

# CXCR4 Negatively Correlates with NKT Cell Infiltration During Gastric Precancerous Lesions Progression to Early Gastric Cancer

**Xiaotao Jiang**

Guangzhou University of Chinese Medicine

**Kunhai Zhuang**

Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital

**Kailin Jiang**

Guangzhou University of Chinese Medicine

**Yi Wen**

Guangzhou University of Chinese Medicine

**Linling Xie**

Guangzhou University of Chinese Medicine

**Senhui Weng**

Guangzhou University of Chinese Medicine

**Xu Chen**

Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital

**Kechao Nie**

Guangzhou University of Chinese Medicine

**Zhijia Zheng**

Guangzhou University of Chinese Medicine

**Jinglin Pan**

Guangzhou University of Chinese Medicine Affiliated Hospital of Traditional Chinese Medicine: Hainan Province Hospital of Traditional Chinese Medicine

**Yanhua Yan**

Guangzhou University of Chinese Medicine

**Peng Liu**

Guangzhou University of Chinese Medicine

**Junhui Zheng**

Guangzhou University of Chinese Medicine

**Yuanheng Huang**

Guangzhou University of Chinese Medicine

**Yufeng Liu**

Guangzhou University of Chinese Medicine

**Peiwu Li**

Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital

**Fengbin Liu** (✉ [liufengbin163@163.com](mailto:liufengbin163@163.com))

Guangzhou University of Traditional Chinese Medicine First Affiliated Hospital

---

## Research article

**Keywords:** CXCR4, NKT cell, early gastric cancer (EGC), weighted gene co-expression analysis (WGCNA)

**Posted Date:** October 12th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-88936/v1>

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

[Read Full License](#)

---

# Abstract

**Background** Immune microenvironment in gastric precancerous lesions (GPL) and early gastric cancer (EGC) still remain largely unknown. This study aims to identify key immune cells and hub genes associated with GPL progression to EGC.

**Methods** Immune Cell Abundance Identifier (ImmuCellAI) algorithm was used to quantify the proportions of immune cells of GPL and GC samples based on gene expression profiles. Key immune cells associated with GPL progression to EGC were identified using one-way analysis of variance (ANOVA) test and Spearman's correlation test. Weighted gene co-expression analysis (WGCNA) and pathway enrichment were adopted to identify hub gene co-expression network and hub genes associated with the key immune cells infiltration. The pattern of key immune cells infiltration, hub genes expression and their correlation were verified in an independent GPL-EGC cohort and GC datasets.

**Results** NKT cell was found gradually decreased during GPL progression to EGC and negatively correlated with tumorigenesis. According to WGCNA and hub genes screening, CXCR4, having a poor prognosis, increased with GPL progression, positively correlated with tumorigenesis and negatively correlated with NKT cell infiltration significantly, was identified as the real hub gene. The negative correlation between CXCR4 and NKT cell infiltration was successfully verified in an independent GPL-EGC cohort and GC datasets.

**Conclusion** CXCR4 and NKT cell are possible to serve as biomarkers in monitoring GPL progression to EGC. Besides, CXCR4 may be involved in regulating NKT cell infiltration during GPL progression to EGC, which may provide a new immunotherapeutic target.

## Background

Gastric cancer (GC) is the fifth most common malignancy and the third leading cause of cancer-related mortality[1]. Histologically, Lauren categorized GC into two subtypes: intestinal and diffuse types [2]. Intestinal-type GC (IGC), with a clear and multistep histological evolution starting from chronic inflammation and progressing to atrophy, intestinal metaplasia, gastric precancerous lesion [GPL, including low-grade intraepithelial neoplasia (LGIN) and high-grade intraepithelial neoplasia (HGIN)] and frank malignancy, is closely related with inflammatory and immune response [3, 4]. As early gastric cancer (EGC) has a far better prognosis than advanced GC, early diagnosis and treatment are essential [5].

Inflammatory cell infiltration exhibits a positive or negative effect on tumor invasion, growth, metastasis, and prognosis [6]. With the coming of immunotherapy era, immunotherapy is gradually playing a vital role in the treatment of GC and much attention has been paid to the specific stage of GC [7]. However, immune microenvironment of GPL and EGC still remain largely unknown.

With the development of public databases, such as the Gene Expression Omnibus (GEO), large amounts of genome, epigenome, transcription and proteomics data have become more accessible [8]. Correspondingly, multiple algorithms for analyzing omics data have been developed over the past few years [9], which provides more opportunities to uncover the immune microenvironment and molecular mechanisms of carcinogenesis. For instance, Immune Cell Abundance Identifier (ImmuCellAI) algorithm can estimate the abundance of immune cell based on transcriptome data [10]. Weighted gene co-expression analysis (WGCNA) can identify key biological targets and therapeutic targets through constructing a co-expressed gene network and exploring its correlation with clinical traits [11].

In current study, we used ImmuCellAI to quantify the proportions of immune cells of GPL and GC sample based on gene expression profiles. Combining with one-way ANOVA and correlation analysis, key immune cells associated with GPL progression to EGC were identified. WGCNA and pathway enrichment were adopted to identify hub gene co-expression network and hub genes associated with the key immune cells infiltration. The pattern of key immune cells infiltration, hub genes expression and their correlation were verified in an independent GPL-EGC cohort and GC datasets.

## Methods

### Data acquisition

The research strategy is shown in Fig. S1. Raw files of three GPL-related gene expression profiles (GSE55696 [12], GSE87666 [13] and GSE130823 [14]) were downloaded from the GEO database on NCBI ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). GSE55696 performed on the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F includes 19 LGIN, 20 HGIN, 19 EGC and 19 chronic gastritis (CG) samples. GSE87666 and GSE130823 both performed on the Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 include 7 LGIN, 9 HGIN, 6 EGC and 17 LGIN, 14 HGIN, 15 EGC, 47 paired inflammation controls respectively. In order to investigate the difference between individuals from different pathological stage, data of 47 paired inflammation controls of GSE130823 were excluded.

Raw files of three GC data sets on the GEO database with clinical information (GSE66229 [15] (normal controls (NC) n = 100, GC n = 300), GSE15459 [16] (GC n = 200), GSE34942 [17] (GC n = 56)) performed on the same microarray platform of [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array were downloaded. RNA sequencing data of 375 stomach adenocarcinoma (STAD) samples with clinical information from The Cancer Genome Atlas (TCGA) and 175 stomach NC from Genotype-Tissue Expression (GTEx) datasets were obtained from the UCSC Xena browser (<https://xenabrowser.net/>) [18].

Characteristics of datasets in this study were displayed in Table 1.

### Data preprocessing

Raw data of the three GPL-related gene expression profiles were extracted. Background correction and normalization were performed by “limma” package of R respectively [19].

Data of the three GC data sets were extracted and normalized using the RMA algorithm. Then they were transformed to the form of log base 2 and consolidated as a GC meta-GEO cohort totally with 100 NC and 556 GC. The batch effects were removed by the “combat” function of “sva” package of R [20].

As for GTEx and TCGA datasets, RNA sequencing data (FPKM values) were transformed into log<sub>2</sub>(FPKM+1). Then, they were merged as a STAD TCGA-GTEx cohort and normalized according to the “normalizeBetweenArrays” function of the package of “limma” in R software so that the expression values have similar distribution across a set of arrays [19].

### **Estimation of immune cells abundance**

Immune Cell Abundance Identifier (ImmuCellAI, <http://bioinfo.life.hust.edu.cn/web/ImmuCellAI/>) is a gene set signature-based method for precisely estimating the abundance of 24 immune cell types, especially T-cell properties (18 T-cell subsets). It is reported that it has the best performance in immune cell abundance estimation [10], compared with other five methods (xCell [21], CIBERSORT [22], EPIC [23], MCP-counter [24], and TIMER [25]). Therefore, gene expression data of GPL-EGC cohorts (GSE55696, GSE87666 and GSE130823), GC meta-GEO cohort and GC TCGA-GTEx cohort were submitted to ImmuCellAI to acquire the estimation of immune cells abundance.

### **Identification of key immune cells associated with GPL progression to EGC**

We used “stage” to represent the severity of the pathology of tissues. In following correlation calculation, CG was represented by “1”, LGIN was represented by “2”, HGIN was represented by “3” and EGC was represented by “4”. The relationship between immune cell abundance with pathological stage was preliminarily analyzed by one-way analysis of variance (ANOVA) test with  $p < 0.05$  set as the cut-off. Then, correlation between immune cell abundance and pathological stage was calculated using Spearman’s correlation test. Common immune cells with a Spearman’s  $r$  greater than 0.5 and a  $p$ -value less than 0.05 in GSE55696 and GSE87666 cohorts were considered as key immune cells. After further verification between GC and NC in the GC meta-GEO cohort and STAD TCGA-GTEx cohort, immune cells with correspondingly consistent alterations were subjected for identifying related gene co-expression network.

### **Construction of gene co-expression networks**

In order to remove outlier samples, sample clustering was performed based on Pearson’s correlation matrices and the average linkage method. Through utilizing the ‘WGCNA’ package in R[11], top 25% genes with the largest variance differences were used to construct weight gene co-expression networks in GSE55696 and GSE87666 datasets respectively. Firstly, a similarity matrix was established based on the Pearson’s correlation value between the paired genes. Then, adjacency matrix was created using following formula [26]:  $a_{mn} = |c_{mn}|^\beta$  ( $a_{mn}$ : adjacency matrix between gene  $m$  and gene  $n$ ,  $c_{mn}$  = Pearson’s correlation between paired genes,  $\beta$ : softpower value) and converted into a topological overlap matrix. The value of soft threshold power was chosen when the scale free topology scale  $R^2$  exceeding 0.85.

Genes with similar expression patterns were categorized into modules by average linkage hierarchical clustering with a module minimum size set as 100.

### **Identification of hub modules associated with key immune cells**

Module eigengenes (MEs) was the major component of a gene module. To identify the modules significantly associated with key immune cells infiltration, we calculated the correlation between MEs and the infiltration level of key immune cells using Pearson's correlation test in GSE55696 and GSE87666 cohorts respectively, with  $p$ -value  $< 0.05$  set as the cut-off. Then, we took intersection of the modules with same correlation direction in GSE55696 and GSE87666 cohorts respectively, and performed KEGG pathway enrichment analysis on the overlapping genes to understand their potential functions to assist in determining the hub modules, with a cut-off criterion of adjusted  $p$ -value  $< 0.05$ .

### **Identification of candidate genes**

After hub modules identification, candidate hub genes were identified. Module hub genes, which are highly connected intra-modular genes, have the high Module Membership (MM) scores to the respective module. MM is the parameter that correlates the expression profile of a gene with the MEs of a module which can measure the distance between genes and a given module. The absolute value of gene significance (GS), measuring the Pearson's correlation between a given gene and the clinical trait, can also serve as the parameter to filter out the hub genes. We set the  $|MM| \geq 0.70$  and the  $|GS| \geq 0.50$  for hub genes in co-expression network. Meanwhile all genes in the hub module were uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING; <https://string-db.org/>) database [27] to construct protein-protein interaction (PPI) network with the species limited to 'Homo sapiens' and confidence  $> 0.9$ . We used Cytoscape to visualize the network (<https://cytoscape.org/>) [28]. Genes with node degree ranking in top 10% were considered as central nodes in PPI network. Both in the hub modules of GSE55696 and GSE87666 datasets, common genes serve as hub genes in co-expression network and central nodes in PPI network were considered as the candidate hub genes, which was visualized in the form of Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

### **Identification of real hub genes**

For the candidate hub genes, in STAD TCGA-GTEX and GC meta-GEO cohort, we compared their expression level between GC and NC and performed prognosis analysis with the quartile setting as the group cutoff. Genes with significant difference between GC and NC and prognostic significance were identified as real hub genes.  $P$ -value was defined as significantly different when it was less than 0.05 in above analysis.

### **Real hub genes expression patterns and relationship with pathological stage and key immune cells in GPL-EGC datasets and GC dataset**

The expression level of real hub genes was plotted as box plots and the relationship with pathological stage was analyzed by one-way ANOVA test and Spearman correlation test in primary GPL-EGC datasets

(GSE55696 and GSE87666). Correlation between hub genes and key immune cells was evaluated using Pearson's correlation test. *P*-value was defined as significantly different when it was less than 0.05 in above analysis.

### **Verification of hub genes and key immune cells as well as their relationship**

In order to confirm the pattern of key immune cells infiltration and hub genes expression as well as their relationship, we plotted the expression level of hub genes and the abundance of key immune cells across different pathological stages of LGIN, HGIN, and EGC in the independent cohort GSE130823 and analyzed by one-way ANOVA test and Spearman's correlation test. Their correlation was also verified using Pearson's correlation test in the independent GPL-EGC cohort GSE130823 and GC datasets (STAD TCGA-GTEx and GC meta-GEO cohort). *P*-value was defined as significantly different when it was less than 0.05 in above analysis.

## **Results**

### **The immune cell infiltration landscape and key immune cells identification during GPL progression to EGC**

24 kinds of immune cell abundance were estimated by ImmuCellAI and analyzed by one-way ANOVA test in terms of pathological stage in GSE55696 and GSE87666 cohort. Combining with the results of above two cohorts, the abundances of CD8 naive cell and NKT cell were significantly different across pathological stages (one-way ANOVA test *p*-value < 0.05) while only NKT consistently showed a linear downward trend in the transformation process in the two cohorts (Fig. 1a-b). The correlation between 24 immune cell abundance and pathological stage in two cohorts was displayed in heatmaps (Fig. 1c-d). Under the criterion of Spearman's *r* greater than 0.5 and a *p*-value less than 0.05, in GSE55696, Effector memory T cell (*r* = -0.58, *p* = 3e-08) and NKT cell (*r* = -0.58, *p* = 3e-08) were significantly negatively correlated with pathological stage while Macrophage (*r* = 0.61, *p* = 3e-09) was opposite. As for GSE87666, only NKT cell (*r* = -0.63, *p* = 0.01) met the criterion of significance. Combining with the results of Spearman's correlation test and one-way ANOVA test in above two cohorts, NKT cell was considered as key immune cell associated with GPL progression to EGC (Fig. 1e). With further verification in the GC meta-GEO cohort and STAD TCGA-GTEx cohort, it was significantly downregulated in GC compared with NC (Fig. 1f-g).

### **Identification Of Hub Modules Correlated With Nkt Cell Infiltration**

To identify hub modules correlated with NKT cell infiltration, we took intersection of the modules with same correlation direction in GSE55696 and GSE87666 datasets respectively [positive: blue module (GSE55696) ∩ red module (GSE87666), yellow module (GSE55696) ∩ red module (GSE87666)]; negative:

pink module (GSE55696)  $\cap$  green module (GSE87666), green module (GSE55696)  $\cap$  green module (GSE87666), red module (GSE55696)  $\cap$  green module (GSE87666)]. Then, KEGG pathway enrichment analysis was performed on the overlapping genes. The overlapping genes of GSE55696 and GSE87666 green modules were enriched in immune-related pathways (Fig. S3a), while others had not found any pathway that can be enriched (Fig. S3b-d) or the enriched pathway had no obvious relationship with immunity (Fig. S3e). Therefore, green modules of GSE55696 and GSE87666 were considered as hub modules correlated with NKT cell infiltration and subjected for further analysis.

## Identification Of Candidate Hub Genes

There were 66 overlapping genes shared in two hub modules (Fig. 3a) and they were enriched in immune-related pathways including 'chemokine signaling pathway', 'cytokine-cytokine receptor interaction', 'leukocyte transendothelial migration' 'primary immunodeficiency', 'hematopoietic cell lineage', 'intestinal immune network for IGA production', 'Leishmania infection' and 'B cell receptor signaling pathway' (Fig. 3b). The correlation between green module's GS and MM in GSE55696 and GSE87666 was calculated and found that they have a positive relationship ( $r = 0.87$ ,  $p = 2.3e - 119$ ;  $r = 0.50$ ,  $p = 9.7e - 19$ ), which indicated that hub genes of module tend to be associated with NKT cell infiltration (Fig. 3c-d). PPI networks of the green modules in GSE55696 and GSE87666 were constructed with a cutoff confidence  $> 0.9$  (Fig. 3e-f). Based on  $|MM| \geq 0.70$  and  $|GS| \geq 0.50$ , 184 genes in the green module of GSE55696 and 94 genes in the green module of GSE87666 were selected as candidate hub genes in co-expression network, respectively. As for PPI network, 34 genes in the green module of GSE55696 and 22 genes in the green module of GSE87666 were considered as central nodes with node degree ranking in top 10%. After conducting Venn analysis, in the green modules of GSE55696 and GSE87666 cohort, 4 genes (CXCR4, CD53, IL10RA and CD19), serving as both candidate hub gene in co-expression network and central node in PPI network, was regarded as the candidate hub gene negatively associated with NKT cell infiltration (Fig. 3g).

## Discussion

In the current study, we estimated the abundance of immune cells in GPL and EGC related datasets using ImmuCellAI algorithm [10]. NKT was found gradually decreased with GPL progression to EGC and was identified as key immune cell associated with tumorigenesis using one-way ANOVA test and Spearman's correlation test. Further verification indicated that it was significantly downregulated in GC in meta-GEO cohort and STAD TCGA-GTEx cohort. Co-expression networks of NKT cell were constructed using WGCNA. Green modules in GSE55696 and GSE87666 having a significantly negative correlation with NKT cell infiltration were identified as hub modules as their overlapping genes were significantly enriched in immune-related pathways. In further screening, CXCR4, CD53, IL10RA and CD19 were identified as the candidate hub genes according to gene network related parameters. Combining expression comparison and prognosis analysis, CXCR4 was determined as the real hub gene. CXCR4 expression was increased with GPL progression, significantly positively correlated with tumorigenesis and negatively correlated with

NKT cell infiltration. The pattern of NKT cell infiltration and CXCR4 expression as well as their relationship stay consistent in the independent GPL-EGC cohort GSE130823 and GC datasets (STAD TCGA-GTEX and GC meta-GEO cohort).

NKT cells (usually defined as CD3 + CD56+), are a broad group of CD3 + T cells that co-express T-cell antigen receptor (TCR) and NK-cell markers [29, 30] They share characteristics from both NK and T cells and possess both innate and acquired immune functions. On the one hand, they can be activated to secrete cytotoxic enzymes and cytokines to kill target cells once TCR ligation [31]. On the other hand, they can mediate non-MHC restricted lysis and cytokine production in the absence of TCR activation. Therefore, they were considered to play an important role in anti-tumor and anti-virus immune response [32, 33]. Zhuang *et al* [34] has found that the frequencies of CD3 + CD56 + NKT-like cells in GC tumors were significantly decreased and the effector function had impaired, which were consistent with our findings. They also found that the mechanism of the impairment of CD3 + CD56 + NKT-like cells from GC patients was not resulted from coinhibitory molecules such as PD-1, Tim-3, LAG-3 and TIGIT but from some soluble inhibitory factors released by tumor cells or tumor-infiltrating immune cells. However, the specific factors remained to be explored.

We found that CXCR4 expression was increased with GPL progression and negatively correlated with NKT cell infiltration during GPL progression to EGC. CXCR4, is a seven-span transmembrane G-protein coupled-receptor that is the primary receptor for CXCL12 [35]. It is reported that CXCR4 plays an important role in tumor biological behaviour, such as growth, metastasis, angiogenesis and cancer cell-microenvironment interaction [36]. In GC, CXCR4 is upregulated and associated with poor prognosis [37, 38]. It was reported that SDF-1/CXCR4 axis could facilitate myeloid-derived suppressor cells (MDSCs) accumulation in tumor microenvironment and suppresses T cell immune responses [39, 40]. In current study, we speculated that some kinds of immunosuppressive cells expressing CXCR4 may gradually accumulate in the tumor microenvironment and inhibit the immune response of NKT cells, leading to GPL progression.

## Conclusions

In conclusion, NKT cell and CXCR4 are possible to serve as biomarkers in monitoring GPL progression. CXCR4 may be involved in NKT cell infiltration during GPL progression to EGC, which may provide a new immunotherapeutic target.

## Abbreviations

GPL

gastric precancerous lesions; EGC:early gastric cancer; ImmuCellAI:Immune Cell Abundance Identifier; ANOVA:analysis of variance; WGCNA:weighted gene co-expression analysis; GC:gastric cancer; IGC:Intestinal-type GC; LGIN:low-grade intraepithelial neoplasia; HGIN:high-grade intraepithelial neoplasia; GEO:Gene Expression Omnibus; NC:normal controls; STAD:stomach adenocarcinoma; TCGA:the cancer

genome atlas; GTEx:genotype-tissue expression; MEs:module eigengenes; MM:module membership; GS:gene significance; PPI:protein-protein interaction; TCR:T-cell antigen receptor

## **Declarations**

### **Acknowledgments**

Not applicable.

### **Authors' contributions**

FL, LY, PL designed the study. XJ, KZ, KJ and carried out the analysis. YW, LX, SW, XC, KN, and ZZ participated in coordination of the study, interpretation of results. JP, YY and PL participated in manuscript writing. JZ and YH revised the manuscript. All authors read and approved the final manuscript

### **Funding**

Natural Science Foundation of China (Grant No. 81973819).

Youth program of Natural Science Foundation of China (Grant No. 81904139, 81904145).

Major Subject Research Projects of Guangzhou University of Traditional Chinese Medicine (Grant No. A1-2606-19-110-007).

Liu Fengbin, Guangdong famous traditional Chinese medicine inheritance studio (Grant No. Guangdong TCM office [2020] no. 1).

“Innovation Foster Hospital” Program of first affiliated hospital of Guangzhou University of Chinese medicine (Grant No. 2017TD05).

Breakthrough Project of TCM Dominant Diseases with Special Funds for Strengthening Traditional Chinese Medicine in 2015 (No.19), Guangdong Administration of Traditional Chinese Medicine (Grant No. Document of TCM of Guangdong [2015] No. 19).

Major Projects of First-Class Disciplines in Guangzhou University of TCM in 2018 (Grant No. A1-AFD018181A27).

National Research Project on Traditional Chinese Medicine (Grant No. 201507001-09);

Natural Science Foundation of Guangdong Province (Grant No. 2017A030310121, 2019A1515011145).

Guangdong Provincial Science and technology plan project (South China traditional Chinese medicine Collaborative Innovation Center) (Grant No. 2014B090902002).

Innovation and strengthening project of the First Affiliated Hospital of Guangzhou University of traditional Chinese Medicine (Grant No. 211010010101).

### **Availability of data and materials**

The data of this study are available from TCGA, GTEx, and GEO database.

### **Ethical approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### **Author details**

<sup>1</sup>Department of Gastroenterology, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong, China. <sup>2</sup>First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong, China. <sup>3</sup>Department of Gastroenterology, Hainan Provincial Hospital of Traditional Chinese Medicine, Haikou 570100, Hainan, China.

## **References**

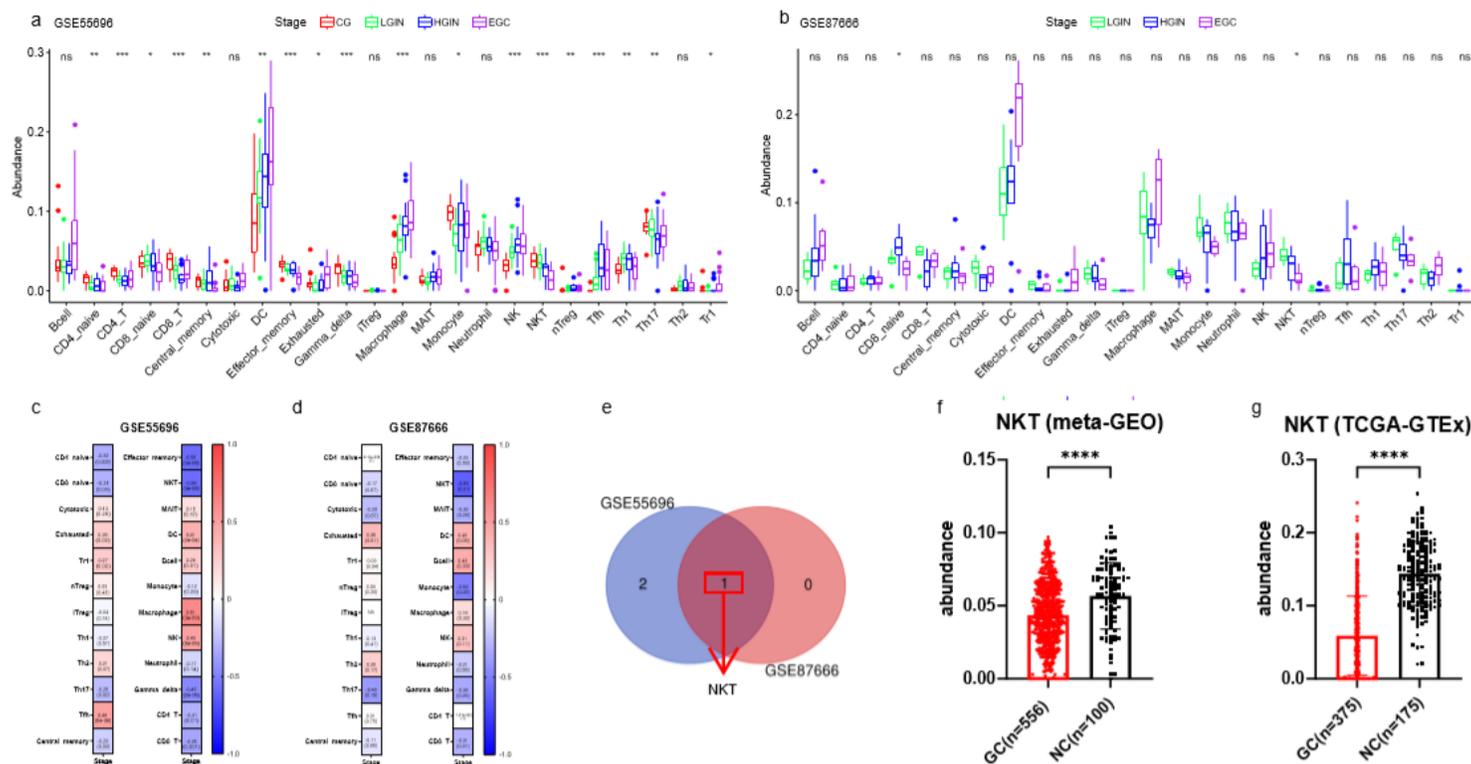
1. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. LAUREN P. The two histological main type gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histological classification. *Acta Pathol Microbiol Scand.* 1965;64:31–49.
3. Correa PA. A human model of gastric carcinogenesis. *Cancer Res.* 1988;48(13):3554–60.
4. Bornschein J, Kandulski A, Selgrad M, Malfertheiner P. From gastric inflammation to gastric cancer. *Dig Dis.* 2010;28(4–5):609–14.
5. Gotoda T, Yanagisawa A, Sasako M, Ono H, Nakanishi Y, Shimoda T, Kato Y. Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers. *Gastric Cancer.* 2000;3(4):219–25.
6. Banin-Hirata B, de Oliveira C, Losi-Guembarovski R, Ozawa P, Vitiello G, de Almeida F, Derossi D, André N, Watanabe M. The prognostic value of regulatory T cells infiltration in HER2-enriched breast cancer

- microenvironment. *Int Rev Immunol*. 2018;37(3):144–50.
7. Magalhães H, Fontes-Sousa M, Machado M. Immunotherapy in Advanced Gastric Cancer: An Overview of the Emerging Strategies. *Can J Gastroenterol & Hepatol*. 2018; 2018:2732408.
  8. Barrett T, Wilhite S, Ledoux P, Evangelista C, Kim I, Tomashevsky M, Marshall K, Phillippy K, Sherman P, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41:D991–5.
  9. Lin E, Lane H. Machine learning and systems genomics approaches for multi-omics data. *Biomark Res*. 2017;5:2.
  10. Miao Y-R, Zhang Q, Lei Q, Luo M, Xie G-Y, Wang H, Guo A-Y, ImmuCellAI. A Unique Method for Comprehensive T-Cell Subsets Abundance Prediction and its Application in Cancer Immunotherapy. *Advanced Science*. 2020;7(7):1902880.
  11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
  12. Xu X, Feng L, Liu Y, Zhou W, Ma Y, Fei G, An N, Li Y, Wu X, Yao F, et al. Differential gene expression profiling of gastric intraepithelial neoplasia and early-stage adenocarcinoma. *World J Gastroenterol*. 2014;20(47):17883–93.
  13. Kuang M, Cheng J, Zhang C, Feng L, Xu X, Zhang Y, Zu M, Cui J, Yu H, Zhang K, et al. A novel signature for stratifying the molecular heterogeneity of the tissue-infiltrating T-cell receptor repertoire reflects gastric cancer prognosis. *Sci Rep*. 2017;7(1):7762.
  14. Zhang Y, Wu X, Zhang C, Wang J, Fei G, Di X, Lu X, Feng L, Cheng S, Yang A. Dissecting expression profiles of gastric precancerous lesions and early gastric cancer to explore crucial molecules in intestinal-type gastric cancer tumorigenesis. *J Pathol*. 2020;251(2):135–46.
  15. Oh S, Sohn B, Cheong J, Kim S, Lee J, Park K, Lee S, Park J, Park Y, Lee H, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nature communications*. 2018;9(1):1777.
  16. Ooi C, Ivanova T, Wu J, Lee M, Tan I, Tao J, Ward L, Koo J, Gopalakrishnan V, Zhu Y, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLoS Genet*. 2009;5(10):e1000676.
  17. Lei Z, Tan I, Das K, Deng N, Zouridis H, Pattison S, Chua C, Feng Z, Guan Y, Ooi C, et al. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology*. 2013;145(3):554–65.
  18. 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 Res*. 2015;43:D812–7.
  19. Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, Smyth G. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
  20. Leek J, Johnson W, Parker H, Jaffe A, Storey J. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.

21. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
22. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
23. Racle J, De Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife.* 2017; 6.
24. Becht E, Giraldo N, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman W, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
25. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu J, Li B, Liu X. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017;77(21):e108–10.
26. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:Article17.
27. Szklarczyk D, Gable A, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva N, Morris J, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:D607–13.
28. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
29. Godfrey DI, Macdonald HR, Kronenberg M, Smyth MJ, Kaer LV. NKT cells: What's in a name? *Nat Rev Immunol.* 2004;4(3):231–7.
30. Kronenberg M. TOWARD AN UNDERSTANDING OF NKT. *CELL BIOLOGY: Progress and Paradoxes.* *Annu Rev Immunol.* 2005;23(1):877–900.
31. Golden-Mason L, Castelblanco N, O'Farrelly C, Rosen H. Phenotypic and functional changes of cytotoxic CD56pos natural T cells determine outcome of acute hepatitis C virus infection. *J Virol.* 2007;81(17):9292–8.
32. Lu PH, Negrin RS. A novel population of expanded human CD3 + CD56 + cells derived from T cells with potent in vivo antitumor activity in mice with severe combined immunodeficiency. *J Immunol.* 1994;153(4):1687–96.
33. Kokordelis P, Krämer B, Boesecke C, Voigt E, Ingiliz P, Glässner A, Wolter F, Srassburg C, Spengler U, Rockstroh J, et al. CD3(+)/CD56(+) Natural Killer-Like T Cells Display Anti-HCV Activity but Are Functionally Impaired in HIV(+) Patients With Acute Hepatitis C. *J Acquir Immune Defic Syndr.* 2015;70(4):338–46.
34. Peng LS, Mao FY, Zhao YL, Wang TT, Zhuang Y. Altered phenotypic and functional characteristics of CD3 + CD56 + NKT-like cells in human gastric cancer. *Oncotarget.* 2016;7(34):55222–30.
35. Loetscher M, Geiser T, O'Reilly T, Zwahlen R, Moser B. Cloning of a human seven-transmembrane domain receptor, LESTR, that is highly expressed in leukocytes. *J Biol Chem.* 1994;269(1):232–7.

36. Sleightholm RL, Neilsen BK, Li J, Steele MM, Singh RK, Hollingsworth MA, Oupicky D. Emerging roles of the CXCL12/CXCR4 axis in pancreatic cancer progression and therapy. *Pharmacol Ther.* 2017;158.
37. Nikzaban M, Hakhamaneshi MS, Fakhari S, Sheikhesmaili F, Roshani D, Ahsan B, Kamali F, Jalili A. The chemokine receptor CXCR4 is associated with the staging of gastric cancer. *Adv Biol Res.* 2014;3:16–6.
38. He H, Wang C, Shen Z, Fang Y, Wang X, Chen W, Liu F, Qin X, Sun Y. Upregulated expression of C-X-C chemokine receptor 4 is an independent prognostic predictor for patients with gastric cancer. *PLoS One.* 2013;8(8):e71864.
39. Obermajer N, Muthuswamy R, Odunsi K, Edwards RP, Kalinski P. PGE2-Induced CXCL12 Production and CXCR4 Expression Controls the Accumulation of Human MDSCs in Ovarian Cancer Environment. *Cancer Res.* 2011;71(24):7463–70.
40. Jiang K, Li J, Zhang J, Wang L, Shan L. SDF-1/CXCR4 axis facilitates myeloid-derived suppressor cells accumulation in osteosarcoma microenvironment and blunts the response to anti-PD-1 therapy. *Int Immunopharmacol.* 2019;75:105818.

## Figures



**Figure 1**

Landscape of immune infiltration cell and key immune cells identification. a, b Differences of 24 subtypes of immune cells across pathological stage in GSE55696 and GSE87666 cohorts. c, d Correlation heatmap of 24 immune infiltration cells abundance with pathological stage in GSE55696 and

GSE87666 cohorts. e Intersection of key immune cells in GSE55696 and GSE87666 cohorts. f, g NKT abundance between GC and NC in GC meta-GEO and STAD TCGA-GTEx cohort. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ , GC, gastric cancer; NC, normal controls; GEO, Gene Expression Database; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression

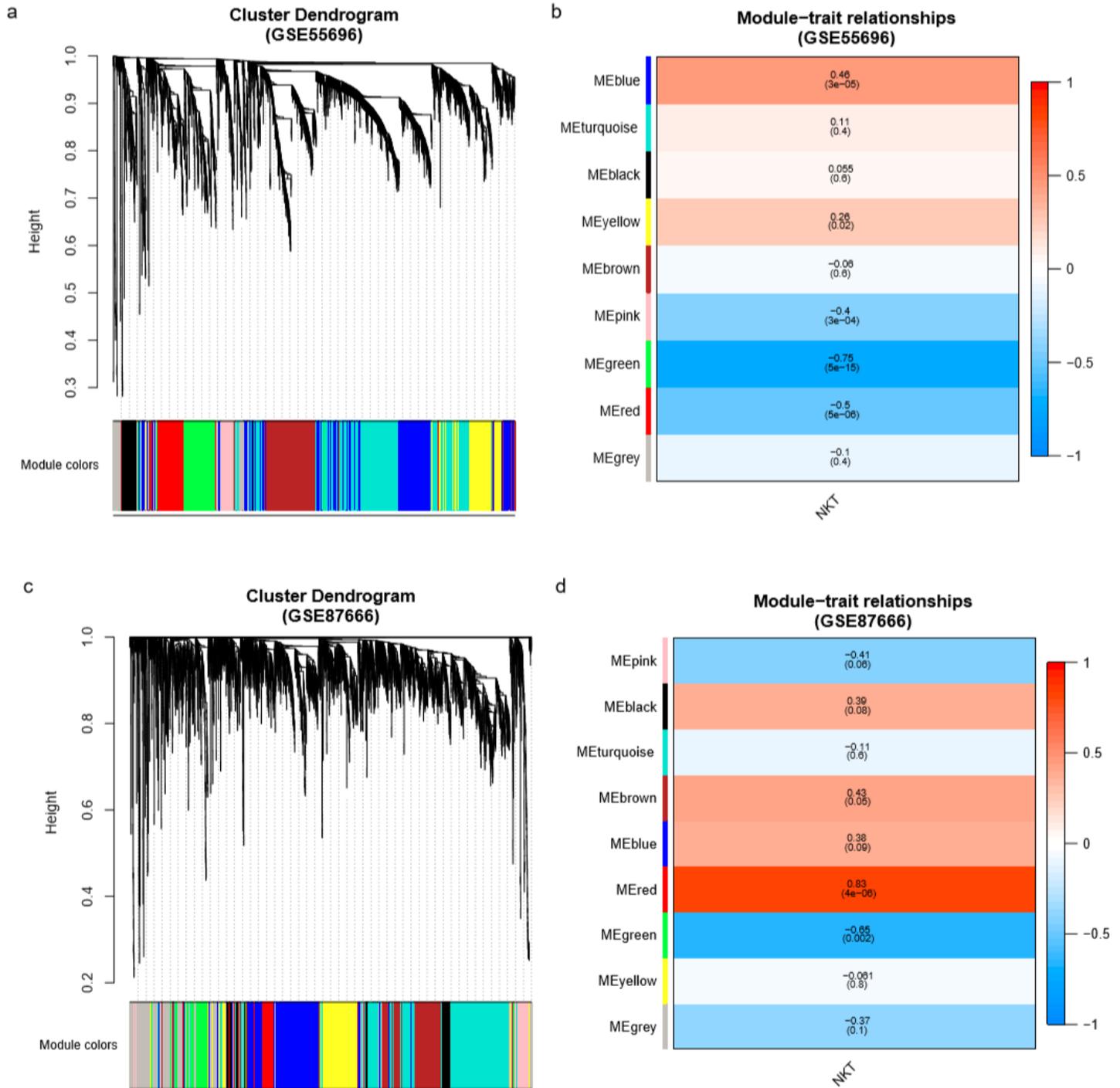
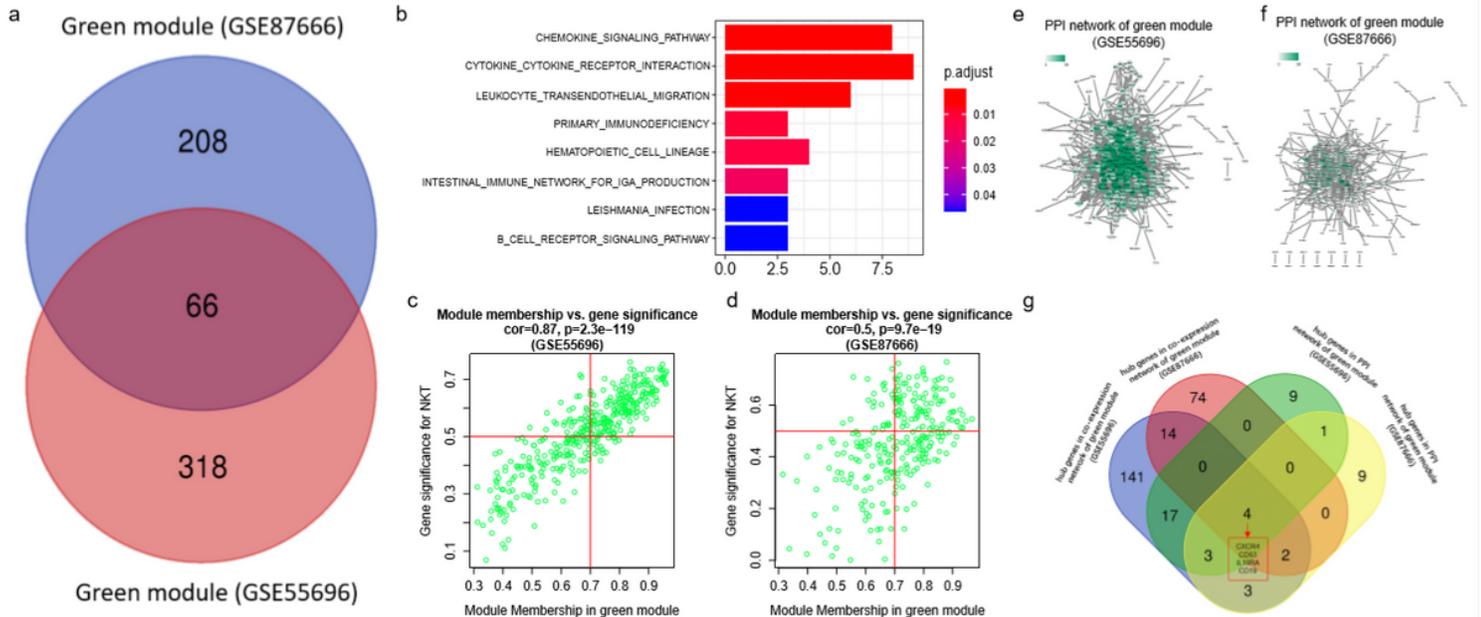


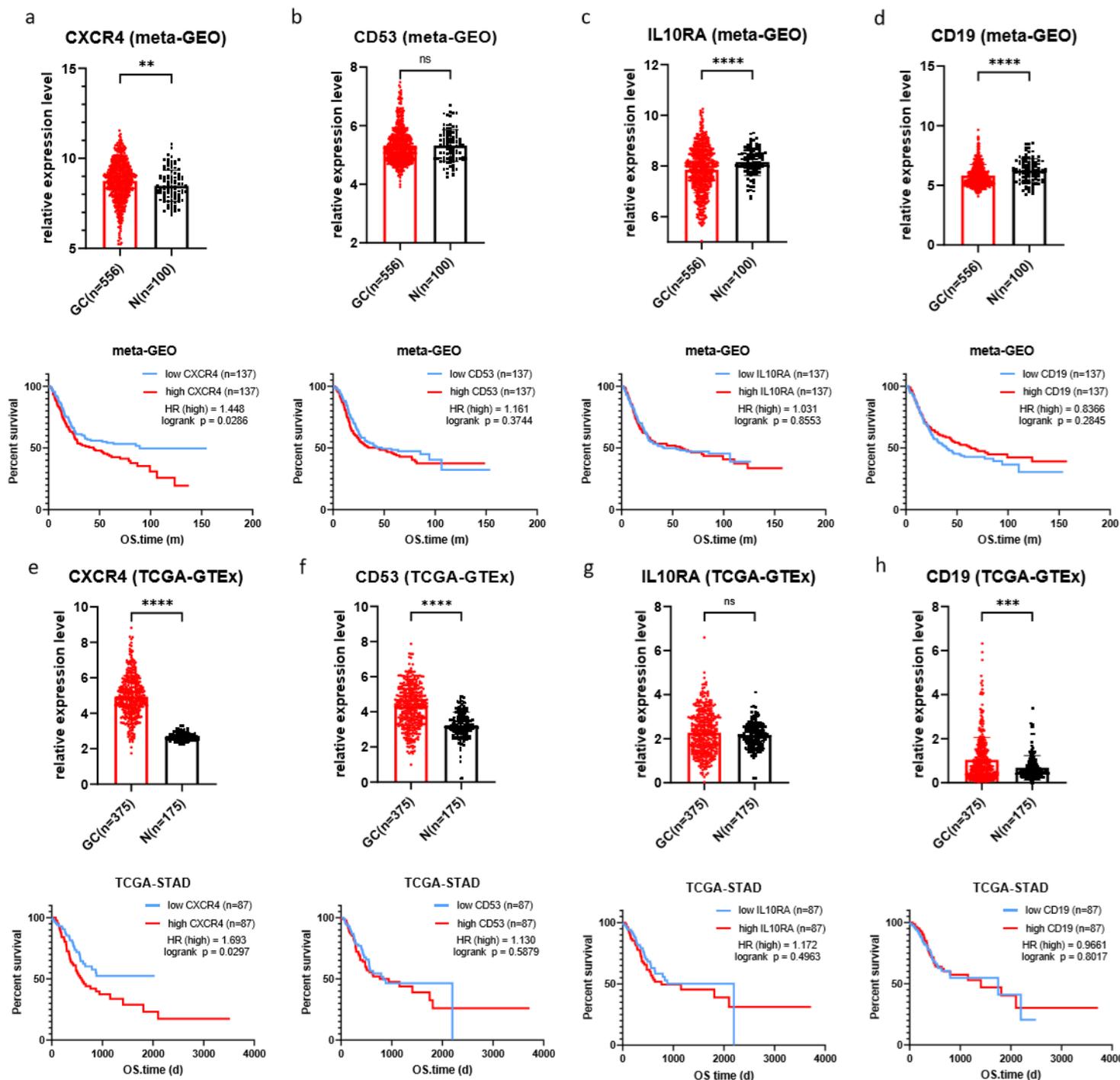
Figure 2

Construction of gene co-expression networks associated with NKT cell infiltration. The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix in a GSE55696 cohort and c GSE87666 cohort. Each module was assigned different colors. Module-trait relationships in b GSE55696 cohort and d GSE87666 cohort. Each row corresponds to a color module and each column correlates to NKT abundance. Each cell contains the corresponding correlation and p-value. TOM, Topological Overlap Matrix



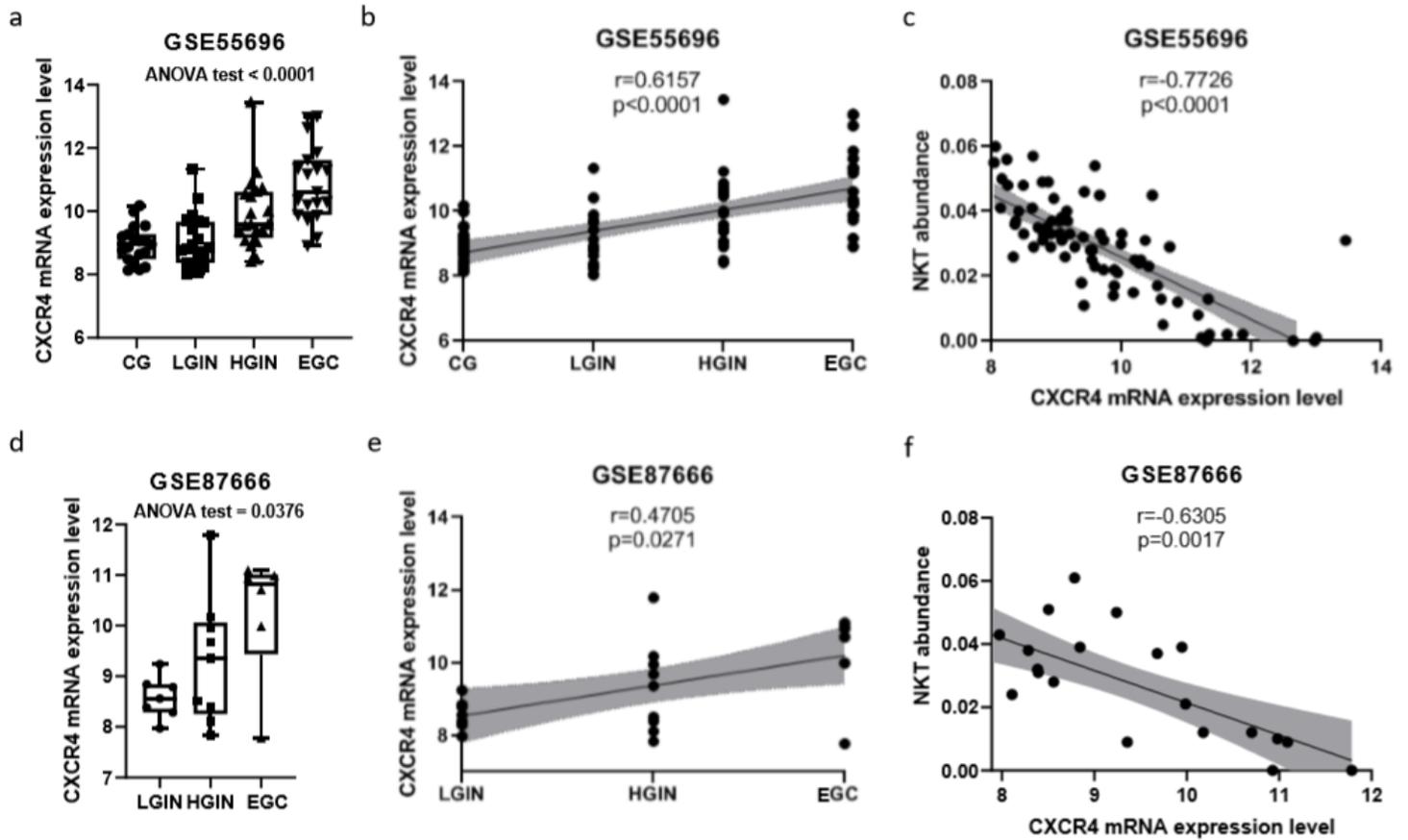
**Figure 3**

Identification of candidate hub genes. a Intersection of hub modules in GSE55696 and GSE87666 cohorts. b KEGG pathway enrichment analysis for overlapping genes of green modules from GSE55696 and GSE87666 cohorts. Scatter plot of module eigengenes in the green module in c GSE55696 cohort and d GSE87666 cohort. Protein–protein interaction network of genes in the green module in e GSE55696 cohort and f GSE87666 cohort. The color intensity in each node was proportional to the node degree. g Intersection of hub genes in PPI network and co-expression network and 4 candidate genes (CXCR4, CD53, IL10RA and CD19) were obtained. PPI, protein-protein interaction



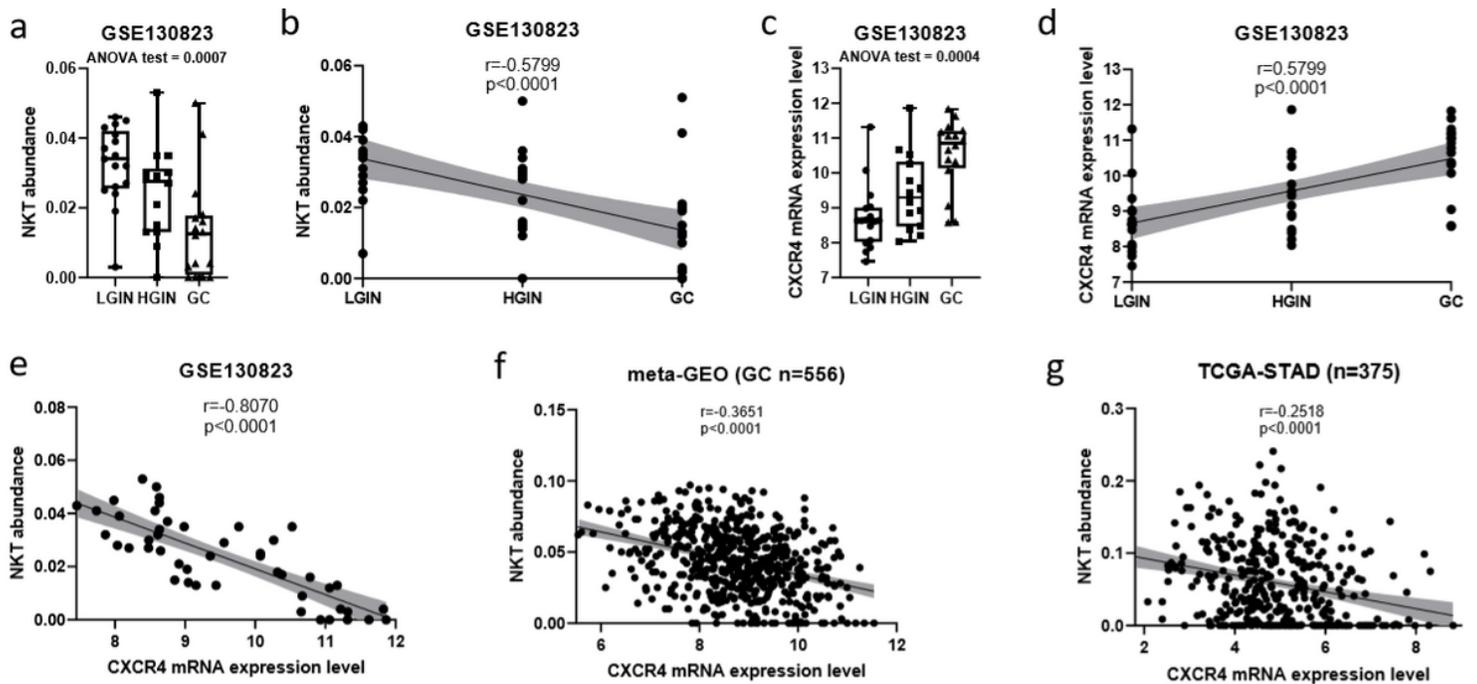
**Figure 4**

Expression comparison and prognosis analysis of 4 candidate hub genes in meta-GEO and TCGA-GTEx cohort. Relative expression level between GC and NC and Kaplan–Meier curve in GC patients of a CXCR4, b CD53, c IL10RA and d CD19 in GC meta-GEO cohort. Relative expression level between GC and NC and Kaplan–Meier curve in GC patients of e CXCR4, f CD53, g IL10RA and h CD19 in STAD TCGA-GTEx cohort. The patients were stratified into high-level group and low-level group according to quartile expression. GC, gastric cancer; NC, normal controls; GEO, Gene Expression Database; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression



**Figure 5**

CXCR4 expression pattern and correlation with NKT cell infiltration. Gene expression value CXCR4 among pathological stage in a GSE55696 cohort and d GSE87666 cohort. Scatter plot showing pathological stage (x-axis) and CXCR4 expression levels (y-axis) in b GSE55696 cohort and e GSE87666 cohort. Scatter plot showing CXCR4 expression levels (x-axis) and NKT cell abundance (y-axis) in c GSE55696 cohort and f GSE87666 cohort.



**Figure 6**

Verification of the pattern of NKT cell infiltration, CXCR4 expression and their relationship. a NKT abundance among pathological stage in GSE130823 cohort. b Scatter plot showing pathological stage (x-axis) and NKT cell abundance (y-axis) in GSE130823 cohort. c Gene expression level of CXCR4 among pathological stage in GSE130823 cohort. d Scatter plot showing pathological stage (x-axis) and CXCR4 expression level (y-axis) in GSE130823 cohort. e Scatter plot showing CXCR4 expression level (x-axis) and NKT cell abundance (y-axis) in GSE130823 cohort, f GC meta-GEO cohort and g STAD TCGA-GTEx cohort. GEO, Gene Expression Database; STAD, stomach adenocarcinoma; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression

## Supplementary Files

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

- [FigureS3.pdf](#)
- [FigureS2.pdf](#)
- [FigureS1.pdf](#)