Am J Cancer Res 2021;11(10):5141-5154
www.ajcr.us /ISSN:2156-6976/ajcr0135879

Original Article

Clinical relevance of stem cell surface markers CD133, CD24, and CD44 in colorectal cancer

Jing Li Huang[1*], Masanori Oshi[1,2*], Itaru Endo[2], Kazuaki Takabe[1,2,3,4,5,6]

[1]Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263, USA; [2]Department of Gastroenterological Surgery, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [3]Department of Surgery, Jacobs School of Medicine and Biomedical Sciences, State University of New York, Buffalo, New York 14263, USA; [4]Division of Digestive and General Surgery, Niigata University Graduate School of Medical and Dental Sciences, Niigata 951-8520, Japan; [5]Department of Breast Surgery, Fukushima Medical University School of Medicine, Fukushima 960-1295, Japan; [6]Department of Breast Surgery and Oncology, Tokyo Medical University, Tokyo 160-8402, Japan. [*]Equal contributors.

Received May 31, 2021; Accepted September 4, 2021; Epub October 15, 2021; Published October 30, 2021

Abstract: Colon cancer stem cells (CSC) identified by cell surface markers CD133, CD24, and CD44, have been shown to be involved with tumor formation, chemotherapy resistance, and the progression of metastatic disease. Using an *in silico* translational approach, we hypothesize that a combination of these CSC markers has prognostic value in a large cohort of patients with colorectal cancer. Clinicopathologic and RNA expression data from a total of 594 colorectal cancer (CRC) patients from TCGA were analyzed. The expression of CD133, CD24, and CD44 was individually defined as "high" or "low" based on the median expression. Disease specific survival (DSS) and overall survival (OS) were not associated with tumors that are CD133-high or CD44-high alone. Patients with CD24-high tumors have significantly better DSS (P<0.001) and OS (P = 0.043). CD24-high, CD44-high and CD133-high tumors were associated with significantly greater EGFR, KRAS and Ki67 expression (all P<0.001). CD133, CD24 and CD44-high tumors were independently enriched for conventional stemness-related signaling pathways such as Wnt/β-catenin and Hedgehog signaling pathways. There was no survival difference linked to CD133-high/CD44-low patients, but CD44-high/CD24-low patients have worse DSS (P = 0.005) compared with CD44-low/CD24-high patients. CD133-high/CD24-low tumors show significant negative enrichment of MYC targets, E2F targets, G2M checkpoint and mitotic spindle gene sets, suggesting less cell proliferation in these tumors. Patients with CD133-high/CD24-low tumors have worse DSS (P = 0.004) and OS (P = 0.044), and are more likely to have early and late recurrences. In conclusion, we demonstrated that CD133-high/CD24-low tumors may predict colorectal cancer prognosis.

Keywords: Cancer stem cells, CD133, CD24, CD44, colon cancer, gene set, cancer biomarker

Introduction

Colon and rectal cancers (CRC) are the 3rd leading cause of cancer-related death in the US. Metastatic disease occurs in up to 15-20% of newly diagnosed CRC, and the 5-year survival for stage IV disease is only 12% [1]. Up to 60-70% of metastatic colon cancer recurs after resection. The survival benefit of adjuvant chemotherapy in stage III node-positive CRC is only as high as 25% [2]. Ongoing efforts in investigating new treatment strategies for CRC include understanding the involvement of cancer stem cells.

Since 2007, the discovery of a subpopulation of tumor-initiating cells in colon cancer has led to greater insight into the metabolic characteristics of colon cancer stem cells (CSC) and reshaped the understanding of cancer development and metastasis [3, 4]. Features of CSCs include a capacity for self-renewal, resistance to chemotherapy and/or radiation, and greater metastatic potential that may be shaped by the tumor microenvironment [5]. Investigations into the functional phenotypes of CSC have been propelled by the availability of inhibitors that may block signaling pathways that traditionally regulate CSC growth, such as Wnt/β-

catenin, Notch, and Hedgehog signaling pathways [6].

Specific CSC surface markers may be prognostic [7-9]. However, the clinical relevance of which surface markers or combination of surface markers has been controversial. Putative stem cell markers include CD133-a hematopoietic stemness marker as known as Prominin 1, CD24-a cell adhesion molecule, and CD44-a hyaluronic receptor. *In vitro* studies have shown that a single colon cancer cell highly expressing CD133 or CD44 can give rise to tumors that have the full array of heterogeneity [3, 4, 10, 11].

To date, the clinical relevance of these CSC surface markers in CRC has been reported only in small institutional studies [12-20]. In addition, the heterogeneity in techniques and antibodies to identify these cell surface markers make comparisons across studies difficult. Therefore, we propose the use of transcriptomic data from The Cancer Genome Atlas to evaluate clinical relevance of these putative CSCs on a large scale. We hypothesize that a combination of CSC surface markers is prognostic in colorectal cancer.

Methods

Clinical and gene expression data of colorectal cancer cohort

The clinicopathologic and RNA sequence data of 594 patients from TCGA-colorectal cancer cohort were obtained and downloaded via Genomic Data Commons Data Portal (GDC). In the cohort, 63.6% of patients had colon cancer and 26% had rectal cancer, and 10% of patients had mucinous adenocarcinoma of the colon or rectum. The mutation status was obtained from cBioportal (https://www.cbioportal.org), as previously performed and described [21-23]. The frequency of APC mutation in the TCGA cohort is 72.5%, 40.8% for KRAS mutation and 11.6% for BRAF mutation. The staging for colorectal cancer was performed in accordance with American Joint Committee on Cancer staging guidelines. The approval of the Roswell Park Institutional Review Board was waived due to the deidentified nature of the data points.

Gene set enrichment analysis (GSEA)

Fifty Hallmarks of Cancer gene sets in the Molecular Signatures Database (MSigDB) [24] collection were analyzed to investigate the biological function using GSEA as previously demonstrated by the Broad Institute (http://www.gsea-msigdb.org/gsea/index.jsp) [25]. Enriched gene sets were categorized in accordance with previous publications [26]. For instance, cell proliferation-related gene sets consist of G2M checkpoints [21, 27], E2F Targets [22], MYC Targets v1 and v2 [28, 29], and Mitotic Spindle [30] as we have previously reported. The False Discovery Rate (FDR) was used for statistical analysis. Because multiple gene sets were analyzed with our study, we used the FDR value of less than 0.25 as the cutoff for significance, which was recommended by the Broad Institute to adjust for gene set size.

Cell composition analysis

Based on the transcriptomic data of the cohort, the web-based computational algorithm, xCell, was used to perform cell types enrichment analysis of immune cells between tumor groups as previously described [22, 27, 31-35]. The xCell algorithm allows researchers to get the transcriptomic data of 64 cell types in tumor microenvironment. These include not only immune cells such as regulatory T cells [36], T helper cells [37], M1 and M2 macrophages [35], CD8+ cells [33], CD4 memory cells [33], dendritic cells [38], and B cells, but also stromal cells [39, 40]. The xCell algorithm was developed at the University of California-San Francisco (https://www.xcell.ucsf.edu).

Statistical analyses

The median expression level of CD133, CD24, and CD44 was used to distinguish between “high” versus “low” expression. Histograms were created based on the expression levels of each marker. One-way ANOVA test was used for statistical comparisons between groups. Fisher’s exact test was used for the recurrence analysis. The *P*-value of less than 0.05 represented a statistically significant difference. Tukey type boxplots showed median and interquartile level values; Mann-Whitney U test was used to calculate *P* values. Kaplan-Meier plots with log-rank test were used for survival analy-

ses. All statistical analyses were performed using R software (version 4.0.1, http://www.r-project.org/).

Results

CD133-high tumors are associated with stemness signaling pathways, but not with clinical or pathologic outcomes

Based on data from *in vitro* studies that support CD133 as a stemness marker, we first expected that CD133-high tumors are associated with pathways and genes that are abnormally regulated in cancer stem cells. In particular, the gene set enrichment analysis (GSEA) of the Hallmarks of cancer genes showed that CD133-high CRC significantly enriched for Wnt/β-catenin signaling (normalized enrichment score (NES) = 1.40, and false discovery rate (FDR) = 0.17), TNF-α signaling via NFκB (NES = 1.33, FDR = 0.16), PI3K AKT MTOR signaling (NES = 1.33, FDR = 0.19), and Hedgehog signaling (NES = 1.31, FDR = 0.14), which are all genes related to conventional stemness signaling pathways (**Figure 1A**). Further, treatment-relevant markers of CRC, such as EGFR, KRAS and proliferation marker Ki67 were all significantly elevated in CD133-high CRC (all P<0.001, **Figure 1B**), This suggests high proliferation of CD133-high CRC tumors. Clinicopathologic features such as AJCC cancer stage, tumor location, tumor histology were not associated with CD133 expression levels (**Figure 1C**). Based on previous studies, we expected that patients with CD133-high tumors would have worse clinical outcomes; however, we found that neither disease specific survival nor overall survival was associated with tumors that are CD133-high (**Figure 1D**).

The association of CD24 and CD44 expression with clinicopathologic features of CRC

We next investigated the clinical relevance of the putative CSC markers, CD24 and CD44. Consistent with stemness, CD24-high CRC tumors were significantly enriched for Wnt/β-catenin signaling (NES = 1.45, FDR = 0.15), Notch signaling (NES = 1.41, FDR = 0.15), PI3K AKT MTOR signaling (NES = 1.39, FDR = 0.14), TNF-α signaling via NFκB (NES = 1.33, FDR = 0.13), and Hedgehog signaling (NES = 1.28, FDR = 0.15) (**Figure 2A**). CD24-high CRC was associated with significantly elevated EGFR, KRAS and Ki67 expressions (all P<0.001, **Figure 2B**). Higher expression level of CD24 correlated with left-sided rather than right-sided colon tumors (P = 0.018) and correlated with adenocarcinoma rather than mucinous histology (P<0.001, **Figure 2C**). However, no significant difference in CD24 expression level was found amongst different stages of CRC (**Figure 2C**). Although previous studies have shown that high cytoplasmic CD24 correlated with poor prognosis, we found that patients with CD24-high CRC had better disease-specific survival (P<0.001) and overall survival in the TCGA population (P = 0.04, **Figure 2D**).

As anticipated, CD44-high CRC were significantly associated with aberrant stemness related gene sets, including Wnt/β-catenin (NES = 1.30, FDR = 0.14), Hedgehog (NES = 1.35, FDR = 0.16), PI3K AKT MTOR (NES = 1.35, FDR = 0.14), and Notch signaling (NES = 1.46, FDR = 0.21) (**Figure 3A**). Treatment-relevant markers; KRAS, EGFR and Ki67 were all elevated in CD44-high tumors (all P<0.001, **Figure 3B**). Higher expression level of CD44 was significantly associated colon tumors (P = 0.003), and tumors with mucinous histology (P = 0.017, **Figure 3C**). Patients with CD44-high CRC were not significantly associated with better disease specific survival nor overall survival (**Figure 3D**).

Tumors with high expression of CSC markers are not associated with immune cell infiltrations in the tumor microenvironment

Due to the recent development on the prognostic value of the Immunoscore in CRC as well as the possibility of immune-escape of CSCs as modulated by immune cells [24, 41, 42], we investigated the association of immune cell fractions in tumors with CSC marker expressions. Tumors that were low in CD24 were significantly associated with higher M1 macrophages (P = 0.003), B cells (P = 0.03) and dendritic cells (P<0.001, Figure S1), suggesting that these tumors may attract an anti-cancer immune response. However, tumors with a combination of CD44 and CD133 or CD24 did not show significant associations with pro-cancer nor anti-cancer immune cells (Figure S2). In addition, tumors with high CD133 and CD44 were also not associated with immune cells (Figure S1).

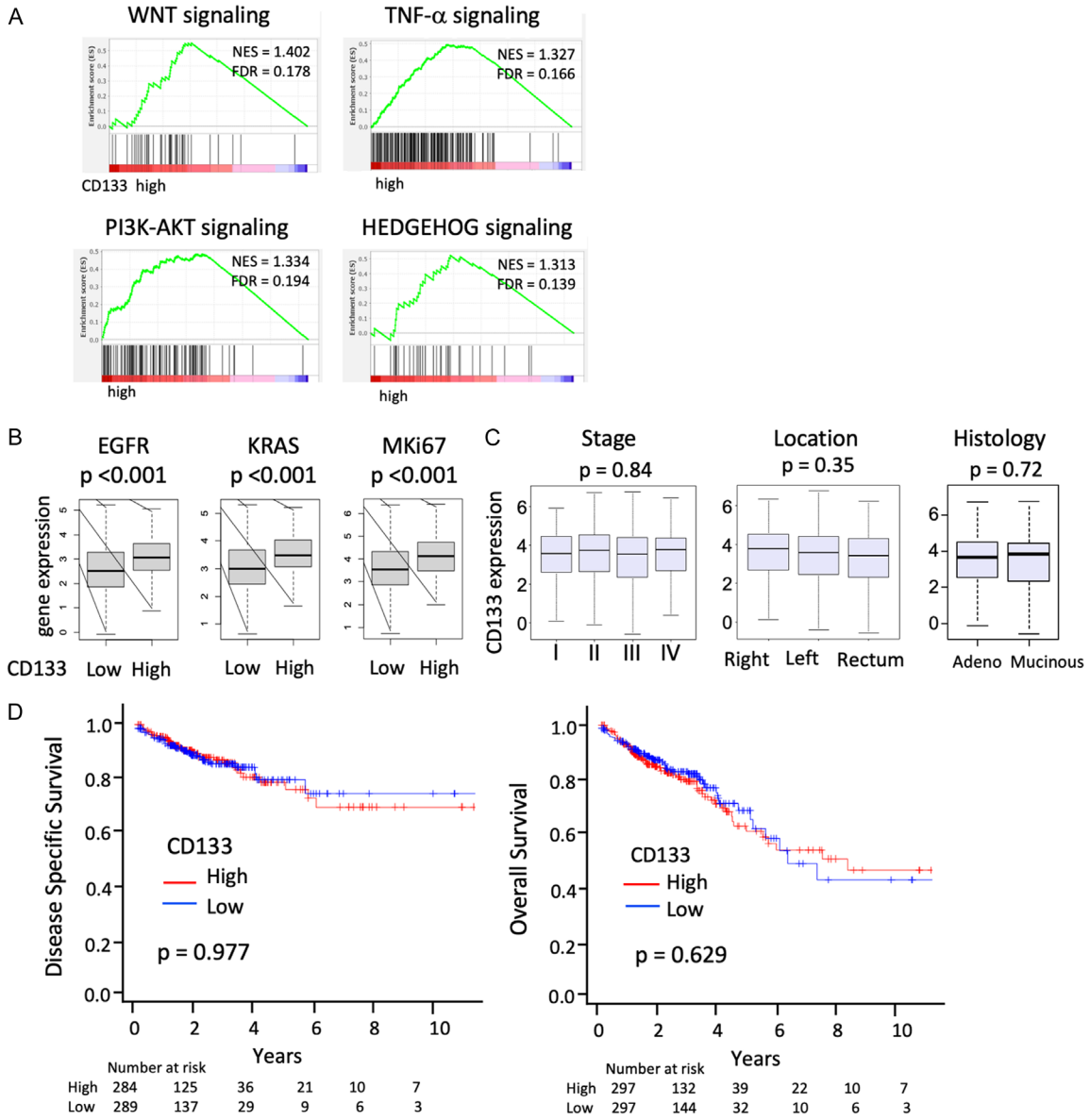


Figure 1. Molecular biological and clinical features of CD133-high colorectal cancer (CRC) in the TCGA cohort. A. GSEA of stemness-related gene sets; WNT, TNF-α, PI3K-AKT, and Hedgehog signaling. NES and FDR were determined with the classical GSEA method, where FDR<0.25 is considered significant. B. Expression levels of CRC treatment related genes, EGFR, KRAS, and MKi67. C. CD133 gene expression levels by AJCC cancer stage, location of cancer, and histology of CRC. D. Kaplan-Meier plots of disease-specific survival and overall survival by expressions of CD133 high (red line) and low (blue line) in the TCGA cohort are demonstrated. Median cut-off was used to divide two groups. Log-rank test was used to calculate *P* value. AJCC, American Joint Committee on Cancer; GSEA, gene set enrichment analysis; FDR, False Discovery Rate; NES, normalized enrichment score.

Mutational analysis

Given that up to 40% of CRC tumors have a KRAS mutation and 60% have an APC mutation [43], we also investigated the relationship between mutational status and CSC marker expressions. We found that there was significant increase in expression of CD133 for tumors with KRAS gain-of-function mutation, but not in tumors with APC loss-of-function mutation or mutant BRAF (**Figure 4**). Consistent with previous reports [44], there were significant increase in expression of CD24 and CD44 in mutation of APC, KRAS and the mutant

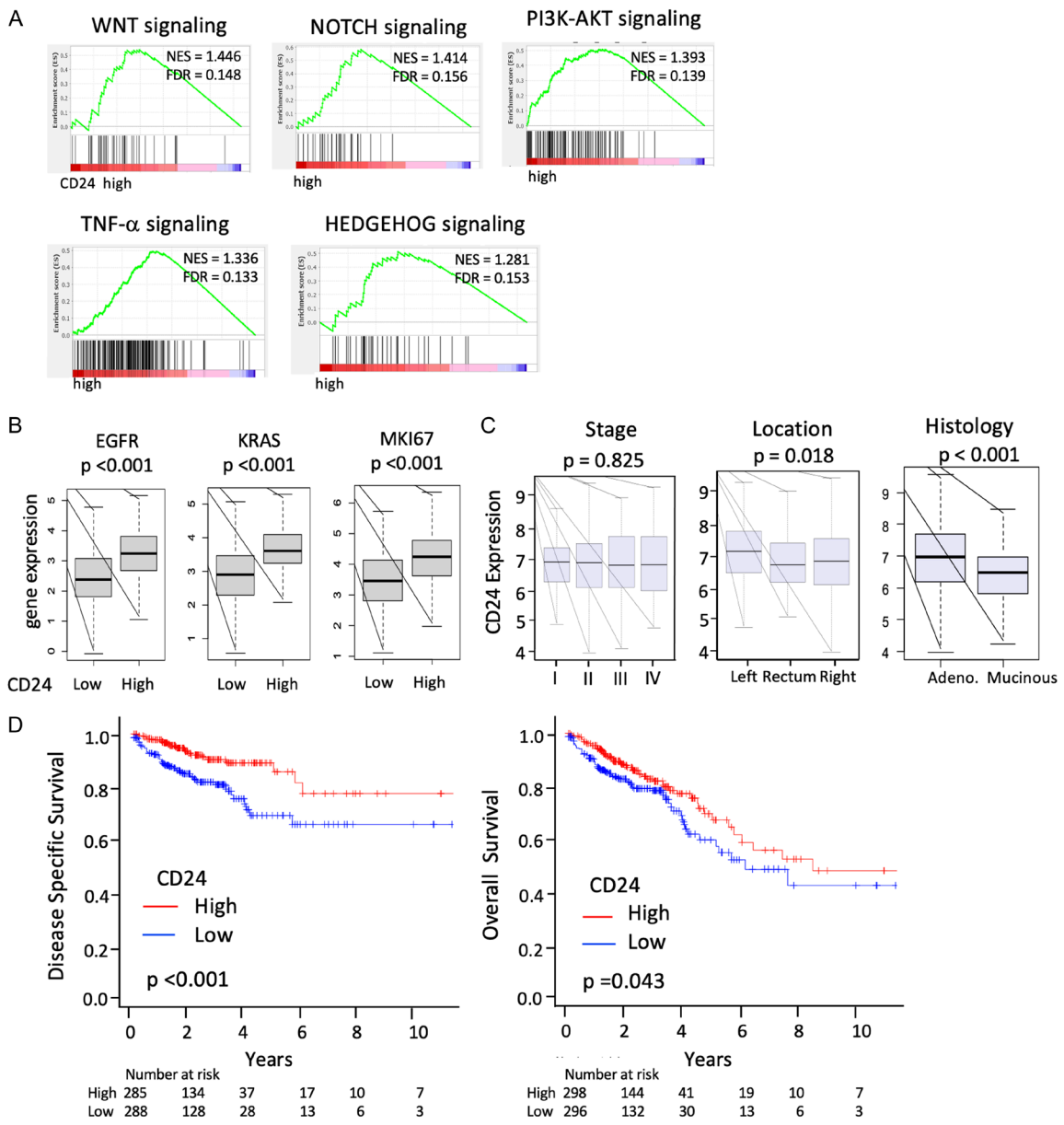


Figure 2. Molecular biological and clinical features of CD24-high CRC in the TCGA cohort. A. GSEA of stemness-related gene sets; WNT, NOTCH, TNF-α, PI3K-AKT, and Hedgehog signaling. NES and FDR were determined with the classical GSEA method, where FDR<0.25 is considered significant. B. Expression levels of CRC treatment related genes, EGFR, KRAS, and MKi67. C. CD24 gene expression levels by AJCC cancer stage, location of cancer, and histology of CRC. D. Kaplan-Meier plots of disease-specific survival and overall survival by expressions of CD24 high (red line) and low (blue line) in the TCGA cohort are demonstrated. Median cut-off was used to divide two groups. Log-rank test was used to calculate *P* value. AJCC, American Joint Committee on Cancer; GSEA, gene set enrichment analysis; FDR, False Discovery Rate; NES, normalized enrichment score.

downstream effector of the Ras-Raf-MEK-ERK pathway, BRAF (**Figure 4**). This demonstrates the potential interaction between germline or sporadic mutations of APC, KRAS and BRAF and aberrant stemness behavior in a large cohort.

Patients with CD44-high/CD24-low tumors have worse disease-specific survival and more likely to have recurrence

With evidence from literature that CSC profiles consistent of more than one stemness surface

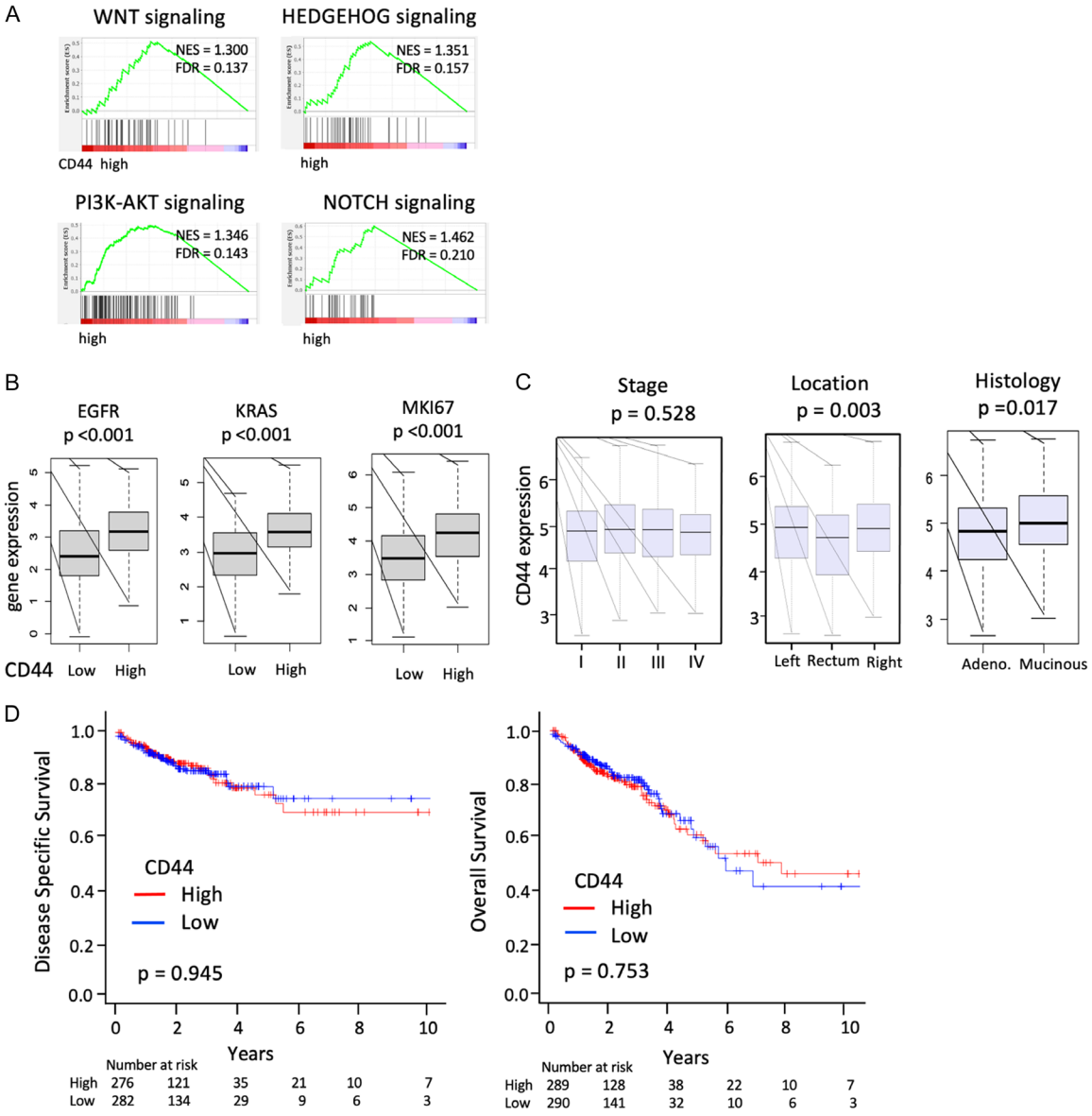


Figure 3. Molecular biological and clinical features of CD44-high CRC in the TCGA cohort. A. GSEA of stemness-related gene sets; WNT, Hedgehog, PI3K-AKT, and NOTCH signaling. NES and FDR were determined with the classical GSEA method, where FDR<0.25 is considered significant. B. Expression levels of CRC treatment related genes, EGFR, KRAS, and MKi67. C. CD44 gene expression levels by AJCC cancer stage, location of cancer, and histology of CRC. D. Kaplan-Meier plots of disease-specific survival and overall survival by expressions of CD44 high (red line) and low (blue line) in the TCGA cohort are demonstrated. Median cut-off was used to divide two groups. Log-rank test was used to calculate *P* value. AJCC, American Joint Committee on Cancer; GSEA, gene set enrichment analysis; FDR, False Discovery Rate; NES, normalized enrichment score.

marker are tumorigenic, we hypothesized that tumors with a combination of putative CSC would be prognostic. Studies suggest that tumors that are CD133/44 high may be prognostic [45]. However, CD133-high/CD44-low tumor was not associated with DSS, OS or disease recurrence in TCGA cohort (Figure S3).

On the other hand, CD44-high/CD24-low CRC tumors were significantly associated with worse DSS (P = 0.005, **Figure 5A**), and disease recurrence (P = 0.016, **Figure 5B**), but not with OS (P = 0.155, **Figure 5A**). However, GSEA did not show significant enrichment of any gene sets for CD44-high/CD24-low CRC (data not shown).

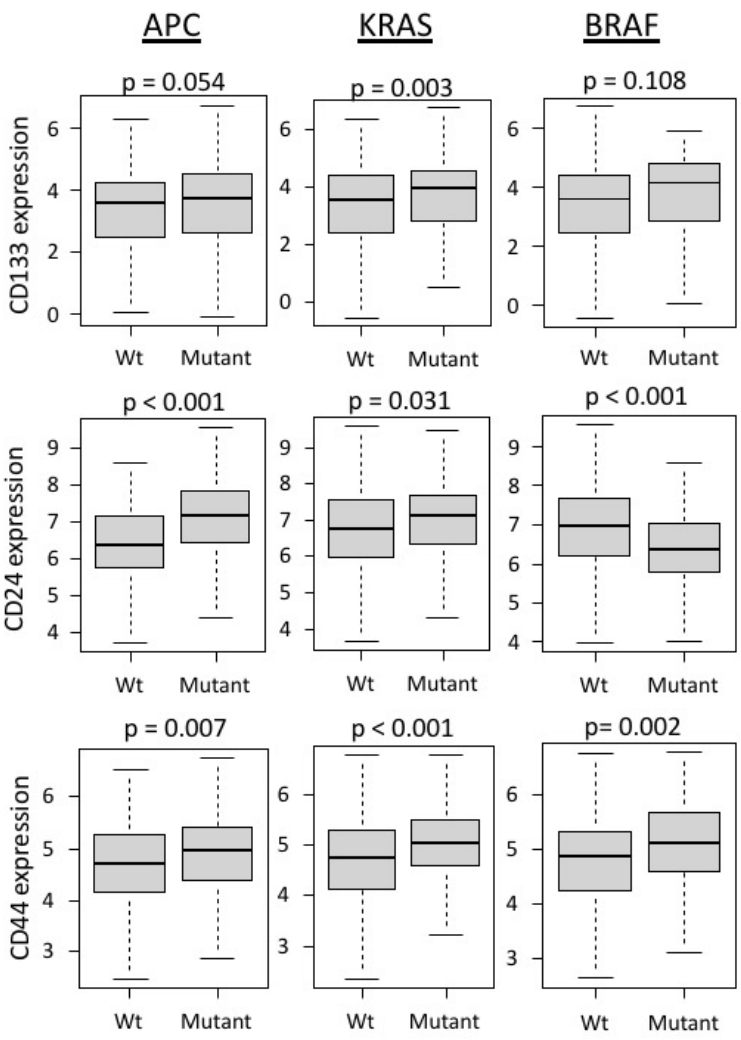


Figure 4. Gene expression levels of cancer stem cell surface markers, CD133, CD24, CD44, by mutation status of APC, KREAS and BRAF genes. A. Boxplots of gene expression levels of cancer stem cell surface markers, CD133, CD24, CD44, by wild type (Wt) or mutation of Adenomatous polyposis coli (APC), KRAS, BRAF genes. Mann-Whitney U test was used to calculate *P* values.

CD133-high/CD24-low tumors are associated with worse survival and recurrence

Previous small studies have also suggested a role of CD133/24-high tumors in predicting recurrence [12, 45, 46]. The GSEA showed negative enrichment of stemness signaling pathways (Figure S4). However, CD133-high/CD24-low tumors demonstrated significant negative enrichment of all the Hallmark cell proliferation-related gene sets: MYC targets (NES-1.71, FDR 0.04), E2F targets (NES -1.63, FDR 0.10), G2M checkpoint pathway (NES -1.55, FDR 0.06), and mitotic spindle gene sets (NES -1.42, FDR 0.12, **Figure 6A**). This suggests lower cell proliferation in these tumors, which is one of the phenotypes of stemness. Interestingly, CD133-high/CD24-low tumors were also negatively enriched for genes related to DNA repair (NES -1.56, FDR 0.06, **Figure 6A**), suggesting possible link with impaired DNA repair mechanisms. Additionally, patients with CD133-high/CD24-low tumors had worse DSS (P = 0.004) and OS (P = 0.04, **Figure 6B**), after a 10-year follow up. The comparison between patients with CD133-high/CD24-low tumors and all other groups is shown in Figure S5. Although comparison of CD44-high/CD24-low CRCs (n = 107) in **Figure 5** and CD133-high/CD24-low CRCs (n = 104) in **Figure 6** would be of interest, it would not be statistically appropriate since they were different grouping of the same cohort and approximately half of patients (n = 54) were included in both groups. On the other hand, in the subpopulation of patients with recurrence, patients with CD133-high/CD24-low tumors tended to have early recurrence within 3 years after initial treatment as well as late recurrence, compared to patients with CD133/low-CD24-high tumors (P = 0.01, **Figure 6C**).

Discussion

In this study, we investigated the link of recognized CSC surface markers to clinically relevant outcomes of colorectal cancer using a large cohort of patients. *In vitro* studies have found that cells that overexpress either CD133, CD24 or CD44 can independently initiate and sustain tumors that can recapitulate the histology of the parent tumor, which is one of the hallmarks of stemness [3, 4, 10]. Our analysis of gene set enrichment from a large number of patient specimen confirmed the enrichment of genes that are abnormally regulated in CSCs, such as Wnt/β-catenin pathway, Notch signaling,

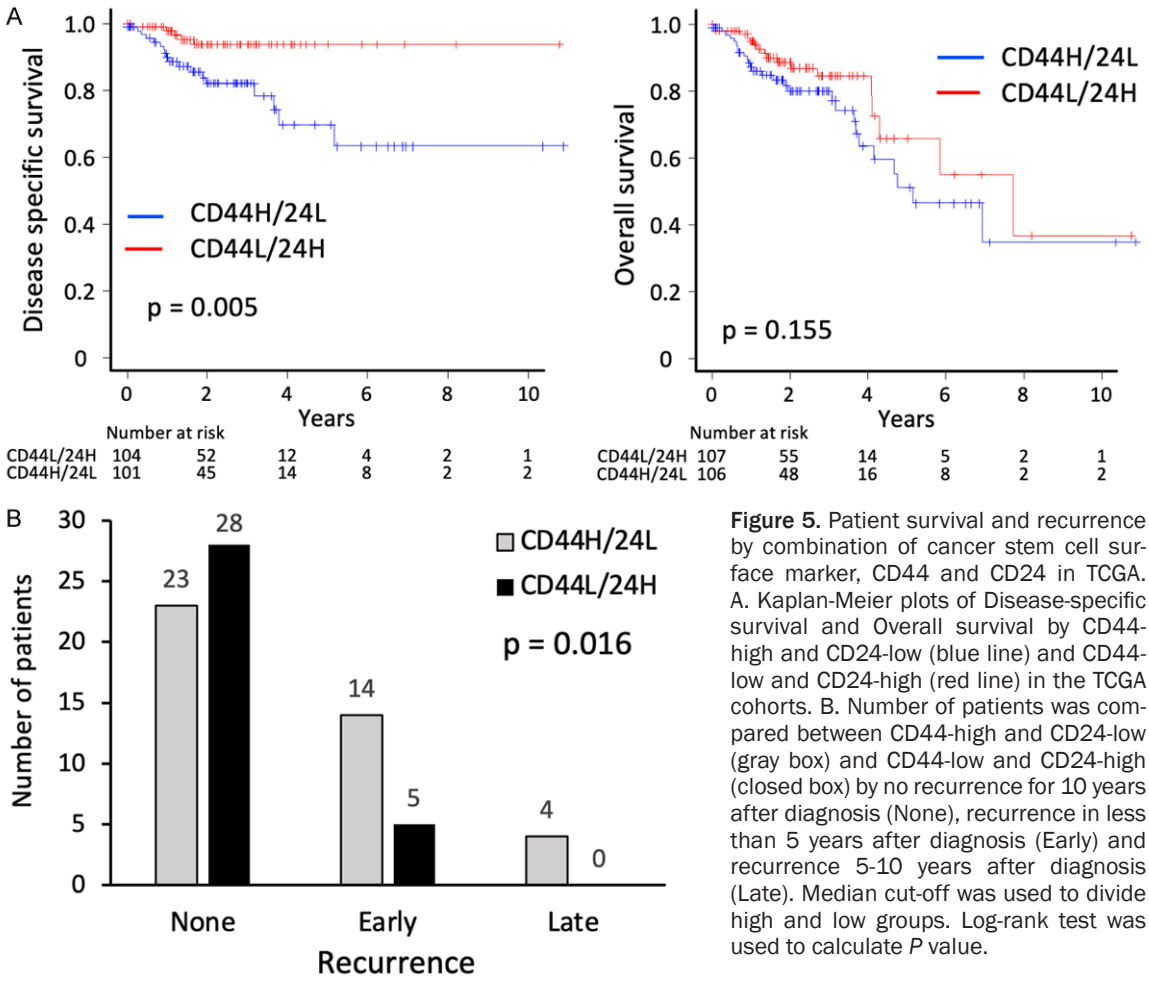


Figure 5. Patient survival and recurrence by combination of cancer stem cell surface marker, CD44 and CD24 in TCGA. A. Kaplan-Meier plots of Disease-specific survival and Overall survival by CD44-high and CD24-low (blue line) and CD44-low and CD24-high (red line) in the TCGA cohorts. B. Number of patients was compared between CD44-high and CD24-low (gray box) and CD44-low and CD24-high (closed box) by no recurrence for 10 years after diagnosis (None), recurrence in less than 5 years after diagnosis (Early) and recurrence 5-10 years after diagnosis (Late). Median cut-off was used to divide high and low groups. Log-rank test was used to calculate *P* value.

Hedgehog signaling and TNF-α via NFκB. Aberrancy in signaling of these pathways could induce CSC self-renewal, exit out of dormancy, dedifferentiation, inhibition of apoptosis, and increase proliferation through MYC [6]. Indeed, we showed that tumors that were CD133-high, CD24-high or CD44-high showed enrichment of proliferation-related gene sets when an enrichment of Wnt/β-catenin signaling was also present.

Interestingly, we showed that higher CD44 expression was found in mucinous carcinomas and higher CD24 expression was found in adenocarcinoma. Indeed, these findings were congruent with a recent study that linked the expression of mucin to CD44-expression colon cancer stem cells and their role in chemoresistance [47]. Furthermore, CD24 has been found in 90% of adenoma and 86% of malignant lesions and has been hypothesized to be involved in early adenoma to carcinoma transition [48].

Studies have shown that colon cancer with chemotherapy and radiation resistance, greater invasive potential, and progression to metastatic disease tend to have a subpopulation of CSCs that express different markers [12, 15, 45, 46, 49]. The CD133-high/CD24-low CRC in this study exhibited decrease in the enrichment of cell proliferation-related pathways, as well as early recurrence and poor survival at the same time. These observations initially may seem to conflict; however, it makes sense if these tumors are in dormancy with less proliferation that leads to less response to chemotherapy, greater resistance to apoptosis and therefore, have greater potential for tumor recurrence. This is supported by studies showing that while CD133 cells are only present in 2% of the bulk tumor in colorectal cancer, they tend to produce interleukin-4 that protect them from apoptosis [50]. In addition, CRC with increased CD133 expression when exposed to neoadjuvant chemotherapy and radiation tend to have greater residual tumor and lower tumor regres-

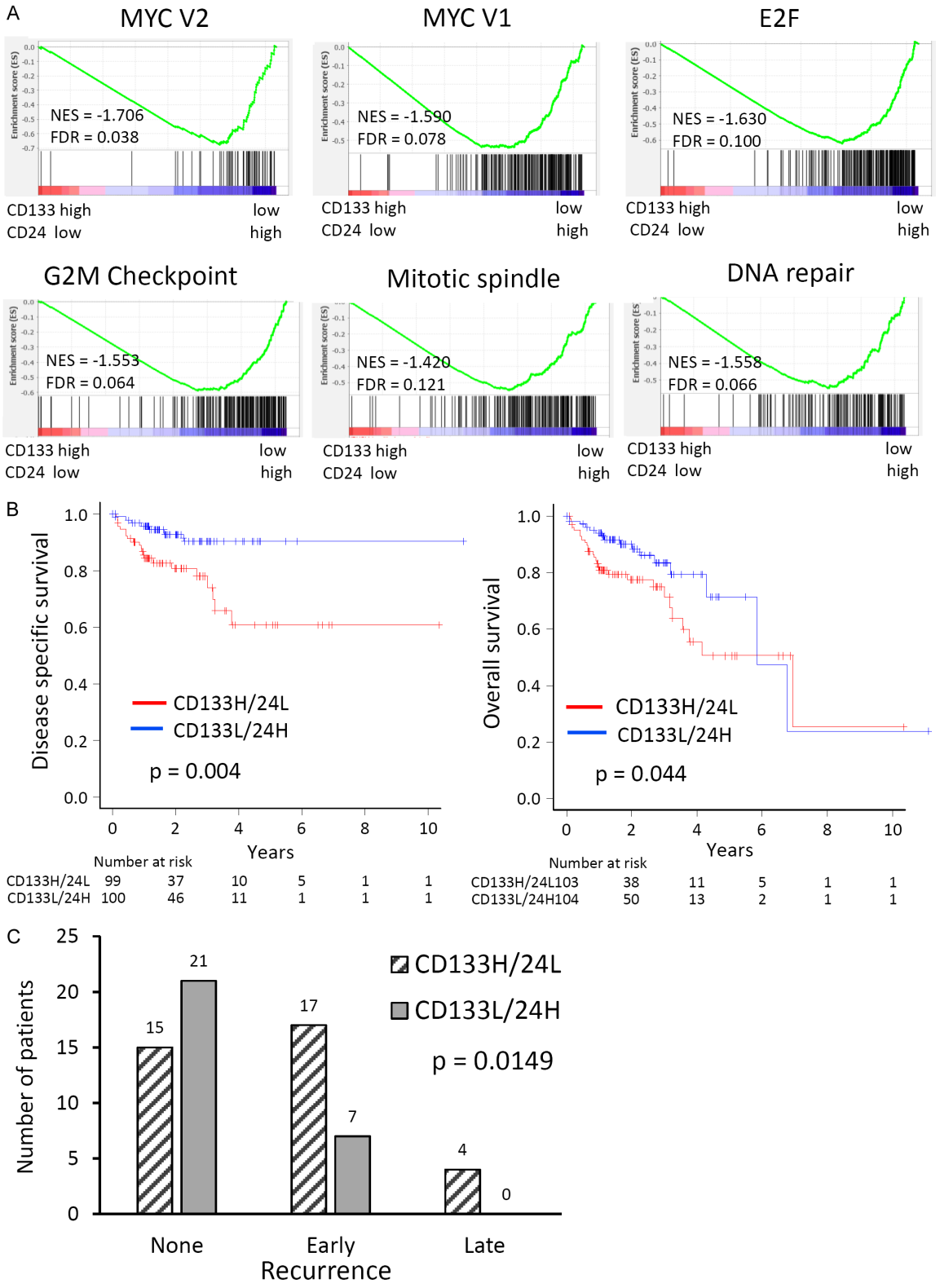


Figure 6. Gene set enrichment analysis (GSEA), patient survival and cancer recurrence by combination expression of CD133 and CD24 in the TCGA cohort. A. GSEA of all cell proliferation-related gene sets in Hallmark collection; MYC signaling version 1 and MYC signaling version 2, E2F targets, G2M checkpoint, Mitotic spindle, as well as DNA repair, were compared between CD133-high/CD24-low vs CD133-low/CD24-high. NES and FDR were determined with the classical GSEA method, where FDR<0.25 is considered significant. B. Kaplan-Meier plots of disease-specific survival and overall survival by expressions of CD133-high/CD24-low high (red line) and CD133-low/CD24-high

(blue line) in the TCGA cohort are demonstrated. C. Number of patients was compared between CD133-high/CD24-low (shaded box) and CD133-low/CD24-high (gray box) by no recurrence for 10 years after diagnosis (None), recurrence in less than 5 years after diagnosis (Early) and recurrence 5-10 years after diagnosis (Late). Median cut-off was used to divide high vs low groups. Log-rank test was used to calculate *P* value.

sion in surgical specimens [14, 18]. Therefore, as systemic chemotherapy has been shown to exhibit greater efficacy on cells that rapidly divide, the negative enrichment of CD133-high/CD24-low tumors to proliferation pathways may indicate poorer response to chemotherapy. Indeed, the association between high CD133 expression with KRAS mutation also suggest worse response or even insensitivity to certain targeted chemotherapy agents, as had previously been demonstrated for cetuximab and panitumumab on tumors with KRAS mutation [51].

Several mechanisms regarding CSC could have contributed to our finding that CD133-high/CD24-low is not a stemness marker. CSCs are known for their plasticity, defined as the ability to transition from states of proliferation and quiescence as well as states of stem and nonstemness, depending on the tumor microenvironment and possibly genetic alternation [5]. Therefore, one snapshot in time as represented by the samples available in TCGA may limit the representation of the effect of stemness state on tumor and the disease process over time. Stem cells and nonstem cells may both lead to tumor development, as studies have previously demonstrated [5, 52, 53].

Our study also did not show that there was a survival difference in CD133-high tumors or CD44-high tumors. However, patients with CD24-high tumors showed significantly better disease-specific survival and overall survival, compared to patients with CD24-low tumors. The multifactorial regulation of CSCs may contribute to these findings. Stochastic factors involving sporadic or germline gene mutations have been shown to regulate stemness properties. Moon et al. demonstrated that in mice with the background of APC loss-of-function, KRAS mutations can activate the Wnt/β-catenin pathway, leading to increase proliferation and likelihood of metastatic disease [54]. Indeed, we showed that mutant KRAS is independently associated with increased expression of CSCs, CD133, CD24 and CD44. Recent *in vitro* reports have also shown the increasing role of epigenetic regulation that affect the stemness phenotype. Tao et al. showed that DNA hypermethylation contributes to the stemness phenotype in the background of BRAFV600E mutation, promoting formation of tumor organoids [55]. Cell lines have demonstrated greater CD133 expression with decreased methylation of CpG islands, and histone H3 acetylation is also involved with regulation of CD44 expression [53]. With our analysis, since we are only able to assess the transcriptional products of these cancer stem cell markers, the impact of epigenetic silencing will require further investigation. Lastly, the tumor microenvironment has been shown to contribute an important role in maintaining stemness and tumorigenicity [56]. Vermeulen et al. demonstrated that tumor-associated myofibroblasts secrete hepatocyte growth factors that stimulate the nearby colon cancer cells to maintain stemness through the Wnt/β-catenin pathway [57]. More recently, the same group demonstrated that functional stem cells that drive tumor growth appear to reside at the edge rather than the center of tumors, and that they do not necessarily express traditional stem cell markers such as CD133. This demonstrated the importance of spatial orientation of cells within the bulk tumor in promoting stemness phenotypes such as chemotherapy resistance and tumor growth [58]. In this respect, CSCs may rely on the optimal tumor microenvironment to be able obtain functional stemness phenotypes that would affect their impact on tumor growth.

Since this is fundamentally a retrospective study, there are several relevant limitations. The data was based on patient samples made available in a public domain. Therefore, the analysis relied on the information that had already been catalogued in the TCGA, which has limited data granularity. In addition, although the spatial relationship of the cancer stem cell relative to the bulk tumor may be important, the location from which the patient sample was derived may be variable from patient to patient. Cancer stem cells that are present at the periphery of the tumor may not have been sampled, which may result in an underrepresentation of the full array and functionality of the CSCs. Nonetheless, this bioinformatics report is one of the first that used a large cohort of patients from the TCGA to link CSC markers to patient outcomes.

Conclusion

In conclusion, our bioinformatics analysis showed that CD133, CD24 and CD44 were individually associated with cell proliferation. This did not translate to a difference in overall survival. However, mucinous histology was significantly more associated with higher CD44 while adenocarcinomas were associated with higher CD24. The combination of CD133-high/CD24-low characterization is associated with poorer prognosis and greater recurrence. This may be due to decreased proliferation and absence of stemness features associated with CD133-high/CD24-low tumors. Further studies will be needed to understand the interactions between these CSC surface markers and their impact on tumor formation.

Acknowledgements

This work was supported by US National Institutes of Health/National Cancer Institute grant R01CA160688, R01CA250412, R37CA248018, US Department of Defense BCRP grant W81XWH-19-1-0674, as well as the Edward K. Duch Foundation and Paul & Helen Ellis Charitable Trust to K.T., and US National Cancer Institute cancer center support grant P30-CA016056 to Roswell Park Comprehensive Cancer Center.

Disclosure of conflict of interest

None.

Abbreviations

CSC, cancer stem cell; TCGA, The Cancer Genome Atlas; GSEA, Gene set enrichment analysis; DSS, Disease-specific survival; OS, Overall survival.

Address correspondence to: Dr. Kazuaki Takabe, Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Elm & Carlton Streets, Buffalo, New York 14263, USA. Tel: 1-619-845-5128; E-mail: Kazuaki.Takabe@RoswellPark.org

References

[1] Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL and Siegel RL. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019; 69: 363-385.

[2] Tamas K, Walenkamp AM, de Vries EG, van Vugt MA, Beets-Tan RG, van Etten B, de Groot DJ and Hospers GA. Rectal and colon cancer: not just a different anatomic site. Cancer Treat Rev 2015; 41: 671-679.

[3] O'Brien CA, Pollett A, Gallinger S and Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 2007; 445: 106-110.

[4] Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C and De Maria R. Identification and expansion of human colon-cancer-initiating cells. Nature 2007; 445: 111-115.

[5] Zeuner A, Todaro M, Stassi G and De Maria R. Colorectal cancer stem cells: from the crypt to the clinic. Cell Stem Cell 2014; 15: 692-705.

[6] Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, Zhang G, Wang X, Dong Z, Chen F and Cui H. Targeting cancer stem cell pathways for cancer therapy. Signal Transduct Target Ther 2020; 5: 8.

[7] Zhu L, Gibson P, Currle DS, Tong Y, Richardson RJ, Bayazitov IT, Poppleton H, Zakharenko S, Ellison DW and Gilbertson RJ. Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. Nature 2009; 457: 603-607.

[8] Todaro M, Gaggianesi M, Catalano V, Benfante A, Iovino F, Biffoni M, Apuzzo T, Sperduti I, Volpe S, Cocorullo G, Gulotta G, Dieli F, De Maria R and Stassi G. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell 2014; 14: 342-356.

[9] de Sousa e Melo F, Kurtova AV, Harnoss JM, Kljavin N, Hoeck JD, Hung J, Anderson JE, Storm EE, Modrusan Z, Koeppen H, Dijkgraaf GJ, Piskol R and de Sauvage FJ. A distinct role for Lgr5(+) stem cells in primary and metastatic colon cancer. Nature 2017; 543: 676-680.

[10] Du L, Wang H, He L, Zhang J, Ni B, Wang X, Jin H, Cahuzac N, Mehrpour M, Lu Y and Chen Q. CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res 2008; 14: 6751-6760.

[11] Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea M, Richel DJ, Stassi G and Medema JP. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. Proc Natl Acad Sci U S A 2008; 105: 13427-13432.

[12] Choi D, Lee HW, Hur KY, Kim JJ, Park GS, Jang SH, Song YS, Jang KS and Paik SS. Cancer stem cell markers CD133 and CD24 correlate with invasiveness and differentiation in colorectal adenocarcinoma. World J Gastroenterol 2009; 15: 2258-2264.

[13] Chen S, Song X, Chen Z, Li X, Li M, Liu H and Li J. CD133 expression and the prognosis of

colorectal cancer: a systematic review and meta-analysis. PLoS One 2013; 8: e56380.
[14] Jao SW, Chen SF, Lin YS, Chang YC, Lee TY, Wu CC, Jin JS and Nieh S. Cytoplasmic CD133 expression is a reliable prognostic indicator of tumor regression after neoadjuvant concurrent chemoradiotherapy in patients with rectal cancer. Ann Surg Oncol 2012; 19: 3432-3440.
[15] Jing F, Kim HJ, Kim CH, Kim YJ, Lee JH and Kim HR. Colon cancer stem cell markers CD44 and CD133 in patients with colorectal cancer and synchronous hepatic metastases. Int J Oncol 2015; 46: 1582-1588.
[16] Lin EH, Hassan M, Li Y, Zhao H, Nooka A, Sorenson E, Xie K, Champlin R, Wu X and Li D. Elevated circulating endothelial progenitor marker CD133 messenger RNA levels predict colon cancer recurrence. Cancer 2007; 110: 534-542.
[17] Iinuma H, Watanabe T, Mimori K, Adachi M, Hayashi N, Tamura J, Matsuda K, Fukushima R, Okinaga K, Sasako M and Mori M. Clinical significance of circulating tumor cells, including cancer stem-like cells, in peripheral blood for recurrence and prognosis in patients with Dukes' stage B and C colorectal cancer. J Clin Oncol 2011; 29: 1547-1555.
[18] Sprenger T, Conradi LC, Beissbarth T, Ermert H, Homayounfar K, Middel P, Ruschoff J, Wolff HA, Schuler P, Ghadimi BM, Rodel C, Becker H, Rodel F and Liersch T. Enrichment of CD133-expressing cells in rectal cancers treated with preoperative radiochemotherapy is an independent marker for metastasis and survival. Cancer 2013; 119: 26-35.
[19] Wang Q, Chen ZG, Du CZ, Wang HW, Yan L and Gu J. Cancer stem cell marker CD133+ tumour cells and clinical outcome in rectal cancer. Histopathology 2009; 55: 284-293.
[20] Weichert W, Denkert C, Burkhardt M, Gansukh T, Bellach J, Altevogt P, Dietel M and Kristiansen G. Cytoplasmic CD24 expression in colorectal cancer independently correlates with shortened patient survival. Clin Cancer Res 2005; 11: 6574-6581.
[21] Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Matsuyama R, Endo I and Takabe K. G2M cell cycle pathway score as a prognostic biomarker of metastasis in estrogen receptor (ER)-positive breast cancer. Int J Mol Sci 2020; 21: 2921.
[22] Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Nagahashi M, Matsuyama R, Endo I and Takabe K. The E2F pathway score as a predictive biomarker of response to neoadjuvant therapy in ER+/HER2- breast cancer. Cells 2020; 9: 1643.
[23] Takahashi H, Katsuta E, Yan L, Dasgupta S and Takabe K. High expression of Annexin A2 is associated with DNA repair, metabolic alteration, and worse survival in pancreatic ductal adenocarcinoma. Surgery 2019; 166: 150-156.
[24] Angell HK, Bruni D, Barrett JC, Herbst R and Galon J. The immunoscore: colon cancer and beyond. Clin Cancer Res 2020; 26: 332-339.
[25] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.
[26] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP and Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015; 1: 417-425.
[27] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Katz MHG and Takabe K. High G2M pathway score pancreatic cancer is associated with worse survival, particularly after margin-positive (R1 or R2) resection. Cancers (Basel) 2020; 12: 2871.
[28] Schulze A, Oshi M, Endo I and Takabe K. MYC targets scores are associated with cancer aggressiveness and poor survival in ER-positive primary and metastatic breast cancer. Int J Mol Sci 2020; 21: 8127.
[29] Katsuta E, Yan L, Takeshita T, McDonald KA, Dasgupta S, Opyrchal M and Takabe K. High MYC mRNA expression is more clinically relevant than MYC DNA amplification in triple-negative breast cancer. Int J Mol Sci 2019; 21: 217.
[30] Oshi M, Gandhi S, Angarita FA, Kim TH, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. A novel five-gene score to predict complete pathological response to neoadjuvant chemotherapy in ER-positive/HER2-negative breast cancer. Am J Cancer Res 2021; 11: 3611-3627.
[31] Tokumaru Y, Oshi M, Katsuta E, Yan L, Huang JL, Nagahashi M, Matsuhashi N, Futamura M, Yoshida K and Takabe K. Intratumoral adipocyte-high breast cancer enrich for metastatic and inflammation-related pathways but associated with less cancer cell proliferation. Int J Mol Sci 2020; 21: 5744.
[32] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Nagahashi M and Takabe K. Intra-tumoral angiogenesis is associated with inflammation, immune reaction and metastatic recurrence in breast cancer. Int J Mol Sci 2020; 21: 6708.
[33] Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, Endo I and Takabe K. CD8 T cell score as a prognostic biomarker for triple negative breast cancer. Int J Mol Sci 2020; 21: 6968.

[34] Oshi M, Newman S, Murthy V, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. ITPKC as a prognostic and predictive biomarker of neoadjuvant chemotherapy for triple negative breast cancer. Cancers (Basel) 2020; 12: 2758.

[35] Oshi M, Tokumaru Y, Asaoka M, Yan L, Satyananda V, Matsuyama R, Matsuhashi N, Futamura M, Ishikawa T, Yoshida K, Endo I and Takabe K. M1 Macrophage and M1/M2 ratio defined by transcriptomic signatures resemble only part of their conventional clinical characteristics in breast cancer. Sci Rep 2020; 10: 16554.

[36] Oshi M, Asaoka M, Tokumaru Y, Angarita FA, Yan L, Matsuyama R, Zsiros E, Ishikawa T, Endo I and Takabe K. Abundance of regulatory T cell (Treg) as a predictive biomarker for neoadjuvant chemotherapy in triple-negative breast cancer. Cancers (Basel) 2020; 12: 3038.

[37] Le L, Tokumaru Y, Oshi M, Asaoka M, Yan L, Endo I, Ishikawa T, Futamura M, Yoshida K and Takabe K. Th2 cell infiltrations predict neoadjuvant chemotherapy response of estrogen receptor-positive breast cancer. Gland Surg 2021; 10: 154-165.

[38] Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Kalinski P, Endo I and Takabe K. Plasmacytoid dendritic cell (pDC) infiltration correlate with tumor infiltrating lymphocytes, cancer immunity, and better survival in triple negative breast cancer (TNBC) more strongly than conventional dendritic cell (cDC). Cancers (Basel) 2020; 12: 3342.

[39] Oshi M, Huyser MR, Le L, Tokumaru Y, Yan L, Matsuyama R, Endo I and Takabe K. Abundance of microvascular endothelial cells is associated with response to chemotherapy and prognosis in colorectal cancer. Cancers (Basel) 2021; 13: 1477.

[40] Katsuta E, Rashid OM and Takabe K. Fibroblasts as a biological marker for curative resection in pancreatic ductal adenocarcinoma. Int J Mol Sci 2020; 21: 3890.

[41] Bruttel VS and Wischhusen J. Cancer stem cell immunology: key to understanding tumorigenesis and tumor immune escape? Front Immunol 2014; 5: 360.

[42] Miller TJ, McCoy MJ, Hemmings C, Bulsara MK, Iacopetta B and Platell CF. The prognostic value of cancer stem-like cell markers SOX2 and CD133 in stage III colon cancer is modified by expression of the immune-related markers FoxP3, PD-L1 and CD3. Pathology 2017; 49: 721-730.

[43] Janssen KP, Alberici P, Fsihi H, Gaspar C, Breukel C, Franken P, Rosty C, Abal M, El Marjou F, Smits R, Louvard D, Fodde R and Robine S. APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. Gastroenterology 2006; 131: 1096-1109.

[44] Kemper K, Versloot M, Cameron K, Colak S, de Sousa e Melo F, de Jong JH, Bleackley J, Vermeulen L, Versteeg R, Koster J and Medema JP. Mutations in the Ras-Raf axis underlie the prognostic value of CD133 in colorectal cancer. Clin Cancer Res 2012; 18: 3132-3141.

[45] Sahlberg SH, Spiegelberg D, Glimelius B, Stenerlow B and Nestor M. Evaluation of cancer stem cell markers CD133, CD44, CD24: association with AKT isoforms and radiation resistance in colon cancer cells. PLoS One 2014; 9: e94621.

[46] Paschall AV, Yang D, Lu C, Redd PS, Choi JH, Heaton CM, Lee JR, Nayak-Kapoor A and Liu K. CD133+CD24lo defines a 5-Fluorouracil-resistant colon cancer stem cell-like phenotype. Oncotarget 2016; 7: 78698-78712.

[47] Pothuraju R, Rachagani S, Krishn SR, Chaudhary S, Nimmakayala RK, Siddiqui JA, Ganguly K, Lakshmanan I, Cox JL, Mallya K, Kaur S and Batra SK. Molecular implications of MUC5AC-CD44 axis in colorectal cancer progression and chemoresistance. Mol Cancer 2020; 19: 37.

[48] Sagiv E, Memeo L, Karin A, Kazanov D, Jacob-Hirsch J, Mansukhani M, Rechavi G, Hibshoosh H and Arber N. CD24 is a new oncogene, early at the multistep process of colorectal cancer carcinogenesis. Gastroenterology 2006; 131: 630-639.

[49] Wielenga VJ, Smits R, Korinek V, Smit L, Kielman M, Fodde R, Clevers H and Pals ST. Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. Am J Pathol 1999; 154: 515-523.

[50] Todaro M, Alea MP, Di Stefano AB, Cammareri P, Vermeulen L, Iovino F, Tripodo C, Russo A, Gulotta G, Medema JP and Stassi G. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. Cell Stem Cell 2007; 1: 389-402.

[51] Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD and Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol 2008; 26: 1626-1634.

[52] Shmelkov SV, Butler JM, Hooper AT, Hormigo A, Kushner J, Milde T, St Clair R, Baljevic M, White I, Jin DK, Chadburn A, Murphy AJ, Valenzuela DM, Gale NW, Thurston G, Yancopoulos GD, D'Angelica M, Kemeny N, Lyden D and Rafii S. CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. J Clin Invest 2008; 118: 2111-2120.

[53] Vincent A, Ouelkdite-Oumouchal A, Souidi M, Leclerc J, Neve B and Van Seuningen I. Colon cancer stemness as a reversible epigenetic state: implications for anticancer therapies. World J Stem Cells 2019; 11: 920-936.
[54] Moon BS, Jeong WJ, Park J, Kim TI, Min do S and Choi KY. Role of oncogenic K-Ras in cancer stem cell activation by aberrant Wnt/beta-catenin signaling. J Natl Cancer Inst 2014; 106: djt373.
[55] Tao Y, Kang B, Petkovich DA, Bhandari YR, In J, Stein-O'Brien G, Kong X, Xie W, Zachos N, Maegawa S, Vaidya H, Brown S, Chiu Yen RW, Shao X, Thakor J, Lu Z, Cai Y, Zhang Y, Mallona I, Peinado MA, Zahnow CA, Ahuja N, Fertig E, Issa JP, Baylin SB and Easwaran H. Aging-like spontaneous epigenetic silencing facilitates Wnt activation, stemness, and braf(V600E)-induced tumorigenesis. Cancer Cell 2019; 35: 315-328, e316.
[56] Medema JP and Vermeulen L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. Nature 2011; 474: 318-326.
[57] Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T, Tuynman JB, Todaro M, Merz C, Rodermond H, Sprick MR, Kemper K, Richel DJ, Stassi G and Medema JP. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol 2010; 12: 468-476.
[58] Lenos KJ, Miedema DM, Lodestijn SC, Nijman LE, van den Bosch T, Romero Ros X, Lourenço FC, Lecca MC, van der Heijden M, van Neerven SM, van Oort A, Leveille N, Adam RS, de Sousa E Melo F, Otten J, Veerman P, Hypolite G, Koens L, Lyons SK, Stassi G, Winton DJ, Medema JP, Morrissey E, Bijlsma MF and Vermeulen L. Stem cell functionality is microenvironmentally defined during tumour expansion and therapy response in colon cancer. Nat Cell Biol 2018; 20: 1193-1202.

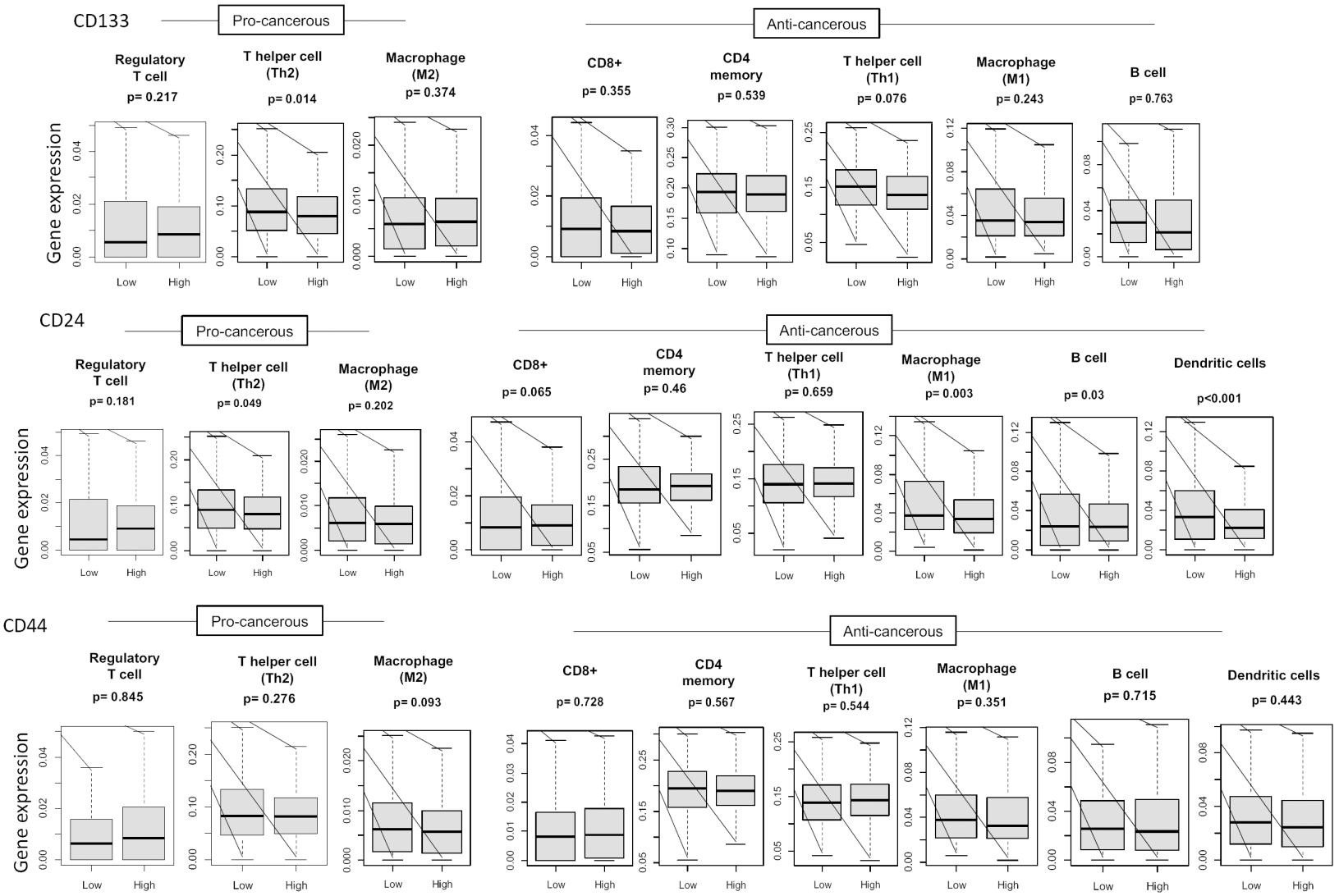


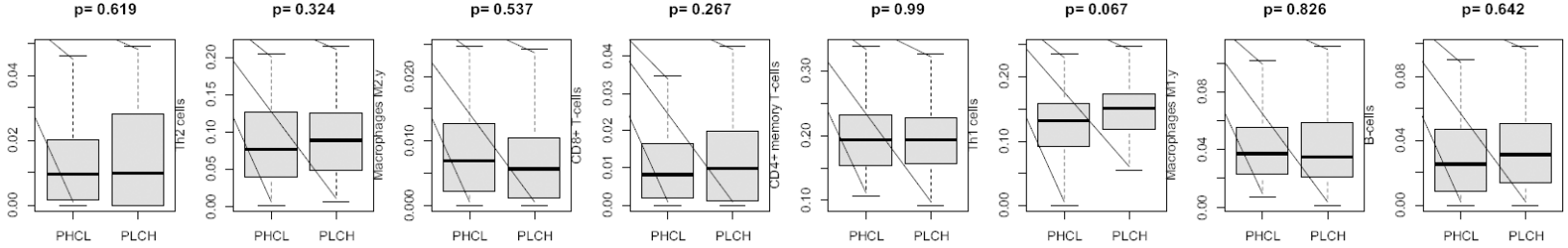
Figure S1. Pro-cancerous and Anti-cancerous immune cell infiltrations by cancer stem cell surface marker gene expressions, CD133, CD24 and CD44. Amount of cell infiltration was estimated by xCell algorithm. Marker gene expression was determined high vs low by median cut-off.

CD133/CD44

PHCL: CD133-high/CD44-low
PLCH: CD133-low/CD44-high

Pro-cancerous | Anti-cancerous

Regulatory T cell	T helper cell (Th2)	Macrophage (M2)	CD8+	CD4 memory	T helper cell (Th1)	Macrophage (M1)	B cell
p= 0.619	p= 0.324	p= 0.537	p= 0.267	p= 0.99	p= 0.067	p= 0.826	p= 0.642

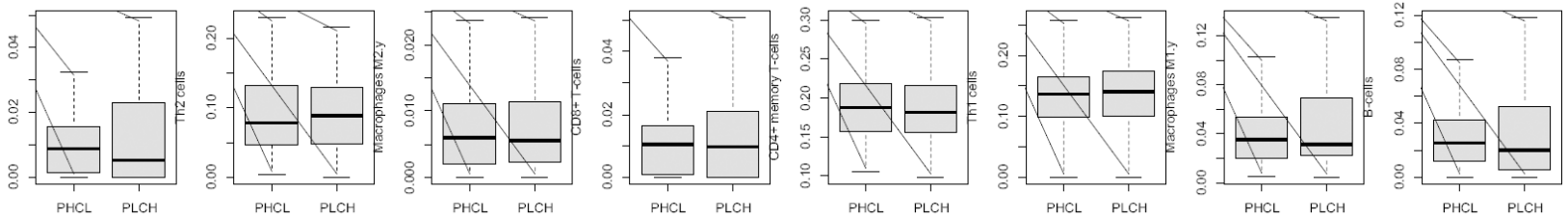


CD24/CD44

PHCL: CD24-high/CD44-low
PLCH: CD24-low/CD44-high

Pro-cancerous | Anti-cancerous

Regulatory T cell	T helper cell (Th2)	Macrophage (M2)	CD8+	CD4 memory	T helper cell (Th1)	Macrophage (M1)	B cell
p= 0.512	p= 0.55	p= 0.778	p= 0.19	p= 0.924	p= 0.947	p= 0.137	p= 0.164



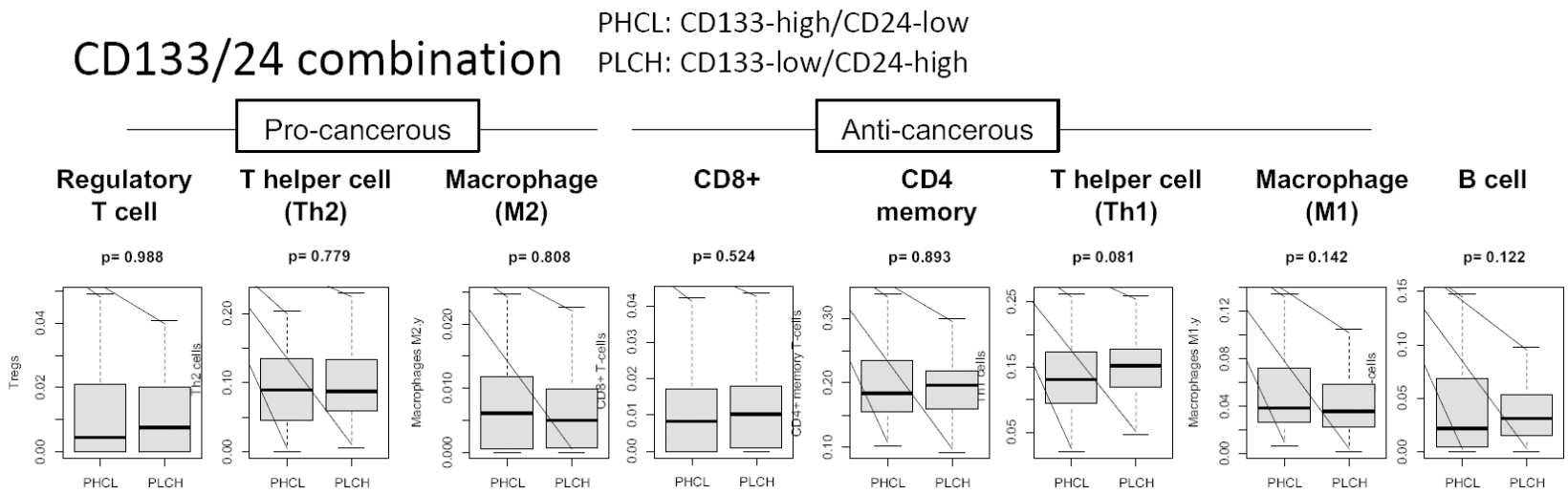


Figure S2. Pro-cancerous and Anti-cancerous immune cell infiltrations by combination of expressions of cancer stem cell surface markers, CD133/CD44, CD24/CD44, and CD133/CD24. Amount of cell infiltration was estimated by xCell algorithm. Marker gene expression was determined high vs low by median cut-off.

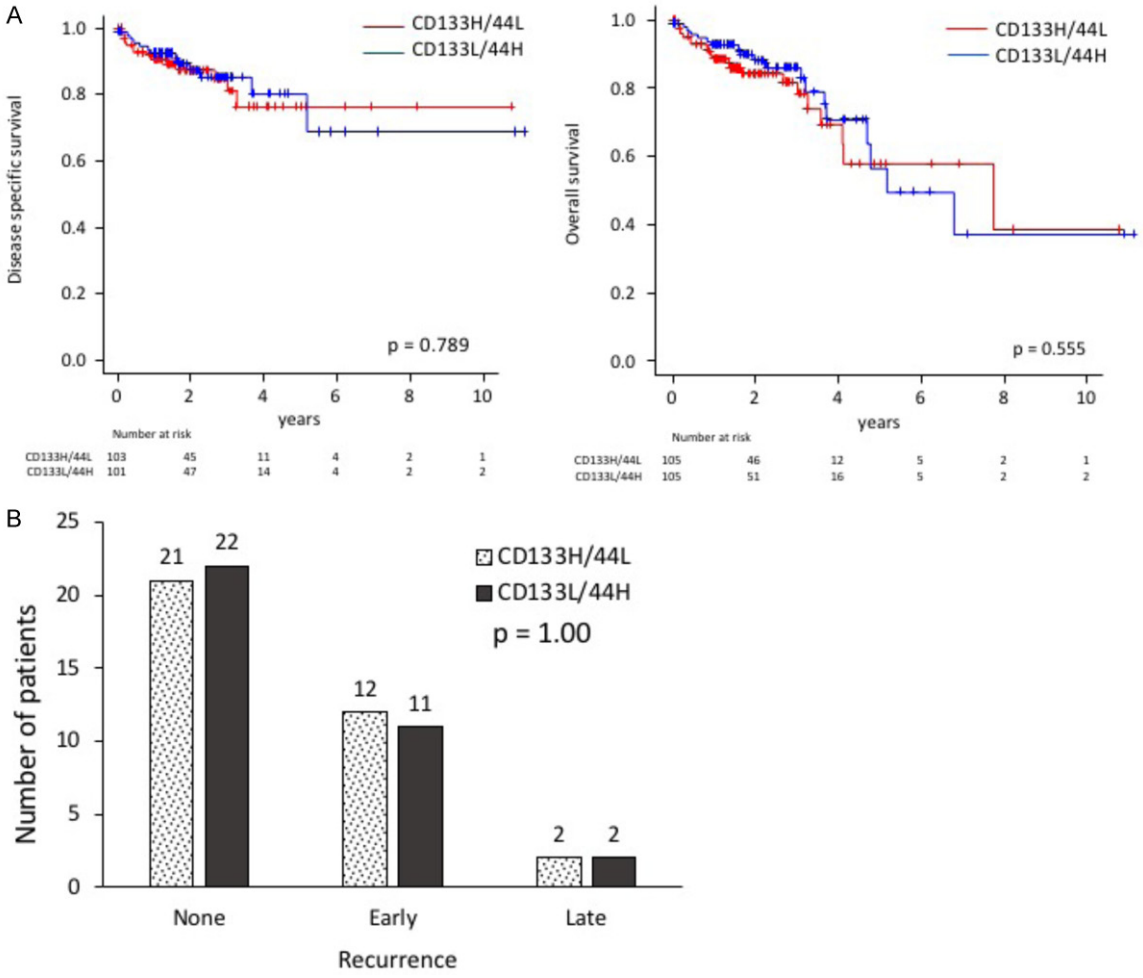


Figure S3. Survival analyses by combination of expressions of cancer stem cell surface markers, CD133/CD44. A. Kaplan-Meier plots of disease-specific survival and overall survival by expressions of CD133-high/CD44-low (red line) and CD133-low/CD44-high (blue line) in the TCGA cohort are demonstrated. B. Number of patients were compared between CD133-high/CD44-low (gray box) and CD133-low/CD44-high (closed box) by no recurrence for 10 years after diagnosis (None), recurrence in less than 5 years after diagnosis (Early) and recurrence 5-10 years after diagnosis (Late). Median cut-off was used to divide high vs low groups. Log-rank test was used to calculate *p* value.

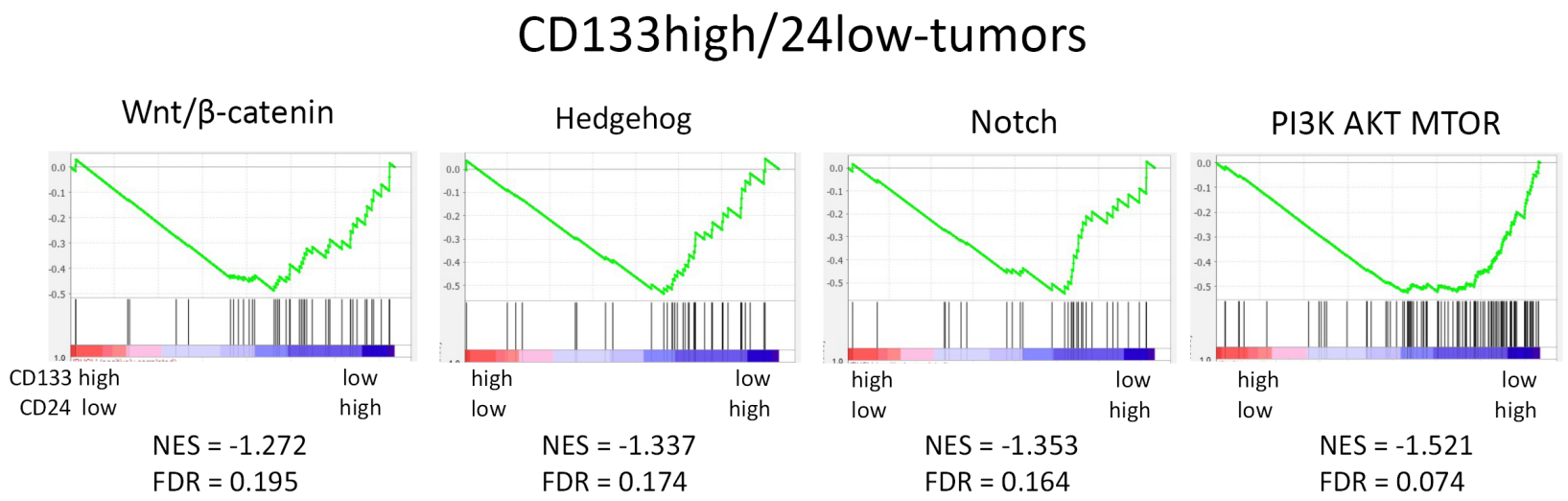


Figure S4. Gene set enrichment analysis (GSEA) of combined expression of CD133 and CD24 in the TCGA cohort. GSEA of cancer stemness-related gene sets; WNT signaling, Hedgehog signaling, Notch signaling and PI3K-AKT MTOR signaling.

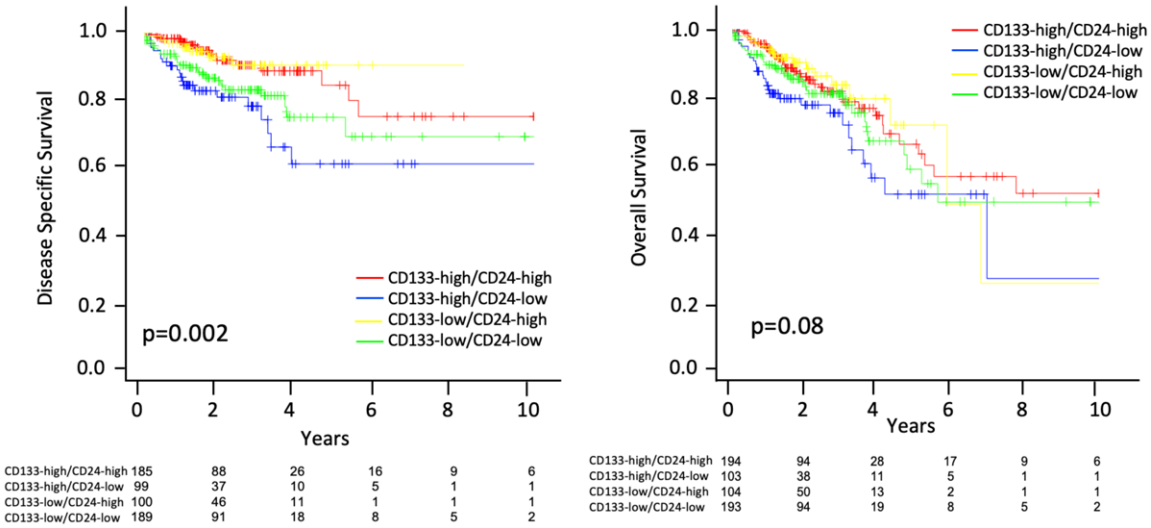


Figure S5. Disease specific survival and overall survival for the CD133/24 combination groups. Kaplan-Meier plots of disease specific survival and overall survival by 4 combinations of tumors with CD133-high/CD24-high, CD133-high/CD24-low, CD133-low/CD24-high, and CD133-low/CD24-low expression. Log-rank test was used to calculate the *p* value.