Multiple Testing Correction and Functional Analysis for Transcriptomic Data

Firstly, adjustments have been made in Excel among sample types to enhance the accuracy of calculations.

a) At this stage,

=T.TEST(tumor_range, normal_range, 2, 2)

adjustments have been made in Excel to enhance the accuracy of calculations for different sample types. The formula used for this purpose is: This formula was employed to calculate p-values. A total of 2352 genes exhibited p-values below 0.05, indicating that these genes significantly altered their expression. Furthermore, 1126 genes had p-values below 0.01, signifying significant changes in expression for this more stringent 0.01 cut-off.

b) In this stage, the first step involved calculating the fold change by taking the mean of expression values for tumor and normal samples separately. Subsequently, the fold change value was determined by dividing the tumor mean by the normal mean.

Since both p-values and fold change values are essential for creating a Volcano plot (Figure 1), their accurate calculation is crucial. To better represent fold change and p-values on the Volcano plot, we computed logFC (logarithm of fold change) and -log10(p-value) values. These values were then added as new columns in the Excel file. Finally, we visualized these values on the Volcano Plot for a comprehensive representation.

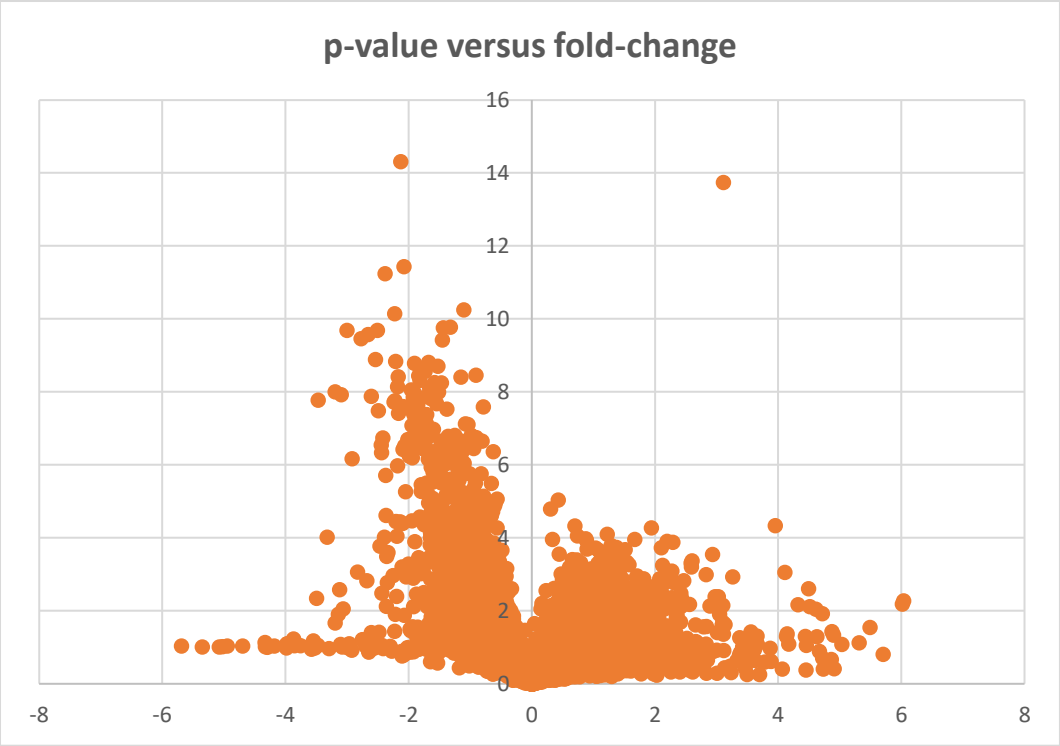

Figure 1. Volcano Plot about p-value versus fold-change.

c) The Bonferroni correction will be applied to the p-values obtained in the first step by examining the relevant genes. For this, PBonferroni is calculated using the formula $\mathbf{P^{Bonferroni}}$ **= original p-value / total test number**. A new column was created for this in Excel. This value is calculated for p-values that are less than the p-value cutoff of 0.05.

d) Benjamini-Hochberg Correction was applied to calculate Q-values for the genes, and a new column was added to Excel for these Q-values. In the formula:

=AK2*12532/(ROW(AK2)-1), the gene's p-value is multiplied by the total number of genes (12532), and then the result is divided by the position of that p-value in the list.

e) Performing multiple testing correction is crucial when conducting analyses involving a large number of statistical tests simultaneously. Without correction, the likelihood of obtaining false positives increases, leading to potentially misleading or inaccurate results. The primary goal of correction methods is to control the overall false discovery rate (FDR) or familywise error rate, maintaining the integrity and reliability of the findings.

In the given case, where transcriptome data from non-small cell lung cancer patients is analyzed for differential gene expression, the application of a correction method is highly recommended. The dataset involves the comparison of gene expression between tumor and normal samples, and as such, it likely includes a considerable number of hypothesis tests. Failing to correct for multiple comparisons might lead to an elevated risk of identifying genes as differentially expressed purely due to chance.

Therefore, to ensure the robustness and validity of the results, a correction method should be employed. In this context, both **Bonferroni** and **Benjamini-Hochberg** corrections are common choices. The Bonferroni correction is conservative but straightforward, adjusting the significance threshold based on the number of tests conducted. On the other hand, the Benjamini-Hochberg correction, being less conservative, is often preferred in genomics studies as it provides a balance between controlling the FDR and maintaining higher statistical power.

f) We extracted the list of genes with p-values less than the 0.05 cut-off from Excel and pasted it into the relevant section for Gene Ontology and Pathway Enrichment Analysis. Additionally, we defined the Benjamini-Hochberg FDR as significance threshold.

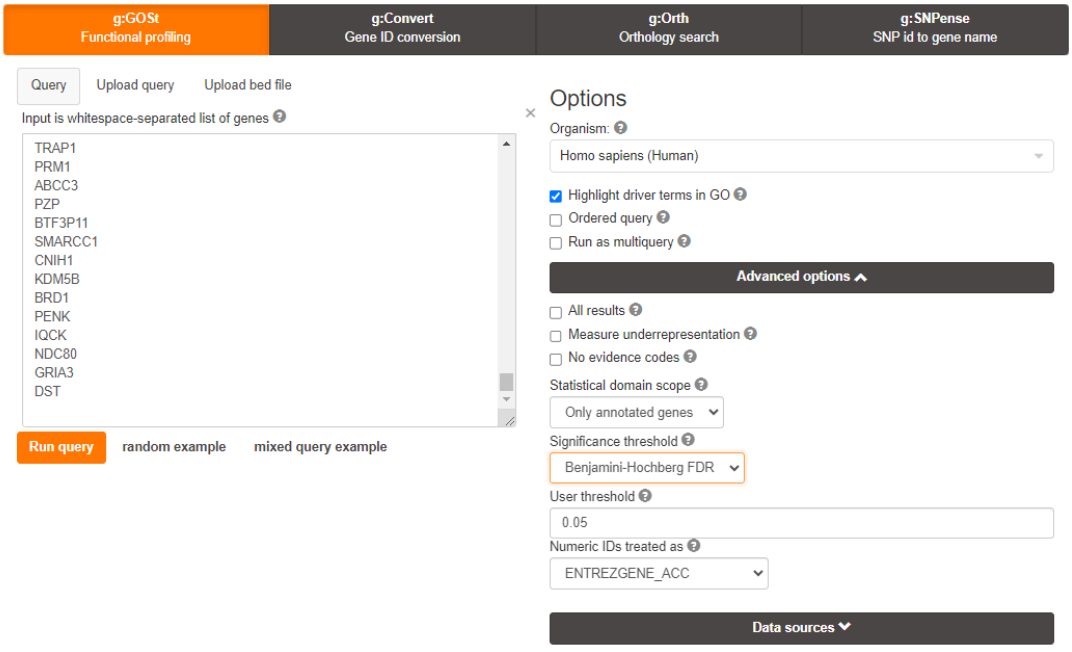

Figure 2. g:Profiler Run query step.

After we run query (Figure 2), we select again the Ensembl ID with the most GO annotations (for all; first if the same)

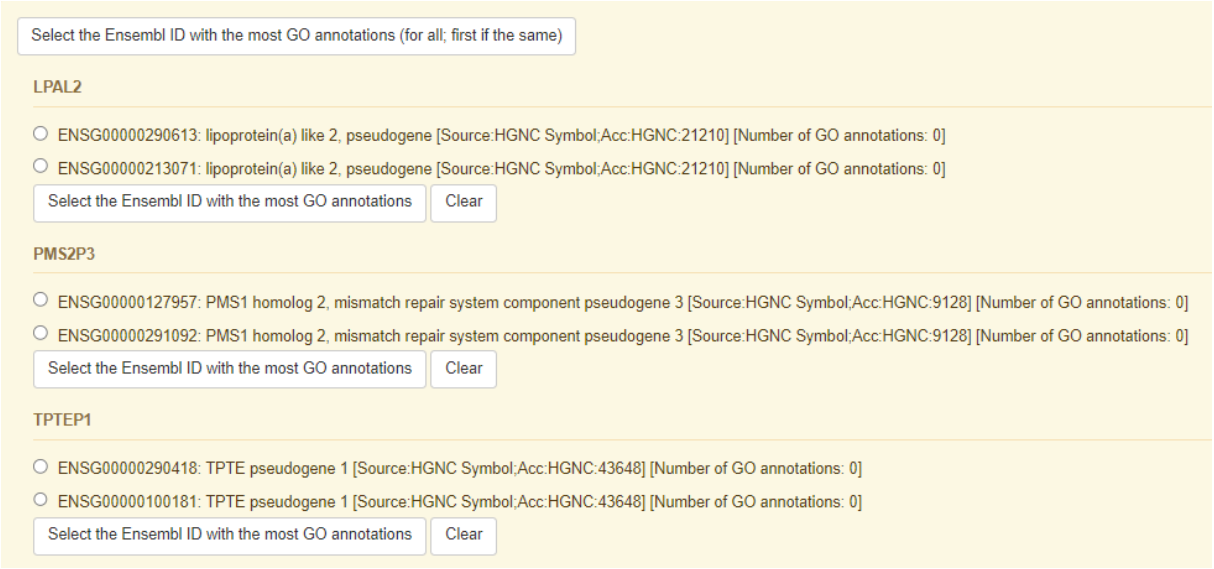

Figure 3. Rerun query step.

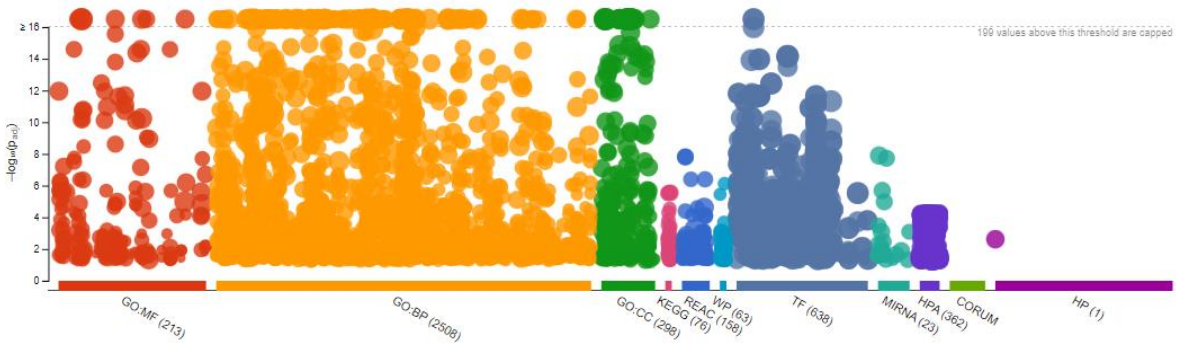

Figure 4. General Results of Our Dataset.

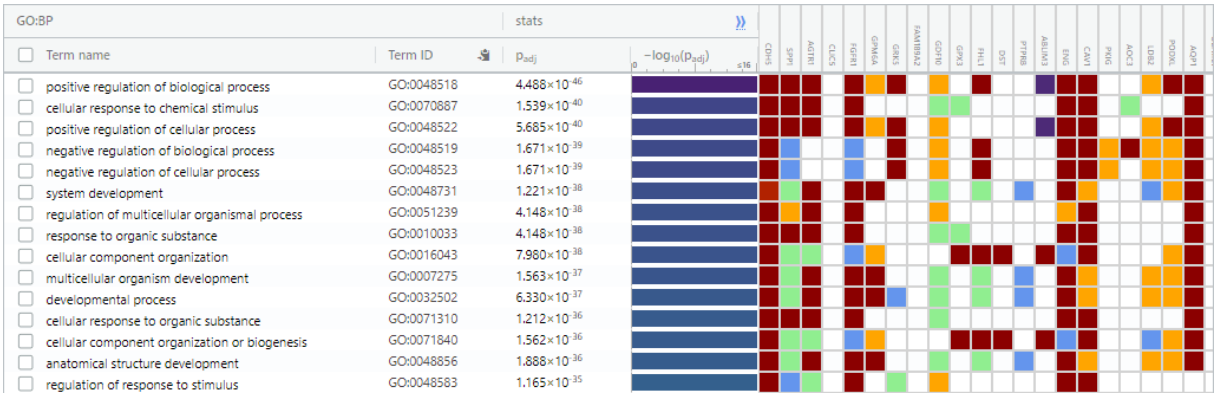

Figure 5. Gene Ontology: Biological Process for Our Dataset.

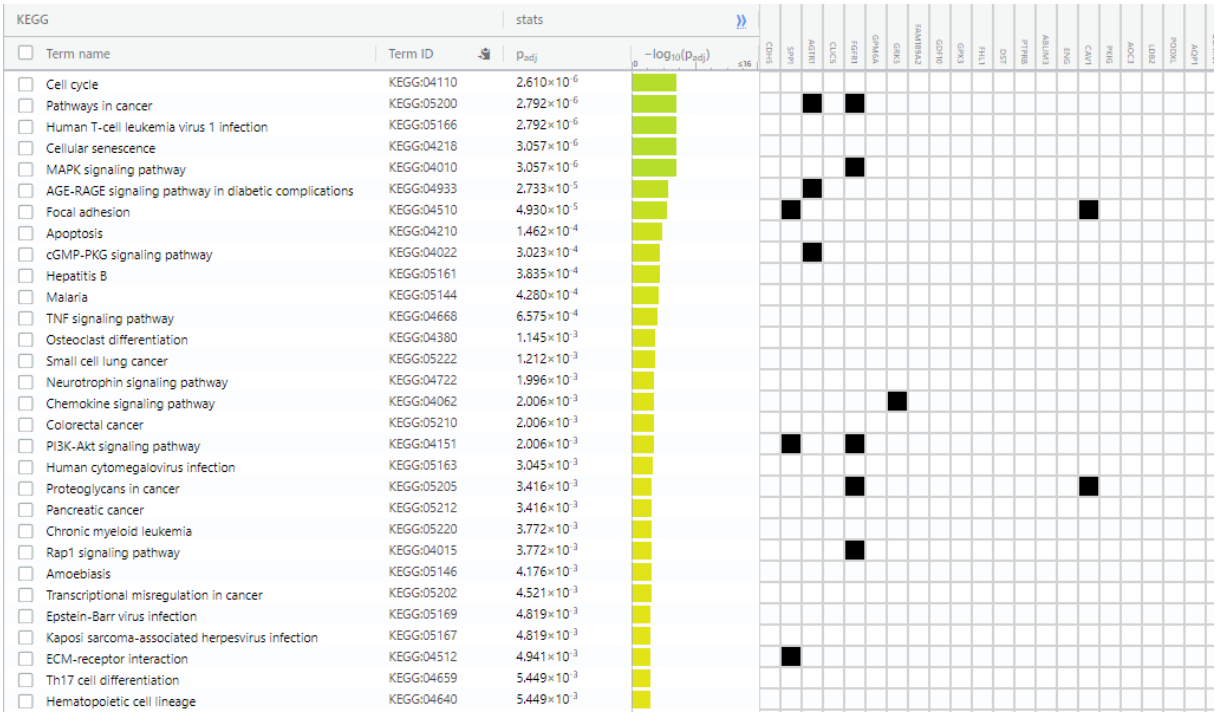

Figure 6. KEGG for Our Dataset.

g) According to the results obtained from the analysis, the data is consistent and logical. Consequently, when the provided terms are researched in the literature, the obtained results are as follows:

MAPK Signaling Pathways and Relatod to Cancer: MAPK signaling pathways are crucial mechanisms that regulate fundamental cellular processes such as cell growth, proliferation, differentiation, and survival. In association with cancer, activated MAPK signaling pathways may contribute to uncontrolled cell proliferation, hinder differentiation, and promote metastasis. Consequently, MAPK signaling pathways play a significant role in cancer biology and are a focal point in cancer research. When comparing the expression data of genes for consistency and their association with cancer, it has been concluded that the expression levels in tumor samples present a consistent pattern.

Focal Adhesion and its Related to Cancer: Focal adhesions are key structural and functional points where cells interact with their surroundings. While playing a crucial role in regulating normal cell behavior, they become a significant focal point concerning cancer. Specifically, cancer cells can attach to these points, playing a crucial role in the metastatic process as they migrate towards surrounding tissues. By disrupting normal regulation, cancer cells may stimulate abnormal focal adhesion signaling, contributing to the development of invasive

characteristics. Additionally, focal adhesions influence biological processes such as cell growth, proliferation, and survival through signal transmission, thereby supporting uncontrolled growth of cancer cells. Thus, the relationship between focal adhesions and cancer highlights their effectiveness in critical cancer features like metastasis, invasion, and signal transmission.

TNF Signaling Pathway and its Relation to Cancer: The TNF (Tumor Necrosis Factor) signaling pathway is a critical network regulating cell apoptosis, controlling inflammation, and influencing immune responses. This pathway is closely associated with cancer, in addition to its involvement in normal biological processes. Specifically, the TNF signaling pathway demonstrates potential efficacy in combating cancer by regulating cell death processes. However, in certain instances, cancer cells can evade this apoptotic signaling pathway, enabling continuous survival. Furthermore, TNF signaling pathways play a role in cancer development through their impact on inflammation control and the immune system. Chronic inflammation can promote the formation of cancer. Ongoing research aims to understand the specific mechanisms of how these signaling pathways are involved in cancer biology and their potential applications in cancer treatments.

When using g:Profiler, the obtained data parallels the information in our Excel file. Examining the expression levels of genes in tumor and healthy samples reveals a significant difference. Genes showing high expression levels in tumor samples also manifest themselves in cancer-related biological pathways.

In the KEGG table, the ***AKT3, RAF1,*** and ***MAPK3*** genes stand out, particularly in many provided pathways. Excessive activation of the ***AKT3*** gene can contribute to the uncontrolled proliferation of cancer cells, promoting their survival and tumor formation. Such abnormal activation of ***AKT3*** is specifically associated with various cancer types, including breast cancer, prostate cancer, and colon cancer. The ***RAF1*** gene emerges as a part of the MAPK signaling pathways, which play a role in cell growth and proliferation. Under normal conditions, it regulates cell growth, but excessive activation, mutation, and interaction with tumor suppressors and oncogenes associate it with cancer. ***MAPK3***, known as a gene regulating fundamental cellular processes associated with cell growth, proliferation, and differentiation, is linked to cancer through abnormal activation and mutations. It can contribute to uncontrolled cell proliferation, tumor formation, and metastasis.