

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

Uma Paul

Tesla STEM High School

Redmond, WA

Acknowledgements

This research was conducted under the mentorship of Dr. Vicente Notario, Professor in the Departments of Radiation Medicine and of Biochemistry and Molecular & Cellular Biology at Georgetown University.

Abstract

Lung cancer is the leading cause of cancer-related deaths worldwide, with lung adenocarcinoma being the most common subtype. This study used statistical analysis to determine the prognostic value of *GYPC*, *NME1*, and *SLIT2* DNA methylation for lung adenocarcinoma patients, which could aid in survival predictions and DNA methylation treatment. Using data from The Cancer Genome Atlas and the Gene Expression Omnibus, ANOVA was used to compare *GYPC*, *NME1*, and *SLIT2* DNA methylation (as beta values) between tumor ($n = 66$) and non-tumor tissue ($n = 60$) and against demographic features (age, sex, cancer stage, smoking history). Kaplan-Meier survival analyses and log-rank tests were utilized to determine CpG sites associated with patient survival and detect differences in survival across demographic groups and DNA methylation levels of the three target genes. There was hypermethylation of two *SLIT2* CpG sites ($p = 4.23\text{e-}04$ and $p = 5.01\text{e-}04$) and one *GYPC* CpG site ($p = 1.24\text{e-}04$) in tumor tissue. Methylation increased by cancer stage ($p = 0.0228$) and varied by smoking history ($p = 0.0478$) for two *SLIT2* CpG sites. Overall survival decreased by cancer stage ($p = 0.0133$), and increased methylation was associated with poor overall survival for a *NME1* CpG site ($p = 0.0226$) and a *GYPC* site ($p = 0.0379$). An additional 1426 survival-associated CpG sites were identified. This study identified two possible lung adenocarcinoma prognostic biomarkers from the three target genes and determined CpG sites that likely play a role in lung adenocarcinoma due to differential methylation in tumor tissue.

Keyword: lung adenocarcinoma, cancer prognosis, *GYPC*, *NME1*, *SLIT2*

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

Lung cancer causes around 160,000 deaths per year in the US alone, and lung adenocarcinoma (LUAD) accounts for approximately 40% of all lung cancers (Zappa & Mousa, 2016). Improving cancer prognosis can help doctors determine the appropriate course of treatment and provide patients with a more accurate prediction of their expected survival.

DNA methylation is a process by which a methyl group is added to a cytosine or adenosine in a DNA molecule and usually results in a lower level of transcription. In human somatic cells, the majority of methylation occurs on cytosine bases that are followed by guanine bases; these are referred to as CpG sites. The genes *GYPC*, *NME1*, and *SLIT2* are associated with smoking-related LUAD, and their expression may be affected by DNA methylation (Wang et al., 2020). *GYPC* is an integral membrane glycoprotein carried by human erythrocytes that regulates cell stability. Its role in cancer is not well-established but it has been associated with the JAK/STAT and cell adhesion signaling pathways, which play a role in tumor progression and metastasis (Wang et al., 2020) (Jin, 2020). *SLIT2* is a secreted glycoprotein that aids in cellular migration and is considered a tumor suppressor due to its involvement in the epithelial-mesenchymal transition, which increases metastatic and invasive activity and is linked to cancer stem cells (Wang et al., 2020) (Jin, 2020). *NME1*, or nucleoside diphosphate kinase 1, is a gene associated with lower mRNA levels in metastatic cells and has been linked to the cell cycle pathway (Wang et al., 2020).

This project sought to explore how the DNA methylation of *GYPC*, *SLIT2*, and *NME1* can be used in LUAD prognosis. The DNA methylation of the three genes was compared for differences in methylation between tumor and non-tumor tissue. Additionally, the DNA

methylation of *GYPC*, *SLIT2*, and *NME1* was compared based on clinical feature; differences in methylation by clinical feature would aid in improving our understanding of the differences in LUAD development and progression between demographic groups. The impact of dichotomized DNA methylation levels of *GYPC*, *SLIT2*, and *NME1* and clinical feature on the overall survival of LUAD patients was evaluated to determine the role of the genes in prognosis. The implications of this research would further the improvement of LUAD prognostic predictions, which could aid doctors in determining more effective, personalized treatments for LUAD patients.

Methods

The methylation of the three target genes was compared using MATLAB. One-way ANOVA tests were used to compare methylation levels (as beta values that range from 0.0-1.0) between tumor and non-tumor tissue and across the clinical features of age, sex, smoking history, and cancer stage. Kaplan-Meier survival analyses and log-rank tests determined survival-associated CpG sites and differences in survival by clinical feature (Fig. 1).

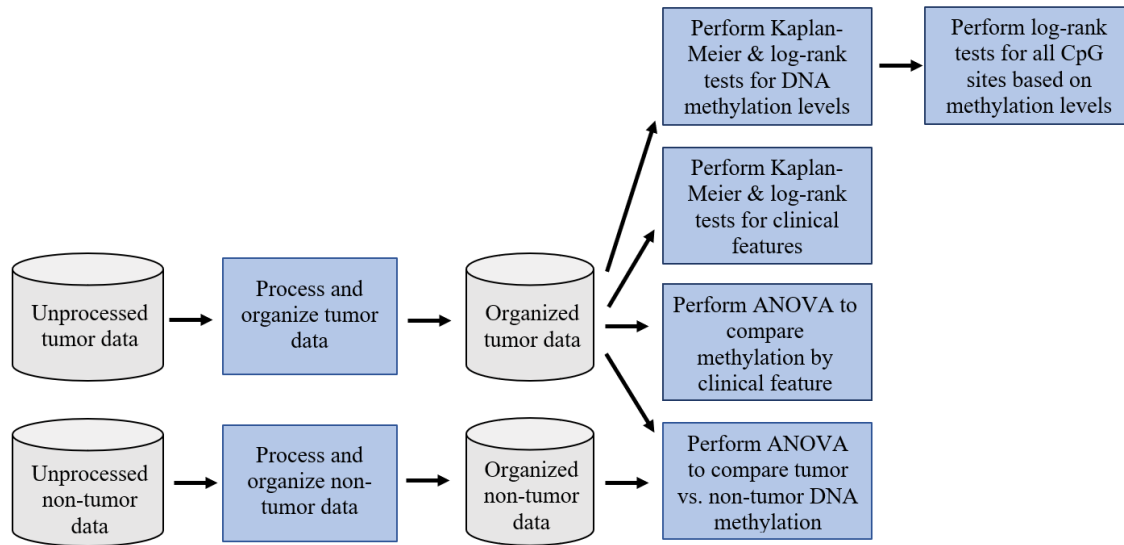


Fig. 1 A self-made flowchart outlining the data analysis

Data Sources

Tumor sample data was collected from The Cancer Genome Atlas database (URL: <https://portal.gdc.cancer.gov/exploration>), and 66 samples were collected in total. Non-tumor tissue was collected from the GSE32861 database (URL: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32861>), and 60 samples were collected in total. The data examined was restricted to 27K methylation arrays. For each of genes *GYPC*, *NME1*, and *SLIT2*, 2 CpG sites were present in the 27K methylation array. The CpG sites examined were cg13901526 and cg17105014 (*GYPC*), cg03742003 and cg18972811 (*SLIT2*), and cg04380669 and cg01063524 (*NME1*).

ANOVA

ANOVA tests were used to compare tumor and non-tumor DNA methylation and identify differences in DNA methylation based on the clinical features of age, sex, smoking history, and cancer stage. For tumor vs. non-tumor comparison, data was organized into separate tables based on the tissue type while examining the same demographic group. For instance, female tumor data was organized into one table and female non-tumor data into another table. The MATLAB ‘*anova1*’ function was used to compare the methylation of *GYPC*, *NME1*, and *SLIT2* between tumor and non-tumor tissue while controlling for clinical features. For DNA methylation comparison by clinical feature, data was organized into separate tables based on the demographic group. The ‘*anova1*’ function was used to determine the P values when comparing the methylation of *GYPC*, *NME1*, and *SLIT2* between demographics. P values < 0.05 were considered statistically significant.

Kaplan-Meier survival analyses and log-rank tests

Kaplan-Meier survival curves were used to illustrate differences in overall survival by methylation level and clinical feature. Log-rank tests were performed to determine the P values for differences in overall survival. For comparison by methylation level, samples were divided into a low methylation table and a high methylation table based on the median methylation value of each CpG site. The MATLAB ‘*ecdf*’ function was applied to graph overall survival for high and low DNA methylation levels. For comparison by demographic group, samples were divided into separate tables based on clinical feature. The ‘*ecdf*’ function was used to graph overall survival based on clinical feature. For all Kaplan-Meier curves, a self-coded, original log-rank test method was applied to determine P values to quantify differences in overall survival. P values < 0.05 were considered statistically significant.

Additional prognostic biomarker determination

Additional possible LUAD prognostic biomarkers were determined by comparing overall patient survival between low and high methylation groups for all CpG sites in the 27K methylation array using the log-rank test method. CpG sites with P values < 0.05 were identified as survival-associated sites.

Results

ANOVA: Tumor vs. Non-tumor

Tumor tissue displayed hypermethylation (a significantly higher level of methylation in comparison to non-tumor tissue) for cg03742003, a *SLIT2* CpG site ($p = 4.23e-04$), cg17105014, a *GYPC* CpG site ($p = 1.24e-04$), and cg18972811, a *SLIT2* CpG site ($p = 5.01e-04$), which can lead to gene silencing. Also, females displayed differential methylation for 4 sites while there were no differentially methylated sites for males. See Table 1 in the appendix for full results.

ANOVA: Clinical Features

The methylation level of cg18972811 varied based on smoking history ($p = 0.0478$), with there being a significant difference in methylation between lifelong nonsmokers and current reformed smokers for more than 15 years. The methylation of cg03742003 increased by cancer stage ($p = 0.0228$). Both of these sites are *SLIT2* CpG sites. See Table 2 in the appendix for full results.

Kaplan-Meier: DNA Methylation Level

Significant differences in survival by methylation level were identified for cg13901526, a *GYPC* CpG site (Fig. 2A) and cg01063524, an *NME1* CpG site (Fig. 2B). High levels of methylation for cg13901526 and cg01063524 were associated with poor overall survival.

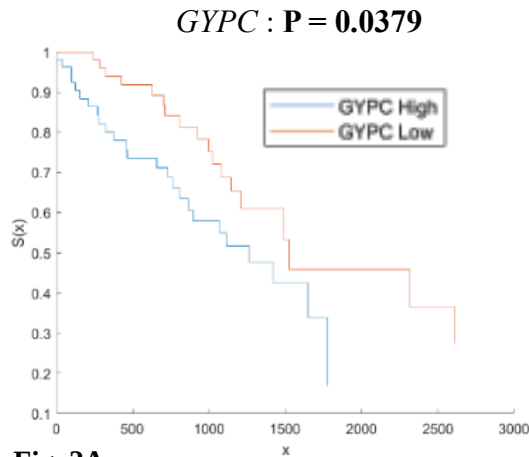
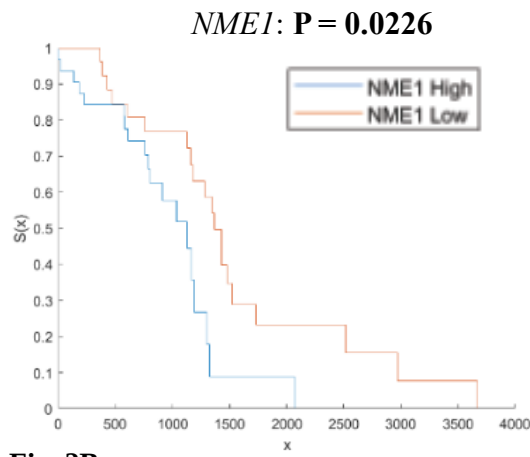
**Fig. 2A****Fig. 2B**

Fig. 2 Kaplan-Meier curves of overall survival for **A)** cg13901526 and **B)** cg01063524

Kaplan-Meier: Clinical Features

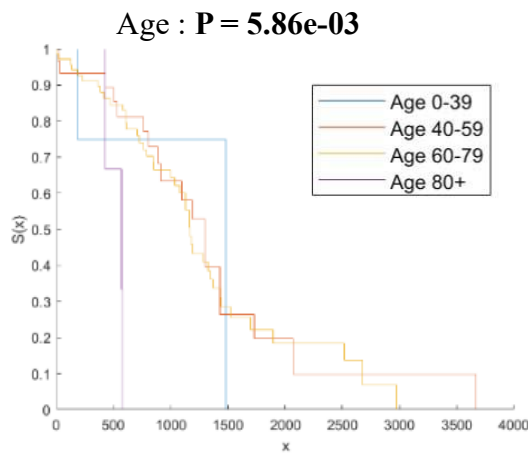
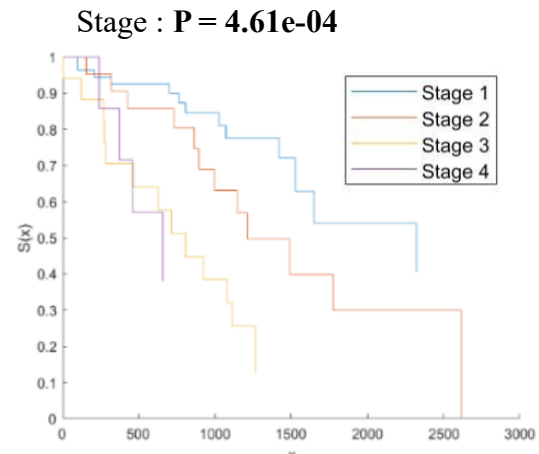
**Fig. 3A****Fig. 3B**

Fig. 3 Kaplan-Meier curves of overall survival for **A)** age group and **B)** cancer stage

Significant differences in overall survival were identified based on cancer stage, with overall survival decreasing as cancer stage increased. However, the significance of the P value for age requires further validation due to a lack of a sufficient number of samples for the age groups 0-39 and 80+.

CpG Sites Associated with Overall Survival

CpG site	<i>P</i> value	Gene Symbol	<i>Associated with poor prognosis</i>
<i>cg24103438</i>	<i>4.369e-06</i>	<i>NLGN4X</i>	<i>Hypermethylation</i>
<i>cg27217148</i>	<i>3.588e-05</i>	<i>PCGF6</i>	<i>Hypermethylation</i>
<i>cg16516400</i>	<i>1.318e-04</i>	<i>FAM89A</i>	<i>Hypermethylation</i>
<i>cg24363955</i>	<i>1.577e-04</i>	<i>NPR3</i>	<i>Hypermethylation</i>
<i>cg01592593</i>	<i>2.780e-04</i>	<i>FGD1</i>	<i>Hypermethylation</i>
<i>cg18710692</i>	<i>2.823e-04</i>	<i>ZMYM5</i>	<i>Hypermethylation</i>
<i>cg20616414</i>	<i>2.92e-04</i>	<i>WNK2</i>	<i>Hypermethylation</i>
<i>cg27652350</i>	<i>3.25e-04</i>	<i>ALDH1A3</i>	<i>Hypermethylation</i>
<i>cg22833292</i>	<i>3.46e-04</i>	<i>RPS6KA4</i>	<i>Hypomethylation</i>
<i>cg05798972</i>	<i>4.48e-04</i>	<i>MED1</i>	<i>Hypomethylation</i>

Fig. 4 Table with the top ten genes associated with LUAD overall survival

After the examination of all CpG sites in the 27K methylation array, 1426 CpG sites were identified that were associated with LUAD overall survival ($P < 0.05$). 622 sites were identified where hypomethylation was associated with poor LUAD overall survival, and in 804 sites, hypermethylation was associated with poor prognosis.

Discussion

This study successfully identified two possible LUAD prognostic biomarkers from the three target genes- *cg13901526* and *cg01063524*. These findings furthered our understanding of the role of *GYPC* and *NME1* in cancer prognosis. High *GYPC* expression is associated with favorable prognosis in leukemia and LUAD and poor prognosis in ovarian cancer (Wang et al., 2020) (Zhu et al., 2017) (Guo et al., 2020). High *NME1* expression is associated with favorable prognosis in breast cancer and melanoma and poor prognosis in neuroblastoma and cervical cancer (Wang et al., 2020). Determining that high *GYPC* and *NME1* methylation, and therefore likely low expression, is associated with poor prognosis aids in understanding the oncological role of the two genes. Additionally, significant differences in DNA methylation parameters between tumor and non-tumor tissue suggest that the DNA methylation of the identified sites may be associated with LUAD and could serve as diagnostic biomarkers. Significant differences

in DNA methylation levels between clinical features may indicate differences in the mechanism of LUAD development and progression by demographic. Also, research has found that *SLIT2*, *GYPC*, and *NME1* are differentially expressed based on smoking history (Wang et al., 2020); therefore, it is hypothesized that there would be differential methylation in their associated CpG sites when comparing samples by smoking history. However, *GYPC*, *NME1*, and *SLIT2* CpG sites other than cg18972811 lacked differential methylation; epigenetic factors such as histone acetylation may have a greater influence on expression levels in comparison to DNA methylation for these sites. As for differential *SLIT2* methylation by cancer stage, *SLIT2* has been found to be a tumor suppressor gene with the effects of inhibiting tumor migration and growth via the epithelial-mesenchymal transition (Wang et al., 2020). This conclusion is supported by the results of this study, which were that *SLIT2* methylation increased as cancer stage increased.

This study was limited by the data sets available for use. Adjacent, non-tumor tissue from LUAD patients was used for non-tumor data. However, adjacent non-tumor tissue can show early signs of abnormal methylation, and therefore, using lung tissue from healthy individuals or tissue from distant lung sites may improve the accuracy of results (Brock et al., 2003). Another limitation is that only the *GYPC*, *NME1*, and *SLIT2* CpG sites found in the 27K methylation were examined, not those in the 450K array, which contains around 17 times more CpG sites. The 450K array has 39 *NME1* sites, 18 *GYPC* sites, and 29 *SLIT2* sites, while the 27K array has 2 CpG sites for each gene. The final limitation is that due to the small sample size of this study and the fact that the majority of samples came from Caucasian or African-American patients, additional analysis with a larger, more diverse dataset would help determine whether the results of this study are consistently seen.

Future applications of this research are the use of cg01063524 and cg13901526 in survival predictions for LUAD and as possible targets for targeted DNA methylation treatment. Further development of targeted methylation treatment like homology directed repair would allow for the possibility of treatment involving the survival-associated CpG sites identified in this study (Cali et al., 2019). Another application of this study would be the investigation of cg18972811, cg03742003, and cg17105014 for use in LUAD predictions – these sites were differentially methylated in Stage 1 LUAD. Additionally, research into the differences in the mechanism of LUAD development and progression between the demographics identified in this study could be used in determining personalized treatment due to the fact that DNA methylation differences have been found to indicate resistance to radiation and chemotherapy (McMahon et al., 2017). In conclusion, this study determined that *GYPC*, *SLIT2*, and *NME1* are possible LUAD diagnostic and prognostic biomarkers, and they may be useful for improving LUAD survival predictions and personalized treatment.

References

- Brock, M. V., Gou, M., Akiyama, Y., Muller, A., Wu, T. T., Montgomery, E., Deasel, M., Germonpré, P., Robinson, 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.
- Christopher P Cali, Daniel S Park, Edward B Lee, Targeted DNA methylation of neurodegenerative disease genes via homology directed repair, *Nucleic Acids Research*, Volume 47, Issue 22, 16 December 2019, Pages 11609–11622, <https://doi.org/10.1093/nar/gkz979>
- 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>
- 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>
- 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>
- 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>

Zappa, C., & Mousa, S. A. (2016). Non-small cell lung cancer: current treatment and future advances. *Translational lung cancer research*, 5(3), 288–300.

<https://doi.org/10.21037/tlcr.2016.06.07>

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>

Appendix

		<i>P value</i>					
		<i>GYPC</i> <i>cg13901526</i>	<i>SLIT2</i> <i>cg03742003</i>	<i>NME1</i> <i>cg04380669</i>	<i>GYPC</i> <i>cg17105014</i>	<i>SLIT2</i> <i>cg18972811</i>	<i>NME1</i> <i>cg01063524</i>
Sex	Male	0.164	0.129	0.996	0.077	0.071	0.101
	Female	0.359	0.002	0.031	5.4529e-04	0.006	0.656
Smoking	Nonsmoker	0.405	0.019	0.043	0.093	0.029	0.813
	Current Smoker	0.105	0.192	0.504	0.338	0.027	0.424
Age	40-59	0.67	0.113	0.013	0.114	0.07	0.583
	60-79	0.083	0.008	0.425	0.003	0.01	0.128
	80+	0.248	0.947	0.7	0.294	0.2	0.844
Stage	I	0.399	0.002	0.179	5.8910e-04	0.002	0.286
	II	0.962	0.546	0.641	0.304	0.287	0.459
	III	0.385	0.143	0.002	0.566	0.198	0.846
	IV	0.341	0.473	0.006	0.298	0.487	0.09
Overall		0.068	4.2279e-04	0.078	1.2376e-04	5.0077e-04	0.236

Table 1: The P values when comparing tumor DNA methylation of *GYPC*, *NME1*, and *SLIT2* to non-tumor tissue

	P value					
	GYPC cg13901526	SLIT2 cg03742003	NME1 cg04380669	GYPC cg17105014	SLIT2 cg18972811	NME1 cg01063524
Sex	0.403	0.912	0.864	0.629	0.52	0.537
Smoking History	0.616	0.928	0.34	0.641	0.048	0.852
Age	0.747	0.718	0.181	0.572	0.893	0.29
Cancer Stage	0.081	0.023	0.852	0.133	0.538	0.753

Table 2: The P values when comparing the DNA methylation of *GYPC*, *NME1*, and *SLIT2* based on clinical feature