

# Triple Negative Breast Cancer

By Michaela Strizencova and Phung Tran

## What is TNBC?

TNBC stands for **Triple-Negative Breast Cancer**, which is a subtype of breast cancer characterized by the **under-expression** of three common receptors found in breast cancer cells: **estrogen receptors (ER)**, **progesterone receptors (PR)**, and **human epidermal growth factor receptor 2 (HER2)**.

The absence of these receptors makes TNBC **more difficult to treat** using hormone-based therapies or drugs that target receptor positive breast cancers. TNBC is typically characterized as more **aggressive** with a **higher risk of metastasis and recurrence**.



**10-20%**

TNBC accounts for about **10-20%** of all breast cancer cases. It tends to occur more frequently in **younger women, African American and Hispanic women**, and individuals with **BRCA1 gene mutations**.

## Vanderbilt Subtypes

In order to gain a deeper understanding of TNBC and produce more specialized treatments, the **genetic expression** of TNBC samples were **analyzed** and **sorted** into unique subgroups: The Vanderbilt subtypes are a **classification tool** based on **3,247 gene expression profiles** from **21 breast cancer data sets**.

**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. **\*\*The dataset used in this research re-classified the types into five groups (B, M, LAR, IM, and UNS (unspecified))**

## Estrogen Receptor Expression

Estrogen Receptor 1 (**ESR1**) is a **gene** that **encodes** an **estrogen receptor** and **ligand-activated transcription factor**.

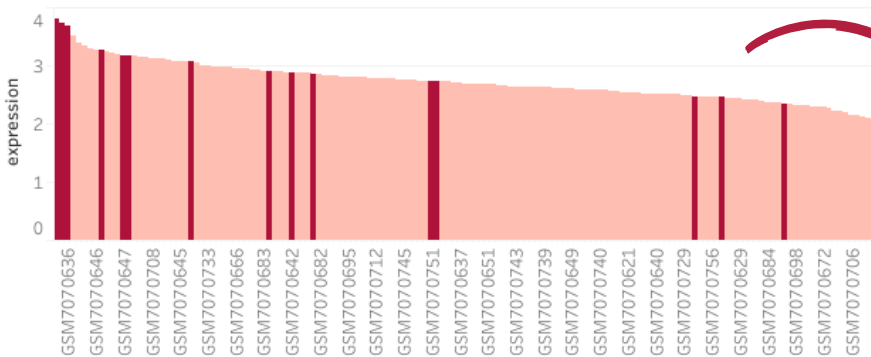
ER prognosis is a big determinant of the severity of breast cancer. While TNBC is characterized by the absence of ER, there are **varying levels of ER expression** among the **molecular subtypes**.

**Gene Expression Omnibus (GEO)**: 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 data utilized in this particular project.

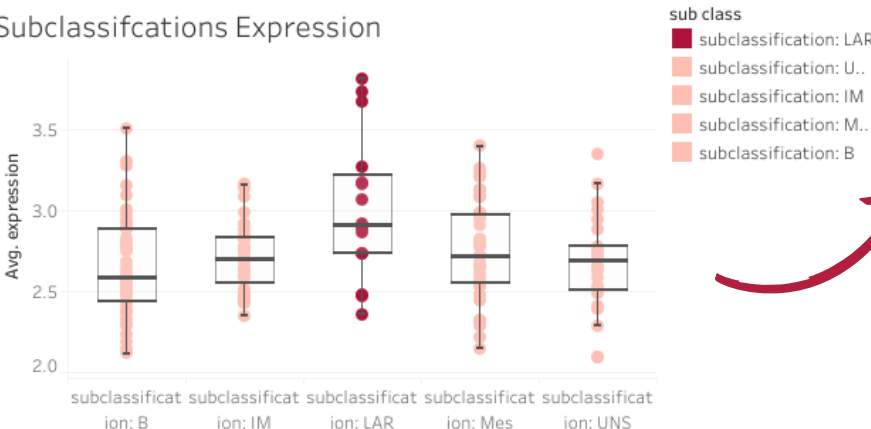
## The LAR Subtype

Data visualization depicting **ESR1** expression across five Vanderbilt subtypes (**LAR, UNS, IM, MES, B**):

probe 17013809 expression



Subclassifications Expression



2x2 contingency table used to calculate two-tailed Fisher's exact test for statistical significance of LAR ESR1 expression:

	ER expression > 3	ER expression <= 3	Total
LAR Samples	7	8	15
Non-LAR Samples	19	113	132
Total	26	121	147

Fisher's exact test

The two-tailed P value equals 0.0060

Estrogen Receptor 1 Probe: 17013809

**GEO** was used to obtain gene expression data for each Vanderbilt subtype. Analysis was conducted with ESR1 probe to **compare ER expression levels** across the **five subtypes**.

**Data Visualization**

**Top bar graph** depicts **ESR1** expression from highest to least across **all samples**. **LAR subtype** (highlighted in **red**) had **highest overall expression level of ESR1** (3.815). Remaining subtypes are in pink.

Each subtype is represented by a **box and whisker plot** to further analyze averages, quartiles, and extremum. The **LAR subtype** had the **highest median (2.915)** and **highest upper quartile**.

**Two-tailed Fisher's exact test** further demonstrates the **overexpression** of **LAR samples** in the pool of samples with **ER expression > 3**. The test yielded a **p-value of 0.0060**, which is **statistically significant**.

**Thus, it can be concluded that LAR samples are associated with higher ESR1 expression levels**