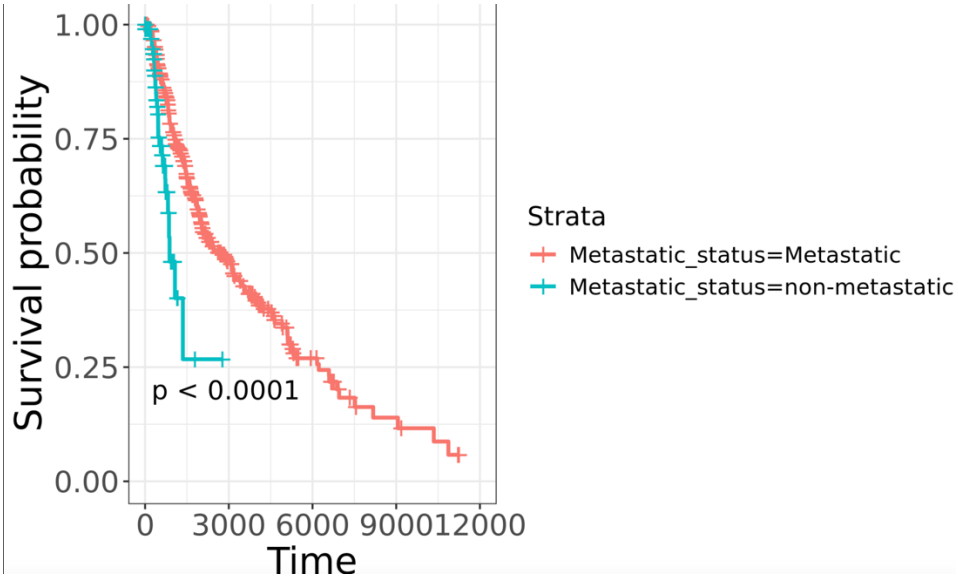
(part 1 question is in the file "review_project.Rmd" together with part 2 coding)

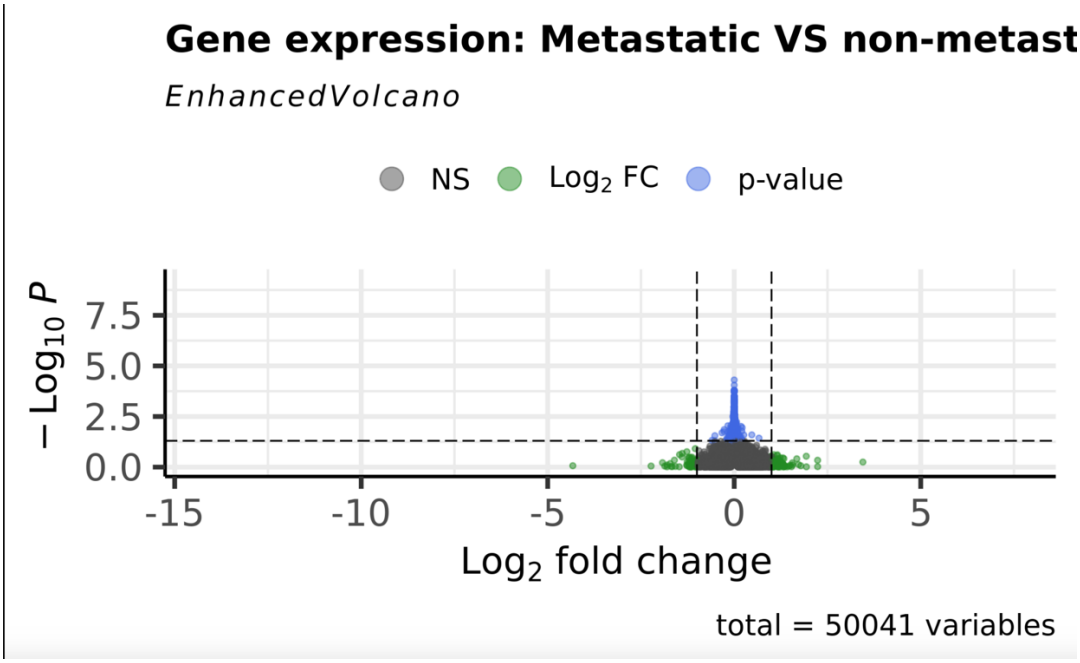
Part 3: Results and Interpretations

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



The p-value ($p < 0.0001$) shows a statistical significance in survival difference between metastatic and non-metastatic patients. The non-metastatic patients seem to have an advantage in early survival, but its survival probability drop significantly after about 1000, being lower than that of metastatic patients, which is somehow counter-intuitive. This can be misleading because the number at risk for non-metastatic patients is much lesser(higher censoring) than that of metastatic patients in TCGA database.

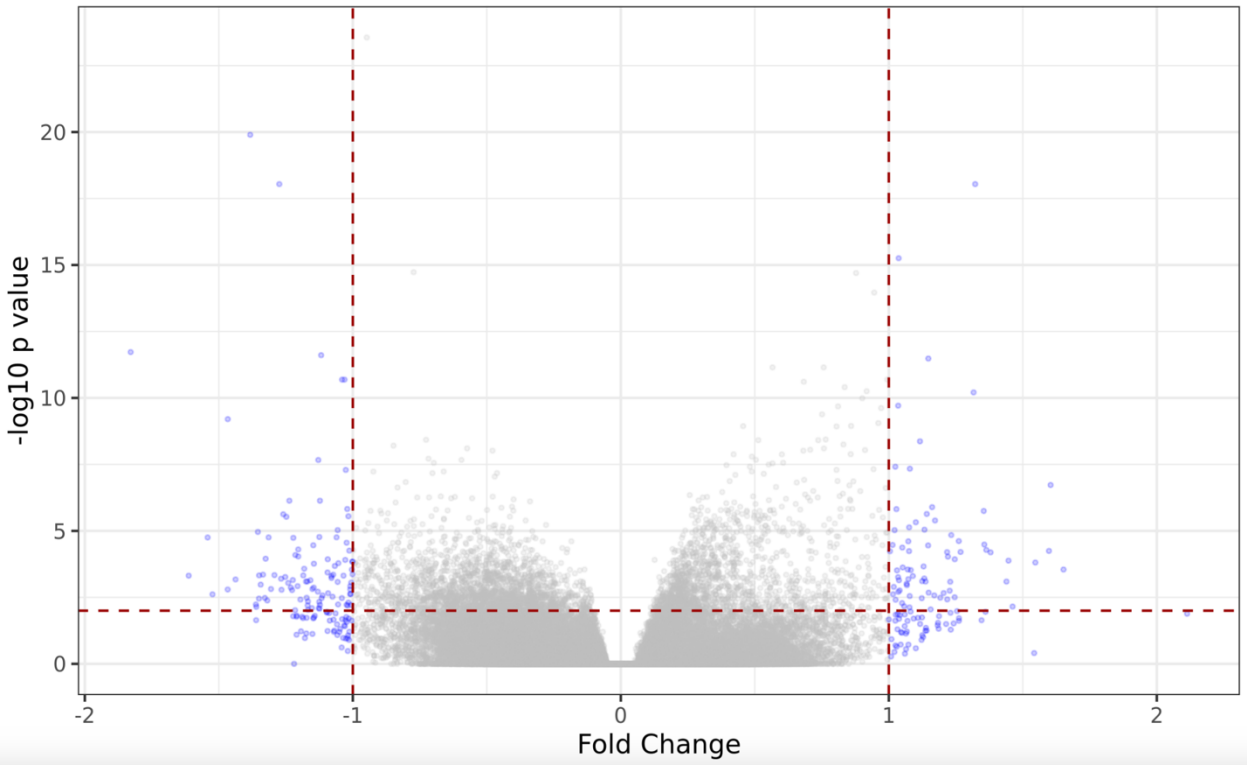
2) Expression differences between metastatic and non-metastatic patients



Threshold padj value is at 0.05 and log2FoldChange at |1|. The metastatic samples are the comparison group, while the non-metastatic samples are the baseline group. The volcano plot shows the upregulation and downregulation of genes in metastatic samples **compared to non-metastatic sample**.

The genes at bottom left corner are those insignificantly downregulated. The genes at the bottom right corner are insignificantly upregulated. Many genes fall in the middle which means they only have a minor change in expression, and most of them are insignificant, though there is a portion of genes in blue have a significant but minor change in expression level.

3) Methylation differences between metastatic and non-metastatic patients



Metastatic sample is set to 1 and non-metastatic sample is 0. Therefore, this volcano plot shows the hypermethylation and undermethylation of genes across CpG sites of metastatic samples relative to non-metastatic sample.

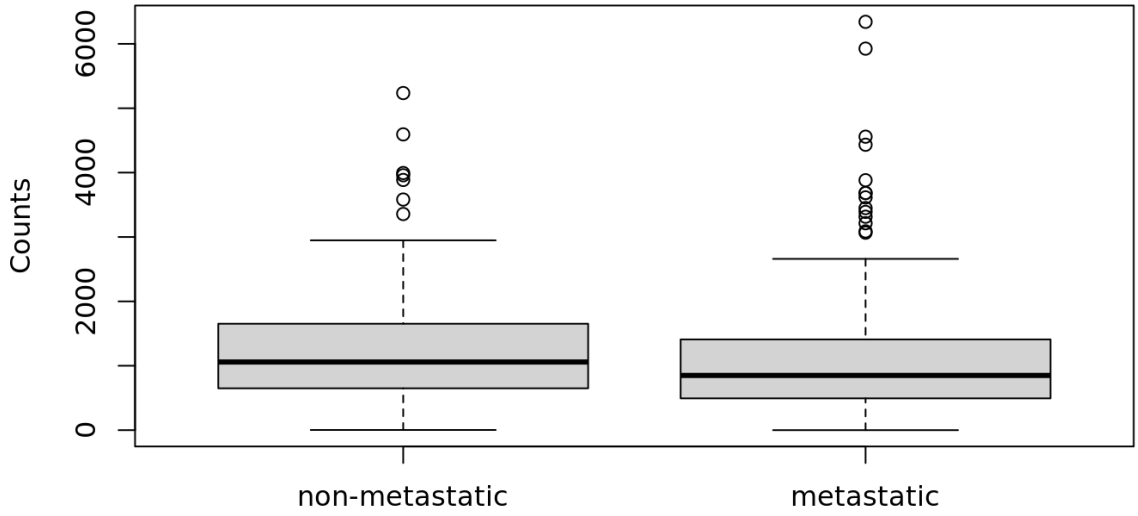
There are a few genes which are in blue shows a difference in methylation in metastatic sample compared to non-metastatic sample. Those at left bottom are insignificantly undermethylated, while those at left top are significantly undermethylated; Those at right bottom are insignificantly undermethylated, while those at right top are significantly hypermethylated. The genes in grey which fall in the middle shows little change in methylation level.

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

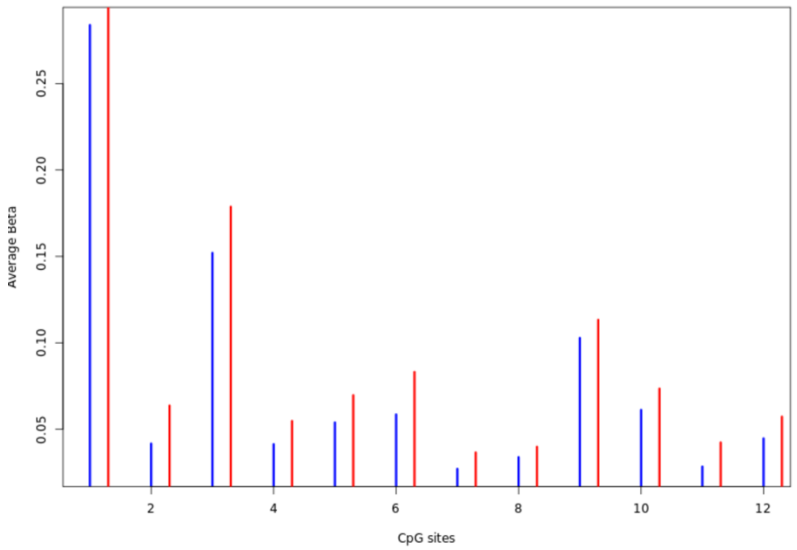
```
Interested genes:
[1] "MAD1L1"    "CD9"       "ZC3H3"     "RUNX3"     "PNPLA6"    "SFSWAP"    "EPN1"
"HDAC4"
 [9] "GRAMD4"    "RAP1GAP"   "TOLLIP"    "SLC7A8"    "HAS3"      "CLCN7"     "MAX"
"KLF16"
[17] "MLLT1"     "ITGB4"     "UNK"       "GYPC"      "GMPR"      "IRF4"
"MAPK8IP3"  "VPS37B"
[25] "PML"       "TBCD"      "COL6A1"    "OCIAD2"    "NPFFR1"    "DIP2C"     "TMEM164"
"AGAP1"
[33] "KRTCAP3"   "NRG2"      "CTBP1"     "RGS12"     "ABR"       "FOXK1"     "RCCD1"
"TBC1D16"
[41] "OTUD3"     "BANP"      "CD164L2"   "MSRA"      "BNIP3"     "IRX5"      "CETN1"
"PLEC"
[49] "APOLD1"    "ALX1"      "TACSTD2"   "LINC00482" "PCGF3"     "P4HB"      "MBP"
"PRRT1"
```

```
[57] "CTAGE1"    "LINC00887"
```

Take gene "ALX1" as an example:

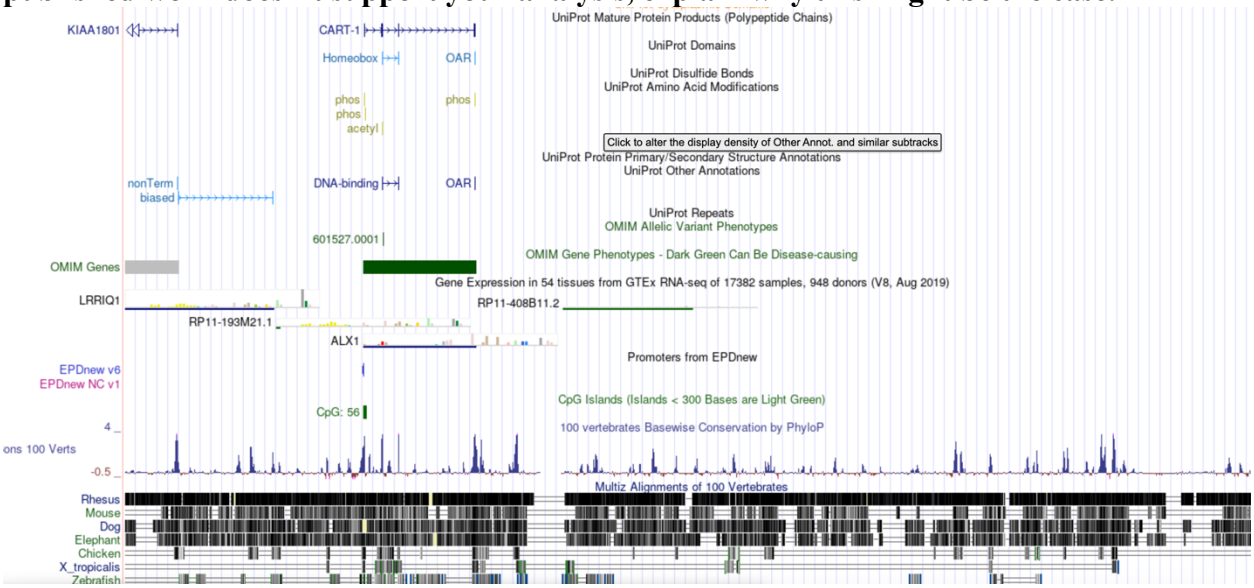


The boxplot shows that gene ALX1 has a slightly lower trasncriptional expression in metastatic samples compare to non-metastatic sample by looking at the median and maximum of two box-and-whisker plot. Also, there is a relatively smaller variability in expression of ALX1 in metastatic sample as it has a smaller interquartile range. Given the trend of lower and more consistent expression of ALX1 in metastatic samples, ALX1 could potentially be a biomarker for metastasis, or it may be involved in pathways that regulate metastatic behavior, although the difference in expression is not that significant between metastatic sample and non-metastatic sample.



No particular large difference in methylation between metastatic and non-metastatic is found across CpG sites, but metastatic samples (red) showed a very consistent hypermethylation (though that difference is small) compared to non-metastatic sample(blue) in all CpG sites

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.



This is a visualization of CpG sites and protein domains for ALX1. There is a CpG site labeled(CpG56), and a promoter region can be found there. This may explain why the hypermethylation in gene ALX1 across CpG sites in metastatic sample lead to a relatively lower transcriptional expression in gene ALX1 compared to non-metastatic sample.

Jiao et al's finding that ALX1 is highly expressed in melanoma tissues and promotes cancer cell proliferation, invasion, and EMT contrasts with my conclusion that ALX1 is downregulated in metastatic samples compared to non-metastatic ones, potentially due to hypermethylation (Jiao et al., 2019). Similarly, Yao et al.'s study shows ALX1 upregulation in lung cancer, especially in metastatic cases, with ALX1 promoting cell proliferation, migration, and invasion, while silencing it inhibits these processes (Yao et al., 2015). Although the role of ALX1 might be context-dependent (which could be different in the case of SKCM), ALX1 is more often found as upregulated in cancer in general and thus contributing to metastasis. The contradiction of my conclusion with these research may be due to the limited number of metastatic samples in my analysis, as most of the samples in `rna_se@colData$definition` are "Primary solid Tumors" and only a few are "Metastatic".

References

Jiao, J.-X., Jiao, L.-J., Yang, S., & Zhao, Y.-J. (2019). Knockdown of aristaless-like homeobox1 inhibits epithelial-mesenchymal transition through Wnt/β-catenin signaling pathway in melanoma cells. *Biochemical and Biophysical Research Communications*, *511*(1), 105–110. https://doi.org/10.1016/j.bbrc.2019.02.050

Yao, W., Liu, Y., Zhang, Z., Li, G., Xu, X., Zou, K., Xu, Y., & Zou, L. (2015). ALX1 promotes migration and invasion of lung cancer cells through increasing snail expression. *PubMed*, *8*(10), 12129–12139.