**Expression of Estrogen Receptor 1 across Vanderbilt Subtypes of**

**Triple Negative Breast Cancer**

by Michaela Strizencova and Phung Tran

*Abstract*

Triple Negative Breast Cancer (TNBC) is a form of breast cancer that is characterized by the absence of estrogen receptor, progesterone receptor, and human epidermal growth factor 2. The lack of receptors makes TNBC particularly difficult to treat, as typical breast cancer treatments often target these receptors via hormone-based therapies and drugs. Therefore, other methods must be explored in order to implement effective treatment for TNBC. One method utilizes the analysis and classification of various forms of TNBC; this research will explore the Vanderbilt subtypes, a TNBC classification method which allows for a deeper understanding of the disease and subsequently more specialized treatments for each unique case. The Gene Expression Omnibus is used to analyze the Vanderbilt subtypes, specifically examining the variation of estrogen receptor expression levels across the subtypes.

*Introduction*

Breast cancer is the most common form of cancer in women worldwide, and can take on various forms. The simplest method of breast cancer classification is based on the presence of three common markers: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). These hormone receptors are proteins on or within cells that detect substances within the blood; for example, ER and PR detect and bind to estrogen hormone and progesterone hormone, respectively. About 80% of all breast cancers are ER positive (ER+), which means that the cancer cells grow in the presence of estrogen hormone. 65% of ER+ cancers are also PR+, in which tumor cells grow in the presence of progesterone hormone (NCBI).

Normal breast cells have both ER and PR, as these receptors serve vital roles in breast development and morphological changes. It is the overexpression of these receptors that promotes the development of breast cancer. However, the presence of these receptors, along with human epidermal growth factor receptor 2 (HER2), make possible various hormone therapies that target those specific receptors (such as selective ER modulators, aromatase inhibitors, and selective ER degraders). In other words, tumors that are ER/PR-positive are much more likely to respond to hormone therapy than tumors that are ER/PR-negative (NCBI).

Not all breast cancers are receptor positive. TNBC is a form of breast cancer that lacks the presence of receptors commonly found in breast cancer; it does not have hormone receptors for ER, PR, or HER2. These circumstances make TNBC generally more aggressive, have faster tumor growth rates, a higher risk of metastasis, and a higher likelihood of recurrence. TNBC accounts for 10-20% of all breast cancer cases and is most common in younger individuals, African American and Hispanic women, and patients with BRCA1 mutations (which is a gene on chromosome 17 that normally helps to suppress cell growth) (Jenkinson et al. 520).

In contrast to ER+/PR+ and HER2+ breast cancers, there are minimal FDA-approved targeted therapies for TNBC due to the complications that arise from not having an obvious drug target (i.e. lack of receptors). Current therapies for TNBC include surgery, chemotherapy, and radiation. However, research is rapidly progressing and it has been discovered that TNBC is not one single disease, but rather a group of breast cancers with distinct biological and clinicopathological characteristics. TNBC can be classified into numerous molecular subtypes based on gene expression profiles, which can be particularly useful for determining prognosis and developing personalized treatments (BCRF).

This particular research project will analyze the Vanderbilt subtypes of TNBC. The Vanderbilt subtypes are a classification tool based on 3,247 gene expression profiles from 21 breast cancer data sets; from this study, six distinct subtypes were discovered. This includes two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem–like (MSL), and a luminal androgen receptor (LAR) subtype. Each of these subtypes have unique gene expression patterns with different sensitivities and responses to targeted therapeutic agents. The researchers of this study developed a web-based prediction tool that utilized the classification methods and gene expression data to predict the subtype of a candidate TNBC sample (Vanderbilt).

*Methods*

The Vanderbilt subtypes were determined based on patients who underwent primary curative surgery for TNBC. Formalin fixation and paraffin embedding (FFPE) of tissues preserves the morphology and cellular details of tissue samples. Thus, it has become the standard preservation procedure for diagnostic surgical pathology. FFPE blocks (a form of preservation and preparation for biopsy specimens that aids in examination, experimental research, and diagnostic/drug development) were used for gene expression analysis. Tissue microarray construction was used for immunohistochemical (IHC) staining. Most clinical laboratories use immunohistochemistry (IHC) to evaluate the expression of ER and PR in patient samples; IHC uses antibodies to check for certain antigens (markers) in a tissue sample (Misra).

The gene expression dataset that was used was published on March 4, 2023. It included expression data for Vanderbilt TNBC subtype classification based on patients that underwent breast cancer surgery between January 2009 and October 2017 at Seoul St. Mary’s Hospital. Expression profiling was done by array; the platform used was [HuGene-2\_0-st] Affymetrix Human Gene 2.0 ST Array [transcript (gene) version]. It uses whole-transcript arrays that include probes to measure both messenger (mRNA) and long intergenic non-coding RNA transcripts (lincRNA). These whole-transcript array designs provide a complete expression profile of mRNA as well as the intermediary lincRNA transcripts that impact the mRNA expression profile (GEO). While six Vanderbilt subtypes were mentioned earlier, this particular dataset utilized re-classifications of the Vanderbilt subtypes into five distinct groups: IM (immunomodulatory), Mes (mesenchymal-like), B (basal-like), LAR (luminal androgen receptor), and UNS (unspecified). Methods of gene expression analysis will be further explained within the Research portion.

*Research*

The focus of this research project was to analyze the expression of ESR1 across the various Vanderbilt subtypes. As mentioned in the background information, estrogen receptor 1 (ESR1) has been the focal point of breast cancer. ESR1 is a gene that encodes an estrogen receptor and ligand-activated transcription factor. ER prognosis is a big determinant of the severity of breast cancer. While TNBC is characterized by the absence of ER, there are varying levels of ER expression among the molecular subtypes. The Gene Expression Omnibus (GEO) aided in researching the differing levels of ER expression across the Vanderbilt subtypes. GEO is a database for gene expression profiling. It utilizes high-throughput screening genomics data that is often derived from microarray technology, as was used in the research and data utilized in this particular project.

Microarray experiments use probes attached to a solid surface to detect gene expression levels. In the context of gene analysis, probes are short DNA or RNA sequences that are designed to bind to a particular gene or transcript of interest. These probes are used to measure the expression levels of genes in various gene expression profiling techniques (such as microarrays and RNA sequencing). In microarrays, labeled RNA or cDNA molecules from samples are then applied to the microarray and bind to specific probes, which reveals gene expression information. In RNA sequencing, probes capture specific RNA molecules that are later sequenced. The resulting sequences are compared to a reference genome or transcriptome. Through computational analysis methods of the gene expression data, researchers are able to identify expressed genes and gain insights into biological processes or diseases.

To analyze the ER expression across the Vanderbilt subtypes, the ESR1 gene information was needed to locate the respective probe. From the dataset, it was discovered that NC000006.11 was the genomic build associated with the genomic coordinates (written as 151977807..152450754) of the ER gene on chromosome 6. These genomic coordinates lead to the probe ID number, which was determined to be 17013809. With this information, the corresponding accession number reveals information about the ER gene. Probe 17013809 was isolated in the series matrix file, and was used to create box plots and a bar chart of the probe values stratified by Vanderbilt subtype. The following data visualizations reflect the results of ESR1 expression across the Vanderbilt subtypes.

<https://public.tableau.com/app/profile/phung.tran3417/viz/ProbeESR1ExpressioninTNBC/Dashboard1>

*Conclusion*

The top bar graph visualizes the level of ER expression levels across five Vanderbilt subtypes (LAR, Mes, B, IM, and UNS). The ID numbers of each sample are depicted on the x-axis; Mes, B, IM, and UNS types are represented by pink bars. The LAR subtype is highlighted in red bars. The level of ER expression is measured on the y-axis. The bar graph is organized in descending order, from highest to lowest level of expression (from left to right). The three samples highest in ER expression level (at 3.677, 3.737, and 3.815) are all classified as the LAR subtype.

The box and whisker graphs on the bottom left give further insight into the ER expression level averages, quartiles, and extremum of each subtype. Each box represents one subtype; the box is divided by a line near the middle which represents the median expression level of ER in that subtype. The upper half of the box represents the upper quartile, while the vertical line (or upper “whisker”) runs to the maximum expression level. The bottom half of the box depicts the lower quartile, with the bottom whisker running down to the minimum point. Although the LAR subtype is the least common subtype among the samples, it has the highest ER expression median, as well as maximum. Subtypes B and Mes were the most common subtypes, but had lower ER expression levels (with medians at 2.587 and 2.721, respectively). The UNS subtype had the lowest expression levels, as well as an outlier. The IM subtype had expression levels similar to Mes, but had the shortest range of expression of all the subtypes.

To further confirm that the LAR subtype is overrepresented for samples with ER expression levels > 3, a 2x2 contingency table was used along with a two-tailed Fisher’s exact test (a statistical test that is used to determine whether the proportions of two categorical variables are different from each other). The exact test yielded a p-value of 0.0060, which is statistically significant. This further confirmed the overrepresentation of LAR samples in all study samples that have expression levels > 3; thus, it can be concluded that LAR samples are associated with higher expression levels.

￼

*Two-tailed Fisher’s exact test to further demonstrate the overexpression of LAR samples in the pool of samples > 3*

*Future Research*

The LAR subtype was determined to have the highest level of ER expression out of all the Vanderbilt subtypes. Since TNBC is classified as being ER-negative by definition, future research could examine the boundaries of a sample being characterized as ER- versus ER+. Expanding upon this, research could also explore the difference of ER expression levels between LAR samples (especially those higher in ER) and typical ER+ breast cancer samples. This comparison could be achieved through a large dataset that incorporates gene expression data for both LAR samples and ER+ breast cancer samples for side-by-side analysis.

Each of the Vanderbilt subtypes have distinct molecular and clinical features that may cause the various subtypes to respond differently to different types of therapy. For example, the LAR subtype is characterized by the expression of androgen receptor (AR) as a notable biomarker; thus, the LAR subtype may be sensitive to anti-androgen therapy. Further research could seek to analyze the expression of AR across LAR samples with respect to the expression level of ESR1 in those same samples. An association between the AR and ESR1 expression levels may provide further insight into the characteristics of TNBC.

Works Cited

BCRF (Breast Cancer Research Foundation) blog article:

"Triple-Negative Breast Cancer: Treatment, Symptoms, Research." BCRF, 2021, [www.bcrf.org/blog/triple-negative-breast-cancer-treatment-symptoms-research/](http://www.bcrf.org/blog/triple-negative-breast-cancer-treatment-symptoms-research/)

"Breast Cancer Types: ER-Positive, HER2-Positive." WebMD, WebMD LLC, 2021. <https://www.webmd.com/breast-cancer/breast-cancer-types-er-positive-her2-positive>

"C-C motif chemokine ligand 2." National Center for Biotechnology Information (NCBI) Gene, National Library of Medicine, National Institutes of Health, <https://www.ncbi.nlm.nih.gov/gene/2099>

Dekker, Tessa J. A., et al. "Physiological variations in the expression of sex hormone receptors and ligands in the female reproductive tract." NCBI (National Center for Biotechnology Information) Journal of Molecular Endocrinology, vol. 66, no. 3, 2021, pp. R143-R158. PMC, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8150273/#:~:text=Physiological%20variations%20in%20the%20expression,even%20after%20menopause%20%5B2%5D>

DeSantis, Carol E., et al. "Breast Cancer Statistics, 2021." NCBI (National Center for Biotechnology Information) CA: A Cancer Journal for Clinicians, vol. 71, no. 1, 2021, pp. 7-33. PMC, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7869562/>

Jenkinson, Caroline E., et al. "The Health Benefits of Volunteering: A Review of Recent Research." Journal of Epidemiology and Community Health, vol. 67, no. 6, 2013, pp. 520-522. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3412597/>

Misra, Nagen K., Babu Misra, Gobardhan Das. "The Cellular and Molecular Mechanisms of Immuno-suppression by Human Type 1 Regulatory T cells". Frontiers in Immunology Volume: 5, 2014, pp. 359. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4077003/>

NCBI (National Center for Biotechnology Information). “Illumina HumanHT-12 V4.0 expression beadchip” Gene Expression Omnibus (GEO); Accession Number: GPL16686. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL16686>

Vanderbilt University Medical Center. "Triple-Negative Breast Cancer Information." Cancer Biomedical Informatics Grid (CBC), 2021, cbc.app.vumc.org/tnbc/