

Write-Up (1-2 page Analysis)

Introduction/Abstract

A biomarker is a measurable characteristic or substance that can indicate the presence or progression of a disease, condition, or biological process. Biomarkers can be found in various types of biological samples, such as blood, urine, tissue, or saliva. They can be used for diagnosis, prognosis, prediction of treatment response, and monitoring of disease progression or recurrence. Some common examples of biomarkers include proteins, genes, metabolites, and imaging markers. Biomarkers play a crucial role in the development of new therapies, as they can help researchers better understand the underlying mechanisms of disease and evaluate the effectiveness of potential treatments. For effective treatment of breast cancer, it is important that the diagnosis is established at an early stage by doing early screening for women who are at high risk. To contribute to our breast cancer biomarker research we focus on estrogen receptors. Estrogen receptors are important in breast cancer research because they play a key role in the development, growth, and progression of most breast cancers. The majority of breast cancers (about 70%) are estrogen receptor-positive (ER+), meaning that the cancer cells have receptors on their surface that bind to estrogen and promote their growth. In these types of breast cancer, estrogen stimulates the cancer cells to grow and divide, leading to the development and spread of tumors.

Methods + Results

We used the Gene Expression Omnibus (GEO) and analyzed a dataset pertaining to ER expression. GEO is a public repository of arrays, sequencing, and functional genomic data. We used GSE17040: "Functional ER alpha transcriptional regulatory network for cell cycle in an ER(+) breast cancer subgroup." There are 57 samples, 46 ER(+) and 11 ER(-). The data takes from both cell line and clinical samples. The platform used is GPL887 Agilent-012097 Human 1A Microarray (V2) G4110B (Feature Number version). We utilized the "Analyze with GEO2R" feature to automatically run statistical analysis. We then downloaded the Excel sheets of all the genes sampled. Next, we located the estrogen receptor gene (estrogen receptor 1 - ESR1) among all of the genes. Then, we analyzed expression level and statistical significance. To better visualize the data, we downloaded the "Series Matrix File" to access the full dataset. We then located the ER gene, and then separated ER(+) vs. ER(-) to create box plots for each of them.

The "Analyzed with GEO2R" feature gave us the result for comparing expression levels of ER(+) vs ER(-). For ESR1 (ID 5561), the log fold change is 4.04669853, and the adjusted P-value is 2.03E-08 (<0.05). The log fold change indicates the expression level of an ER group

relative to the other. The P-value indicates statistical significance (adjusted t-test, <0.05 indicates statistical significance). The result means that the ER(+) variant has an expression level around 4 times higher than the ER(-), and that this difference is statistically significant. This finding corroborates with the hypothesized pattern: ER(+) expresses estrogen at a statistically higher level than ER(-).

Glossary:

Gene Expression: the process by which the information encoded in a gene is turned into a function.

Estrogen Receptor: activate transcriptional processes and/or signaling events that result in the control of gene expression

Progesterone: A protein bind to the receptors inside the cells and may cause the cells to grow