Tumour Marker Prognostic Study

INTERNATIONAL JOURNAL OF SURGERY

OPEN

Prediction of hematogenous metastasis risk from molecular classification in gastric cancer

Seungho Lee, MD, PhD[a], Jaeun Yoo, BS[b], Seungbok Lee, MD, PhD[c], Sujin Oh, MS[d], Hyun Myong Kim, MS[b], Kyoungyun Jeong, MS[b], Yie-Ri Yoo, MS[b], Ji-Yeon Shin, BS[b], Hye Seung Lee, MD, PhD[b,e,*], Kyoung Un Park, MD, PhD[f], Seong-Ho Kong, MD, PhD[a,b], Do Joong Park, MD, PhD[a,b,*], Hyuk-Joon Lee, MD, PhD[a,b], Han-Kwang Yang, MD, PhD[a,b]

Background: Hematogenous metastasis (HM) poses a significant challenge in the treatment and outcome of gastric cancer (GC). This study identified and characterized molecular subtypes of GC associated with HM to develop predictive models and enhance precision medicine.

Methods: We analyzed bulk-RNA sequencing data from 64 micro-dissected primary GC samples to identify distinct molecular subtypes and validated our findings using a patient-derived xenograft (PDX) model. Signature genes associated with HM were identified using a machine-learning survival model and validated across multiple external cohorts. To evaluate potential therapeutic agents for high-risk HM, we analyzed data from the Cancer Cell Line Encyclopedia.

Results: Two molecular subtypes were identified based on gene expression profiles in GC: stemness and gastric subtypes. The stemness subtype was associated with a significantly increased risk of HM. Validation in the PDX model confirmed accelerated HM in the stemness subtype. The machine survival learning model and differential gene expression analysis identified 10 stemness signature genes and 7 gastric signature genes, which stratified HM risk groups based on a 17-gene signature score. High-risk HM groups exhibited significantly decreased hematogenous metastasis-free survival in three external cohorts.

Conclusion: Stemness subtype gastric cancer is associated with an elevated risk of HM. The identification of molecular characteristics and the development of prediction tools offer practical tools in the clinical field and have a high potential for precision medicine.

Keywords: gastric cancer, hematogenous metastasis, molecular subtype, stemness

Introduction

Gastric cancer (GC), the fifth most common malignancy worldwide, presents significant challenges in clinical management owing to its heterogeneity and poor prognosis[1]. Despite recent advances in surgical technology, chemotherapy, and immunotherapy for GC, the outcomes of patients with advanced GC remain low[2]. Patient survival significantly decreases when tumor cells are detected in distant organs or the peritoneal cavity[3,4]. Although understanding the mechanisms underlying metastasis is crucial for developing new treatment modalities, the heterogeneity of GC hinders the unveiling of progression mechanisms. Several studies on the molecular profiling of GC have recently reported its molecular heterogeneity; however, only a few have demonstrated the molecular characteristics associated with poor survival[5–7]. For example, although the

HIGHLIGHTS

- A molecular subtype favoring hematogenous metastasis in gastric cancer was identified.
- A machine learning-based single patient risk classifier using a 17-gene signature predicts hematogenous metastasis after curative gastrectomy.
- Potential therapeutic agents for high-risk hematogenous metastasis gastric cancer were discovered.

[a]Department of Surgery, Seoul National University Hospital, Seoul, Republic of Korea, [b]Cancer Research Institute, Seoul National University, Seoul, Republic of Korea, [c]Department of Genomic Medicine, Seoul National University Hospital, Seoul, Republic of Korea, [d]Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea, [e]Department of Pathology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea and [f]Department of Laboratory Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Seungho Lee and Jaeun Yoo contributed equally.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding authors. Address: Department of Surgery and Cancer Research Institute, Seoul National University College of Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. Tel.: +82-2-2072-1490. E-mail: djparkmd@snu.ac.kr (D. J. Park); Department of Pathology and Cancer Research Institute, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. Tel.: +82-2-740-8269. E-mail: hye2@snu.ac.kr (H.S. Lee).

International Journal of Surgery (2025) 111:5311–5324

Received 14 February 2025; Accepted 15 May 2025

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.lww.com/international-journal-of-surgery.

Published online 28 May 2025

http://dx.doi.org/10.1097/JS9.0000000000002587

Cancer Genome Atlas (TCGA) group has proposed four molecular subtypes of GC based on multi-omics analysis, their findings had no discernible clinical relevance[5]. Additionally, the Asian Cancer Research Group (ACRG) reported four molecular subtypes and demonstrated distinct survival outcomes across these subtypes[6]. However, none of the studies revealed an association between the molecular characteristics and metastasis of GC.

Gastric cancer (GC) exhibits three major metastatic pathways: lymphatic, hematogenous, and peritoneal dissemination[8]. Lymphatic metastasis, involving regional lymph nodes, is the most common and often the earliest route of spread. Hematogenous metastasis refers to the dissemination of tumor cells through the bloodstream to distant organs such as the liver and lungs. In contrast, peritoneal metastasis involves the transcoelomic spread of tumor cells within the peritoneal cavity, often resulting in extensive local dissemination and ascites, which severely impact patient quality of life and survival. These metastatic patterns reflect fundamentally distinct mechanisms of disease progression. Therefore, it is evident that different treatment approaches should be established according to metastasis patterns, which should be preceded by an understanding of the molecular characteristics of both.

In this study, we analyzed bulk-RNA seq data from microdissected primary GC samples and identified a molecular subtype and its molecular characteristics that favor hematogenous metastasis. These results were validated using an independent patient-derived xenograft (PDX) model. To identify molecular features associated with hematogenous metastasis, we developed a machine-learning survival model using the gradient-boosted algorithm. We discovered signature gene sets in the primary tumor and validated their prognostic significance associated with hematogenous metastasis in three independent GC cohorts. These results provide new insights into the molecular characteristics associated with hematogenous metastasis and target discovery for future therapeutic exploitation.

Materials and methods

This study was conducted and reported in accordance with the REMARK guidelines for prognostic biomarker studies[9].

Patients and sample preparations

GC samples were obtained from patients who received curative gastrectomy following the Declaration of Helsinki. The Institutional Review Board approved the study (approval number: 2111-093-1272). Total RNA was extracted from formalin-fixed paraffin-embedded (FFPE) blocks stored in the pathology department. Hematoxylin–eosin-stained slides produced using FFPE blocks of primary tumor tissues were reviewed and annotated by an expert pathologist to select appropriate areas with sufficient tumor cellularity and minimal contamination from benign cells for microdissection. Subsequently, 8-μm thick FFPE sections were deparaffinized using the deparaffinization solution (Qiagen, Germantown, MD, USA). Total RNA was extracted using the RNeasy FFPE Kit (Qiagen) according to the manufacturer's protocols and quantified using the Qubit 1X dsDNA high-sensitivity assay kit (Qiagen). The RNA integrity was assessed using the DV200 value (%), which represents the percentage of RNA fragments of > 200 nucleotides, using the Agilent RNA 6000 nano kit (Agilent Technologies, USA) with the 2100 bioanalyzer system (Agilent).

Whole transcriptome sequencing (WTS)

WTS was conducted to examine the gene expression profiles of tumor tissues. Libraries were generated using the Illumina stranded total RNA prep with ribo-zero plus kit (Illumina, San Diego, CA, USA). After rRNA depletion using the rRNA-targeted DNA probes, first-strand cDNA was synthesized using random hexamers, followed by second-strand synthesis with deoxyuridine triphosphate. Subsequently, the cDNA was subjected to adapter ligation using the Integrated DNA Technology for Illumina DNA/RNA Unique Dual Indexes (Illumina) and enriched with polymerase chain reaction. The final libraries were evaluated using the Agilent DNA 1000 kit (Agilent) on the 2100 bioanalyzer system (Agilent) and sequenced with a 75-bp paired-end run on a NextSeq 550 instrument (Illumina). Next, the sequencing quality was evaluated using FastQC, and adapter and over-represented sequences were trimmed using Trimmomatic. Trimmed reads were aligned to the human reference genome (GRCh38) using STAR aligner v.2.7.10 with Ensembl annotations. The gene count matrix was generated using the quantMode option of the STAR aligner. mRNA-based gene expression was analyzed as log2(TPM + 1), where TPM represents transcripts per million.

Generation of PDX models and RNA sequencing

For PDX model generation, we used same the method described by Na *et al*[10]. Tissues obtained from surgical resections were cut into approximately 2-mm fragments and implanted subcutaneously into the flanks of 6-week-old female NOD/SCID/IL-2γ-receptor null NOD scid gamma (NSG) mice (The Jackson Laboratory, Farmington, CT, USA). Tumor formations exceeding 500 mm^3 at the implantation site were considered successful engraftments. Mice with successful engraftments were sacrificed, and the tumor tissues were harvested and preserved. Total RNA was extracted from non-tumor tissues, patient tumors, and PDX tumors using TRIzol™ (Invitrogen). Only samples with an RNA integrity number > 5 were further processed. The 101-bp paired-end libraries were created using the TruSeq RNA Sample prep kit v2 (Illumina) with 1 μg RNA. WTS and raw count matrix generation were conducted using the same methods as for primary tumor WTS data generation.

Profiling molecular subtype and characteristics of GC

To identify molecular subtypes in GC, we performed non-negative matrix factorization (NMF) using the NMF R package. Log2-normalized TPM + 1 data were used for clustering, and the input genes were determined by selecting genes that showed over 2 standard deviations across the samples. The stability of clustering associated with a given *k* cluster was measured using the cophenetic correlation coefficient[11]. Differentially expressed genes (DEGs) were identified using Student's *t*-test. The normalized gene expression profile was used to perform single-sample gene set variation analysis (GSVA) using the GSVA R package (v.1.42.0). The following parameters were

used to evaluate the GSVA function: mx.diff = TRUE, kcdf = "Gaussian," min.sz = 1, max.sz = Inf, tau = 0.25, ssgsea_norm = TRUE, and abs.ranking = FALSE. To analyze the gene network, the STRING database[12] was used to identify protein interactions among input DEGs (P-value < 10^{-4}, t-test) corresponding to different subtypes. Cytoscape software was used to visualize the gene network result. For the transcription factor analysis, we used the Decoupler Python package (v. 1.6.0) with default parameters.

Estimation of tumor microenvironment (TME) cells in bulk-level transcriptome data

To calculate immune, endothelial, and epithelial cells and fibroblast fractions in the primary GC transcriptome and PDX host transcriptome dataset, we used CIBERSORTx for relative cell fractions[13]. For signature matrix generation, the single-cell RNA count matrix from GSE183904 was used, as it represents the largest publicly available gastric cancer single-cell dataset. Unlike pre-defined LM22 or TR4 signatures, which are derived from non-gastric tissues or immune cell-focused contexts, this dataset enabled the construction of a gastric cancer-specific signature matrix encompassing four major cell types: epithelial, stromal, endothelial, and immune cells[14]. To derive the TME's four-cell-type signature, we randomly selected 50 000 cells, encompassing immune, epithelial, endothelial, and fibroblast cells. We set CIBERSORTx to single-cell mode with 100 permutations, a q-value of 0.01, a G.min of 300, and a G.max of 1 000. For estimating the TME component cell in PDX host transcriptome data, we converted the mouse gene symbols into their human counterparts.

PDX transcriptome data analysis

To separate human and mouse NGS reads from the bulk-RNA sequencing FASTQ data of PDX samples, we used Xengsort[15] with a default setting. Human-specific reads were aligned and quantified in the same manner as for the primary GC transcriptome, and mouse-specific reads were aligned to mm38 and quantified similarly. For the subtype classification of PDX graft (human) transcriptome data, we used the Pycaret Python library (3.0.4). Transcriptome data from our institution were used as a training dataset. Input genes in the training data were determined by selecting those with P-values of Student's t-test between stemness and gastric subtypes lower than 10^{-6}. Transcriptome data were randomly divided in a 7:3 ratio for model training and testing. The model was tuned using the area under the curve (AUC). Finally, PDX graft transcriptome data were classified using the constructed model.

Machine learning survival analysis

To construct the machine learning survival model that predicts hematogenous metastasis-free survival (HMFS), we used gradient boost models using the scikit-survival Python library (v. 0.22.2). For the comparison of the c-index between survival model, we also used the Cox proportional model and the random forest model by using the scikit-survival Python library. For the generation of the survival model, we first calculated the standard variation of each gene and then normalized the gene expression dataset with the mean and standard variation. Next, we randomly divided the training and test datasets in a 7:3 ratio and calculated the c-index according to the standard variation cut-off value. The following parameters were used to generate the survival model: learning rate = 1.0; max_depth = 1; random_state = 0. Finally, we generated the survival model with genes whose expression showed over 1.0 standard deviation between samples. PDX dataset validation was performed using PDX graft transcriptome data with the same genes that were used as input data for model generation among the stage I, II, and III samples. GSVA, using predictive genes, was performed using Fisher's exact test as follows:

$$Odds\ Ratio = \frac{Predictive\ genes\ (AUC > 0.6)\ in\ each\ gene\ set}{Non{-}predictive\ genes\ (AUC < 0.6)\ in\ each\ gene\ set}$$

Generation and validation of hematogenous metastasis risk score

Among the highly predictive genes in the gradient boost survival model (mean time-dependent AUC > 0.8), we selected 10 genes that showed significantly elevated levels in the stemness subtype (P < 0.05, t-test) as the stemness signature genes. For gastric subtype signature, we selected genes that showed lower mean time-dependent AUC < 0.2 and showed significantly elevated expression in the gastric subtype ($P < 10^{-4}$, t-test). Stemness and gastric scores were calculated using the single-sample gsva (ssGSVA) function in the GSVA R package. Finally, we defined the hematogenous metastasis risk score as the difference between the stemness and gastric scores.

$$\text{Hematogenous metastasis risk score} = \text{stemness signature gsva score} - \text{gastric signature gsva score}$$

For defining the cut-off value of the hematogenous metastasis risk score to identify hematogenous metastasis risk groups, we conducted K-means clustering for the stemness and gastric scores from GSE66229 microarray data using the scikit-learn Python library (version 1.3.1).

External data

Gene expression data from TCGA were downloaded from the University of California, Santa Cruz Genomics Institute (https://xenabrowser.net/). Microarray data were obtained from the gene expression omnibus database (GSE66229 and GSE84437). Gene expression, PRISM repurposing drug screening data, and sample meta data of the Cancer Cell Line Encyclopedia (CCLE) were downloaded from DepMap (https://depmap.org, Public 22Q4 + and 24Q2 for PRISM data).

Survival and statistical analyses

All statistical analyses were primarily performed using the SciPy library of Python. The Lifelines Python library was used for all survival analyses, including the log-rank test and multivariate Cox proportional-hazard analysis. Statistical significance was defined as a two-sided P-value < 0.05. For TCGA dataset survival analysis, we analyzed 63 samples with complete records of progression-free interval time and metastasis site information. For the survival analysis of GSE84437, we analyzed 330 samples with complete records of metastasis information including metastasis sites and dates.

The hazard ratios were calculated using the multivariate Cox proportional hazard analysis model. For the survival analysis, we excluded stage IV, i.e., metastasis before the curative gastrectomy. Hematogenous metastasis was defined as recurrence at distant organs, including the liver, bone, adrenal gland, and lung. Metastasis to the peritoneal cavity and ovary was defined as peritoneal metastasis.

Result

Identification of molecular subtypes of primary GC

We conducted unsupervised clustering of bulk RNA-seq data from 64 micro-dissected primary gastric cancer (GC) specimens using non-negative matrix factorization (NMF). The optimal number of clusters was determined to be two, based on the highest cophenetic coefficient (Figure S1A. http://links.lww.com/JS9/E243), resulting in two distinct molecular subtypes (Fig. 1A, Figures S1B. http://links.lww.com/JS9/E243 and S2A. http://links.lww.com/JS9/E243). One cluster exhibited the elevated expression of canonical cancer stem cell-associated genes, including *STMN1, NES,* and *BMI1* ($P = 1.20 \times 10^{-4}$, 1.70×10^{-2}, and 2.18×10^{-2}, *t*-test, respectively; Fig. 1B, Figure S1C. http://links.lww.com/JS9/E243). This cluster also demonstrated significant enrichment of stemness-related gene sets according to single-sample gene set variation analysis (ssGSVA) ($P = 7.62 \times 10^{-3}$ and 5.69×10^{-5}, *t*-test, respectively; Fig. 1B) [16]. The other cluster showed upregulation of gastric mucosal lineage markers such as *TFF1, MUC5AC*, and *PSCA* ($P = 9.76 \times 10^{-5}$, 5.99×10^{-6}, and 2.76×10^{-3}, *t*-test, respectively; Fig. 1B, Figure S1C. http://links.lww.com/JS9/E243) [17,18]. Based on these transcriptional characteristics, we defined the first group as the stemness subtype and the second as the gastric subtype (Fig. 1A).

Molecular characteristics of stemness subtype GC

To identify molecular characteristics of the stemness subtype in GC, we performed gene network analysis using differentially expressed genes from the stemness subtype (Figure S2A. http://links.lww.com/JS9/E243). Notably, members of the melanoma-associated antigen A *(MAGEA)* family, such as *MAGEA3, MAGEA6, MAGEA10*, and *MAGEA12*, were significantly upregulated in the stemness subtype ($P = 1.62 \times 10^{-13}$, 2.15×10^{-16}, 6.24×10^{-10}, and 4.11×10^{-12}, *t*-test, respectively, Figure S2B. http://links.lww.com/JS9/E243). Moreover, the expression of the MAGEA family members exhibited a binary pattern; in the stemness subtype, these genes were highly expressed in most samples, whereas in the gastric subtype, their expression was minimal or absent (Figure S2B. http://links.lww.com/JS9/E243). Additionally, the expression of *GABRA3*, which shares a promoter with *MAGEA6*, was significantly increased in the stemness subtype ($P = 1.61 \times 10^{-13}$, *t*-test, Figure S2B. http://links.lww.com/JS9/E243) [19]. Furthermore, in the transcription factor activity analysis, *FOXM1*, a master regulator of tumor metastasis, was the most activated transcription factor in stemness subtype GC[20–22] (Fig. 1C). These findings suggested that the stemness subtype of gastric cancer exhibits a dedifferentiated transcriptional state, characterized by the reactivation of embryonic or cancer-testis antigens.

In gene set variation analysis (gsva), single sample gsva (ssgsva) scores of the hallmark gene set, associated with a cell cycle, were significantly elevated in stemness GC, including E2F targets and G2M checkpoint ($P = 9.76 \times 10^{-3}$ and 7.15×10^{-3}, *t*-test, respectively; Fig. 1D) [17]. We found that angiogenesis-related genes were significantly upregulated in the stemness subtype of GC ($P = 1.24 \times 10^{-3}$, *t*-test; Fig. 1D). This suggests an enhanced capability of the tumor to develop new blood vessels to support tumor growth and metastasis. To ensure that this observation was not due to contamination from the TME in our bulk-RNA sequencing samples, we used CIBERSORTx to deconvolute the fractions of four major cell types within the TME, namely epithelial and immune cells, fibroblasts, and endothelial cells. The analysis showed no significant differences in the fractions of these cell types between the stemness and gastric subtypes, confirming that the observed upregulation of angiogenesis genes was intrinsic to the cancer cells ($P = 0.81$, 0.59, 0.79, and 0.91, *t*-test, respectively; Figure S3A. http://links.lww.com/JS9/E243). Thus, we inferred that the elevated angiogenesis observed was not due to contamination by endothelial cells in the micro-dissected samples but rather from the cancer transcriptome. Therefore, the stemness subtype of GC is characterized by a dedifferentiated transcriptional state, marked by stem-like features, enhanced proliferative activity, angiogenic potential, and high metastatic propensity.

Decreased hematogenous metastasis-free survival in stemness subtype GC

Next, we analyzed the clinicopathologic characteristics of the 64 GC patients (Table S1. http://links.lww.com/JS9/E244). Between the two molecular subtypes, most clinicopathologic features did not show significant differences, except for pathologic gross subtypes ($P = 0.037$, chi-square test; Fig. 1A; Table 1). The stemness subtype predominantly exhibited non-infiltrated types, macroscopically, such as Borrmann type I and II. Additionally, T, N, and TNM stages did not show significant differences between the two groups ($P = 0.347$, 0.176, and 0.229, respectively, chi-square test; Fig. 1A; Table 1). The history of adjuvant chemotherapy was also similar between the two groups ($P = 0.191$; Fig. 1A; Table 1). Subsequently, we conducted survival analysis and observed a substantial difference in HMFS between the two GC subtypes ($P = 0.005$, log-rank test; Fig. 2A). We found that hematogenous metastasis was accelerated in the stemness subtype GC. However, peritoneal metastasis-free survival ($P = 0.709$, log-rank test; Fig. 2A) and overall survival (OS) ($P = 0.219$, log-rank test; Fig. 2A) did not show significant differences between the two groups. Since all early gastric cancer samples, defined as mucosal or submucosal invasion, were classified into the gastric subtype, we conducted a survival analysis using advanced gastric cancer (AGC) samples to eliminate pathologic stage bias. In the AGC samples, HMFS was also significantly decreased in the stemness subtype ($P = 0.013$, log-rank test; Fig. 2B). Furthermore, peritoneal metastasis-free survival and OS did not show significant differences between the two groups ($P = 0.595$ and 0.333, respectively, log-rank test; Fig. 2B). Finally, multivariable Cox regression analysis revealed that the stemness subtype was the most significant independent prognostic feature for

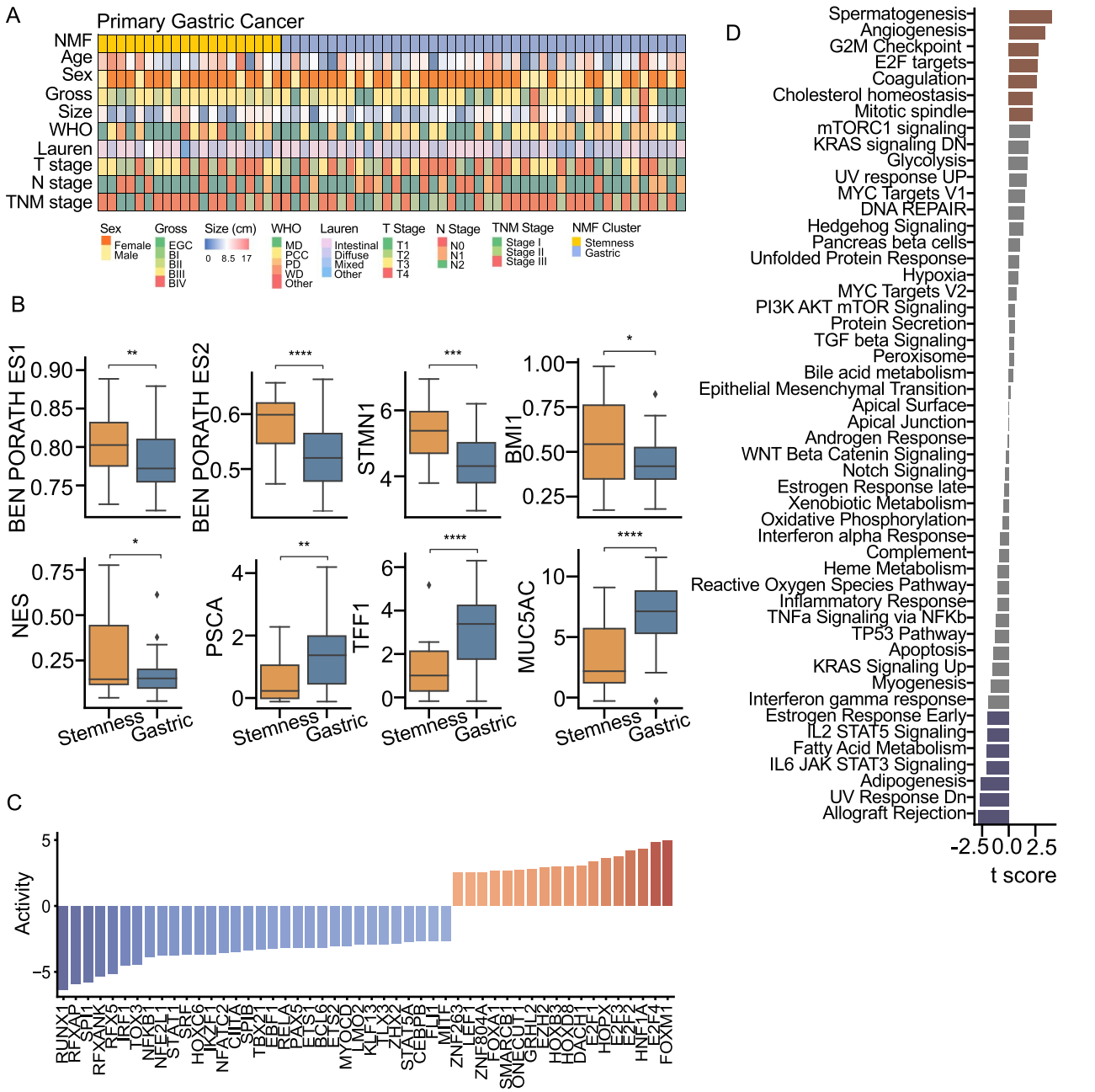

Figure 1. Identification of the molecular subtypes in gastric cancer (GC). (A) Oncoprint table showing the clinicopathologic characteristics of the NMF clusters. Each column represents a patient. Each feature is annotated based on the clinical variable (bottom). (B) Single-sample gene set variation analysis (GSVA) score of Ben Porath ES1 and ES2 gene sets and log2-normalized TPM of STMN1, BMI1, NES, PSCA, TFF1, and MUC5AC genes in the stemness and gastric subtypes. Boxes represent the 25th, 50th, and 75th percentiles, and whiskers show the 10th and 90th percentiles. Asterisk indicates the *P*-value of the *t*-test; *: $0.01 < P \leq 0.05$, **: $0.001 < P \leq 0.01$, ***: $0.0001 < P \leq 0.001$, ****: $P \leq 0.0001$. (C) Bar plot of the normalized enrichment score of the top 50 transcription factors. The normalized enrichment score is colored from blue to red according to the score. (D) Bar plot shows t-values of hallmark gene sets in MsigDB relatively up- or downregulated in the stemness subtype compared with those in the gastric subtype. Dark red bar indicates significantly upregulated gene sets ($P < 0.05$, t-test), and dark blue bar indicates significantly downregulated gene sets ($P < 0.05$, t-test). Grey bars represent no significant differences.

hematogenous metastasis, considering age, sex, and T and N stages, which are well-known prognostic factors after curative gastrectomy (P = 0.008; hazard ratio (HR) = 2.87; Fig. 2C). To evaluate whether TME components influenced HMFS, we conducted a multivariate Cox proportional hazards analysis using representative TME signature genes, including CD3G, PTPRC, THY1, FAP, PECAM1, and PLVAP. None of these genes were found to be significant predictors of HMFS (Figure S3B. http://links.lww.com/JS9/E243). Therefore, we inferred that hematogenous metastasis in GC is driven by tumor-intrinsic molecular characteristics, rather than by components of the tumor microenvironment.

Table 1

Clinico-pathologic characteristics between molecular subtype of gastric cancer

Subtype	Gastric (N = 44)	Stemness (N = 20)	*P*-value
Age	61.4 ± 10.9	65.4 ± 10.7	0.171
Sex			0.663
Male	29 (65.9%)	15 (75.0%)	
Female	15 (34.1%)	5 (25.0%)	
Gross			0.037
BI	1 (2.3%)	2 (10.0%)	
BII	4 (9.1%)	5 (25.0%)	
BIII	26 (59.1%)	13 (65.0%)	
BIV	2 (4.5%)	0 (0.0%)	
EGC	11 (25.0%)	0 (0.0%)	
Size1	5.4 ± 3.5	5.4 ± 2.8	0.986
WHO			0.084
MD	16 (36.4%)	9 (45.0%)	
PCC	16 (36.4%)	4 (20.0%)	
PD	11 (25.0%)	3 (15.0%)	
WD	1 (2.3%)	2 (10.0%)	
Other	0 (0.0%)	2 (10.0%)	
Lauren			0.254
Diffuse	21 (47.7%)	7 (35.0%)	
Intestinal	19 (43.2%)	10 (50.0%)	
Mixed	4 (9.1%)	1 (5.0%)	
Other	0 (0.0%)	1 (5.0%)	
T stage			0.347
I	5 (11.4%)	0 (0.0%)	
II	11 (25.0%)	6 (30.0%)	
III	10 (22.7%)	7 (35.0%)	
IV	18 (40.9%)	7 (35.0%)	
N stage			0.176
N0	8 (18.2%)	6 (30.0%)	
N1	10 (22.7%)	1 (5.0%)	
N2	26 (59.1%)	13 (65.0%)	
TNM			0.229
Stage I	8 (18.2%)	5 (25.0%)	
Stage II	13 (29.5%)	2 (10.0%)	
Stage III	23 (52.3%)	13 (65.0%)	

BI, Borrman type I; BII, Borrman type II; BIII, Borrman type III; BIV, Borrman type IV; EGC, early gastric cancer; MD, moderate differentiated; PCC, poorly cohesive carcinoma; PD, poorly differentated; WD, well differentiated; WHO, World Health Organization.

Validation of accelerated hematogenous metastasis in stemness subtype GC using PDX model

Subsequently, we analyzed independent PDX transcriptome data to validate the accelerated hematogenous metastasis in patients with stemness subtype GC. Additionally, we evaluated the TME characteristics from the mouse transcriptomes. Since human tumor cells are engrafted while stromal and endothelial cells are replaced by their mouse counterparts, the PDX transcriptome sequencing data represent a mixture of tumor-specific and stromal/endothelial-specific transcriptomes derived from human (graft) and mouse (host) cells, respectively[23–25]. First, we performed *in silico* deconvolution transcriptional signal of human malignant cell (graft) components and mouse (host) TME components (Method, Figure S4. http://links.lww.com/JS9/E243). Next, we performed machine-learning classification of PDX graft transcriptome data using primary GC transcriptome data as a training data set (Method, Figure S5A. http://links.lww.com/JS9/E243). The model's classification performance exhibited significantly high, with a receiver operating characteristic (AUC) of 1.0 (Figure S5B. http://links.lww.com/JS9/E243 and S5C. http://links.lww.com/JS9/E243). Finally, cancer-specific transcriptome data from 51 bulk-RNAseq of PDX samples were classified into 14 stemness and 37 gastric subtypes (Table S2. http://links.lww.com/JS9/E244). When patients from the PDX graft expression dataset were stratified according to stemness or gastric subtypes, clinicopathologic characteristics showed similar results, including age, sex, WHO and Lauren classifications, and T, N, and TNM stages (Fig. 3A, Table 2). In survival analysis, the stemness subtype exhibited a significantly decreased HMFS (P = 0.0001, log-rank test; Fig. 3B). In addition, no differences were observed in peritoneal metastasis survival and OS (P = 0.324 and 0.726, log-rank test, respectively; Fig. 3B). To evaluate the TME, we calculated the cellular fraction, including endothelial cells, fibroblasts, and immune cells in the deconvoluted mouse transcriptome data using CIBERSORTx. The fraction of endothelial cells in the host microenvironment showed significantly higher differences in the stemness subtype PDX cancer (P = 0.035, t-test; Fig. 2C); however, immune cell infiltration and fibroblast fraction did not show differences between the two subtypes (P = 0.902 and 0.461, t-test, respectively; Fig. 2C). Thus, PDX data analysis validated decreased hematogenous metastasis survival in stemness subtypes, as well as increased endothelial cell in TME.

Development of machine learning model for prediction of HMFS

To identify the molecular characteristics associated with hematogenous metastasis in GC, we developed a machine-learning survival model using the gradient boost algorithm due to its higher predict performance between three machine learning survival models including Cox's proportional hazard's model, random forest survival model, and gradient boost survival model (Figure S6A. http://links.lww.com/JS9/E243). For model generation, we first divided our bulk-RNAseq samples into 70% for training and 30% for test datasets. To construct the most optimal survival model without supervision, we only used transcriptome data as input data, and input gene selection was determined based on the standard deviation of gene expression across samples by comparing the concordance index (c-index). Notably, the highest c-index was achieved when the cut-off standard deviation was set at 1.0 or 1.2 (Fig. 4A). For final model generation, we excluded genes with a standard deviation of less than 1.0 in their expression among 64 samples. Consequently, the prediction of hematogenous metastasis in our test dataset showed good performance (c-index = 0.732; Fig. 4B). In this survival model, the expression of *CERKL, GABRB2, CYP2C18,* and *CEACAM5* was the top discriminating feature (Figure S6B. http://links.lww.com/JS9/E243). Next, we computed the c-index using the PDX graft transcriptome to validate the hematogenous metastasis survival model. The PDX dataset demonstrated higher hematogenous metastasis prediction performance compared to the internal test dataset (c-index = 0.759; Fig. 4B).

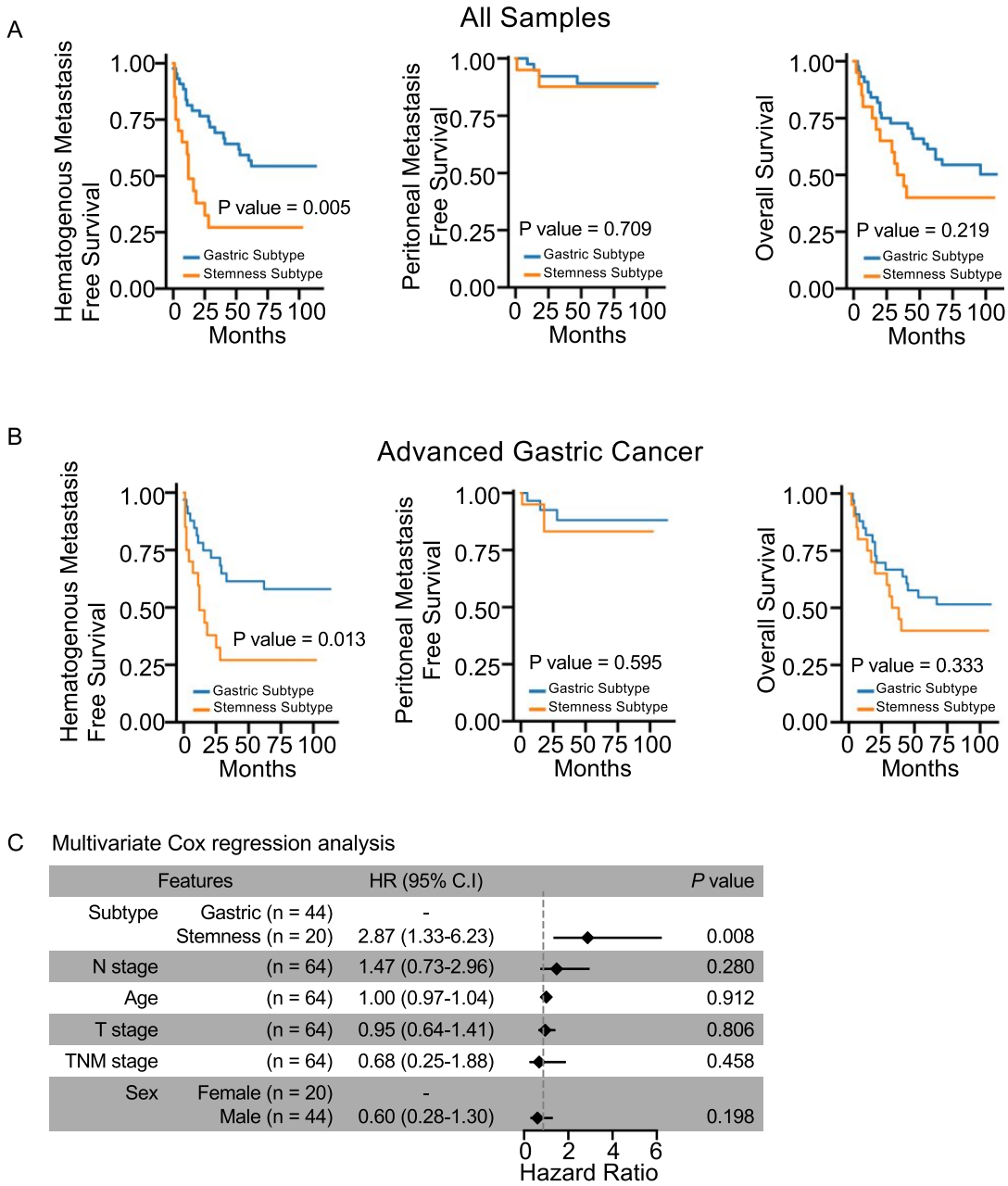

Figure 2. Prognostic significance associated with hematogenous metastasis of stemness subtype gastric cancer. (A) The Kaplan–Meier plot illustrates hematogenous metastasis-free survival (left), peritoneal metastasis-free survival (middle), and overall survival (right) between the stemness and gastric subtypes in primary GC patients (n = 64), color-coded using molecular subtypes. (B) The Kaplan–Meier plot illustrates hematogenous metastasis-free survival (left), peritoneal metastasis-free survival (middle), and overall survival (right) between the stemness (n = 20) and gastric subtypes (n = 33) in patients with advanced gastric cancer patients, color-coded using molecular subtypes. (C) Multivariate Cox proportional analysis used to identify independent prognostic features. For each variable, the reference level is the first one. The diamond represents the hazard ratio, and error bars represent the 95% confidence intervals. HR: hazard ratio; CI: confidence interval.

Identification of hematogenous metastasis-related feature from the machine learning survival model

Next, time-dependent evaluation of the AUC was performed to assess the predictive power of each gene associated with hematogenous metastasis. Several significantly expressed genes in the gastric subtype GC, such as *MUC5AC* and *GKN1*, showed low time-dependent AUC, whereas the stemness subtype significantly expressed genes, including *MAGEA3, TOP2A, GGH, SPP1*, and *OLR1* demonstrated high time-dependent AUC (Fig. 4C; Table S3. http://links.lww.com/JS9/E244). Moreover, a moderate correlation was found between the mean time-dependent AUC of each gene and the gene expression difference between the stemness and gastric subtypes (r = 0.476; Fig. 4D). Thus, stemness subtype-specific genes were predictive of hematogenous metastasis. Additionally, using Fisher's exact test, GSEA was conducted on genes with mean time-dependent AUC exceeding 0.6. This

Table 2

Clinico-pathologic characteristics between molecular subtypes in PDX data

Classification	Stemness (N = 14)	Gastric (N = 37)	*P*-value
Age	62.4 ± 12.0	61.5 ± 10.5	0.778
Sex			0.454
Male	8 (57.1%)	27 (73.0%)	
Female	6 (42.9%)	10 (27.0%)	
Gross			0.396
BI	1 (7.1%)	2 (5.4%)	
BII	4 (28.6%)	7 (18.9%)	
BIII	7 (50.0%)	19 (51.4%)	
BIV	1 (7.1%)	0 (0.0%)	
EGC	0 (0.0%)	5 (13.5%)	
Size	6.5 ± 2.4	5.8 ± 2.3	0.351
WHO			0.71
MD	6 (42.9%)	9 (24.3%)	
PCC	2 (14.3%)	8 (21.6%)	
PD	5 (35.7%)	17 (45.9%)	
WD	1 (7.1%)	2 (5.4%)	
Lauren			0.362
Diffuse	4 (28.6%)	11 (29.7%)	
Intestinal	7 (50.0%)	14 (37.8%)	
Mixed	2 (14.3%)	11 (29.7%)	
Other	1 (7.1%)	0 (0.0%)	
T stage			0.319
I	0 (0.0%)	5 (13.5%)	
II	1 (7.1%)	2 (5.4%)	
III	7 (50.0%)	9 (24.3%)	
IV	6 (42.9%)	20 (54.1%)	
N stage			0.265
N0	1 (7.1%)	4 (10.8%)	
N1	1 (7.1%)	6 (16.2%)	
N2	7 (50.0%)	7 (18.9%)	
N3	5 (35.7%)	19 (51.4%)	
TNM			0.445
Stage I	0 (0.0%)	6 (16.2%)	
Stage II	3 (21.4%)	6 (16.2%)	
Stage III	10 (71.4%)	22 (59.5%)	
Stage IV	1 (7.1%)	3 (8.1%)	

analysis revealed that the hallmark angiogenesis gene set exhibited the highest odds ratio among the hallmark gene sets, indicating a strong association with genes predictive of hematogenous metastasis (odds ratio = 6.98; $P = 0.011$; Fig. 4E). Furthermore, cell cycle-related gene sets, including hallmark E2F and G2M targets, were significantly associated with hematogenous metastasis (odds ratios = 5.34 and 4.46, respectively; $P = 3.15 \times 10^{-6}$ and 8.02×10^{-5}, respectively; Fig. 4E).

Identification of the hematogenous metastasis-associated signature genes

We next sought to define a signature gene set that would be useful in the clinical setting. Among the highly predictive hematogenous metastasis survival genes (mean time-dependent AUC > 0.8) in a gradient boost survival model, we curated 10 genes, namely *MCM10, GGH, WDR72, KIF4A, SLC7A5, PROX1, RAD51, DLX4, NT5C3B,* and *QPCT* that showed significantly elevated expression in the stemness subtype for stemness signature genes, and seven genes, namely *ADH1B, ZNF831, CAMK4, FAM189A2, MS4A1, KIT*, and *FCMR,* that were significantly elevated in the gastric subtype and showed low mean time-dependent AUC for gastric signature genes (Fig. 5A, Table S4. http://links.lww.com/JS9/E244, Method. http://links.lww.com/JS9/E244). We validated this gene set in two independent bulk-level microarray-based GC transcriptome cohorts (GSE66229 and GSE84437) and the TCGA bulk-level RNAseq cohort[5,6,26]. In three external cohorts, the stemness score and gastric score showed moderated to strong negative correlations (Pearson $r = -0.57$; GSE66228, $r = -0.67$; TCGA, $r = -0.53$; GSE84437, Fig. 5B and Figure S7A. http://links.lww.com/JS9/E243).

Next, we performed survival analysis to validate prognostic significance of 10 stemness signature genes. In the survival analysis, patients with a higher stemness score, classified using the maxstat R package, exhibited significantly decreased hematogenous metastasis survival in the GSE66229 ($P = 0.005$, log-rank test, Figure S7B. http://links.lww.com/JS9/E243). Consistent with previous results, OS did not show differences based on the stemness score difference ($P = 0.986$, log-rank test, Figure S7B. http://links.lww.com/JS9/E243). Although we validated the prognosis significance of the stemness score, we could not access the risk at the single-sample level using only the stemness score, owing to the difference in the ssgsva score according to different gene expression analysis methods, such as RNAseq and microarray (Figure S7C. http://links.lww.com/JS9/E243). Therefore, we defined the difference between the stemness and gastric scores as the hematogenous metastasis risk score to facilitate risk identification in individual samples (Fig. 5B, Figure S7C. http://links.lww.com/JS9/E243, Method. http://links.lww.com/JS9/E243).

To identify the optimal cut-off value of the signature gene score, we performed unsupervised clustering using the *K*-means clustering algorithm to identify survival groups using stemness and gastric scores in the GSE66229 cohort. In the *K*-means clustering, the optimal number of clusters was three (Figure S8A. http://links.lww.com/JS9/E243 and S8B. http://links.lww.com/JS9/E243). Of note, clusters A and C showed decreased HMFS compared to cluster B (Figure S8C. http://links.lww.com/JS9/E243). Since nearly all samples in the GSE66229 cohort from clusters A and B could be distinguished when the risk score threshold was set at 0.15 (Fig. 5C), we finally defined the hematogenous metastasis high-risk group as those with a risk score above 0.15 and the low-risk group as those with a risk score below 0.15 (Fig. 5C).

Validation of the hematogenous metastasis risk score

To validate the hematogenous-metastasis risk score, we performed survival analyses in three independent cohorts – GSE66229, TCGA, and GSE84437. In every dataset, patients classified as high risk displayed significantly shorter HMFS than those in the low-risk group (GSE66229, $P = 0.002$; TCGA, $P = 0.040$; GSE84437, $P = 0.030$; log-rank test; Fig. 5D). A pooled analysis of the external cohorts confirmed this association ($P = 2.57 \times 10^{-5}$, Figure S9A. http://links.lww.com/JS9/E243).

Multivariate Cox regression in GSE66229 showed that the risk group was the most powerful prognostic factor for

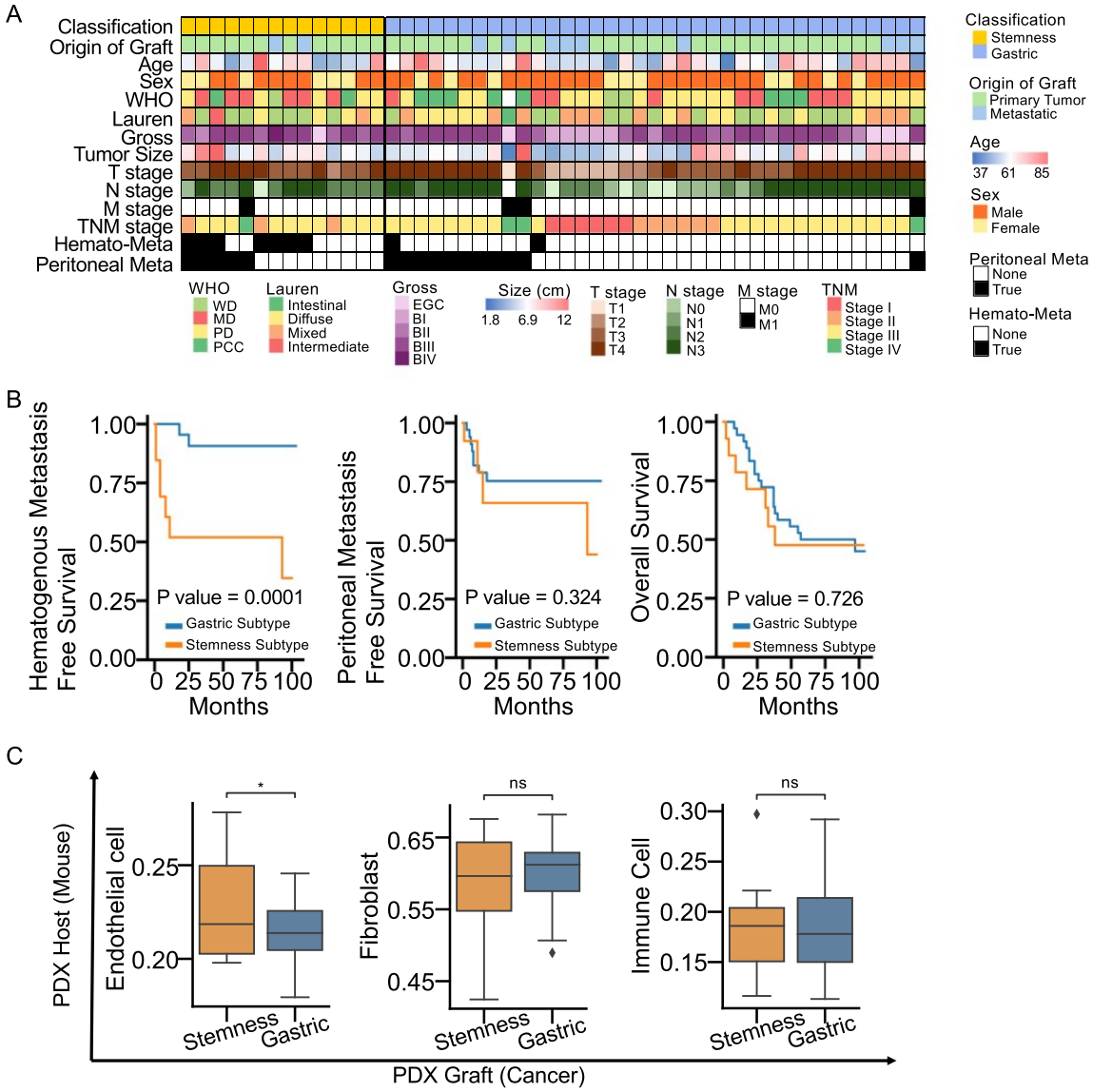

Figure 3. Validation of decreased hematogenous metastasis survival of stemness subtype GC in the patient-derived xenograft (PDX). (A) Oncoprint table showing the clinicopathologic characteristics using the molecular subtype in the PDX dataset (n = 51). Each column represents the corresponding patient of xenograft cancer. Each feature is annotated based on clinical variables (bottom). (B) Kaplan–Meier plot of hematogenous metastasis, peritoneal metastasis, and overall free survival of corresponding patients of xenograft cancer. Orange and blue lines represent stemness and gastric subtypes, respectively. P-value (p) of the log-rank test is shown. (C) Boxplots of the CIBERSORTx fraction result of tumor microenvironment cells, including endothelial (left), fibroblast (middle), and immune (right) cells, in PDX host transcriptome (mouse). Boxes represent the 25th, 50th, and 75th percentiles, and whiskers indicate the 10th and 90th percentiles. An asterisk indicates the P-value of the t-test; *: $0.01 < P \leq 0.05$; ns: $P \geq 0.05$.

HMFS, outperforming previously reported molecular subtypes (EMT and mesenchymal phenotype) as well as age, sex, and TNM stage (Fig. 5E) [6,7]. The risk score also retained its significance as the strongest independent predictor in both the TCGA and GSE84437 cohorts after adjustment for age, sex, and pathologic stage (Figures S9B. http://links.lww.com/JS9/E243 and S9C. http://links.lww.com/JS9/E243).

Next, we analyzed the effect of adjuvant chemotherapy on hematogenous metastasis-free survival in the GSE84437 cohort. However, the history of adjuvant chemotherapy did not affect the oncology outcome of the hematogenous metastasis high-risk group (P= 0.23, log-rank test, Fig. 5F). These results indicated that hematogenous metastasis in gastric cancer was affected by molecular characteristics of cancer cells, and current adjuvant chemotherapy was not effective for these molecular subtypes GC.

Discovery of potential agents for hematogenous metastasis high-risk group

Since the HM high-risk group did not show benefit from adjuvant chemotherapy, we analyzed cancer cell line data to

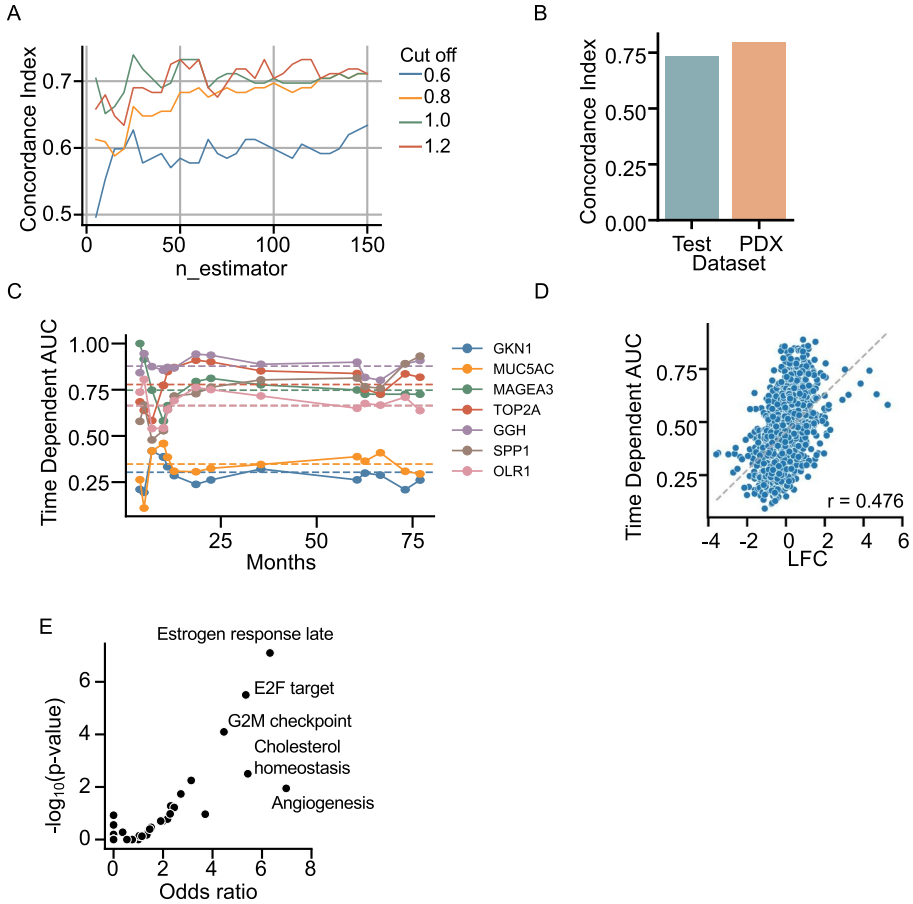

Figure 4. Machine learning survival analysis predicting hematogenous metastasis-free survival in GC. (A) A line plot showing the concordance index (c-index) of the survival model using the number of estimators. The colors of the plot represent the results of the c-index at each cut-off condition, labeled right for input gene selection. (B) The bar plot of the c-index value in the test dataset of the primary GC transcriptome data and PDX transcriptome dataset. (C) Time-dependent AUC result of gene expression in the survival model. The colors of the plot represent each gene labeled on the right side, and the dashed line represents the mean time-dependent AUC. (D) Correlation of log2 fold change of input gene expression between stemness and gastric subtypes and the time-dependent AUC value. The Pearson *r* value is shown. (E) Gene set enrichment analysis of higher survival prediction genes in primary GC data (AUC > 0.6). The hallmark gene set from MSigDB was used for this analysis. The top five enriched gene sets are labeled.

discover potential agents decreasing hematogenous metastasis in GC. Among the 39 gastric cancer cell lines from the CCLE data, a significant correlation between the stemness score and gastric score was also identified (Pearson $r = -0.47$, $P = 0.003$, Fig. 6A). In addition, the hematogenous metastasis risk score was significantly higher in the cells that was harvested from liver metastasis compared to other sites such as the primary tumor, lymph node, and ascites ($P = 0.005$, $P = 0.012$, $P = 0.007$, *t*-test, respectively, Fig. 6B). These findings not only validated our results at the cell line level but also demonstrated that both the stemness score and gastric score accurately reflect cancer cell-specific characteristics. Next, we analyzed drug sensitivity data from the CCLE to discover sensitive drugs for HM high-risk GC. Several drugs, including olutasidenib, a mutant IDH1 target agent, showed significant sensitivity to liver metastasis cells compared with ascites cells ($P = 0.018$, *t*-test, Fig. 6C, Table S5. http://links.lww.com/JS9/E244). In addition, several clinical trial agents, such as MMV-390 048, PI4K inhibitor, and IACS-10 759, selective oxidative phosphorylation inhibitor, also showed significant sensitivity to liver metastasis cells ($P = 0.03$, both, *t*-test, Fig. 6C). These cell line results underscore the need for further investigation in pre-clinical and clinical settings to discover potential therapeutic agents.

Discussion

Recent advances in molecular biology have enabled the development of new therapeutic modalities and increased survival in patients with cancer across various tumor types[27–29]. However, GC is considered as incurable especially with distant organ metastasis or peritoneal seeding owing to its resistance to chemotherapy and rapid progression[30]. Therefore, a deeper investigation of molecular characteristics is crucial for understanding the mechanisms of metastasis in GC and to develop new treatment strategies. In this study, we analyzed bulk-RNAseq data from 64 micro-dissected primary GC samples and 51 GC PDX models combined with machine learning survival analysis to uncover molecular characteristics associated with hematogenous metastasis in GC.

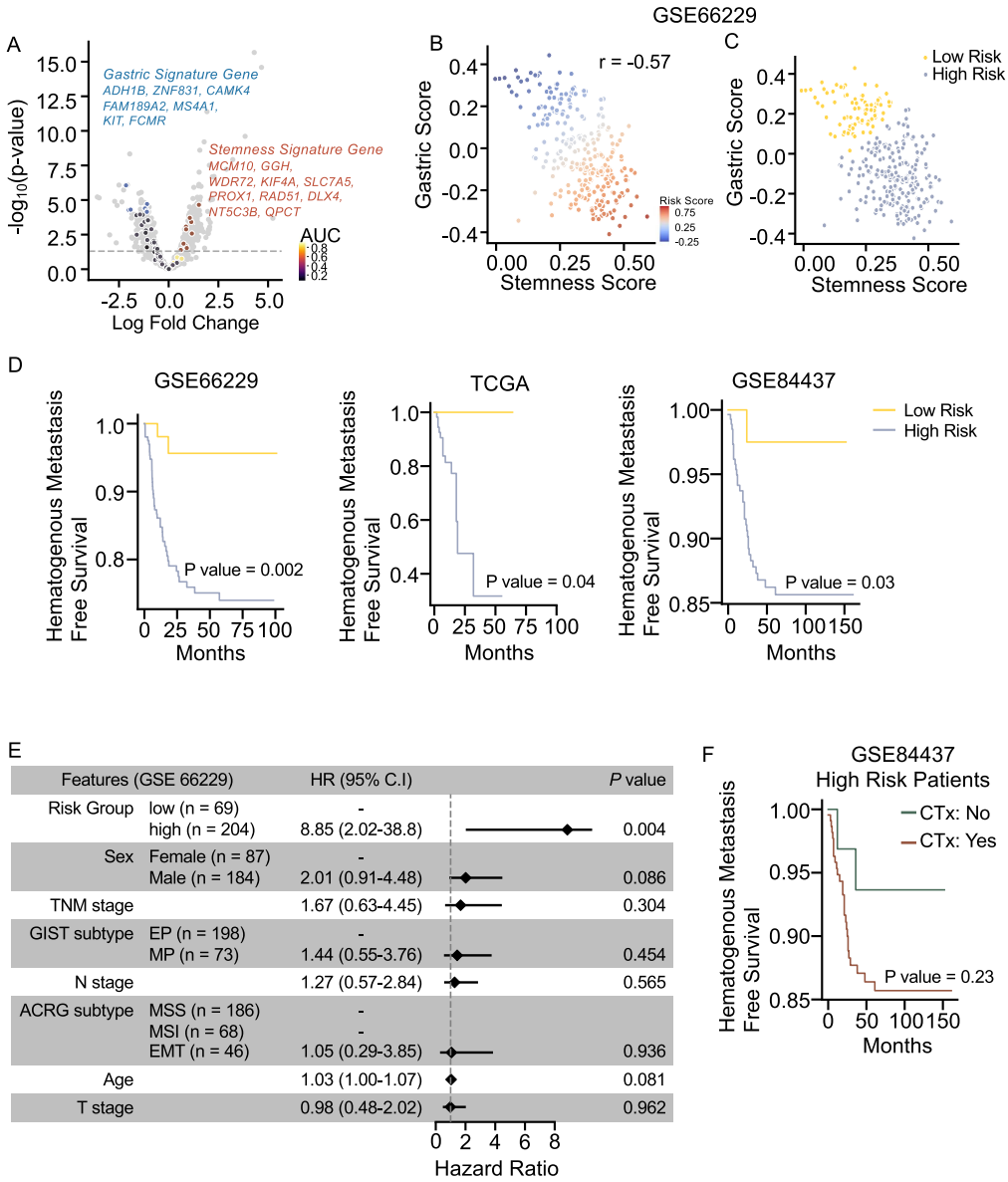

Figure 5. Identification and validation of hematogenous metastasis risk score. (A) Volcano plot of the fold change of gene expression between stemness subtype and gastric subtype. Fold change of gene expression and -$\log_{10}$(*P*-value) of the *t*-test are shown. Genes that showed highly predictive (mean time-dependent AUC > 0.8 and AUC < 0.2) are colored from dark purple to yellow according to the AUC value. Stemness signature genes and gastric signature genes were colored by red and blue, respectively. (B) Scatter plot showing the correlation between stemness and gastric scores in the GSE66229 cohort and colored from blue to red according to the hematogenous metastasis risk score. (C) Scatter plot showing the correlation between stemness and gastric scores in the GSE66229 cohort. The yellow color represents the hematogenous metastasis low-risk group, and the blue color represents the hematogenous metastasis high-risk group. The Pearson *r* value is shown. (D) Kaplan–Meier plot of hematogenous metastasis-free survival in the GSE66229 (left), the Cancer Genome Atlas (TCGA) (middle), and GSE84437 (right) cohorts. Yellow and blue colors represent hematogenous metastasis low-risk and high-risk groups, respectively. *P*-value of the log-rank test is shown. (E) Result of multivariate Cox regression analysis in the GSE66229 cohort. For each variable, the reference level is the first indicated. The diamond represents the hazard ratio, and the error bars represent the 95% confidence intervals. HR: hazard ratio; CI: confidence interval. (F) Kaplan–Meier plot of hematogenous metastasis-free survival according to adjuvant chemotherapy history among the high-risk HM patients in the GSE84437 cohort. *P*-value of the log-rank test is shown.

To the best of our knowledge, this is the first study to explore the molecular characteristics of the GC subtype that favors hematogenous metastasis. We found the elevated expression of stemness markers in this subtype, such as *STMN1* and the *MAGE-A* gene family. *STMN1*, a well-characterized protein involved in maintaining the cytoskeleton and regulating the cell cycle, has been linked to stemness and poor prognosis in various cancers[31–33]. Moreover, single-cell RNA sequencing data analysis revealed that *STMN1* is also the isthmus stem cell marker[34]. Similarly, several studies have reported the association between the upregulated expression of *MAGE-A* gene family and stemness of cancer[35,36]. Therefore, this molecular GC subtype exhibits a loss of differentiated phenotype and an acquisition of progenitor cell characteristics.

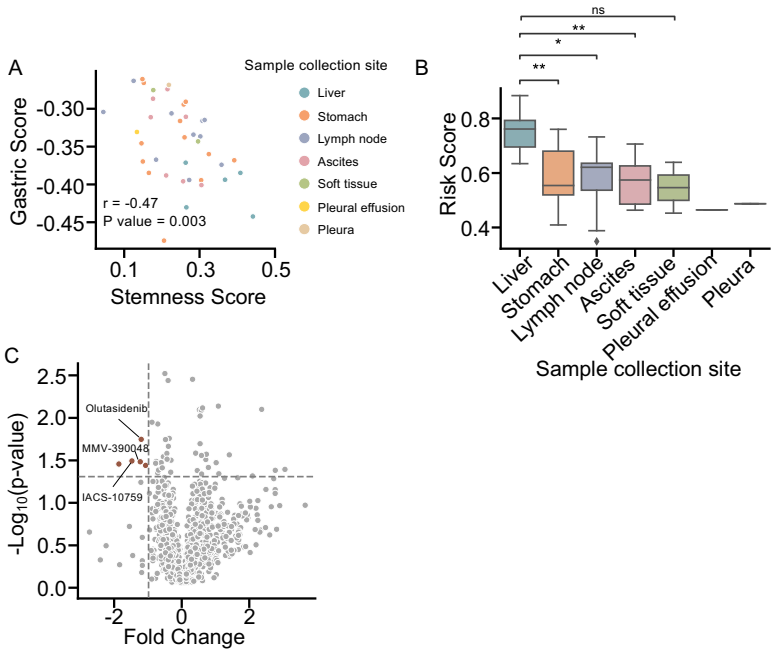

Figure 6. Validation of hematogenous metastasis risk scores in cancer cell line data (A) Scatter plot of correlation between stemness and gastric scores in stomach cancer cell lines from CCLE data, color-coded using the sample collection site. Pearson *r* and *P*-value are shown. (B) Boxplot of the hematogenous metastasis risk score in stomach cancer cell lines from CCLE data between the sample collection sites. Boxes represent the 25th, 50th, and 75th percentiles, and whiskers indicate the 10th and 90th percentiles. An asterisk indicates the *P*-value of the *t*-test; *: $0.01 < p \leq 0.05$. **: $0.001 < P \leq 0.01$. Ns: $p \geq 0.05$. (C) Volcano plot of the fold change of PRISM drug sensitivity between liver metastasis and ascites gastric cancer cell lines, fold change of PRISM LFC, and *P*-value of *t*-test are shown. Horizontal dashed line indicated $P = 0.05$, and vertical dashed line indicated fold change = −1. Red color indicated cells that showed fold change < − 2 and $P < 0.05$.

When compared with previously reported molecular classifications of gastric cancer based on stemness and gastric scores, the stemness subtype identified in our study appears to partially overlap with the MSS or MSI subtypes in the ACRG classification, as well as the epithelial phenotype (EP) described by Oh *et al*. In contrast, both the EMT subtype in the ACRG classification and the mesenchymal phenotype (MP) defined by Oh *et al* exhibited elevated gastric scores, indicating potential molecular similarity with our gastric subtype in GSE66229 microarray data (Figure S10A. http://links.lww.com/JS9/E243 and S10B. http://links.lww.com/JS9/E243) [6,7]. Notably, neither the ACRG molecular subtypes nor the EP/MP classification showed a significant association with hematogenous metastasis or HM-free survival (Figure S10C. http://links.lww.com/JS9/E243 and S10D. http://links.lww.com/JS9/E243). In contrast, our hematogenous metastasis (HM) risk group demonstrated a robust and consistent association with HM-free survival, independent of previously defined subtypes, suggesting that our classification more effectively captures the metastatic potential of gastric cancer (Fig. 5A, S10C. http://links.lww.com/JS9/E243 and S10D. http://links.lww.com/JS9/E243). This is likely because, unlike previous classifications that primarily focused on pathological features and overall prognosis, our study was specifically designed to investigate transcriptomic characteristics associated with hematogenous metastasis-free survival (HMFS). Previously reported molecular subtypes of gastric cancer were largely derived from bulk samples that contained substantial tumor microenvironment (TME) components, potentially limiting their ability to capture tumor cell-specific transcriptomic signals. This contrasts with our RNA-seq data derived from microdissected samples, where EMT-related gene expression – a surrogate for stromal infiltration – did not show significant differences between subtypes[23,24]. This finding stands in contrast to the ACRG classification, in which EMT scores were among the most discriminatory features, suggesting that prior classifications may have been confounded by stromal cell infiltration[6].

In addition, most molecular GC subtypes were determined using transcriptome data lack single-sample resolution. For example, the ACRG conducted principal component analysis (PCA) to identify molecular subtypes and identified four subtypes associated with distinct clinical outcomes[6]. However, PCA, used to evaluate and select principal components explaining data variability, does not provide an absolute criterion for molecular subtyping. Similarly, Oh *et al* identified a mesenchymal GC phenotype associated with poor survival; however, they did not present an absolute standard for determining molecular subtypes[7]. In contrast, our hematogenous metastasis risk score can be applied to single-sample gene expression data, allowing for the evaluation of hematogenous metastasis risks. In addition, to enable stratified hematogenous metastasis risk after curative gastrectomy, we are releasing hematogenous metastasis risk classifiers (hmrisk) that can be used by the community as research tools (Data and Software availability)

Although the detailed metastasis mechanisms are unclear, both the increased angiogenesis and cellular proliferation of the stemness subtype presumably promote hematogenous metastasis. This is because increased angiogenesis around tumor cells can enhance interactions between vessels and cancer cells. In addition, highly proliferative cells can invade blood vessels by disrupting endothelial cells[37,38]. Therefore, the higher hematogenous metastasis rate in the stemness subtype GC is likely attributed to its elevated angiogenetic and mitotic state.

However, further experimental studies are necessary to reveal the detailed mechanism of hematogenous metastasis in GC.

Most hematogenous metastases in five GC cohorts including our primary GC transcriptome data, PDX, GSE66229, TCGA, and GSE84437 were observed within 24 months after curative resection. This observation suggests that intravascular dissemination was already present at the time of surgery, or that microscopic metastasis had already occurred in distant organs in cases of hematogenous metastasis. In pancreas ductal adenocarcinoma, considered as systemic disease from the beginning, recent studies have demonstrated that neoadjuvant systemic chemotherapy decreases metastasis and increases oncologic outcome[39,40]. Therefore, neoadjuvant chemotherapy can be considered for hematogenous metastasis high-risk GC patients. In addition, in the GSE84437 cohort, current adjuvant chemotherapy is ineffective for the prevention of hematogenous metastasis in the HM high-risk group (Fig. 5F). Therefore, new chemotherapy agents or regimens specifically targeting HM high-risk patients should be developed to prevent hematogenous metastasis. Therefore, it is necessary not only to validate several possible compounds that showed sensitivity to liver metastasis cell lines in CCLE data but also to discover and evaluate new drugs in pre-clinical models such as patient-derived organoids or animal models[41].

Possible limitations of our study include the following points: First, although we generated bulk-level RNAseq data using microdissected FFPE blocks and our data did not have TME differences between molecular groups, the bulk-RNA-seq sample inevitably included the TME component. Therefore, single-cell RNA sequencing data analysis could enhance our results. Second, although our molecular subtyping robustly predicted hematogenous dissemination, peritoneal metastasis-free survival did not differ between subtypes in our discovery and PDX cohorts, and a comprehensive evaluation of transcoelomic spread will require larger, prospectively annotated datasets. Third, although the 17-gene risk score demonstrated consistent prognostic value across three large, independent external cohorts, we acknowledge that prospective validation in real-world clinical settings remains necessary. Notably, the external cohorts used in this study already provide a degree of geographic and ethnic diversity. Nevertheless, future multicenter studies involving broader patient populations and comprehensive clinical annotations will be essential to further evaluate the universality and clinical applicability of the scoring system. This will help establish the clinical relevance of our model and its potential to guide personalized treatment strategies in patients with GC. Lastly, while our *in silico* drug sensitivity analysis identified several promising compounds, including olutasidenib, MVV-390 048, and IACS-10 759, preclinical validation using gastric cancer-specific PDX or organoid models will be necessary to confirm subtype-specific therapeutic efficacy. This study will facilitate improvement in oncologic outcomes in patients with gastric cancer.

In conclusion, by identifying new molecular characteristics of GC associated with hematogenous metastasis, we provide a comprehensive characterization of the hematogenous metastasis-associated characteristics and its clinical application. To the best of our knowledge, currently, no such signature has been reported in clinical or biological studies; thus, our results have a high potential of being used to stratify patients for more effective therapies and personalized medicine.

Ethical approval

GC samples were obtained from patients who received curative gastrectomy following the Declaration of Helsinki. The Institutional Review Board approved the study (approval number: 2111-093-1272).

Consent

This study is a retrospective study and qualifies for a waiver of informed consent. Additionally, no additional procedures were performed on patients for research purposes, ensuring that no additional risks were introduced. Furthermore, all patient data used in this study were anonymized and did not contain any personally identifiable information.

Sources of funding

This work was supported by the New Faculty Startup Fund from Seoul National University and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1A2C1093217).

Author contributions

DJ Park and HS Lee: Conceptualization, Methodology, Investigation, Supervision, Project administration, Funding acquisition, Writing original draft. Seungho Lee: Methodology, Formal analysis, Investigation, Data Curation, Visualization, Writing original draft. JE Yoo: Investigation, Data Curation, Visualization, Writing original draft. Seungbok Lee: Methodology, Investigation. Oh,HM Kim, K Jeong, YR Yoo, JY Shin, KU Park, SH Kong, HJ Lee, HK Yang: Data Curation.

Conflicts of interest disclosure

Dr H.S.Lee reported receiving grants from Astellas Pharma Inc., Ono Pharma Korea Co., Ltd., and SparkrBioPharma Inc. outside the submitted work. Dr D.J. Park reported receiving grants from Medtronic, Daewoong Pharmaceutical Company, HK inno.N Corporation, and JW Pharmaceutical Corporation outside the submitted work. Seoul National University has filed a patent application based on this work, in which S.L., J.Y., H.S.L., and D.J.P. are listed as inventors. The other author declares no competing interests.

Research registration unique identifying number (UIN)

This study does not require registration as it is a retrospective observational study based on anonymized patient data.

Guarantor

Do Joong Park.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Data availability statement

We have submitted the SNUH RNA sequencing data and PDX RNA sequencing data to the National Center of Biotechnology Information's Sequence Read Archive under accession numbers PRJNA1119255 and PRJNA1120352, respectively. The hematogenous metastasis risk classifier (hmrisk) can be downloaded from GitHub (https://github.com/oreolic/hmrisk).

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7–30.

[2] Rau B, Brandl A, Thuss-Patience P, *et al*. The efficacy of treatment options for patients with gastric cancer and peritoneal metastasis. Gastric Cancer 2019;22:1226–37.

[3] Lin Z, Wang R, Zhou Y, *et al*. Prediction of distant metastasis and survival prediction of gastric cancer patients with metastasis to the liver, lung, bone, and brain: research based on the SEER database. Ann Transl Med 2022;10:16.

[4] Sawaki K, Kanda M, Ito S, *et al*. Survival times are similar among patients with peritoneal, hematogenous, and nodal recurrences after curative resections for gastric cancer. Cancer Med 2020;9:5392–99.

[5] Cancer Genome Atlas Research, Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–09.

[6] Cristescu R, Lee J, Nebozhyn M, *et al*. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449–56.

[7] Oh SC, Sohn BH, Cheong JH, *et al*. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun 2018;9:1777.

[8] Verstegen MH, Harker M, van de Water C, *et al*. Metastatic pattern in esophageal and gastric cancer: influenced by site and histology. World J Gastroenterol 2020;26:6037–46.

[9] McShane LM, Altman DG, Sauerbrei W, *et al*. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93:387–91.

[10] Na D, Chae J, Cho S-Y, *et al*. Predictive biomarkers for 5-fluorouracil and oxaliplatin-based chemotherapy in gastric cancers via profiling of patient-derived xenografts. Nat Commun 2021;12:4840.

[11] Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101:4164–69.

[12] Szklarczyk D, Gable AL, Nastou KC, *et al*. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 2021;49:D605–12.

[13] Newman AM, Steen CB, Liu CL, *et al*. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37:773–82.

[14] Kumar V, Ramnarayanan K, Sundar R, *et al*. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discov 2022;12:670–91.

[15] Zentgraf J, Rahmann S. Fast lightweight accurate xenograft sorting. Algorithms Mol Biol 2021;16:2.

[16] Ben-Porath I, Thomson MW, Carey VJ, *et al*. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499–507.

[17] Malta TM, Sokolov A, Gentles AJ, *et al*. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell 2018;173:338–354.e15.

[18] Wang R, Dang M, Harada K, *et al*. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. Nat Med 2021;27:141–51.

[19] Fain JS, Van Tongelen A, Loriot A, De Smet C. Epigenetic coactivation of MAGEA6 and CT-GABRA3 defines orientation of a segmental duplication in the human X chromosome. Cytogenet Genome Res 2019;159:12–18.

[20] Raychaudhuri P, Park HJ. FoxM1: a master regulator of tumor metastasis. Cancer Res 2011;71:4329–33.

[21] Khan MA, Khan P, Ahmad A, Fatima M, Nasser MW. FOXM1: a small fox that makes more tracks for cancer progression and metastasis. Semin Cancer Biol 2023;92:1–15

[22] Kopanja D, Chand V, O'Brien E, *et al*. Transcriptional repression by FoxM1 suppresses tumor differentiation and promotes metastasis of breast cancer. Cancer Res 2022;82:2458–71.

[23] Isella C, Terrasi A, Bellomo SE, *et al*. Stromal contribution to the colorectal cancer transcriptome. Nat Genet 2015;47:312–19.

[24] Fujii E, Kato A, Suzuki M. Patient-derived xenograft (PDX) models: characteristics and points to consider for the process of establishment. J Toxicol Pathol 2020;33:153–60.

[25] Qian W, Chen X, Sheng Y, *et al*. Tumor purity in preclinical mouse tumor models. Canc Res Commun 2022;2:353–65.

[26] Yoon S-J, Park J, Shin Y, *et al*. Deconvolution of diffuse gastric cancer and the suppression of CD34 on the BALB/c nude mice model. BMC Cancer 2020;20:314.

[27] Herbst RS, Giaccone G, de Marinis F, *et al*. Atezolizumab for first-line treatment of PD-L1–selected patients with NSCLC. N Engl J Med 2020;383:1328–39.

[28] Colombo N, Dubot C, Lorusso D, *et al*. Pembrolizumab for persistent, recurrent, or metastatic cervical cancer. N Engl J Med 2021;385:1856–67.

[29] Morris VK, Kennedy EB, Baxter NN, *et al*. Treatment of metastatic colorectal cancer: ASCO guideline. J Clin Oncol 2023;41:678–700.

[30] Chen Z, Zhang P-F, Xi H-Q, Wei B, Chen L, Tang Y. Recent advances in the diagnosis, staging, treatment, and prognosis of advanced gastric cancer: a literature review. Front Med 2021;8:744839.

[31] Handa T, Yokobori T, Obayashi S, Fujii T, Shirabe K, Oyama T. Association between high expression of phosphorylated-STMN1 and mesenchymal marker expression and cancer stemness in breast cancer. Anticancer Res 2023;43:5341–48.

[32] Suzuki K, Watanabe A, Araki K, *et al*. High STMN1 expression is associated with tumor differentiation and metastasis in clinical patients with pancreatic cancer. Anticancer Res 2018;38:939–44.

[33] Zhang R, Gao X, Zuo J, *et al*. STMN1 upregulation mediates hepatocellular carcinoma and hepatic stellate cell crosstalk to aggravate cancer by triggering the MET pathway. Cancer Sci 2020;111:406–17.

[34] Han S, Fink J, Jörg DJ, *et al*. Defining the identity and dynamics of adult gastric isthmus stem cells. Cell Stem Cell 2019;25:342–356.e7.

[35] Wei Y, Wang Y, Gong J, *et al*. High expression of MAGE-A9 contributes to stemness and malignancy of human hepatocellular carcinoma. Int J Oncol 2018;52:219–30.

[36] Yin B, Zeng Y, Liu G, Wang X, Wang P, Song Y. MAGE-A3 is highly expressed in a cancer stem cell-like side population of bladder cancer cells. Int J Clin Exp Pathol 2014;7:2934–41.

[37] Wong AD, Searson PC. Mitosis-mediated intravasation in a tissue-engineered tumor-microvessel platform. Cancer Res 2017;77:6453–61.

[38] Wong AD, Searson PC. Live-cell imaging of invasion and intravasation in an artificial microvessel platform. Cancer Res 2014;74:4937–45.

[39] Patel SH, Katz MHG, Ahmad SA. The landmark series: preoperative therapy for pancreatic cancer. Ann Surg Oncol 2021;28:4104–29.

[40] Mizrahi JD, Surana R, Valle JW, Shroff RT. Pancreatic cancer. Lancet 2020;395:2008–20.

[41] Zhao Y, Li S, Zhu L, *et al*. Personalized drug screening using patient-derived organoid and its clinical relevance in gastric cancer. Cell Rep Med 2024;5:101627.