

# Identification of novel prognosis biomarkers through integrating multi-omics data in gastric cancer

**Nannan Liu**

Harbin Medical University

**Yun Wu**

Harbin Medical University

**Weipeng Cheng**

Harbin Medical University

**Yuxuan Wu**

Harbin Medical University

**Liguo Wang**

Harbin Medical University

**Liwei Zhuang** (✉ [zhuangliweiyd@126.com](mailto:zhuangliweiyd@126.com))

Harbin Medical University

---

## Research Article

**Keywords:** Gastric cancer, Multi-omics data, Genomic variation, Prognostic efficacy, Drug response

**Posted Date:** December 4th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-110215/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Cancer on April 26th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-08210-y>.

# Abstract

**Background:** Gastric cancer is a fatal gastrointestinal cancer with high morbidity and poor prognosis. The dismal 5-year survival rate urges reliable biomarkers to assess and improve prognosis of gastric cancer. Distinguishing driver mutations that were required for the cancer phenotype from passenger ones poses a formidable challenge for cancer genomics.

**Methods:** We integrated the multi-omics data of 293 primary gastric cancer patients from The Cancer Genome Atlas (TCGA) projects to identify key driver genes by establishing the prognosis model of the patients. The combination of copy number alteration and somatic mutation analysis helped us to comprehensively reveal molecular markers of genomic variation. Integrating the transcription level of genes provides a unique perspective for us to discover dysregulated factors in transcriptional regulation.

**Results:** The results, we identified 31 molecular markers of genomic variation comprehensively. For instance, the copy number alteration of WASHC5 (also known as KIAA0196) frequently recurrence in gastric cancer patients, which cannot be discovered using traditional methods based on significant mutations. Further, we revealed several dysregulatory factors played a "hubs" regulatory role in the process of biological metabolism based on dysregulatory networks. Cancer hallmark and functional enrichment analysis showed that these key driver genes (KD genes) played a vital role in regulating programmed cell death. The drug response pattern and transcriptional signatures of KD genes reflected its clinical application value.

**Conclusions:** These findings indicated that the KD genes could serve as novel prognostic biomarkers for further research on the pathogenesis of gastric cancers, while elucidating a multidimensional and comprehensive genomic landscape and highlighting the molecular complexity.

## Background

Gastric cancer (GC) is the fourth most common and remains the second cause of death of all malignancies worldwide [1,2]. Accumulating advances in the early diagnosis and treatment of GCs contribute to timely curative resection or chemotherapy for patients [3,4]. However, Metastasis and recurrence of GCs gradually formed due to tumor evolution, resulting in a poor prognosis with dismal 5-year survival rate, which is only about 29.3% [3,4]. Genetic factors play an important role in GCs due to genomic variation and aberrant gene expression, leading to a malignant phenotype [5,6].

Several major cancer sequencing projects, such as The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC) and the Therapeutically Applicable Research to Generate Effective Treatments (TARGET), have created a comprehensive catalog of genomic variations across all major cancer types, including GC [7,8]. Cancer genomics has produced extensive information on cancer-associated genes. In recent decades, basing on rapid development of high throughput technology, numerous biomarkers have been identified and applied to targeted treatment [9,10], such as *HER2* (human epidermal growth factor receptor type 2) in breast cancer and *EGFR* (epidermal growth factor

receptor) in lung cancer [11]. Driver mutations are required for the cancer phenotype, whereas passenger mutations are irrelevant to tumor development and accumulate through DNA replication [7,12]. Although many bioinformatics tools dedicated to driver genes identification have been developed [13,14], the number and specificity of cancer-driver genes remains a matter of debate, thus, distinguishing driver genes from passenger ones poses a formidable challenge for cancer genomics [15].

Owing to the clinically and genetically heterogeneity of cancer, the current screened driver genes of GCs are far from preventing and treating this fatal disease. Therefore, it is of critical to identify more reasonable biomarkers for assessing the response to therapy and predicting prognosis in GC patients. In this study, we integrated multi-omics data of GCs from TCGA cohort to identify prognostic-related key driver genes (KD genes), which drive the development and progression of GCs. We reveal the biological functions of KD genes, like programmed cell death, and the clinical characteristics, including the drug response patterns and the prognostic efficacy of expression signatures in GC patients.

## Methods

### Data source

We obtained multi-omics data of 293 TCGA primary gastric cancer patients from cBioPort data sources (<https://www.cbioportal.org/>), including genome-wide human SNP array 6.0 copy number alteration (level 3), IlluminaGA DNaseq mutation (level 2), IlluminaHiSeq RNAseqV2 mRNA expression (level 3) and miRNA expression (level 3), as well as clinical data of 265 patients. In addition, we acquired two data sets of gene expression profile matched by the disease and normal samples, as well as corresponding clinical data, from Gene Expression Omnibus (GEO) database (accession code were GSE13911 and GSE29272, the sample size is 69 and 268, respectively). The relationships of transcriptional factor (TF) targeting mRNA was from Transfac [16], UCSC [17] and ChIPbase[18], while that of miRNA regulating mRNA was from miTarbase [19] and Starbase [20]. Based on previous study [21], we found there were no major batch effects for the expression data or copy number data. In addition, we downloaded an extra data for gastric cancer patients from the Firhose database (<https://gdac.broadinstitute.org/>).

### Building binary genomic variation profile of coding genes

For mRNA expression data from RNAseqV2, we performed  $\log_2(\text{RESM}+1)$ . For DNA copy number alteration (CNA), we retained CNA gain and loss, as well as high-level amplification and homozygous deletion discretized by GISTICv2 [22]. For mutation, silent (synonymous) substitutions were discarded. We built a binary somatic CNA profile of protein coding genes, where copy number altered was 1 and wild-type was 0. By integrating the CNA spectrum and the genomic mutations, a binary spectrum of the genomic variation of the coding genes are formed (1, 0 represents the variation and the wild type, respectively). Four requirements were for candidate key genes selection. (i) gene should have a dominant CNA type (amplification or deletion, binomial test,  $p \leq 0.05$ ); (ii) RSEM of gene was above 0 in more than 75% of cancer samples; (iii) gene should have concordant changes between CNA and expression, that is,

copy number amplification upregulates its own gene expression level, and copy number deletion downregulates expression level; (iv) in order to improve the accuracy of the calculation, we require the frequency of genomic variation to exceed 0.1.

### **Identification of prognosis-related Key Driver genes (KD genes)**

We have downloaded two datasets of matched disease and normal sample gene expression profiles from GEO (GSE13911 and GSE29272), and performed the differential expression analysis using R package DESeq2 at  $FDR \leq 0.05$  and  $pvalue \leq 0.01$ . Then, we obtained the intersection of the differentially expression genes and the candidate key genes with genomic variation, and then used the clinical data in cBioPort to train clinical model. Considering the genomic variation and patient's survival information, including overall survival (OS) and disease-free survival (DFS), we screened for the prognostic genes whose genomic variation has an impact on patients survival (in both types of survival, OS and DFS, the minimum pvalue is required to be less than 0.05) used log rank test. Combining the above candidate key genes, we determined prognostic related key driver genes (KD genes).

### **Construction of transcriptional dysregulation network**

The experimental confirmation data of miRNA-target gene regulatory relationships from known miRNA-target interaction databases (miTarbase and Starbase), and the experimental confirmation and prediction data of the regulatory relationship between TF and target genes from known TF-target interaction databases (Transfac, UCSC and ChIPbase). Based on known knowledge of expression regulation [23], only regulatory relationships that satisfied the following requirements were remained: 1) miRNA's expression has negative (Pearson's correlation coefficient  $< 0$ ,  $FDR \leq 0.05$ ) regulatory effect on its target gene's expression; 2) The relationship between TF and target gene must be significant ( $FDR \leq 0.05$ ). Benjamini & Hochberg correction were be used for among all relationships between regulators (miRNAs and TFs) and KD genes.

### **Functional enrichment analysis of KD genes**

A web server g:Profiler can be used to genes functional enrichment analysis, including Gene Ontology terms (like molecular function (MF), biological process (BP) and cellular component (CC)) and pathways from KEGG, Reactome and WikiPathways. This web server can also be used to gene set enrichment analysis, including miRNA targets from miRTarBase and regulatory motif matches from TRANSFAC. We inputted KD genes list on g:Profiler web server, and used  $FDR < 0.05$  to screen the functions and target factors (miRNAs and TFs) , which were thought to be enriched by KD genes.

### **Drug sensitivity analysis**

Gastric cancer drug sensitivity data were generated from ongoing high-throughput screening performed by the Cancer Cell Line Encyclopedia (CCLE) from Broad Institute [24] and the Cancer Genome Project (CGP) at the Wellcome Trust Sanger Institute [25]. The values we obtained on the websites include the half maximal inhibitory concentration (IC50), genes expression level across cancer cell lines for each

experiment. Totally, we acquired 24 anticancer compounds screened across 38 gastric cancer cell lines and 251 compounds across 25 gastric cancer cell lines from CCLE and CGP, respectively. The effect measures are the spearman's rank correlation coefficient between the drug IC50s and gene expressions (for example, an effect of -0.5 or 0.5 indicates decrease or increase in drug concentration, respectively).

### KD genes construction grouping signatures

Using the Pearson correlation coefficient of all KD genes expression level, we measured the similarity of the patients with gastric cancer. Based on the matrix of the Pearson similarity, the patients were classified by using hierarchical clustering. To determine the optimal number of categories for the gastric cancer patients, we calculated the Elbow method for the number of categories from 1 to 10. The Elbow method was calculated the total within-cluster sum of square (WSS) using R function `fviz_nbclust` from packages "factoextra" and "NbClust". The WSS value was consistent with the classification performance of the model.

## Results

### Identification of candidate genes driving the development of gastric cancer based on multi-omics data

Cancer genomics has produced extensive information on cancer-associated genes [26]. Driver mutations of cancer confers tumor cell growth advantages during carcinogenesis and disease progression, however, distinguishing driver mutations from passenger ones poses a formidable challenge for cancer genomics [7,15]. In order to identify the driver events in GCs and explore its effects in the progression of the tumor, we analyzed multi-omics data of 293 patients from the cBioPort data resource (see Materials and methods), including genomic data (copy number alteration and somatic mutation) and transcriptome data (the expressions of mRNAs, miRNAs and TFs) (Table 1). We constructed a genomic variation binary spectrum of protein-coding genes in GCs by pre-processing multi-omics data and integrating gene CNA, somatic mutation and gene expression (Fig. 1; see Materials and methods).

**Table 1.** Basic information on gastric cancer multi-omics data

	CNA (SNP 6.0)	Expression (RNA-Seq)	Mutation	miRNA (RNA-Seq)	TF
Samples	293	265	289	395	265
Genes	20558	20132	17172	-	-

After multiple steps of screening, we obtained a total of 2318 candidate genes, which were accounted potentially driving the development and progression of GCs (Fig. 1, Fig. 2A). In the process of integrating multi-omics data, we required that the candidate genes should have concordant changes between copy number alteration (CNA) and mRNA expression. For example, the candidate gene *DERL1* was recognized to be copy number amplification, which have significantly higher mRNA expression level in patients with

CNA compared to wild-type patients ( $p \leq 0.001$ , two-sided Wilcoxon's rank-sum test; Fig. 2BD). Conversely, for candidate gene *NAA15*, patients with copy number deletion have significantly lower mRNA expression levels compared to wild-type patients ( $p \leq 0.001$ , two-sided Wilcoxon's rank-sum test; Fig. 2CE). As a result, the gene size decreases rapidly with the sample size increases when implemented the step that CNA concordant affects mRNA expression itself (Fig. 2AF). After integrating CNA and mutation profile, we found that the candidate genes have high variation frequency, which ranged from 16.9% to 69.8% (the average is 39.1%), and the variant samples size of most genes were between 11 and 20 (Fig. 2G). Also, we found that the number of genes with more than 30 variant samples reached 38 (Fig. 2G). This suggests that adding somatic mutation information can help us to conduct a more comprehensive genomic variation research.

### Determining prognostic-related key driver genes (KD genes)

Studies have shown that cancer driver mutations can provide tumor cell with growth advantage, thus contribute to tumor initiation, progression or metastasis [27,28]. The key driver genes should have an impact on the patient's survival time. Therefore, it is urgent to identify key driver genes from numerous tumor genome events. First, we obtained two expression profile datasets of GCs from the GEO database (GSE13911 and GSE29272), and performed gene differential expression analysis (see Materials and methods in detail). Hierarchical clustering analysis showed that candidate genes in GC patients were significant expression disorders compared to normal patients (Fig. S1AB). By overlapping the differentially expressed genes with the above-identified candidate genes, we obtained the candidate driver genes with different expression levels of tumor tissues and normal tissues (388 in total). For each candidate driver gene, the cancer samples were split into two groups according to its copy number status (one group with CNA and the other without). Then, we used the clinical data of GCs for survival analysis based on CNA status (see Materials and methods in detail). Finally, we have obtained a total of 31 prognostic-related key driver genes (KD genes) (Supplementary Table 1). For example, the patients with KD gene *ABCE1* CNA (deletion) had significantly better disease-free survival (DFS) ( $p < 0.0096$ , Log-rank test; Fig. 3A). In contrast, patients with KD gene *SHFM1* CNA (amplification) had significantly worse overall survival (OS) ( $p < 0.033$ , Log-rank test; Fig. 3B). Further, we mapped the genomic variation landscape of the 31 prognosis-related KD genes in GC patients (Fig. 3C). We found that these KD genes have high genomic variation frequencies (including CNA and somatic mutation). For example, *KIAA0196* (also known as *WASHC5*) has the highest genomic variation frequency (65.5%), which exhibited copy number amplification in 189 patients (where 23 patients were high-level amplification), and somatic mutations occurred in 12 patients (Fig. 3C). In addition, the variation frequency of KD gene *ABCE1* (ATP-binding cassette E1) was 42.2%, which included copy number deletion in 103 patients (all were copy number loss) and somatic mutations occurred in 5 patients (Fig. 3C). *ABCE1* is a member of the ATP-binding cassette transporters and regulates a broad range of biological functions including viral infection, cell proliferation, and anti-apoptosis. Previous research have shown *ABCE1* plays an essential role in lung cancer progression and metastasis [29]. In our study, however, *ABCE1* as a key driver gene associated with prognosis in GCs.

## Construction of transcriptional dysregulatory networks of KD genes

Recent studies have confirmed that regulatory factors can affect the transcriptional activity of protein-coding genes, and thus contribute to disease [30,31]. Regulatory factors miRNA functions in RNA silencing and post-transcriptional regulation of gene expression, which via base-pairing with complementary sequences within mRNA molecules [32]. Transcriptional factors that are activators can help turn specific genes "on" or "off" by binding to nearby DNA [31]. For example, the oncogenic TF *TAL1* can produce a modified autoregulatory circuitry that drives the oncogenic program in T-cell acute lymphoblastic leukemia [31]. Thus, KD genes should have closely functional association on biological network with regulatory factors. We obtained experimental and/or predicted miRNA-target gene pairs and TF-target gene pairs from known databases (Table 2; see Materials and methods in detail). After calculation, we determined the regulatory factors (TFs and miRNAs) with significant regulatory coefficients for each KD gene (Fig. S2AB). Simultaneously, using the regulatory relationship between regulatory factors and KD genes, we constructed a transcriptional dysregulatory network (Fig. 3D; see Materials and methods in detail). Consistent with previous studies, we found that regulatory factors, particularly miRNAs, play a crucial role in cell growth and development [33,34]. For example, KD gene *SPOCK1* was regulated by 12 miRNAs, including miR-7-5p, miR-155-5p, miR-326, and miR-107 (Fig. 3DE). Research shows that *SPOCK1* can promote the invasion and metastasis of gastric cancer through Slug-induced epithelial-mesenchymal transition [35]. While miR-155-5p can form a regulatory feedback loop with *STAT1* and might trigger cancer immunoediting in order to allow tumor cells to escape from immunosurveillance and even to promote tumorigenesis [36]. Also, KD gene *NR3C2* was regulated by 9 miRNAs including miR-7d-5p, miR-155-5p, miR-421 and miR-32-5p (Fig. 3DE). Studies have shown that *NR3C2* play a role in transcription regulator and molecular transducer activity [37], and can be inhibited by migration inhibitory factor (MIF) induced signaling pathway, which was a key driver of disease aggressiveness in patients [38]. From the dysregulatory network, we found that miR-7d-5p, miR-155-5p and miR-135b-5p were synergistic regulated to KD genes *SPOCK1* and *NR3C2* (Fig. 3DE). In addition, several miRNAs could simultaneously regulate multiple KD genes, for instance, miR-140-5p regulated to KD genes *SPOCK1* and *ABCE1*, and miR-26b-5p regulated to *SPCS3* and *SMARCA5* (Fig. 3D). Studies have shown that miR-140-5p has been characterized as a tumor suppressor in gastrointestinal cancer [39,40]. Especially, gastric cancer, miR-140-5p could suppress the proliferation, migration and invasion of tumor cells by regulating *YES1* [41]. This indicates that these miRNAs play a "hubs" regulatory role in the process of biological metabolism and have a biological significance.

**Table 2.** Information of TF and miRNA regulating mRNA used in this study.

	Interaction	Regulator	Target	Resources
TF-gene	626,331	752	19,257	Transfac[16], UCSC[17], Chipbase[18]
miRNA- gene	680,008	2,562	16,339	miRTarbase[19],starbase[20]

In the dysregulatory network, a total of 7 TFs are involved (Fig. 3DE, Fig. S2B), including TF *MYC* positive regulated to KD gene *NAA15* (Padj =  $9.95e-08$ , R = 0.32), *CUX1* negative regulated to *MRPL13* (Padj = 0.001, R = -0.20), and *EGR1* negatively regulated to *SMARCA5* (Padj = 0.022, R = -0.14). Similarly, we found that multiple TFs were synergistic regulated to KD gene. For example, TFs *HNF4A*, *SREBF1* and *POU2F1* simultaneously regulated the transcription of KD gene *ETFDH*, in which up-regulated by *HNF4A* (Padj =  $3.5e-04$ , R = 0.22) and *SREBF1* (Padj = 0.02, R = 0.14), while down-regulated by *POU2F1* (Padj =  $7.05e-06$ , R = -0.27) (Fig. 3DE, Fig. S2B). Previous researches have shown *SREBF1* (also known as *SREBP-1*) is a key regulator of fatty acid metabolism and plays a pivotal role in the transcriptional regulation of different lipogenic genes that mediate lipid synthesis, which acts as a cancer promoter in human diseases [42,43]. The TFs as the dysregulatory factor of KD genes, we also found the "hubs" in the dysregulatory network. For instance, TF *HNF4A* regulated KD genes *ETFDH* and *ACADVL* (Padj = 0.33, R = 0.33), and *ARNT* regulated KD genes *NR3C2* (Padj =  $3.34e-04$ , R = 0.22) and *DERL1* (Padj =  $3.51e-05$ , R = 0.25) simultaneously (Fig. 3D). Wang H et al. generated a miRNA-TF regulatory network, and discovered 5 regulators might have critical roles in colorectal cancer pathogenesis, which was helpful to understand the complex regulatory mechanisms and guide clinical treatment [44]. Interestingly, in our study, KD genes were regulated by multiple types of dysregulatory factors. For example, KD gene *ACADVL* was regulated by TF *HNF4A* and miRNAs miR-100-5p and miR-99a-5p, while, KD gene *NAA15* was regulated by *MYC*, miR-497-5p and miR-145-5p (Fig. 3DE). Hao S et al. revealed that 5 miRNAs (including miR-145, miR-497, miR-30a, miR-31, and miR-20a) were considered to regulate tumor cells proliferation through TFs [45]. These findings suggested that transcriptional regulators play a crucial role in the dysregulation of KD genes, and studies of these dysregulatory factors may facilitate biomarker discovery.

### Functional mechanism portrays of KD genes and dysregulatory factors

In order to characterize the molecular mechanisms of KD genes, we first use the g:profiler online tool for functional enrichment analysis (see Materials and methods in detail). As a result, our KD genes were enriched in many types of biological functions, including Gene Ontology (MF, BP, CC), pathway (REAC and WP) (Fig. 4A). In details, KD genes were enriched in apoptosis-associated functions, like "Apoptosis", "Programmed cell death", "Apoptosis-related network due to altered Notch3 in cancer", and "Apoptosis induced DNA fragmentation", etc.. (Fig. 4B). In addition, several KD genes were enriched in immune-associated function, such as "Antigen processing: Ubiquitination & Proteasome degradation" (Fig. 4B). To further characterize our results, we sought to characterize cancer hallmark landscape of KD genes. In brief, we calculated the semantic similarity score between KD genes-related GO terms and known cancer hallmark-related GO terms [46]. We found that the semantic similarities between our KD genes and the of apoptotic-related cancer hallmark, "Evading Apoptosis", were 0.35. Also, the semantic similarities with two immune-related cancer hallmark ("Evading Immune Detection" and "Tumor Promoting Inflammation") were 0.3 and 0.48, respectively (Fig. 4C). For cancer hallmark "Genome Instability and Mutation", the semantic similarity with KD genes is the highest, reaching 0.65 (Fig. 4C), which indicates that genomic variation of KD genes have important functional mechanisms, including genome instability occurred, and thus play a carcinogenic role in biology [47]. Interestingly, our KD genes were enriched with the functions related to the synthesis and secretion of gastric hormones, such as "Synthesis, secretion, and deacylation

of Ghrelin" (Fig. 4C). Ghrelin is an endogenous peptide hormone mainly produced in the stomach. Previous research have shown that ghrelin can be a promising therapeutic option for cancer cachexia [48]. In addition, KD genes are also enriched with functions related to cell growth, development and metabolism, such as "Biosynthetic process", "Negative regulation of biological process" and "Cellular macromolecule metabolic process".

Our above results have stated that the regulators of KD genes plays an important regulatory role in the carcinogenesis. Using g:profiler, we also observed that KD genes were enriched in the regulatory relationships with TFs and miRNAs (Fig. 4A). In order to study the functional mechanism of regulators, we used the enrichment analysis of KD genes to further characterize. By comparing the regulators we identified with the TFs and miRNAs enriched by KD genes, we found that 4 TFs and 5 miRNAs were confirmed by enrichment (Fig. 4D). The confirmed 4 TFs include *EGR1* (enrichment significance  $P= 1.42e-04$ ), *HNF4A* ( $P= 4.42e-03$ ), *POU2F1* ( $P= 7.67e-03$ ) and *MYC* ( $P= 4.70e-04$ ) (Fig. 4D; Supplementary table S2). Where *EGR1* regulation of 10 KD genes such as *SMARCA5*, *POLR3C*, and *MRPL13*, etc. (Fig. 4D). TF *HNF4A* simultaneously regulated two KD genes *ACADVL* and *ETFDH*. In addition, TFs *HNF4A* and *POU2F1* combination regulated KD gene *HMGB2* (Fig. 3D; Supplementary table S2). It is worth noting that TF *MYC* and miR-139-5p combination regulated KD gene *NAA15* (Fig. 3D; Supplementary table S2). While miR-139-5p ( $P= 0.028$ ) acts as a confirmed miRNA by enrichment, it also regulated two KD genes *HMGB2* and *DERL1* (Fig. 3D; Supplementary table S2). In addition, miR-30a-5p ( $P= 0.021$ ) regulated KD genes *SPCS3*, *PPID* and *DERL1* (Fig. 3D; Supplementary table S2). The miRNAs confirmed by KD gene enrichment also includes miR-26b-5p ( $P= 0.013$ ), miR-32-5p ( $P= 0.037$ ) and miR-186-5p ( $P= 0.044$ ) (Fig. 4D). In summary, functional enrichment analysis indicated that the KD genes were involved in vital biological functions, and further demonstrated that the regulators can be used as potential biomarkers for further experimental studies.

### Drug response effects in preclinical cell models of GC

To explore the potential effects of KD genes on drug response, we evaluated whether their expression level could influence drug response across 38 preclinical cell models of GC from Cancer Cell Line Encyclopedia (CCLE). We found that multiple KD genes presented strong correlations with the drugs response in GC cells (Fig. 5A; see Materials and methods in detail). For example, irinotecan, as a broad spectrum anticancer drug, showed a significant positive correlation with the expression levels of 5 KD genes in GC cells, including *KIAA0196* ( $R= 0.81$ ,  $P= 0.02$ ), *POLR3C* ( $R= 0.86$ ,  $P= 0.01$ ), *RNF139* ( $R= 0.74$ ,  $P= 0.04$ ), *DERL1* ( $R= 0.88$ ,  $P= 0.007$ ), *TRMT12* ( $R= 0.79$ ,  $P= 0.02$ ) (Fig. 5A-C). This result indicated that these KD genes expression levels could enhance the resistance of irinotecan in GC cells. Studies have shown that patients with advanced gastric cancer are often treated with irinotecan monotherapy as salvage-line therapy [49].

Sorafenib is a multi-kinase inhibitor with activity against angiogenesis and RAF-MEK-ERK pathway, it could inhibit proliferation of human gastric cancer cell line, and may reverse resistance to cisplatin through down-regulating MDR1 expression [50]. In our study, however, the drug response of sorafenib in

GC cells showed a significant negative correlation with the expression level of KD gene *NAA15* ( $R = -0.6$ ,  $P = 0.0092$ ) (Fig. 5A-B). Moreover, paclitaxel, as a widely used anticancer drug, exhibited multiple response patterns at the expression level of KD genes. For instance, paclitaxel showed strong resistance in GC cells with upregulated *SPOCK1* expression ( $R = 0.59$ ,  $P = 0.01$ ), while showing strong sensitivity in GC cells with upregulated *HMGB2* ( $R = -0.67$ ,  $P = 0.002$ ) (Fig. 5AB and Fig. 5D). Studies have shown that nab-paclitaxel as second-line treatment in locally advanced inoperable or metastatic gastric and gastroesophageal junction carcinoma is an active chemotherapy regimen [51,52].

Interestingly, the KD genes we identified have response effects in multiple anticancer drugs (Fig. 5AE). For example, in addition to sorafenib, RAF265 ( $R = -0.67$ ,  $P = 0.0023$ ), Nultin-3 ( $R = -0.65$ ,  $P = 0.0037$ ) and L-685458 ( $R = -0.64$ ,  $P = 0.0042$ ) all showed strong drug sensitivity in GC cells with upregulated *NAA15* expression (Fig. 5AE). However, the small molecule compound ZD-6474 showed significant drug resistance at the expression level of *NAA15* ( $R = 0.48$ ,  $P = 0.045$ ). In addition to irinotecan, multiple drugs showed strong resistance at the expression level of KD gene *DERL1*. For instance, drug responses, including AZD6244 ( $R = 0.55$ ,  $P = 0.019$ ), PD-0325901 ( $R = 0.54$ ,  $P = 0.020$ ) and panobinostat ( $R = 0.49$ ,  $P = 0.041$ ), were presented significant positive correlation with *DERL1* (Fig. 5AE). Study shown that pan-histone deacetylase inhibitor panobinostat sensitizes gastric cancer cells to anthracyclines via induction of *CITED2* [53]. Moreover, the drug response of multiple small molecules exhibited strong drug resistance on KD gene *SPOCK1* (response effects from 0.47 to 0.59), while several small molecules showed strong drug sensitivity on KD gene *PFAS* (effect from 0.48 to 0.51) (Fig. 5AE). Besides, we further evaluated whether KD genes could influence drug response across 23 gastric cancer cell lines from cancer genome project (CGP). Indeed, the drug response patterns of several KD genes (such as *SPCK1*, *PFAS*) in CGP were similar to those in CCLE (Fig. S3). Noteworthy, drug response patterns of multiple KD genes were complementary in two cell line models, such as *NAA15*, *RNF139*, and *ETFD* (Fig.5E, Fig. S3). This result suggested that it is necessary to use a combination of two cell line models to explore drug response [54]. In summary, we used cellular models to study the drug response mechanisms of KD genes on the transcriptional level. The effect of small molecule compounds on KD genes can guide researchers for new drug research and development, and has potential application value.

### **Clinical application of KD gene signatures in gastric cancer patients**

Among the above results, we have identified 31 prognostic-related KD genes (Supplementary Table 1). In order to explore the global clinical application value of all KD genes, we constructed sample grouping signatures based on the transcriptional level of KD genes and verified it in multiple sets of data. In briefly, we grouped patients based on the expression of KD genes using hierarchical clustering method (see Materials and methods). We found that when 265 GC patients were clustered into 4 groups, the model was best classified (Fig. 6A, see Materials and methods). The number of patients in these four groups were 146, 63, 16 and 40 in group 1-4, respectively (Fig. 6B). By calculating the Pearson dissimilarity between the samples, we find that the distance inside the group was relatively close, and the distance between the groups was far, which was in line with our research (Fig. 6B).

To reveal the clinical benefits of KD gene signature in GCs, the log rank test were be used to explore the survival outcomes between patient groups. The KM curve showed that the patients within group 2 have the best disease-free survival (DFS) time (Fig. 6C). By observing the patient's genomic variation events, we found that the patients in group 2 mainly carry variations in *KIAA0196* (also known as *WASHC5*), *EIF3H*, *MRPL13*, *MTSS1*, *DERL1*, and *RNF139* (Fig. S4A). Compared with group 2 patients, the patients within group 1 presented significantly worse DFS ( $P= 0.02$ , log rank test; Fig. 6C), and these patients mainly carried genomic events in *TRMT12*, *TRPA1*, *TCEB1*, *MRPS28*, *PON2* and *SHFM1* (Fig. S4A). Also, both group 3 and group 4 have shown significantly worse DFS (Fig. 6C). However, the four groups of KD gene signature did not show significant prognostic efficacy on the patient's overall-survival (OS) time (Fig. S6A). In addition, we found that group 3, with worst survival, showed a higher proportion of LAUREN intestinal class (proportion 87.5%) and WHO tubular class (68.8%) (Fig. 6DE). The other three groups all showed a lower proportion of LAUREN intestinal classes (58.9%, 77.8%, 60.0% for group 1, group 2, group 4, respectively) and WHO tubular class (41.1%, 55.6%, 40.0% for group 1, group 2, group 4, respectively). In contrast, group 3 shows a lower percentage of LAUREN diffuse class and WHO poorly cohesive class (both were 6.25%) (Fig. 6DE). We also counted other clinical features of GCs between these groups, and found that the patients in group 3 showed older age, fewer tumor cells, and more tumor lymphatic infiltration, however, these features did not show statistical significance (Fig. S5).

To further elucidate the clinical value of our KD genes in GCs, we used extra data sets for validation. Similarly, through the expression level of KD genes, we subtyped 443 patients into 4 groups based on hierarchical clustering (Fig. S6BC; see Materials and methods). The KM curve indicated that these four groups of KD gene signature have significant prognostic efficacy in patient's overall-survival (OS) time ( $P= 0.016$ , log rank test; Fig. 6F). Compared with group 1, the patients within both other three groups showed significantly worse OS time ( $P= 0.011$ , 0.003 and 0.042 for group 2, group 3 and group 4, respectively). Using multivariate COX proportional hazard model to correct some clinical features, including age, tumor stage, and number of lymph nodes, we found that the patients within both other three groups were risk factors compared to group 1 (group 2 (HR= 1.22, 95% IC [0.77-1.92],  $P= 0.039$ ), group 3 (HR= 1.75, 95% IC [1.07-2.88],  $P= 0.026$ ), and group 4 (HR= 1.27, 95% IC [0.65-2.50],  $P= 0.042$ ) compared to the patients within group 1, respectively; Fig. 6G). This shows that the signatures of our KD gene were independent prognostic factors. The COX model also showed, compared to tumor stage I, both stage III (HR = 2.83, 95% IC [1.52-5.27],  $p = 0.001$ ) and stage IV (HR = 6.23, 95% IC [2.99- 12.97],  $P= 1e-04$ ) were risk factors (Fig. 6G). In addition, the patient's age were showed correlation with the patient's OS time (HR= 1.03, 95% IC [1.01-1.05],  $P= 8e-04$ ). Similarly, our results suggest that the patients within group 3, as a poorer OS time group, have higher proportion of patients who enriched in more malignant tubular intestinal type subclass (Fig. S6D), and more advanced TNM stage T3 and N2 (Fig. S6E). In summary, we used KD genes to construct the group signatures for gastric cancer patients and validated their association with patient survival in both inherent data sets and extra data sets, which suggested that the KD genes we identified can be used as prognostic biomarkers for further study by basic experimental and clinical researchers.

## Discussion

To date, although many bioinformatics tools dedicated to driver mutations identification have been developed [13,14], distinguishing driver mutations from passenger ones poses a formidable challenge [15]. Therefore, it is quite urgent to identify cancer drivers and understand it from a functional level. In this study, we integrated multi-omics data for identifying the cancer drivers and their dysregulatory factors in patients with gastric cancer. Applying it to 293 patients from TCGA, we identified 31 prognostic-related key driver genes (KD genes). By means of functional enrichment analysis of KD genes, we characterized their affected cancer hallmarks and the biological functions they involved, such as programmed cell death and antigen processes. The drug response pattern and transcriptional signatures of KD genes reflect its clinical application value. Combining DNA copy number alteration and mutation can help us avoid the limitations of traditionally identifying driver events. This study proves the integration of multi-omics data enable to discover novel driver molecules and their dysregulatory mechanisms during tumorigenesis.

Cancer genes generally induce deregulated by their regulators, and exert driver roles in cancer. Based on the regulatory relationships between regulators and target genes [23], we constructed a transcriptional dysregulatory network of KD genes (Fig. 3D; see Materials and methods in detail). The dysregulatory factors of KD genes play a crucial role in cell growth and development [33,34]. In our study, KD gene *SPOCK1* was regulated by miR-155-5p, which can form a regulatory feedback loop with *STAT1* and might trigger cancer immunoediting in order to allow tumor cells to escape from immunosurveillance and even to promote tumorigenesis [36]. In addition, the transcription of KD gene *ETFDH* were up-regulated by TF *SREBF1*. Previous researches have shown *SREBF1* is a key regulator of fatty acid metabolism and plays a pivotal role in the transcriptional regulation of different lipogenic genes that mediate lipid synthesis, which acts as a cancer promoter in human diseases [42,43]. Our study also shown that both miRNAs and TFs play a "hubs" regulatory role in the dysregulation of KD genes, which suggested that dysregulatory factors play a crucial role in the process of biological metabolism. Therefore, studying these dysregulatory factors may facilitate discovery of biomarkers.

During process of genomic variation and natural selection, several driver events exhibited different combination mutational patterns to drive cancer formation, and formed evolutionary dependence. These evolutionary dependency drivers are always highly functional associated, such as participating in similar biological processes and mediating pathway crosstalk, and corporately promote clonal expansion or selective sweep [28,55]. Functional enrichment analysis results show that some KD genes are enriched apoptosis-associated (like "Programmed cell death" and "Apoptosis-related network due to altered Notch3 in cancer") and immune-associated function (like "Antigen processing: Ubiquitination & Proteasome degradation"), which was related to the corresponding cancer hallmarks, including "Evading Apoptosis", "Evading Immune Detection" and "Tumor Promoting Inflammation", respectively. In addition, these KD genes offered shed new insights into molecular mechanisms and provided novel prognostic and drug response potential for clinical practice. These findings suggested that the KD genes and dysregulated factors we identified by integrating multi-omics data could have important implications for understanding

cancer evolution as well as for diagnostic and therapeutic approaches, and might play crucial roles and are worthy to be further explored.

## Conclusions

This study integrates multi-omics data to discover novel driver molecules and their dysregulatory mechanisms based on copy number alteration, somatic mutation analysis and transcription level. We revealed clinical application value of KD genes through drug response pattern and transcriptional signatures. These results will pave the way towards understanding the potential mechanisms that govern the GC progression, which will be useful in clinical practice and might prompt the development of novel therapeutic target for GC patients.

## Abbreviations

GC:Gastric cancer, CNA: Copy number alteration, OS: Overall survival, DFS: Disease-free survival, KD genes: Key driver genes, TF: Transcriptional factor, GEO: Gene Expression Omnibus, MF: Molecular function, BP:Biological process, CC:Cellular component, CCLE:Cancer Cell Line Encyclopedia, CGP:Cancer Genome Project, WSS:Within-cluster sum of square, MIF:migration inhibitory factor.

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Availability of data and materials

The datasets used and analysed during the current study are available in the cBioPort (<https://www.cbioportal.org/>), GEO(<https://www.ncbi.nlm.nih.gov/geo/>), Transfac(<http://genexplain.com/transfac/>), UCSC(<https://genome.ucsc.edu/>), Chipbase(<http://rna.sysu.edu.cn/chipbase/>), miTarbase(<http://mirtarbase.cuhk.edu.cn/php/index.php>), Starbase(<http://starbase.sysu.edu.cn/>) and Firhose (<https://gdac.broadinstitute.org/>) repository.

### Competing interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

### Funding

This work was supported by the Key Research Project from The Fourth Affiliated Hospital of Harbin Medical University (HYDSYYZ201507).

## Authors' contributions

LWZ and LGW contributed to conceive and design the research. NNL and YW performed data analysis and wrote the paper. WPC and YXW participated in the design of the study and mathematical analysis of data. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable

## References

1. Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *CA: a cancer journal for clinicians* **2011**, *61*, 69-90, doi:10.3322/caac.20107.
2. Wright, N.A.; Poulsom, R.; Stamp, G.; Van Noorden, S.; Sarraf, C.; Elia, G.; Ahnen, D.; Jeffery, R.; Longcroft, J.; Pike, C., et al. Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology* **1993**, *104*, 12-20, doi:10.1016/0016-5085(93)90830-6.
3. Oba, K.; Paoletti, X.; Alberts, S.; Bang, Y.J.; Benedetti, J.; Bleiberg, H.; Catalano, P.; Lordick, F.; Michiels, S.; Morita, S., et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. *Journal of the National Cancer Institute* **2013**, *105*, 1600-1607, doi:10.1093/jnci/djt270.
4. Wei, J.; Wu, N.D.; Liu, B.R. Regional but fatal: Intraperitoneal metastasis in gastric cancer. *World journal of gastroenterology* **2016**, *22*, 7478-7485, doi:10.3748/wjg.v22.i33.7478.
5. Fu, D.G. Epigenetic alterations in gastric cancer (Review). *Molecular medicine reports* **2015**, *12*, 3223-3230, doi:10.3892/mmr.2015.3816.
6. Smith, M.G.; Hold, G.L.; Tahara, E.; El-Omar, E.M. Cellular and molecular aspects of gastric cancer. *World journal of gastroenterology* **2006**, *12*, 2979-2990, doi:10.3748/wjg.v12.i19.2979.
7. Kumar, R.D.; Swamidass, S.J.; Bose, R. Unsupervised detection of cancer driver mutations with parsimony-guided learning. *Nature genetics* **2016**, *48*, 1288-1294, doi:10.1038/ng.3658.
8. Korthauer, K.D.; Kendzierski, C. MADGiC: a model-based approach for identifying driver genes in cancer. *Bioinformatics* **2015**, *31*, 1526-1535, doi:10.1093/bioinformatics/btu858.
9. Hou, J.Y.; Wang, Y.G.; Ma, S.J.; Yang, B.Y.; Li, Q.P. Identification of a prognostic 5-Gene expression signature for gastric cancer. *Journal of cancer research and clinical oncology* **2017**, *143*, 619-629, doi:10.1007/s00432-016-2324-z.
10. Chen, X.; Yang, Y.; Liu, J.; Li, B.; Xu, Y.; Li, C.; Xu, Q.; Liu, G.; Chen, Y.; Ying, J., et al. NDRG4 hypermethylation is a potential biomarker for diagnosis and prognosis of gastric cancer in Chinese

- population. *Oncotarget* **2017**, *8*, 8105-8119, doi:10.18632/oncotarget.14099.
11. Lazar, D.C.; Taban, S.; Cornianu, M.; Faur, A.; Goldis, A. New advances in targeted gastric cancer treatment. *World journal of gastroenterology* **2016**, *22*, 6776-6799, doi:10.3748/wjg.v22.i30.6776.
  12. Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G.L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C., et al. Patterns of somatic mutation in human cancer genomes. *Nature* **2007**, *446*, 153-158, doi:10.1038/nature05610.
  13. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A., et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214-218, doi:10.1038/nature12213.
  14. Han, Y.; Yang, J.; Qian, X.; Cheng, W.C.; Liu, S.H.; Hua, X.; Zhou, L.; Yang, Y.; Wu, Q.; Liu, P., et al. DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies. *Nucleic acids research* **2019**, *47*, e45, doi:10.1093/nar/gkz096.
  15. Iranzo, J.; Martincorena, I.; Koonin, E.V. Cancer-mutation network and the number and specificity of driver mutations. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, E6010-E6019, doi:10.1073/pnas.1803155115.
  16. Wingender, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Briefings in bioinformatics* **2008**, *9*, 326-332, doi:10.1093/bib/bbn016.
  17. Goldman, M.; Craft, B.; Swatloski, T.; Cline, M.; Morozova, O.; Diekhans, M.; Haussler, D.; Zhu, J. The UCSC Cancer Genomics Browser: update 2015. *Nucleic acids research* **2015**, *43*, D812-817, doi:10.1093/nar/gku1073.
  18. Yang, J.-H.; Li, J.-H.; Jiang, S.; Zhou, H.; Qu, L.-H. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic acids research* **2013**, *41*, D177-D187.
  19. Hsu, S.D.; Lin, F.M.; Wu, W.Y.; Liang, C.; Huang, W.C.; Chan, W.L.; Tsai, W.T.; Chen, G.Z.; Lee, C.J.; Chiu, C.M., et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research* **2011**, *39*, D163-169, doi:10.1093/nar/gkq1107.
  20. Li, J.H.; Liu, S.; Zhou, H.; Qu, L.H.; Yang, J.H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research* **2014**, *42*, D92-97, doi:10.1093/nar/gkt1248.
  21. Zhang, H.; Deng, Y.; Zhang, Y.; Ping, Y.; Zhao, H.; Pang, L.; Zhang, X.; Wang, L.; Xu, C.; Xiao, Y., et al. Cooperative genomic alteration network reveals molecular classification across 12 major cancer types. *Nucleic acids research* **2017**, *45*, 567-582, doi:10.1093/nar/gkw1087.
  22. Mermel, C.H.; Schumacher, S.E.; Hill, B.; Meyerson, M.L.; Beroukhi, R.; Getz, G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **2011**, *12*, R41, doi:10.1186/gb-2011-12-4-r41.
  23. Ping, Y.; Deng, Y.; Wang, L.; Zhang, H.; Zhang, Y.; Xu, C.; Zhao, H.; Fan, H.; Yu, F.; Xiao, Y., et al. Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional

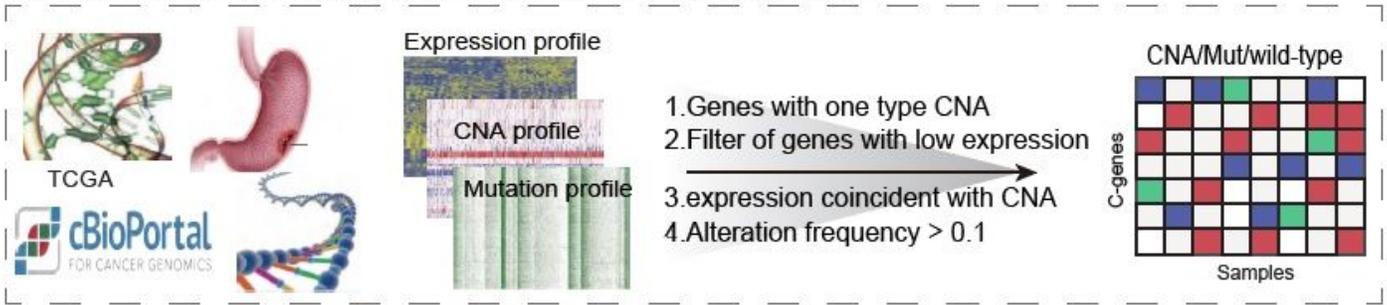
- regulatory networks through integrating multi-dimensional genomic data. *Nucleic acids research* **2015**, *43*, 1997-2007, doi:10.1093/nar/gkv074.
24. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Wilson, C.J.; Lehar, J.; Kryukov, G.V.; Sonkin, D., et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483*, 603-607, doi:10.1038/nature11003.
  25. Garnett, M.J.; Edelman, E.J.; Heidorn, S.J.; Greenman, C.D.; Dastur, A.; Lau, K.W.; Greninger, P.; Thompson, I.R.; Luo, X.; Soares, J., et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012**, *483*, 570-575, doi:10.1038/nature11005.
  26. Ding, L.; Bailey, M.H.; Porta-Pardo, E.; Thorsson, V.; Colaprico, A.; Bertrand, D.; Gibbs, D.L.; Weerasinghe, A.; Huang, K.L.; Tokheim, C., et al. Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell* **2018**, *173*, 305-320 e310, doi:10.1016/j.cell.2018.03.033.
  27. Vogelstein, B.; Papadopoulos, N.; Velculescu, V.E.; Zhou, S.; Diaz, L.A., Jr.; Kinzler, K.W. Cancer genome landscapes. *Science* **2013**, *339*, 1546-1558, doi:10.1126/science.1235122.
  28. Bozic, I.; Antal, T.; Ohtsuki, H.; Carter, H.; Kim, D.; Chen, S.; Karchin, R.; Kinzler, K.W.; Vogelstein, B.; Nowak, M.A. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 18545-18550, doi:10.1073/pnas.1010978107.
  29. Tian, Y.; Tian, X.; Han, X.; Chen, Y.; Song, C.Y.; Jiang, W.J.; Tian, D.L. ABCE1 plays an essential role in lung cancer progression and metastasis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **2016**, *37*, 8375-8382, doi:10.1007/s13277-015-4713-3.
  30. Andersson, R.; Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nature reviews. Genetics* **2019**, doi:10.1038/s41576-019-0173-8, doi:10.1038/s41576-019-0173-8.
  31. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237-1251, doi:10.1016/j.cell.2013.02.014.
  32. Ambros, V. The functions of animal microRNAs. *Nature* **2004**, *431*, 350-355, doi:10.1038/nature02871.
  33. Oliveto, S.; Mancino, M.; Manfrini, N.; Biffo, S. Role of microRNAs in translation regulation and cancer. *World journal of biological chemistry* **2017**, *8*, 45-56, doi:10.4331/wjbc.v8.i1.45.
  34. Gulyaeva, L.F.; Kushlinskiy, N.E. Regulatory mechanisms of microRNA expression. *Journal of translational medicine* **2016**, *14*, 143, doi:10.1186/s12967-016-0893-x.
  35. Chen, D.; Zhou, H.; Liu, G.; Zhao, Y.; Cao, G.; Liu, Q. SPOCK1 promotes the invasion and metastasis of gastric cancer through Slug-induced epithelial-mesenchymal transition. *Journal of cellular and molecular medicine* **2018**, *22*, 797-807, doi:10.1111/jcmm.13357.
  36. Lin, C.C.; Jiang, W.; Mitra, R.; Cheng, F.; Yu, H.; Zhao, Z. Regulation rewiring analysis reveals mutual regulation between STAT1 and miR-155-5p in tumor immunosurveillance in seven major cancers. *Scientific reports* **2015**, *5*, 12063, doi:10.1038/srep12063.
  37. Zhang, W.; Zhang, Q.; Zhang, M.; Zhang, Y.; Li, F.; Lei, P. Analysis for the mechanism between the small cell lung cancer and non-small cell lung cancer combing the miRNA and mRNA expression

- profiles. *Thoracic cancer* **2015**, *6*, 70-79, doi:10.1111/1759-7714.12135.
38. Yang, S.; He, P.; Wang, J.; Schetter, A.; Tang, W.; Funamizu, N.; Yanaga, K.; Uwagawa, T.; Satoskar, A.R.; Gaedcke, J., et al. A Novel MIF Signaling Pathway Drives the Malignant Character of Pancreatic Cancer by Targeting NR3C2. *Cancer research* **2016**, *76*, 3838-3850, doi:10.1158/0008-5472.CAN-15-2841.
  39. Yu, J.; Zhang, W.; Tang, H.; Qian, H.; Yang, J.; Zhu, Z.; Ren, P.; Lu, B. Septin 2 accelerates the progression of biliary tract cancer and is negatively regulated by mir-140-5p. *Gene* **2016**, *589*, 20-26, doi:10.1016/j.gene.2016.05.005.
  40. Zhang, W.; Zou, C.; Pan, L.; Xu, Y.; Qi, W.; Ma, G.; Hou, Y.; Jiang, P. MicroRNA-140-5p inhibits the progression of colorectal cancer by targeting VEGFA. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **2015**, *37*, 1123-1133, doi:10.1159/000430237.
  41. Fang, Z.; Yin, S.; Sun, R.; Zhang, S.; Fu, M.; Wu, Y.; Zhang, T.; Khaliq, J.; Li, Y. miR-140-5p suppresses the proliferation, migration and invasion of gastric cancer by regulating YES1. *Molecular cancer* **2017**, *16*, 139, doi:10.1186/s12943-017-0708-6.
  42. Bao, J.; Zhu, L.; Zhu, Q.; Su, J.; Liu, M.; Huang, W. SREBP-1 is an independent prognostic marker and promotes invasion and migration in breast cancer. *Oncology letters* **2016**, *12*, 2409-2416, doi:10.3892/ol.2016.4988.
  43. Lin, L.; Zheng, X.; Qiu, C.; Dongol, S.; Lv, Q.; Jiang, J.; Kong, B.; Wang, C. SIRT1 promotes endometrial tumor growth by targeting SREBP1 and lipogenesis. *Oncology reports* **2014**, *32*, 2831-2835, doi:10.3892/or.2014.3521.
  44. Wang, H.; Luo, J.; Liu, C.; Niu, H.; Wang, J.; Liu, Q.; Zhao, Z.; Xu, H.; Ding, Y.; Sun, J., et al. Investigating MicroRNA and transcription factor co-regulatory networks in colorectal cancer. *BMC bioinformatics* **2017**, *18*, 388, doi:10.1186/s12859-017-1796-4.
  45. Hao, S.; Huo, S.; Du, Z.; Yang, Q.; Ren, M.; Liu, S.; Liu, T.; Zhang, G. MicroRNA-related transcription factor regulatory networks in human colorectal cancer. *Medicine* **2019**, *98*, e15158, doi:10.1097/MD.00000000000015158.
  46. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **2011**, *144*, 646-674, doi:10.1016/j.cell.2011.02.013.
  47. Basu, A.K. DNA Damage, Mutagenesis and Cancer. *International journal of molecular sciences* **2018**, *19*, doi:10.3390/ijms19040970.
  48. Soleyman-Jahi, S.; Sadeghi, F.; Pastaki Khoshbin, A.; Khani, L.; Roosta, V.; Zendehtel, K. Attribution of Ghrelin to Cancer; Attempts to Unravel an Apparent Controversy. *Frontiers in oncology* **2019**, *9*, 1014, doi:10.3389/fonc.2019.01014.
  49. Makiyama, A.; Arimizu, K.; Hirano, G.; Makiyama, C.; Matsushita, Y.; Shirakawa, T.; Ohmura, H.; Komoda, M.; Uchino, K.; Inadomi, K., et al. Irinotecan monotherapy as third-line or later treatment in advanced gastric cancer. *Gastric cancer : official journal of the International Gastric Cancer*

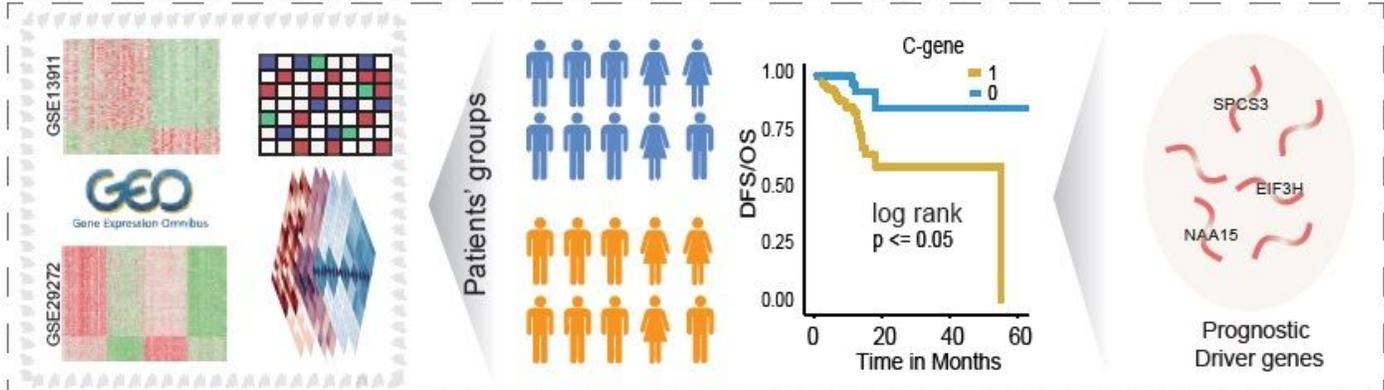
- Association and the Japanese Gastric Cancer Association* **2018**, *21*, 464-472, doi:10.1007/s10120-017-0759-9.
50. Huang, Y.S.; Xue, Z.; Zhang, H. Sorafenib reverses resistance of gastric cancer to treatment by cisplatin through down-regulating MDR1 expression. *Medical oncology* **2015**, *32*, 470, doi:10.1007/s12032-014-0470-1.
51. Katsaounis, P.; Kotsakis, A.; Kentepozidis, N.; Polyzos, A.; Bakogeorgos, M.; Koinis, F.; Vamvakas, L.; Vardakis, N.; Kalbakis, K.; Boukovinas, I., et al. Nab-paclitaxel as second-line treatment in advanced gastric cancer: a multicenter phase II study of the Hellenic Oncology Research Group. *Annals of gastroenterology* **2018**, *31*, 65-70, doi:10.20524/aog.2017.0215.
52. Hironaka, S.; Zenda, S.; Boku, N.; Fukutomi, A.; Yoshino, T.; Onozawa, Y. Weekly paclitaxel as second-line chemotherapy for advanced or recurrent gastric cancer. *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* **2006**, *9*, 14-18, doi:10.1007/s10120-005-0351-6.
53. Regel, I.; Merkl, L.; Friedrich, T.; Burgermeister, E.; Zimmermann, W.; Einwachter, H.; Herrmann, K.; Langer, R.; Rocken, C.; Hofheinz, R., et al. Pan-histone deacetylase inhibitor panobinostat sensitizes gastric cancer cells to anthracyclines via induction of CITED2. *Gastroenterology* **2012**, *143*, 99-109 e110, doi:10.1053/j.gastro.2012.03.035.
54. Safikhani, Z.; Smirnov, P.; Freeman, M.; El-Hachem, N.; She, A.; Rene, Q.; Goldenberg, A.; Birkbak, N.J.; Hatzis, C.; Shi, L., et al. Revisiting inconsistency in large pharmacogenomic studies. *F1000Research* **2016**, *5*, 2333, doi:10.12688/f1000research.9611.3.
55. Gatenby, R.A.; Cunningham, J.J.; Brown, J.S. Evolutionary triage governs fitness in driver and passenger mutations and suggests targeting never mutations. *Nature communications* **2014**, *5*, 5499, doi:10.1038/ncomms6499.

## Figures

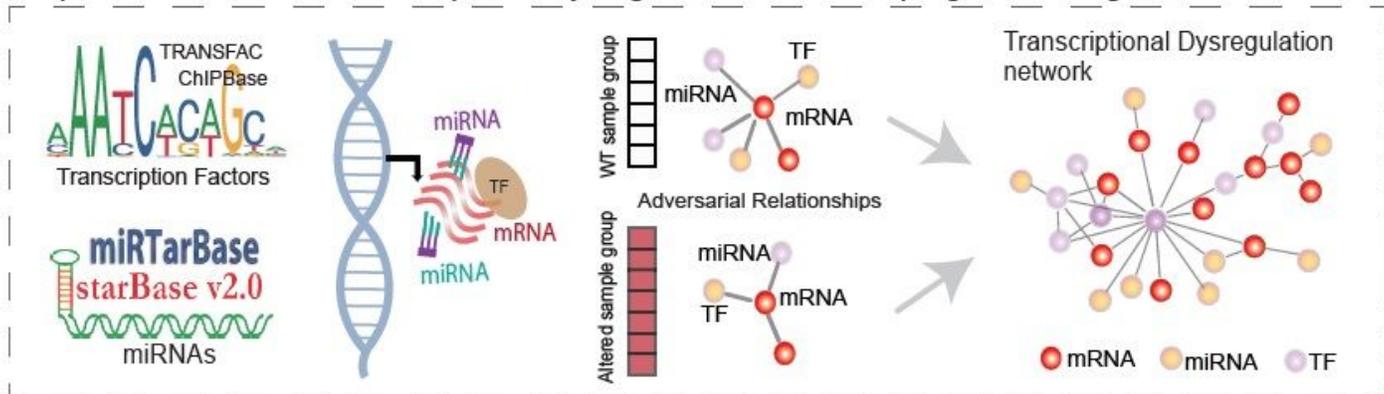
**Step1: Search for candidate genes based on multi-omics data.**



**Step2: Identify the prognostic key driver genes (KD genes) combined with GEO data.**

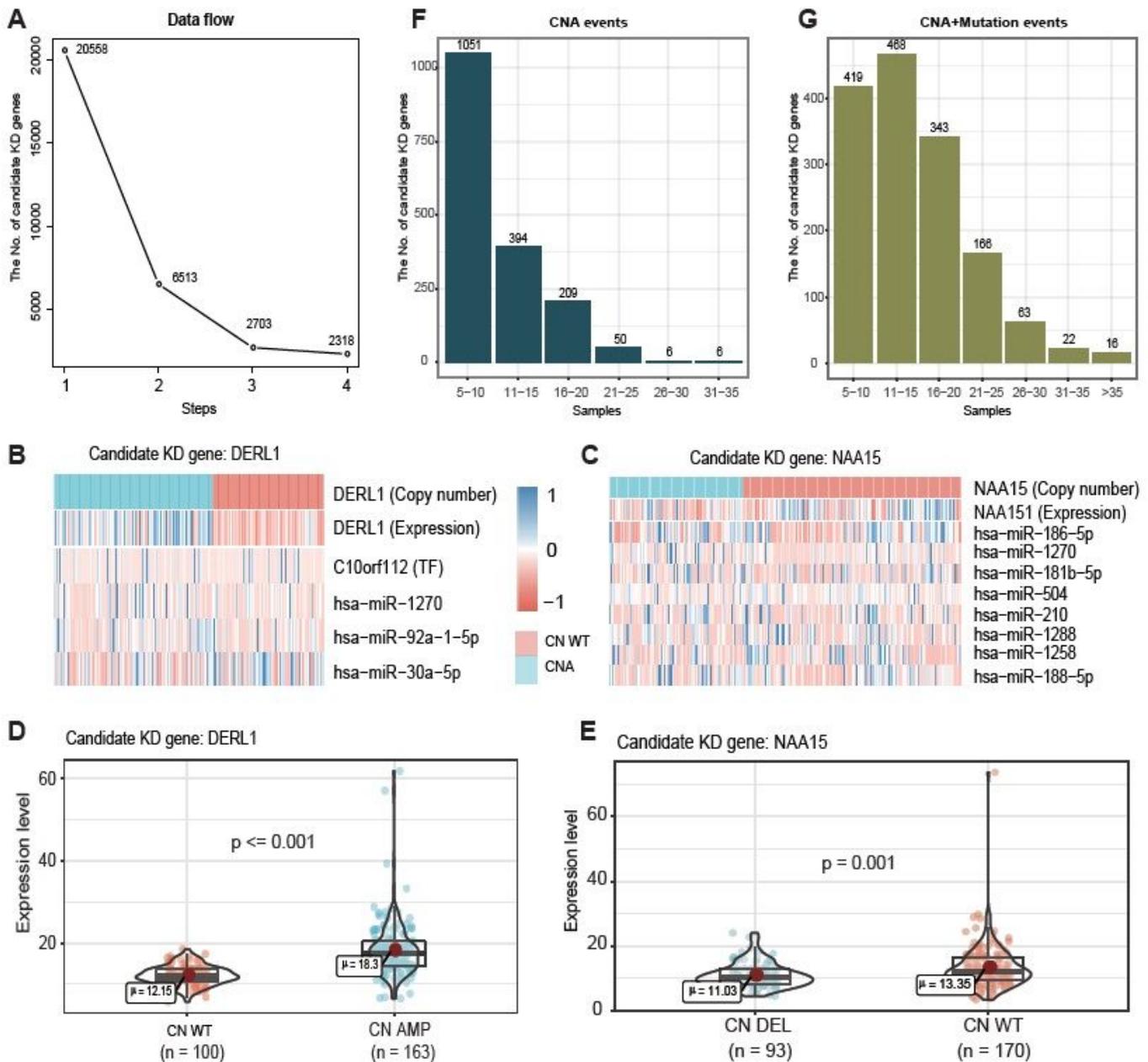


**Step3: Construction the transcriptional dysregulation network of prognostic KD genes.**



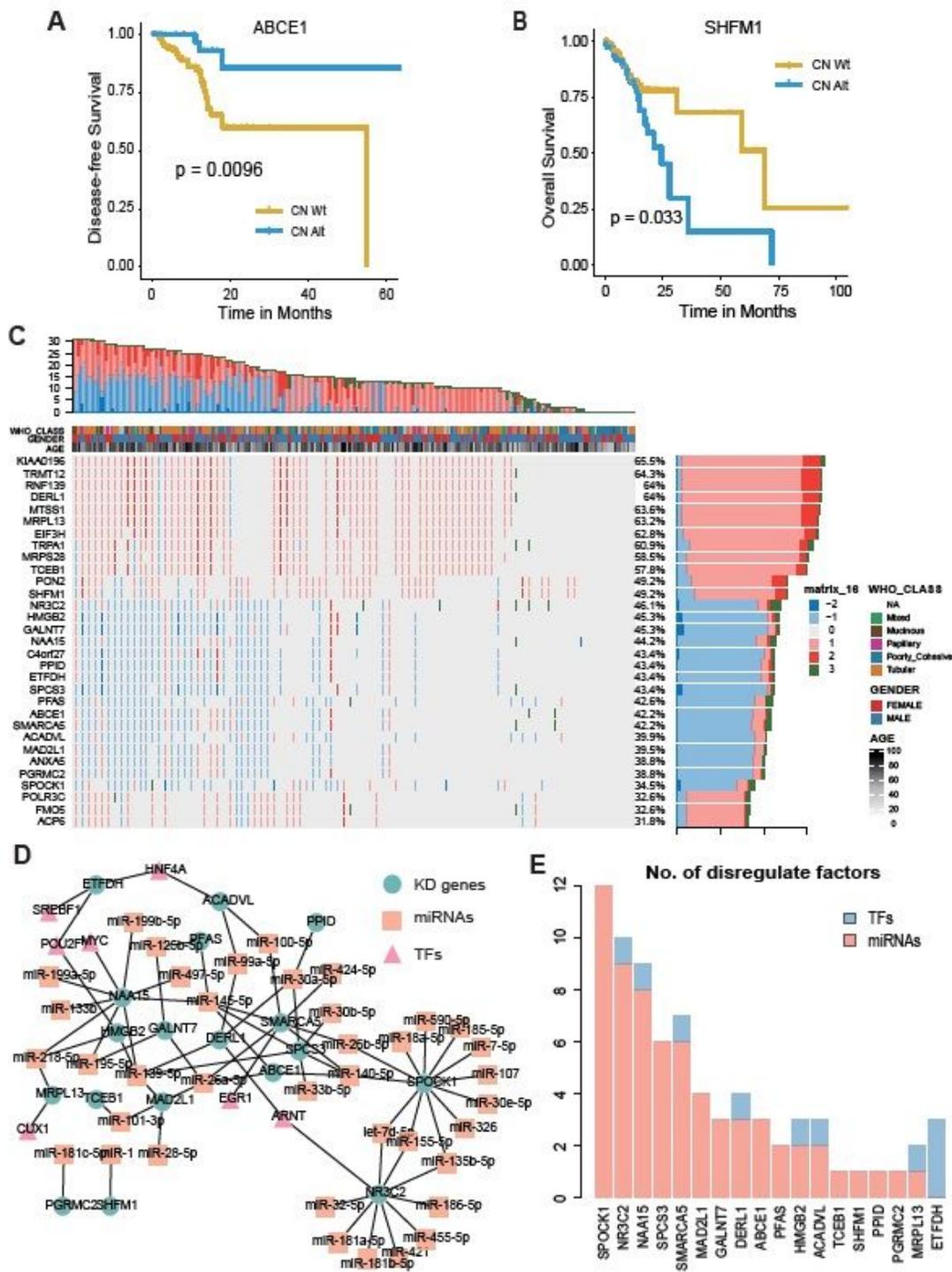
**Figure 1**

The overview Step1, The genomic variation spectrum of gastric cancer patients were constructed using TCGA multi-omics data (including copy number alteration, mutation and mRNA expression level), and the candidate genes were screened. Step2, Identification of key driver genes (KD genes) related to prognosis in gastric cancer patients based on expression data and clinical data. Step3, The transcriptional dysregulatory network of gastric cancer patients was constructed based on the relationship between known regulatory factors (miRNAs and TFs) and target genes.



**Figure 2**

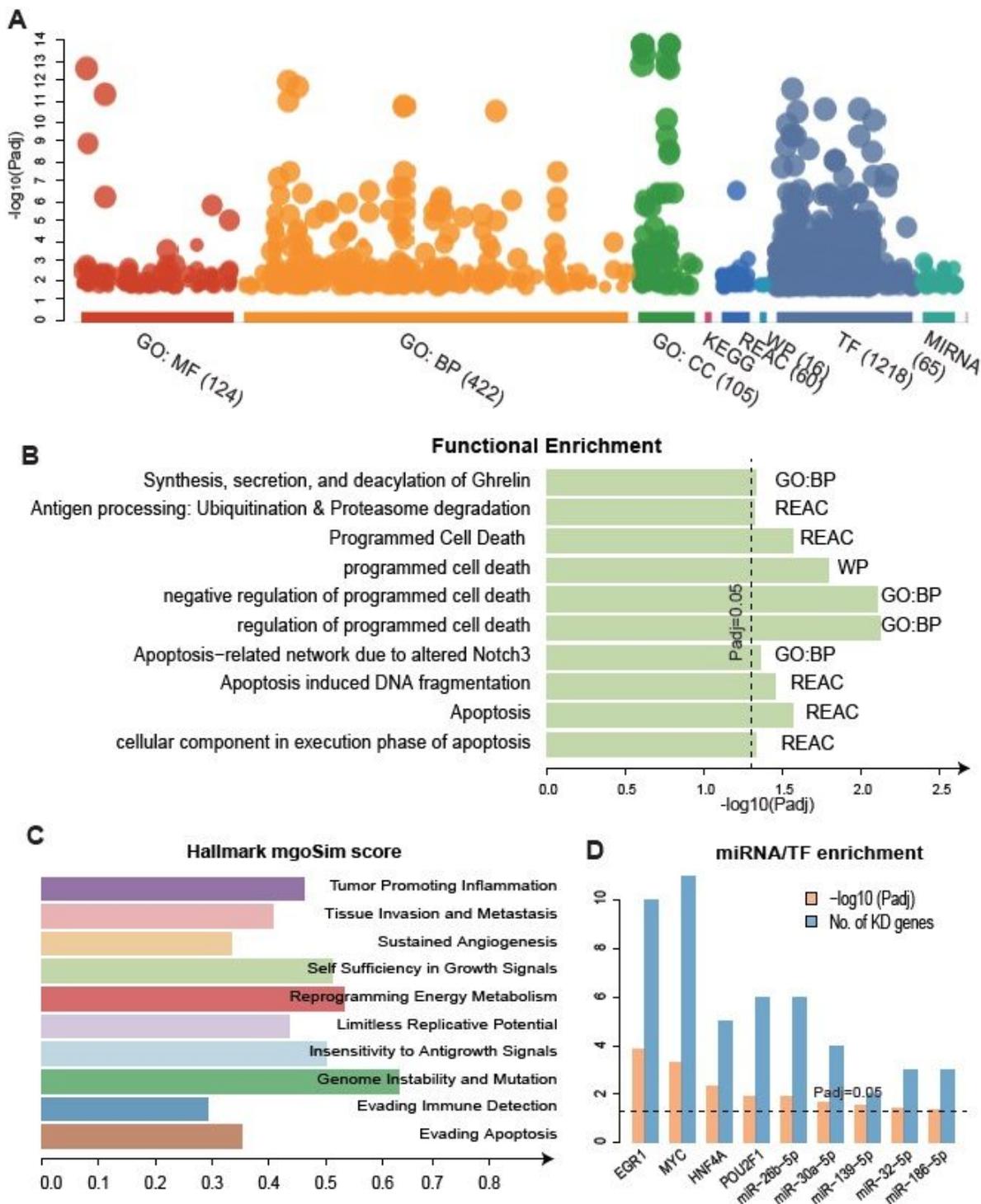
Screening genomic variation candidate genes using multi-omics data A, Data flow of the candidate genes screening process. BC, The examples show the copy number alteration of the candidate KD genes DERL1 (B) and NAA15 (C) on DNA and the expression level on the mRNA, as well as the expression level of its dysregulatory factors. DE, Examples showing the effects of candidate KD genes DERL1 (D) and NAA15 (E) copy number alteration on expression level (Wilcoxon Rank Sum Test). FG, sample distribution of candidate KD genes, including before (F) and after (G) the addition of mutation information.



**Figure 3**

Construction of prognostic-related transcriptional dysregulatory network of KD genes AB, Examples show the prognostic efficacy of KD genes ABCE1 (A) and SHFM1 (B) gastric cancer patients. C, Genomic variation of all recognized KD genes in gastric cancer patients, including high copy number copy (dark red), low level amplification (bright red), homozygous copy number deletion (dark blue), heterozygous

deletion (Bright blue), as well as somatic mutation (green). D, The transcriptional dysregulatory network of KD genes. E, The number of dysregulatory factors for KD genes.



**Figure 4**

Functional enrichment analysis of KD genes A, Functional enrichment analysis for KD genes using g:profiler online tool, including Gene Ontology (MF, BP, CC) and pathway (REAC and WP) in different colors. The Y axis represents  $-\log_{10}(\text{Padj})$ . B, The functions enriched by KD genes. The X axis represents  $-\log_{10}(\text{Padj})$ .

log10 (Padj), and the dotted line indicates Padj=0.05. C, The semantic similarity between the functions enriched by KD genes and the known cancer hallmarks. D, The regulatory factors for KD genes confirmed by enrichment analysis. Blue indicates the number of KD genes, and orange indicates significance, -log10 (Padj).

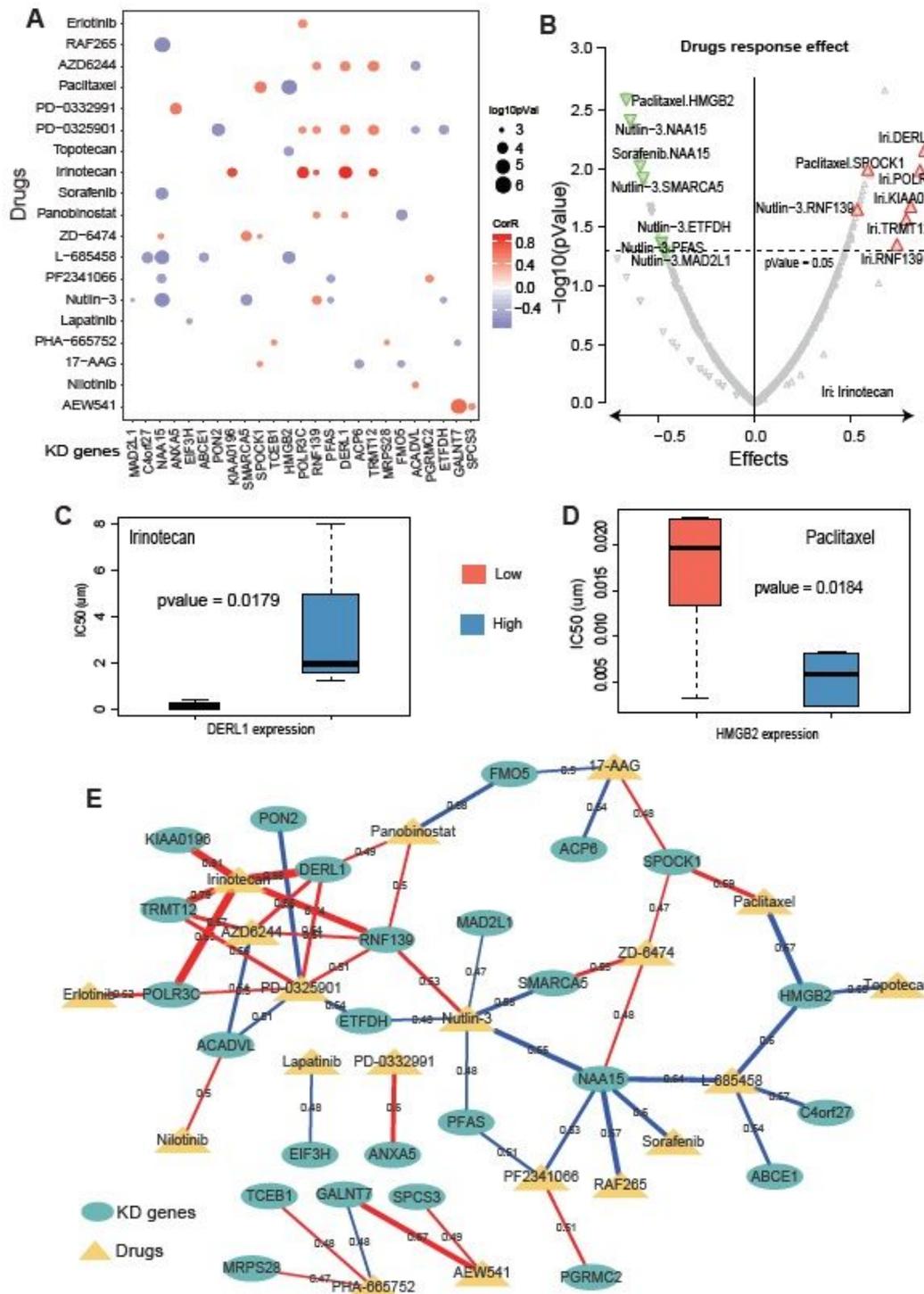
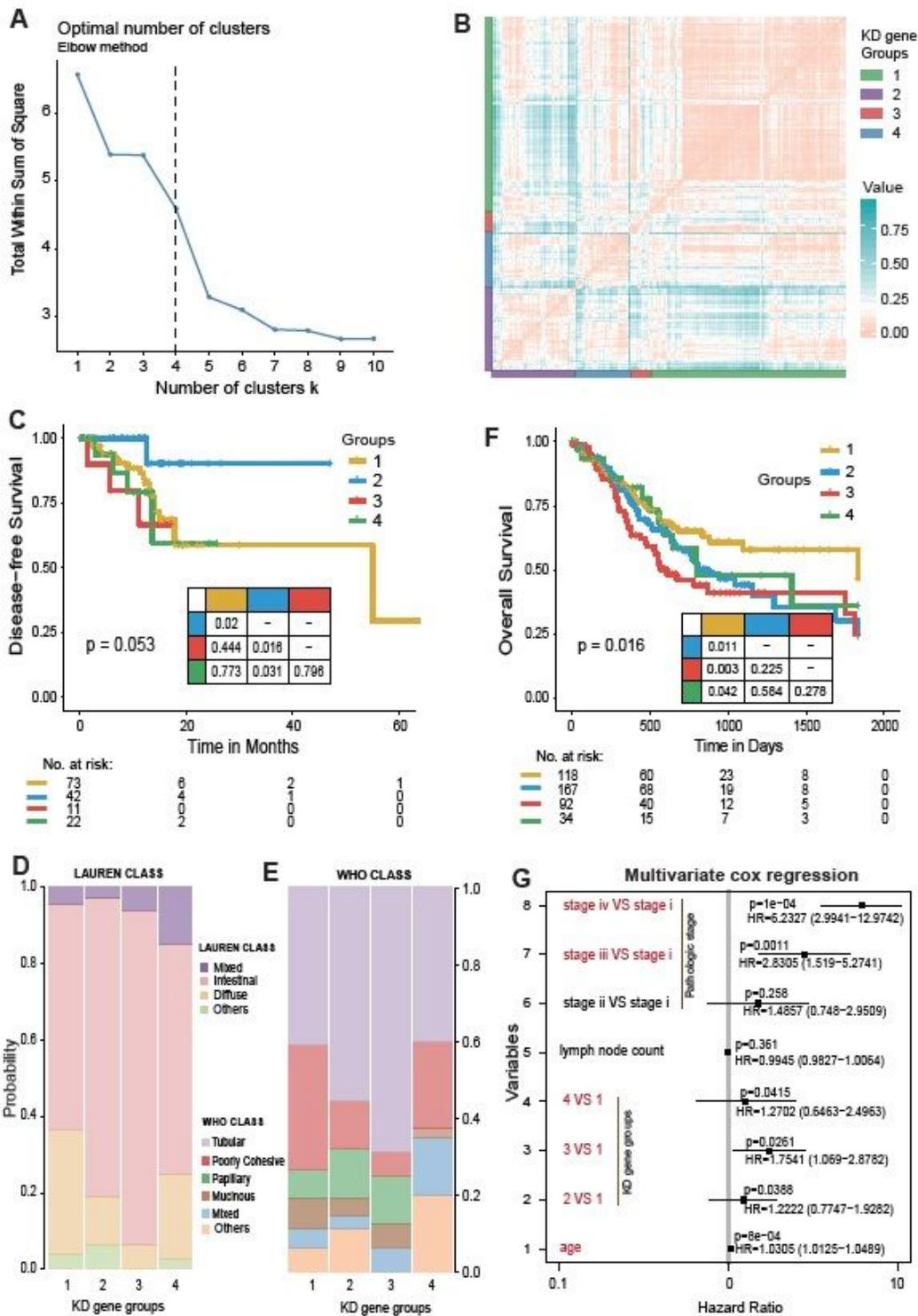


Figure 5

Analysis of drug response effects of KD genes A, Correlation analysis between the expression level of KD genes and drug IC50 in CCLE cell model. The point size indicates the level of significance, and the color indicates the correlation coefficient, red (positive), blue (negative). Spearman's Rank correlation. B, The volcano map shows the response of particular drugs, red indicates resistance, and green indicates sensitivity. CD, The response pattern of specific drugs on KD genes, irinotecan showed drug resistance on DERL1 (C), and paclitaxel showed drug sensitivity on HMGB2 (D). E, The drug response pattern of the drugs in all KD genes in the CCLE cell model, including drug resistance (red) and sensitivity (blue), and line thickness indicates response effect.



**Figure 6**

Prognostic efficacy of KD gene signatures A, Determine the optimal number of clusters in hierarchical clustering using Elbow method. B, Clustering analysis were performed based on the expression level of KD genes, which showed pearson dissimilarity between patients with gastric cancer. C, The KM curve shows disease-free survival (DFS) time of patients across all groups, log rank test. DE, The proportion of patient groups of KD gene signatures in clinical features, including LAUREN (D) and WHO class (E). F, The

KM curve shows the overall-survival (OS) time of patients in extra data. G, Multivariate COX regression analysis (corrected tumor stage, lymph node count, and age). Red means  $p \leq 0.05$ .

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfile.docx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)