Strigolactone Analogues and their Effect on Apoptosis

By: Ammar Abdullah, Arun Jairam, Joel John

**Abstract**

In the field of cancer study, there’s a potential chemical that could help with anticancer therapy known as Strigolactones and they are plant hormones that are produced in the roots and they’re known for inhibiting self-renewal of undifferentiated meristem cells. This means that they can cause apoptosis of certain cells and stop their growth. While they have been shown to affect cancer cells exponentially, their effect on normal cells is being tested here.

This study was trying to find how much Strigolactone analogues affect the growth of normal cells to see how much their growth is inhibited and, from what we found, there’s little-to-no effect that the analogues have on the growth of normal cells. Previous studies have shown that Strigolactone analogues halt the growth of cancer cells significantly and the fact that they don’t affect normal cells as much is also what a previous study said. Thanks to the previous study, along with this study, we know that Strigolactone analogues are good to use for anticancer treatment. From this current study, we found that there isn’t a lot of variation between the growth of the cells that are being tested compared to the cells that aren’t being tested. It supports the studies that were done by the previous study and proves it to be true. Because of this, it means that any testing using Strigolactones won’t affect normal cells when performing tests with them. In addition to this, and experiments done before proving that it does potently halt the development of cancer cells, it means that this can help in studies relating to anticancer research by limiting the abilities of cancer cells while not harming normal cells, therefore, slowing cancer enough to make them easier to deal with in terms of experimentation and control.

**Introduction**

The data in this experiment involved the usage of strigolactones. They are made in the roots of plants and they regulate shoot and root development for the plant [1, 2]. They regulate the growth process of their own cells by taking the growth of cells that branch away from their expected growth and/or are cancerous in terms of development and inhibiting them [3]. Because of this information, several scientists have studied these strigolactones and their effect on cancerous human cells. These studies have shown that the strigolactones inhibit breast cancer growth [4], and the growth of human myeloid leukemia as their growth had not been factored into any developing DNA or RNA nearby [5]. These have been connected to the fact that the strigolactones cause apoptosis within these cells, making it so that their reproduction is limited and the cancerous cells become detached from the overall development cycle of standard cells [6, 7]. Along with this, there are also a few studies that talk about how they have little effect on the reproduction of normal cells while inhibiting the cancerous ones [3, 8]. Combining these factors together, strigolactones became widely seen as a potential anticancer drug and are being widely tested and used for the sake of anticancer research [9, 10].

For this experiment, the question that was sought to be answered was how much strigolactone analogues could affect the growth of noncancerous human cells. The data was gathered by utilizing a variety of non-cancerous human cells, including skin cells, liver cells and lung cells and placing them within 5ppm of ST-362, a specific analogue of strigolactone, for six hours and calculating how much RNA growth was made in that sample compared to a sample of the same types of human cells without the ST-362 and calculating RNA growth when put by themselves after six hours. With the data received from the six hours of both the cells with the ST-362 and without the ST-362, several methods were used to determine the results of the six hour growth and make sure that they were accurate by either differentially expressing them, analyzing the enrichment of the gene sets, clustering them to see how unique they were from one another and using statistics to see how accurate the clustering methods were.

**Methods**

All code and analysis can be found in our github repo: <https://github.com/joeljohn12/Bioinformatics-Project>

**Differential Expression**

We created a density plot using the gene and gene read count information to find how much variation there was between the entirety of the data and their outcomes of how much variation there is between the growth of the RNA of each of the datasets. We then created a PCA plot, which shows clusters of samples based on their similarity, and thus shows the variance of the samples between the ST-362 and control, with the input being the same data as the density plot. Since there were three people in the group, we also created a UMAP plot, which is used to show how separable some classes are with respect to certain features. We ran the data in, with the certain feature being the counts, and the classes being control and ST-362. We then created a volcano plot showing the relationship between the fold changes and its p values, this was in order to understand what the most statistically significant genes are. And as the same went with all of the previous plots, we used the counts data for each gene as the input for creating this plot as well. Using the information we received by creating the volcano plot, we also created a heatmap, showing how statistically significant the top 14 genes were for each class. So we made the heatmap with the statistical significance of the top 14 statistically significant genes and found that all of them completely coincide, and have the same significance in both of the classes, suggesting the similarity between the two classes are high.

| **Density Plot** | **PCA Plot** | **UMAP Plot** |
| --- | --- | --- |
|  |  |  |
| **Volcano Plot** | | **Significant Genes Heatmap** |
|  |  |  |

**Gene Set Enrichment Analysis**

Due to time constraints, we were only able to do one enrichment analysis, and it was the gProfiler2 with gene ontology. The gene enrichment was to find the most significant genes in the datasets as well, and the input was the tables of genes and their respective counts during the testing phase. Unfortunately we weren’t able to further study the enrichment analysis due to a lack of time.

| **gProfiler2 Enrichment Analysis** |
| --- |
|  |

**Clustering**

Cluster plots were made for the data sorting the 5000 most variable genes and determining how they would react in a given cluster. The first clustering method used was the HClust method which sorted the 5000 most variable genes as an input parameter, as well as having k equal to 10, to see how the most variable genes clustered together given both euclidean and average parameters. Then clustering was done with the PAM method by making a pamResult data frame from the 5000 most variable genes, and then using fviz\_cluster to create a graph of the clusters, with the input parameter being that k equals 2. GMM Clustering was done by predetermining and sorting the 5000 most variable genes into a scatter plot with no k value needing to be inputted as it was automatically calculated. Afterwards, alluvial graphs were done with each of these clustering methods by comparing them to the 10, 100, 1000 and 10000 most variable genes and seeing how they were being clustered as well as how the clusters changed given more these variable genes to work with. Finally, heatmaps were done with the 5000 most variable genes showing the clusters that could be identified through sorting them on the heatmap and seeing how similarly the clusters aligned with the with the clustering algorithms.

| **HClust Plot** | **PAM Plot** | **GMM Plot** |
| --- | --- | --- |
|  |  |  |
| **HClust Alluvial Comparison** | **PAM Alluvial Comparison** | **GMM Alluvial Comparison** |
|  |  |  |
| **HClust Heatmap** | **PAM Heatmap** | **GMM Heatmap** |
|  |  |  |

**Statistics**

In order to check the correlation between each of the clustering methods, Chi squared tests were done on each of the methods, when they had the parameters of the 5000 most variable genes and checking their p values to determine if there was a significant pattern between specific genes in specific clusters. Then, in order to test the accuracy of the clustering methods with one another, chi squared tests were done comparing the clustering of each of the methods with one another, subdividing these categories by testing the clusters and how they are made with the 50 most variable genes and the 1000 most variable genes of each clustering method. Finally, an enrichment plot was made using all of these chi squared tests to show the statistical results of all the metadata and clustering methods.

| **Chi Squared Test of HClust Cluster of the 5000 most variable genes** | **Chi Squared Test of the PAM Cluster of the 5000 most variable genes** | **Chi Squared Test of the GMM Cluster of the 5000 most variable genes:** |
| --- | --- | --- |
|  |  |  |
| **Chi Squared Test of the GMM cluster of 1000 genes vs. the PAM cluster of 50 genes** | **Chi Squared Test of the GMM cluster of 1000 genes vs. the PAM cluster of 1000 genes** | **Chi Squared Test of the Hclust cluster of 50 genes vs. the GMM cluster of 1000 genes** |
|  |  |  |
| **Chi Squared Test of the HClust cluster of 50 genes vs. the HClust cluster of 1000 genes** | **Chi Squared Test of the HClust cluster of 50 genes vs. the PAM cluster of 50 genes** | **Chi Squared Test of the HClust cluster of 50 genes vs. the PAM cluster of 1000 genes** |
|  |  |  |
| **Chi Squared Test of the GMM cluster of 1000 genes vs. the HClust cluster of 1000 genes** | **Chi Squared Test of the PAM cluster of 1000 genes vs. the HClust cluster of 1000 genes** | **Chi Squared Test of the PAM cluster of 50 genes vs. the HClust cluster of 1000 genes** |
|  |  |  |
| **Chi Squared Test of the PAM cluster of 50 genes vs. the PAM cluster of 1000 genes** | **Enrichment plot comparing the statistical test results of the chi-squared test analyses** | |
|  |  | |

**Results**

Looking at the density plot, we can see that there is a good amount of variation between tests, but not so much between the actual classes, proving a lack of variation in the growth of cells between the two groups. The same can be said when looking at the PCA and UMAP plots, which basically plots the variation and difference between the two groups. The volcano plot shows a vast majority of the genes are very close together and aren’t super significant in general, although there are few genes that are very uncharacteristically high. The heatmap then shows that a lot of the genes that were significant in the control group were also significant in the ST-362 group. The gene enrichment, agreed with the prior statement, although we weren’t able to conduct any further observations or analysis with the enrichment due to a lack of time.

Looking at the clustering, chi squared tests and the enrichment plot we can see some interesting results. For the Hclust plots, it showed that the genes with the most variance had smaller clusters and the genes with least variance had larger clusters. With the PAM method, it was shown that ther were two clusters found with this method and they had a similar amount of genes are within them. These combined help to support the hypothesis by showing that, when more clusters are made, it becomes clearer that the ones with the highest variance are great outliers compared to the variability of the rest of the variables. The GMM method plotted a scatterplot of the 5000 most variable genes which showed that the variable genes for all of them were linearly aligned with similar values, proving the hypothesis by showing that the control and variable groups aren’t consistently different in variability and showing that the data between them doesn’t show significant difference. In terms of chi squared tests, the first thing to note is the fact that the Hierarchical clustering method seemed to have statistical significance due to the fact that the X-squared value was so much higher than its p-value. However, the PAM and GMM clustering algorithms did not show statistical significance as their X-squared values were lower than their p-values. In terms of the relationship between the clustering algorithms and whether or not they were statistically significant, surprisingly, all the other chi-squared tests showed statistical significance indicating that clustering algorithms got similar results to each other. This is shown further in the enrichment plot as well. Overall, this supports the hypothesis by showcasing that, even when clustering with variance, the only ones with massive variance are outliers and that the majority of them were clustered together showing little variance.

In terms of the success of our project, a lot of things went to plan and we were able to manipulate the data enough to get solid evidence to support our hypothesis. We were able to find an interesting project to work on and managed to manipulate and clean up the data given in order to find the variability of the genes and perform enrichment analysis on said data. After that, we were able to successfully run three clustering algorithms on the data to see the natural groupings of the genes and perform chi squared tests to find if the clustering was statistically significant or not. And finally, we were able to successfully create all of the graphs and tables requested of us to display all the data and results we found.

When it comes to the weaknesses in our project, we found that they’re are two main ones. The first failure we had was we were only able to do one enrichment analysis and this limited our data analysis by a decent amount. This was mainly due to a time crunch we had at the end of the second assignment and this ended up affecting the rest of our project pretty significantly. The second failure we had was our clustering didn’t work exactly as intended and some of our graphs and plots came out slightly incorrect. This is clear in some of the alluvial diagrams and volcano plots as some of the values are incorrect.

The only bioethical concern we had with our project is how the Strigolactones was being tested on the human cells as there are different ways of testing it. Thankfully, after conducting more research on how the experiment was being conducted, we found that the cells that were taken were taken in the appropriate scientific means as all the cells were taken from subjects with their knowledge, consent, and full transparency on what the test was being used for. In terms of future work that could be done with this kind of research, considering our hypothesis was supported by the data, the next logical step would be to test Strigolactone analogues on actual human subjects or at least something closer to actual humans than just cells.

**Conclusions**

Our original hypothesis was that we could determine how the amount of strigolactone analogues can affect its rate of apoptosis in a cell as well as the fact that it would not affect normal cells if given them. From the tests that we have done, we can conclude that the strigolactone analogues don’t affect the growth of a cell in any drastic way.

While there is variance within the development of each cell themselves and the groups, as seen with the UMAP and the PCA Plot, the volcano and density plots of the differential expression of the two groups showed that the two groups are similar in terms of their distribution and comparison to one another. It was also seen when extracting the most differentially expressed genes and showing that there was only one outlier that was extremely different while the rest weren’t noticeably different from one another.

By comparing the 5000 most variable genes, it was also found that, the cluster distribution was uneven overall, showing results similarly to the heatmap with the most variable ones being grouped with themselves while the ones with lesser variability were clumped together in the end, showing that a significant amount of variability was an outlier. Most of the clustering methods also concluded that the variability within specific clusters weren’t related to what group they belonged to, meaning that they were generally outliers from the data with the rest of the data supporting the hypothesis.

In terms of improvements that could also be made to the project there are two main ones. The first big improvement that could be made is using different clustering algorithms to analyze the data. We found that the clustering algorithms we used gave us different results and didn’t really work well with our data and so potentially using different clustering algorithms could have yielded better results. For example, Mean Shift is another popular clustering algorithm that finds and adapts cendroids using mean values and we believe this would have yielded a good clustering plot. The other big improvement that could have been made to the project is more gene enrichment analysis could have been done as only one was conducted. This could easily be fixed in future iterations of the project as more enrichment analysis could have been done with different gene analysis packages.

**References**

1. Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC, Bouwmeester H, Becard G, Beveridge CA, Rameau C, Rochange SF. Strigolactone inhibition of shoot branching. Nature. 2008;455(7210):189–194
2. Clouse SD, Sasse JM. BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. Annual review of plant physiology and plant molecular biology. 1998;49:427–451
3. Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, Magome H, Kamiya Y, Shirasu K, Yoneyama K, Kyozuka J, Yamaguchi S. Inhibition of shoot branching by new terpenoid plant hormones. Nature. 2008;455(7210):195–200
4. Pollock CB, Koltai H, Kapulnik Y, Prandi C, Yarden RI. Strigolactones: a novel class of phytohormones that inhibit the growth and survival of breast cancer cells and breast cancer stem-like enriched mammosphere cells. Breast cancer research and treatment. 2012;134(3):1041–1055
5. Ishii Y, Sakai S, Honma Y. Cytokinin-induced differentiation of human myeloid leukemia HL-60 cells is associated with the formation of nucleotides, but not with incorporation into DNA or RNA. Biochimica et biophysica acta. 2003;1643(1-3):11–24
6. Steigerova J, Oklestkova J, Levkova M, Rarova L, Kolar Z, Strnad M. Brassinosteroids cause cell cycle arrest and apoptosis of human breast cancer cells. Chemico-biological interactions. 2010;188(3):487–496
7. Goldin N, Arzoine L, Heyfets A, Israelson A, Zaslavsky Z, Bravman T, Bronner V, Notcovich A, Shoshan-Barmatz V, Flescher E. Methyl jasmonate binds to and detaches mitochondria-bound hexokinase. Oncogene. 2008;27(34):4636–4643
8. Skoog F, Strong FM, Miller CO. Cytokinins. Science. 1965;148(3669):532–533
9. Newman DJ, Cragg GM. Advanced preclinical and clinical trials of natural products and related compounds from marine sources. Current medicinal chemistry. 2004;11(13):1693–1713.
10. Cohen S, Flescher E. Methyl jasmonate: a plant stress hormone as an anti-cancer drug. Phytochemistry. 2009;70(13-14):1600–1609