International Journal of Clinical Science and Medical Research

ISSN(print): 2770-5803, ISSN(online): 2770-582X

Volume 06 Issue 04 April 2026

DOI: https://doi.org/10.55677/IJCSMR/V6I4-07/2026, Impact Factor: 8.005

Page No : 119-123

Evaluating Dataset-Specific Variability and Reproducibility of CDH1 (E-cadherin) Expression in Breast Cancer Using GEO2R

Muhammad Salman Ashghar[1], Muhammad Hamza Mubarak[2]

[1,2] Faculty of Medicine, Universitas Islam Indonesia

ABSTRACT

Breast cancer is a major global health burden, with alterations in gene expression contributing to tumor progression. The CDH1 gene, encoding E-cadherin, plays a critical role in maintaining epithelial integrity, and its loss has been associated with increased invasiveness and metastasis in breast cancer. This study aimed to evaluate the differential expression of CDH1 using the GSE15852 dataset. Gene expression data were obtained from the Gene Expression Omnibus (GEO) and analyzed using the GEO2R platform, comparing tumor and normal samples. Differential expression was assessed based on log fold change (logFC) and statistical significance. The analysis revealed that CDH1 is significantly downregulated in breast cancer, with a logFC of −2.386, a p-value of 2.72×10^{-11}, and an adjusted p-value of 3.68×10^{-14}. These findings indicate a substantial and statistically significant reduction in CDH1 expression in tumor tissues. In conclusion, CDH1 is markedly downregulated in breast cancer, supporting its role as a tumor suppressor. GEO2R provides a reliable approach for preliminary gene expression analysis; however, further validation is required to confirm these findings.

Published Online: April 27, 2026

KEYWORDS:

Breast Cancer, CDH1, Gene Expression Analysis, GEO2R, Differential Expression

INTRODUCTION

Breast cancer is one of the most commonly diagnosed malignancies and a leading cause of cancer-related mortality among women worldwide. Despite significant advances in early detection and targeted therapies, the heterogeneity of breast cancer continues to pose major challenges in diagnosis, prognosis, and treatment selection. At the molecular level, breast cancer is characterized by complex alterations in gene expression that drive tumor initiation, progression, and metastasis. Understanding these gene expression patterns is essential for identifying reliable biomarkers and therapeutic targets, particularly in the era of precision medicine (Narod et al., 2024).

CDH1 (cadherin 1) is one of the genes that are involved in the pathogenesis of breast cancer. CDH1 is a gene that produces E-cadherin, a calcium-dependent transmembrane glycoprotein which ensures cell to cell adhesion and the structure of the tissue in the epithelial cell. E-cadherin is an important constituent of adherens junctions, which serves to stabilize intercellular contact between e-cadherin and catenins to regulate intracellular signaling pathways. The correct functionality and expression of E-cadherin is critical to the integrity of the epithelial tissues. Loss or dysfunction of CDH1 interferes with cellular adhesion resulting in enhanced cellular motility, invasion as well as metastatic dissemination-characteristics of cancer progression (Algebaly et al., 2021).

In breast cancer, down-regulation or impaired activity of the CDH1 gene has been consistently linked with tumor aggressiveness specifically in invasive and metastatic disease subtypes. The CDH1 downregulation is linked closely to epithelial-mesenchymal transition (EMT) whereby epithelial cells lose their polarity and adhesion and develop mesenchymal-like properties that enable them to migrate more effectively. This conversion promotes tumor invasion and leads to poor clinical outcome. The mechanisms that regulate the down-regulation of CDH1 include genetic mutations, hypermethylation of promoters, transcriptional repression and post-translational modifications. CDH1 is

Corresponding Author: Muhammad Salman Ashghar

Cite this Article: Ashghar, M.S., Mubarak, M.H. (2026). Evaluating Dataset-Specific Variability and Reproducibility of CDH1 (E-cadherin) Expression in Breast Cancer Using GEO2R. International Journal of Clinical Science and Medical Research, 6(4), 119-123. https://doi.org/10.55677/IJCSMR/V6I4-07/2026

considered an important tumor suppressor gene in breast cancer since it has a central role in ensuring the integrity of the epithelial cells and inhibiting tumor growth (Dines et al., 2020).

The development of high-throughput technologies, especially, microarray-based gene expression profiling has facilitated the examination of the expression patterns of a large number of genes under varying biological conditions. The Gene Expression Omnibus (GEO) is a publicly accessible repository of an extensive amount of transcriptomic data produced using disparate experimental designs and groups of patients. These data provide useful prospects of secondary data analysis whereby researchers can be capable of examining the changes in gene expression without necessarily having to generate primary data. Nonetheless, inconsistency in datasets (due to sample heterogeneity, experimental systems, and data pre-processing) may affect reproducibility and interpretation of findings (Sadeghi et al., 2020).

To facilitate accessible analysis of GEO datasets, the GEO2R online tool has been developed as an interactive platform for comparing gene expression between predefined sample groups. GEO2R utilizes statistical methods based on the limma (Linear Models for Microarray Data) framework to calculate differential expression, providing outputs such as log fold change (logFC), p-values, and adjusted p-values. Its user-friendly interface allows researchers to perform differential gene expression analysis without requiring advanced computational expertise. As a result, GEO2R has become widely used for preliminary bioinformatics investigations and hypothesis generation (Silvestri et al., 2020).

Even though tools such as GEO2R are convenient, there are still concerns about the variability and reproducibility of gene expression results among datasets. Inconsistent results can occur even in the study of the same gene and related disease situations due to differences in study design, sample heterogeneity, and technical reasons. Assessing dataset-specific variability is thus critical to establish the strength of observed patterns of gene expression. Applied to CDH1, the evaluation of its levels in individual datasets can help analyze its stability as a biomarker and its applicability in a particular cohort (Kotsopoulos et al., 2024).

In this study, we aim to evaluate the expression profile of CDH1 in breast cancer using a GEO2R-based analytical approach, focusing on the GSE15852 microarray dataset. By comparing tumor and normal samples within this dataset, we seek to determine the direction and significance of CDH1 differential expression. Furthermore, this study highlights the importance of considering dataset-specific factors when interpreting gene expression results derived from public repositories. Through this approach, we aim to contribute to the understanding of CDH1's role in breast cancer while emphasizing the strengths and limitations of GEO2R-based analyses in bioinformatics research (Gruber et al., 2022).

CDH1 (cadherin 1) and Breast Cancer:

Breast cancer is a heterogeneous disease characterized by diverse histopathological and molecular subtypes, making it a complex condition to diagnose and treat. It accounts for a substantial proportion of cancer incidence and mortality worldwide, with its burden continuing to rise in both developed and developing countries. The progression of breast cancer is driven by a combination of genetic, epigenetic, and environmental factors that collectively alter cellular behavior. Among these, changes in gene expression play a central role in regulating key processes such as cell proliferation, differentiation, apoptosis, and metastasis. The identification and analysis of differentially expressed genes are therefore essential for understanding the molecular mechanisms underlying breast cancer and for discovering potential diagnostic and prognostic biomarkers (Albert et al., 2020).

The gene, CDH1 (cadherin 1) is one of the most important genes in the maintenance of normal epithelial structure and it encodes the protein known as E-cadherin. E-cadherin is a calcium-dependent cell-cell adhesion protein (transmembrane glycoprotein) that plays a role in epithelial tissues. It is a key constituent of adherens junctions, in which it connects with intracellular proteins like 2 -catenin and 2 -catenin to stabilize the cytoskeleton and to preserve tissue integrity. These interactions do not only maintain structural cohesion, but also engage E-cadherin in intracellular signaling pathways controlling cell growth and differentiation. E-cadherin expression and localization is thus crucial in preserving normal epithelial functions (Lee et al., 2020).

The downregulation or loss of CDH1 has been well-known as a marker of tumor progression in the cancer context, especially breast cancer. The lower level of E-cadherin expression interferes with cellular adhesion and results in growth of cellular detachment, migration and invasion. This de-adhesion is one of the major stages in an epithelial-mesenchymal transition (EMT), a biological event during which epithelial cells gain mesenchymal functions, such as increased motility and anti-apoptotic ability. EMT is closely linked with metastasis of cancer and bad clinical outcomes. A loss of CDH1 expression has been commonly reported in invasive ductal carcinoma and lobular carcinoma in breast cancer, demonstrating the involvement in facilitating tumor aggressiveness (Lubinski et al., 2024).

Several molecular pathways are involved in downregulation of CDH1 in breast cancer. These comprise genetic mutations in the CDH1 gene, epigenetic changes like promoter

hypermethylation and transcriptional suppression by EMT-inducing transcriptional factors, including Snail, Slug, and Twist. Also, the E-cadherin functions can be further worsened by post-translational modifications and alterations in protein trafficking. The combination of the mechanisms leads to the down-regulated expression or functional suppression of CDH1, which promotes tumor development and spread. Due to its key position in cancer biology, CDH1 has been of significant interest with respect to being used as a prognostic marker and a potential therapeutic target (Markowski & Antonarakis, 2020).

With the development of methods of high-throughput gene expression, including microarrays, cancer research has been transformed as it has become possible to analyze thousands of genes at once. These technologies have generated vast amounts of transcriptomic data, much of which is publicly available in repositories like the Gene Expression Omnibus (GEO). GEO is a useful tool that can be used by the researchers to access the datasets covering a broad spectrum of experimental conditions, tissues, and disease states. Using these datasets, researchers are able to conduct secondary analyses, to determine differentially expressed genes and to study molecular pathways of disease processes (Song et al., 2020).

To facilitate the analysis of GEO datasets, the GEO2R tool has been developed as an accessible, web-based platform for differential gene expression analysis. GEO2R allows users to compare predefined groups of samples within a dataset and generates statistical outputs, including log fold change (logFC), p-values, and adjusted p-values. The tool is based on the limma statistical framework, which applies linear models and empirical Bayes methods to improve the reliability of differential expression analysis. Due to its ease of use and minimal computational requirements, GEO2R is widely used for preliminary bioinformatics studies and hypothesis generation (Cheng et al., 2024).

However, despite the advantages of using publicly available datasets and tools like GEO2R, challenges remain in ensuring the reliability and reproducibility of gene expression findings. Variability between datasets can arise from differences in sample size, patient demographics, experimental design, microarray platforms, and data processing methods. Such variability may lead to inconsistencies in the identification of differentially expressed genes, even for well-characterized genes like CDH1. Therefore, careful interpretation of results is necessary, particularly when analyses are limited to a single dataset (Abida et al., 2020).

In this context, analyzing CDH1 expression within a specific dataset provides an opportunity to assess its role in breast cancer while also highlighting the potential influence of dataset-specific factors. By utilizing GEO2R to evaluate CDH1 expression in the GSE15852 dataset, this study aims to contribute to the growing body of evidence supporting the involvement of CDH1 in breast cancer progression. At the same time, it underscores the importance of considering variability and methodological limitations when interpreting results derived from bioinformatics analyses (de Jonge et al., 2021).

METHODOLOGY

This study employed a retrospective bioinformatics design to evaluate the differential expression of the CDH1 gene in breast cancer using publicly available microarray data. Gene expression data were obtained from the Gene Expression Omnibus (GEO), a repository maintained by the National Center for Biotechnology Information (NCBI). The dataset GSE15852 was selected for analysis due to the availability of both breast cancer (tumor) and normal breast tissue samples, allowing direct comparison of gene expression profiles. Relevant dataset information, including sample classification and experimental conditions, was accessed through the GEO database interface.

The analysis was conducted using the GEO2R online tool, an interactive platform that enables comparison between predefined groups within a dataset. Samples in GSE15852 were divided into two groups based on their biological status: tumor samples and normal control samples. No additional preprocessing or normalization steps were performed beyond the default processing integrated within GEO2R, ensuring methodological consistency and reproducibility within a GEO2R-only analytical framework.

Differential gene expression analysis was performed using GEO2R, which is based on the limma (Linear Models for Microarray Data) statistical package. This method applies linear modeling and empirical Bayes moderation to estimate gene-wise variance and improve statistical reliability. For each gene, including CDH1, GEO2R generated key statistical parameters, including log fold change (logFC), p-value, adjusted p-value, and t-statistic. The logFC value was used to determine both the magnitude and direction of gene expression, where negative values indicate downregulation and positive values indicate upregulation in tumor samples relative to normal controls. Statistical significance was assessed using p-values, while adjusted p-values were calculated using the Benjamini–Hochberg false discovery rate (FDR) method to account for multiple comparisons.

The gene CDH1 (cadherin 1) was identified within the GEO2R output using its official gene symbol and corresponding probe ID (201131_s_at). The associated statistical values, including logFC, p-value, adjusted p-value, and t-statistic, were extracted for analysis. Differential expression was interpreted based on standard criteria, with statistical significance defined as a p-value less than 0.05 and strong significance indicated by an adjusted p-value below

0.05. Additionally, a logFC threshold of ≥1 or ≤−1 was considered indicative of biologically meaningful changes in gene expression.

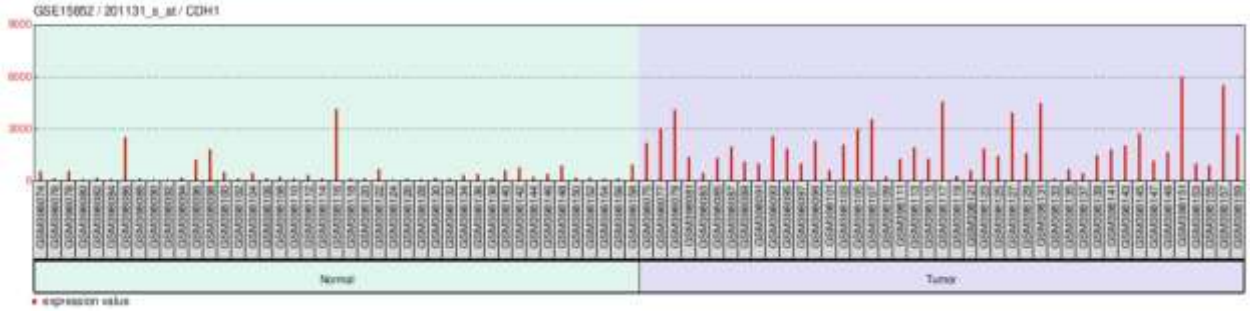

Figure 01: Expression Profile of CDH1 in GSE15852

The results were summarized in tabular form, presenting the extracted statistical parameters for CDH1, and supported by graphical visualization to illustrate differences in expression between tumor and normal samples. As this study was limited to a single dataset and relied exclusively on GEO2R for analysis, no external validation or additional computational methods were applied. Consequently, the findings should be interpreted with consideration of potential dataset-specific variability and methodological limitations.

RESULTS

Differential gene expression analysis of the GSE15852 dataset was performed using the GEO2R platform to compare breast cancer (tumor) samples with normal breast tissue controls. The analysis identified the CDH1 gene (probe ID: 201131_s_at) as significantly differentially expressed between the two groups.

CDH1 exhibited a log fold change (logFC) of −2.386, indicating a marked reduction in expression in tumor samples relative to normal controls. The statistical analysis demonstrated a p-value of 2.72×10^{-11}, confirming a highly significant difference in expression. After adjustment for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) method, the adjusted p-value was 3.68×10^{-14}, further supporting the robustness of the result. Additionally, the gene showed a t-statistic of −8.94, reflecting a strong standardized difference in expression between the two groups.

Table 1: Differential Expression of CDH1 in GSE15852

Gene	logFC	p-value	adj. p-value	t-statistic	Regulation
CDH1	-2.386	2.72×10^{-11}	3.68×10^{-14}	-8.94	Downregulated

The magnitude of downregulation (|logFC| > 2) indicates a substantial decrease in CDH1 expression in breast cancer samples. Based on the predefined criteria for significance (p-value < 0.05 and |logFC| ≥ 1), CDH1 meets both statistical and biological thresholds for differential expression. These findings confirm that CDH1 is significantly and consistently downregulated in the GSE15852 breast cancer dataset.

Overall, the results demonstrate a strong and statistically significant alteration in CDH1 expression, supporting its potential role in breast cancer pathology within this dataset.

CONCLUSION

CDH1 is significantly downregulated in breast cancer in the GSE15852 dataset (logFC = −2.386; adjusted p-value = 3.68×10^{-14}), supporting its role as a tumor suppressor. These findings highlight its potential involvement in tumor progression, although further validation is required.

REFERENCES

1. Abida, W., Patnaik, A., Campbell, D., Shapiro, J., Bryce, A. H., McDermott, R., Sautois, B., Vogelzang, N. J., Bambury, R. M., Voog, E., Zhang, J., Piulats, J. M., Ryan, C. J., Merseburger, A. S., Daugaard, G., Heidenreich, A., Fizazi, K., Higano, C. S., Krieger, L. E., … Chowdhury, S. (2020). Rucaparib in Men With Metastatic Castration-Resistant Prostate Cancer Harboring a *BRCA1* or *BRCA2* Gene Alteration. *Journal of Clinical Oncology*, *38*(32), 3763–3772. https://doi.org/10.1200/JCO.20.01035
2. Albert, C., Haase, M., Albert, A., Kropf, S., Bellomo, R., Westphal, S., Westerman, M., Braun-Dullaeus, R. C., & Haase-Fielitz, A. (2020). Urinary Biomarkers may Complement the Cleveland Score for Prediction of Adverse Kidney Events After Cardiac Surgery: A Pilot Study. *Annals of Laboratory Medicine*, *40*(2), 131–141. https://doi.org/10.3343/alm.2020.40.2.131

3. Algebaly, A. S., Suliman, R. S., & Al-Qahtani, W. S. (2021). Comprehensive study for BRCA1 and BRCA2 entire coding regions in breast cancer. *Clinical and Translational Oncology*, *23*(1), 74–81. https://doi.org/10.1007/s12094-020-02385-9
4. Cheng, H. H., Shevach, J. W., Castro, E., Couch, F. J., Domchek, S. M., Eeles, R. A., Giri, V. N., Hall, M. J., King, M.-C., Lin, D. W., Loeb, S., Morgan, T. M., Offit, K., Pritchard, C. C., Schaeffer, E. M., Szymaniak, B. M., Vassy, J. L., Katona, B. W., & Maxwell, K. N. (2024). *BRCA1, BRCA2* , and Associated Cancer Risks and Management for Male Patients. *JAMA Oncology*, *10*(9), 1272. https://doi.org/10.1001/jamaoncol.2024.2185
5. de Jonge, M. M., de Kroon, C. D., Jenner, D. J., Oosting, J., de Hullu, J. A., Mourits, M. J. E., Gómez Garcia, E. B., Ausems, M. G. E. M., Margriet Collée, J., van Engelen, K., van de Beek, I., Smit, V. T. H. B. M., Rookus, M. A., de Bock, G. H., van Leeuwen, F. E., Bosse, T., Dekkers, O. M., & van Asperen, C. J. (2021). Endometrial Cancer Risk in Women With Germline *BRCA1* or *BRCA2* Mutations: Multicenter Cohort Study. *JNCI: Journal of the National Cancer Institute*, *113*(9), 1203–1211. https://doi.org/10.1093/jnci/djab036
6. Dines, J. N., Shirts, B. H., Slavin, T. P., Walsh, T., King, M.-C., Fowler, D. M., & Pritchard, C. C. (2020). Systematic misclassification of missense variants in BRCA1 and BRCA2 "coldspots." *Genetics in Medicine*, *22*(5), 825–830. https://doi.org/10.1038/s41436-019-0740-6
7. Gruber, J. J., Afghahi, A., Timms, K., DeWees, A., Gross, W., Aushev, V. N., Wu, H.-T., Balcioglu, M., Sethi, H., Scott, D., Foran, J., McMillan, A., Ford, J. M., & Telli, M. L. (2022). A phase II study of talazoparib monotherapy in patients with wild-type BRCA1 and BRCA2 with a mutation in other homologous recombination genes. *Nature Cancer*, *3*(10), 1181–1191. https://doi.org/10.1038/s43018-022-00439-1
8. Kotsopoulos, J., Gronwald, J., Huzarski, T., Møller, P., Pal, T., McCuaig, J. M., Singer, C. F., Karlan, B. Y., Aeilts, A., Eng, C., Eisen, A., Bordeleau, L., Foulkes, W. D., Tung, N., Couch, F. J., Fruscio, R., Neuhausen, S. L., Zakalik, D., Cybulski, C., … Rastelli, A. (2024). Bilateral Oophorectomy and All-Cause Mortality in Women With *BRCA1* and *BRCA2* Sequence Variations. *JAMA Oncology*, *10*(4), 484. https://doi.org/10.1001/jamaoncol.2023.6937
9. Lee, Y.-C., Lee, Y.-C., Li, C.-Y., Lee, Y.-L., & Chen, B.-L. (2020). BRCA1 and BRCA2 Gene Mutations and Lung Cancer Risk: A Meta-Analysis. *Medicina*, *56*(5), 212. https://doi.org/10.3390/medicina56050212
10. Lubinski, J., Kotsopoulos, J., Moller, P., Pal, T., Eisen, A., Peck, L., Karlan, B. Y., Aeilts, A., Eng, C., Bordeleau, L., Foulkes, W. D., Tung, N., Couch, F. J., Fruscio, R., Ramon y Cajal, T., Singer, C. F., Neuhausen, S. L., Zakalik, D., Cybulski, C., … Blum, J. (2024). MRI Surveillance and Breast Cancer Mortality in Women With *BRCA1* and *BRCA2* Sequence Variations. *JAMA Oncology*, *10*(4), 493. https://doi.org/10.1001/jamaoncol.2023.6944
11. Markowski, M. C., & Antonarakis, E. S. (2020). BRCA1 Versus BRCA2 and PARP Inhibitor Sensitivity in Prostate Cancer: More Different Than Alike? *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *38*(32), 3735–3739. https://doi.org/10.1200/JCO.20.02246
12. Narod, S. A., Metcalfe, K., Finch, A., Chan, A.-W., Armel, S. R., Aeilts, A., Eisen, A., Karlan, B., Bordeleau, L., Tung, N., Foulkes, W. D., Neuhausen, S. L., Eng, C., Olopade, O., Zakalik, D., Couch, F., Cullinane, C., Pal, T., Sun, P., … Cohen, S. (2024). The risk of skin cancer in women who carry BRCA1 or BRCA2 mutations. *Hereditary Cancer in Clinical Practice*, *22*(1), 7. https://doi.org/10.1186/s13053-024-00277-5
13. Sadeghi, F., Asgari, M., Matloubi, M., Ranjbar, M., Karkhaneh Yousefi, N., Azari, T., & Zaki-Dizaji, M. (2020). Molecular contribution of BRCA1 and BRCA2 to genome instability in breast cancer patients: review of radiosensitivity assays. *Biological Procedures Online*, *22*(1), 23. https://doi.org/10.1186/s12575-020-00133-5
14. Silvestri, V., Leslie, G., Barnes, D. R., Agnarsson, B. A., Aittomäki, K., Alducci, E., Andrulis, I. L., Barkardottir, R. B., Barroso, A., Barrowdale, D., Benitez, J., Bonanni, B., Borg, A., Buys, S. S., Caldés, T., Caligo, M. A., Capalbo, C., Campbell, I., Chung, W. K., … Ottini, L. (2020). Characterization of the Cancer Spectrum in Men With Germline *BRCA1* and *BRCA2* Pathogenic Variants. *JAMA Oncology*, *6*(8), 1218. https://doi.org/10.1001/jamaoncol.2020.2134
15. Song, Y., Barry, W. T., Seah, D. S., Tung, N. M., Garber, J. E., & Lin, N. U. (2020). Patterns of recurrence and metastasis in *BRCA1/BRCA2* -associated breast cancers. *Cancer*, *126*(2), 271–280. https://doi.org/10.1002/cncr.32540