Determining the Prognostic Value of the DNA Methylation of the *GYPC*, *NME1*, and *SLIT2* Genes in Human Lung Adenocarcinoma

TMED040

Uma Paul

Tesla STEM High School

Redmond, Washington, USA

All images/charts/graphs were created by the finalist unless otherwise referenced

- Introduction: Relevant Research -

Lung cancer is responsible for the highest number of cancer-related deaths worldwide, causing around 160,000 deaths per year in the United States alone. Lung adenocarcinoma (LUAD) is the most common subtype of non-small cell lung cancer and accounts for 40% of all lung cancers.

Prognosis helps doctors determine the appropriate course of treatment for patients and provides patients and their families with a more accurate prediction of the patient's expected survival.

The Genes

The genes *GYPC, NME1, and SLIT2* have been found to be associated with smoking-related LUAD, and their expression may be controlled by DNA methylation [1].

GYPC is an integral membrane glycoprotein carried by human erythrocytes that regulates cell stability. It has been associated with the JAK/STAT and cell adhesion signaling pathways, which aid in tumor progression and metastasis [1][2].

SLIT2 is a secreted glycoprotein that aids in cellular migration and has displayed tumor suppressing activity due to its involvement in the epithelial-mesenchymal transition (EMT), which increases metastatic and invasive activity and is linked to cancer stem cells [1][2].

NME1, or nucleoside diphosphate kinase 1, is a protein associated with lower mRNA levels in metastatic cells. It has been shown to be associated with the cell cycle pathway [1].

- Introduction: Goals -

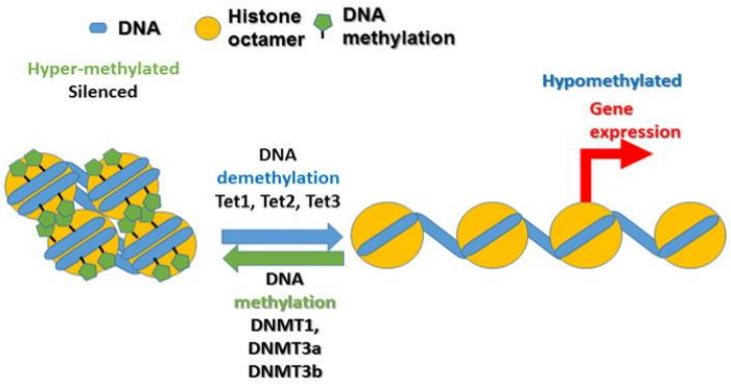
What is the prognostic value of the DNA methylation of *GYPC*, *NME1*, and *SLIT2* in human lung adenocarcinoma?

1. The first goal is to create a program to analyze DNA methylation data and clinical data to determine significant differences in *GYPC*, *NME1*, and *SLIT2* DNA methylation between clinical features and between tumor and non-tumor tissue using ANOVA.

2. The second goal is to determine the impact of the DNA methylation levels of *GYPC*, *NME1*, and *SLIT2* and demographic on overall survival using Kaplan-Meier survival estimates and log-rank statistical tests.

3. The final goal is to analyze all of the CpG sites in the 27K methylation array to determine which sites are associated with lung adenocarcinoma overall survival through the log-rank test.

- Framework -



A diagram of the process and function of DNA methylation from "The roles of DNA methylation in the stages of cancer" [7] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5657558/

Key Terms:

CpG sites: Most DNA methylation occurs at sites in the DNA sequence where a cytosine is followed by a guanine

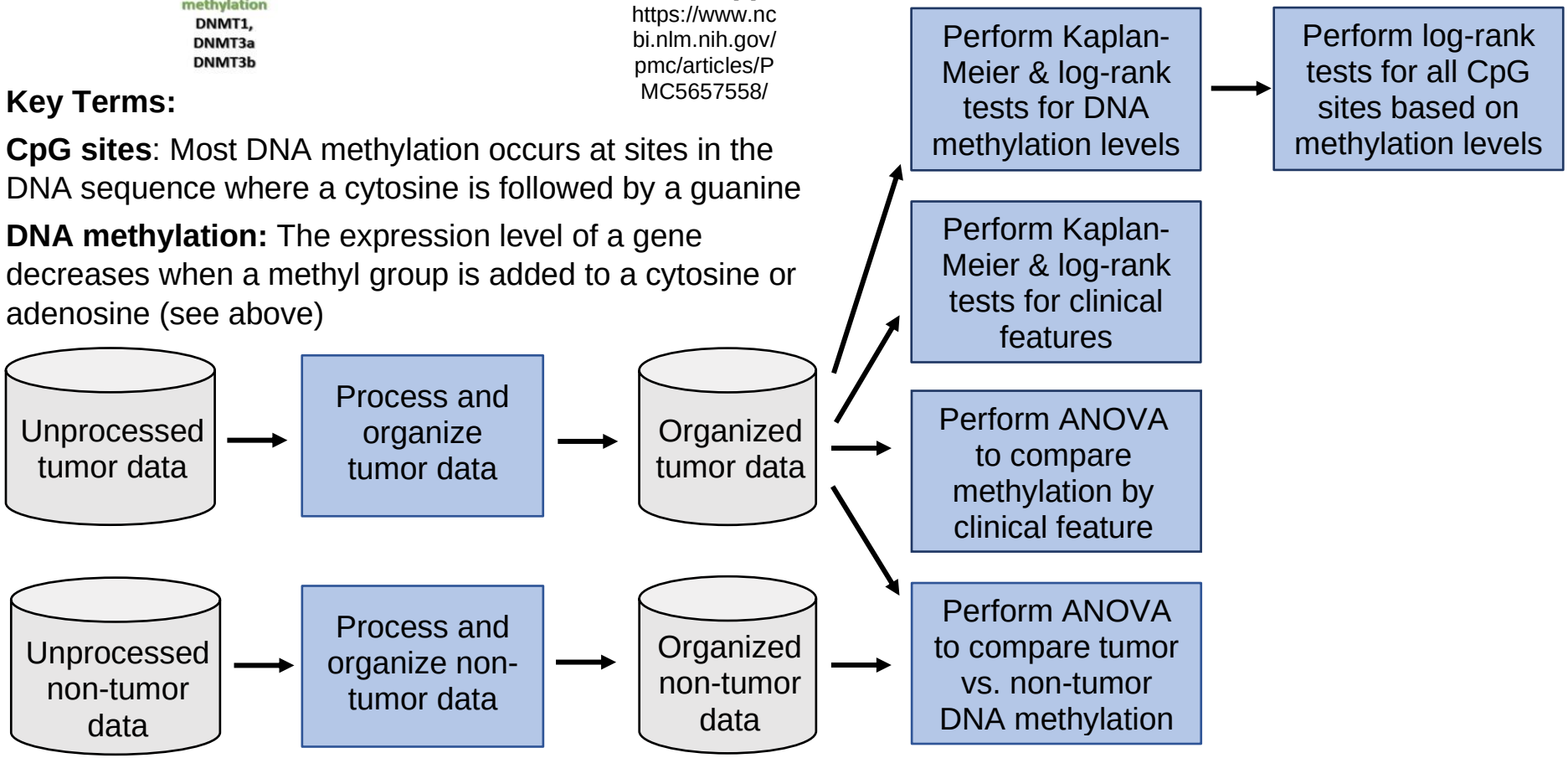
DNA methylation: The expression level of a gene decreases when a methyl group is added to a cytosine or adenosine (see above)

Programming:

- Done in MATLAB
- ANOVA: MATLAB anova1 method
- Kaplan-Meier: MATLAB ecdf method
- Log-rank test: original logranktest method

Databases:

- TCGA (tumor) - 66 samples
- GSE32861 (non-tumor) - 60 samples



Self-made flowchart of the data analysis program

- Findings -

ANOVA: Tumor vs. Non-tumor

		***p*-value**					
		GYPC *(cg13901526)*	***SLIT2*** *(cg03742003)*	***NME1*** *(cg04380669)*	***GYPC*** *(cg17105014)*	***SLIT2*** *(cg18972811)*	***NME1*** *(cg01063524)*
Sex	*Male*	*0.1640*	*0.1290*	*0.9963*	*0.0769*	*0.0712*	*0.1006*
	Female	*0.3586*	*0.0015*	*0.0307*	*5.4529e-04*	*0.0059*	*0.6557*
Smoking	*Nonsmoker*	*0.4050*	*0.0191*	*0.0430*	*0.0925*	*0.0285*	*0.8127*
	Current Smoker	*0.1053*	*0.1919*	*0.5042*	*0.3377*	*0.0266*	*0.4240*
Age	*40-59*	*0.6700*	*0.1129*	*0.0134*	*0.1137*	*0.0699*	*0.5829*
	60-79	*0.0831*	*0.0084*	*0.4246*	*0.0030*	*0.0098*	*0.1280*
	80+	*0.2478*	*0.9467*	*0.6995*	*0.2937*	*0.1995*	*0.8444*
Stage	*I*	*0.3988*	*0.0015*	*0.1793*	*5.8910e-04*	*0.0018*	*0.2860*
	II	*0.9620*	*0.5456*	*0.6409*	*0.3042*	*0.2867*	*0.4588*
	III	*0.3848*	*0.1434*	*0.0017*	*0.5659*	*0.1975*	*0.8463*
	IV	*0.3406*	*0.4730*	*0.0060*	*0.2978*	*0.4871*	*0.0902*
	Overall	*0.0678*	*4.2279e-04*	*0.0781*	*1.2376e-04*	*5.0077e-04*	*0.2361*

Table 1.1: The *p*-values when comparing tumor DNA methylation of *GYPC*, *NME1*, and *SLIT2* to non-tumor tissue

Significant differences in tumor vs. non-tumor tissue

- Hypermethylation of genes like *GYPC*, *NME1*, and *SLIT2* can lead to gene silencing
- *SLIT2* is considered a tumor suppressor and *NME1*, a metastasis suppressor
- This study found that several *GYPC*, *NME1*, and *SLIT2* CpG sites were hypermethylated in tumor tissue but only in certain demographics such as females, who showed differential methylation for 4 sites while males did not for any
    - It has been found that females have a significantly lower level of DNA methylation than males. One possible explanation is X-chromosome inactivation in females
- Differences in the *p*-values across a clinical feature may indicate differences in the role of the gene and the CpG site between demographic groups

ANOVA: Clinical Features

	***p*-value**					
	GYPC *(cg13901526)*	***SLIT2*** *(cg03742003)*	***NME1*** *(cg04380669)*	***GYPC*** *(cg17105014)*	***SLIT2*** *(cg18972811)*	***NME1*** *(cg01063524)*
Sex	*0.4026*	*0.9116*	*0.8644*	*0.6285*	*0.5200*	*0.5374*
Smoking History	*0.6159*	*0.9276*	*0.3401*	*0.6414*	*0.0478*	*0.8516*
Age	*0.7471*	*0.7176*	*0.1806*	*0.5722*	*0.8933*	*0.2900*
Cancer Stage	*0.0808*	*0.0228*	*0.8521*	*0.1326*	*0.5384*	*0.7526*

Table 1.2: The *p*-values when comparing the DNA methylation of *GYPC*, *NME1*, and *SLIT2* based on clinical features

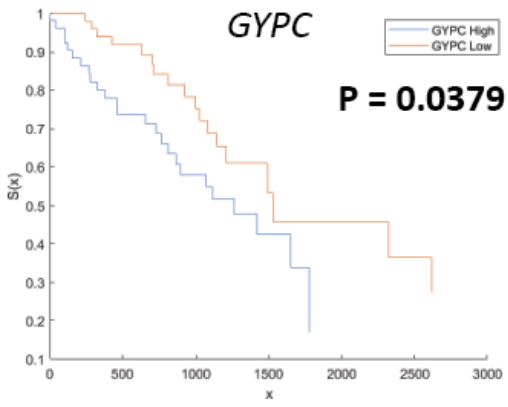
Significant differences in *SLIT2* methylation by smoking history

- These three genes were determined by a study to be differentially expressed based on smoking history [1]
- Therefore, it is likely that there would be differential methylation by smoking history
- The lack of differential methylation for other CpG sites could be due to the fact that many epigenetic factors other than DNA methylation affect expression levels such as histone acetylation

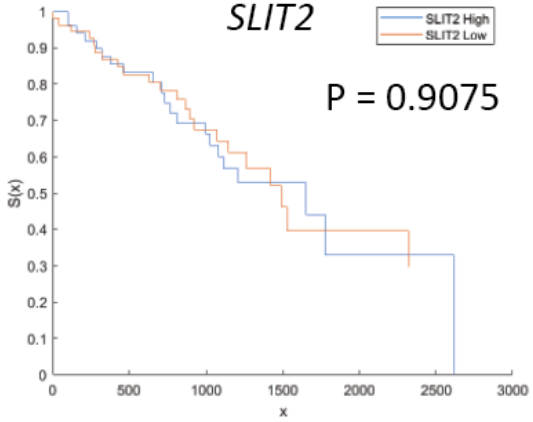
Significant results *SLIT2* methylation by cancer stage

- *SLIT2* plays a role in the EMT as a tumor suppressor gene that inhibits tumor migration and growth
- Lower *SLIT2* expression has been associated with poor prognosis in breast cancer patients [1]
- In this study, *SLIT2* methylation increased as the cancer stage increased

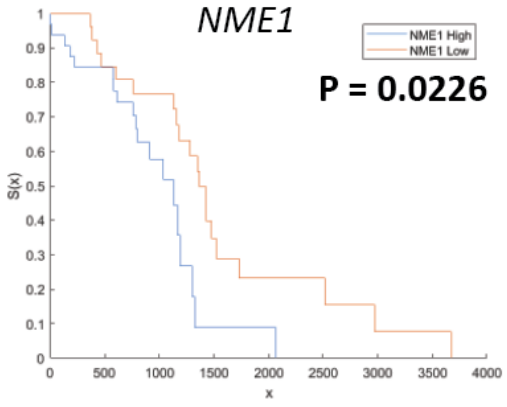
Kaplan – Meier Results



Kaplan-Meier curve comparing low and high cg13901526 methylation levels



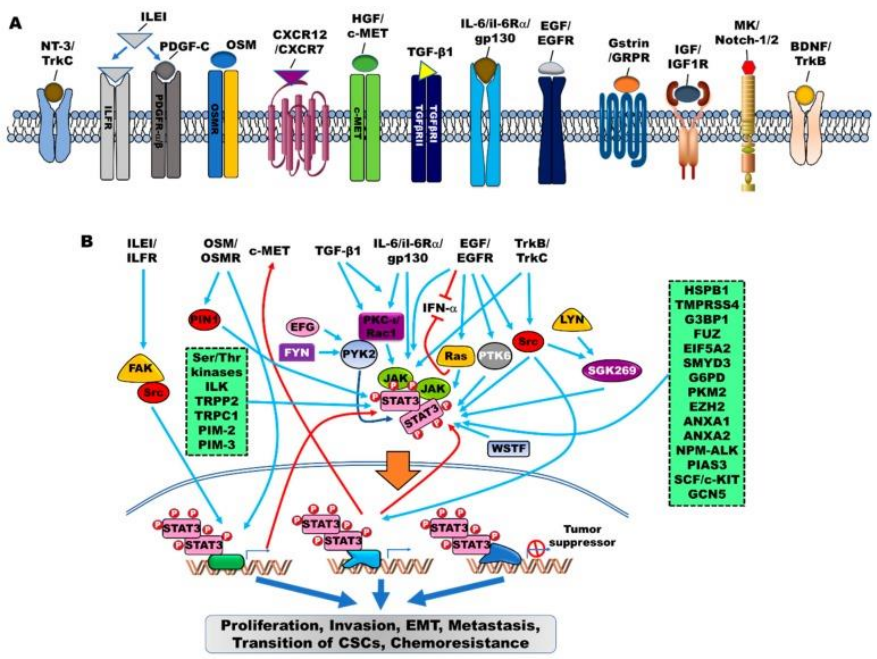
Kaplan-Meier curve comparing low and high cg18972811 methylation levels



Kaplan-Meier curve comparing low and high cg01063524 methylation levels

Significant results in survival by *NME1* methylation

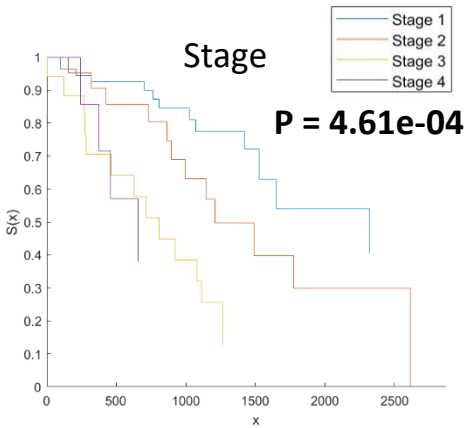
- There have been differing results regarding the correlation between *NME1* expression and prognosis for different types of cancer
    - High expression associated with favorable prognosis in breast cancer and melanoma [1]
    - High expression associated with poor prognosis in neuroblastoma and cervical cancer [1]
- This study found that higher *NME1* methylation, and therefore likely lower *NME1* expression, is associated with poor prognosis



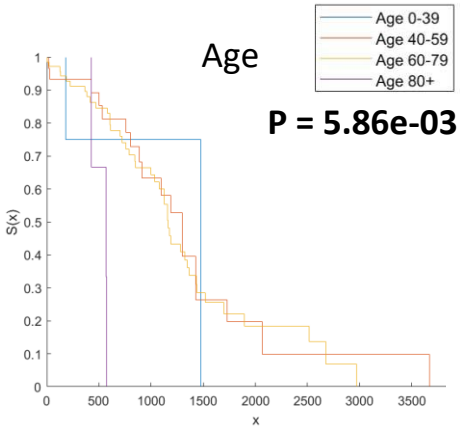
A diagram illustrating the JAK/STAT pathway from “Role of JAK/STAT3 Signaling in the Regulation of Metastasis, the Transition of Cancer Stem Cells, and Chemoresistance of Cancer by Epithelial-Mesenchymal Transition” [2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7017057/

Significant result in survival by *GYPC* methylation

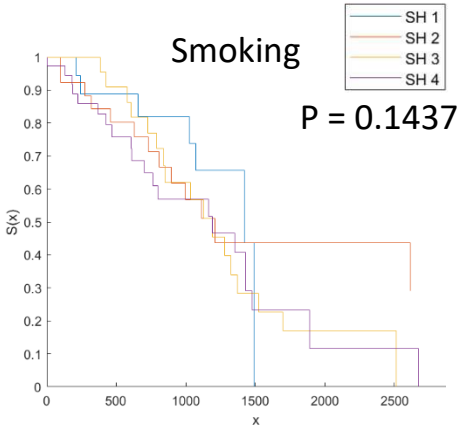
- The role of *GYPC* is not well researched
- However, *GYPC* has been found to be associated with JAK/STAT pathway (see left) and cell adhesion signaling pathways, whose activation leads to an increase in the metastatic abilities and invasiveness of tumors [1]
- High expression is associated with favorable prognosis in leukemia and LUAD and poor prognosis in ovarian cancer [1][3][4]
- This study found that high *GYPC* methylation is associated with poor prognosis



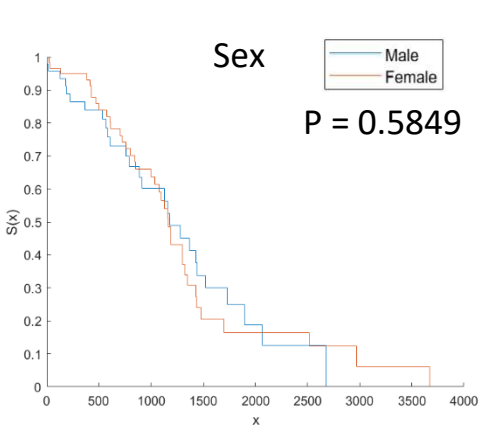
Kaplan-Meier curve comparing survival by cancer stage



Kaplan-Meier curve comparing survival by age group



Kaplan-Meier curve comparing survival by smoking history



Kaplan-Meier curve comparing survival by sex

CpG Sites Associated with LUAD Prognosis

CpG site	***p*-value**	***Associated with poor prognosis***	***Chromosome***	***Gene Symbol***
cg24103438	*4.369e-06*	*Hypermethylation*	*chrX*	*NLGN4X*
cg27217148	*3.588e-05*	*Hypermethylation*	*chr10*	*PCGF6*
cg16516400	*1.318e-04*	*Hypermethylation*	*chr1*	*FAM89A*
cg24363955	*1.577e-04*	*Hypermethylation*	*chr5*	*NPR3*
cg01592593	*2.780e-04*	*Hypermethylation*	*chrX*	*FGD1*
cg18710692	*2.823e-04*	*Hypermethylation*	*chr13*	*ZMYM5*
cg20616414	*2.92e-04*	*Hypermethylation*	*chr9*	*WNK2*
cg27652350	*3.25e-04*	*Hypermethylation*	*chr15*	*ALDH1A3*
cg22833292	*3.46e-04*	*Hypomethylation*	*chr11*	*RPS6KA4*
cg05798972	*4.48e-04*	*Hypomethylation*	*chr17*	*MED1*

The top ten CpG sites associated with LUAD survival based on *p*-value

Overall:

1426 CpG sites were identified that were associated with LUAD overall survival ($p < 0.05$).

In **622 sites**, hypomethylation was associated with poor prognosis.

In **804 sites**, hypermethylation was associated with poor prognosis

- Conclusion -

Limitations

- This study used adjacent, non-tumor tissue from LUAD patients for non-tumor data. However, it has been found that adjacent non-tumor tissue can show early signs of abnormal methylation [5]
    - In the future, it would likely be more accurate to compare tumor samples with lung tissue from an individual without cancer or to tissue from a distant lung site
- Also, this study had a small sample size – 66 tumor samples and 60 non-tumor samples.
    - Additional testing should be done with a larger dataset to see if the results seen in this study are consistently present
- Lastly, not all *GYPC, NME1,* and *SLIT2* CpG sites were investigated – only the sites in the 27K methylation array were analyzed, not the sites in the 450K array
    - The 450K array has 39 *NME1* sites, 18 *GYPC* sites, and 29 *SLIT2* sites, while the 27K array has 2 CpG sites for each of the three genes

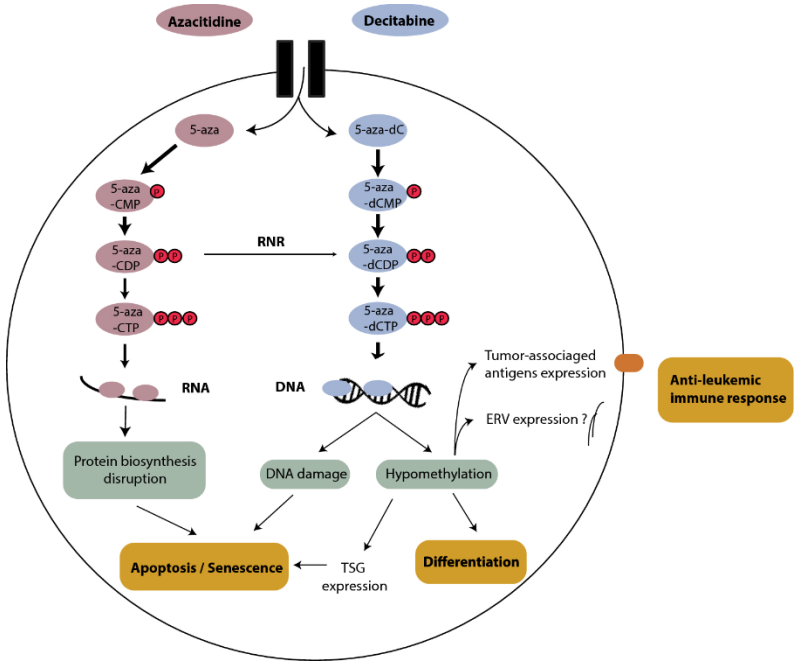
Accomplishments

- This study identified two possible LUAD prognostic biomarkers - cg13901526 and cg01063524
- Significant differences ($p < 0.05$) in methylation between tumor and non-tumor tissue may indicate that the methylation of *GYPC, NME1,* and *SLIT2* CpG sites plays a role in lung adenocarcinoma and can be a possible diagnostic biomarker for demographics in which the sites are differentially methylated

- Significant differences in DNA methylation levels between demographics may also indicate differences in the mechanism of LUAD development and progression

Future Applications

- Use the methylation of cg01063524 and cg13901526 in survival predictions for LUAD and possibly as targets for DNA methylation treatment
    - Azacitidine and decitabine, both hypomethylating agents, are approved by the FDA for use in treatment for myelodysplastic syndromes [6]
        - See the diagram to the right
    - Targeted DNA methylation treatment for controlling the methylation of specific CpG sites has been researched but not extensively
- Investigate the use of cg18972811, cg03742003, and cg17105014 in LUAD prediction - this study found that they were differentially methylated in Stage 1 LUAD



A diagram illustrating the mechanism of azacytidine and decitabine from “Clinical update on hypomethylating agents” [6] link.springer.com/article/10.1007/s12185-019-02651-9

- Research further into differences in the mechanism of LUAD development and progression between demographics identified in this study, which could be used in determining personalized treatment

- Bibliography -

[1] Wang, J., Chen, T., Yu, X., OUYang, N., Tan, L., Jia, B., Tong, J., & Li, J. (2020). Identification and validation of smoking-related genes in lung adenocarcinoma using an in vitro carcinogenesis model and bioinformatics analysis. Journal of translational medicine, 18(1), 313. https://doi.org/10.1186/s12967-020-02474-x

[2] Jin W. (2020). Role of JAK/STAT3 Signaling in the Regulation of Metastasis, the Transition of Cancer Stem Cells, and Chemoresistance of Cancer by Epithelial-Mesenchymal Transition. Cells, 9(1), 217. https://doi.org/10.3390/cells9010217

[3] Zhu, G. Z., Yang, Y. L., Zhang, Y. J., Liu, W., Li, M. P., Zeng, W. J., Zhao, X. L., & Chen, X. P. (2017). High Expression of AHSP, EPB42, *GYPC* and HEMGN Predicts Favorable Prognosis in FLT3-ITD-Negative Acute Myeloid Leukemia. Cellular Physiology and Biochemistry, 42(5), 1973–1984. https://doi.org/10.1159/000479837

[4] Guo, Y., Wang, Y. L., Su, W. H., Yang, P. T., Chen, J., & Luo, H. (2020). Three Genes Predict Prognosis in Microenvironment of Ovarian Cancer. Frontiers in genetics, 11, 990. https://doi.org/10.3389/fgene.2020.00990

[5] Brock, M. V., Gou, M., Akiyama, Y., Muller, A., Wu, T. T., Montgomery, E., Deasel, M., Germonpré, P., Rubinson, L., Heitmiller, R. F., Yang, S. C., Forastiere, A. A., Baylin, S. B., & Herman, J. G. (2003). Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. Clinical cancer research, 9(8), 2912–2919.

[6] Duchmann, M., & Itzykson, R. (2019). Clinical update on hypomethylating agents. International Journal of Hematology, 110, 161–169. https://doi.org/10.1007/s12185-019-02651-9

[7] McMahon, K. W., Karunasena, E., & Ahuja, N. (2017). The Roles of DNA Methylation in the Stages of Cancer. *Cancer journal (Sudbury, Mass.)*, *23*(5), 257–261. https://doi.org/10.1097/PPO.0000000000000279