

CD133 expression is associated with less DNA repair, better response to chemotherapy and survival in ER-positive/HER2-negative breast cancer

Takumi Sato

University of Tokyo Hospital

Masanori Oshi

Yokohama City University Graduate School of Medicine

Jing Li Huang

Roswell Park Comprehensive Cancer Center

Kohei Chida

Roswell Park Comprehensive Cancer Center

Arya Mariam Roy

Roswell Park Comprehensive Cancer Center

Itaru Endo

Yokohama City University Graduate School of Medicine

Kazuaki Takabe

Kazuaki.Takabe@RoswellPark.org



Roswell Park Comprehensive Cancer Center

Research Article

Keywords: CD133, cancer stem cell, cell surface marker, HR positive breast cancer

Posted Date: March 27th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4148608/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Purpose: CD133, a cancer stem cells (CSC) marker, has been reported to be associated with treatment resistance and worse survival in triple-negative breast cancer (BC). However, the clinical relevance of CD133 expression in ER-positive/HER2-negative (ER+/HER2-) BC, the most abundant subtype, remains unknown.

Methods: The BC cohorts from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, n = 1904) and The Cancer Genome Atlas (TCGA, n = 1065) were used to obtain biological variables and gene expression data.

Results: Epithelial cells were the exclusive source of CD133 gene expression in a bulk BC. CD133-high ER+/HER2- BC was associated with CD24, NOTCH1, DLL1, and ALDH1A1 gene expressions, as well as with WNT/ β -Catenin, Hedgehog, and Notch signaling pathways, all characteristic for CSC. Consistent with a CSC phenotype, CD133-low BC was enriched with gene sets related to cell proliferation, such as G2M Checkpoint, MYC Targets V1, E2F Targets, and Ki67 gene expression. CD133-low BC was also linked with enrichment of genes related to DNA repair, such as BRCA1, E2F1, E2F4, CDK1/2. On the other hand, CD133-high tumors had proinflammatory microenvironment, higher activity of immune cells, and higher expression of genes related to inflammation and immune response. Finally, CD133-high tumors had better pathological complete response after neoadjuvant chemotherapy in GSE25066 cohort and better disease-free survival and overall survival in both TCGA and METABRIC cohorts.

Conclusion: CD133-high ER+/HER2- BC was associated with CSC phenotype such as less cell proliferation and DNA repair, but also with enhanced inflammation, better response to neoadjuvant chemotherapy and better prognosis.

Introduction

Although cancer stem cells (CSC) comprise only 0.1-1% of the cancer cells within a bulk tumor, they possess properties of self-renewal, initiate tumors from a single cell, and differentiate to resist treatments, thereby being implicated in causing relapses [1–3]. The roles of CSC in breast cancer (BC) have been studied [4–6]; however, the results were sometimes contradictory or incongruent due to heterogeneity in the techniques used to detect CSC populations [7–9]. For example, some studies utilized protein expression, often detected by immunohistochemistry, while others relied on gene expressions of multiple markers.

While many studies used CD44 and CD24 cell surface markers as well as Aldehyde Dehydrogenase (ALDH) to identify CSCs [4, 10–12], CD133 has also been well-characterized as a cell surface marker of CSCs [13]. CD133 was shown to activate WNT/ β Catenin pathway in vitro, which is an essential signaling pathway for cell proliferation of CSCs [14]. CD133 also activates the NOTCH and Hedgehog pathways which are other characteristic pathways related to CSCs [15].

Expression of CD133 assessed by immunohistochemistry in 67 patients was shown to correspond to the aggressiveness of triple-negative breast cancer (TNBC) [9]. CD133 was associated with a less response to neoadjuvant chemotherapy (NAC) in 102 BC patients when measured by immunohistochemistry [16]. CD133

mRNA overexpression was linked with a poor prognosis in invasive BC [17]. On the other hand, the clinical significance of CD133 expression in the ER-positive/HER2-negative (ER+/HER2-) BC remains unexplored. Despite the distinct biology of each BC subtype and considering that ER+/HER2- is the most prevalent subtype, there is a gap in our understanding of the specific implications of CD133 in this context.

Our group has been employing an in-silico approach to conduct translational research, investigating the clinical relevance of gene expression. Unlike experiments involving cell lines or animals, we have gained comprehensive and reliable insights by analyzing multiple independent large patient cohorts of transcriptomes associated with clinical parameters [18–21]. In this study, we hypothesize that high CD133 expression would be associated with the prognosis of ER+/HER2- BC. To better elucidate the association between the prognosis of ER+/HER2- BC and the CSC surface marker CD133, we utilized two distinct cohorts: The Cancer Genome Atlas (TCGA) cohort, which includes 1065 breast cancer patients, and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort, consisting of 1904 breast cancer patients.

Methods

Acquisition of BC patients' data

The gene expression data associated with clinical parameters was obtained from 1065 patients of The Cancer Genome Atlas (TCGA) BC cohort and 1904 patients of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort [22]. In accordance with previous studies, cBioportal was utilized to identify the cohorts which included mutation status (<https://www.cbioportal.org>) [13, 18, 23–35]. According to the staging guidelines of the American Joint Committee on Cancer, BC staging was conducted. Other BC cohorts analyzed in this study include cohorts of patients who received neoadjuvant chemotherapy (NAC); GSE25066 [36], GSE20194 [37], GSE32646 [38], and single cell sequence cohorts; SCP1039 (https://singlecell.broadinstitute.org/single_cell/study/SCP1039/a-single-cell-and-spatially-resolved-atlas-of-human-breast-cancers) and SCP1106 (https://singlecell.broadinstitute.org/single_cell/study/SCP1106/stromal-cell-diversity-associated-with-immune-evasion-in-human-triple-negative-breast-cancer). We obtained a waiver of the Roswell Park Institutional Review Board for approval, given the de-identified nature of the data points.

Gene set enrichment analysis (GSEA)

GSEA investigates the extent to which the expression of genes related to a certain pathway differs between groups. In this study, the cohort was categorized into 'high' and 'low' expression groups based on the median value as the cutoff. Fifty Hallmarks of Cancer gene sets in the Molecular Signatures Database (MSigDB) [39] were studied, as previously demonstrated by the Broad Institute (<http://www.gsea-msigdb.org/gsea/index.jsp>) [40]. The Normalized Enrichment Score (NES) was employed to assess the strength of the correlation. The False Discovery Rate (FDR) was utilized for statistical analysis, considering a cutoff for significance as an FDR value of less than 0.25. This choice aligns with the recommendation by the Broad Institute to adjust for gene set size, considering the multiple gene sets analyzed in our study.

Cell composition analysis

Thorsson et al. computed and reported cell proliferation score, homologous recombination defects score, mutation rate score, neoantigens score, and immune activity score [41]. We utilized xCell, the web-based computational algorithm developed at the University of California San Francisco (<https://www.xcell.ucsf.edu>), to analyze the association between the transcriptomic data and the enrichment of immune cells in different BC groups, as previously described [24–25, 27–31, 42–44]. Transcriptomic data of 64 cell types in the tumor microenvironment (TME) can be analyzed by the xCell algorithm. In this study, we analyzed transcriptomic data of immune cells such as T helper cells [45], regulatory T cells, M1 and M2 macrophages, CD8 + cells, CD4 memory cells [23], dendritic cells [24], and B cells.

Statistical Analysis

Data downloading, analysis, organization, and visualization were performed using R software (version 4.0.1. <http://www.r-project.org/>). Histograms were created to describe differences between “high” and “low” CD133 tumors. Two-sided test was employed to calculate p-values, and the cutoff of less than 0.05 was regarded as statistically significant. Median and interquartile level values were displayed using Tukey-type boxplots. Survival analyses were performed using Kaplan-Meier plots with log-rank tests.

Results

CD133 gene was predominantly expressed in epithelial cells in the BC tumor microenvironment (TME) of single-cell sequence cohorts

Given that CD133 is a well-characterized cell surface marker of CSCs [13], we first investigated which type of cells in the TME express CD133. This analysis was conducted using two independent single-cell sequence cohorts of BC patients, SCP1039 and SCP1106. We found that epithelial cells, from which BC cells arise, predominantly express CD133 in the TME ($p < 0.001$, Fig. 1). Since epithelial cells in bulk tumors are almost exclusively cancer cells, the observation of high CD133 expression in “normal” epithelial cells in SCP1039 likely represents less proliferative cancer cells.

CD133 expression was associated with CSC markers and stemness-related signaling pathways

Several CSC markers and their characteristic signaling pathways have been reported [6], and CD133 was reported to be one of them in colon [13, 46] and ovarian cancer [47]. Consequently, we sought to explore whether CD133 expression is associated with other CSC markers and stemness-related signaling pathways in ER-positive/HER2-negative (ER+/HER2-) BC. We found that CD133-high BC was significantly associated with higher expression levels of CSC marker genes: CD24, NOTCH1, DLL1, and ALDH1A1, consistently in both TCGA and METABRIC cohorts (all $p < 0.001$, Fig. 2A). Furthermore, we found that CD133-high BC significantly enriched signaling pathways that are activated in CSCs: WNT/ β catenin signaling (normalized enrichment score (NES) = 1.77, FDR = 0.21), Hedgehog signaling (NES = 1.27, FDR = 0.24), and NOTCH signaling (NES = 1.63, FDR = 0.12) in TCGA, which was validated in the METABRIC cohort: WNT/ β -catenin signaling (NES = 1.62, FDR = 0.03), Hedgehog signaling (NES = 1.28, FDR = 0.16), and NOTCH signaling (NES

= 1.39, FDR = 0.10) (Fig. 2B). These results suggest that CD133 is associated with other CSC markers and their signaling pathways in ER+/HER2- negative BC.

CD133 expression was negatively associated with cancer cell proliferation

Since CSCs are known to be less proliferative compared to differentiated cancer cells [3, 48–51], we were interested in whether the same trend holds true for CD133 expression in ER+/HER2- BC. Indeed, we found that Hallmark gene sets related to cell proliferation: E2F Targets (NES = -1.77, FDR = 0.05), G2M Checkpoint (NES = -1.52, FDR = 0.11), and MYC Targets V1 (NES = -1.64, FDR = 0.07) were all significantly enriched in CD133-low BC in the TCGA cohort (Fig. 3A). Enrichment of E2F Targets (NES = -1.57, FDR = 0.08) and G2M Checkpoint (NES = -1.48, FDR 0.11), but not MYC Targets V1, was validated in the METABRIC cohort. Furthermore, CD133-low tumors were associated with higher Ki67 gene (MKI67) expression compared to CD133-high tumors consistently in both cohorts (both $p < 0.001$, Fig. 3B). In concurrence with this, the pre-calculated Proliferation Score by Thorsson et al. [41] demonstrated an inverse relationship with CD133 expression in the TCGA cohort ($p < 0.001$, Fig. 3C). These findings compellingly suggest that elevated CD133 expression is linked to lower cell proliferation in ER+/HER2- BC.

CD133 expression was associated with less DNA repair activity and less mutation rate

We have previously demonstrated that DNA repair pathway enhancement is linked with cell proliferation [18, 45, 52]. Given that CD133 expression was negatively associated with cell proliferation in ER+/HER2- BC, it was of interest to explore whether CD133 gene expression was associated with the DNA repair pathway and its related gene expressions. We found that DNA repair gene set was significantly enriched in CD133-low tumors in both TCGA (NES = -1.83, FDR = 0.07) and METABRIC cohorts (NES = -1.60, FDR = 0.10, Fig. 4A), aligning with the relationship between CD133 expression and cell proliferation. Expressions of genes related to DNA repair such as BRCA1 and E2F7 were significantly higher in CD133-low tumors in both TCGA and METABRIC cohorts (all $p < 0.001$, Fig. 4B), whereas E2F1 and CDK1/2 expressions were higher in only one cohort (TCGA and METABRIC, respectively, both $p < 0.001$), not validated by other. Although E2F4 is also known as a gene related to DNA repair [53], there was no relationship with CD133 expression in our study. Homologous recombination deficiency (HRD) scores were negatively correlated with CD133 expression in the TCGA cohort (Fig. 4C), and silent and non-silent mutation rate was slightly enriched in CD133-low tumors (Fig. 4D). On the other hand, there were no differences in single nucleotide mutations (SNV) neoantigens and indel neoantigens by CD133 expression in the TCGA cohort. These results overall suggest CD133-high ER+/HER2- BC is associated with less DNA repair activity.

CD133-high tumors were associated with inflammation and immune response-related pathways

Upon discerning diminished DNA repair in CD133-high tumors, our focus shifted towards investigating a potential association between CD133 expression and cancer immunity. This inquiry is motivated by our previous findings, which demonstrated a correlation between reduced DNA repair, heightened

immunogenicity, and enhanced cancer immunity [45]. Utilizing the scores pre-calculated by Thorsson et al. [41], we found that CD133-high ER+/HER2- BC was associated with greater lymphocyte infiltration, TGF- β response (both $p < 0.001$), and T cell Receptor (TCR) richness ($p = 0.038$), suggesting a proinflammatory microenvironment in the TCGA cohort (Fig. 5A). We observed difference in the infiltration of several types of immune cells between CD133-high and low BC, such as Th1 cells, M1 and M2 macrophages which infiltrated less, and classical dendritic cells (cDC) which infiltrated more in CD133-high BC, while many types of immune cells, such as CD8 + T cells, CD4 + memory T cells, Th2 cells, Mast cells, and B-cells, presented no significant difference in both cohorts (Fig. 5B). Cytolytic activity, which encapsulates the collective killing prowess of immune cells, exhibited a positive correlation with CD133 expression in the METABRIC cohort ($p < 0.001$). The same trend held true in the TCGA cohort, although insignificantly ($p = 0.126$, Fig. 5C). We further found that many Hallmark gene sets related to inflammation and immune response were enriched in CD133-high tumors in the TCGA cohort: inflammatory response (NES = 1.51, FDR = 0.10), TNF- α signaling via NF κ B (NES = 1.57, FDR = 0.12), complement (NES = 1.52, FDR = 0.10), allograft rejection (NES = 1.32, FDR = 0.22), IL6 JAK STAT3 signaling (NES = 1.53, FDR = 0.12) and TGF- β signaling (NES = 1.65, FDR = 0.14), all of which were validated in the METABRIC cohort: inflammatory response (NES = 1.47, FDR = 0.06), TNF- α signaling via NF κ B (NES = 1.56, FDR = 0.04), complement (NES = 1.19, FDR = 0.23), allograft rejection (NES = 1.22, FDR = 0.20), IL6 JAK STAT3 signaling (NES = 1.42, FDR = 0.08) and TGF- β signaling (NES = 1.51, FDR = 0.05, Fig. 5D). These results suggest that CD133 expression is associated with inflammation and immune response in TME of ER+/HER2- BC.

CD133-high tumors were associated with better response to neoadjuvant chemotherapy (NAC) and survival

Given that CD133 expression was associated with less DNA repair but enhanced inflammation and immune response, we hypothesized that CD133-high ER+/HER2- BC may be vulnerable to cellular insult and respond better to chemotherapy. We utilized GSE25066, GSE20194, and GSE32646 cohorts, which included ER+/HER2- BC patients who underwent NAC: taxane and anthracycline in GSE25066, paclitaxel, 5-fluorouracil, cyclophosphamide, and doxorubicin in GSE20194, and paclitaxel followed by 5-fluorouracil, epirubicin, cyclophosphamide (P-FEC) combination in GSE32646. CD133-high tumors responded significantly better than CD133-low tumors with pathological complete response (pCR) rate of 6.6% vs 14.9% after NAC in the GSE25066 cohort ($n = 278$, $p = 0.03$), the largest cohort of the three. Although statistical significance was not achieved, the trend that CD133-high tumors achieved a higher pCR rate was consistent in GSE20194 with a pCR rate of 3.1% vs 7.7% and GSE32646 cohorts with a pCR rate of 3.7% vs 14.3% ($n = 129$ and 55 , respectively, Fig. 6A). CD133-high tumors were associated with better disease-free survival (DFS, $p = 0.015$) and overall survival (OS, $p = 0.05$) in the TCGA cohort, and both DFS ($p = 0.027$) and OS ($p < 0.001$) were validated by the METABRIC cohort (Fig. 6B).

Discussion

In summarizing our study, CD133 was exclusively expressed in cancer cells compared to stromal and immune cells and was associated with other CSC markers (CD24, NOTCH1, DLL1, and ALDH1A1), as well as enriched WNT/ β -Catenin, Hedgehog, and NOTCH signaling, validating CD133 as a CSC marker. We found

that the expression of the cancer stem cell marker CD133 is associated with reduced cell proliferation and DNA repair, yet heightened inflammation, and is linked to a more favorable outcomes after NAC and improved survival among ER-positive/HER2-negative BC patients.

Based on the fact that the CSCs are less proliferative than other types of cells in the tumor, we expected the expression of the CSC marker CD133 to be related to less cell proliferation. However, Joseph et al. reported that CD133 is associated with greater cell proliferation, less response to NAC, and worse prognosis in invasive BC [17]. Our data was consistent with our expectation and contradicted Joseph et al.'s report, which analyzed invasive BC as a whole, as opposed to our study that specifically investigated the ER+/HER2- subtype based on the understanding that biology and characteristics are significantly different by subtypes. It may be worth noting that CD133 protein expression evaluated by flow cytometry did not correlate with its mRNA expression level [54].

Our team, alongside other investigators, has reported an association between DNA repair enhancement and cell proliferation [18, 52]. The same trend has been shown by Oshi et al. in hepatocellular carcinoma [45], who found that enhanced DNA repair was associated with a worse prognosis and more cell proliferation but not with the fraction of immune cell infiltration nor immune response. Consistently, high expressions of RAD51 [18] or BRCA2 [52], both of which play a critical part in DNA repair, were associated with increased cell proliferation and aggressive biology in BC. Given that CD133-high BC was associated with less cell proliferation, its association with less DNA repair may explain its mechanism. On the other hand, Cheah et al. reported that CD133-marked putative CSCs correlated with proficient mismatch repair [55], thus multiple mechanisms may be involved in the relationship between CD133 expression and DNA repair.

We also found that inflammation and immune response were enriched in CD133-high TME. The number of many types of infiltrating cells in TME were not significantly different between high and low CD133 tumors and, interestingly, some types of cells were negatively correlated with CD133 expression. However, cytolytic activity, which represents the overall activity of the immune cells and thus cancer immunity, was significantly and positively correlated with high CD133 expression. It remains unclear precisely how and, after all, whether low DNA repair leads to high inflammation in the TME. Several previous studies reported that in several cell lines and cancer types, low DNA repair led to a higher neoantigen load, therefore high immunogenicity, and, as a result, more lymphocytes infiltration and richer inflammation [38, 56]. Nevertheless, while we observed slightly higher silent and non-silent mutation rates in CD133-low tumors, no discernible difference was noted in SNV neoantigens and indel neoantigens based on CD133 expression. This observation diminishes the persuasiveness of that explanation in our study. However, there are still several possible mechanisms that impaired DNA repair results in richer inflammation in TME, although not in higher loads of neoantigens. One is through the accumulation of DNA damage and subsequent activation of several signaling pathways such as the ATM/ATR pathway and the DNA/PK pathway, which can lead to the activation of NF κ B and other pro-inflammatory transcription factors that induce the production of pro-inflammatory cytokines, chemokines, and growth factors by cancer cells and surrounding immune cells [57]. This hypothesis is further supported by the fact that TNF α signaling via NF κ B is enriched in CD133-high tumors in our study (Fig. 5C), which is also known to enrich inflammation [58]. Another possible explanation is that impaired DNA repair results in the accumulation of damaged or misfolded proteins in the endoplasmic reticulum of cancer cells, leading to

endoplasmic reticulum stress and activation of the unfolded protein response (UPR). The UPR can also activate pro-inflammatory pathways, leading to the production of pro-inflammatory cytokines and chemokines [59–60], although several previous studies suggest that the chronic activation of UPR is considered a mechanism of tumor progression [61–63], going against better DFS and OS observed in our study, which may be due to the difference in cohorts.

Finally, and most importantly, we found that CD133-high BC carried a better survival outcome. We cannot help but speculate that while CD133-high tumors have a poor prognosis, as previous studies suggest accordingly with the cancer stem cell concept that involves self-renewal, differentiation, and the initiation of tumorigenesis, CD133-low tumors may carry even worse prognosis due to their ability to repair DNA, more cell proliferation, decreased immunogenicity, hence less response to NAC and worse survival outcome. The correlation between inflammation and pCR in invasive BC has been proposed by Hatzis et al. [36]. Furthermore, less cell proliferation in CD133-high BC may explain better prognosis, going along with some prior findings that showed an association between more expression of genes related to proliferation such as G2M [26–27], E2F [23, 25], and MYC [64] and worse prognosis in ER+/HER2- BC. In summary, the association of elevated CD133 expression in breast cancer cells with diminished DNA repair, improved response to NAC, and enhanced survival underscores CD133's potential role as a marker for predicting the treatment response in ER+/HER2- subtype BC.

Our method is subject to certain limitations inherent in the essentially retrospective nature of this study. Firstly, the utilization of patient sample data from a public domain means the analysis relies on information that had previously been cataloged, resulting in limited granularity. Secondly, the origin of the sample within the bulk tumor may vary among patients, even though the spatial relationship of CSCs in the bulk tumor may be of importance. It has been indicated that CSCs at the periphery of the bulk tumor may not have been sampled [65], although CD133 is known to be particularly upregulated in low O2 tissues [66]. These biases may have resulted in an underrepresentation of the full array and functionality of the CSCs.

Conclusion

CD133, a cancer stem cell (CSC) marker, was associated with less cell proliferation and less DNA repair, but with enhanced inflammation and better response to NAC and enhanced survival in ER+/HER2- BC.

Abbreviations

CSC, cancer stem cell; GSEA, Gene set enrichment analysis; TCGA, The Cancer Genome Atlas; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; OS, overall survival; DFS, disease-free survival.

Declarations

Funding:

This work was supported by US National Institutes of Health grants R37CA248018, R01CA-250412, R01CA251545, R01EB029596, as well as US Department of Defense BCRP grants W81XWH-19-1-0674 and W81XWH-19-1-0111 to K.T. This work was also supported by Roswell Park Comprehensive Cancer Center support grant P30CA016056 by National Cancer Institute cancer center. K.C was supported by Rotary Club District 2830.

Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval:

We obtained a waiver of the Roswell Park Institutional Review Board for approval, given the de-identified nature of the data points.

Data Availability:

Publicly available datasets were analyzed in this study. TCGA data can be found here:

[<https://www.cbioportal.org/> /Breast Invasive Carcinoma (TCGA, PanCancer Atlas)]. METABRIC data can be found here: [<https://www.cbioportal.org/> / Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)].

Data sets from each of the GEO databases can be downloaded from the following sites and access numbers:

[<https://www.ncbi.nlm.nih.gov/geo/> /GSE96058/GSE2034/GSE124647/GSE159956/GSE12276/GSE110590/GSE20194/GSE25066/GSE163882/ GSE20271].

Author Contributions:

The first draft of the manuscript was written by Takumi Sato and all authors commented on previous versions of the manuscript. All authors contributed to the study conception and design. Data collection and analysis were performed by Masanori Oshi. All authors read and approved the final manuscript.

References

1. Walcher L, Kistenmacher AK, Suo H, Kitte R, Dluczek S, Strauss A, Blandszun AR, Yevsa T, Fricke S, Kossatz-Boehlert U (2020) Cancer Stem Cells-Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. *Front Immunol* 11:1280. <https://doi.org/10.3389/fimmu.2020.01280>
2. Butti R, Gunasekaran VP, Kumar TVS, Banerjee P, Kundu GC (2019) Breast cancer stem cells: Biology and therapeutic implications. *Int J Biochem Cell Biol* 107:38-52. <https://doi.org/10.1016/j.biocel.2018.12.001>
3. Han J, Won M, Kim JH, Jung E, Min K, Jangili P, Kim JS (2020) Cancer stem cell-targeted bio-imaging and chemotherapeutic perspective. *Chem Soc Rev* 49(22):7856-7878. <https://doi.org/10.1039/d0cs00379d>
4. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 100(7):3983-3988.

<https://doi.org/10.1073/pnas.0530291100>

5. Idowu MO, Kmiecziak M, Dumur C, Burton RS, Grimes MM, Powers CN, Manjili MH (2012) CD44(+)/CD24(-/low) cancer stem/progenitor cells are more abundant in triple-negative invasive breast carcinoma phenotype and are associated with poor outcome. *Hum Pathol* 43(3):364-373. <https://doi.org/10.1016/j.humpath.2011.05.005>
6. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, Zhang G, Wang X, Dong Z, Chen F, Cui H (2020) Targeting cancer stem cell pathways for cancer therapy. *Signal Transduct Target Ther* 5(1):8. <https://doi.org/10.1038/s41392-020-0110-5>
7. Elbaiomy MA, Akl T, Atwan N, Elsayed AA, Elzaafarany M, Shamaa S (2020) Clinical Impact of Breast Cancer Stem Cells in Metastatic Breast Cancer Patients. *J Oncol* 2020:2561726. <https://doi.org/10.1155/2020/2561726>
8. Blackford AN, Jackson SP (2017) ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell* 66(6):801-817. <https://doi.org/10.1016/j.molcel.2017.05.015>
9. Xia P (2017) CD133 mRNA may be a suitable prognostic marker for human breast cancer. *Stem Cell Investig* 4:87. <https://doi.org/10.21037/sci.2017.10.03>
10. Tanei T, Morimoto K, Shimazu K, Kim SJ, Tanji Y, Taguchi T, Tamaki Y, Noguchi S (2009) Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin Cancer Res* 15(12):4234-4241. <https://doi.org/10.1158/1078-0432.CCR-08-1479>
11. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S, Schott A, Hayes D, Birnbaum D, Wicha MS, Dontu G (2007) ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* 1(5):555-567. <https://doi.org/10.1016/j.stem.2007.08.014>
12. Honeth G, Bendahl PO, Ringner M, Saal LH, Gruvberger-Saal SK, Lovgren K, Grabau D, Ferno M, Borg A, Hegardt C (2008) The CD44+/CD24- phenotype is enriched in basal-like breast tumors. *Breast Cancer Res* 10(3):R53. <https://doi.org/10.1186/bcr2108>
13. Huang JL, Oshi M, Endo I, Takabe K (2021) Clinical relevance of stem cell surface markers CD133, CD24, and CD44 in colorectal cancer. *Am J Cancer Res* 11(10):5141-5154.
14. Barzegar Behrooz A, Syahir A, Ahmad S (2019) CD133: beyond a cancer stem cell biomarker. *J Drug Target* 27(3):257-269. <https://doi.org/10.1080/1061186X.2018.1479756>
15. Akbari M, Shomali N, Faraji A, Shanehbandi D, Asadi M, Mokhtarzadeh A, Shabani A, Baradaran B (2020) CD133: An emerging prognostic factor and therapeutic target in colorectal cancer. *Cell Biol Int* 44(2):368-380. <https://doi.org/10.1002/cbin.11243>
16. Aomatsu N, Yashiro M, Kashiwagi S, Takashima T, Ishikawa T, Ohsawa M, Wakasa K, Hirakawa K (2012) CD133 is a useful surrogate marker for predicting chemosensitivity to neoadjuvant chemotherapy in breast cancer. *PLoS One* 7(9):e45865. <https://doi.org/10.1371/journal.pone.0045865>
17. Joseph C, Arshad M, Kurozomi S, Althobiti M, Miligy IM, Al-Izzi S, Toss MS, Goh FQ, Johnston SJ, Martin SG, Ellis IO, Mongan NP, Green AR, Rakha EA (2019) Overexpression of the cancer stem cell marker

- CD133 confers a poor prognosis in invasive breast cancer. *Breast Cancer Res Treat* 174(2):387-399. <https://doi.org/10.1007/s10549-018-05085-9>
18. Oshi M, Tokumaru Y, Benesch MG, Sugito N, Wu R, Yan L, Yamada A, Chishima T, Ishikawa T, Endo I, Takabe K (2022) High miR-99b expression is associated with cell proliferation and worse patient outcomes in breast cancer. *Am J Cancer Res* 12(10):4840-4852.
 19. Wu R, Patel A, Tokumaru Y, Asaoka M, Oshi M, Yan L, Ishikawa T, Takabe K (2022) High RAD51 gene expression is associated with aggressive biology and with poor survival in breast cancer. *Breast Cancer Res Treat* 193(1):49-63. <https://doi.org/10.1007/s10549-022-06552-0>
 20. Wu R, Roy AM, Tokumaru Y, Gandhi S, Asaoka M, Oshi M, Yan L, Ishikawa T, Takabe K (2022) NR2F1, a Tumor Dormancy Marker, Is Expressed Predominantly in Cancer-Associated Fibroblasts and Is Associated with Suppressed Breast Cancer Cell Proliferation. *Cancers (Basel)* 14(12). <https://doi.org/10.3390/cancers14122962>
 21. Wu R, Gandhi S, Tokumaru Y, Asaoka M, Oshi M, Yan L, Ishikawa T, Takabe K (2022) Intratumoral PDGFB gene predominantly expressed in endothelial cells is associated with angiogenesis and lymphangiogenesis, but not with metastasis in breast cancer. *Breast Cancer Res Treat* 195(1):17-31. <https://doi.org/10.1007/s10549-022-06661-w>
 22. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Group M, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346-352. <https://doi.org/10.1038/nature10983>
 23. Oshi M, Asaoka M, Tokumaru Y, Angarita FA, Yan L, Matsuyama R, Zsiros E, Ishikawa T, Endo I, Takabe K (2020) Abundance of Regulatory T Cell (Treg) as a Predictive Biomarker for Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Cancers (Basel)* 12(10). <https://doi.org/10.3390/cancers12103038>
 24. Oshi M, Asaoka M, Tokumaru Y, Yan L, Matsuyama R, Ishikawa T, Endo I, Takabe K (2020) CD8 T Cell Score as a Prognostic Biomarker for Triple Negative Breast Cancer. *Int J Mol Sci* 21(18). <https://doi.org/10.3390/ijms21186968>
 25. Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Nagahashi M, Matsuyama R, Endo I, Takabe K (2020) The E2F Pathway Score as a Predictive Biomarker of Response to Neoadjuvant Therapy in ER+/HER2- Breast Cancer. *Cells* 9(7). <https://doi.org/10.3390/cells9071643>
 26. Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Matsuyama R, Endo I, Takabe K (2020) G2M Cell Cycle Pathway Score as a Prognostic Biomarker of Metastasis in Estrogen Receptor (ER)-Positive Breast Cancer. *Int J Mol Sci* 21(8). <https://doi.org/10.3390/ijms21082921>
 27. Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Katz MHG, Takabe K (2020) High G2M Pathway Score Pancreatic Cancer is Associated with Worse Survival, Particularly after Margin-Positive (R1 or R2) Resection. *Cancers (Basel)* 12(10). <https://doi.org/10.3390/cancers12102871>
 28. Oshi M, Newman S, Tokumaru Y, Yan L, Matsuyama R, Endo I, Nagahashi M, Takabe K (2020) Intra-Tumoral Angiogenesis Is Associated with Inflammation, Immune Reaction and Metastatic Recurrence in

- Breast Cancer. *Int J Mol Sci* 21(18). <https://doi.org/10.3390/ijms21186708>
29. Tokumaru Y, Oshi M, Katsuta E, Yan L, Huang JL, Nagahashi M, Matsuhashi N, Futamura M, Yoshida K, Takabe K (2020) Intratumoral Adipocyte-High Breast Cancer Enrich for Metastatic and Inflammation-Related Pathways but Associated with Less Cancer Cell Proliferation. *Int J Mol Sci* 21(16). <https://doi.org/10.3390/ijms21165744>
 30. Oshi M, Newman S, Murthy V, Tokumaru Y, Yan L, Matsuyama R, Endo I, Takabe K (2020) ITPKC as a Prognostic and Predictive Biomarker of Neoadjuvant Chemotherapy for Triple Negative Breast Cancer. *Cancers (Basel)* 12(10). <https://doi.org/10.3390/cancers12102758>
 31. Oshi M, Tokumaru Y, Asaoka M, Yan L, Satyananda V, Matsuyama R, Matsuhashi N, Futamura M, Ishikawa T, Yoshida K, Endo I, Takabe K (2020) M1 Macrophage and M1/M2 ratio defined by transcriptomic signatures resemble only part of their conventional clinical characteristics in breast cancer. *Sci Rep* 10(1):16554. <https://doi.org/10.1038/s41598-020-73624-w>
 32. Takahashi H, Katsuta E, Yan L, Dasgupta S, Takabe K (2019) High expression of Annexin A2 is associated with DNA repair, metabolic alteration, and worse survival in pancreatic ductal adenocarcinoma. *Surgery* 166(2):150-156. <https://doi.org/10.1016/j.surg.2019.04.011>
 33. Cancer Genome Atlas N (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61-70. <https://doi.org/10.1038/nature11412>
 34. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5):401-404. <https://doi.org/10.1158/2159-8290.CD-12-0095>
 35. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269):pl1. <https://doi.org/10.1126/scisignal.2004088>
 36. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, Vidaurre T, Holmes F, Souchon E, Wang H, Martin M, Cotrina J, Gomez H, Hubbard R, Chacon JI, Ferrer-Lozano J, Dyer R, Buxton M, Gong Y, Wu Y, Ibrahim N, Andreopoulou E, Ueno NT, Hunt K, Yang W, Nazario A, DeMichele A, O'Shaughnessy J, Hortobagyi GN, Symmans WF (2011) A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* 305(18):1873-1881. <https://doi.org/10.1001/jama.2011.593>
 37. Andruska N, Zheng X, Yang X, Helferich WG, Shapiro DJ (2015) Anticipatory estrogen activation of the unfolded protein response is linked to cell proliferation and poor survival in estrogen receptor alpha-positive breast cancer. *Oncogene* 34(29):3760-3769. <https://doi.org/10.1038/onc.2014.292>
 38. Sicchieri RD, da Silveira WA, Mandarano LR, de Oliveira TM, Carrara HH, Muglia VF, de Andrade JM, Tiezzi DG (2015) ABCG2 is a potential marker of tumor-initiating cells in breast cancer. *Tumour Biol* 36(12):9233-9243. <https://doi.org/10.1007/s13277-015-3647-0>
 39. Angell HK, Bruni D, Barrett JC, Herbst R, Galon J (2020) The Immunoscore: Colon Cancer and Beyond. *Clin Cancer Res* 26(2):332-339. <https://doi.org/10.1158/1078-0432.CCR-18-1851>

40. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43):15545-15550. <https://doi.org/10.1073/pnas.0506580102>
41. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS, Cancer Genome Atlas Research N, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG, Shmulevich I (2018) The Immune Landscape of Cancer. *Immunity* 48(4):812-830 e814. <https://doi.org/10.1016/j.immuni.2018.03.023>
42. Oshi M, Roy AM, Yan L, Sasamoto M, Tokumaru Y, Wu R, Yamada A, Yamamoto S, Chishima T, Narui K, Endo I, Takabe K (2023) Accelerated glycolysis in tumor microenvironment is associated with worse survival in triple-negative but not consistently with ER+/HER2- breast cancer. *Am J Cancer Res* 13(7):3041-3054.
43. Chida K, Oshi M, Roy AM, Sato T, Endo I, Takabe K (2023) Pancreatic ductal adenocarcinoma with a high expression of alcohol dehydrogenase 1B is associated with less aggressive features and a favorable prognosis. *Am J Cancer Res* 13(8):3638-3649.
44. Chida K, Oshi M, Roy AM, Yachi T, Nara M, Yamada K, Matsuura O, Hashizume T, Endo I, Takabe K (2023) E2F target score is associated with cell proliferation and survival of patients with hepatocellular carcinoma. *Surgery* 174(2):307-314. <https://doi.org/10.1016/j.surg.2023.04.030>
45. Oshi M, Huyser MR, Le L, Tokumaru Y, Yan L, Matsuyama R, Endo I, Takabe K (2021) Abundance of Microvascular Endothelial Cells Is Associated with Response to Chemotherapy and Prognosis in Colorectal Cancer. *Cancers (Basel)* 13(6). <https://doi.org/10.3390/cancers13061477>
46. Todaro M, Francipane MG, Medema JP, Stassi G (2010) Colon cancer stem cells: promise of targeted therapy. *Gastroenterology* 138(6):2151-2162. <https://doi.org/10.1053/j.gastro.2009.12.063>
47. Ghoneum A, Gonzalez D, Abdulfattah AY, Said N (2020) Metabolic Plasticity in Ovarian Cancer Stem Cells. *Cancers (Basel)* 12(5). <https://doi.org/10.3390/cancers12051267>
48. Skvortsov S, Jimenez CR, Knol JC, Eichberger P, Schiestl B, Debbage P, Skvortsova I, Lukas P (2011) Radioresistant head and neck squamous cell carcinoma cells: intracellular signaling, putative biomarkers for tumor recurrences and possible therapeutic targets. *Radiother Oncol* 101(1):177-182. <https://doi.org/10.1016/j.radonc.2011.05.067>
49. Arnold CR, Mangesius J, Skvortsova I, Ganswindt U (2020) The Role of Cancer Stem Cells in Radiation Resistance. *Front Oncol* 10:164. <https://doi.org/10.3389/fonc.2020.00164>
50. Moore N, Lyle S (2011) Quiescent, slow-cycling stem cell populations in cancer: a review of the evidence and discussion of significance. *J Oncol* 2011. <https://doi.org/10.1155/2011/396076>
51. Skvortsova I, Skvortsov S, Stasyk T, Raju U, Popper BA, Schiestl B, von Guggenberg E, Neher A, Bonn GK, Huber LA, Lukas P (2008) Intracellular signaling pathways regulating radioresistance of human prostate

- carcinoma cells. *Proteomics* 8(21):4521-4533. <https://doi.org/10.1002/pmic.200800113>
52. Satyananda V, Oshi M, Endo I, Takabe K (2021) High BRCA2 Gene Expression is Associated with Aggressive and Highly Proliferative Breast Cancer. *Ann Surg Oncol* 28(12):7356-7365. <https://doi.org/10.1245/s10434-021-10063-5>
53. Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD (2002) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* 16(2):245-256. <https://doi.org/10.1101/gad.949802>
54. Sgambato A, Puglisi MA, Errico F, Rafanelli F, Boninsegna A, Rettino A, Genovese G, Coco C, Gasbarrini A, Cittadini A (2010) Post-translational modulation of CD133 expression during sodium butyrate-induced differentiation of HT29 human colon cancer cells: implications for its detection. *J Cell Physiol* 224(1):234-241. <https://doi.org/10.1002/jcp.22124>
55. Cheah PL, Li J, Looi LM, Teoh KH, Ong DB, Arends MJ (2018) DNA mismatch repair and CD133-marked cancer stem cells in colorectal carcinoma. *PeerJ* 6:e5530. <https://doi.org/10.7717/peerj.5530>
56. Green AR, Aleskandarany MA, Ali R, Hodgson EG, Atabani S, De Souza K, Rakha EA, Ellis IO, Madhusudan S (2017) Clinical Impact of Tumor DNA Repair Expression and T-cell Infiltration in Breast Cancers. *Cancer Immunol Res* 5(4):292-299. <https://doi.org/10.1158/2326-6066.CIR-16-0195>
57. Roos WP, Kaina B (2013) DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Lett* 332(2):237-248. <https://doi.org/10.1016/j.canlet.2012.01.007>
58. Balkwill F (2006) TNF-alpha in promotion and progression of cancer. *Cancer Metastasis Rev* 25(3):409-416. <https://doi.org/10.1007/s10555-006-9005-3>
59. Hetz C, Zhang K, Kaufman RJ (2020) Mechanisms, regulation and functions of the unfolded protein response. *Nat Rev Mol Cell Biol* 21(8):421-438. <https://doi.org/10.1038/s41580-020-0250-z>
60. Oshi M, Roy AM, Gandhi S, Tokumaru Y, Yan L, Yamada A, Endo I, Takabe K (2022) The clinical relevance of unfolded protein response signaling in breast cancer. *Am J Cancer Res* 12(6):2627-2640.
61. Sisinni L, Pietrafesa M, Lepore S, Maddalena F, Condelli V, Esposito F, Landriscina M (2019) Endoplasmic Reticulum Stress and Unfolded Protein Response in Breast Cancer: The Balance between Apoptosis and Autophagy and Its Role in Drug Resistance. *Int J Mol Sci* 20(4). <https://doi.org/10.3390/ijms20040857>
62. Minakshi R, Rahman S, Jan AT, Archana A, Kim J (2017) Implications of aging and the endoplasmic reticulum unfolded protein response on the molecular modality of breast cancer. *Exp Mol Med* 49(11):e389. <https://doi.org/10.1038/emm.2017.215>
63. Patel A, Oshi M, Yan L, Matsuyama R, Endo I, Takabe K (2021) The Unfolded Protein Response Is Associated with Cancer Proliferation and Worse Survival in Hepatocellular Carcinoma. *Cancers (Basel)* 13(17). <https://doi.org/10.3390/cancers13174443>
64. Schulze A, Oshi M, Endo I, Takabe K (2020) MYC Targets Scores Are Associated with Cancer Aggressiveness and Poor Survival in ER-Positive Primary and Metastatic Breast Cancer. *Int J Mol Sci* 21(21). <https://doi.org/10.3390/ijms21218127>

65. Damjanov I, Katic V, Stevens LC (1975) Ultrastructure of ovarian teratomas in LT mice. *Z Krebsforsch Klin Onkol Cancer Res Clin Oncol* 83(4):261-267. <https://doi.org/10.1007/BF00573012>
66. Ivanovic Z, Hermitte F, Brunet de la Grange P, Dazey B, Belloc F, Lacombe F, Vezon G, Praloran V (2004) Simultaneous maintenance of human cord blood SCID-repopulating cells and expansion of committed progenitors at low O2 concentration (3%). *Stem Cells* 22(5):716-724. <https://doi.org/10.1634/stemcells.22-5-716>

Figures

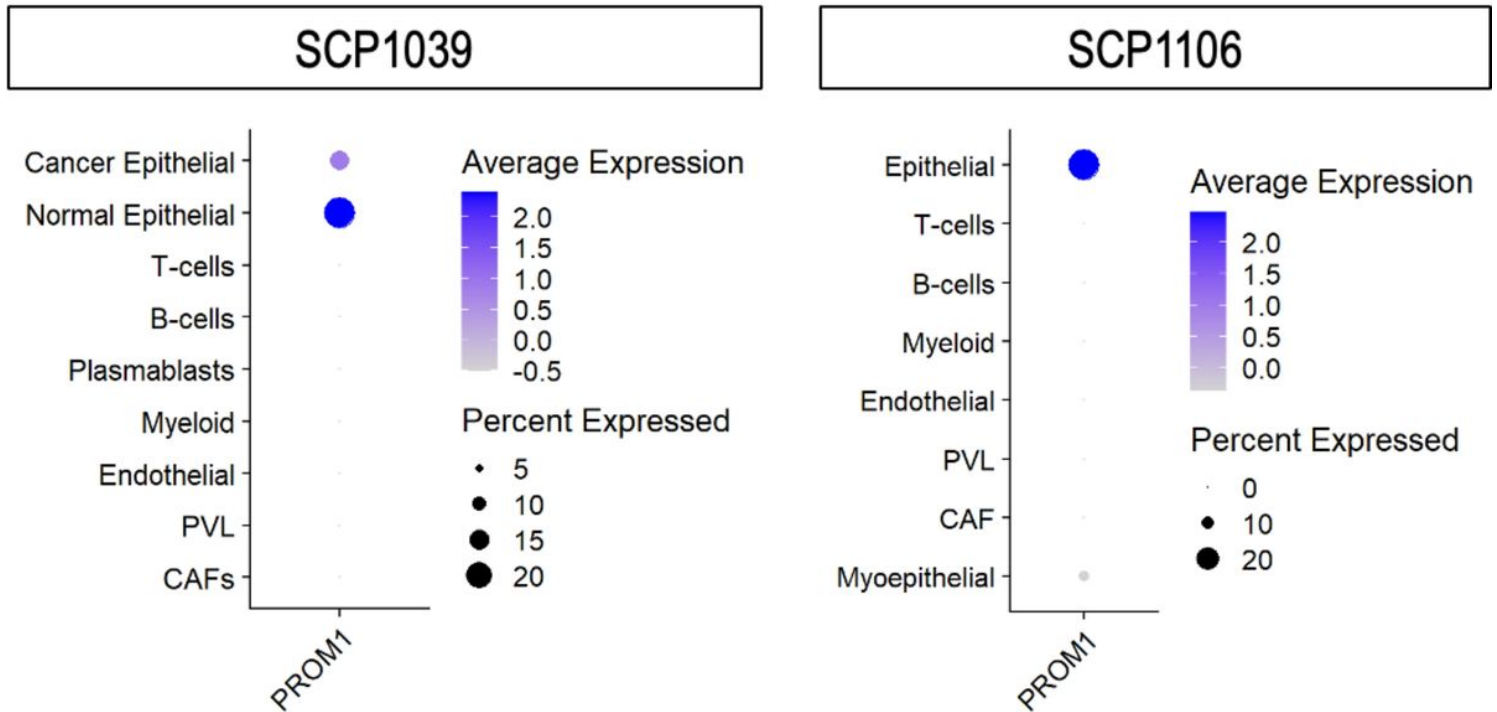


Figure 1

CD133 gene expression levels and percentages of cells expressed in single-cell sequence cohorts, SCP1039 and SCP1106.

CD133 expression pattern in various cell traits (cancer/normal epithelial cells, plasmablasts, T-cells, B-cells, endothelial cells, myeloid cells, PVL, and CAFs) in SCP1039 and SCP1106 cohorts. PVL, perivascular-like subpopulations; CAFs, Cancer-Associated Fibroblasts.

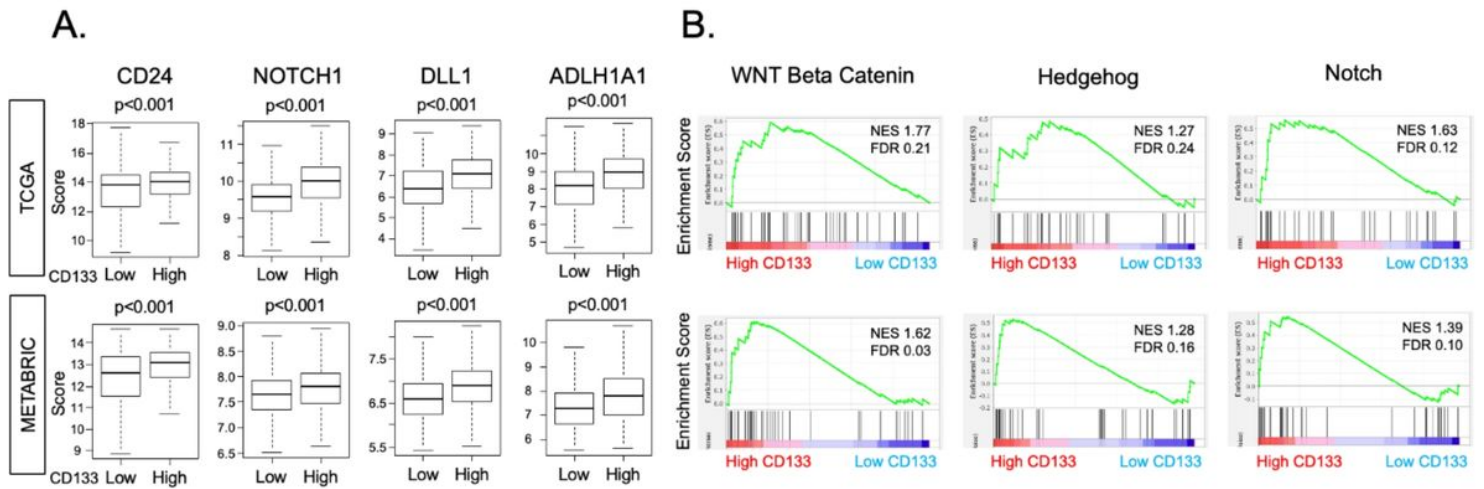


Figure 2

The relationship between CD133 expression and CSC surface marker and stemness-related signaling pathways in TCGA and METABRIC cohorts. (A) Gene expression levels of CSC markers (CD24, NOTCH1, DLL1, ALDH1A1) by high or low CD133 expression in TCGA and METABRIC cohorts are given in boxplots. The cohort was divided into “high” and “low” expression groups by the median value. Two-sided test was employed to calculate p-values, and the cutoff of less than 0.05 was regarded as statistically significant. The vertical line in the box shows the median, and top and bottom show the 25th and the 75th percentiles, respectively. **(B)** Gene set enrichment analysis (GSEA) of stemness-related signaling pathways; WNT/beta-catenin, Hedgehog, NOTCH. The classical GSEA method was used to compute NES and FDR values, with a cutoff of FDR < 0.25 considered statistically significant. CSC, cancer stem cells: NES, normalized enrichment score: FDR, false discovery rate.

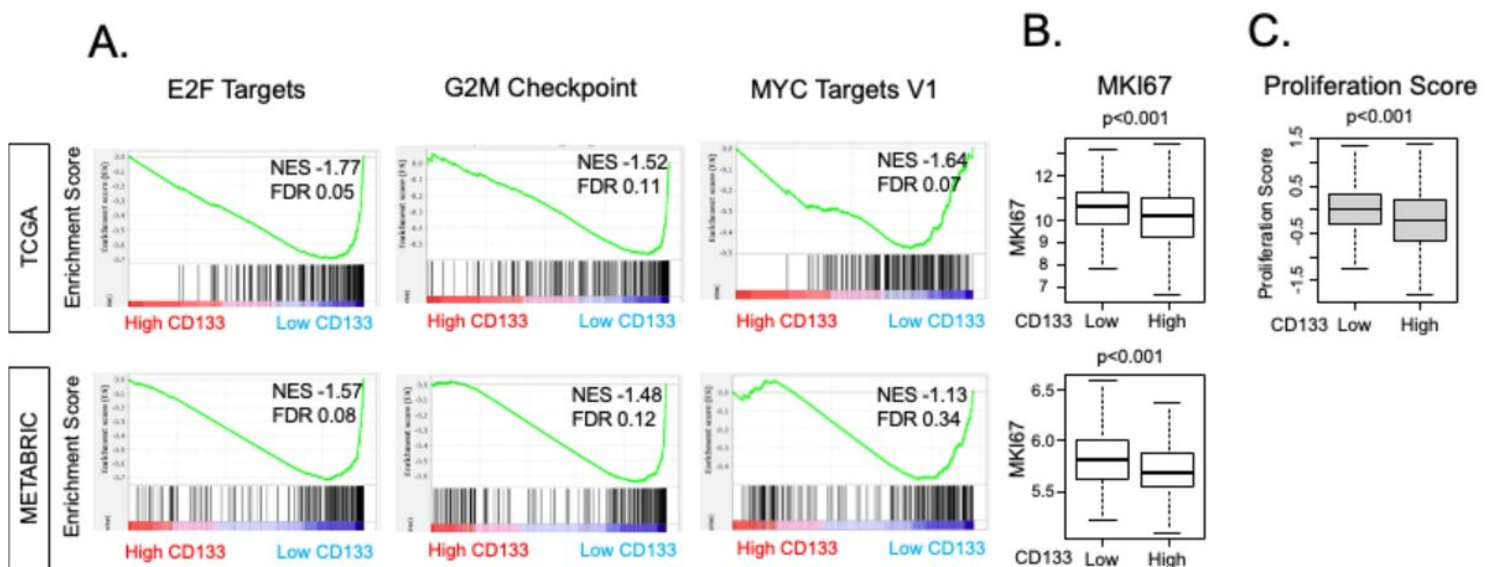


Figure 3

Association of CD133 expression with cell proliferation-related pathways, gene expression, and a score. (A) Gene set enrichment analysis (GSEA) of cell proliferation-related hallmark gene sets; E2F Targets, G2M Checkpoint, and MYC Targets version 1. The classical GSEA method was employed to calculate NES and FDR values, and the cutoff of $FDR < 0.25$ was regarded as statistically significant. **(B)** Boxplots showing the expression level of MKI67 of low and high CD133 expression groups in the TCGA and METABRIC cohorts. The cohort was divided into “high” and “low” expression groups by the median value. A two-sided test was employed to calculate p-values, and the cutoff of less than 0.05 was regarded as statistically significant. **(C)** Boxplot of the proliferation score that Thorson et al. pre-calculated.

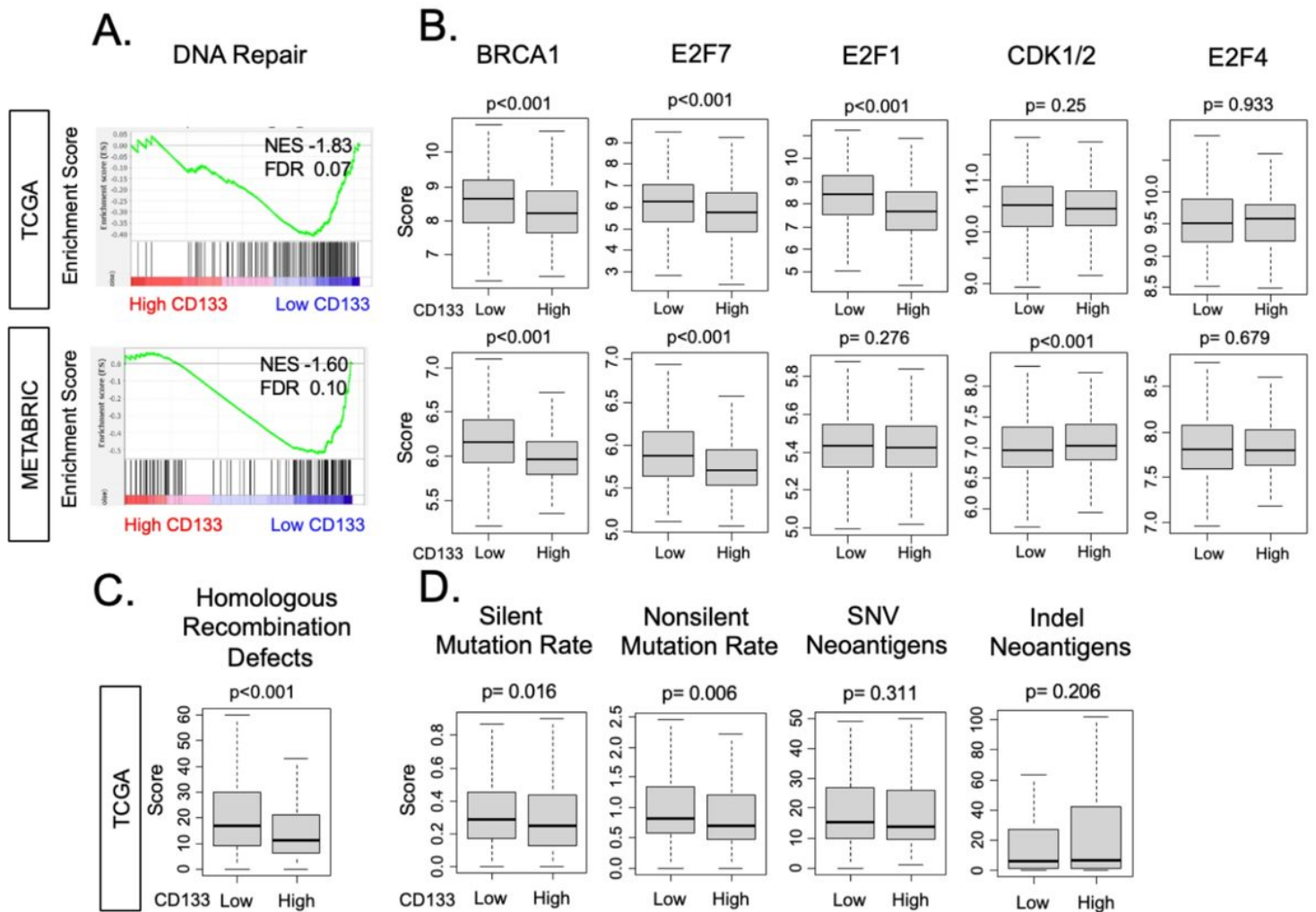


Figure 4

Association of CD133 expression with DNA-repair pathway, related gene expression, and a score. (A) Gene set enrichment analysis (GSEA) of DNA-repair pathway. The classical GSEA method was employed to calculate NES and FDR values, and the cutoff of $FDR < 0.25$ was regarded as statistically significant. **(B)** Boxplots illustrating the gene expression levels of DNA repair-related genes (BRCA1, E2F1, E2F4, E2F7, and CDK 1/2) in the low and high CD133 expression groups within the TCGA and METABRIC cohorts. Two-sided test was employed to calculate p-values, and the cutoff of less than 0.05 was regarded as statistically significant. **(C)** The boxplot of Homologous Recombination Defects scores that Thorson et al. pre-calculated, by low and

high CD133 expression in the TCGA cohort. The cohort was divided into “high” and “low” expression groups by the median value. **(D)** Boxplots illustrating the levels of SNV neoantigens, indel neoantigens, silent mutation rate, and non-silent mutation rate in the low and high CD133 expression groups within the TCGA cohort.

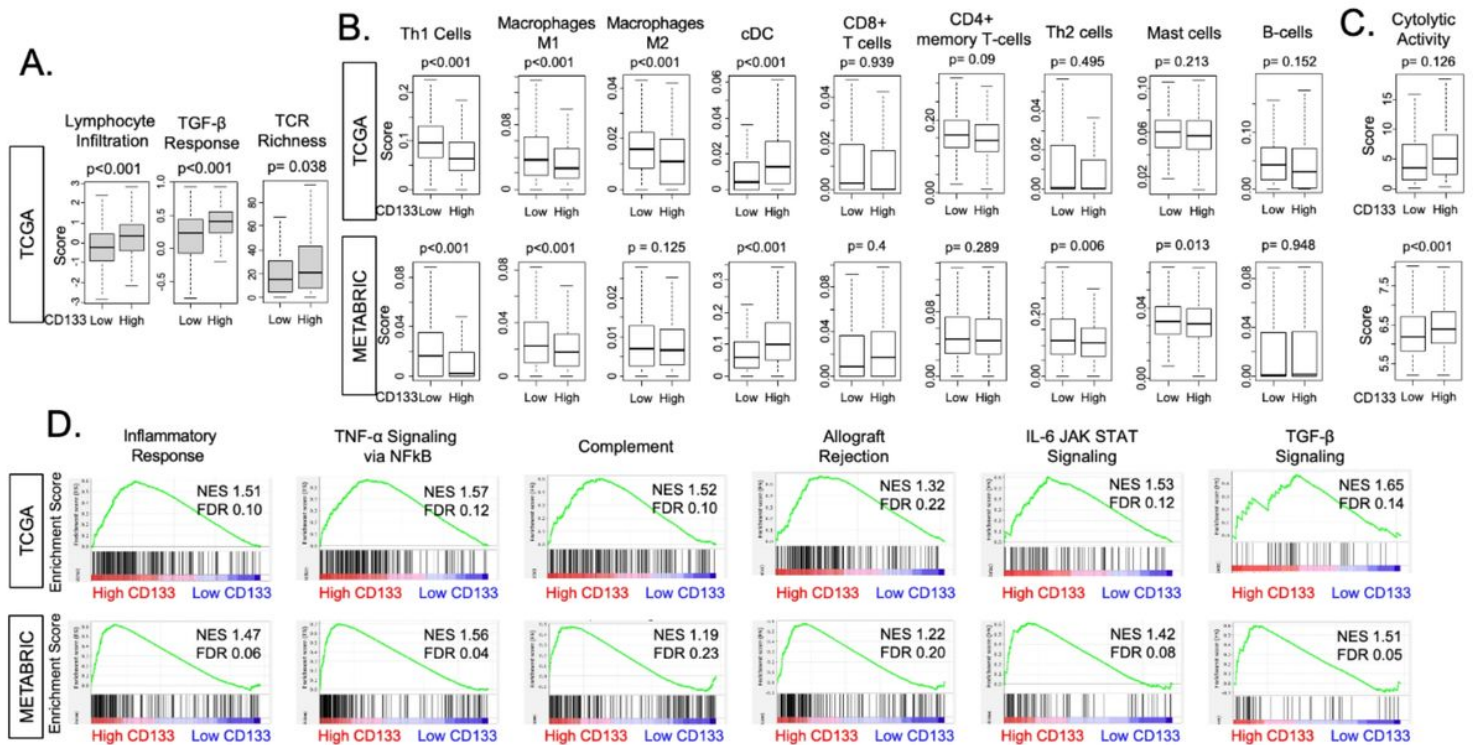


Figure 5

Association of CD133 expression with inflammation and immune response-related pathways. (A) The boxplots of scores Thorson et al. pre-calculated regarding pro-inflammatory microenvironment; Lymphocyte infiltrating signature, TGF- β response, and TCR richness. The TCGA cohort was divided into low and high CD133 expression groups by median cut-off. Two-sided test was employed to calculate p-values, and the cutoff of less than 0.05 was regarded as statistically significant. **(B)** Boxplots showing the levels of infiltration by several types of immune cells; Th1 cells, M1 and M2 macrophages, cDC in the TCGA and METABRIC cohorts. **(C)** The boxplots showing cytolytic activity, which summarizes the overall killing ability of immune cells. **(D)** Gene set enrichment analysis (GSEA) of inflammation and immune response-related pathways; inflammatory response, TNF α signaling via NF κ B, complement, allograft rejection, IL-6 JAK STAT signaling pathway, and TGF β signaling pathway. The classical GSEA method was employed to calculate NES and FDR values, and the cutoff of FDR < 0.25 was regarded as statistically significant. TCR, T cell Receptor; cDC, classical dendritic cells.

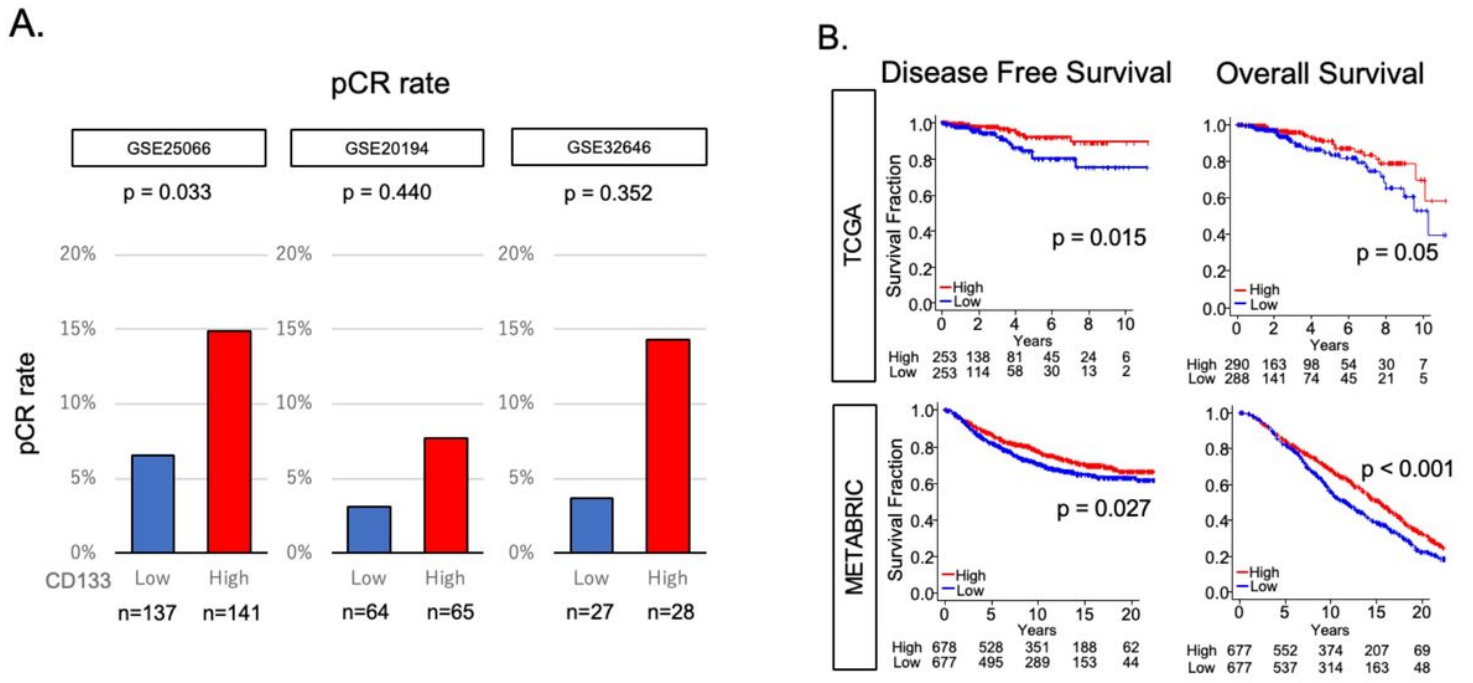


Figure 6

Association of CD133 expression with response to chemotherapy and survival. (A) Bar graphs of pCR rate by low and high CD133 expressing patients in GSE25066, GSE20194, and GSE32646 cohorts. The cohorts were divided into “high” and “low” expression groups by the median value. **(B)** Kaplan-Meier curves demonstrating DFS and OS in TCGA and METABRIC cohorts. The blue lines depict the low CD133 expression group, while the red lines represent the high CD133 expression group. Log-rank tests were employed for calculating p-values. pCR, pathological complete response: DFS, disease free survival: OS, overall survival.