance.

**Results**
**Stratification of GC patients into two NET-based clusters**
To characterize the heterogeneity of NETs across GCs, we conducted consensus
clustering analysis of TCGA-STAD samples based on the expressions of known NET-associated genes. After obtaining a consensus matrix, the TCGA-STAD samples were clearly stratified into two clusters: cluster 1 and cluster 2 (Figure 1A). The consistency and reliability of this molecular classification system were externally proven in the GSE88433 and GSE88437 cohorts (Figure 1B,C). Additionally, heterogeneous clinical traits were investigated between the clusters. In particular, cluster 1 had more advanced tumor grade, stage, and TNM stage (Figure 1D). Cluster 1 patients also presented shorter OS and DFS time than did cluster 2 patients (Figure 1E,F). Taken together, these findings suggested that the two NET-based clusters are clinically heterogeneous in GC.

Differences in heterogeneous somatic variants and immune compositions between NET-based clusters
Analysis of somatic variants revealed that TP53, CDKN2A, PEG3, DMXL1, TRPV4, FMO2, and LRRC3 had more frequent mutations in cluster 1 than in cluster 2, with more frequent mutations in SOX5, CDH1, and NCDN occurring in cluster 2 (Figure 2A,B), revealing the unique molecular mutation traits in each cluster. In addition, most immune components appeared to be more abundant in cluster 1 than in the other clusters (Figure 2C,D). Higher immune, stromal and ESTIMATE scores were found in cluster 1 (Figure 2E‒G), further reflecting the richer stromal and immune cell admixtures in this cluster.

Heterogeneous molecular mechanisms between NETs-based clusters
The molecular mechanisms underlying the NET-based classification were subsequently explored. Consequently, immunity-associated mechanisms were notably activated in cluster 1, with remarkable activation of metabolic processes in cluster 2 (Figure 3A‒D). In addition, epithelial–mesenchymal transition (EMT) activity was greater in cluster 1, and cell cycle progression (CCP) activity was greater in cluster 2 (Figure 3E,F). The above findings reflected the heterogeneity in the molecular mechanisms of the two NET-based clusters.

Heterogeneity in treatment responses between NET-based clusters
Next, we investigated whether two NET-based clusters elicited diverse therapeutic responses. Among the small-molecule compounds, cluster 2 was more sensitive to BI-2536, with greater sensitivity to KU-55933, ribociclib, AZD8186, AZD8055, NU7441, SB216763, BMS-754807, AZ960, entospletinib, ZM447439, CZC24832, AMG-319, GSK269962A, WZ4003, JQ1, RVX-208, and PRT062607 in cluster 1 (Supplementary Figure S1A). Three ICB predictors, TMB, T-cell-inflamed status and TIDE score, were adopted for inferring the ICB response. Cluster 2 tumours exhibited greater TMB than cluster 1 tumours did, but the difference was not significant (Supplementary Figure S1B,C). Moreover, higher T-cell-inflamed and TIDE scores were detected in cluster 1 than in another cluster (Supplementary Figure S1D,E). Hence, the two NET-based clusters exhibited diverse responses to targeted therapy and ICB.
Establishment of a stable and reliable classifier for differentiating NET-based clusters

To differentiate the two NET-based clusters, we first selected genes differentially expressed between the clusters (Figure 4A–C). With an AUC>0.9, twelve feature genes of the NET-based classification (C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDa, MPEG1, PLEK, SRGN, and STAB1) were chosen for building five classifiers (linearSVM, RBFSVM, GaussianNB, and logistic models). All the feature genes had higher expression in cluster 1 than in cluster 2 (Supplementary Figure S2). To verify the efficacy of these classifiers in differentiating the two NET-based clusters, receiver operating characteristic (ROC) analysis was conducted. The AUC of the random forest model was 1 for TCGA-STAD cohort, indicating possible data overfitting (Figure 4D). The AUC values of the logistic model were relatively higher, and were 0.97, 0.81 and 0.80 in the TCGA-STAD, GSE88433 and GSE88437, respectively (Figure 4E,F and Table 2). Thus, the logistic model was eventually utilized for differentiating the two NET-based clusters. Table 3 shows the weight of each feature gene in the logistic model. These feature genes were found to be strongly related to each other (Figure 4G). To further validate the performance of the classifier in determining GC prognosis, a real-world cohort of GC patients (n=106) was included in our study. Patients were classified into cluster 1 and cluster 2 according to the weight and expression of feature genes. The results showed that cluster 1 patients had shorter OS than did cluster 2 patients (Figure 4H), and the AUC was 0.7253 (Figure 4I), further confirming the stability and reliability of the classifier. In addition, C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDa, MPEG1, PLEK, and SRGN exhibited notable upregulation in GC cells (AGS, SGC7901 and MKN28) compared with GES-1 cells, with remarkable downregulation of STAB1 (Figure 5).

Knockdown of the feature gene C5AR1 prevents GC cell growth

However, the biological role of C5AR1 in GC has not been determined. Thus, we next investigated the function of C5AR1 in the malignant behaviours of GC cells through experimental validation. Three siRNAs targeting C5AR1 or its controls were transfected into AGS, SGC7901 and MKN28 cells. Consequently, si-C5AR1#2 was proven to have the optimal efficacy in knocking out C5AR1 in three GC cell lines (Figure 6A–D). C5AR1-knockout AGS, SGC7901 and MKN28 cells exhibited impaired proliferative capacity (Figure 6E–G). Thus, it was inferred that targeting C5AR1 potentially prevents GC cell growth.

C5AR1 knockdown elicits intracellular ROS accumulation in GC cells

ROS are implicated in the formation of NETs [34]. Next, we investigated whether C5AR1 influenced intracellular ROS accumulation. Consequently, in the context of C5AR1 knockdown, intracellular ROS levels were notably elevated in AGS, SGC7901 and MKN28 cells (Figure 7), thus demonstrating the function of C5AR1 in...
modulating ROS accumulation within GC cells.

**C5AR1 knockdown hinders the aggressiveness of GC cells**

Based upon the results of wound healing experiments, C5AR1 knockdown effectively elevated the relative wound distance in AGS, SGC7901 and MKN28 cells (Figure 8). Taken together, these findings suggested that suppression of C5AR1 hinders the aggressiveness of GC cells.

**C5AR1 knockdown impairs the EMT phenotype in GC cells**

E-cadherin, a transmembrane glycoprotein, participates in cell-cell adhesion and EMT, while Vimentin is highly expressed in mesenchymal cells and is positively associated with tumor metastasis [35]. C5AR1 knockdown effectively elevated E-cadherin expression and attenuated Vimentin expression in AGS, SGC7901 and MKN28 cells (Figure 9), indicating that aberrantly expressed C5AR1 might be involved in mediating the EMT phenotype of GC cells and even tumor metastasis.

**Discussion**

GC has notable molecular heterogeneity, aggressive behaviours and therapeutic resistance and thus remains a barrier to clinical management [36]. NET formation has been proven to be crucial for GC growth [16], the metastatic cascade [17–19] and cancer-associated thrombosis [20]. Nonetheless, the heterogeneity in NETs remains unclear. Herein, this work proposed a novel NET-based molecular classification system for GC and stratified GC patients into two clusters, which might assist in choosing anti-NET therapy for GC patients.

Clinically, the two NET-based clusters exhibited heterogeneous clinical traits and prognostic outcomes. Carcinogenic signaling pathway activity was greater in cluster 1 than in cluster 2, reflecting worse survival outcomes in these patients. Somatic variants were also heterogeneous in two clusters. In particular, mutant TP53, CDKN2A, PEG3, DMXL1, TRPV4, FMO2, and LRRRC3 were more frequent in cluster 1, with more frequent mutations in SOX5, CDH1, and NCDN in cluster 2. To quantify the NETs-based classification, a logistic regression model classifier composed of C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDA, MPEG1, PLEK, SRGN, and STAB1 was eventually built, which was shown to possess possible clinical implications in NET-based subtyping, thus assisting personalized anti-NET therapy.

Currently, platinum-containing regimens are still the most common treatment for advanced GC, but chemical resistance remains one of the major reasons for treatment failure despite extensive research on the molecular mechanisms underlying chemotherapy resistance [37]. Thus, novel small-molecule agents are urgently needed for GC patients. Here, it was found that cluster 2 exhibited stronger sensitivity to BI-2536, while cluster 1 was more sensitive to KU-55933, ribociclib, AZD8186, AZD8055, NU7441, SB216763, BMS-754807, AZ960, entospletinib, ZM447439, CZC24832, AMG-319, GSK269962A, WZ4003, JQ1, RVX-208, and PRT062607. Preclinical and clinical evidence has proven the therapeutic efficacy of these agents in
GC. For instance, BI2536 acts as an effective and selective inhibitor of PLK1 and synergizes with cisplatin to treat GC [38]. KU-55933 suppresses ataxia telangiectasia mutated, thus preventing p53 phosphorylation [39]. A reduction in ATM kinase activity induced by KU55933 enhances olaparib sensitivity in GCs with depleted or inactivated p53 [40]. Ribociclib functions as a cyclin-dependent kinase 4/6 inhibitor for first-line therapy against HR+/HER2− advanced breast tumours, and its bioavailability is not influenced by gastric pH alterations or food intake [41]. Based upon a phase Ib/II cohort (KCSG ST18-20), AZD8186 (a selective PI3Kβ/δ inhibitor) plus paclitaxel was well tolerated by advanced GC patients [42].

ICB therapy has resulted in a durable clinical response in a minority of GC patients [43]. The two NET-based clusters were heterogeneous in terms of the immune microenvironment, with more abundant immune and stromal components in cluster 1. The T-cell-inflamed phenotype is positively associated with the effectiveness of ICB [44]. In addition, the TIDE score was negatively linked with the ICB response. In accordance with the lower T-cell-inflamed score and lower TIDE score, cluster 2 patients were inferred to respond better to ICB but more likely to experience therapeutic resistance. Hence, more reliable biomarkers of ICB are urgently needed.

The feature genes exhibited strong interactions and heterogeneous expression between NET-based clusters in GC tumours. Among them, C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDA, MPEG1, PLEK, and SRGN were proven to be prominently upregulated in GC cells, with notable downregulation of STAB1. Several feature genes have been experimentally verified in prior studies. For instance, CSF1R participates in modulating cancer stemness, the immunosuppressive microenvironment and metastasis in GC [45]. A genetic reduction in the SRC kinase HCK prevents STAT3-based GC growth [46]. Blockade of the complement receptor C5AR1 enables tumor-associated GC growth to be reprogrammed and reinvigorated CD8+ T-cell-mediated antitumour immunity, thereby synergizing with anti-PD-1 therapy for GC tumor eradication [47]. After further verification of C5AR1, we found that specific suppression of C5AR1 prevented GC growth and provoked intracellular ROS generation. Numerous studies have explored the molecular mechanisms governing the process of GC invasion and metastasis; for example, CORO1C enables the recruitment of phospho-PAK4 at serine 99 to the leading edge and promotes the migration of GC cells [48]. This study revealed that C5AR1 inhibition impaired the aggressiveness and EMT phenotype of GC cells. Taken together, these findings indicate that C5AR1 functions as a treatment target for GC. Notably, C5AR1 presented higher expression in NET-based cluster 1 than in cluster 2. Thus, therapeutically targeting C5AR1 might be effective for GC patients in cluster 1. Our findings could assist in choosing appropriate treatment targets and regimens for GC.

Several limitations need to be acknowledged. Despite the external verification, additional GC samples are needed to prove the reliability of the NET-based classification. Moreover, the biological roles of C5AR1 and other feature genes need to be investigated in additional experiments.

Collectively, the results of this study revealed that NET formation is heterogeneous in GC, and a novel NET-based classification was proposed, which can be estimated
by a machine learning-based classifier. Additionally, C5AR1 was proven to mediate GC growth and metastatic spread. Despite this, our conclusions warrant further in-depth investigations.

Supplementary Data
Supplementary data is available at Acta Biochimica et Biophysica Sinica online.

Funding
This work was supported by the grants from the Science and Technique Foundation of Henan Province (No. 212102310778), the Medical Science and Technique Foundation of Henan Province (No. LHGJ20210201) and the Henan Medical Science and Technology Research Program Joint Construction Project (No. LHGJ20220409).

References


2021; 9.


[48] Li X, Chen M, Yuan Y et al. CORO1C, a novel PAK4 binding protein, recruits phospho-PAK4 at serine 99 to the leading edge and promotes the migration of gastric cancer cells. Acta Biochim Biophys Sin (Shanghai) 2022; 54:673-685.

Figure legends

Figure 1. Stratification of GC patients into two NET-based clusters

(A) Consensus matrix heatmap (k=2) representing the possibility of two patients belonging to the same cluster across TCGA-STAD samples based upon NET-associated genes. The value ranges from 0 to 1, and the colour ranges from white to dark blue. (B,C) Verification of the NET-based consensus clustering analysis in GSE88433 and GSE88437. (D) Heterogeneous clinical traits of two NET-based clusters in the TCGA-STAD cohort. ns, not significant; *P<0.05, **P<0.01, and ***P<0.001. (E,F) Distinct OS and DFS probabilities between the two clusters.

Figure 2. Differences in heterogeneous somatic variants and immune compositions between the two NET-based clusters

(A) Oncoplots illustrating the somatic landscape of two NET-based clusters. Genes are ranked based upon mutational frequency. (B) Forest plot of genes differentially mutated between clusters. (C) Immune composition landscape across TCGA-STAD specimens. (D) Visualization of the fractions of immune compositions in TCGA-STAD patients stratified by NET-based classification. (E-G) Differential immune, stromal or ESTIMATE scores between NET-based clusters. *P<0.05, **P<0.01, ***P<0.001 and ****P<0.0001.

Figure 3. Heterogeneous molecular mechanisms in the two NET-based clusters

(A–D) GSEA for the enrichment analysis of biological processes, cellular components, molecular functions, and KEGG pathways in the two clusters. (E,F) Characterization and comparison of the enrichment scores of EMT and CCP in the two clusters. ***P<0.001.

Figure 4. Definition of a stable and reliable classifier for differentiating between two NET-based clusters

(A,B) Selection of genes with differential expression between clusters. The cut-off was |log2-fold change|>1 and q<0.05. (C) The first twenty genes in each cluster. (D) ROC analysis of the established linear SVM,
RBFsvm, GaussianNB, and logistic models in the TCGA-STAD cohort. (E) Fivefold cross-validation curves of the logistic model. (F) External verification of the logistic model in the GSE88433 and GSE88437 cohorts. (G) Correlation analysis of the twelve feature genes. (H) OS probability of patients in clusters 1 and 2 in a real-world GC patient cohort \((n=106)\). (I) ROC curves for the classification of OS in the cohort.

**Figure 5. Experimental verification of feature genes in GC and normal gastric epithelial cells**  
(A–L) RT-qPCR was used to measure the expression of C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MNDA, MPEG1, PLEK, and SRGN in GES-1, AGS, SGC7901 and MKN28 cells. ns, not significant; **\(P<0.01\), ****\(P<0.0001\).

**Figure 6. C5AR1 knockdown prevents GC cell growth**  
(A–D) Western blot analysis of C5AR1 expressions in AGS, SGC7901 and MKN28 cells transfected with si-NC or si-C5AR1. (E–G) CCK-8 assays for 0, 24, 48, 72 and 96 h-cell viability in AGS, SGC7901 and MKN28 cells transfected with si-NC or si-C5AR1. ****\(P<0.0001\).

**Figure 7. C5AR1 knockdown elicits intracellular ROS accumulation**  
(A–D) Flow cytometry for estimation of DCFH-DA-labelled ROS levels in AGS, SGC7901 and MKN28 cells transfected with si-NC or si-C5AR1. **\(P<0.01\), and ***\(P<0.001\).

**Figure 8. C5AR1 suppression impairs the aggressiveness of GC cells**  
(A) Representative 0- and 48-h wound healing images of AGS, SGC7901 and MKN28 cells transfected with si-NC or si-C5AR1. Bar, 500 \(\mu\)m. (B–D) Quantification of the relative wound healing distance in the three GC cell lines transfected with si-NC or si-C5AR1. **\(P<0.01\), and ****\(P<0.0001\).

**Figure 9. C5AR1 inhibition impairs the EMT phenotype in GC cells**  
(A,B) Immunofluorescence analysis of E-cadherin and Vimentin expression in AGS, SGC7901 and MKN28 cells transfected with si-NC or si-C5AR1. Scale bar, 50 \(\mu\)m.
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>HGNC symbol</th>
<th>Entrez Gene</th>
<th>Ensembl</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT1</td>
<td>207</td>
<td>ENSG00000142208</td>
<td></td>
</tr>
<tr>
<td>AKT2</td>
<td>208</td>
<td>ENSG00000105221</td>
<td></td>
</tr>
<tr>
<td>ATG7</td>
<td>10533</td>
<td>ENSG00000197548</td>
<td></td>
</tr>
<tr>
<td>CLEC6A</td>
<td>93978</td>
<td>ENSG00000205846</td>
<td></td>
</tr>
<tr>
<td>CSF3</td>
<td>1440</td>
<td>ENSG00000108342</td>
<td></td>
</tr>
<tr>
<td>CTSG</td>
<td>1511</td>
<td>ENSG00000100448</td>
<td></td>
</tr>
<tr>
<td>CYBB</td>
<td>1536</td>
<td>ENSG00000165168</td>
<td></td>
</tr>
<tr>
<td>DNAS1E</td>
<td>1773</td>
<td>ENSG00000213938</td>
<td></td>
</tr>
<tr>
<td>ELANE</td>
<td>1991</td>
<td>ENSG00000197561</td>
<td></td>
</tr>
<tr>
<td>ENTPD4</td>
<td>14573</td>
<td>ENSG00000197217</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>2152</td>
<td>ENSG00000117525</td>
<td></td>
</tr>
<tr>
<td>HMGB1</td>
<td>3146</td>
<td>ENSG00000189403</td>
<td></td>
</tr>
<tr>
<td>IL17A</td>
<td>3605</td>
<td>ENSG00000112115</td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>3553</td>
<td>ENSG00000125538</td>
<td></td>
</tr>
<tr>
<td>IL6</td>
<td>3569</td>
<td>ENSG00000136244</td>
<td></td>
</tr>
<tr>
<td>IL8</td>
<td>3576</td>
<td>ENSG00000169429</td>
<td></td>
</tr>
<tr>
<td>IRAK4</td>
<td>51135</td>
<td>ENSG00000198801</td>
<td></td>
</tr>
<tr>
<td>ITGAM</td>
<td>3684</td>
<td>ENSG00000169896</td>
<td></td>
</tr>
<tr>
<td>ITGAM</td>
<td>3689</td>
<td>ENSG00000160255</td>
<td></td>
</tr>
<tr>
<td>KCNN3</td>
<td>3782</td>
<td>ENSG00000143603</td>
<td></td>
</tr>
<tr>
<td>MAPK1</td>
<td>5594</td>
<td>ENSG00000100030</td>
<td></td>
</tr>
<tr>
<td>MAPK3</td>
<td>5595</td>
<td>ENSG00000102882</td>
<td></td>
</tr>
<tr>
<td>MMP9</td>
<td>4218</td>
<td>ENSG00000100985</td>
<td></td>
</tr>
<tr>
<td>MPO</td>
<td>4353</td>
<td>ENSG00000005381</td>
<td></td>
</tr>
<tr>
<td>MTOF</td>
<td>2475</td>
<td>ENSG00000198793</td>
<td></td>
</tr>
<tr>
<td>PADI4</td>
<td>23569</td>
<td>ENSG00000159339</td>
<td></td>
</tr>
<tr>
<td>PTAFR</td>
<td>5724</td>
<td>ENSG00000169403</td>
<td></td>
</tr>
<tr>
<td>PIK3CA</td>
<td>5290</td>
<td>ENSG00000121879</td>
<td></td>
</tr>
<tr>
<td>RIPK1</td>
<td>8737</td>
<td>ENSG00000137275</td>
<td></td>
</tr>
<tr>
<td>RIPK3</td>
<td>11035</td>
<td>ENSG00000129465</td>
<td></td>
</tr>
<tr>
<td>SELP</td>
<td>6403</td>
<td>ENSG00000174175</td>
<td></td>
</tr>
<tr>
<td>SELPLG</td>
<td>6404</td>
<td>ENSG00000110876</td>
<td></td>
</tr>
<tr>
<td>SIGLEC14</td>
<td>10049587</td>
<td>ENSG00000254415</td>
<td></td>
</tr>
<tr>
<td>TLR2</td>
<td>7097</td>
<td>ENSG00000137462</td>
<td></td>
</tr>
<tr>
<td>TLR4</td>
<td>7099</td>
<td>ENSG00000136869</td>
<td></td>
</tr>
<tr>
<td>TLR7</td>
<td>51284</td>
<td>ENSG00000196664</td>
<td></td>
</tr>
<tr>
<td>TLR8</td>
<td>51311</td>
<td>ENSG00000101916</td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>7124</td>
<td>ENSG00000232810</td>
<td></td>
</tr>
</tbody>
</table>
Description
(coding protein)
AKT serine/threonine kinase 1
AKT serine/threonine kinase 2
Autophagy related 7
Dectin-2
Granulocyte colony stimulating factor
Cathepsin G
NADPH oxidase
Deoxyribonuclease I
Neutrophil elastase
Ectonucleosome Triphosphate Diphosphohydrolase 4
Coagulation Factor III, tissue factor
High mobility group box 1
Interleukin 17
Interleukin 1 beta
Interleukin 6
Interleukin 8
Interleukin 1 receptor associated kinase 4
Complement component 3 receptor 3 subunit
Complement component 3 receptor 3 and 4 subunit
Potassium channel, calcium activated
Mitogen-activated protein kinase 1
Mitogen-activated protein kinase 3
Matrix metalloprotease 9
Myeloperoxidase
Mechanistic target of rapamycin kinase
Peptidyl arginine deiminase 4
Platelet activation factor receptor
Phosphatidylinositol-4,5-bisphosphate 3-kinase
Receptor interacting serine/threonine kinase 1
Receptor interacting serine/threonine kinase 3
P-selectin
P-selectin receptor
Sialic acid binding Ig like lectin 4
Toll like receptor 2
Toll like receptor 4
Toll like receptor 7
Toll like receptor 8
Tumor necrosis factor-alpha
Table 1. Sequences of primers and siRNAs used in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence (5′→3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5AR1</td>
<td>F: TCCTTCAATTATACCACCCCTGA R: ACGCAGCGTGTAGAAGTTTTAT</td>
</tr>
<tr>
<td>CSF1R</td>
<td>F: GGGAACTCCAGTGATAGAGCC  R: TTGGAAGGTAGCTTGTGCTTG</td>
</tr>
<tr>
<td>CSF2RB</td>
<td>F: AGCGGGCTTCAGGACTCTTTG  R: CTGGGCATAGGTGCTCTG</td>
</tr>
<tr>
<td>CYBB</td>
<td>F: ACCGGGTTCATGATATTCCACCT R: GATTTCCAGACTGGAAGGA</td>
</tr>
<tr>
<td>HCK</td>
<td>F: CCCTGTATGATTACGAGCCCA  R: CACTCCCCAGATTCTCTAGG</td>
</tr>
<tr>
<td>ITGB2</td>
<td>F: TCGGCCTCTCTCTCAGGAGT   R: GTTCCCATGATTGGCTAGCC</td>
</tr>
<tr>
<td>LILRB2</td>
<td>F: GCATCTTGGATTACACCGGATCG R: CTGACAGCCATATCGCCTG</td>
</tr>
<tr>
<td>MNDA</td>
<td>F: AACTGACATCAGGGAAGCAAGAG R: CCTGATTCCGAGTAAACGAAG</td>
</tr>
<tr>
<td>MPEG1</td>
<td>F: CGGCAGCATGGGCTAAATCA   R: TGTCCACATTCGCAGATTTG</td>
</tr>
<tr>
<td>PLEK</td>
<td>F: AAGAAAGGGGACGCTGGTTCAAT R: TCAGCGGATCATCCTTGG</td>
</tr>
<tr>
<td>SRGN</td>
<td>F: AGGTTATCTACGCGGAGAG    R: GTCTTTGAAAAAGGTCAGTCT</td>
</tr>
<tr>
<td>STAB1</td>
<td>F: AACCACGTTCGGCACTCTAGT  R: CGGCAGTCTCGGTTATCTG</td>
</tr>
<tr>
<td>GAPDH</td>
<td>F: ACAACTTTGGATCTGGGAGG   R: GCCATACGCGCAGTTCGT</td>
</tr>
<tr>
<td>si-NC</td>
<td>UUCUCCGAACGUGUCACGUTT</td>
</tr>
<tr>
<td>si-C5AR1 (1)</td>
<td>GCUUUCUGCUUGUGUUAAC</td>
</tr>
<tr>
<td>si-C5AR1 (2)</td>
<td>GGUCCACCAAGACACUCAAGG</td>
</tr>
<tr>
<td>si-C5AR1 (3)</td>
<td>GCUCCUGUAAGUGAGUUAUU</td>
</tr>
</tbody>
</table>

Table 2. Analysis of the efficacy of five machine learning classifiers in differentiating two NET-based clusters

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classifier</th>
<th>AUC</th>
<th>95%CI</th>
<th>P</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA-STAD</td>
<td>linearSVM</td>
<td>0.969</td>
<td>0.955–0.983</td>
<td>0</td>
<td>0.958</td>
<td>0.859</td>
</tr>
<tr>
<td>TCGA-STAD</td>
<td>RBFsvm</td>
<td>0.966</td>
<td>0.949–0.983</td>
<td>0</td>
<td>0.948</td>
<td>0.87</td>
</tr>
<tr>
<td>TCGA-STAD</td>
<td>RandomForest</td>
<td>1</td>
<td>1–1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TCGA-STAD</td>
<td>GaussianNB</td>
<td>0.949</td>
<td>0.929–0.968</td>
<td>0</td>
<td>0.874</td>
<td>0.87</td>
</tr>
<tr>
<td>Dataset</td>
<td>Algorithm</td>
<td>Weight (Mean)</td>
<td>Weight (Min–Max)</td>
<td>p-value</td>
<td>SE (Mean)</td>
<td>SE (Min–Max)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>---------------</td>
<td>------------------</td>
<td>---------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>TCGA-STAD</td>
<td>Logistic</td>
<td>0.97</td>
<td>0.956–0.984</td>
<td>0</td>
<td>0.885</td>
<td>0.924</td>
</tr>
<tr>
<td>GSE88433</td>
<td>linearSVM</td>
<td>0.798</td>
<td>0.757–0.839</td>
<td>2.09×10^-46</td>
<td>0.595</td>
<td>0.845</td>
</tr>
<tr>
<td>GSE88433</td>
<td>RBFsvm</td>
<td>0.785</td>
<td>0.742–0.829</td>
<td>8.88×10^-38</td>
<td>0.713</td>
<td>0.769</td>
</tr>
<tr>
<td>GSE88433</td>
<td>RandomForest</td>
<td>0.818</td>
<td>0.779–0.857</td>
<td>2.22×10^-57</td>
<td>0.779</td>
<td>0.727</td>
</tr>
<tr>
<td>GSE88433</td>
<td>GaussianNB</td>
<td>0.838</td>
<td>0.801–0.874</td>
<td>1.76×10^-73</td>
<td>0.749</td>
<td>0.773</td>
</tr>
<tr>
<td>GSE88433</td>
<td>Logistic</td>
<td>0.806</td>
<td>0.766–0.846</td>
<td>1.66×10^-50</td>
<td>0.738</td>
<td>0.723</td>
</tr>
<tr>
<td>GSE88437</td>
<td>linearSVM</td>
<td>0.791</td>
<td>0.745–0.837</td>
<td>1.44×10^-35</td>
<td>0.545</td>
<td>0.891</td>
</tr>
<tr>
<td>GSE88437</td>
<td>RBFsvm</td>
<td>0.774</td>
<td>0.726–0.823</td>
<td>3.39×10^-28</td>
<td>0.724</td>
<td>0.736</td>
</tr>
<tr>
<td>GSE88437</td>
<td>RandomForest</td>
<td>0.803</td>
<td>0.758–0.847</td>
<td>1.98×10^-40</td>
<td>0.667</td>
<td>0.776</td>
</tr>
<tr>
<td>GSE88437</td>
<td>GaussianNB</td>
<td>0.829</td>
<td>0.788–0.871</td>
<td>9.06×10^-55</td>
<td>0.712</td>
<td>0.811</td>
</tr>
<tr>
<td>GSE88437</td>
<td>Logistic</td>
<td>0.796</td>
<td>0.751–0.842</td>
<td>3.42×10^-37</td>
<td>0.577</td>
<td>0.876</td>
</tr>
</tbody>
</table>

**Table 3. The weight of each feature gene in the logistic model**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5AR1</td>
<td>1.301465</td>
</tr>
<tr>
<td>CSF1R</td>
<td>0.798827</td>
</tr>
<tr>
<td>CSF2RB</td>
<td>1.691757</td>
</tr>
<tr>
<td>CYBB</td>
<td>−0.93195</td>
</tr>
<tr>
<td>HCK</td>
<td>0.176422</td>
</tr>
<tr>
<td>ITGB2</td>
<td>0.370265</td>
</tr>
<tr>
<td>LILRB2</td>
<td>0.393421</td>
</tr>
<tr>
<td>MNDA</td>
<td>0.055314</td>
</tr>
<tr>
<td>MPEG1</td>
<td>−0.74223</td>
</tr>
<tr>
<td>PLEK</td>
<td>0.56748</td>
</tr>
<tr>
<td>SRGN</td>
<td>0.51999</td>
</tr>
<tr>
<td>STAB1</td>
<td>0.911245</td>
</tr>
</tbody>
</table>
Supplementary Figure S1. Heterogeneous therapeutic responses in the two NET-based clusters

(A) Differential IC50 values of small-molecule agents between clusters. The threshold was set as $|t|>10$ and $P<0.05$. (B) Distribution of TMB across TCGA-STAD tumor types. The median TMB was 2.3/MB. (C) Comparison of TMB between the two clusters. (D,E) Differential T-cell-inflamed and TIDE scores between clusters. ****$P<0.0001$. 
Supplementary Figure S2. Heterogeneous expressions of C5AR1, CSF1R, CSF2RB, CYBB, HCK, ITGB2, LILRB2, MND, MPEG1, PLEK, and SRGN in two NET-based clusters. **** \( p < 0.0001 \).

Supplementary Table S1. Information on 35 NET-associated genes
**Highlights**

Neutrophil extracellular traps (NETs) are implicated in gastric cancer (GC) growth, metastatic dissemination, cancer-associated thrombosis, etc. This work was conducted for elucidating the heterogeneity of NETs in GC via consensus clustering algorithm, with subsequent external verification. These findings collectively provide a theoretical basis for anti-NETs in GC treatment.

1. To reflect the transcriptome heterogeneity of NETs in GC, we proposed a novel NETs-based classification called consensus clustering algorithm.
2. NETs function as a main contributor to intra-tumoral heterogeneity and differential drug sensitivity in GC.
3. C5AR1 is evidenced to trigger GC growth and metastatic spread.