Cancer Stem Cell Transcriptome Landscape Reveals Biomarkers Driving Breast Cancer Heterogeneity

Zhifa Zhang
Jiangnan university
Xiao Chen
Jiangnan university
Jianying Zhang
Zhengzhou University
Xiaofeng Dai (xiaofengteam@163.com)
Jiangnan university https://orcid.org/0000-0002-0006-4042

Research article

Keywords: Cancer stem cell, transcriptome, breast cancer, biomarker

Posted Date: September 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-80712/v1

Abstract

Background

Breast cancers are heterogeneous diseases with distinct clinical outcomes and cancer stem cell percentages. Exploring breast cancer stem cell landscape could help understand the heterogeneity of such cancers with profound clinical relevance.

Methods

We conducted transcriptional profiling of cancer stem cells and non-cancer stem cells isolated from 3 triple negative breast cancer cell lines, analyzed the cancer stem cell transcriptome landscape that drives breast cancer heterogeneity through differential expressed gene analysis, gene ontology and pathway enrichment as well as network construction, and performed experimental validations on the network hub gene.

Results

We identified a cancer stem cell feature panel consisting of 122 and 381 over-represented and under-expressed genes capable of differentiating breast cancer subtypes. We also underpin the prominent roles of the PI3K-AKT pathway in empowering cancer cells with uncontrolled proliferative and migrative abilities that ultimately foster cancer stemness, and reveal the potential promotive roles of *ATP6V1B1* on breast tumor stemness through functional *in vitro* studies.

Conclusions

Our study contributes in identifying a cancer stem cell feature panel for breast tumors that drives breast cancer heterogeneity at the transcriptional level, which provides a reservoir for diagnostic marker and/or therapeutic target identification once experimentally validated as demonstrated by *ATP6V1B1*.

Introduction

Breast cancers are highly heterogeneous regarding the molecular feature, pathogenesis and clinical outcome, which can be classified into at least luminal A, luminal B, human epithelial receptor 2 (HER2) positive, and triple negative breast cancer (TNBC) subtypes [1]. TNBCs are the most aggressive and heterogeneous among all subtypes, which can be further distinguished into, e.g., luminal androgen receptor (LAR), basal-like immunosuppressed (BLIS), basal-like immune-activated (BLIA), and mesenchymal (MES) subclasses [2], rendering the precise diagnosis difficult to make. Further, TNBCs lack targeted therapies due to the lack of surficial receptors estrogen receptor (ER), progestrogen receptor (PR) and HER2 [1], and patients carrying TNBCs suffer from severe side effects if treated with chemo- or radio-therapies [3]. Therefore, exploring effective therapeutic strategies with little side effects against TNBCs is an important yet difficult task for breast cancer management.

TNBCs contain higher percentage of cancer stem cells (CSCs) than the other subtypes [4]. Therefore, targeting CSCs, which are considered to promote DNA mutation during carcinogenesis and the generation of heterogeneous bulk cancer cells, brought up new opportunities to the therapeutics of TNBCs. CSCs are not homogeneous across cancer types [5]. Thus, exploring the transcriptomic pattern of CSCs of TNBCs can advance our understandings on the molecular features of TNBCs and help us identify the diagnostic marker and/or therapeutic targets for establishing novel medical platform for precise TNBC management.

By exploring the molecular differences between CSCs and non-CSCs of breast cancer cells, following pathway and gene ontology enrichment analysis, as well as network construction, we identified a CSC feature panel consisting of 503 genes that is capable of differentiating breast cancer subtypes as validated using two publicly available datasets. We hypothesize that CSCs drive the molecular heterogeneity of breast cancers and modulating key CSC relevant genes in this panel may alter cancer cell stemness with experimental validations confirmed using *ATP6V1B1* that has not been previously associated with cancer stemness. Our study contributes in providing a marker panel capturing breast cancer stemness and stimulating in-depth studies on the novel roles of *APT6V1B1* in CSC control.

Data And Methods

Differential expression analysis

The GSE132083 data stored at Gene Expression Omnibus (GEO) was used in this study. Differential expression analysis was performed based on student T test and Bayes theorem using the 'llimma' library from the 'Bioconductor' package (http://www.bioconductor.org/). Genes differentially expressed in stem versus non-stem cells were identified according to FPKM (fragments per kb per million reads, **Equation 1**) using the empirical Bayes approach (the 'eBayes' function), with the significance criteria set to log2FoldChange ≥ 2 and Benjamin-Hochberg adjusted p value≤0.05

$$\text{FPKM (gene A)} = \frac{\text{Fragment counts matched to gene A}}{\text{Fragment counts matched to all genes} \times \text{length of gene A}} \times 10^9$$

Hierarchical clustering analysis

Hierarchical clustering was performed to evaluate the differential expression pattern between stem and non-stem cell cohorts, where Euclidean distance and the Ward linkage were used [6]. An additional microarray gene expression data comprised of 56 cell-lines [7] was used for computational validation. The data was publicly available and stored at European Bioinformatics institution (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-181/).

Functional enrichment analysis

Functional enrichment analysis was performed based on Gene Ontology (GO; http://www.geneontology.org) and Kyoto Encyclopedia of Genes and Genomes database (KEGG; http://www.genome.jp/kegg/pathway.html) using the R package 'clusterProfiler'. Fisher's exact test was utilized to measure the significance of GO terms and biological pathways. The p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR), and the threshold of $p<0.01$ was used to assess the statistical significance of each test[8].

Gene interference and cell transfection assay

SKBR3 cells were plated in the 6-well plate, and three siRNAs (RIBOBIO, Guangzhou, China) were designed to silence *ATP6V1B1* (**supplementary Table 1**). SKBR3 cells were transfected with *ATP6V1B1* siRNA-1, siRNA-2, siRNA-3 and non-targeting siRNA (negative control) separately and in combination using siRNA mate transfection agent (RIBOBIO, Guangzhou, China). The siRNA concentrations used for transfection were 50 nm for single siRNA. Cells were incubated in 5% CO_2 at 37°C for 36 h and 48 h, respectively, before testing the silencing effect at transcriptional and translation levels using qRT-PCR and Western blot. Primer pairs of *ATP6V1B1* used in this study are provided in **Supplementary Table 2**. Triplicates were conducted, with statistical significance cutoff being set at $p \leq 0.05$ from student T-test.

Flow cytometry assay

SKBR3 cells after siRNA transfection for 36h were harvested and resuspended in PBS. 200,000 cells were incubated with the antibodies in recommended concentrations at 37°C for 0.5 h, followed by PBS washing for two times. Cells were sorted and analyzed by Flow cytometry (BD Biosciences). The ALDEFLUOR Kit (StemCell Technologies) was used to test ALDH activity following the manufacturer's protocol. The Diethylaminobenzaldehyde (DEAB), a specific ALDH inhibitor, was used as negative control in the ALDH analysis. Flow cytometry data processing was performed with flowJov10(Tree Star ,USA).

Transwell invasion and migration assays

Cell invasion and migration were examined in 24-well Transwell chambers (8 μm; Corning, USA). 50 ul Matrigel (2 mg/mL; Sigma-Aldrich) was used to cover the surface of the upper chambers in the invasion assay but not in the migration assay. SKBR3 cells (1×10^4) in 100 μL FBS-free medium were seeded in the upper chamber followed by the addition of 500 μL complete medium (with 10% FBS) to the bottom well. After 24-h incubation, cells invaded to the lower surfaces were gently cleaned and fixed with 4% Paraformaldehyde for 30 min and stained using 0.1% crystal violet for 30 min. Five random fields from the membrane were selected, and the number of migrated cells was counted under the ortho microscope (Nikon; ×100 magnification) and photographed.

Results

Selection of CSC feature genes

There are 212 and 522 genes concomitantly up- and down-regulated, respectively, in both SUM149 and HCC1937 cells (**Figures 1A, 1B**). Genes over- and under-expressed in CSCs of SUM149PT and HCC1937 were reduced to 122 and 381, respectively, when being mapped to the upper and lower 50% quantiles of SUM159PT (which is comprised primarily of CSCs). These 503 genes were selected as candidates capturing the features of breast cancer stem cells and named as 'CSC feature genes' (**Figure 1C, supplementary Table 3**).

Our experimental samples could be nicely clustered based on their cancer stemness following cell line-wise difference (**Figure 2A**), suggesting that these CSC feature genes primarily capture differences on cancer stemness. Such a clustering pattern could be well-reproduced using the 56 breast cancer cell line data retrieved from [7] (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-181/), where cell lines were grouped into luminal, basal-A and basal-B cohorts in the increasing order of cancer stemness (**Figure 2B**).

Gene Ontology, KEGG enrichment and network analysis of CSC feature genes

CSC feature genes are enriched in cell-cell adhesion and developmental related GO terms, with the top up- and down-regulated GO terms being 'extracellular matrix organization' and 'epidermis development', respectively (**Figures 3A, 3B**), implicating the importance of cell migration and differentiation in fostering CSC features. 'PI3K/AKT signaling' and 'tight junction' are the topmost altered KEGG pathways (**Figures 3C, 3D**), where the critical roles of PI3K/AKT pathway in cancer cell migration [9, 10] and differentiation [11] and migration are well-known, and losing tight junctions represents an important feature of cancer metastasis [12].

Regulatory networks were constructed using up- and down-regulated CSC feature genes (**Figure 4**). Overall, more CSC feature genes were down-regulated than up-regulated, which is in accordance with the less differentiated state that CSCs represent. PRKCA and ESR1 each has the most connectivity among the up- and down-regulated CSC feature genes, where PRKCA is a central signaling node and therapeutic target for breast cancer stem cells [13] and co-inhibiting ESR1, mTORC1, HDAC suppresses breast cancer stemness [14].

Functional studies on one CSC feature gene

Among these CSC feature genes, selected *ATP6V1B1*, a gene with the highest connectivity next to *ESR1* and being connected with *FA2H* (our recently identified gene with suppressive role on cancer stemness [15]), and tested its functionalities in transiting cells between states having different CSC and non-CSC percentages.

The siRNA technique was used to silence *ATP6V1B1* in SKBR3 cells to explore the functional roles of ATP6V1B1 on breast cancer stem cell regulation. By testing the efficiency of the designed siRNAs separately and in combination, we found that siRNA-1 and siRNA-3 could effectively silence *ATP6V1B1* in SKBR3 cells at the transcriptional level (**Figure 5A**, p=0.0012 for siRNA-1, p=0.0064 for siRNA-3) and

translational level (**Figures 5B**, p=0.015 for siRNA-1, p=0.007 for siRNA-3). We used siRNA-1 in the following assays. Flow cytometry analysis was conducted following the procedure aforementioned to examine the CSC percentage of cancer cells. The results showed a substantial reduction on CSC percentage after silencing *ATP6V1B1*, with and without supplementing cells with ALDH inhibitor (DEAB) in SKBR3 cells (**Figure 5C**). Transwell assays were performed to test cell invasion and migration according to procedures described in [16]. With non-transfected cells as the control, silencing *ATP6V1B1* significantly reduced cell invasion (p=0.0335) and migration (p=0.0213) in SKBR3 cells (**Figure 5D**). In addition, breast cancer patient overall survival could be significantly stratified by ATP6V1B1 protein expression using the 'Tang_2018 (n=118)' dataset (p=0.0067, **Figure 5E**).

In summary, by silencing *ATP6V1B1* in SKBR3 (considered to be comprised of non-CSCs), cells exhibited increased cancer stem cell percentage, cell invasion and migration abilities, which is consistent with our observation from the transcriptome data that *ATP6V1B1* is lowly expressed in CSCs.

Discussion

More genes were suppressed than over-expressed in CSCs, i.e., 381 versus 122 genes, in the CSC feature panel, suggesting a more plastic state that CSCs represent. Thus, CSCs are similar with stem cells in a sense that they are pluripotent, but differ from stem cells as they are chaotic and occur under pathological conditions.

Down-regulated genes were enriched in 'tight junction' and 'epidermis development' (**Figures 3A, 3C**), attracting CSCs in a less differentiated state. Up-regulated genes were enriched in the 'PI3K-AKT signaling pathway' and cell migration related GO terms (**Figures 3B, 3D**), implicating the more proliferative and aggressive feature of CSCs than the bulk tumor cells.

The regulatory network constructed using down-regulated CSC feature genes was centered around *ESR1* and *ATP6V1B1*. *ESR1*, also known as *ER*, is the primary biomarker used for breast cancer subtyping and typically under-expressed in triple negative breast cancers that harbor higher cancer stemness. Several genes connected with *ESR1* have been associated with breast cancer stemness. For example, we previously reported the suppressive role of *FOXA1* on breast cancer stemness [17]; *SOX9* is as a stem cell factor[18] that drives the epithelial mesenchymal transition (EMT) in non-small cell lung cancer through Wnt pathway[19] and maintains human breast luminal progenitor and breast cancer stem cells through *SOX9* mediated signaling [20], and both the *SOX9/FXYD3/SRC* [21] and the *SOX9/SOX2* [20] axes are critical for breast cancer stem cell functionalities; *PLA2G7* is associated with ER negativity in clinical breast cancer samples and regulates EMT *in vitro*[22]. The prognostic values of several markers have been reported, including *PTPN6* and *KRT19* in breast cancers[23-25], *EPPK1* in non-small cell lung cancers[26], *SERPINB5* in colorectal cancers[27], and *NCCRP1* in squamous cell carcinoma[28]. Besides, the roles of *S100A7* and *SLC37A1* was implicated in breast cancers[29, 30], that of *SERPINB4* was reported in squamous cell carcinoma[31], and that of *MYH14* was lately identified in pancreatic

cancers[32]. Also identified down-regulated in the CSC feature panel include several famous genes in cancer stem cell maintenance or signaling such as *CDH1* [33], *EPCAM* and *CLDN7* [34].

While the ER centered cluster has already been well-associated with cancer stemness (i.e., 12 out of 14 densely connected genes already have well-supported evidences), relatively less has been reported on the association between the cluster centered at *ATP6V1B1* and CSC features. Most genes in this cluster are involved in cell energy production and metabolic regulations, suggesting the pivotal roles of metabolic reprogramming in CSC feature maintenance. Out of the 11 genes, 5 have known roles in cancer initiation and progression, i.e., *SH3YL1* together with *DOCK4* regulate breast cancer cell migration [35], *ATP6V1C2* is an early prognostic marker for colorectal cancers [36], *PRKAR2B* promotes EMT and is oncogenic in prostate cancers [37, 38], *MYO5B* is associated with gastric cancers[39, 40], and the physiological role of *AKR1B15* and its involvement in cancer development has been characterized in [41]. Among the rest 6 members in the cluster, *FAT2* encodes a cadherin family member and is aliased as FAT tumor suppressor homolog 2, *MYO5C* encodes a member of the same family with *MYO5B* that has already been implicated in cancers, *ATP6V1B1*, *ATP6V1C2* and *ATP2C2* encode ATPase subunits, and *FA2H* encodes fatty acid 2-hydroxylase, where the association between *FA2H* and cancers has been recently reported[15].

Up-regulated genes are centered at *PRKCA* which is a key component in PI3K-AKT signaling. This is consistent with the pathway analysis and suggests the prominent roles of uncontrolled proliferation and increased migration ability in CSC feature initiation and maintenance. *VIM* [42], *MMP2*[43], *FGF2*[44], *ITBG3*[45, 46], *ANXA6*[47] are metastasis-associated genes. The oncogenic roles of *KCNMA1* has been revealed in breast, prostate, ovarian and colorectal cancers [48-52], and *RRAD* is associated with glucose uptake in a human ovarian cancer model[53] and implicated in non-small cell lung cancers[54].

This CSC feature panel can clearly distinguish samples with high and low stemness in GSE132083, and characterize basal-A, basal-B and luminal cell lines in the E-MTAB-181 dataset. Though cell line wise difference exists in the molecular profiling of GSE132083, it is secondary to the differences characterized by cancer stemness (**Figure 2**). These not only suggest the validity of identified CSC feature in differentiating breast cancer cells with different CSC percentages, but also implicate the prominent roles of CSCs in driving breast cancer heterogeneity.

We experimentally explored the functionalities of *ATP6V1B1* and its association with breast cancer stemness, through which the novel role of *ATP6V1B1* in suppressing breast cancer stemness has been revealed. However, in-depth investigations on the regulatory mechanism of *ATP6V1B1* in breast cancer progression and its potential prognostic value on breast cancer outcome await to be conducted.

Conclusion

We identify a CSC feature panel consisting of 503 genes which can be used for breast cancer cell subtype differentiation and offer a reservoir for diagnostic marker and therapeutic target identification. We propose that CSCs are the driving force of breast cancer heterogeneity empowering cells with differential uncontrolled proliferative and migrative abilities. We also reveal the potential encouraging role of

ATP6V1B1 on cancer stemness with explorations on the underlying mechanisms await in-depth elucidation.

Declarations

Availability of supporting data

The data is stored at Gene Expression Omnibus (GEO) as GSE132083(https://www.ncbi.nlm.nih.gov/geo/).

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

The authors are consent for the publication of this paper.

Competing Interests

The authors declare no conflict of interests

Funding

This study was funded by the National Natural Science Foundation of China (Grant No. 81972789), the National Science and Technology Major Project (Grant No. 2018ZX10302205-004-002), Technology Development Funding of Wuxi (Grant No. WX18IVJN017). These funding sources have no role in the writing of the manuscript or the decision to submit it for publication.

Authors' Contributions

XF Dai designed this study and prepared the draft. ZF Zhang conducted experimental validations under the supervision of XF Dai. XF Dai and X Chen conducted the computational analysis. X.F. Dai and J.Y. Zhang financed this project.

Acknowledgements

Not applicable.

References

1. Dai, X., et al., *Breast cancer intrinsic subtype classification, clinical use and future trends.* Am J Cancer Res, 2015. **5**(10): p. 2929-43.
2. Burstein, M.D., et al., *Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer.* Clin Cancer Res, 2015. **21**(7): p. 1688-98.

3. Bergh, J., et al., *First-line treatment of advanced breast cancer with sunitinib in combination with docetaxel versus docetaxel alone: results of a prospective, randomized phase III study.* J Clin Oncol, 2012. **30**(9): p. 921-9.
4. Dai, X., et al., *Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes.* J Cancer, 2016. **7**(10): p. 1281-94.
5. Guenechea, G., et al., *Distinct classes of human stem cells that differ in proliferative and self-renewal potential.* Nature Immunology, 2001. **2**(1): p. 75-82.
6. S, W. and Z. J, *Variable selection for model-based high-dimensional clustering and its application to microarray data*. Vol. 64. 2008. 440-448.
7. Shen, L.L., et al., *Metabolic reprogramming in triple-negative breast cancer through Myc suppression of TXNIP.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(17): p. 5425-5430.
8. Y, W., et al., *Identification of methylated genes and miRNA signatures in nasopharyngeal carcinoma by bioinformatics analysis.* Molecular Medicine Reports, 2018. **17**(4): p. 4909-4916.
9. Tang, Z., et al., *KIAA1199 promotes invasion and migration in non-small-cell lung cancer (NSCLC) via PI3K-Akt mediated EMT.* J Mol Med (Berl), 2019. **97**(1): p. 127-140.
10. Xie, P., et al., *TRAF4 promotes endometrial cancer cell growth and migration by activation of PI3K/AKT/Oct4 signaling.* Exp Mol Pathol, 2019. **108**: p. 9-16.
11. Zhang, R., et al., *RXRalpha provokes tumor suppression through p53/p21/p16 and PI3K-AKT signaling pathways during stem cell differentiation and in cancer cells.* Cell Death Dis, 2018. **9**(5): p. 532.
12. Yang, Y., et al., *The chromatin remodeling protein BRM regulates the transcription of tight junction proteins: Implication in breast cancer metastasis.* Biochim Biophys Acta Gene Regul Mech, 2019. **1862**(5): p. 547-556.
13. Tam, W.L., et al., *Protein Kinase C alpha Is a Central Signaling Node and Therapeutic Target for Breast Cancer Stem Cells.* Cancer Cell, 2013. **24**(3): p. 347-364.
14. Sulaiman, A., et al., *Co-inhibition of mTORC1, HDAC and ESR1alpha retards the growth of triple-negative breast cancer and suppresses cancer stem cells.* Cell Death Dis, 2018. **9**(8): p. 815.
15. Dai, X., et al., *FA2H Exhibits Tumor Suppressive Roles on Breast Cancers via Cancer Stemness Control.* Front Oncol, 2019. **9**: p. 1089.
16. Thompson, C.C., et al., *Pancreatic cancer cells overexpress gelsolin family-capping proteins, which contribute to their cell motility.* Gut, 2007. **56**(1): p. 95-106.
17. Dai, X., et al., *FOXA1 is Prognostic of Triple Negative Breast Cancers by Transcriptionally Suppressing SOD2 and IL6.* Int J Biol Sci, 2019. **15**(5): p. 1030-1041.
18. Aguilar-Medina, M., et al., *SOX9 Stem-Cell Factor: Clinical and Functional Relevance in Cancer.* Journal of Oncology, 2019.

19. Huang, J.Q., et al., *SOX9 drives the epithelial-mesenchymal transition in non-small-cell lung cancer through the Wnt/-catenin pathway.* Journal of Translational Medicine, 2019. **17**.
20. Domenici, G., et al., *A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells.* Oncogene, 2019. **38**(17): p. 3151-3169.
21. Xue, Y., et al., *SOX9/FXYD3/Src Axis Is Critical for ER thorn Breast Cancer Stem Cell Function.* Molecular Cancer Research, 2019. **17**(1): p. 238-249.
22. Lehtinen, L., et al., *PLA2G7 associates with hormone receptor negativity in clinical breast cancer samples and regulates epithelial-mesenchymal transition in cultured breast cancer cells.* Journal of Pathology Clinical Research, 2017. **3**(2): p. 123-138.
23. Youssef, G., et al., *Phosphorylation of NTRK1 at Y674/Y675 induced by TP53-dependent repression of PTPN6 expression: A potential novel prognostic marker for breast cancer.* Modern Pathology, 2014. **27**(3): p. 361-374.
24. Wang, X.M., et al., *KRT19 and CEACAM5 mRNA-marked circulated tumor cells indicate unfavorable prognosis of breast cancer patients.* Breast Cancer Research and Treatment, 2019. **174**(2): p. 375-385.
25. Saha, S.K., et al., *Opposing Regulation of Cancer Properties via KRT19-Mediated Differential Modulation of Wnt/beta-Catenin/Notch Signaling in Breast and Colon Cancers.* Cancers, 2019. **11**(1).
26. Arimura, K., et al., *Prognostic Value of EPPK1 Molecular Alterations in Non-Small Cell Lung Cancer.* American Journal of Respiratory and Critical Care Medicine, 2019. **199**.
27. Baek, J.Y., et al., *Serpin B5 is a CEA-interacting biomarker for colorectal cancer.* International Journal of Cancer, 2014. **134**(7): p. 1595-1604.
28. Miwa, T., et al., *Identification of NCCRP1 as an epigenetically regulated tumor suppressor and biomarker for malignant phenotypes of squamous cell carcinoma of the esophagus.* Oncol Lett, 2017. **14**(4): p. 4822-4828.
29. Wilkie, T., et al., *The role of S100A7 in microbiota mediated inflammation and breast cancer progression.* Cancer Research, 2018. **78**(4).
30. Iacopetta, D., et al., *SLC37A1 Gene expression is up-regulated by epidermal growth factor in breast cancer cells.* Breast Cancer Research and Treatment, 2010. **122**(3): p. 755-764.
31. Catanzaro, J.M., et al., *Oncogenic Ras induces inflammatory cytokine production by upregulating the squamous cell carcinoma antigens SerpinB3/B4.* Nat Commun, 2014. **5**: p. 3729.
32. Surcel, A., et al., *Targeting Mechanoresponsive Proteins in Pancreatic Cancer: 4-Hydroxyacetophenone Blocks Dissemination and Invasion by Activating MYH14.* Biophysical Journal, 2019. **116**(3): p. 260a-260a.
33. Shokraii, F., et al., *Histone Modification Marks Strongly Regulate CDH1 Promoter in Prostospheres as A Model of Prostate Cancer Stem Like Cells.* Cell J, 2019. **21**(2): p. 124-134.
34. Kyuno, D., et al., *Therapeutic Targeting Cancer-Initiating Cell Markers by Exosome miRNA: Efficacy and Functional Consequences Exemplified for claudin7 and EpCAM.* Transl Oncol, 2019. **12**(2): p.

191-199.

35. Kobayashi, M., et al., *Dock4 forms a complex with SH3YL1 and regulates cancer cell migration.* Cell Signal, 2014. **26**(5): p. 1082-8.
36. Zumwalt, T.J., et al., *The ATP6V1C2 (a vacuolar-ATPase gene) is a novel early prognosticator for colorectal cancer.* Cancer Research, 2016. **76**.
37. Sha, J.J., et al., *PRKAR2B plays an oncogenic role in the castration-resistant prostate cancer.* Oncotarget, 2017. **8**(4): p. 6114-6129.
38. Sha, J.J., et al., *PRKAR2B promotes prostate cancer metastasis by activating Wnt/-catenin and inducing epithelial-mesenchymal transition.* Journal of Cellular Biochemistry, 2018. **119**(9): p. 7319-7327.
39. Dong, W.J., L.P. Wang, and R.Z. Shen, *MYO5B Is Epigenetically Silenced and Associated with MET Signaling in Human Gastric Cancer.* Digestive Diseases and Sciences, 2013. **58**(7): p. 2038-2045.
40. Dong, W.J., et al., *Inactivation of MYO5B Promotes Invasion and Motility in Gastric Cancer Cells.* Digestive Diseases and Sciences, 2012. **57**(5): p. 1247-1252.
41. Gimenez-Dejoz, J., et al., *Substrate Specificity, Inhibitor Selectivity and Structure-Function Relationships of Aldo-Keto Reductase 1B15: A Novel Human Retinaldehyde Reductase.* PLoS One, 2015. **10**(7): p. e0134506.
42. Karamagkiolas, S., et al., *Expression of vimentin (VIM) and metastasis-associated 1 (MTA1) protein in laryngeal squamous cell carcinoma are associated with prognostic outcome of patients.* Am J Otolaryngol, 2019.
43. Wu, W., et al., *beta-hCG promotes epithelial ovarian cancer metastasis through ERK/MMP2 signaling pathway.* Cell Cycle, 2019. **18**(1): p. 46-59.
44. Wang, Y., et al., *FGF2 promotes metastasis of uveal melanoma cells via store-operated calcium entry.* Onco Targets Ther, 2017. **10**: p. 5317-5328.
45. Hu, L., et al., *G9A promotes gastric cancer metastasis by upregulating ITGB3 in a SET domain-independent manner.* Cell Death Dis, 2018. **9**(3): p. 278.
46. Zhang, N., et al., *Insufficient Radiofrequency Ablation Treated Hepatocellular Carcinoma Cells Promote Metastasis by Up-Regulation ITGB3.* J Cancer, 2017. **8**(18): p. 3742-3754.
47. O'Sullivan, D., et al., *A novel inhibitory anti-invasive MAb isolated using phenotypic screening highlights AnxA6 as a functionally relevant target protein in pancreatic cancer.* Br J Cancer, 2017. **117**(9): p. 1326-1335.
48. Oeggerli, M., et al., *Role of KCNMA1 in breast cancer.* PLoS One, 2012. **7**(8): p. e41664.
49. Khaitan, D., et al., *Role of KCNMA1 gene in breast cancer invasion and metastasis to brain.* BMC Cancer, 2009. **9**: p. 258.
50. Bloch, M., et al., *KCNMA1 gene amplification promotes tumor cell proliferation in human prostate cancer.* Oncogene, 2007. **26**(17): p. 2525-34.

51. Samuel, P., et al., *Over-expression of miR-31 or loss of KCNMA1 leads to increased cisplatin resistance in ovarian cancer cells.* Tumour Biol, 2016. **37**(2): p. 2565-73.
52. Basile, M.S., et al., *KCNMA1 Expression is Downregulated in Colorectal Cancer via Epigenetic Mechanisms.* Cancers (Basel), 2019. **11**(2).
53. Wang, Y., et al., *Ras-induced epigenetic inactivation of the RRAD (Ras-related associated with diabetes) gene promotes glucose uptake in a human ovarian cancer model.* J Biol Chem, 2014. **289**(20): p. 14225-38.
54. Wei, C.C., et al., *The pseudogene DUXAP10 promotes an aggressive phenotype through binding with LSD1 and repressing LATS2 and RRAD in non small cell lung cancer.* Oncotarget, 2017. **8**(3): p. 5233-5246.

Figures

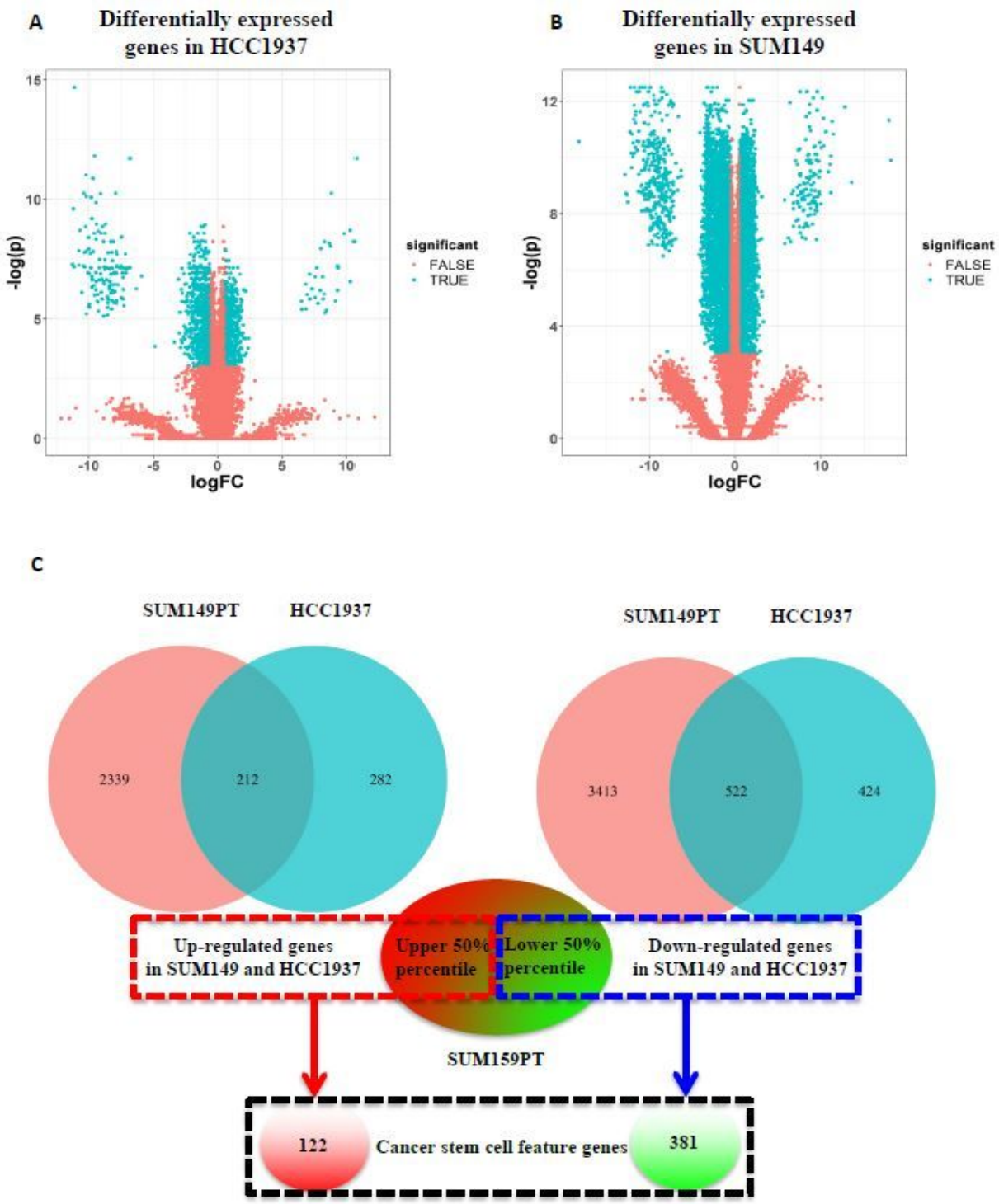

Figure 1

Visualization of genes differentially expressed in cancer stem cells. Volcano plots of differentially expressed genes in (A) SUM149 and (B) HCC1937 cells. (C) Cancer stem cell feature genes include 122 genes up-regulated in SUM149PT, HCC1937, and upper 50% percentile of SUM159PT cells, and 381 genes down-regulated in SUM149, HCC1937 cells and lower 50% percentile of SUM159PT cells.

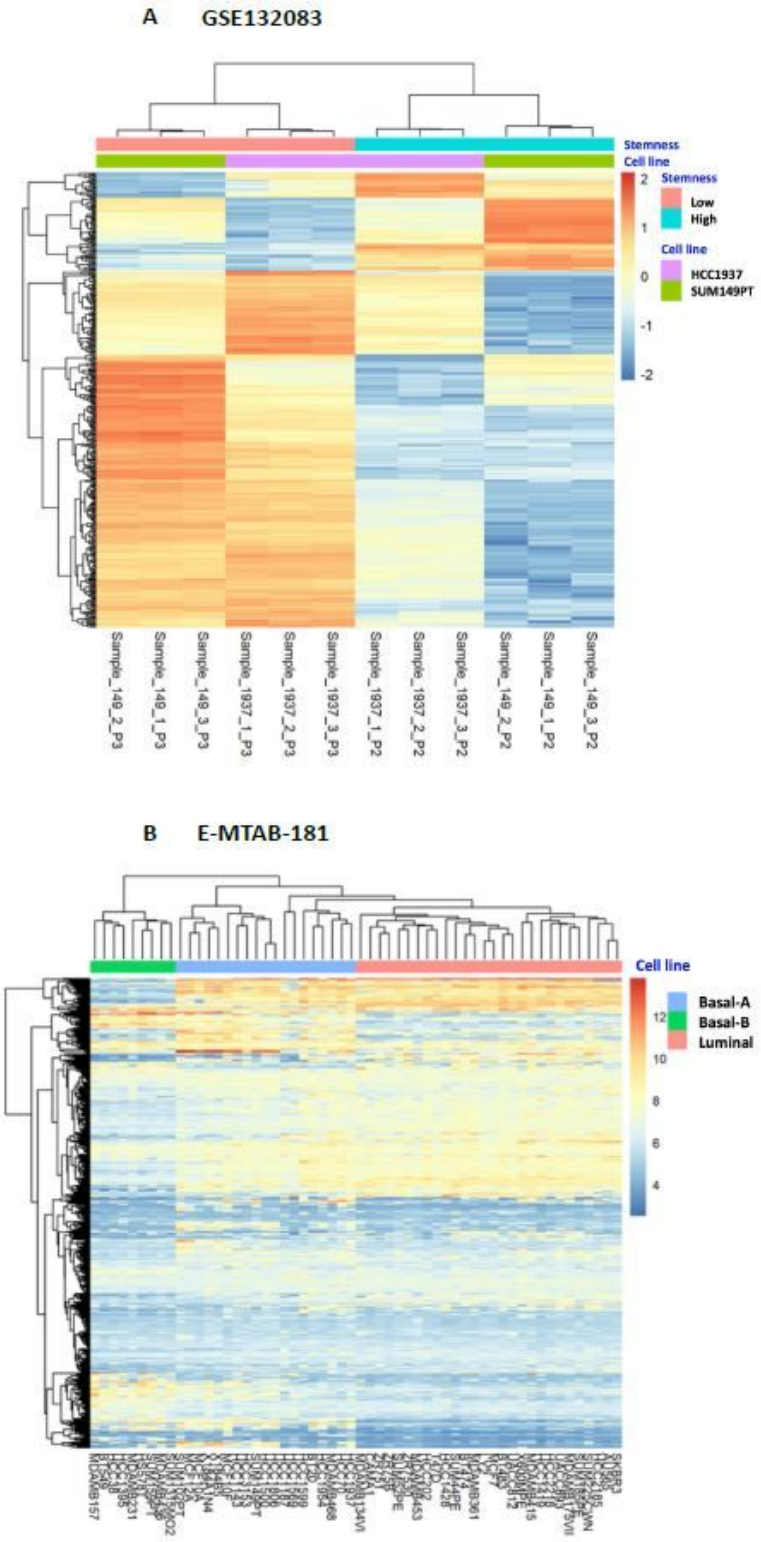

Figure 2

Heatmaps constructed using CSC feature genes and applied in (A) self-produced transcriptome data, and in (B) public GeneChip array data. The public GeneChip array data is comprised of 56 cell lines and retrieved from [7] (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-181/).

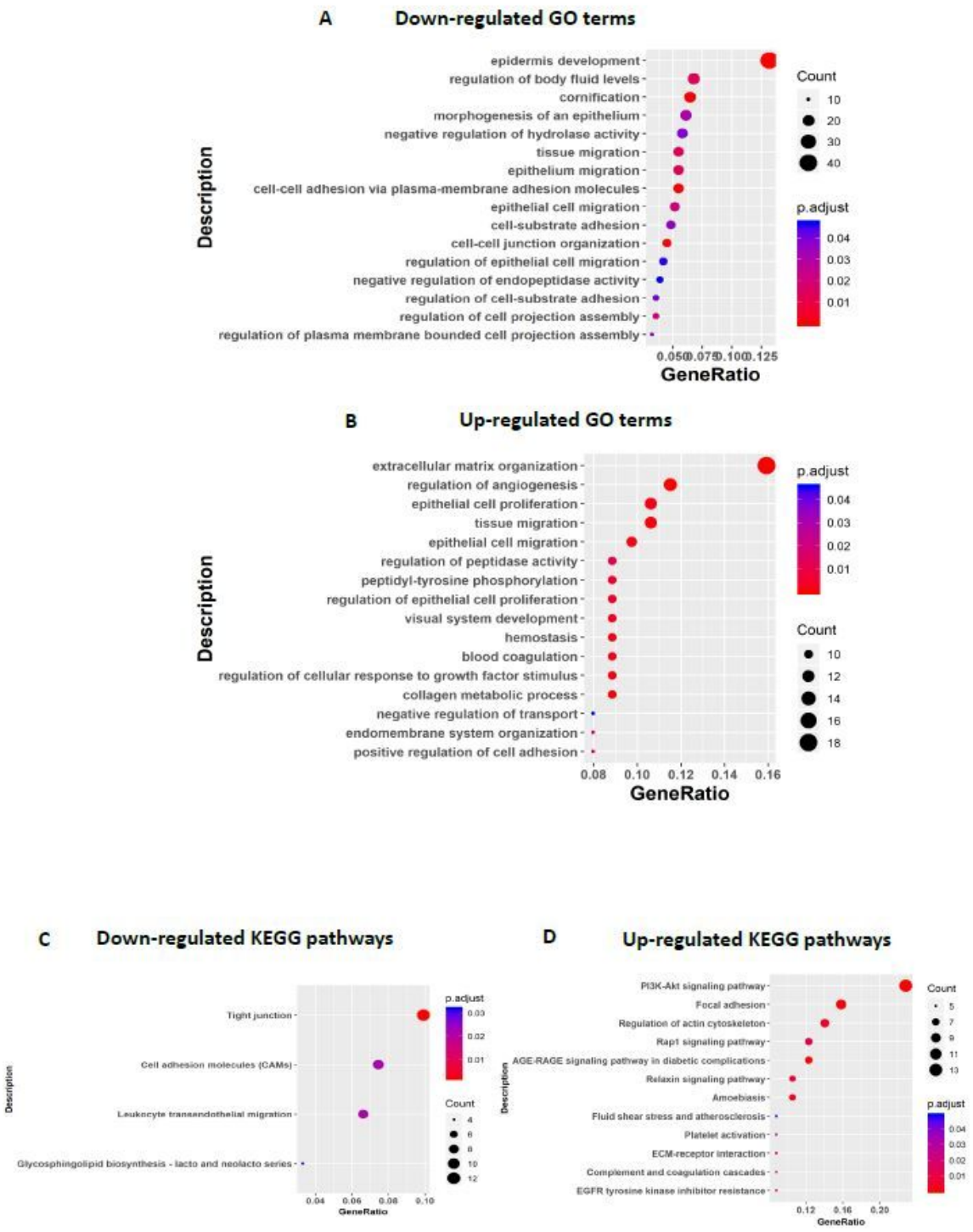

Figure 3

Biological functionalities of genes differentially expressed in cancer stem cells. Gene ontology enrichment of (A) up-regulated and (B) down-regulated genes in cancer stem cells. KEGG pathways differentially (C) up-regulated and (D) down-regulated in cancer stem cells.

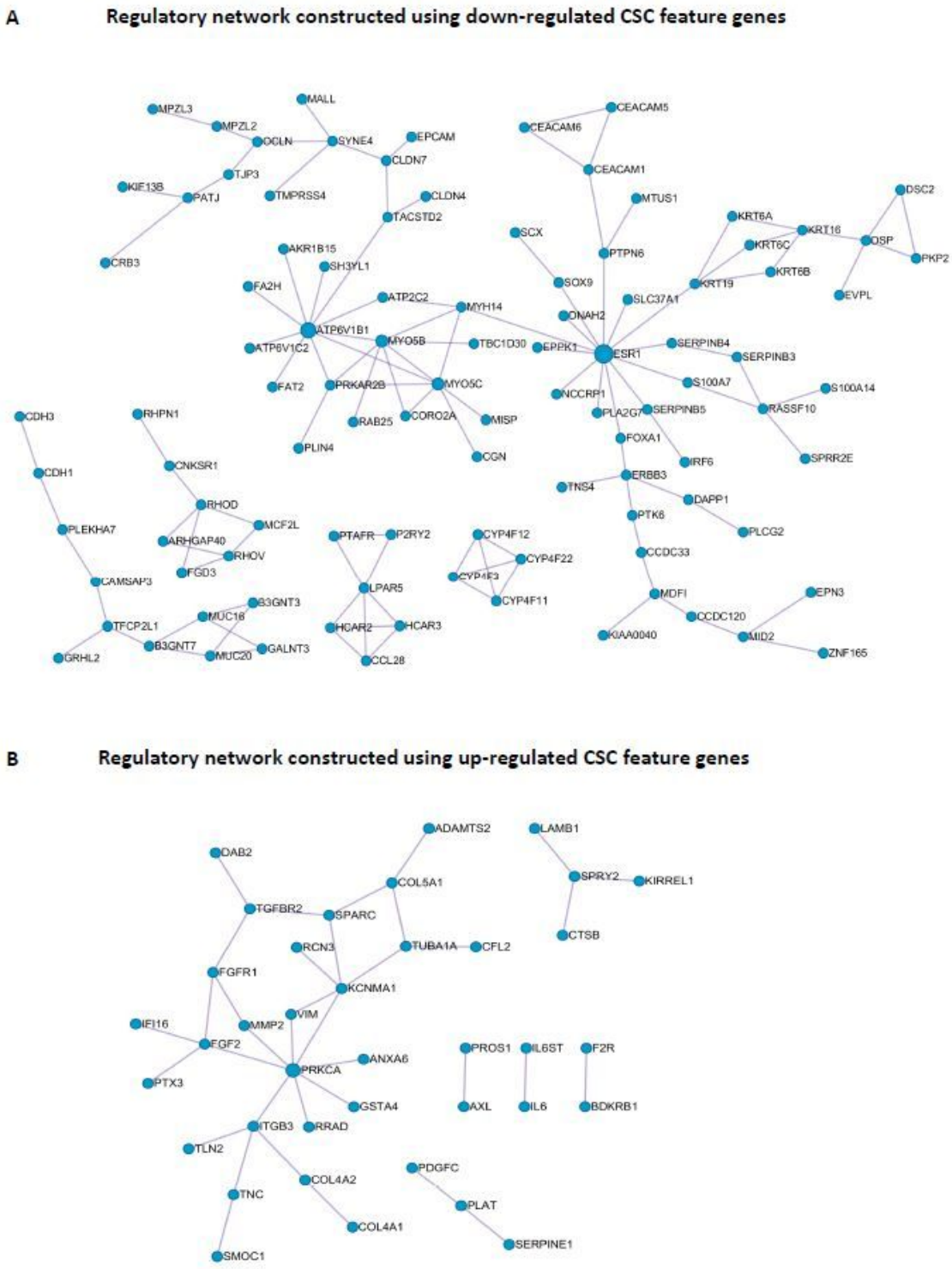

Figure 4

Protein-protein interaction networks constructed using (A) up-regulated and (B) down-regulated CSC feature genes.

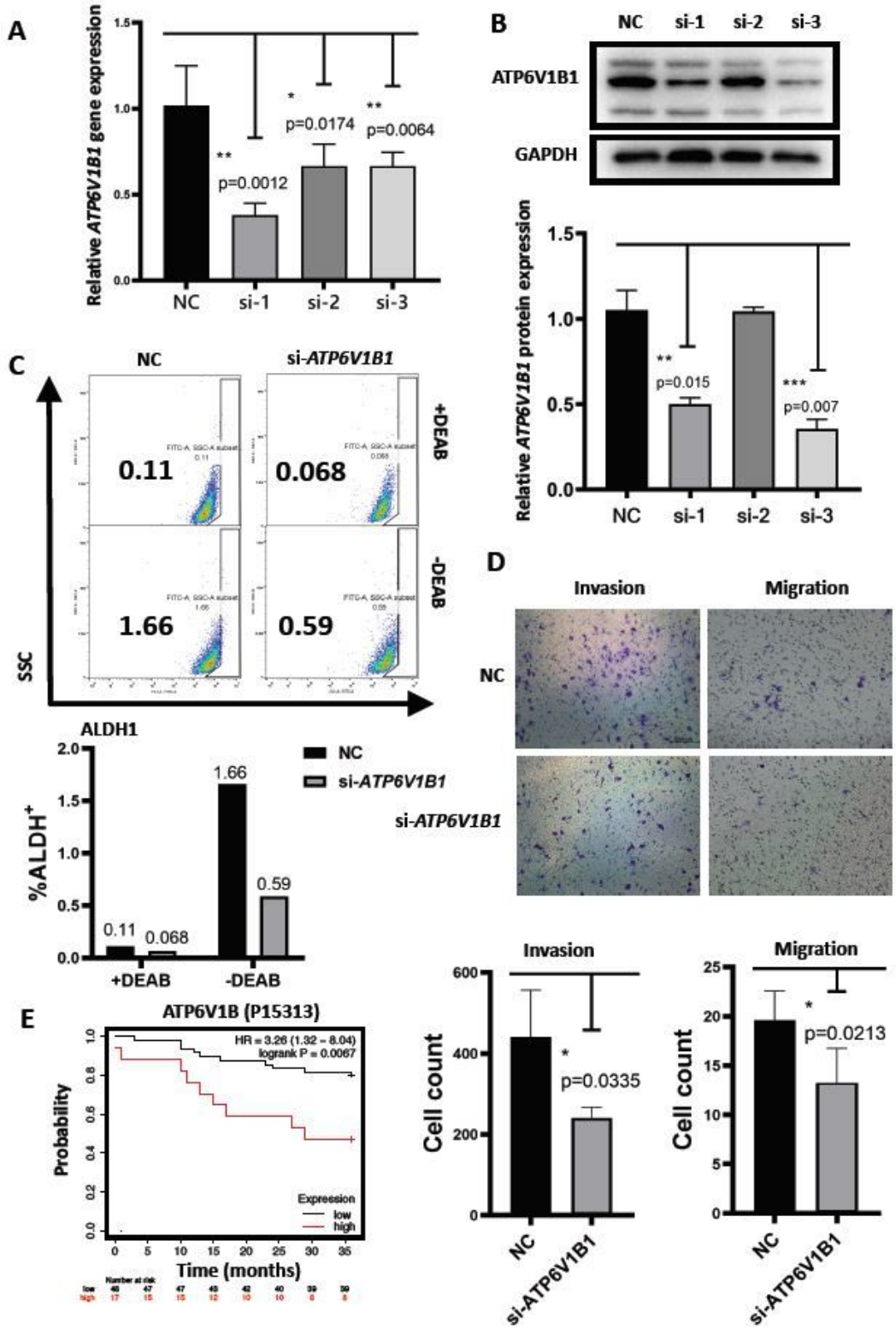

Figure 5

Functional study of a CSC feature gene ATP6V1B. (A) Silencing efficiency of siRNAs against ATP6V1B1 expression in SKBR3 cells. The qRT-PCR results showing ATP6V1B1 gene expression after the transfection of different siRNAs separately (si-1, si-2, si-3) at 36h. (B) Western blotting results and their quantifications showing ATP6V1B1 protein expression after transfecting cells with different siRNAs (si-1, si-2, si-3) at 48h. (C) Flow cytometry images and their quantifications showing CSC percentage before

and after silencing ATP6V1B1 with and without supplementing cells with DEAB (ALDH inhibitor). (D) Transwell assay results and their quantifications showing cell invasion and migration abilities before and after silencing ATP6V1B1. All experiments were conducted in SKBR3 cells. (E) Breast cancer overall survival as stratified using ATP6V1B1 protein expression. The plot was drawn using the 'Tang_2018 (n=118)' dataset in Kaplan-Meier Plotter (http://kmplot.com/analysis/index.php?p=service&cancer=breast_protein).

Supplementary Files

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

- SupplementaryTables.xlsx