DNA Repair-related Gene Signature In Predicting Prognosis Of Colorectal Cancer Patients

Min-Yi Lv (lvmy8@mail2.sysu.edu.cn)
Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. https://orcid.org/0000-0003-0155-9204

Wei Wang
Huzhou Maternity and Child Care Hospital

Min-Er Zhong
Sun Yat-sen University Sixth Affiliated Hospital

Du Cai
Sun Yat-sen University Sixth Affiliated Hospital

Dejun Fan
Sun Yat-sen University Sixth Affiliated Hospital

Cheng-Hang Li
Sun Yat-sen University Sixth Affiliated Hospital

Wei-Bin Kou
Sun Yat-sen University Sixth Affiliated Hospital

Ze-Ping Huang
Sun Yat-sen University Sixth Affiliated Hospital

Xin Duan
Sun Yat-sen University Sixth Affiliated Hospital

Qi-Qi Zhu
Sun Yat-sen University Sixth Affiliated Hospital

Chuling Hu
Sun Yat-sen University Sixth Affiliated Hospital

Xiaosheng He
Sun Yat-sen University Sixth Affiliated Hospital

Feng Gao
Sun Yat-sen University Sixth Affiliated Hospital

Primary research

Keywords: DNA repair-related genes, Prognostic, Colorectal cancer, Prediction model

Posted Date: January 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-143768/v1

Version of Record: A version of this preprint was published at Frontiers in Genetics on April 11th, 2022.
See the published version at https://doi.org/10.3389/fgene.2022.872238.

DNA repair-related gene signature in predicting prognosis of colorectal cancer patients

Min-Yi Lv[1,4#], Wei Wang[2#], Min-Er Zhong[1,4], Du Cai[1,4], Dejun Fan[1,3,4], Cheng-Hang Li[1,4], Wei-Bin Kou[1,4], Ze-Ping Huang[1,4], Xin Duan[1,4], Chuling Hu[1,4], Qi-Qi Zhu[1,4], Xiaosheng He[1,4,*], Feng Gao[1,4,*]

[1]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [2]Department of Gynecology, Huzhou Maternity & Child Health Care Hospital, Huzhou, Zhejiang Province, China. 3Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong Province, China. [4]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangzhou, China

# These authors contributed equally to this study

*Correspondence: hexsheng@mail.sysu.edu.cn; gaof57@mail.sysu.edu.cn

Abstract

Background: Increasing evidence has depicted that DNA repair-related genes (DRGs) are associated with the prognosis of colorectal cancer (CRC) patients. Thus, the aim of this study was to evaluate the impact of DNA repair-related gene signature (DRGS) in predicting the prognosis of CRC patients.

Method: In this study, we retrospectively analyzed the gene expression profiles from six CRC cohorts. A total of 1,768 CRC patients with complete prognostic information were divided into training cohort (n=566) and 2 validation cohorts (n=624 and 578, respectively). LASSO-Cox model was applied to construct a prediction model.

Results: Among 1,376 DRGs, a prognostic DRGS consisting of 11 distinct genes stratified patients into high and low -risk groups. In all cohorts, patients in the high -risk groups had significantly worse disease-free survival (DFS) compared with those in the low-risk groups (training cohort: hazard ratio (HR) = 2.40, 95% confidence interval (CI) = 1.67-3.44, P < 0.001; validation-1: HR = 2.20, 95% CI = 1.38-3.49, P < 0.001; validation-2 cohort: HR = 2.12, 95% CI = 1.40-3.21, P < 0.001). After adjusting for clinical features and molecular types, DRGS still remained as an independent prognostic marker in multivariable analysis (training cohort: HR = 1.80; 95% CI = 1.22-2.64, P = 0.0028; validation-1: HR = 1.85, 95% CI = 1.13-3.02, P = 0.015; validation-2 cohort: HR = 1.75, 95% CI = 1.15-2.65, P = 0.0085). Gene Set Enrichment Analysis (GSEA) showed significant dysregulated pathways in the high-risk involved in angiogenesis, KRAS signaling, epithelial mesenchymal transit (EMT) and myogenesis (P < 0.001).

Conclusions: DNA repair-related gene signature is a favorable prognostic model for patients with CRC, and further studies are necessary to validate the exact biological mechanism.

Keywords: DNA repair-related genes, Prognostic, Colorectal cancer, Prediction model

Background

With the third highest incidence rate in the world, colorectal cancer (CRC) is a serious threat to human health[1]. Nowadays, due to lifestyle changes, there is an increasingly high incidence of mortality from CRC[2]. As one of the most common gastrointestinal tumors in general surgery, colorectal cancer is a multifactorial disease with extremely complex pathogenesis [3]. At present, the early diagnosis of CRC has involved epigenetics, genomics and so on [4]. DNA repair is a series of processes by which a cell recognizes and corrects damage to the DNA molecules that encode its genome, [5, 6] and it is extremely important for maintaining the stability of the genome and protecting the genome from damage by endogenous and environmental agents [7]. It is estimated that human cells suffer more than 2×10^4 DNA damage events per day,[8] but generally speaking, cells can respond to this damage through efficient and highly regulated DNA repair mechanisms[8, 9]. Repair mechanisms include nuclear excision repair (NER), base excision repair (BER), mismatch repair (MMR) and double strand break repair (DSBR) [9]. As we all know, genomic instability caused by the destruction of DNA damage and repair mechanism can lead to cancer progression and DNA repair genes were often found mutate in cancer [10-14]. Recently, Knijnenburg et al. discovered mutations related to DNA damage response genes by analyzing TCGA data, and found that cases in colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) datasets carried multiple mutations of DNA damage response and repair gene[11].

Due to limited options for capturing the molecular heterogeneity of the disease and the lack of consideration and sufficient validation of other gene expressions, few of the prognostic model of early-stage CRC have been applied in clinical practice[15, 16] . Thus, an accurate method is needed to identify effective prognostic models to assess the disease-free survival (DFS) of patients with CRC and provide guidance for clinicians in treatment. The aim of the present study was to examine the inter-relationships between DNA repair-related genes and colorectal cancer, in order to identify effective prognostic models to assess the disease-free survival (DFS) of patients with CRC and provide guidance for clinicians in early diagnosis and treatment.

Materials and methods

Patients

We retrospectively analyzed the gene expression profiles of CRC samples from 6 public cohorts. Totally, 1,768 samples were available for analysis in the current study. The CIT/GSE39582 (n = 566) was used for training the model, and The Cancer Genome Atlas colorectal cancer (TCGA, n = 624) was selected to serve as a validation-1 cohort. The remaining 4 microarray datasets (GSE14333, GSE33113, GSE37892 and GSE39084) were merged into a validation-2 cohort (n = 578) (Table 1). The transcriptome RNA-sequencing data of CRC samples was obtained from the TCGA data portal, and other microarray datasets were acquired directly from GEO database through. The Institutional Review Board (IRB) of our hospital approved this study, and data were collected from May 12 to October 10, 2020.

Construction and validation of DNA repair-related gene signature

Firstly, a comprehensive list of DRGs was obtained online from the MSigDB (version 6.2, https://www.gsea-msigdb.org/gsea/msigdb). We identified a list of candidate genes differentially expressed between relapsed samples and non-relapsed samples by using the "limma" R package [17]. The genes with an absolute log2-fold change of more than 1 and an adjusted $P < 0.05$ were considered for subsequent analysis. In order to minimize over-fitting risk, we applied a Cox proportional hazards regression model on CRC samples combined with the least absolute shrinkage and selection operator (LASSO) [18]. The penalty parameter was estimated by 10-fold cross-validation in the training dataset at the minimum partial likelihood deviance.

To stratify patients into high and low risk groups, the optimal cutoff value was determined by a time-dependent receiver operating characteristic (ROC) curve (survivalROC, version 1.0.3) at 5 years in the training dataset. The ROC curve was estimated by the Kaplan-Meier estimation method. To verify that the 11-DRG signature was independent of other clinical characteristics, univariate and multivariate Cox regression analyses were applied to the cohorts.

Functional annotation analysis

To evaluate the biological functions of the DRGS, enrichment analysis for differentially expressed genes (DEGs) in different groups was applied using the R package "gProfileR". We used the Bioconductor package "HTSanalyzeR" to perform Gene Set Enrichment Analysis (GSEA) to predict significant

dysregulated pathways [19, 20]. Gene sets of cancer hallmarks from MSigDB[21] were examined.

Statistical analysis

All the statistical analyses were performed on R (version3.4.3, www.r-project.org). Hazard ratios were calculated using the "survcomp" package28 (version: 1.28.4) [22]. The LASSO regression was implemented using "glmnet" R package (version: 2.0.16). Cox regression analysis was used for single-factor and multifactor analysis of the results, and the Receiver operating characteristic (ROC) curve and C-index were used to evaluate the model. A *P* value of less than 0.05 was defined as statistical significance in all tests.

Results

Construction and definition of the DNA repair-related gene signature (DRGS)

A total of 1,768 CRC patients were included in the analysis. The CIT dataset (GSE39582, *n*=566) was used as training cohort and genes with relatively high variation were kept as candidates (Table 1, Fig. 1). With median absolute deviation (MAD) > 0.5 and excluding the genes expressed less median expression level, 1,286 genes were screened out of 1,376 DRGs measured on all platforms from the datasets. In addition, in order to improve the robustness of the identification for the limited sample size, we further selected DRGs by using the Cox proportional hazards regression against 1000 randomized trials (80% portion of samples each time) to assess the correlation between each candidate gene and patients' disease-free survival (DFS) in the training cohort. 46 DRGs were robustly associated with individual patients' DFS. In order to minimize over-fitting risk, we applied a Cox proportional hazards regression model on CRC samples combined with the least absolute shrinkage and selection operator (LASSO). By using LASSO Cox regression, 11 prognostic DRGs were selected and combined for the construction of DNA repair-related gene signature (DRGS) (Fig. 2a, b). Risk scores were calculated by the formula designed by Cox regression model. The total risk score was imputed as follows: (−0.1145 × POLR2B) + (−0.0653 × RAD1) + (0.0370 × CDA) + (0.1711 × NPR2) + (−0.0328 × UBE2D2) + (−0.0992 × BCL2) + (−0.0473 × PLD6) + (0.0896 × ERBB2) + (0.1220 ×ARPC1B) + (−0.1086 × FUT4) + (−0. 0765×PSME2). Time-dependent ROC curve analysis showed that the optimal cutoff to stratify high and low risk groups was -0.147 (Fig. 2c).

Prognostic evaluation of the DRGS

Six colorectal cancer transcription datasets containing prognostic data were selected to assess prognostic ability of the DRGS. The entire CIT/GSE39582 dataset (*n*= 566) was used as a training dataset (Fig. 2d). TCGA CRC dataset was enrolled as validation-1 cohort (*n*=624), and additional datasets from GEO were combined as validation-2 cohort (n=578). Patients in the training and validation cohorts, more recurrences were found in the high-risk group compared to the low-risk group (Fig. 3a,d,g).When applied to a follow-up duration of 2, 3 and 5 years, promising prognostic values were also found base on the time-dependent ROC curve analysis in the training cohort (AUC = 0.640 at 2 years; AUC = 0.664 at 3 years; AUC = 0.653 at 5 years) , validation-1 cohort (AUC = 0.620 at 2 years; AUC = 0.628 at 3 years; AUC = 0.606 at 5 years) and validation-2 cohort (AUC = 0.645 at 2 years; AUC = 0.631 at 3 years; AUC = 0.638 at 5 years) (Fig. 3b,e,h). DRGS significantly stratified patients into high and low risk groups in the training cohort (HR=2.40, 95% CI= 1.67-3.44, $P < 0.001$), validation-1 cohort (HR=2.20, 95% CI= 1.38-3.49, $P < 0.001$), and validation-2 cohort (HR=2.12, 95% CI= 1.40-3.21, $P < 0.001$) (Fig. 3c, f, i). Besides, the overall survival (OS) in the low risk group was better than the high risk group (Additional file 1: Figure S1).

When compared with the risk scores calculated using the algorithm in the FDA-approved assay Oncotype DX colon, we found that the DRGS achieved an improved survival correlation in the training cohort (C-index, 0.78 vs. 0.60), validation-1 cohort (C-index, 0.65 vs. 0.51) and validation-2 cohort (C-index, 0.66 vs. 0.62) (Table 2).

To further investigate whether the DRGS could serve as an independent predictor of prognosis, univariate and multivariate Cox proportional hazards regression analyses were performed. As expected, age, sex, tumor stage, tumor location and pathologic gene status were associated with outcomes for CRC patients (Table 3). In the univariate analysis, DRGS, MMR status and KRAS mutation status were significantly correlated with worse prognosis in the training cohort. After adjusting for clinical features such as age, gender, tumor location and molecular types, DRGS remained an independent prognostic factor in multivariate analyses in both validation cohorts.

Functional annotation of the DRGS

A gene set enrichment analysis was performed to further investigate the potential biological processes

and examine the associated mechanisms of these 2 groups. Gene Ontology analyses revealed that some biological process pathways (extracellular region, cell proliferation, and cell adhesion) were the main enriched pathways in the high-risk group (Fig. 4a). In addition, the GSEA pathway enrichment analysis in the high-risk compared with the low-risk groups shown that metastasis-related pathways (ie, angiogenesis, KRAS signaling, epithelial mesenchymal transit and myogenesis pathways) enriched in the high-risk group (Fig. 4b, Additional file 2: Table S1). These findings suggested that the enrichment of pathways provided evidence of molecular mechanisms affected by the DRGS and thus can predict the prognosis of CRC.

Discussion

Colorectal cancer is the leading cause of death among gastrointestinal cancers. The incidence and mortality of colorectal cancer are increasing year by year, and its prognosis is closely related to early diagnosis [23, 24]. Numerous studies have highlighted the biomarkers are associated with the pathogenesis and biology of CRC, [25-28] and a lot of multigene prognostic signatures have been developed for CRC [28-32]. Unfortunately, the accuracy of their prognosis predictions remains uncertain[33]. More efforts are needed to achieve a good prognosis for CRC, which is still considered a challenge.

Studies on DNA repair pathways and DRGs have found some new results. Inactivation of DRGs can disrupt genome integrity, which can increase the risk of the accumulation of gene mutations associated with cancer development[34]. Some reports suggest that the DNA repair process is involved in the intrinsic response of the body to chemotherapeutic agents and has been shown to be associated with the mechanisms of resistance acquired during treatment[34, 35]. In this study, we were aimed to identify and validate a robust and reliable DNA repair-related gene signature (DRGS) and thus improve the accuracy of survival prediction for CRC patients.

This study consisted of a training cohort and 2 validation cohorts, which included 1,768 patients with CRC. Our prognostic DRGS can stratify CRC patients into two groups with different survival outcomes. A multivariate analysis suggested that DRGS remained an independent prognostic factor and significantly associated with poor prognosis in CRC. Furthermore, the C index results of the DRGS showed its clinical superiority to Oncotype DX[36]. Thus, it offers a significantly promising prognostic

biomarker potential compared to the clinicopathological risk factors that are currently in use. The GSEA revealed that metastasis-related pathways (ie, angiogenesis, KRAS signaling, epithelial mesenchymal transit and myogenesis pathways) were enriched in the high-risk group, all of which were well-known to play a crucial role in the progression and proliferation of CRC in numerous studies[37-39]. Further studies are needed to clarify the effects of DNA repair in order to identify more targets and improve the prognosis of CRC patients.

There some limitations to our study. First, this is a retrospective study, although we validated the signature in independent datasets. In addition, samples from primary tumor or metastatic disease may have inconsistent genetic heterogeneity, which could lead to sampling bias[40, 41]. What's more, systematic errors result from analyzing samples of disparate databases, and not all batch effects can be eliminated based on their complexity. Although we investigated as many genes as possible, further clinical, and pharmacological tests are needed to validate our results.

Conclusion

In summary, our work provides an accurate prognostic approach for estimating survival outcomes of CRC patients. Further prospective studies are needed to evaluate the clinical application of this signature for the prognosis of CRC.

Abbreviations

DRGs: DNA repair-related genes; DRGS: DNA repair-related gene signature; CRC: colorectal cancer; DEGs: differentially expressed genes; HR: hazard ratio; GEO: Gene Expression Omnibus; GSEA: Gene Set Enrichment Analysis; LASSO: least absolute shrinkage and selection operator; OS: overall survival; ROC: receiver operating characteristic; TCGA: The Cancer Genome Atlas.

Acknowledgements

Not applicable.

Authors' contributions

MYL and WW contributed equally to this study. MYL, WW, XSH and FG contributed to the study concept and design, the acquisition, analysis, and interpretation of data, and the drafting of the manuscript. MEZ, DC, DJF, CHL, WBK, ZPH, XD, CLH and QQZ contributed to the data collections and manuscript reviews. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No. 82002221, FG), the Fundamental Research Funds for the Central Universities (No.20ykpy05, FG), the Sun Yat-sen University 100 Top Talent Scholars Program – China (No. P20190217202203617, FG), Project funded by China Postdoctoral Science Foundation (No. 2020M683121, MEZ).

Availability of data and materials

The datasets generated and analyzed during the current study are available in the TCGA cohort data was downloaded from Broad GDAC Firehose (http://gdac. broadinstitute.org/). Other microarray datasets were acquired directly from GEO database. (https://www.ncbi.nlm.nih.gov/geo/query).

Ethics approval and consent to participate

This is a retrospective trial from public datasets with minimal risk and we petition for waiver of ethics consent.

Consent for publication

We have obtained consents to publish this paper from all participants of this study.

Competing interests

The authors declared no financial conflict of interests.

Author details

[1]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. [2]Department of Gynecology, Huzhou Maternity & Child Health Care Hospital, Huzhou, Zhejiang Province, China.[3]Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong Province, China. [4]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangzhou, China

Reference

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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. Zheng ZX, Zheng RS, Zhang SW, Chen WQ: Colorectal cancer incidence and mortality in China, 2010. Asian Pac J Cancer Prev 2014, 15(19):8455-8460.
3. Migliore L, Migheli F, Spisni R, Coppede F: Genetics, cytogenetics, and epigenetics of colorectal cancer. J Biomed Biotechnol 2011, 2011:792362.
4. Marcuello M, Vymetalkova V, Neves RPL, Duran-Sanchon S, Vedeld HM, Tham E, van Dalum G, Flugen G, Garcia-Barberan V, Fijneman RJ et al: Circulating biomarkers for early detection and clinical management of colorectal cancer. Mol Aspects Med 2019, 69:107-122.
5. Zinovkina LA: Mechanisms of Mitochondrial DNA Repair in Mammals. Biochemistry (Moscow) 2018, 83(3):233-249.
6. Burdak-Rothkamm S, Rothkamm K: DNA Damage Repair Deficiency and Synthetic Lethality for Cancer Treatment. Trends Mol Med 2020.
7. Friedberg EC: How nucleotide excision repair protects against cancer. Nat Rev Cancer 2001, 1(1):22–33.
8. Wood TLaRD: Quality Control by DNA Repair. 1999, Science 286 (5446), 1897–1905.
9. Iyama T, Wilson DM, 3rd: DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) 2013, 12(8):620-636.
10. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA: DNA repair pathways as targets for cancer therapy. Nat Rev Cancer 2008, 8(3):193-204.
11. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, Fan H, Shen H, Way GP, Greene CS et al: Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Reports 2018, 23(1):239-254.e236.
12. Reilly NM, Novara L, Di Nicolantonio F, Bardelli A: Exploiting DNA repair defects in colorectal cancer. Mol Oncol 2019, 13(4):681-700.
13. Turgeon MO, Perry NJS, Poulogiannis G: DNA Damage, Repair, and Cancer Metabolism. Front Oncol 2018, 8:15.
14. Young Kwang Chae JFA, Aparna Kalyan, Giles aFJ: Genomic landscape of DNA repair genes in cancer. 2016.
15. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P et al: The consensus molecular subtypes of colorectal cancer. Nat Med 2015, 21(11):1350-1356.
16. Phipps AI, Limburg PJ, Baron JA, Burnett-Hartman AN, Weisenberger DJ, Laird PW, Sinicrope FA, Rosty C, Buchanan DD, Potter JD et al: Association between molecular subtypes of colorectal cancer and patient survival. Gastroenterology 2015, 148(1):77-87 e72.
17. Diboun I, Wernisch L, Orengo CA, Koltzenburg M: Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics 2006, 7:252.
18. R T: The lasso method for variable selection in the Cox model. Stat Med 1997, 16(4):385-395.
19. Wang X, Terfve C, Rose JC, Markowetz F: HTSanalyzeR: an R/Bioconductor package for integrated network analysis of high-throughput screens. Bioinformatics 2011, 27(6):879-880.
20. Aravind Subramanian PT, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert: Gene set

enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. 2005.
21. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P: The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015, 1(6):417-425.
22. Schroder MS, Culhane AC, Quackenbush J, Haibe-Kains B: survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics 2011, 27(22):3206-3208.
23. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2016. CA Cancer J Clin 2016, 66(1):7-30.
24. Siegel RL, Fedewa SA, Anderson WF, Miller KD, Ma J, Rosenberg PS, Jemal A: Colorectal Cancer Incidence Patterns in the United States, 1974-2013. J Natl Cancer Inst 2017, 109(8).
25. Lech G, Slotwinski R, Slodkowski M, Krasnodebski IW: Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. World J Gastroenterol 2016, 22(5):1745-1755.
26. Das V, Kalita J, Pal M: Predictive and prognostic biomarkers in colorectal cancer: A systematic review of recent advances and challenges. Biomedicine & Pharmacotherapy 2017, 87:8-19.
27. De Rosa M, Rega D, Costabile V, Duraturo F, Niglio A, Izzo P, Pace U, Delrio P: The biological complexity of colorectal cancer: insights into biomarkers for early detection and personalized care. Therap Adv Gastroenterol 2016, 9(6):861-886.
28. Reena Shah EJ, Victoire Vidart, Peter J.K. Kuppen, John A. Conti, and Nader K. Francis: Biomarkers for Early Detection of Colorectal Cancer and Polyps: Systematic Review. 2014.
29. Gao F, Wang W, Tan M, Zhu L, Zhang Y, Fessler E, Vermeulen L, Wang X: DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. Oncogenesis 2019, 8(9):44.
30. Kandimalla R, Gao F, Matsuyama T, Ishikawa T, Uetake H, Takahashi N, Yamada Y, Becerra C, Kopetz S, Wang X et al: Genome-wide Discovery and Identification of a Novel miRNA Signature for Recurrence Prediction in Stage II and III Colorectal Cancer. Clin Cancer Res 2018, 24(16):3867-3877.
31. Kandimalla R, Ozawa T, Gao F, Wang X, Goel A, Group TCCS: Gene Expression Signature in Surgical Tissues and Endoscopic Biopsies Identifies High-Risk T1 Colorectal Cancers. Gastroenterology 2019, 156(8):2338-2341 e2333.
32. Ozawa T, Kandimalla R, Gao F, Nozawa H, Hata K, Nagata H, Okada S, Izumi D, Baba H, Fleshman J et al: A MicroRNA Signature Associated With Metastasis of T1 Colorectal Cancers to Lymph Nodes. Gastroenterology 2018, 154(4):844-848 e847.
33. Fung KY, Nice E, Priebe I, Belobrajdic D, Phatak A, Purins L, Tabor B, Pompeia C, Lockett T, Adams TE et al: Colorectal cancer biomarkers: to be or not to be? Cautionary tales from a road well travelled. World J Gastroenterol 2014, 20(4):888-898.
34. Bouwman P, Jonkers J: The effects of deregulated DNA damage signalling on cancer chemotherapy response and resistance. Nat Rev Cancer 2012, 12(9):587-598.
35. Badura M, Braunstein S, Zavadil J, Schneider RJ: DNA damage and eIF4G1 in breast cancer cells reprogram translation for survival and DNA repair mRNAs. Proc Natl Acad Sci U S A 2012, 109(46):18767-18772.
36. <Clark-langone-2010-Translating-tumor-biology-into-pers.pdf>.
37. Cooks T, Pateras IS, Tarcic O, Solomon H, Schetter AJ, Wilder S, Lozano G, Pikarsky E, Forshew T, Rosenfeld N et al: Mutant p53 prolongs NF-kappaB activation and promotes

chronic inflammation and inflammation-associated colorectal cancer. Cancer Cell 2013, 23(5):634-646.

38. De Simone V, Franze E, Ronchetti G, Colantoni A, Fantini MC, Di Fusco D, Sica GS, Sileri P, MacDonald TT, Pallone F et al: Th17-type cytokines, IL-6 and TNF-alpha synergistically activate STAT3 and NF-kB to promote colorectal cancer cell growth. Oncogene 2015, 34(27):3493-3503.

39. Lu YX, Ju HQ, Wang F, Chen LZ, Wu QN, Sheng H, Mo HY, Pan ZZ, Xie D, Kang TB et al: Inhibition of the NF-kappaB pathway by nafamostat mesilate suppresses colorectal cancer growth and metastasis. Cancer Lett 2016, 380(1):87-97.

40. Intratumor Heterogeneity and Branched Evolution. The New England Journal of Medicine 2012.

41. Mimori K, Saito T, Niida A, Miyano S: Cancer evolution and heterogeneity. Ann Gastroenterol Surg 2018, 2(5):332-338.

Figure legends

Fig.1 Schema flow chart of the study.

Fig.2 (**a**) Identification and selection of prognostic genes by LASSO Cox proportional hazards regression. (**b**) The establishment of 11 DNA repair-related genes signature from the LASSO COX regression. (**c**) The optimal cutoff point of prognostic gene signature at 5-year OS endpoint from ROC curve. (**d**) Heatmap of the 11 DNA repair-related genes in two risk groups.

Fig.3 (**a, d, and g**) Distribution of the DRGS risk score and its correlation to recurrence in the training, validation-1, and validation-2 cohort. (**b, e, and h**) Time-dependent ROC analysis of disease-free survival for CRC patients in the training, validation-1, and validation-2 cohorts at the time points of 2, 3 and 5 years. (**c, f, and i**) Kaplan–Meier curves comparing survival of patients within the low and high risk groups in training cohort, validation-1, and validation-2 cohorts. *P*-values were calculated using log-rank tests.

Fig.4 (**a**) Gene ontology of the differentially expressed genes between the two risk groups. "GeneRatio" is the percentage of total differential genes in the given GO term. (**b**) GSEA showed several metastasis-related processes enriched in the high risk group, including angiogenesis, KRAS signaling, epithelial mesenchymal transit (EMT) and myogenesis signal pathways.

Supplementary information

Additional file 1: Figure S1 (**a**) Distribution of the DRGS risk score and its correlation to survival status. (**b**) Time-dependent ROC analysis of overall survival for CRC patients at the time points of 5 and 10 years. (**c**) Kaplan–Meier curves comparing overall survival of patients within the low and high risk groups. *P*-values were calculated using log-rank tests.

Additional file 2: Table S1. GSEA results for the comparison of high- vs. low- risk groups.

Figures

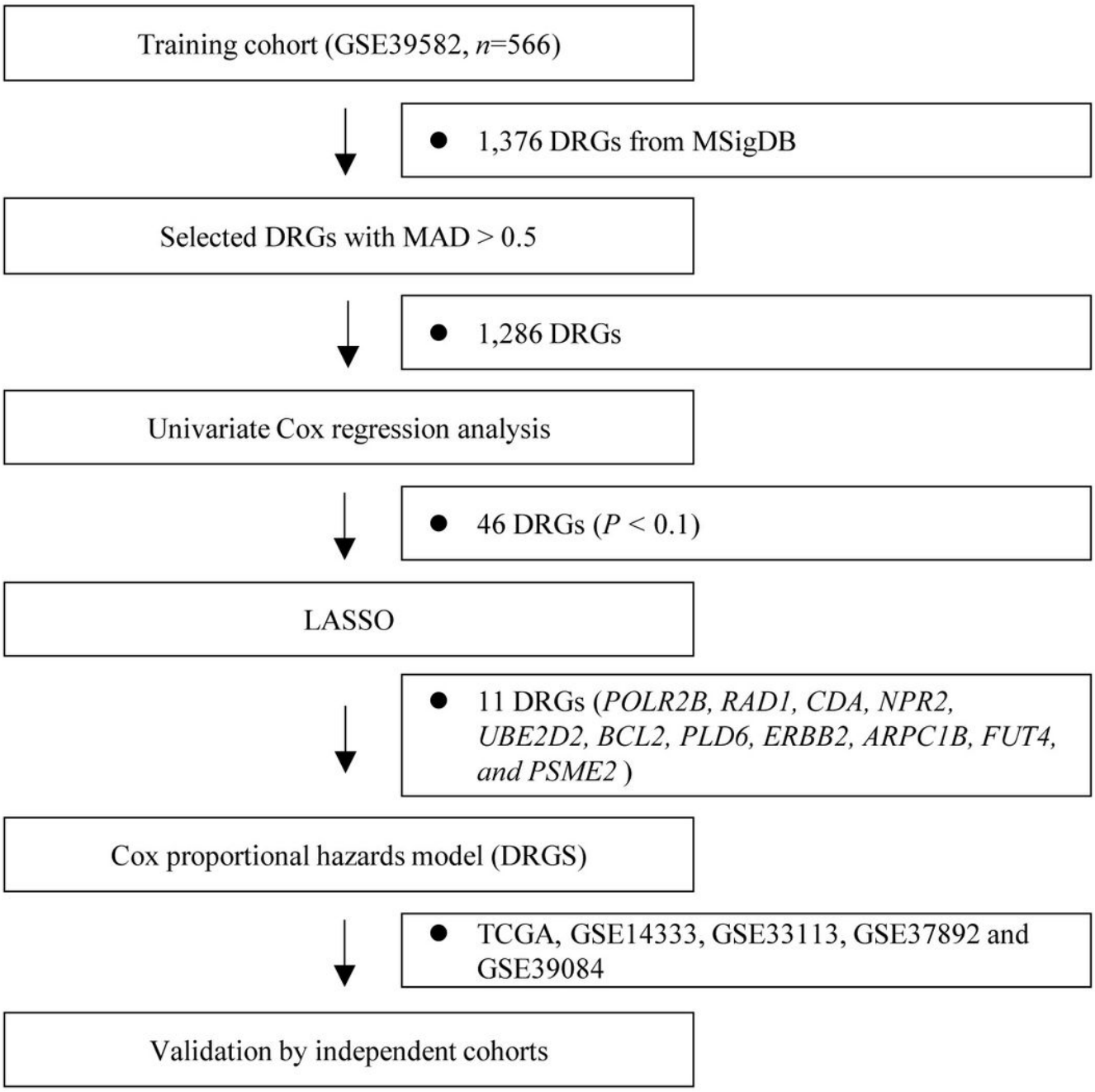

Figure 1

Schema flow chart of the study.

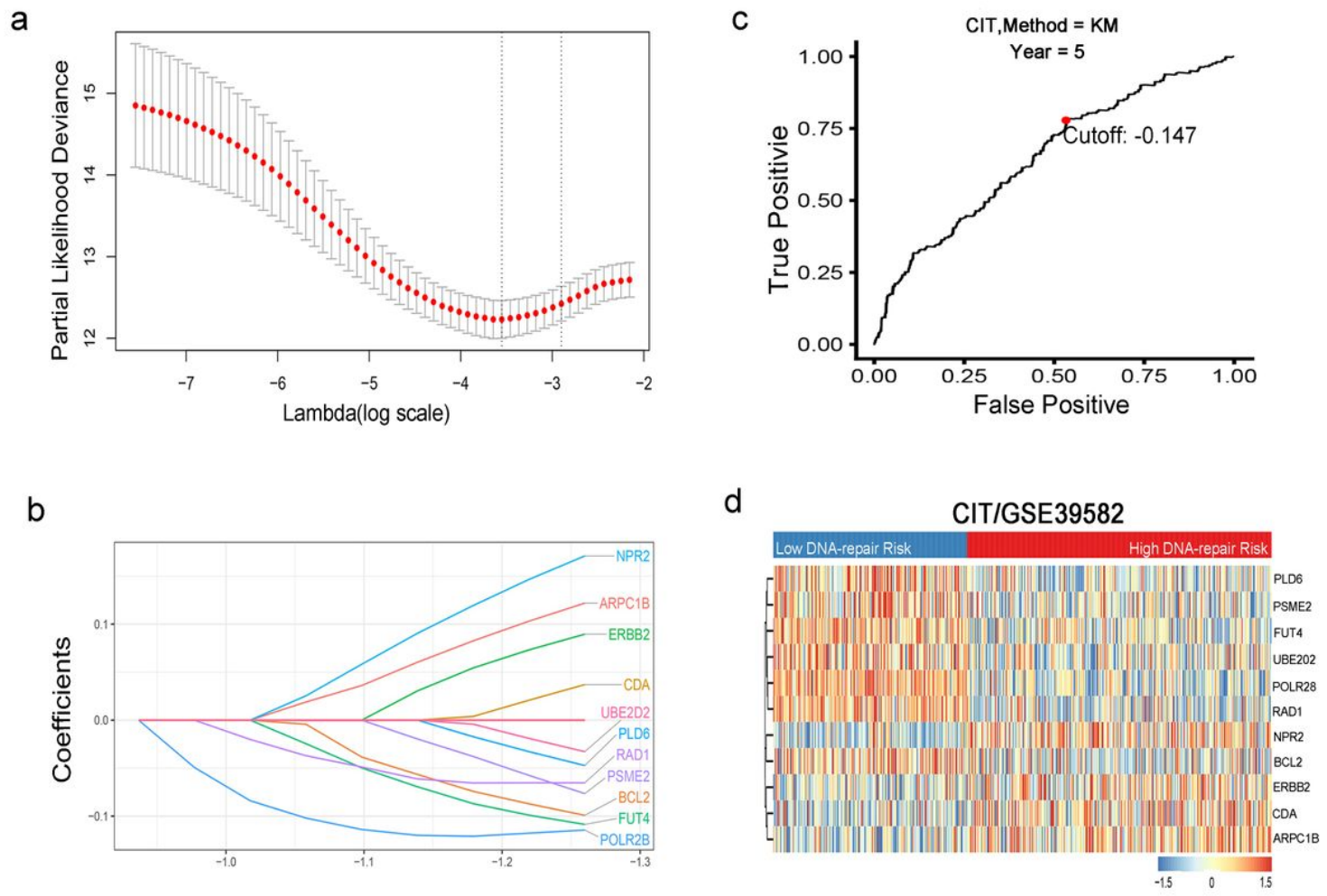

Figure 2

(a) Identification and selection of prognostic genes by LASSO Cox proportional hazards regression. (b) The establishment of 11 DNA repair-related genes signature from the LASSO COX regression. (c) The optimal cutoff point of prognostic gene signature at 5-year OS endpoint from ROC curve. (d) Heatmap of the 11 DNA repair-related genes in two risk groups.

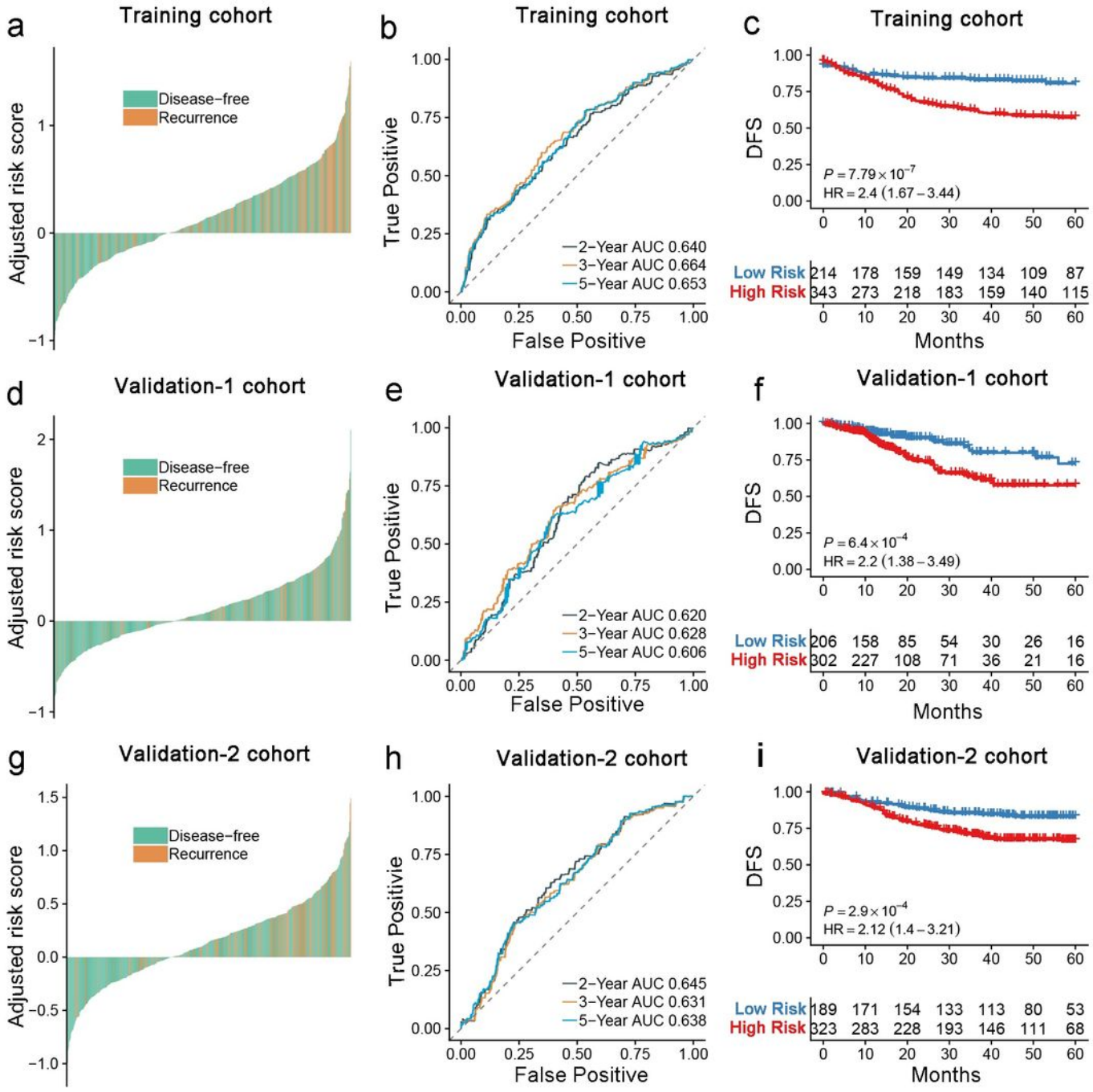

Figure 3

(a, d, and g) Distribution of the DRGS risk score and its correlation to recurrence in the training, validation-1, and validation-2 cohort. (b, e, and h) Time-dependent ROC analysis of disease-free survival for CRC patients in the training, validation-1, and validation-2 cohorts at the time points of 2, 3 and 5 years. (c, f, and i) Kaplan–Meier curves comparing survival of patients within the low and high risk groups in training cohort, validation-1, and validation-2 cohorts. P□values were calculated using log□rank tests.

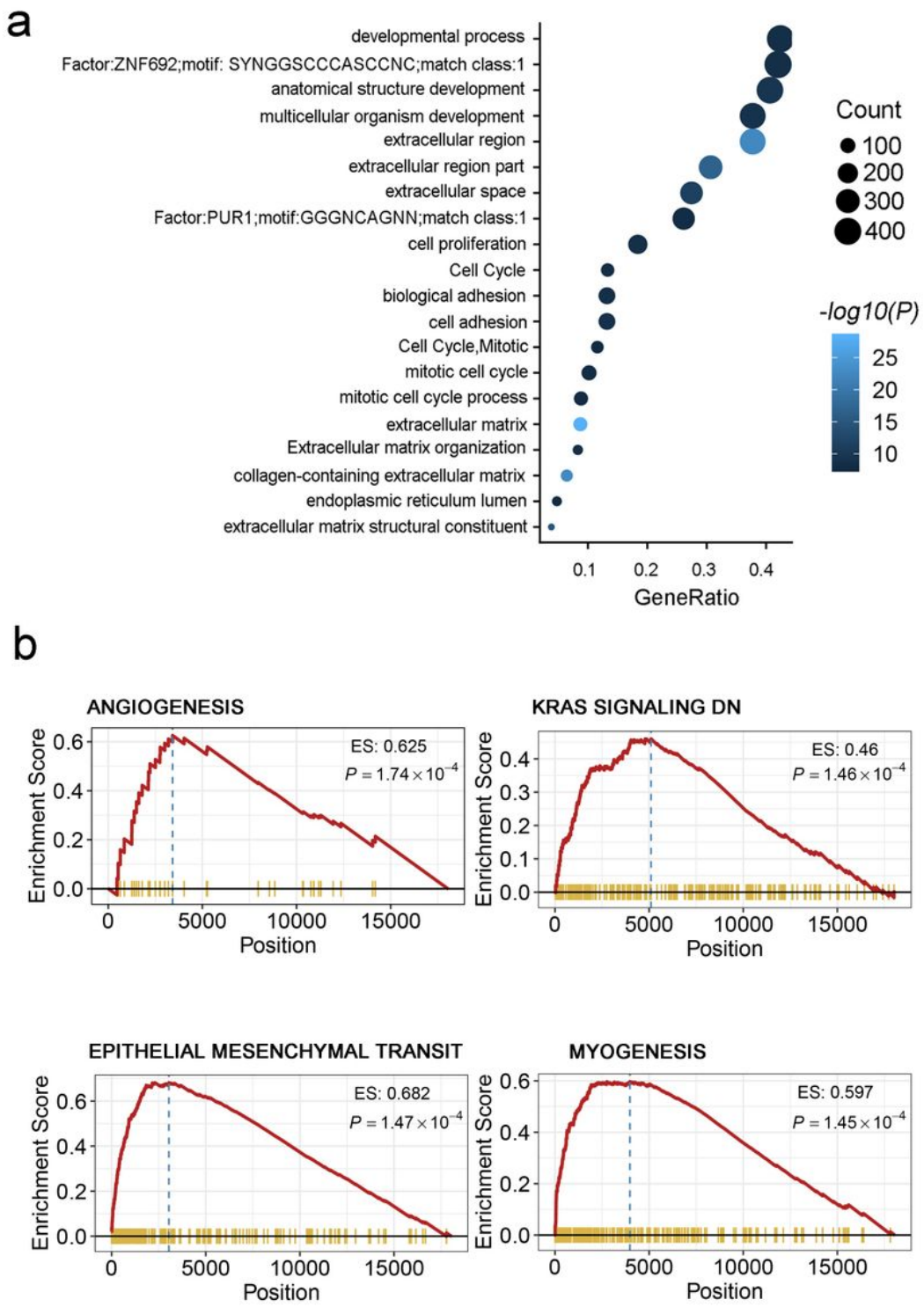

Figure 4

(a) Gene ontology of the differentially expressed genes between the two risk groups. “GeneRatio” is the percentage of total differential genes in the given GO term. (b) GSEA showed several metastasis-related processes enriched in the high risk group, including angiogenesis, KRAS signaling, epithelial mesenchymal transit (EMT) and myogenesis signal pathways.

Supplementary Files

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

- Additionalfile1FigureS1.tif
- Additionalfile2TableS1.docx
- Table1.docx
- Table2.docx
- Table3.docx