

Say Yes To Drugs, and No To Melanoma

Sarah August, Wanying Yang, Cynthia Zhou, and Shiqi Zuo (TUT108-Jenny-2)

December 8th, 2022

Introduction

Homeostasis; An incredibly important idea in cancer research. More specifically, returning a group of cells to homeostasis, and keeping it there. We will look at a data set of Melanoma cancer cells, and the effects that different cancer combating drugs have on the cells under different conditions. We will be analysing the data using a variety of different statistical testing methods in order to answer three research questions developed by our group. Our goal is to identify any patterns in the data, aiming to draw conclusions about the cellular plasticity in melanoma cells. This will hopefully point us in the direction of achieving “good” homeostasis from bad with these treatments, and possibly, a step towards the cure to Melanoma.

Presentation Format

Background Information

Data Summary

Research Questions Analyzed

- Is there a difference between phenotype protein levels before and after the drug is introduced?
- What are the correlations between each phenotype protein?
- How do phenotype protein levels change over time 't' with and without drugs?

Final Conclusion

Limitation & Future Direction

References & Acknowledgments

Background Knowledge

About The Data: The data set we used was developed as evidence in an article about AP-1 Transcription Factors in the context of Melanoma treatment research. The goal of the article was to prove that the AP-1 Transcription factor network can explain cellular plasticity patterns in Melanoma cells. We developed our own questions based on these findings, specifically the phenotype protein results.

What is Melanoma? Melanoma is a serious type of skin cancer that affects the cells in your skin that produce melanin, otherwise known as pigment. Melanoma is known for its high phenotypic plasticity rate, meaning it changes phenotypes rather quickly, making it difficult to treat. Melanocytic cells, or differentiated cells, are mature cancer cells, which mean they proliferate faster than normal skin cells, which is considered “bad” homeostasis.

Data Summary

Initial Data Set has 22 AP-Transcription Factors, 4 Phenotype Proteins.

They are in different conditions: Drug(Vem, Vem+Tram), Dosage(0-3.16uM), Time(0.5-120hr), Rep(1-3).

Three Different Tidy Data Sets, One Per Question

- ① 4 Proteins + 0uM + 3.16uM + Vem + Rep 3 + **24hr time period**
- ② 4 Proteins + 3.16uM + Vem + Rep 3 + **24hr time period**
- ③ 4 proteins + 0uM + 3.16uM + Vem + Rep 3 + **1-24hr time period(time id 1-5)**

Reasoning

Phenotype Proteins: All were included to ensure that all possible phenotypic results were analysed, since each one implies a different result.

AP-1 Transcription Factors: Since we were focusing on the end results of the cells, we only used the phenotype results, and not the transcription factors.

Time Period: Time IDs 6 and 7 were missing variables, specifically in cells that met our outlined conditions, so they were omitted to avoid any error due to missing data.

Dosage: Dose IDs 1 and 5 were used, since this allowed us to compare the cells before they interacted with the drug, and after they had interacted with the maximum dosage of the drug.

Drug: We chose to focus on Vemurafenib as a baseline drug to test the statistical testing methods used throughout this project. After finalising our test, we would be able to test other drugs with the same outline in any further development of our study.

Rep: We used data with exclusively 3 repetitions to ensure the error for each value was as small as possible.

Question 1

Are the four proteins' levels after 24 hours (`timpoint_id = 5`) different between the control group without the drug (`dose_id = 1` and `drug_id = 1`) and the treatment group 3.16uM Vem (`dose_id = 5` and `drug_id = 1`)?

We want to test the effect of the drug on the proteins. Therefore, we choose the maximum amount of the drug and the longest interaction time to test whether the drug can have a difference on the amount of the four proteins.

Statistical Methods & Procedure (Q1)

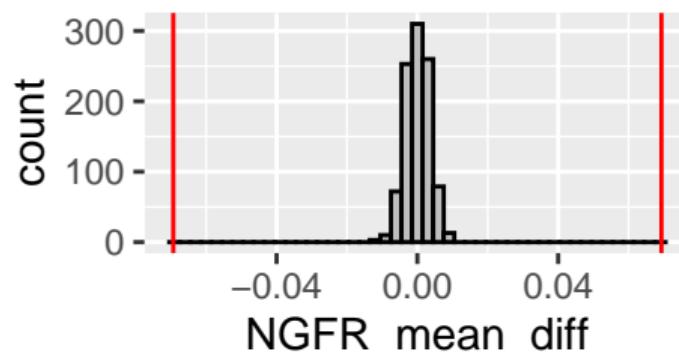
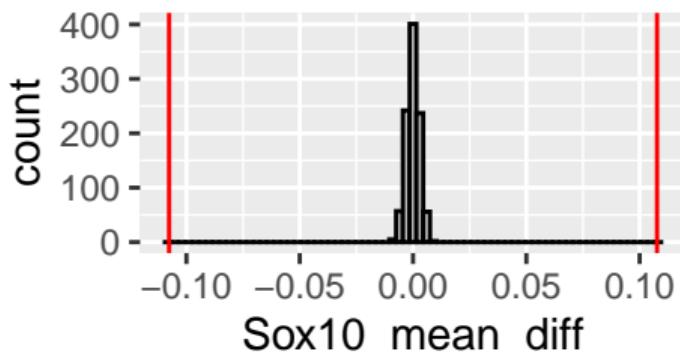
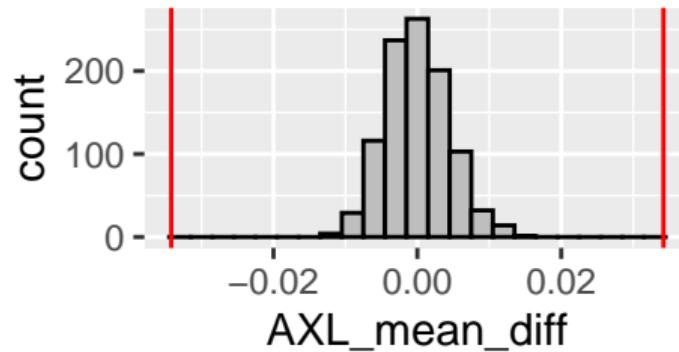
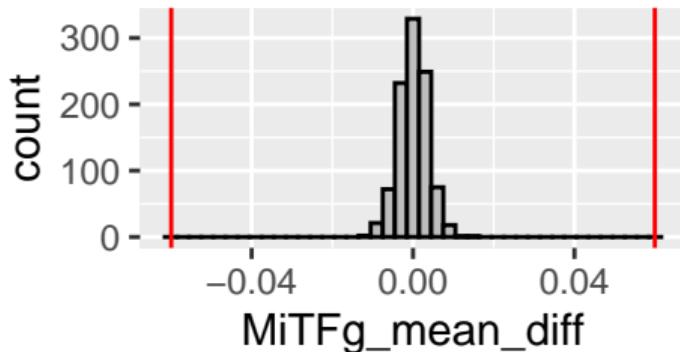
Hypothesis Testing: a method of statistical inference used to decide whether the data sufficiently support the null hypothesis.

- ① Set **null hypothesis**: there is no difference between the means of proteins in the control group and the treatment group.
- ② Calculate the difference between the means of proteins in the control group and the treatment group ($D_1 = \bar{x}_{treatment} - \bar{x}_{control}$).
- ③ Simulate **Sampling Distribution** (the distributions of the difference of two groups' means (D_2) when assuming the null hypothesis is true).
- ④ Compute **p-value** (the probability of observing a D_2 from the sampling distribution which is “as or more extreme” than D_1).
- ⑤ With this table, we can measure the strength of evidence from the perspective of p-value and decide whether the null hypothesis is true.

p-value	above 0.1	0.05 to 0.1	0.01 to 0.05	0.0001 to 0.01	below 0.0001
evidence against H_0	None	Weak	Moderate	Strong	Very Strong

Result (Q1)

P-value Evaluation for Sampling Distribution (red line: $\text{abs}(D1)$ and $-\text{abs}(D1)$)



Conclusion (Q1)

NGFR_p-value	AXL_p-value	Sox10_p-value	MiTFg_p-value
0	0	0	0

For all four proteins, the p-value is very close to zero, because none of our simulations under the null hypothesis were as large as D1 in project data.

This means, according to the above tables, we have very strong evidence against the null hypothesis, and we should reject the null hypothesis.

In other words, there is a difference between the means of proteins in the control group and the treatment group, showing that drugs do affect the amount of 4 phenotype indicator proteins.

Question 2

What are the correlations between each phenotype protein? How do they differ between different times?

Through the hypothesis tests, we know the drug has a significant impact on proteins. We want to find the dynamics of the cell by finding how each phenotype protein is related to the other. This can help us to understand if the changes in each phenotype protein are based on the effect of the drug or other proteins. We compare the result at different timepoints to verify our findings.

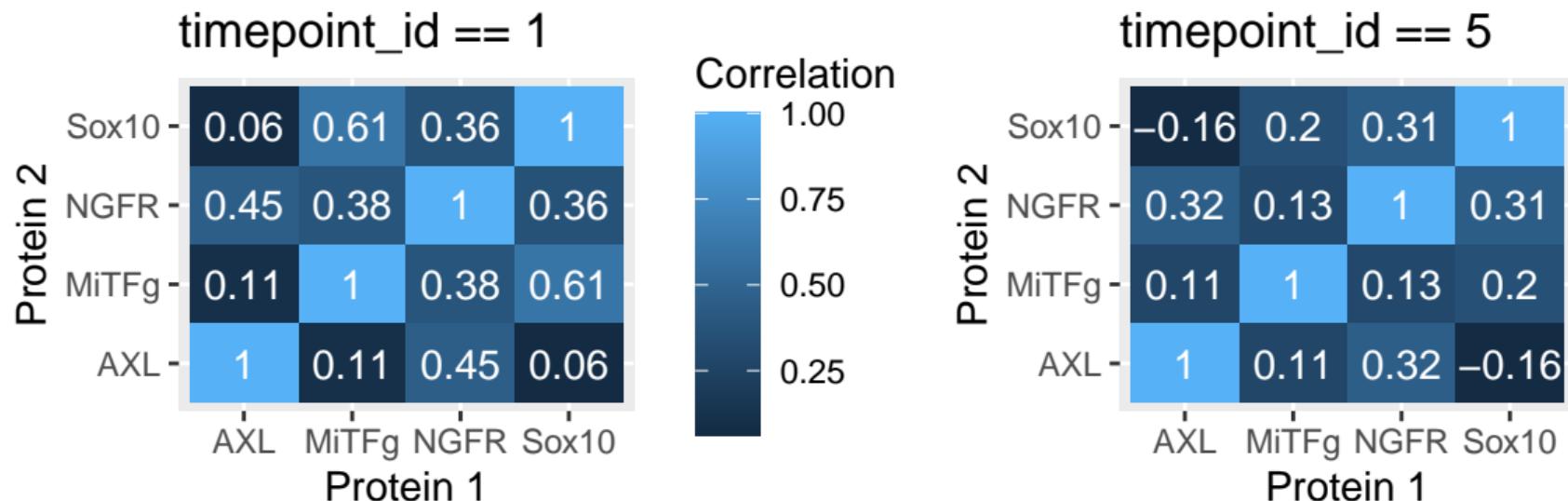
Statistical Method & Procedure(Q2)

We use correlation matrices to find if there exists any linear association between each protein. Furthermore, we have to compute two graphs with different timepoints to see if there is any change in the correlation between proteins.

If the correlation is positive, as one protein increases, the other also increases; as one decreases, the other also decreases. If the correlation is negative, as one protein increases, the other decreases. The closer the correlation to 1 or -1, the stronger the association. The closer the correlation to 0, the weaker the association. However, if one protein stays relatively constant despite changes in another protein, the correlation is also 0.

Result (Q2)

Correlation Matrices of Two Different Timepoints With the Same Drug (Vem) and Dosage (3.16 uM)



Results & Conclusion (Q2)

The majority of the correlations are less than 0.5 meaning a relatively small correlation between each protein or one protein may be constant. As time progresses, the correlation also decreases, meaning proteins are less correlated to each other. This suggests that the effects of the drug vary between each protein. For instance, the drug increases one protein but decreases others. Therefore, to truly understand how the drug affects the proteins, we must evaluate each protein individually.

Note that the correlation is just a linear association. Two proteins might correlate in other ways, such as exponentially. For future research, we can look at other types of associations.

Question 3

How do protein levels change over time t with and without drugs? What are the differences between the two conditions?

Based on the current discovery that the drug Vem is effective for changing protein levels and there are small correlations between proteins, we aim to further find out how the drug changes the protein level. Will the drug increase or decrease the overall protein level? Will the protein level changes bring any corresponding cellular phenotype change?

Statistical Method & Procedure(Q3)

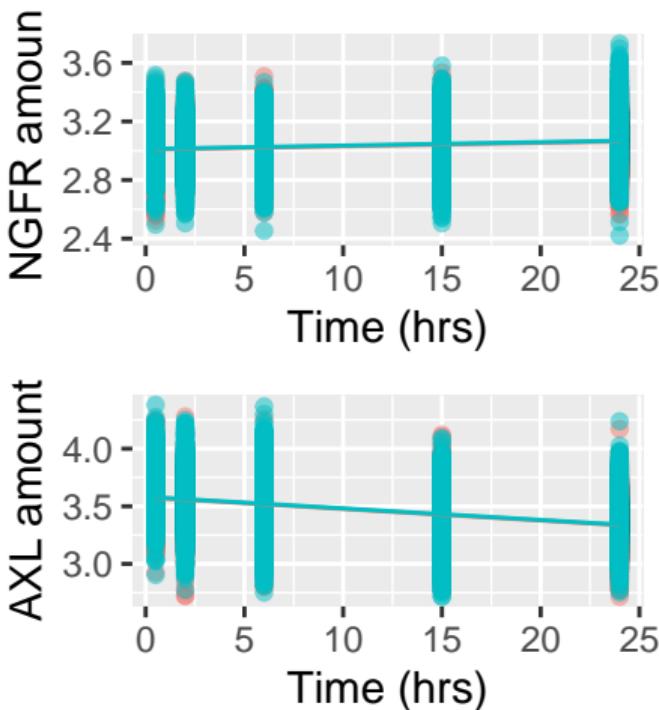
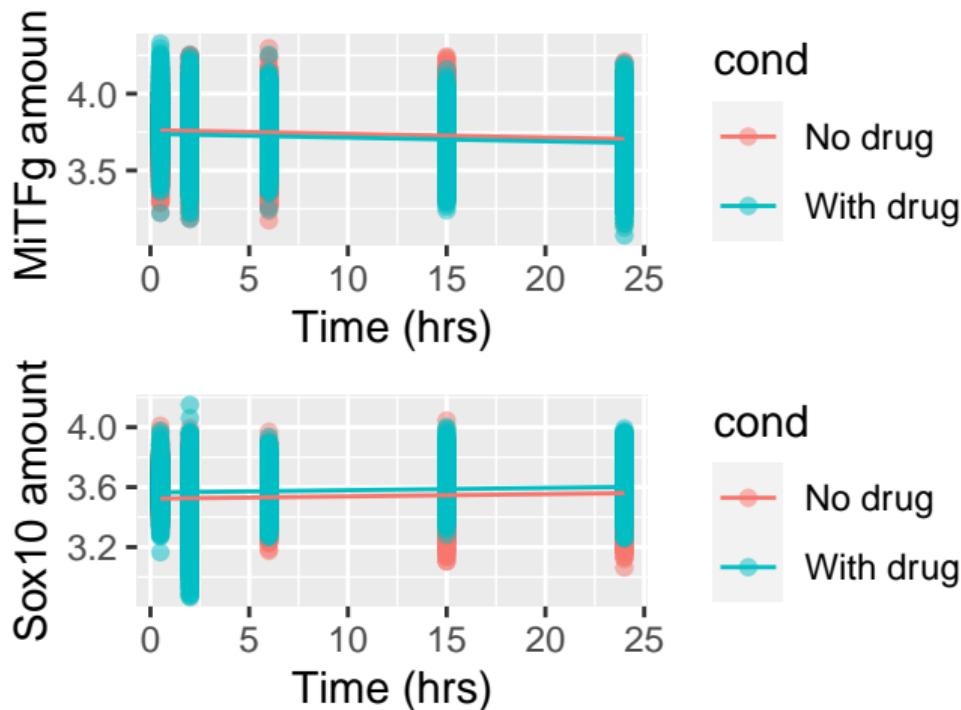
We use **multivariate regression** to find the relationship because multiple variables are considered to influence the protein level.

- One predictor is timepoint.
- The other one is the indicator variable “drug condition” which can be “No drug” ($dose_id = 1$) or “With drug”($dose_id = 5$).

We use the `lm()` function to find the fitted linear regression line and draw a graph to show the result of the multivariate regression for every protein.

Results (Q3)

Multivariate Regression Models for Phenotype Proteins



Fitted regression lines for 4 phenotype proteins

$$MiTFg_i = 3.76 - 0.00240t_i - 0.0243I[cond_i = \text{'Withdrug'}] + \epsilon_i$$

$$NGFR_i = 3.01 + 0.00233t_i + 0.00436I[cond_i = \text{'Withdrug'}] + \epsilon_i$$

$$SOX10_i = 3.52 + 0.00152t_i + 0.0415I[cond_i = \text{'Withdrug'}] + \epsilon_i$$

$$AXL_i = 3.58 - 0.0100t_i + 0.00425I[cond_i = \text{'Withdrug'}] + \epsilon_i$$

Phenotype protein	Association	Protein level after using drug
MiTFg	Negative	Decrease
NGFR	Positive	Increase
SOX10	Positive	Increase
AXL	Negative	Increase

The indicator variable changes the intercept of regression lines. To see the change in the graph, regression lines after using the drug shift upwards or downwards.

Conclusion(Q3)

Cellular Phenotype	Gene	MiTFg	NGFR	SOX10	AXL
Undifferentiated		LOW	LOW	LOW	HIGH
Neural crest-like		LOW	HIGH	HIGH	HIGH
Transitory		HIGH	HIGH	HIGH	LOW
Melanocytic		HIGH	HIGH	LOW	LOW

Based on the table above, we can find out that the risky phenotype melanocytic has a high level of MiTFg and NGFR, and a low level of SOX10 and AXL. According to our regression models, the use of the drug Vem could decrease MiTFg and increase NGFR, Sox10, and AXL.

Therefore, it is very likely that using the drug Vem can effectively turn the cellular phenotype from risky melanocytic to healthier neural crest-like or undifferentiated. In other words, taking the drug Vem may be an effective treatment to shift a bad cellular homeostasis to a good one.

However, we are concerned that though there is statistical significance by doing hypothesis testing, while comparing the fitted regression lines with and without drug use, the change in protein level brought by the drug Vem may still be too small to have practical significance.

Possible explanation:

- Our analysis only uses data up to 24 hours after the treatment. However, the drug Vem may take a longer time to show its effect. Ex: the drug starts taking significant effects after a week.
- All data are collected after only taking the drug once. However, the drug Vem may need to take regularly for a period to maximize its effect. Ex: taking the drug twice a day, for at least 2 weeks. The accumulative effect of the drug may be much stronger.

Final Conclusion

We were able to prove the effectiveness of the drug through our hypothesis testing analysis. The smaller correlation between proteins over time indicates that the drug imposes various effects on every protein. The linear regression model indicates that there is a relationship between protein levels and multiple variables such as time and the indicator variable drug condition. The difference of overall protein levels caused by the drug reflects its ability to return the bad melanocytic cell back to good homeostasis.

Limitation & Future Direction

Limitations: Our research focused on the phenotype proteins which are considered as “outcomes” and lacked the analysis of the “causes” 22 transcription factors. The research would be more inclusive if we analyse the relationship between “causes” and “outcomes”.

Only the effectiveness of drug Vem was tested in our research.

Variables were missing from time IDs 6 and 7 under the condition that no drug was introduced to the sample, therefore we were unable to compare phenotype protein level changes with and without the effects of the drug at these timepoints, limiting our analysis to a 24hr time frame.

Future Direction: Do more analysis on transcription factors, like their correlation with certain phenotype proteins.

Apply our model to a different drug to test its effectiveness, as well as compare it to the drug we built our model around to potentially identify a similar or even more effective treatment.

References and Acknowledgements

This data was developed by three professors from the University of Virginia and their research teams. Their names are Natacha Comandante-Lou, Douglas G. Baumann, and Mohammad Fallahi-Sichani.

We also would like to thank Jenny Oh and Scott Schwartz for their helpful suggestions and comments that improved the presentation of this poster.