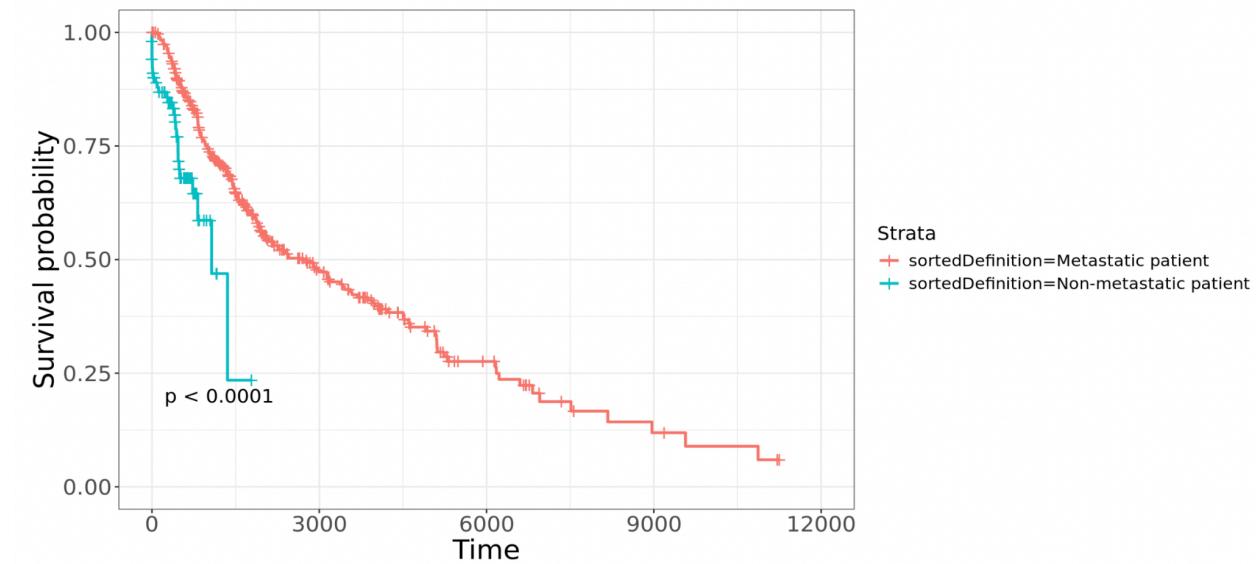


Part 3: Results and Interpretations

For each analysis, include an image of the relevant plot you created in Part 2 and a 3-4 sentence description answering the following question:

- Analyze the plot. What conclusions can you and can you not draw about differences between metastatic and non-metastatic TCGA SKCM patients? Why?

1) Difference in survival between metastatic and non-metastatic patients

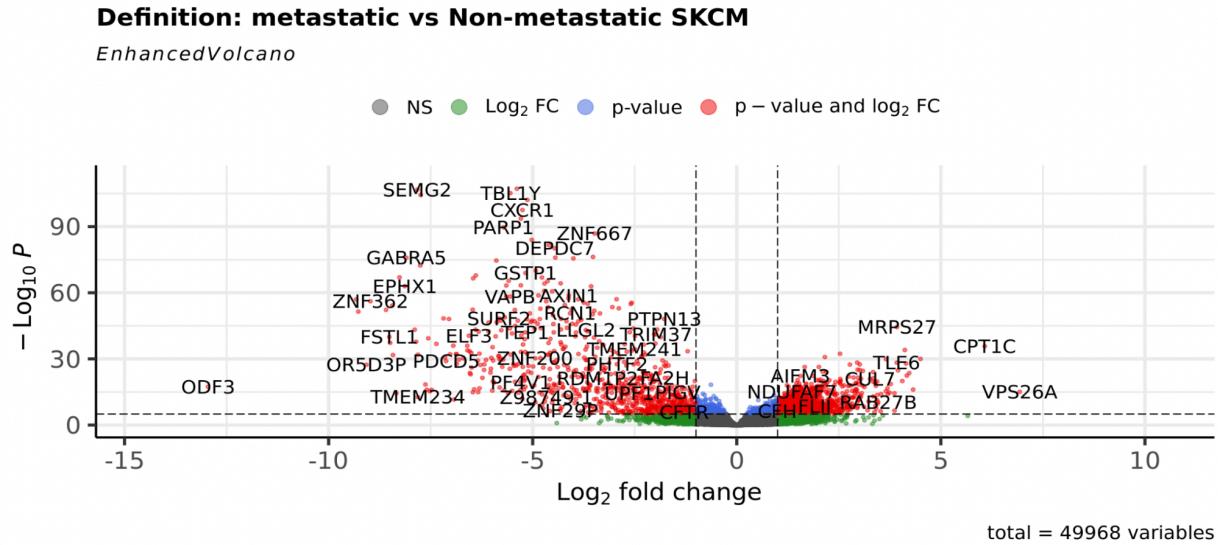


This Kaplan-Meier (KM) plot compares the survival probabilities of patients with metastatic SKCM to those with non-metastatic SKCM. The p-value is less than 0.0001, indicating statistical significance. Based on this graph, one might conclude that patients with metastatic SKCM have a relatively higher survival probability compared to those with non-metastatic SKCM. However, this interpretation is incorrect, given what we know about cancer progression. Non-metastatic cancer is generally less invasive and less lethal than metastatic cancer across all cancer types.

The reason why the plot appears to show that patients with non-metastatic cancer have a lower survival probability is due to one of the limitations of the KM survival plot—censored data. Censored data occurs when survival information is incomplete. In this analysis, survival time for patients with NA in the death_days_to column is taken as the days_to_last_follow_up value instead. This is because, in non-metastatic SKCM, the lower lethality rate and higher cure rate mean that many patients survive and remain cancer-free after treatment. For these patients, the death_days_to column is NA, and we can only use their days_to_last_follow_up as the survival time.

The issue with the data is that, over time, many of these patients stop updating their follow-up information once they are cancer-free, leading to a loss of data for non-metastatic patients as the study progresses. As a result, the KM plot shows an artificially higher survival probability for metastatic SKCM patients. In reality, however, non-metastatic SKCM patients have a better survival prognosis over time.

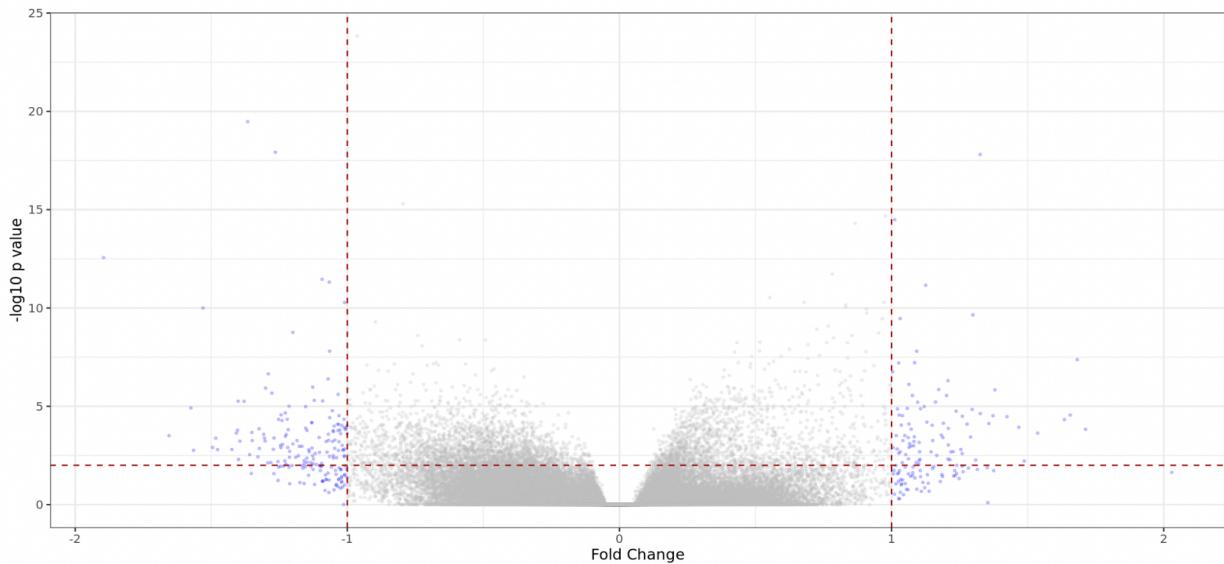
2) Expression differences between metastatic and non-metastatic patients



In this enhanced volcano plot, the threshold for Log2 fold change is set at ± 1 , and the threshold for -Log10 P-value is 1.3. The genes represented by red dots on the upper right are significantly upregulated in metastatic SKCM compared to non-metastatic SKCM. Examples include MRPS27, CPT1C, VPS26A, and others. Genes in red dots on the upper left are significantly downregulated in metastatic SKCM compared to non-metastatic SKCM, with examples such as SEMG2, ODF3, and GABRA5. Those genes are the genes we can draw conclusions from.

The blue dots represent genes that are significant, but their Log2 fold change is not large enough to be classified as either upregulated or downregulated. The green dots represent genes that are not statistically significant but show sufficient upregulation or downregulation to be noteworthy. Finally, the grey dots represent genes that are neither statistically significant nor sufficiently upregulated or downregulated. Those genes are the genes that we cannot draw conclusions from.

3) Methylation differences between metastatic and non-metastatic patients



In this ggplot illustrating the methylation differences between metastatic and non-metastatic patients, the blue dots in the upper right portion represent CpG sites that are hypermethylated in metastatic cancer compared to non-metastatic cancer. Similarly, the blue dots in the upper left portion represent CpG sites that are hypomethylated in metastatic cancer relative to non-metastatic cancer. Those other dots represent CpG sites that are either statistically insignificant or show insufficient methylation changes (either hyper- or hypomethylation), or even both to be considered noteworthy.

4) Direct comparison of transcriptional activity to methylation status for 10 genes

The graphs are all in the qbio490_coding4_pictures folder, I didn't put them on here for the sake of space.

ABCG5: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

AP1M2: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

CETN1: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

DIXDC1: For expression of this gene, there is a clear difference in expression among the non-metastatic and metastatic groups. The metastatic groups are express more than the non-metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites but there are exceptions in some sites.

IBA57: For expression of this gene, it does not look like there is a significant difference in expression amone the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

GHSR: For expression of this gene, it does not look like there is a significant difference in expression amone the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

H1C1: For expression of this gene, there is a clear difference in expression among the non-metastatic and metastatic groups. The metastatic groups are express more than the non-metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites but there are exceptions in some sites.

K1F13A: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypomethylated in most of the CpG sites.

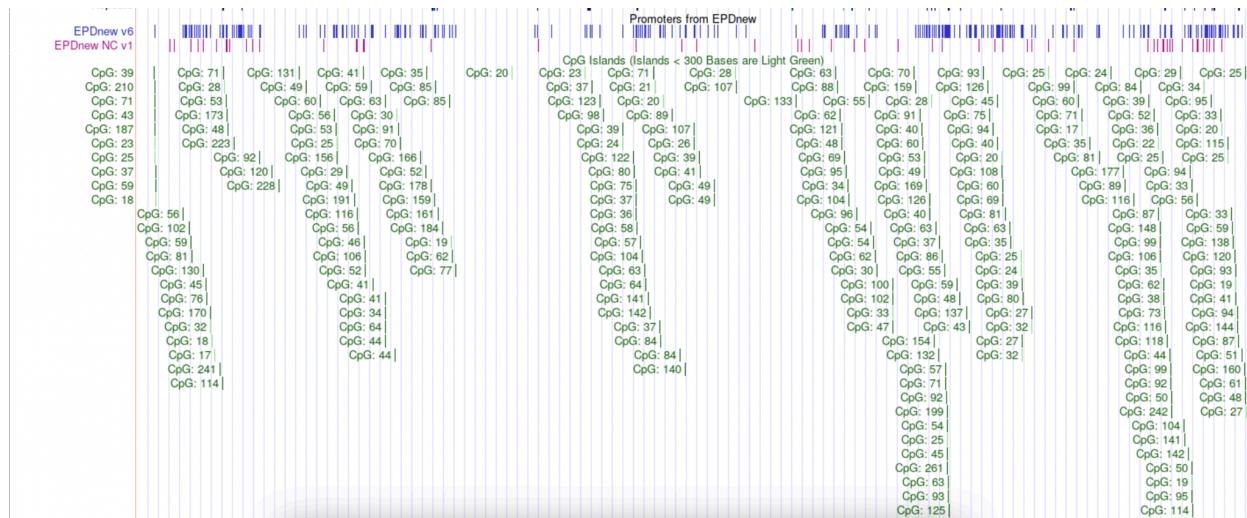
LINC00482: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

MT1E: For expression of this gene, it does not look like there is a significant difference in expression among the non-metastatic and metastatic groups. In terms of methylation, metastatic cancer is relatively more hypermethylated in most of the CpG sites.

Overall, the metastatic and non-metastatic patients didn't have a huge difference as for gene expression, but metastatic patients' genes are more methylated than non-metastatic patients.

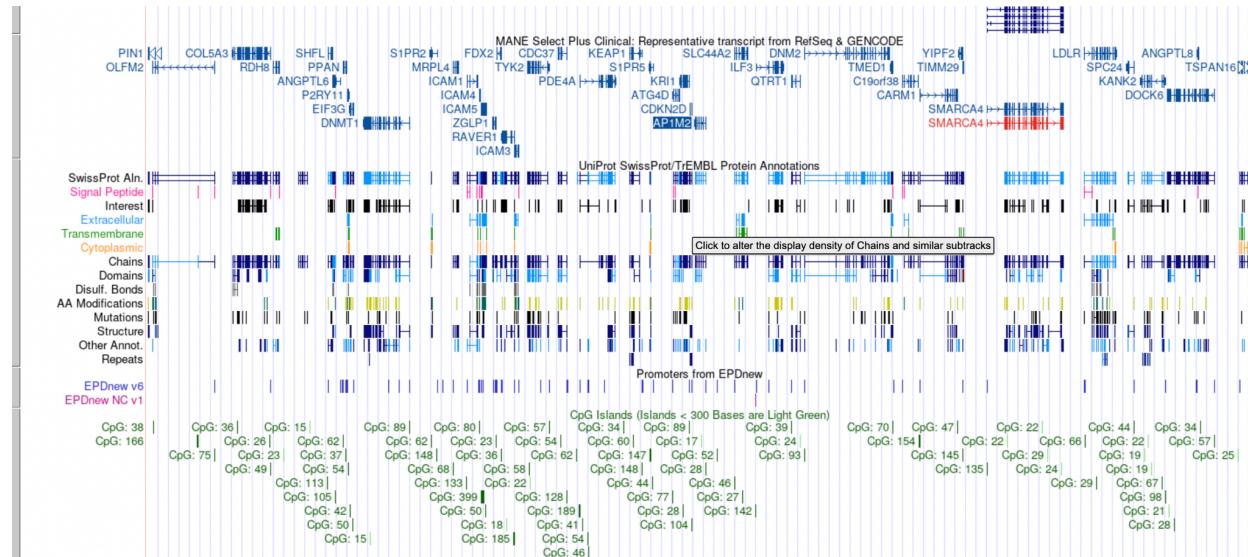
5) Visualization of CpG sites and protein domains for 3 genes (use UCSC genome browser) for a few genes. Describe at least one academic article (research or review) that either supports or doesn't support your final conclusion for one of the genes. If previously published work doesn't support your analysis, explain why this might be the case.

GHSR:



Based on this image on the char3 (q24-29) region on the GHSR gene, we can see a lot of CpG sites aligning with promoter regions, indicating the hypermethylation and downregulation of the expression and methylation on the gene.

AP1M2:



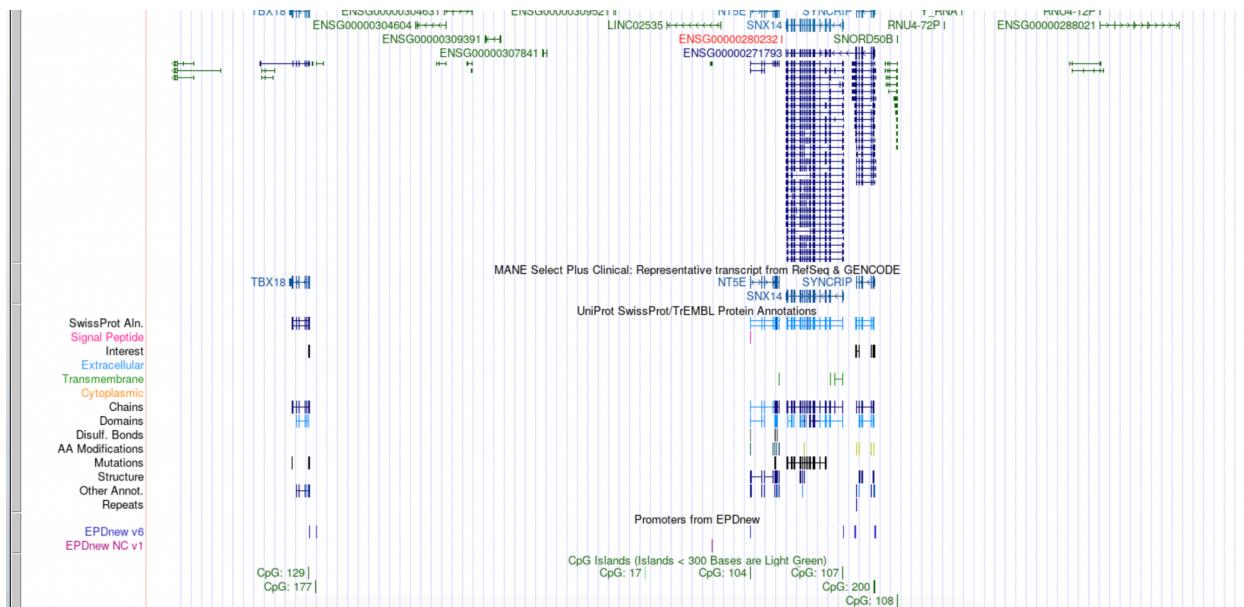
Based on this image on the chr19 (q13.2) region on the AP1M2 gene, we can see a lot of CpG sites aligning with a lot of promoter regions, indicating the hypermethylation and downregulation of the expression and methylation on the gene.

CETN1:



Based on this image on the char18 (p11.32-q11.1) region on the CETN1 gene, we can see a lot of CpG sites aligning with a lot of promoter regions, indicating the hypermethylation and downregulation of the expression and methylation on the gene.

KIF13A:



Based on this image on the char6 (p14.3) region on the KIF13A gene, we can see a lot of CpG sites aligning with a lot of promoter regions, indicating the hypermethylation and downregulation of the expression and methylation on the gene.

Research article: In the paper "Revisiting miRNA Association with Melanoma Recurrence and Metastasis from a Machine Learning Point of View", Korfiati et al. wrote a paper summarizes existing studies on miRNAs as diagnostic and prognostic biomarkers in cutaneous melanoma (CM) and develops a prediction model using publicly available NGS data. By combining network analytics and machine learning, the authors identify two miRNA signatures that accurately predict CM recurrence and metastasis, with validation accuracy of 73.85% and

82.09%, respectively. In the research they found that : “Similarly, the putative target genes with the most negative correlation to metastasis were RNF44, CDK20, HDAC5, IGSF11, KCNJ13, KCTD15, **KIF13A**, LDB1, LRP6, PLXNB1, RBMS2, RNF144A, SOCS7, SOX4, TTC28, TTPA, UTP25, ZKSCAN8, ZNF264, ZNF713 (Figure 1B, red circles)”(Korfiati et all). KIF13A is involved in the putative target genes with the most negative correlation to metastasis, meaning that their expression are linked with less metastasis. This kind of supports my assumption that KIF13A’s expression and methylation is linked with cancer’s metastasis, which made this gene one of the “interested genes” .

Works Cited

- 1.Korfiati A, Grafanaki K, Kyriakopoulos GC, et al. Revisiting miRNA Association with Melanoma Recurrence and Metastasis from a Machine Learning Point of View. *International journal of molecular sciences*. 2022;23(3):1299-.
International journal of molecular sciences. 2022;23(3):1299-.
doi:10.3390/ijms23031299