Stromal Score-based Gene Signature: A Prognostic Prediction Model for Colon Cancer

Jing Jia
Nanjing Medical University Second Clinical Medical College: Nanjing Medical University Second Affiliated Hospital

Yuhan Dai
Nanjing Medical University

Qing Zhang
Nanjing Medical University Second Clinical Medical College: Nanjing Medical University Second Affiliated Hospital

Peiyu Tang
Nanjing Medical University

Qiang Fu
Nanjing Medical University Second Affiliated Hospital

Guanying Xiong (xionggy_nj@126.com)
The First School of clinical medicine, Nanjing Medical University, Nanjing, China.
https://orcid.org/0000-0003-0780-8013

Primary research

Keywords: Colon cancer, Tumor microenvironment, Stromal, Prediction, Prognosis, Gene signature, Tumor biomarkers

Posted Date: December 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-125354/v1

Abstract

Background

Growing evidence has revealed the crucial roles of stromal cells in the microenvironment of various malignant tumors. However, efficient prognostic signatures based on stromal characteristics in colon cancer have not been well-established yet. The present study aimed to construct a stromal score-based multigene prognostic prediction model for colon cancer.

Method

Stromal scores were calculated based on the expression profiles of a colon cancer cohort from TCGA database applying the ESTIMATE algorithm. Linear models were used to identify differentially expressed genes between low-score and high-score groups by limma R package. Univariate and multivariate CoxPH regression analyses were used successively to select prognostic gene signature. An independent dataset from GEO was used as a validation cohort.

Results

Low stromal score was demonstrated to be a favorable factor to overall survival of colon cancer patients in TCGA cohort (log-rank test $p = 0.0046$). Three hundred and seven stromal score-related differentially expressed genes were identified. Through univariate and multivariate CoxPH regression analyses, a gene signature consisting of *LEP*, *SYT3*, *NOG* and *IGSF11* was recognized to build a prognostic prediction model. Based on the predictive values estimated by the established integrated model, patients were divided into two groups with significantly different overall survival outcomes (log-rank test $p < 0.0001$). Time-dependent Receiver operating characteristic curve analyses suggested the satisfactory predictive efficacy for 5-year overall survival of the model (AUC value = 0.736). A nomogram with great predictive performance combining the multigene prediction model and clinicopathological factors was developed. The established model was verified to be of significant prognostic value for different subgroups in an independent colon cancer cohort from GEO database, which was demonstrated to be especially accurate for young patients (AUC value = 0.752).

Conclusion

The well-established model based on stromal score-related gene signature might serve as a promising tool for the prognostic prediction of colon cancer.

Background

Colon cancer is one of the leading causes of cancer-related morbidity and mortality worldwide [1]. To date, the AJCC stage, determined according to the tumor, node, and metastasis (TNM) system, has been generally acknowledged as the most important tool for making clinical decisions and routine prognostication for colon cancer [2]. However, quite diverse prognostic outcomes have been reported

among colon cancer patients with the same TNM stages and similar clinical characteristics, mainly because of the high levels of heterogeneity found in colon cancer, which indicates that the current TNM stage system is not enough to provide prognostic information for colon cancer. Therefore, it is necessary to seek other efficient prognostic factors to improve prognosis stratification and survival outcome prediction in addition to the current staging system.

As a kind of solid malignant tumor, the tumor microenvironment (TME) of colon cancer is composed of immune cells and stromal cells besides tumor cells, which play vital roles in cancer initiation and development, as well as drug resistance [3, 4, 5]. During recent years, growing evidence has revealed the crucial roles of stromal cells in various malignant tumors [6, 7]. In colon cancer, it has been reported that tumor-associated stromal cells can support T-cell suppression by PD-L1 induction, revealing a key role of stromal cells in suppression of CD8 + antitumor immune responses of colon cancer [8].

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) is a newly developed method that infers stromal and immune cells based on gene expression profiles of cancer tissues [9]. Immune score, stromal score, estimate score and tumor purity were calculated through ESTIMATE algorithm to predict the level of infiltrating immune cells and stromal cells using the expression data of specific gene signature associated with immune and stromal components of TME. Up to date, numerous researchers have taken advantage of ESTIMATE algorithm in studies involving varieties of cancers, such as glioblastoma, gastric cancer, breast cancer and prostate cancer [10, 11, 12, 13], suggesting the effectiveness of such a big-data based algorithm. However, the role of ESTIMATE algorithm in colon cancer remains to be elucidated.

In this study, the ESTIMATE algorithm was applied to calculated the immune scores, stromal scores, estimate scores and tumor purity of a series of colon cancer tissues based on their expression profiles and survival analyses indicated that stromal score was prognostic for colon cancer. Then, a novel gene signature based on stromal score was developed subsequently for prognostic prediction in colon cancer.

Materials And Methods

Data source and application of ESTIMATE algorithm

The Cancer Genome Atlas (TCGA) level 3 gene expression RNA-seq data (standardized reads per kilobase per million mapped reads) of tissues from patients with colon cancer, along with corresponding clinicopathological information were downloaded from TCGA database (https://tcga-data.nci.nih.gov/tcga/) on Oct 15, 2020. The expression profiles for tumors with “Colon” as the primary site and the disease type of “Adenocarcinomas” from a “TCGA-COAD (Colon adenocarcinoma)” project were included. Besides expression data, only patients with significant clinicopathological information such as survival information, age, gender and pathological TNM stage were included in this study. Four ESTIMATE scores: immune score, stromal score, estimate score and tumor purity were calculated from the expression matrix applying the ESTIMATE algorithm for each patient, respectively. An independent

dataset GSE39582 from the Gene Expression Omnibus (GEO) database was used for external validation in this study. The gene expression array profiles and clinicopathological data of GSE39582 were downloaded by GEOquery R package and a total of 521 colon cancer patients with survival follow-up and essential clinicopathological information were included into the validation cohort. Next, the Probe IDs were transferred to gene symbol by hug133plus2.db R package. The probe with maximum mean was reversed when more than one probe had a same matched gene name. Access to the de-identified linked dataset was obtained from TCGA and GEO in accordance with the database policy. For analyses of de-identified data from the TCGA and GEO databases, institutional review board approval and informed consent were not required.

Correlation between prognosis and four ESTIMATE scores

Overall survival was used as the primary prognosis endpoint. Based on the four ESTIMATE scores calculated for each patient, best cut-off value for each score was determined by Survminer R package (V.0.4.6) and patients were divided into high-score and low-score groups according to corresponding best cut-off value respectively, and survival prognosis for each group was examined by Kaplan-Meier analysis. The survival outcomes of the two groups were compared by log-rank tests.

Identification of differentially expressed genes (DEGs)

Patients in training cohort were divided into two groups, namely the low stromal score group and the high stromal score group. Linear models were used to identify DEGs between the two groups (low-score group vs. high-score group) by limma R package [14]. A p-value < 0.0001 combined with a simultaneously absolute value of $\log_2$ (fold-change (FC)) > 1 was set as the threshold for DEG identification. Genes downregulated in the low stromal score group compared with the high stromal score group were considered as “downregulated DEGs” and those upregulated in the low stromal score group were considered as “upregulated DEGs.” The DEGs reached the threshold we set were presented on a volcano plot, and expression patterns of significant DEGs were visualized on a heatmap with unsupervised hierarchical clustering analysis.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses

Enrichment analyses of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for identified DEGs was performed using clusterProfiler R package. An FDR (false discovery rate) adjusted p-value < 0.05 was considered to be statistically significant for GO and KEGG pathway over-representation tests.

Definition of the stromal score-based gene signature and prognostic model Univariate and multivariate CoxPH (Cox proportional hazards) regression analyses were used to study the correlation between gene expression levels and overall survival (OS) of patients in training cohort. Firstly, we applied univariate CoxPH regression analysis to identify genes associated with OS. Then, for the prognostic genes recognized in univariate CoxPH regression analysis, multivariate CoxPH regression analysis was used to

select independent prognostic factors for OS of colon cancer patients. A multigene marker-based predictive value was calculated for each patient based on the expression level of the selected gene signature by predict.glm R function. Then, 279 patients in TCGA cohort were divided into high-risk and low-risk groups according to the best cut-off value of predictive value. Kaplan-Meier survival curves and time-dependent Receiver operating characteristic (ROC) curve analyses were operated to evaluate the predictive efficacy of the model.

Building and evaluation of the nomogram for OS prediction of colon cancer

The nomogram is an effective method to predict the prognosis of patients with malignant tumors, which simplifies the complicated statistical prediction model into a readable chart to evaluate the probability of OS for individual patients [15]. Taking advantage of rms R package, in this study, we included the selected gene signature through multivariate CoxPH regression analysis together with age, gender and pathological TNM stage to build a nomogram which could predict the probability of 5-year OS for colon cancer patients. The predicted probability of the nomogram was compared with the actual probability by the calibration curve to verify the accuracy of the nomogram. A predictive line overlapping with the actual line suggests an ideal model.

Validation of the gene signature in an independent cohort

To find out whether the gene signature identified from the TCGA cohort were of prognostic significance for other colon cancer cases as well, we used the dataset GSE39582 from the GEO database mentioned above as an independent validation cohort. Likely, Kaplan-Meier survival analyses were applied for every single prognostic gene and the integrated model on the validation cohort. To be more specific, patients in validation cohort were divided into subgroups depending on their clinicopathological characteristics, then Kaplan-Meier survival analyses and ROC analyses were operated on each subgroup to confirm the predictive capacity of the established model.

Statistical analyses

Survival curves were compared using the Kaplan-Meier method and the log-rank test. DEGs were compared with Student's t-test and those with p values less than 0.05 and fold-changes larger than two were viewed as dramatically dysregulated. Clinicopathological characteristics were compared by χ^2 tests or Wilcoxon tests for the TCGA and GSE39582 cohorts. All tests were two-sided, and a p value less than 0.05 was considered to be significant unless noted otherwise. All data were analyzed using R (Version 4.0.3).

Results

Study design and brief summary of patients' information

The study design is shown as a flowchart in Fig. 1. In this chart, we showed the detailed construction process of the OS prediction model for colon cancer patients. Patients' clinicopathological data in TCGA and GSE39582 cohorts was briefly summarized in Table 1.

Table 1
Clinical and pathological characteristics of COADpatients in TCGA cohort and GSE39582 cohort

Characteristics	TCGA cohorts (n = 279)	GSE39582 cohorts (n = 521)	P value
Age			
≤65	132(47.3%)	202(38.8%)	0.020
65	147(52.7%)	319(61.2%)	
Gender			
Male	154(55.2%)	284(54.5%)	0.853
Female	125(44.8%)	237(45.5%)	
TNM stage			
	46(16.5%)	32(6.1%)	0.110
	111(39.8%)	245(47.1%)	
	81(29.0%)	186(35.7%)	
	41(14.7%)	58(11.1%)	

Stromal scores are significantly associated with overall survival of colon cancer patients in TCGA cohort

We downloaded RNA-seq data of 279 primary colon cancer patients with survival information and significant clinicopathological data from TCGA database. Based on gene expression profiles, immune scores, stromal scores, estimate scores and tumor purity were calculated for each patient respectively using ESTIMATE algorithm. In order to find out the potential correlation between overall survival and the four ESTIMATE scores, we divided the 279 patients into high and low groups according to their four ESTIMATE scores respectively. Kaplan-Meier survival curves indicated that colon cancer patients with higher stromal scores showed poorer overall survival than lower ones (Fig. 2b, log-rank test $p = 0.0046$). However, Kaplan-Meier survival analyses didn't show significant differences in OS between groups with different levels of immune scores, estimate scores or tumor purity (Fig. 2a, c and d).

Comparison of gene expression profiles by stromal scores in colon cancer

In order to identify the key genes contributing to the opposing survival outcomes related to stromal scores, the TCGA expression profiles of colon cancer patients with lower stromal scores were compared to those of ones with higher stromal scores. A total of 307 DEGs were identified as stromal score-related DEGs. Interestingly, among them, 306 DEGs were downregulated in patients with lower stromal scores ($\log_2$FC < -1, p < 0.0001) while only 1 DEG was upregulated ($\log_2$FC > 1, p < 0.0001). The expression profiles of stromal score-related DEGs are visualized on the heatmap (Fig. 3a) and volcano plot (Fig. 3b). Apparently, unsupervised hierarchical clustering analysis showed identified DEGs could clearly distinguish patients with high and low stromal scores, and DEGs in patients with low stromal scores were mostly downregulated, which suggested that further digging might find out crucial genes responsible for poor outcomes of colon cancer patients with high stromal scores. GO and KEGG enrichment analyses found that the DEGs mainly took part in neurosynapse assembly, biosynthesis of synapse specific membrane and function of ion channel on synaptic membrane (Fig. 3c and d).

Identification of prognostic gene signature in colon cancer

Still, we used the TCGA cohort as a training dataset and the 307 DEGs above were subjected to univariate CoxPH regression analysis to identify markers that closely associated with OS. As a result, 64 genes with a p value less than 0.05 were selected as candidates (Additional file 1: Fig. S1). Then, all of the 64 genes together with gender, age and pathological TNM stage were included in the following multivariate CoxPH regression analysis. A total of four genes were recognized as independent prognostic factors (p < 0.05), including *LEP* (HR = 2.16, 95% CI 1.39 to 3.35), *SYT3* (HR = 1.72, 95% CI 1.09 to 2.70), *NOG* (HR = 1.77, 95% CI 1.09 to 2.87) and *IGSF11* (HR = 1.72, 95% CI 1.04 to 2.85) (Fig. 4a). Moreover, age and TNM stage were also showed to be independent prognostic factors (Not shown).

Furthermore, the TCGA cohort was subdivided into high-expression and low-expression subgroups according to the best expression cut-off levels of the four genes (best cut-off values were calculated by Survminer R package (V.0.4.6)). Kaplan-Meier curves revealed that high expression levels of all of the four genes were associated with inferior overall survival (log-rank test p < 0.0001 for *LEP*, p = 0.0095 for *SYT3*, p < 0.0001 for *NOG* and p = 0.023 for *IGSF11*) (Fig. 4b-e). Based on such results, we next established a multiplex prediction model encompassing the transcript expression levels of the four genes. As shown in the Kaplan-Meier curves for prediction model, patients of high risk had much worse OS rates than those of low risk (Fig. 4f). ROC curves showed that the four-gene expression integrated prediction model had an area under the curve (AUC) value of 0.736 in evaluating 5-year OS, which was much higher than any individual factor alone (Fig. 4g), demonstrating a satisfactory predictive efficacy of the model.

Building of a nomogram to predict OS in colon cancer patients

In order to establish a clinically applicable method to predict overall survival of patients with primary colon cancer, a nomogram for 5-year survival prediction was further built by integrating the stromal score-associated four-gene signature, age, gender and TNM stage in TCGA cohort (Fig. 5a). Further calibration

plot illustrated that the nomogram performed well-compared against the performance of an ideal model (Fig. 5b), which confirmed the great predictive accuracy of the newly constructed nomogram.

Validation of stromal score-based gene signature in GSE39582 cohort

The stromal score-based gene signature and the four-gene integrated predictive model were further validated in an independent dataset from the GEO database (Series GSE39582). Consistent with our results got from TCGA cohort, among the 521 included patients diagnosed with primary colon cancer from GSE39582, each of the four genes showed to be a significant unfavorable factor to overall survival by Kaplan-Meier survival analysis (log-rank test $p = 0.0047$ for *LEP*, $p = 0.00048$ for *SYT3*, $p < 0.0001$ for *NOG* and $p = 0.0038$ for *IGSF11*) (Additional file 2: Fig. S2a-d). Similarly, the multiplex model encompassing the four signature genes was built and patients with high risk had significantly worse OS rates than those with low risk (log-rank test $p < 0.0001$) (Fig. 6g). ROC analysis showed that the AUC value of the four-gene signature model was 0.575 in evaluating 5-year OS for patients from GSE39582 cohort (Fig. 6h). Considering the heterogeneity of the two cohorts, we further explored the predictive ability of the four-gene signature in different subgroups divided by age, TNM stage and gender. As shown in Kaplan-Meier curves for different subgroups, the four-gene signature model could well distinguish opposing prognostic outcomes in all of the subgroups (log-rank test $p < 0.0001$ for patients $\leq$ 65Y, $p = 0.00071$ for patients > 65Y, $p = 0.01$ for patients of stage 1 ~ 2, $p < 0.0001$ for patients of stage 3 ~ 4, $p = 0.0057$ for male patients and $p = 0.0019$ for female patients), which suggested huger differences of OS in patients $\leq$ 65Y, patients of stage 3 ~ 4 and female patients (Fig. 6a-f). Also, ROC analyses were operated on the six subgroups later. To our surprise, the AUC value of the four-gene signature model in evaluating 5-year OS for patients younger than 65 showed to be remarkable 0.752, whereas the AUC value was 0.595 for elder patients. Besides, the AUC values were 0.577, 0.661, 0.564 and 0.624 for patients of stage 1 ~ 2, patients of stage 3 ~ 4, male patients and female patients, respectively. In this part of study, we validated the great value of our established four-gene signature model in predicting OS for colon cancer patients, which might have especially promising prognostic value for young patients.

Discussion

Tumor microenvironment is a crucial concept in tumor immunology, which includes immune cells, stromal cells, epithelial cells, fibroblasts, vascular cells and signaling molecules that closely interact with the development and metastasis of tumors [16, 17, 18, 19]. As the most important components of TME, immune cells and stromal cell have been deeply concerned by scientists. Currently, several gene signatures based on characteristics of immune and stromal components in TME have been reported [20, 21, 22, 11]. For colon cancer, Galon et al. built an immunoscore system based on the amounts of infiltrated CD3+, CD8+, or CD45RO + lymphocytes in the central- and peri- tumoral areas, finding the prognostic ability of the immunoscore system stronger than TNM stage [23]. Nonetheless, the latest multi-central clinical research showed deficiency in predictive accuracy of the system [24], probably due

to the neglection of stromal components in TME. Quite recently, Cai et al. reported *SNAP25* as a prognostic gene based on stromal-immune score [25]. However, no concrete model or algorithm was built to predict survival outcomes in that study, which might largely reduce the practicability and reliability of the results. In order to make up for the lack in this aspect for colon cancer, we devised the current study.

In this study, through a specific view of the tumor microenvironment, we explored on the basis of immune and stromal scores applying well-established ESTIMATE algorithm. Different from the results of some studies, we found the stromal score to be the only one significantly related to survival outcomes of colon cancer patients among the four output values of the algorithm. For this reason, we next mined for prognostic genes based on the stromal score but the immune score. Through a series of successive, organized and targeted analyses for transcriptomic data and survival information, this study identified a set of stromal score-related prognostic DEGs and built a stromal score-based multigene prognostic prediction model for colon cancer, which was demonstrated to be highly efficient by ROC curves. Additionally, it was interesting to be suggested in our validation section that, to some extent, the established prediction model might be much more accurate for young patients. Considering the increasing mortality among young patients with colon cancer [26], and that prognosis of colon cancer among young patients is not well known so that it is difficult to advise about adjuvant chemotherapy [27, 28], our results might be of additional value for prognostic prediction and treatment decision among young patients.

All of the four signature genes identified in this study have been reported to play vital roles in tumorigenesis, development and metastasis of varieties of malignant tumors, including colon cancer. *LEP*, which encodes leptin, is well-known because of its significant role in obesity. Besides energy homeostasis, recent studies have shown its extended properties involving numerous aspects including the high risk of colon cancer [29, 30]. Potter et al. suggested that different leptin and leptin receptor genotypes might influence the risk of colon cancer [31]. Although the synaptotagmins (SYTs) have not been well studied in malignant tumors yet, their roles in vesicle trafficking and fusion of vesicles have been pointed out [32], which might be associated with cell migration. Particularly, Roos et al. reported that SYT3 was essential for migration of T cells [33], which might indirectly take part in the tumor microenvironment. As an antagonist of bone morphogenetic proteins (BMP), noggin encoded by gene *NOG* plays a role in both normal development and cancers. It was lately reported that noggin was responsible for poor prognosis of gastric cancer by promoting the proliferation of tumor cells via upregulating EGFR (epidermal growth factor receptor) [34]. In colon, Hardwick et al. found noggin inhibited apoptosis and proliferation in mouse colonic epithelium *in vivo* [35]. The dysregulation of *IGSF11* was reported in colorectal cancer, hepatocellular carcinoma and intestinal-type gastric cancer, which might serve as a target for cancer immunotherapy of gastrointestinal and hepatocellular carcinoma [36]. Mostly consistent with the results of the previous studies, high expression levels of the four signature genes we identified all represents poor survival outcomes of patients with colon cancer.

Amounts of investigations have been done to identify key targets that play crucial roles in tumorigenesis and cancer development, some of them even explored the detailed inner mechanisms. However, it is

necessary to transfer the preclinical medical findings into clinical applications. A recent study utilizing the profiles of clinical samples and TCGA cases established a prognostic prediction model of high quality for head and neck squamous cell carcinoma by encompassing both immune-related gene signature and clinicopathological factor, providing a powerful tool for the prognostic prediction in clinical practice in addition to TNM staging [37]. Meanwhile, with the rapid development in techniques of next-generation sequencing, it may become more and more convenient and popularized to help evaluate prognostic outcomes and make treatment decisions for cancer patients based on transcriptomic profiles of target genes in their tumor tissues. Therefore, if correctly applied in clinical practice, the multigene prediction model constructed in our present study might have potential value for clinical management of colon cancer.

The lack of clinical specimens limited our study to some degree. Further efforts should be done to validate the model in an independent cohort of clinical samples, on both gene and protein levels. In addition, for deeper explorations, the mechanisms of the signature genes regulating tumorigenesis remain to be further investigated both *in vivo* and *in vitro*.

Conclusions

In conclusion, we demonstrated the correlation between stromal characteristics of tumor microenvironment and survival outcomes of colon cancer patients. The novel prognostic prediction model we established based on stromal score-related gene signature might be of value to stratify patients and make clinical decisions for colon cancer.

Abbreviations

AUC
Area under the curve
BMP
Bone morphogenetic proteins
BP
biological process
CC
cellular component
COAD
Colon adenocarcinoma
CoxPH
Cox proportional hazards
DEGs
Differentially expressed genes
EGFR
Epidermal growth factor receptor

ESTIMATE
Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data
FC
Fold-change
FDR
False discovery rate
GEO
The Gene Expression Omnibus
GO
Gene Ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
MF
molecular function
OS
Overall survival
ROC
Receiver operating characteristic
SYTs
synaptotagmins
TCGA
The Cancer Genome Atlas
TME
Tumor microenvironment

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Publicly available datasets were used in this study. TCGA data can be found here: https://tcga-data.nci.nih.gov/tcga/; GEO data were downloaded by GEOquery R package.

Competing interests

The authors have declared that they have no competing interests.

Funding

Not applicable.

Authors' contributions

JJ, YD and GX conceived and designed the study. JJ and YD performed the analyses. JJ, YD, QZ, PT and QF wrote the paper. GX reviewed the data and manuscript. All authors read and approved the final manuscript. Jing Jia and Yuhan Dai contributed equally to this work.

Acknowledgements

Thanks to Senlin Lian for her help in R learning.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a Cancer Journal for Clinicians. 2020; 70:7-30.
2. Hu H, Krasinskas A, Willis J. Perspectives on current tumor-node-metastasis (TNM) staging of cancers of the colon and rectum. Semin Oncol. 2011; 38: 500-510.
3. Hui L, Chen Y. Tumor microenvironment: sanctuary of the devil. Cancer Lett. 2015; 368:7-13.
4. Kobayashi H, Enomoto A, Woods SL, Burt AD, Takahashi M, Worthley DL. Cancer-associated fibroblasts in gastrointestinal cancer. Nat Rev Gastroenterol Hepatol. 2019; 16:282–95.
5. Straussman R, Morikawa T, Shee K, Barzily-Rokni M, Qian ZR, Du J, et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. Nature. 2012; 487:500–4.
6. Xu M, Zhou BP, Tao M, Liu J, Li W. The role of stromal components in pancreatic cancer progression. Anticancer Agents Med Chem. 2016;16(9):1117-1124.
7. Zhang Y, Nojima S, Nakayama H, Jin Y, Enza H. Characteristics of normal stromal components and their correlation with cancer occurrence in human prostate. Oncol Rep. 2003;10(1):207-211.
8. O'Malley G, Treacy O, Lynch K, Naicker SD, Leonard NA, Lohan P, Dunne PD, Ritter T, Egan LJ, Ryan AE. Stromal Cell PD-L1 Inhibits CD8+ T-cell Antitumor Immune Responses and Promotes Colon Cancer. Cancer Immunol Res. 2018; 6(11):1426-1441.
9. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
10. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging. 2018; 10:592-605.
11. Wang H, Wu X, Chen Y. Stromal-Immune Score-Based Gene Signature: A Prognosis Stratification Tool in Gastric Cancer. Front Oncol. 2019; 9:1212.

12. Priedigkeit N, Watters RJ, Lucas PC, et al. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. JCI Insight. 2017;2(17).
13. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, Rockowitz S, Drier Y, Bernstein BE, Long HW, Freedman ML, Arora VK, Zheng D, Sawyers CL. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. Elife. 2017; 6:e27861.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43(7):e47.
15. Park, S. Y. Nomogram: An analogue tool to deliver digital knowledge. J. Thorac. Cardiovasc. Surg. 2018; 155(4):1793.
16. Vitale I, Manic G, Coussens LM, Kroemer G, Galluzzi L. Macrophages and metabolism in the tumor microenvironment. Cell Metab. 2019; 30(1):36-50.
17. Meurette O, Mehlen P. Notch signaling in the tumor microenvironment. Cancer Cell. 2018; 34(4):536-548.
18. Ren B, Cui M, Yang G, Wang H, Feng M, You L, Zhao Y. Tumor microenvironment participates in metastasis of pancreatic cancer. Mol Cancer. 2018; 17(1):108.
19. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast Cancer Res. 2016; 18(1):84.
20. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017; 18(1):248-262.
21. Chifman J, Pullikuth A, Chou JW, Bedognetti D, Miller LD. Conservation of immune gene signatures in solid tumors and prognostic implications. BMC Cancer. 2016; 16(1):911.
22. Hao D, Liu J, Chen M, Li J, Wang L, Li X, Zhao Q, Di LJ. Immunogenomic Analyses of advanced serous ovarian cancer reveal immune score is a strong prognostic factor and an indicator of chemosensitivity. Clin Cancer Res. 2018; 24(15):3560-3571.
23. Pages F, Kirilovsky A, Mlecnik B, Asslaber M, Tosolini M, Bindea G, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. J Clin Oncol. 2009; 27(35): 5944-5951.
24. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018; 391(10135):2128-2139.
25. Zou J, Duan D, Yu C, Pan J, Xia J, Yang Z, Cai S. Mining the potential prognostic value of synaptosomal-associated protein 25 (SNAP25) in colon cancer based on stromal-immune score. PeerJ. 2020; 8:e10142.
26. Schoen RE, Pinsky PF, Weissfeld JL, et al; PLCO Project Team. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. N Engl J Med. 2012; 366(25):2345-2357.

27. Kneuertz PJ, Chang GJ, Hu CY, et al. Overtreatment of young adults with colon cancer: more intense treatments with unmatched survival gains. JAMA Surg. 2015;150(5):402-409.
28. Manjelievskaia J, Brown D, McGlynn KA, Anderson W, Shriver CD, Zhu K. Chemotherapy use and survival among young and middle-aged patients with colon cancer. JAMA Surg. 2017; 152(5):452-459.
29. Tamakoshi K, Toyoshima H, Wakai K, Kojima M, Suzuki K, Watanabe Y, Hayakawa N, Yatsuya H, Kondo T, Tokudome S, Hashimoto S, Suzuki S, et al. Leptin is associated with an increased female colorectal cancer risk: a nested case-control study in Japan. Oncology. 2005; 68(4-6): 454–61.
30. Stattin P, Lukanova A, Biessy C, Soderberg S, Palmqvist R, Kaaks R, Olsson T, Jellum E. Obesity and colon cancer: does leptin provide a link? Int J Cancer. 2004; 109(1): 149–52.
31. Slattery ML, Wolff RK, Herrick J, Caan BJ, Potter JD. Leptin and leptin receptor genotypes and colon cancer: gene-gene and gene-lifestyle interactions. Int J Cancer. 2008 ;122(7):1611-7.
32. Chapman ER. Synaptotagmin: a Ca(2+) sensor that triggers exocytosis? Nat Rev Mol Cell Biol. 2002; 3(7):498-508.
33. Masztalerz A, Zeelenberg IS, Wijnands YM, de Bruijn R, Drager AM, Janssen H, Roos E. Synaptotagmin 3 deficiency in T cells impairs recycling of the chemokine receptor CXCR4 and thereby inhibits CXCL12 chemokine-induced migration. J Cell Sci. 2007 ;120(Pt 2):219-28.
34. Sun Z, Gao X, Zabkiewicz C, Ruge F, Xie M, Cai S, Sun PH, Griffiths P, Pugh MR, Ji J, Ji J, et al. Noggin is associated with a poor prognosis of gastric cancer by promoting the proliferation of gastric cancer cells via the upregulation of EGFR. Int J Oncol. 2020; 57: 813-824.
35. Hardwick JC, Van Den Brink GR, Bleuming SA, Ballester I, Van Den Brande JM, Keller JJ, Offerhaus GJ, Van Deventer SJ, Peppelenbosch MP. Bone morphogenetic protein 2 is expressed by, and acts upon, mature epithelial cells in the colon. Gastroenterology. 2004; 126(1):111-21.
36. Watanabe T, Suda T, Tsunoda T, Uchida N, Ura K, Kato T, Hasegawa S, Satoh S, Ohgi S, Tahara H, Furukawa Y, Nakamura Y. Identification of immunoglobulin superfamily 11 (IGSF11) as a novel target for cancer immunotherapy of gastrointestinal and hepatocellular carcinomas. Cancer Sci. 2005; 96(8):498-506.
37. Yao Y, Yan Z, Lian S, et al. Prognostic value of novel immune-related genomic biomarkers identified in head and neck squamous cell carcinoma. Journal for ImmunoTherapy of Cancer. 2020;8:e000444.

Figures

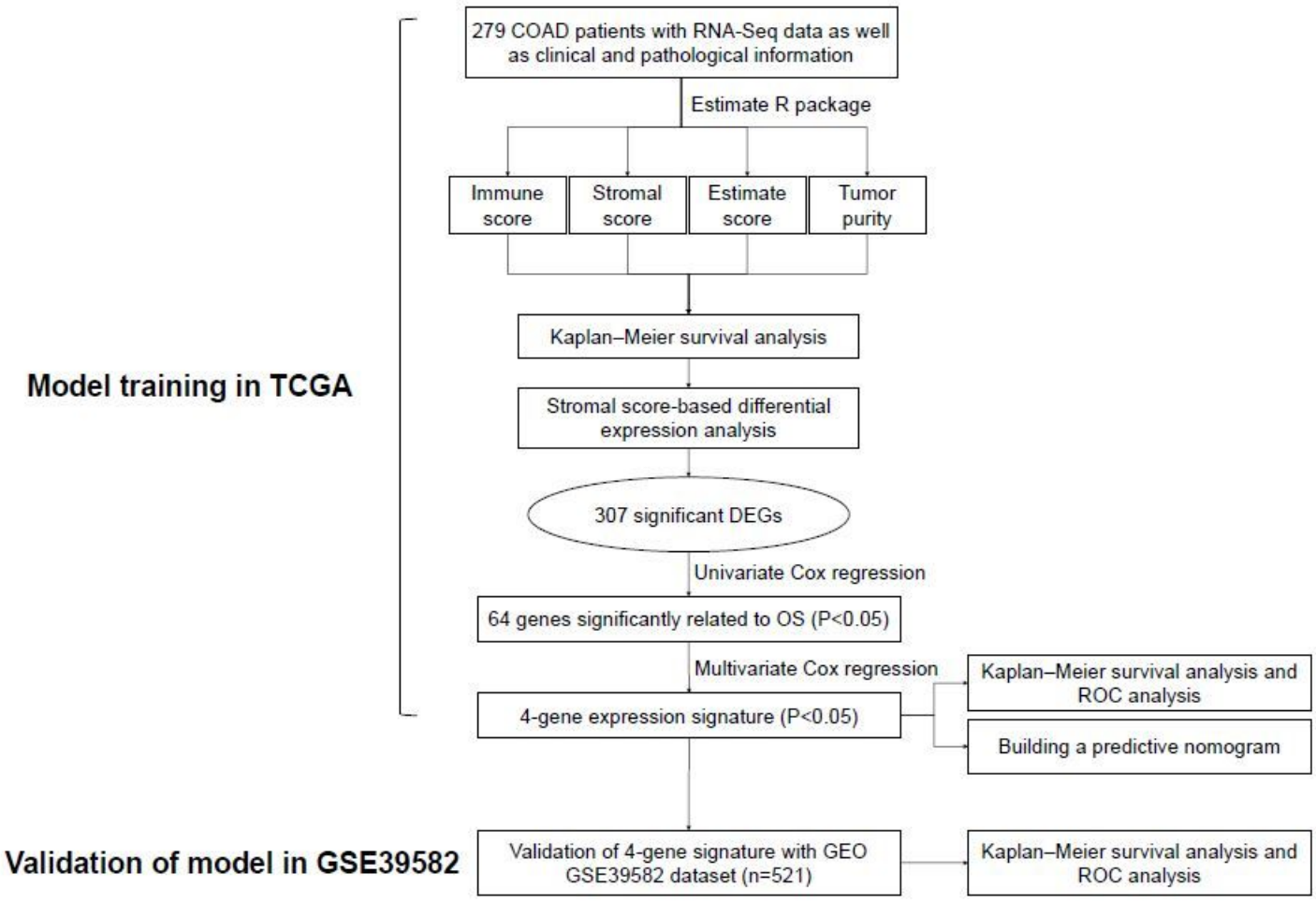

Figure 1

Overall flowchart of this study.

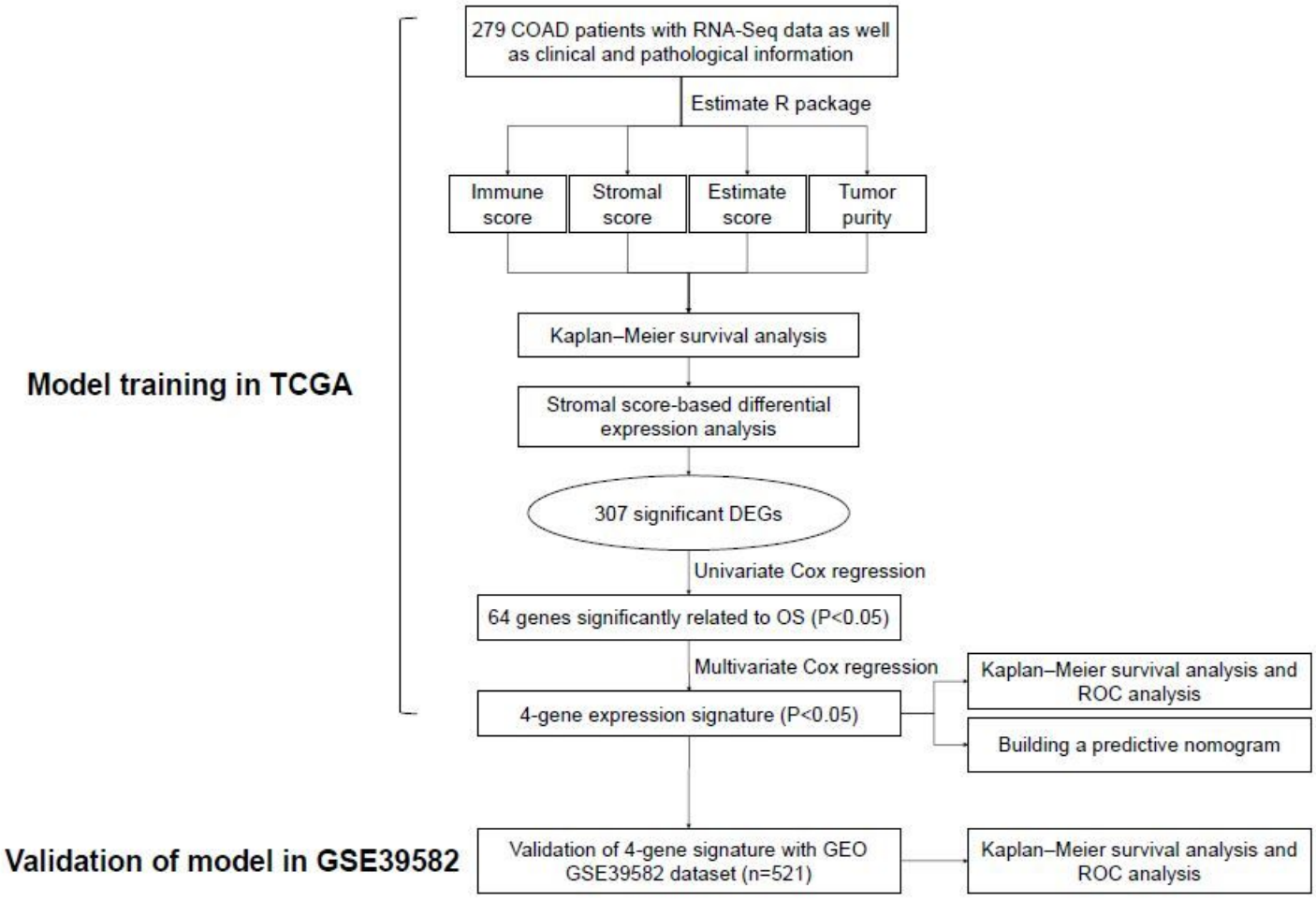

Figure 1

Overall flowchart of this study.

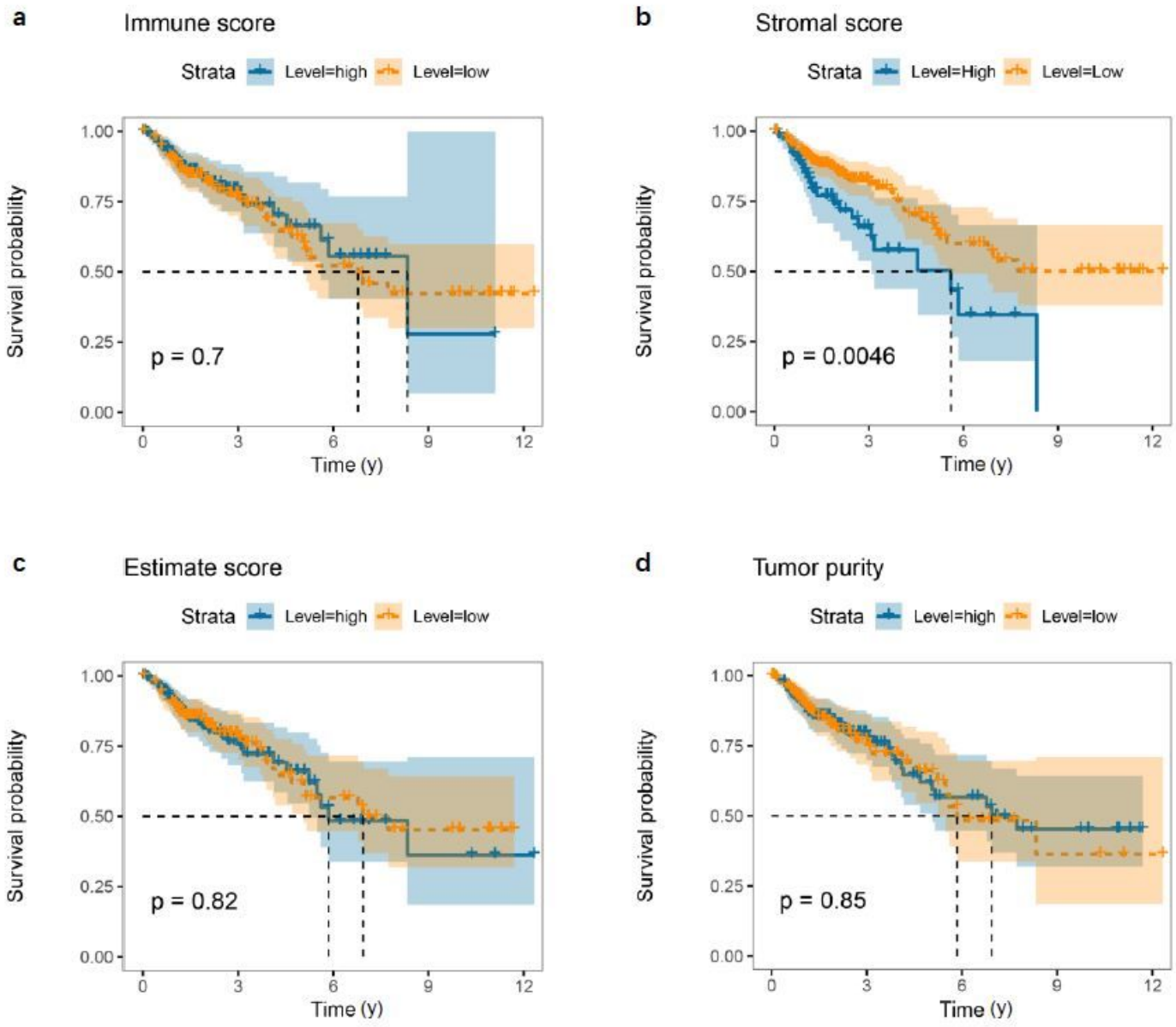

Figure 2

Stromal scores are significantly associated with overall survival of colon cancer patients in TCGA cohort. a Kaplan-Meier curves of overall survival for patients with high vs. low immune scores. b Kaplan-Meier curves of overall survival for patients with high vs. low stromal scores. c Kaplan-Meier curves of overall survival for patients with high vs. low estimate scores. d Kaplan-Meier curves of overall survival for patients with high vs. low tumor purity.

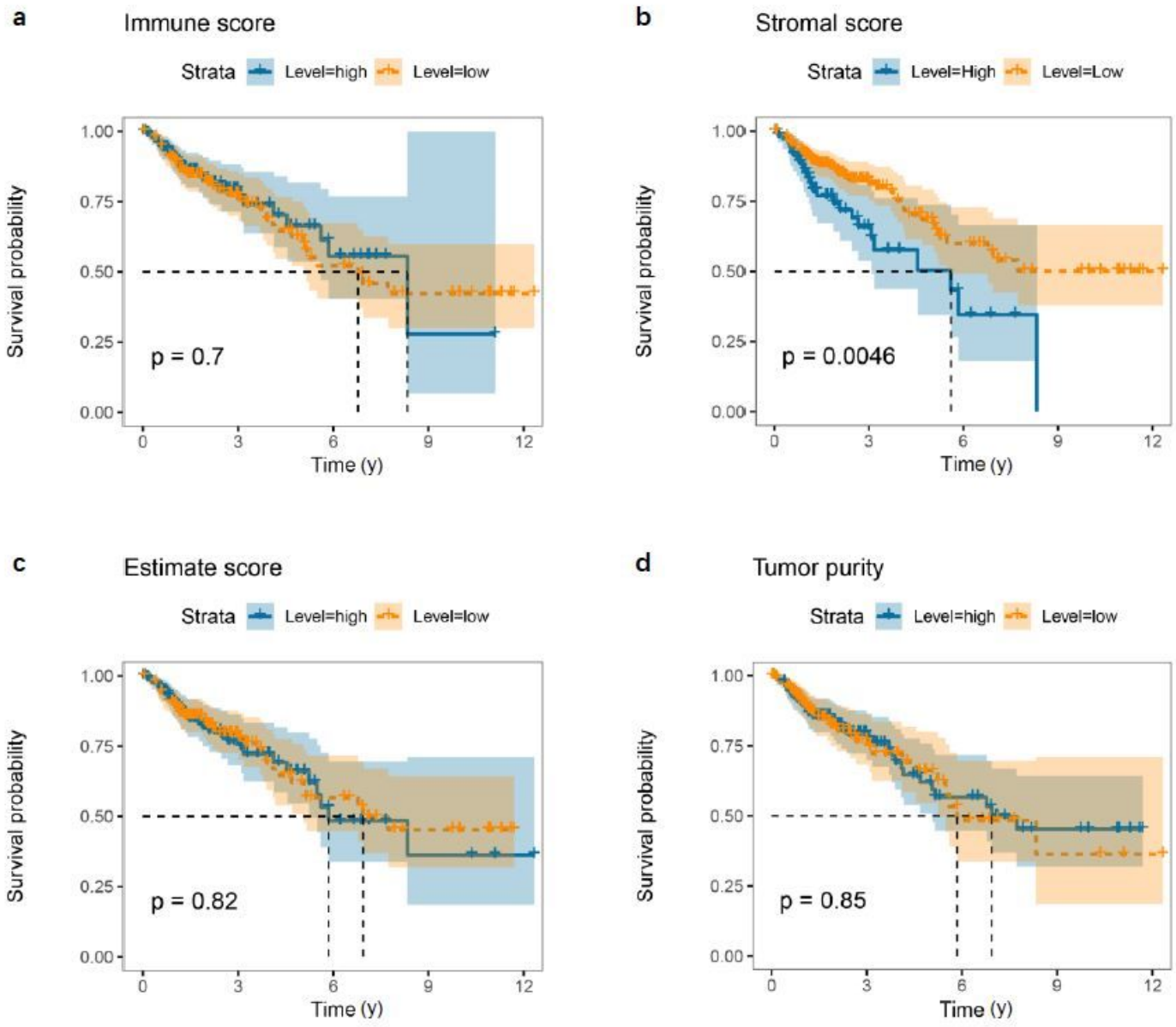

Figure 2

Stromal scores are significantly associated with overall survival of colon cancer patients in TCGA cohort. a Kaplan-Meier curves of overall survival for patients with high vs. low immune scores. b Kaplan-Meier curves of overall survival for patients with high vs. low stromal scores. c Kaplan-Meier curves of overall survival for patients with high vs. low estimate scores. d Kaplan-Meier curves of overall survival for patients with high vs. low tumor purity.

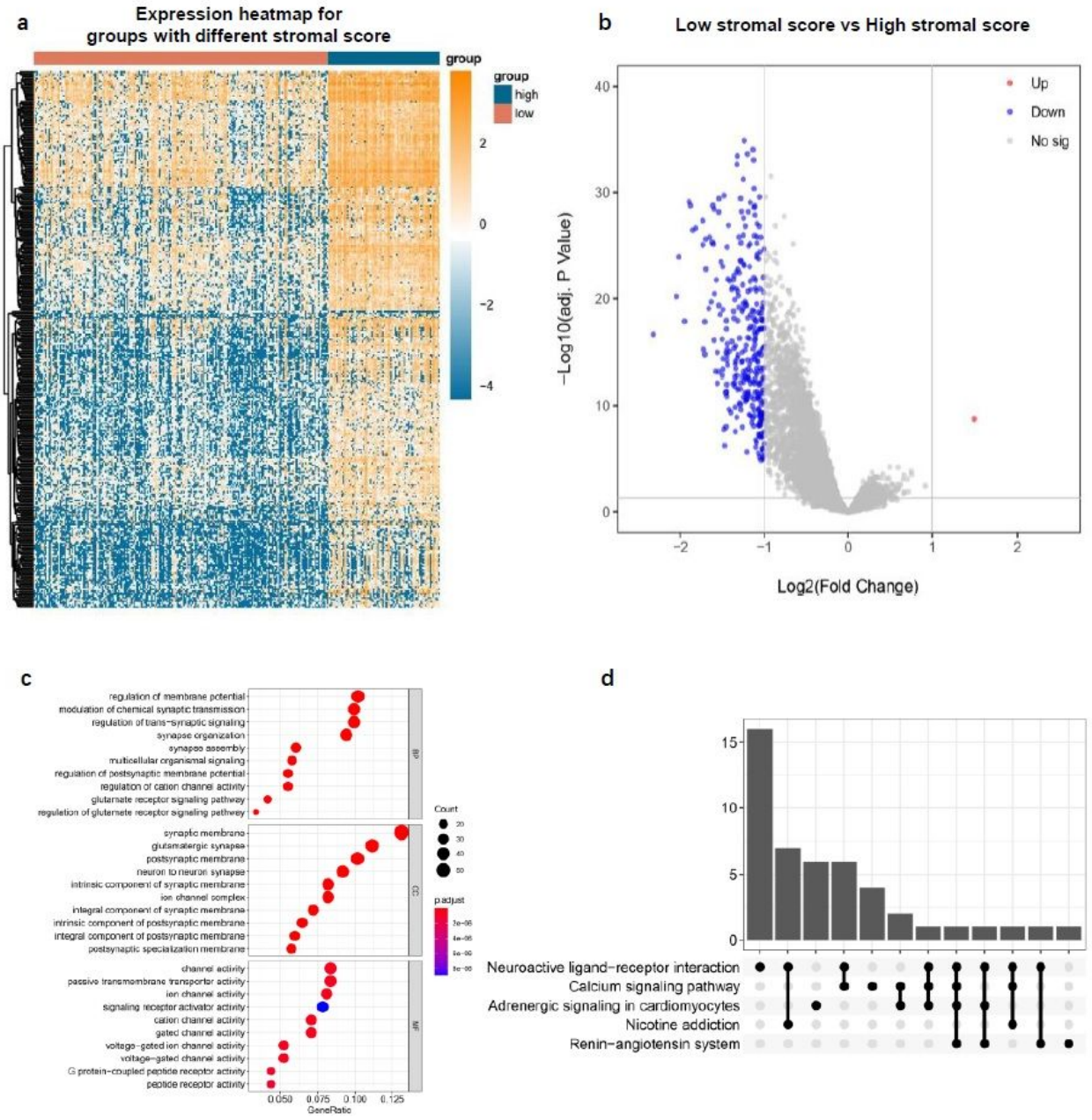

Figure 3

Comparison of gene expression profiles by stromal scores in colon cancer. a The heatmap showing the expression patterns of stromal score-related DEGs with unsupervised hierarchical clustering analysis. b The volcano plot visualizing the expression profiles of stromal score-related DEGs. c,d GO and KEGG enrichment analyses revealed the most significant biological process (BP), cellular component (CC), molecular function (MF) and pathways correlated to DEGs.

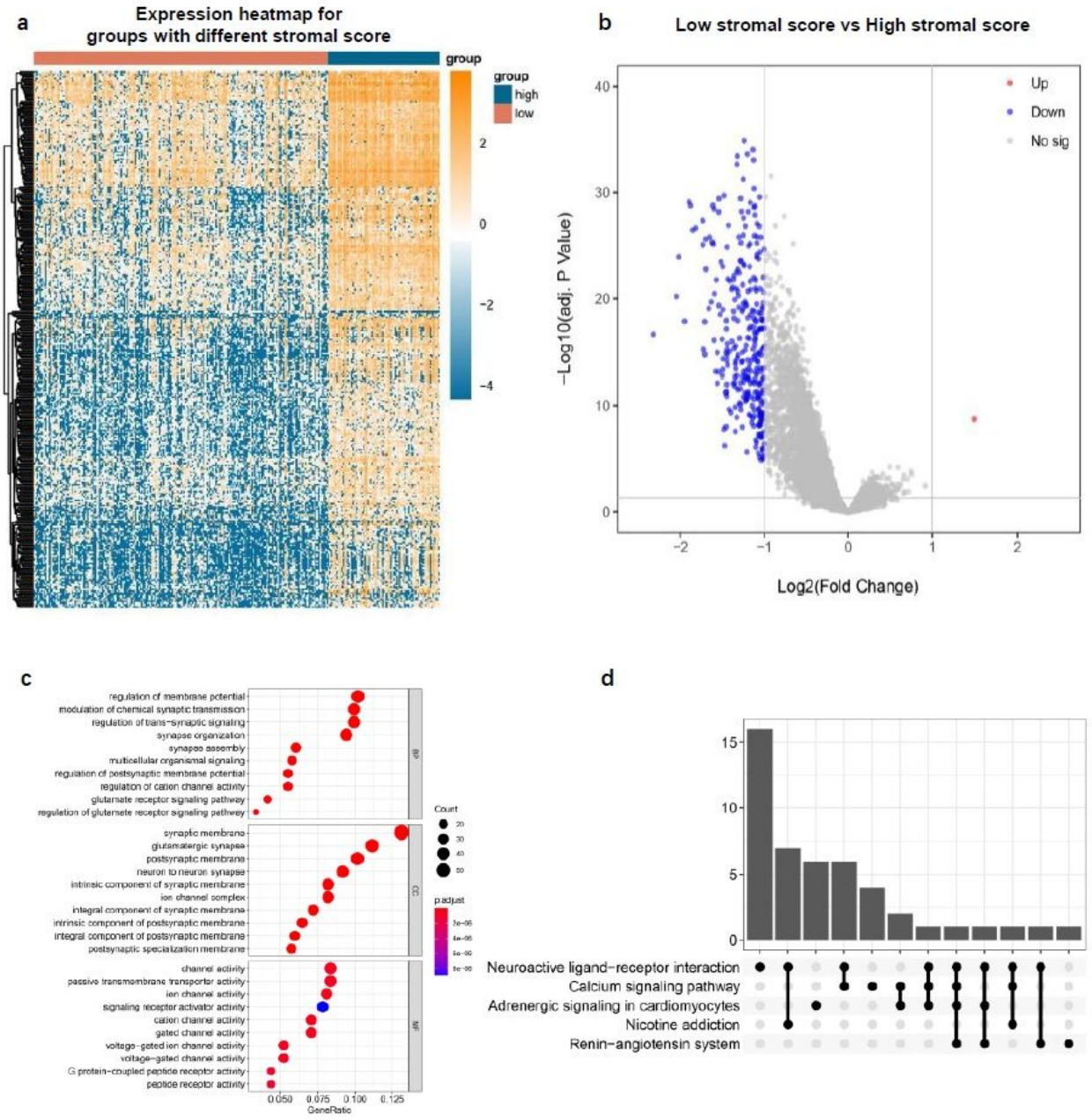

Figure 3

Comparison of gene expression profiles by stromal scores in colon cancer. a The heatmap showing the expression patterns of stromal score-related DEGs with unsupervised hierarchical clustering analysis. b The volcano plot visualizing the expression profiles of stromal score-related DEGs. c,d GO and KEGG enrichment analyses revealed the most significant biological process (BP), cellular component (CC), molecular function (MF) and pathways correlated to DEGs.

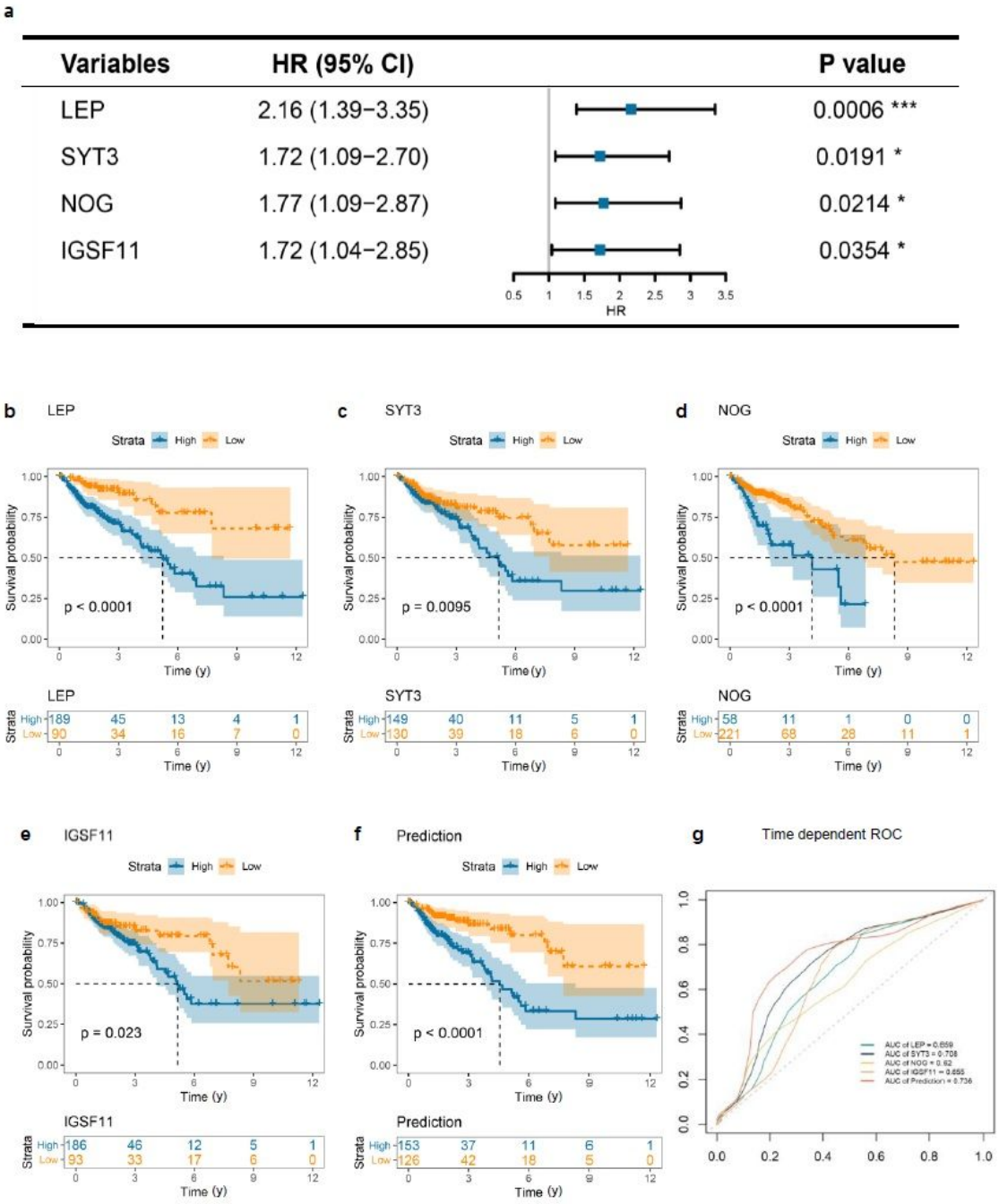

Figure 4

Identification of prognostic gene signature in colon cancer. a The forest plot of hazard ratios for prognostic gene signature selection applying the multivariate CoxPH regression. b-e Kaplan-Meier curves of overall survival for patients grouped by expression levels of four signature genes: LEP, SYT3, NOG and IGSF11. f Kaplan-Meier curves of overall survival for the integrated prediction model encompassing the four genes. g ROC curves for each parameter, with AUC values.

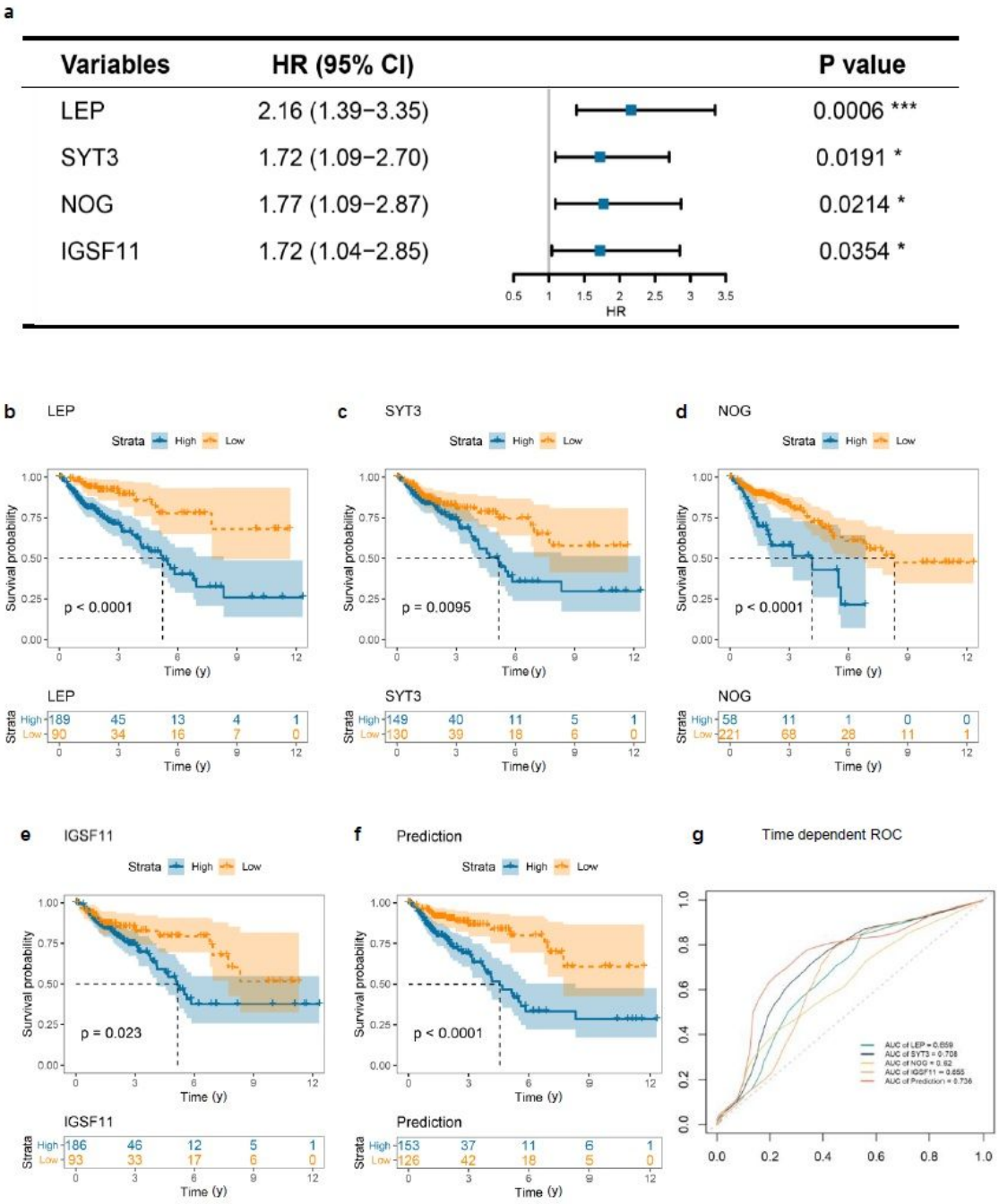

Figure 4

Identification of prognostic gene signature in colon cancer. a The forest plot of hazard ratios for prognostic gene signature selection applying the multivariate CoxPH regression. b-e Kaplan-Meier curves of overall survival for patients grouped by expression levels of four signature genes: LEP, SYT3, NOG and IGSF11. f Kaplan-Meier curves of overall survival for the integrated prediction model encompassing the four genes. g ROC curves for each parameter, with AUC values.

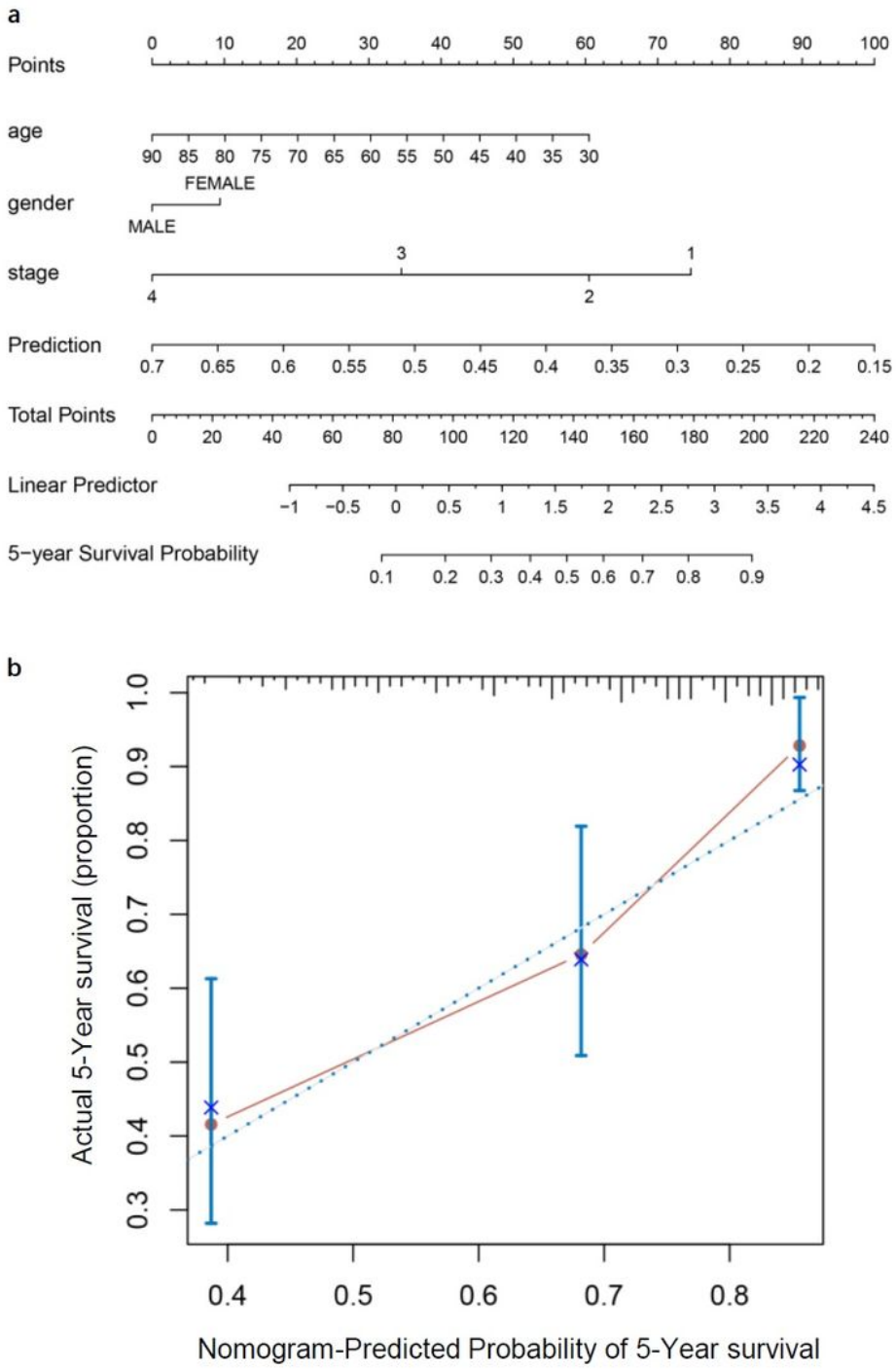

Figure 5

Building of a nomogram to predict OS in colon cancer patients. a The nomogram to predict 5-year overall survival probability by integrating the four-gene prediction model with age, gender and TNM stage. b The plot depicting the calibration of the nomogram in terms of the consistency between predicted and actual outcomes. Nomogram performance is shown as the plot relative to the dotted line, which represents an ideal model.

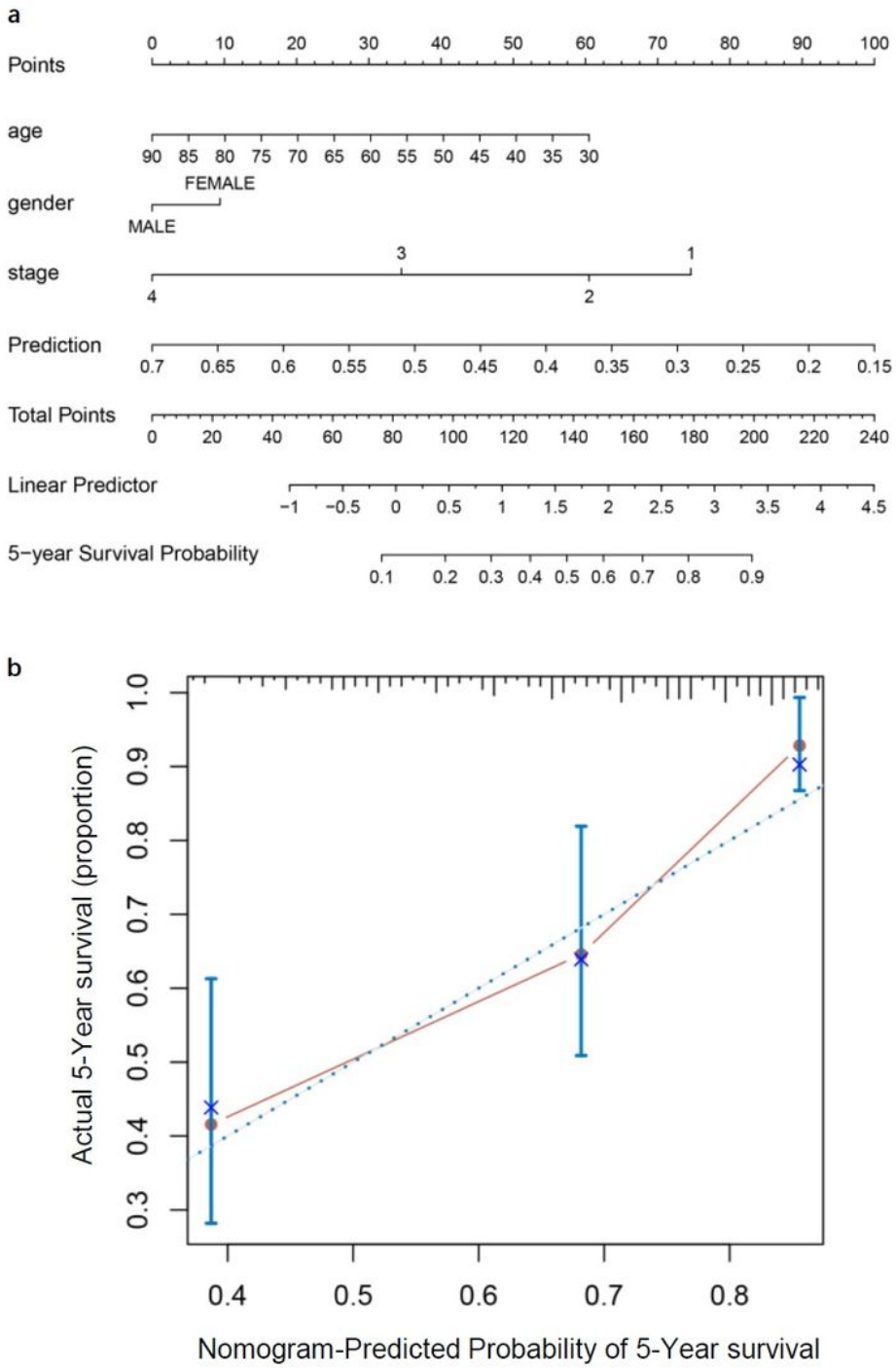

Figure 5

Building of a nomogram to predict OS in colon cancer patients. a The nomogram to predict 5-year overall survival probability by integrating the four-gene prediction model with age, gender and TNM stage. b The plot depicting the calibration of the nomogram in terms of the consistency between predicted and actual outcomes. Nomogram performance is shown as the plot relative to the dotted line, which represents an ideal model.

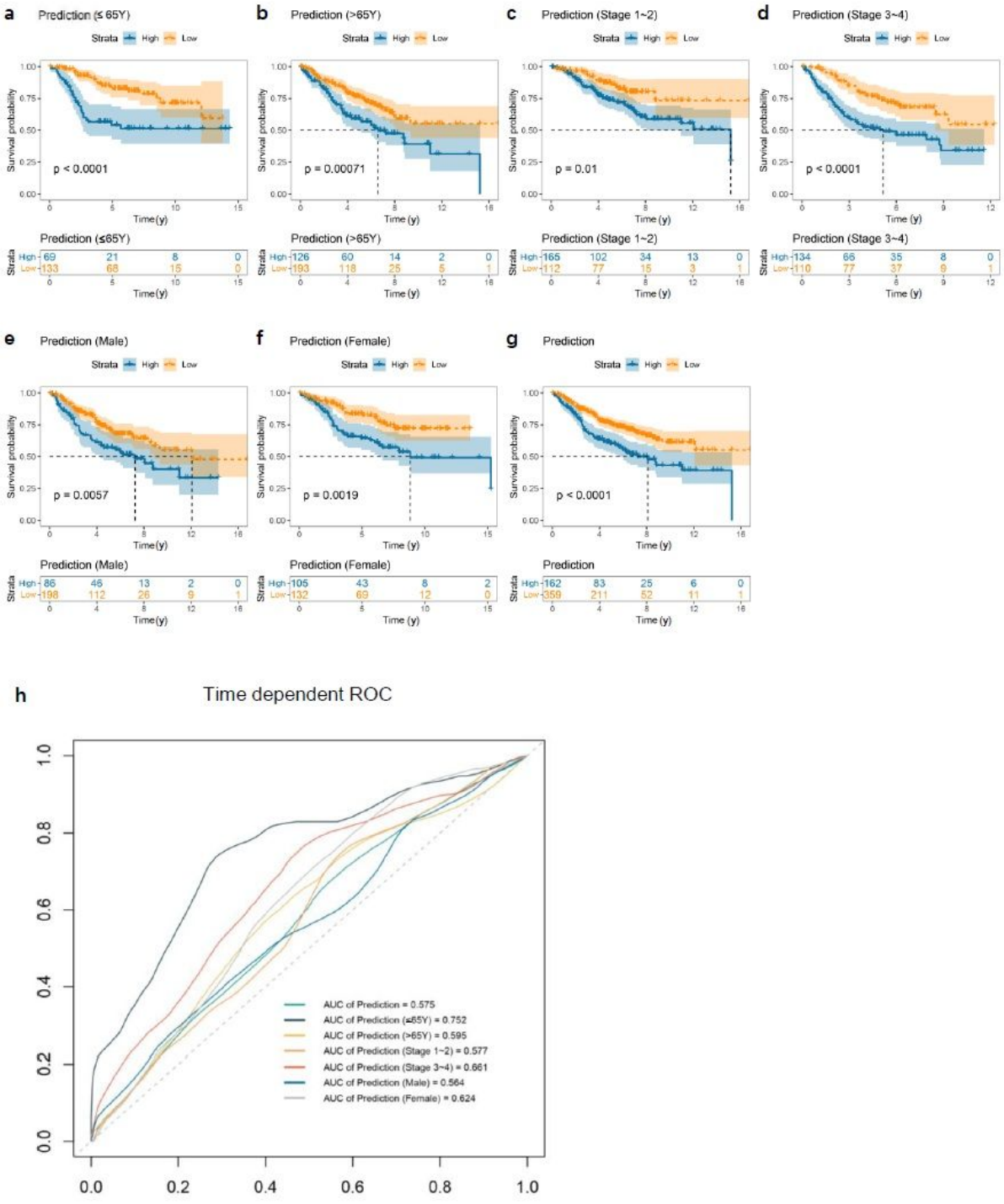

Figure 6

Validation of stromal score-based gene signature in GSE39582 cohort. Subgroup analysis of OS based on a,b age, c,d TNM stage, and e,f gender of colon cancer patients in GSE39582 cohort. g Kaplan-Meier curves of OS for the integrated prediction model encompassing the four genes in GSE39582 cohort. h ROC curves of the multigene prediction model for the whole GSE39582 cohort and different subgroups, with AUC values.

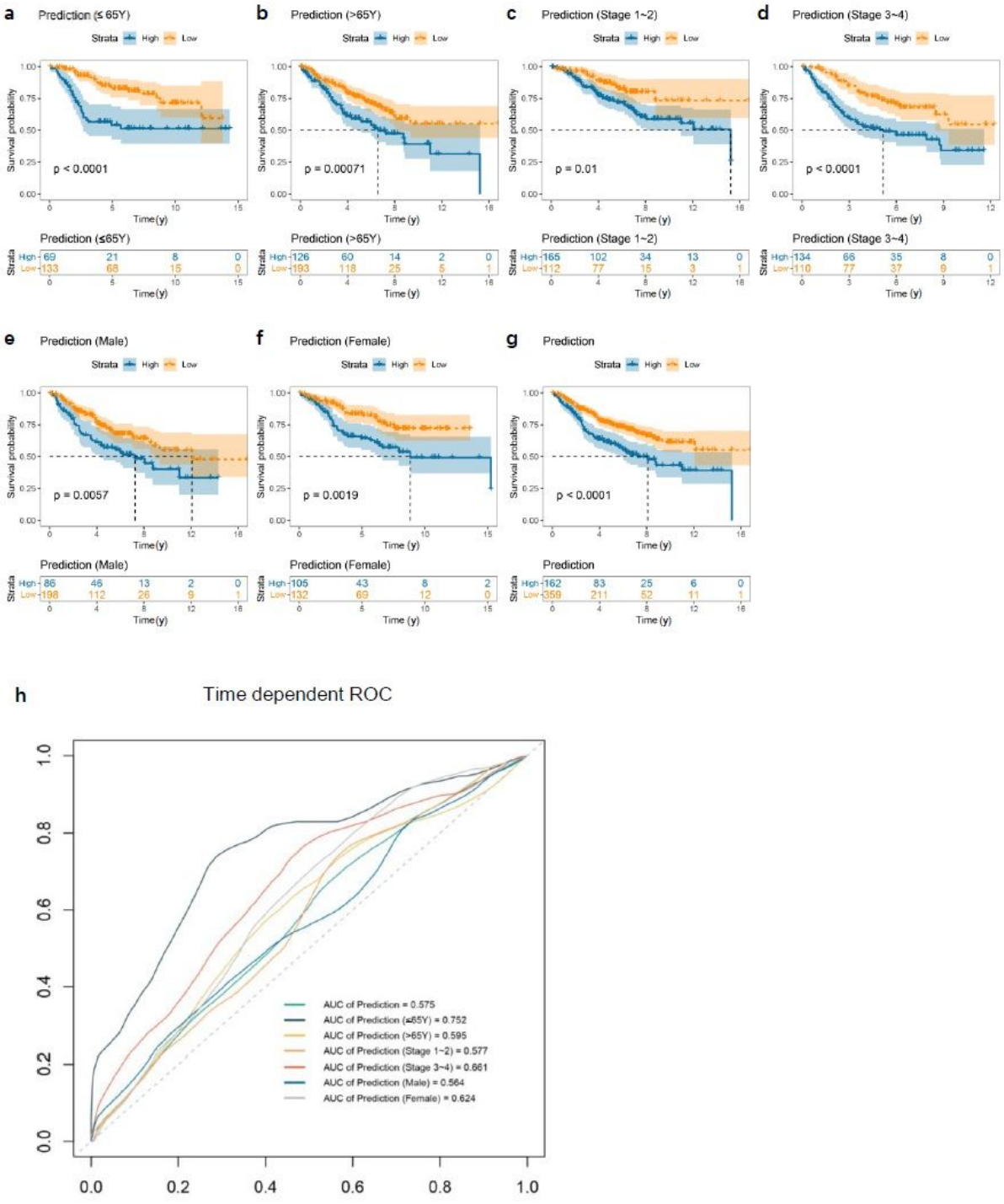

Figure 6

Validation of stromal score-based gene signature in GSE39582 cohort. Subgroup analysis of OS based on a,b age, c,d TNM stage, and e,f gender of colon cancer patients in GSE39582 cohort. g Kaplan-Meier curves of OS for the integrated prediction model encompassing the four genes in GSE39582 cohort. h ROC curves of the multigene prediction model for the whole GSE39582 cohort and different subgroups, with AUC values.

Supplementary Files

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

- Additionalfile1FigureS1.pdf
- Additionalfile1FigureS1.pdf
- Additionalfile2FigureS2.pdf
- Additionalfile2FigureS2.pdf