DOI:10.31557/APJCN.2804.20260629

RESEARCH ARTICLE

Principal Component Analysis for the Expression of Angiogenesis-Related Genes in Breast Cancer

Nasim Beigi Boroujeni[1], Seyyed Amir Yasin Ahmadi[2]

[1]Razi Herbal Medicines Research Center, School of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran. [2]Health Systems Engineering Research Center, Milad Hospital, Tehran, Iran.

Abstract

Objectives: This study aimed to systematically classify angiogenesis-related gene expression profiles in breast cancer using principal component analysis (PCA), with the goal of identifying latent patterns and gene clusters that may underlie distinct angiogenic pathways. **Methods:** Tissue samples from 11 Iranian breast cancer patients were analyzed for the expression of eight angiogenesis-related genes (VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGFR-1, VEGFR-2, VEGFR-3, and PlGF) using real-time PCR. PCA was applied to reduce data dimensionality and uncover underlying patterns, with components retained based on eigenvalues greater than one. **Results:** PCA identified two principal components, collectively explaining 70.6% of the total variance. The first component was strongly associated with VEGF-A, VEGF-D, VEGFR-1, VEGFR-2, and PlGF, while the second component was linked to VEGF-B, VEGF-C, and VEGFR-3. This clear separation suggests that the angiogenic signaling network in breast cancer is not monolithic but rather composed of distinct gene clusters, potentially reflecting different biological functions or regulatory mechanisms. **Conclusion:** Our findings highlight the utility of PCA in revealing the molecular heterogeneity of angiogenesis-related gene expression in breast cancer. The identification of distinct gene clusters provides valuable insight into the complexity of angiogenic networks and lays the groundwork for future studies aiming to integrate multi-omics data into unified angiogenic indices. Ultimately, this approach could improve patient stratification and inform the development of targeted anti-angiogenic therapies tailored to specific molecular subtypes of breast cancer.

Keywords: Apoptosis- cancer- inflammation- systematic review

Asian Pac J Cancer Nursing, 245-248

Submission Date: 04/16/2026 Acceptance Date: 05/20/2026

Introduction

Breast cancer remains the most prevalent malignancy among women globally, with over 2.3 million new cases and 685,000 deaths reported in 2020 [1]. Recent epidemiological studies highlight a concerning rise in incidence rates, particularly among women under 50 years of age and in transitioning countries with limited healthcare infrastructure [2]. While advancements in early detection and targeted therapies have reduced mortality rates by 43% in high-income regions since 1989, disparities persist: women in low- and middle-income countries face 40% lower survival rates compared to their counterparts in developed nations [3]. These inequities underscore the urgent need to unravel molecular mechanisms driving tumor progression, particularly those involving angiogenesis a process critical for metastatic spread and treatment resistance [4].

Angiogenesis, the formation of new blood vessels, is orchestrated by complex interactions between pro-angiogenic factors like vascular endothelial growth factor (VEGF) and angiopoietins [5]. In breast cancer, elevated VEGF expression correlates with increased microvessel density and poor prognosis [6]. Despite the clinical success of anti-angiogenic agents such as bevacizumab, which targets VEGF-A, these therapies often fail to improve overall survival due to redundant signaling pathways and adaptive resistance mechanisms [7]. For instance, tumors frequently upregulate alternative angiogenic factors like fibroblast growth factor (FGF)

Corresponding Author:
Dr. Seyyed Amir Yasin Ahmadi
Health Systems Engineering Research Center, Milad Hospital, Tehran, Iran.
Email: yasin_ahmadi73@yahoo.com

or stromal-derived factor-1 (SDF-1) to bypass VEGF inhibition [7]. This biological plasticity highlights the need to identify robust biomarkers that capture the heterogeneity of angiogenic networks and enable patient stratification for personalized therapies [8].

Current classifications of angiogenesis-related genes in breast cancer remain fragmented, with limited integration of multi-omics data [8]. While over 50 genes including VEGF, HIF-1α, and PDGFR have been implicated in angiogenic pathways, their functional roles vary across molecular subtypes [6]. Traditional clustering methods struggle to resolve this complexity due to high-dimensional noise in gene expression datasets [9]. Principal component analysis (PCA) offers a promising solution by reducing data dimensionality while preserving variance patterns [10]. By transforming correlated gene expression variables into orthogonal principal components (PCs), PCA can uncover latent structures within angiogenic gene sets that correlate with clinical outcomes [10]. However, existing studies applying PCA to breast cancer transcriptomics have focused broadly on tumor subtypes rather than angiogenesis-specific signatures, leaving a critical gap in mechanistic interpretation [9].

This study aims to systematically classify angiogenesis-related genes in breast cancer using PCA-driven approaches, with three primary objectives: First, to identify conserved gene expression patterns across molecular subtypes and metastatic stages; second, to evaluate the prognostic value of PCA-derived angiogenic signatures in predicting treatment response and survival; and third, to develop a framework for integrating multi-omics data (e.g., mRNA, miRNA, and methylation profiles) into unified angiogenic indices [10]. By addressing these goals, our work will provide a computational toolkit for deconvolving angiogenic heterogeneity, ultimately guiding the development of subtype-specific anti-angiogenic therapies and combinatorial treatment strategies.

Materials and Methods

A case series study was undertaken to investigate the molecular and clinical characteristics of breast cancer tissue samples obtained from 11 patients treated at Firoozgar Teaching Hospital, which is a major academic medical center associated with Iran University of Medical Sciences. The research protocol was carefully reviewed and received official approval from the Ethics Committee of Iran University of Medical Sciences, as documented under the approval code IR.IUMS.REC.1397.983.

After the tumor was removed or a complete mastectomy was performed in the operating room, the researcher transported the excised tissue to the pathology department. In the pathology lab, approximately 4 mm of tumor tissue was separated. In cases of complete mastectomy, the tumor was simply excised with a scalpel. In partial mastectomy (tumor resection) cases, the tumor was first used for routine procedures by the pathologist, including frozen section analysis of the margins and touch preparation slides of the margins. If the size of the grossly visible tumor was sufficient, the researcher then collected a tissue sample using a scalpel. Samples of fibroadenoma and normal fibroglandular tissue were collected directly in the operating room. The collected samples were washed in phosphate-buffered saline (PBS) and immediately stored at -80°C until gene expression analysis.

Approximately 200 µg of tissue was aliquoted from each sample. The tissue was then cryogenically homogenized using liquid nitrogen in a mortar. Following this, the homogenized material was lysed in SK buffer for column-based RNA extraction, adhering to the manufacturer's protocol (Norgen, Canada). Post-extraction, RNA concentration was measured via nanodrop spectrophotometry, yielding values between 200–300 ng/µl. For cDNA synthesis, 4 µl of RNA was reverse-transcribed using random hexamer primers as specified by the kit guidelines (Yekta Tajhiz Azma, Iran). Primers were designed based on Masood et al. [11], with sequences verified through the NCBI database and Oligo7 software. To account for VEGF gene alternative splicing and multiple isoforms, primers were selected to target conserved regions across most variants. Beta-actin served as the endogenous control. Amplification was performed using SYBR Green master mix (Norgen, Canada) in a real-time PCR system to quantify gene expression levels.

The -ΔCT values based on GAPDH were used for data analysis. The target genes were VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGFR-1, VEGFR-2, VEGFR-3 and PlGF. Principal component analysis (PCA) was used for data analysis. The components with Eigenvalue more than one were considered. All the statistical analyses were performed in Stata 17 software (Stata Corp. LLC, TX, US).

Results

A total of 11 Iranian breast cancer patients were enrolled in the study, and their tissue samples were analyzed. The mean age of the patients was 46.18 years, with an age range of 31 to 64 years. estrogen receptor was positive in 8 patients. Progesterone receptor was positive in 6 patients. HER2 was positive in 5 patients. One patient was triple negative. A family history of breast cancer was reported in 3 patients. One patient had metastasis. Six patients had lymph node involvement.

According to the PCA results, two components were found with Eigenvalues 3.95 and 1.69. The cumulative explained variance was 0.706 up to these two components. The table of item-component correlation is shown (Table 1). Hence, component 1 was associated with VEGF-A, VEGF-D, VEGFR-1, VEGFR-2 and PlGF, while, component 2 was associated with VEGF-B, VEGF-C and VEGFR-3.

Discussion

Principal component analysis (PCA) is a powerful statistical technique widely used in oncology for reducing the dimensionality of complex gene expression datasets while preserving the underlying patterns of variance. In the context of this study, PCA was applied to the expression

Table 1. Correlation of Variables with Components after PCA

Variable	Component 1	Component 2
VEGF-A	0.361*	-0.177
VEGF-B	0.252	-0.390*
VEGF-C	0.125	0.684*
VEGF-D	0.451*	-0.062
VEGFR-1	0.458*	0.012
VEGFR-2	0.359*	0.158
VEGFR-3	0.256	0.523*
PlGF	0.428*	-0.215

* Absolute value > 0.3

profiles of eight angiogenesis-related genes VEGF-A, VEGF-B, VEGF-C, VEGF-D, VEGFR-1, VEGFR-2, VEGFR-3, and PlGF across 11 breast cancer tissue samples. The analysis revealed two principal components with eigenvalues greater than one, which together accounted for approximately 70.6% of the total variance in the data. The first component was predominantly associated with VEGF-A, VEGF-D, VEGFR-1, VEGFR-2, and PlGF, while the second component was most strongly linked to VEGF-B, VEGF-C, and VEGFR-3. This clear separation suggests that the angiogenic signaling network in breast cancer is not monolithic but rather composed of distinct gene clusters that may reflect different biological functions or regulatory mechanisms. Such findings are in line with recent research emphasizing the utility of PCA for uncovering latent structures within high-dimensional omics data, particularly when traditional clustering methods fail to resolve the inherent complexity and noise in gene expression profiles [12].

The identification of these two principal components provides valuable insight into the molecular heterogeneity of angiogenesis in breast cancer. The association of specific VEGF and receptor genes with each component implies that there may be separate pathways or co-regulated gene modules involved in tumor vascularization. For example, the genes clustered in the first component may be involved in primary angiogenic signaling and tumor progression, while those in the second component could play roles in alternative or compensatory angiogenic mechanisms, such as lymphangiogenesis or adaptive responses to therapy. This PCA-driven classification not only aids in deconvolving the complexity of angiogenic networks but also lays the groundwork for future studies aiming to integrate multi-omics data into unified angiogenic indices. Ultimately, such approaches could enhance patient stratification, improve prognostic assessment, and inform the development of targeted anti-angiogenic therapies tailored to specific molecular subtypes of breast cancer [12].

The current study and the RF-PCA approach (Front. Genet. 2020) both employ principal component analysis for breast cancer data but differ significantly in objectives and methodology. While RF-PCA focuses on accelerating categorical data classification through a hybrid random forest-PCA algorithm optimized for computational efficiency in diagnostic applications, the present work applies traditional PCA specifically to angiogenesis-related gene expression profiles to uncover biological patterns rather than improve processing speed. The RF-PCA framework demonstrated superior performance in handling high-dimensional clinical datasets, achieving 94% accuracy in tumor subtype classification, whereas our angiogenesis-focused PCA identified two biologically distinct components (VEGF-A/D/VEGFR-1/2/PlGF vs. VEGF-B/C/VEGFR-3) explaining 70.6% variance, providing mechanistic insights into angiogenic pathway segregation. This contrast highlights how PCA's utility spans both diagnostic optimization (RF-PCA) and biological discovery (current study), with the former prioritizing clinical decision support and the latter elucidating molecular drivers of tumor vascularization [10].

The present study, despite its valuable insights into the expression patterns of angiogenesis-related genes in breast cancer using PCA, is subject to several limitations. Foremost among these is the small sample size, with only 11 patients included in the analysis, which restricts the statistical power and generalizability of the findings [13]. This limited cohort may not fully capture the biological and clinical heterogeneity of breast cancer, nor account for the variability seen across different molecular subtypes or disease stages. Additionally, the study's reliance on tissue samples from a single institution and its focus on a specific set of angiogenesis-related genes may further limit the broader applicability and robustness of the identified gene signatures. The absence of external validation in independent datasets or integration with multi-omics data, as well as the lack of longitudinal follow-up for clinical outcomes, also constrains the ability to assess the prognostic or predictive value of the PCA-derived components. These limitations underscore the need for larger, multicenter studies with comprehensive clinicopathological and molecular profiling to validate and extend these preliminary results.

In conclusion, this study highlights the utility of PCA as a robust tool for uncovering molecular heterogeneity within angiogenesis-related gene expression profiles in breast cancer. The identification of two distinct principal components one strongly associated with VEGF-A, VEGF-D, VEGFR-1, VEGFR-2, and PlGF, and the other linked to VEGF-B, VEGF-C, and VEGFR-3 suggests the presence of separate angiogenic pathways or co-regulated gene modules that may underlie tumor vascularization and progression. These findings not only advance our understanding of the complexity and diversity of angiogenic signaling networks in breast cancer but also provide a foundation for future research aiming to integrate multi-omics data into unified angiogenic indices. Ultimately, such approaches could improve patient stratification, refine prognostic assessments, and inform the development of targeted anti-angiogenic therapies tailored to specific molecular subtypes of breast cancer.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

The author acknowledges the use of Perplexity Pro for research, information synthesis, and initial drafting support during the preparation of this manuscript.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram L, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians. 2021 05;71(3). https://doi.org/10.3322/caac.21660
2. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, Vignat J, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast (Edinburgh, Scotland). 2022 Dec;66. https://doi.org/10.1016/j.breast.2022.08.010
3. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, Jemal A, Siegel RL. Breast Cancer Statistics, 2022. CA: a cancer journal for clinicians. 2022 Nov;72(6). https://doi.org/10.3322/caac.21754
4. Folkman J. Tumor angiogenesis: therapeutic implications. The New England journal of medicine. 1971 Nov 18;285(21). https://doi.org/10.1056/NEJM197111182852108
5. Byrne GJ, McDowell G, Agarawal R, Sinha G, Kumar S, Bundred NJ. Serum vascular endothelial growth factor in breast cancer. Anticancer research. 2007 Oct;27(5B).
6. Ferrara N, Kerbel RS. Angiogenesis as a therapeutic target. Nature. 2005 Dec 15;438(7070). https://doi.org/10.1038/nature04483
7. Ayoub NM, Sardiah S, Al-Share QY, Alkader MS. Exploring angiogenic pathways in breast cancer: Clinicopathologic correlations and prognostic implications based on gene expression profiles from a large-scale genomic dataset. PloS one. 2024 09 20;19(9). https://doi.org/10.1371/journal.pone.0310557
8. Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics (Oxford, England). 2001 09;17(9). https://doi.org/10.1093/bioinformatics/17.9.763
9. Ma S, Dai Y. Principal component analysis based methods in bioinformatics studies. Briefings in bioinformatics. 2011 Nov;12(6). https://doi.org/10.1093/bib/bbq090
10. Bian K, Zhou M, Hu F, Lai W. RF-PCA: A New Solution for Rapid Identification of Breast Cancer Categorical Data Based on Attribute Selection and Feature Extraction. Frontiers in genetics. 2020 09 09;11. https://doi.org/10.3389/fgene.2020.566057
11. Masood R, Cesarman E, Smith DL, Gill PS, Flore O. Human herpesvirus-8-transformed endothelial cells have functionally activated vascular endothelial growth factor/vascular endothelial growth factor receptor. The American journal of pathology. 2002 01;160(1). https://doi.org/10.1016/S0002-9440(10)64344-1
12. Xu P, Liu S, Song S, Yao X, Li X, Zhang J, Liu Y, et al. Identification and validation of a novel angiogenesis-related gene signature for predicting prognosis in gastric adenocarcinoma. Frontiers in oncology. 2023 01 16;12. https://doi.org/10.3389/fonc.2022.965102
13. Sun Z, Wang CY, Lawson DA, Kwek S, Velozo HG, Owyong M, Lai MD, et al. Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells. Oncotarget. 2017 Dec 29;9(13). https://doi.org/10.18632/oncotarget.23760