BIOS 24150 Final Report: An Analysis of

"Downregulation of VPS13C promotes cisplatin resistance in cervical cancer by upregulating GSTP1"

By: Tan X, Wang X, Liao X, et. all

Rachel Shmidt, Ryan Gallagher, and Aninda Roy

Introduction

This document is written in fulfillment of the final report requirement for the course BIOS 24150 – Bioinformatics & Omics Analysis at MCW. This report analyses the bioinformatics related paper "Downregulation of VPS13C promotes cisplatin resistance in cervical cancer by upregulating GSTP1" by Tan X, Wang X, Liao X, et. All. This report will provide a background on the concepts within this paper and outline the challenges the authors seek to overcome, with emphasis on the data and biological goal. This report will then describe the methods employed by the authors to accomplish their outlined aims. There will then be a discussion of the authors' results and accompanying figures. Finally, we will provide an extension of their analysis by employing methods not featured in the original analysis and discuss their contributions.

Background

Up to 99.7% of cervical cancer is due to persistent human papillomavirus (HPV) infection by 12 high risk HPV types (Okunade et al 2019). Methods to prevent cervical cancer have recently been developed in the past decade by a 9-valent vaccine called GARDASIL 9 that protects against 9 of the top high risk HPV types. Despite a readily available vaccine, there is still a need to study cervical cancer, especially in people living low-income countries and patients who do not respond well to standard treatment.

The standard of care for cervical cancer patients includes biopsy or colposcope followed by treatment with chemotherapy agents such as cisplatin and then radiation. Treatment of patients with chemoresistance, especially in cervical cancer has been difficult to navigate. Therefore, methods to study the cause of chemoresistance in cervical cancer is important. In this paper, the authors wish to study the gene, Vacuolar protein sorting 13 homolog C (VPS13C), previously identified by the author in an earlier publication to be associated with chemoresistance in cervical cancer (Tian et. Al, 2021). This paper identified one differentially expressed gene among 3 groups (LinkedOmics, SiHa, and ME180 gene expression) called Glutathione S-Transferase Pi 1 (GSTP1). Further experiments performed in the paper combined cisplatin with a

GSTP1 inhibitor NDBHEX that showed cervical cancer cells were rescued from cisplatin resistance. The data provided was uploaded to GEO Gene Expression Omnibus (GEO) via excel sheets (GSE235466, SiHa), (GSE235366, ME180). The data included transcriptome and gene expression data from SiHa cells with vector only or with vector and shVPS13C and 20uM cisplatin or with ME180 cells with vector only or with vector and sgVPS13C and 2uM cisplatin.

Aims of the Paper

The paper aims to investigate the mechanisms behind cisplatin resistance in cervical cancer, which is a significant challenge limiting the effectiveness of chemotherapy. To do this, they demonstrate that deficiency in vacuolar protein sorting 13 homolog C (VPS13C) promotes cisplatin resistance in cervical cancer. The author also wished to identify any genes associated with VPS13C that could be causing this chemoresistance. They identified and further investigated one negatively correlated gene glutathione S-transferase pi gene (GSTP1) with VPS13C. Furthermore, they wished to evaluate the therapeutic potential of targeting GSTP1 with the inhibitor NBDHEX to counteract the cisplatin resistance induced by VPS13C deficiency.

Aims of our Analysis

We sought out to repeat some of the analysis done in the paper and expand upon it. To do this we generated a PCA plot comparing the gene expression of the two cell lines (SiHa and ME180). We hoped to repeat the differential gene expression analysis done in the paper and we repeated the survival KM-plots from the supplemental figures.

Methods

Data Processing

Data was sourced from Gene Expression Omnibus (GEO) and retrieved in .csv files. Analyses and preprocessing were conducted in `R` (version 4.3.1) using the `Tidyverse` and `Survival` packages from the CRAN, and the `impute` package from Bioconductor. The analysis file will be attached to this report for reference.

The GEO data received was segmented for SiHa and ME180, respectively. When combined, there were a total of 2438 missing values out of ~35,000 genes. K-Nearest Neighbors Imputation was used to fill missing data. KNN imputation is the process of identifying the 'k' nearest neighbors to a data point with missing values, where 'k' is a user-specified parameter. The nearest neighbors are typically determined based on a distance metric, such as the Euclidean distance, although other metrics like Manhattan or Minkowski distance can also be used depending on the dataset's characteristics. Once the nearest neighbors are identified, the missing values are imputed. Our implementation used `k=10` and the Euclidean distance metric.

Statistical Methods

Principal Component Analysis (PCA)

PCA is employed in genomic research due to its effectiveness in simplifying the complexity of large genetic datasets by transforming the original variables into a new set of orthogonal variables, known as principal components, that capture the most significant variance in the data. In gene expression analysis, PCA helps to reveal the primary modes of variation among different conditions or treatments which allow for more obvious grouping of data. PCA also serves as a powerful visualization tool by reducing data to two or three principal components. This allows for the graphical representation of complex data sets in a two- or three-dimensional space, making it easier to detect trends and outliers. PCA can be instrumental in uncovering new associations between genetic components and phenotypic traits, which might not be immediately apparent due to the high dimensionality of the data.

Survival Analysis

Survival analysis is utilized through this research in the form of Kapalin Meier (KM) curves to show survival probability over time. The KM estimator creates a step function that changes value only at the time of each event. The survival probability is plotted on the y-axis against time on the x-axis. Each time a particular event, like death, occurs, the survival probability takes a step downward. The size of the drop at each event reflects the impact of each event on the overall survival probability, considering the number of individuals still at risk of the event at that time. Kaplan-Meier curves can also be used to compare the survival functions of different groups within a study. This was particularly useful in this analysis to show contract between high and low expression within the survival plots.

Pearson Correlation

Pearson correlation is a measure of the linear correlation between two variables that quantifies the degree to which a relationship between two variables can be described by a line. The Pearson correlation coefficient provides information about both the strength and the direction of the linear relationship. It ranges between -1 and 1. A value of 1 implies a perfect positive linear relationship, where increases in one variable exactly correspond with increases in the other. A value of -1 indicates a perfect negative linear relationship, where increases in one variable exactly correspond with decreases in the other. A correlation of 0 suggests no linear relationship between the variables.

Gene Set Enrichment Analysis (GSEA) / Pathway Analysis

Enrichment analysis in the scope of genomics data is a statistical approach used to determine whether a set of genes or genetic variants is over-represented with certain annotations or functional categories more than would be expected by chance. Particularly, pathway enrichment analysis is used to identify biological pathways that are overrepresented in a list of genes or proteins of interest. These pathways could be related to metabolism, signaling, disease mechanisms, or other cellular processes.

Biological Methods

Cell Lines

In this study the commercially available cervical cancer cell line SiHa was used and ME180 which is an epidermoid carcinoma were used. [Cell Lines]

Plasmid construction

The researchers used short hairpin RNA (shRNA) to reduce the levels of VPS13C in SiHa cells and employed the CRISPR/Cas9 system to completely knock out VPS13C in ME180 cells. They designed and synthesized guide RNAs (sgRNAs) to target the VPS13C gene, which were incorporated into lentiviral plasmids.

Cell Viability Assays

Cells were plated and incubated for 24 hours before adding various concentrations of cisplatin. Then cell viability was measured by CCK-8 assay 2 hours after adding the reagent using an optical density of 450nm. The IC50 of each cell line with cisplatin was then determined to be 20uM for SiHa and 2uM for ME180.

Apoptosis Assays and Flow Cytometry

SiHa and ME180 cells were treated with cisplatin for 48hrs with and without the GSTP1 inhibitor NBDHEX. Apoptosis was measured by counting the cells and analyzing them using flow cytometry to sort cells based on expression of apoptosis markers in the cells.

Results

Authors' Results

In this study, they showed the effect of VPS13C on cisplatin resistance on two cervical cancer cell lines SiHa and ME180, which they used for subsequent experiments. They confirmed knockout of VPS13C with qPCR, Western blot (protein assay) and Sanger (for ME180 only). Both lines exhibited lower expression of VPS13C. (A, B, C, D, and supplemental figure for qPCR that confirms knockout in SiHa and me180).

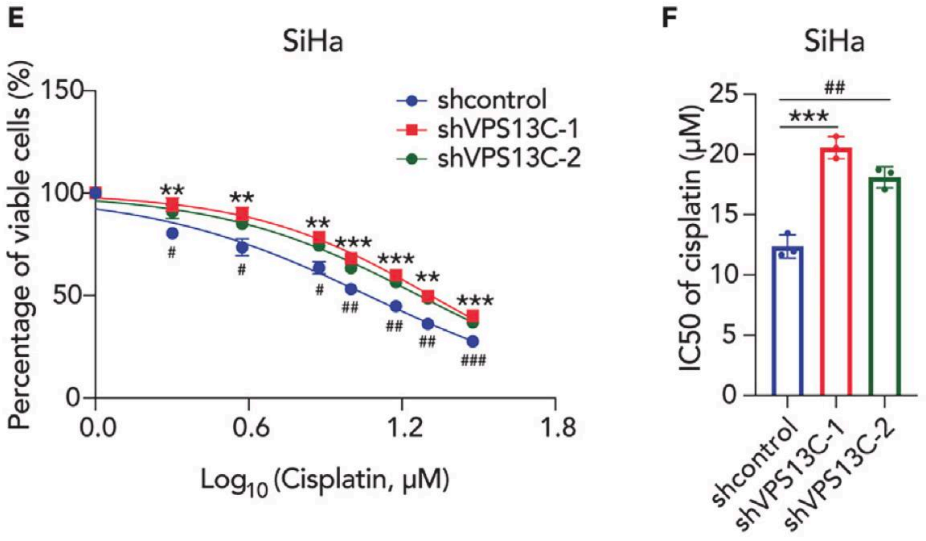


Figure 1: The viability of SiHa cells treated with different concentrations of cisplatin

From Figure 1E we can see that cell viability increases with the increasing concentration of cisplatin and a higher half-maximal inhibitory concentration (IC50) of cisplatin (Figure 1F) on SiHa cells with VPS13C knockout and Cisplatin treatment suggesting Cisplatin resistance.

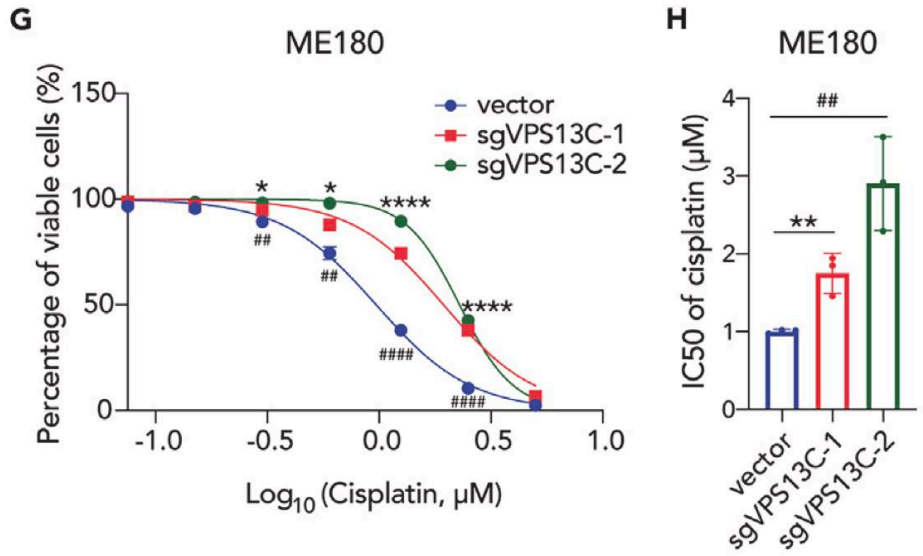


Figure 2: The viability ME180 cells treated with different concentrations of cisplatin.

Here from Figure 2G we can see that cell viability increases with the increasing concentration of cisplatin and a higher half-maximal inhibitory concentration (IC50) of cisplatin (Figure 2H) on ME180 cells. It determines that IC50 (concentration Cisplatin of 50% cells death) for both Cell lines 20um for SiHa, and Me180.

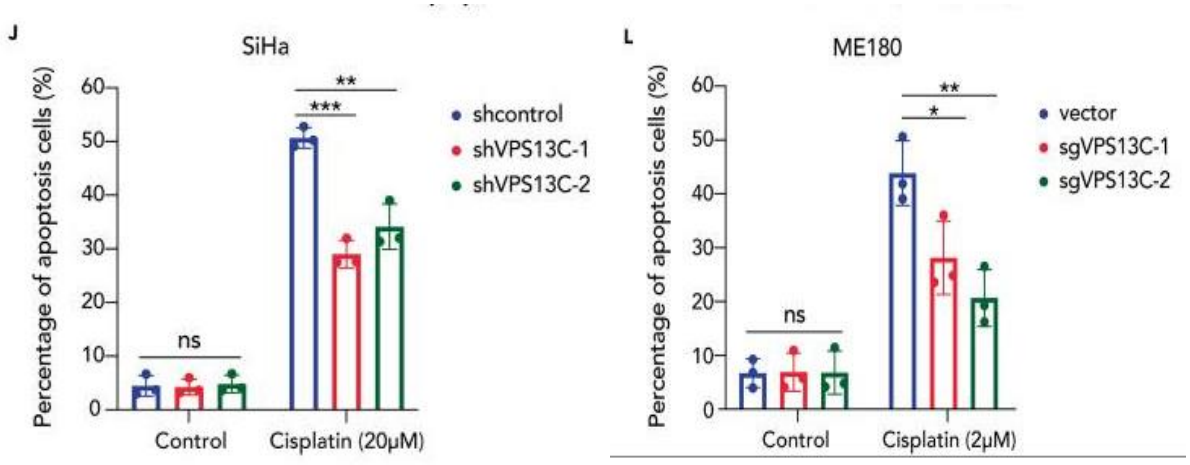
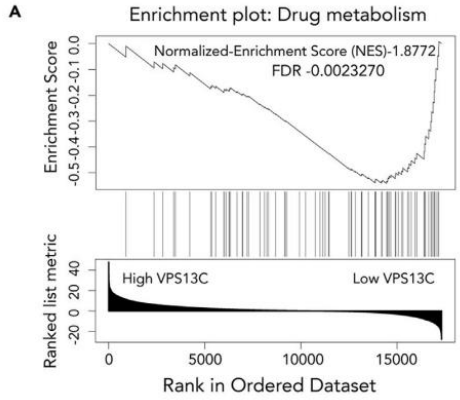


Figure 3: Quantification of the apoptosis rate in SiHa and ME180 cells treated with cisplatin for 48 h.

Here we can see that loss of VPS13C did not alter the apoptosis rate but significantly decreased cisplatin-induced apoptosis in both SiHa and ME180 cells. On subsequent figures, they showed in depth explanation of DE gene expression methods and result. They emphasize Significant p-value and negative Pearson correlation of VPS13C and GSTP1. To delve deeper into the molecular mechanisms underlying the role of

VPS13C in promoting cisplatin resistance in cervical cancer, they conducted gene set enrichment analysis (GSEA). This involved sorting genes based on their Pearson correlation with VPS13C expression in the TCGA dataset.



B

Gene Symbol	Entrez Gene	Score
NME2	4831	-15.257
GSTP1	2950	-14.3941
UPP1	7378	-12.8491
ITPA	3704	-12.1593
UCK2	7371	-11.8619
GSTO1	9446	-11.178
TYMP	1890	-10.5027
NME1	4830	-10.0998
NME1-NME2	654364	-9.3628
UMPS	7372	-8.4891
TK1	7083	-8.1751
UCK1	83549	-7.9144
IMPDH2	3615	-6.0217
UGT1A7	54577	-5.6931
GMPS	8833	-5.0404

Figure 4: Enrichment Analysis (A) and Top genes (B)

Notably, there was a significant enrichment of drug metabolism gene sets in the cohort exhibiting low VPS13C expression (depicted in Figure 4A). The top genes display a negative correlation with VPS13C expression and their association with drug metabolism pathways are enumerated in Figure 4B.

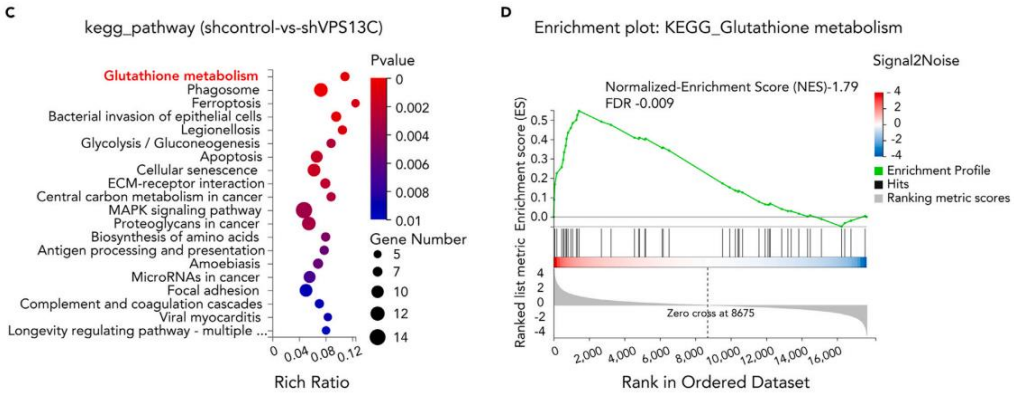


Figure 5: KEGG pathway enrichment analysis on SiHa (C) and GSEA plot of metabolism enrichment in ME180 cells (D) treated with cisplatin.

The outcomes revealed that genes exhibiting differential expression in VPS13C-deficient SiHa cells compared to control cells were notably enriched in glutathione metabolic pathways (depicted in Figure 5C). These pathways are known to potentially induce drug resistance by bolstering the detoxification of anticancer drugs. This finding was corroborated by GSEA, confirming the enrichment of glutathione metabolic pathways in ME180 cells (illustrated in Figure 5D).

To further refine the identification of potential targets contributing to VPS13C-induced cisplatin resistance, they employed Venn diagrams to analyze the overlap of differentially expressed genes within

glutathione metabolic pathways from both the TCGA dataset and RNA-seq analysis. Within this analysis,

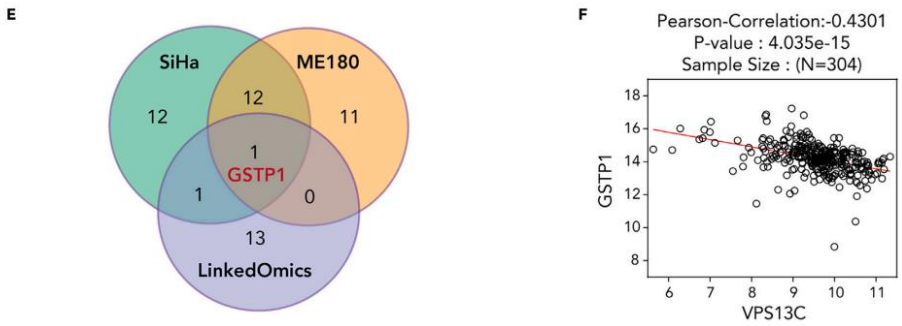


Figure 6: (E)RNA-seq Analysis of SiHa and ME180, (F)Correlation among GSTP1 and VPS13C

the glutathione S-transferase pi gene (GSTP1), recognized for its pivotal role in cisplatin resistance, emerged as a notable downstream molecule exhibiting a significant negative correlation with VPS13C (depicted in Figure 6E). This observation underscored the identified negative correlation between VPS13C and GSTP1 mRNA levels in cervical cancer cell lines (Figure 6F). The correlation is –0.4301 which is not so strong but still it shows the negative correlations between GSTP1 and VPS13C which leads to the conclusion that VPS13C deficiency indeed has influence in GSTP1.

Based on their results, such as GSTP1 is negatively correlated to VPS13C, assays in vitro (cell culture) were done to confirm the results. Assays included qPCR (mRNA expression) Western Blot (protein expression), cell staining. And IHC (immunohistochemistry aka tissue staining).

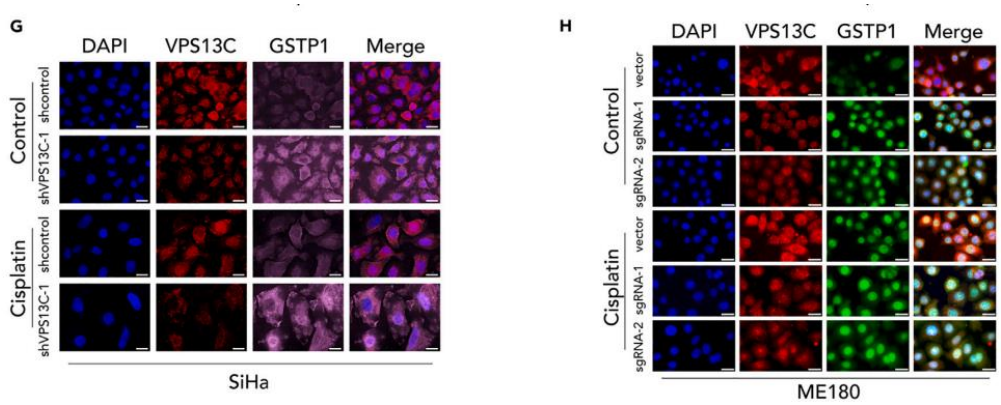


Figure 7: (G-H) GSTP1 and VPS13C protein Expression

In agreement with the findings from RNA sequencing, the expression of both mRNA and protein levels of GSTP1 showed an increase in SiHa and ME180 cells following the knockdown of VPS13C, as illustrated in figures 3(A-D) on the paper. Here, we can see that comparable trends were observed in immunofluorescence staining results, as depicted in Figures 7G and 7H.

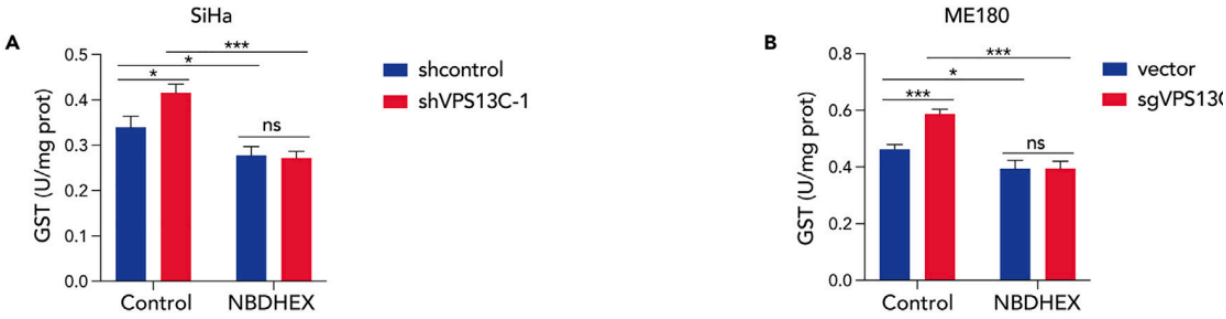


Figure 8: (A-B) NBDHEX reduce the enzyme activity in cells with VPS13C Knockdown

The previous results provide evidence supporting the notion that VPS13C impacts cisplatin efficacy in cervical cancer through its influence on GSTP1 expression. Based on this, now authors postulated that targeting GSTP1 could serve as a potential strategy to overcome cisplatin resistance arising from VPS13C deficiency.

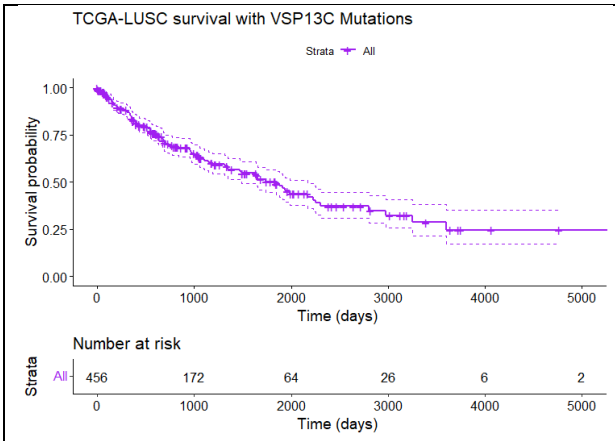
NBDHEX, a specific GSTP1 inhibitor, hinders GSTP1 by diminishing its enzyme activity and disrupting its interaction with crucial signaling factors. This hypothesis was initially tested by assessing the impact of NBDHEX on GST enzyme activity using a GST enzyme kit. The results indicated that NBDHEX effectively reduced the enzyme activity, particularly in cervical cancer cells with VPS13C knockdown, as demonstrated in Figures 8A and 8B. Subsequently, they employed a CCK-8 assay to evaluate the effect of NBDHEX on the cell viability of cervical cancer cells where they showed NBDHEX reversed the heightened cell viability induced by VPS13C deficiency during cisplatin treatment. These findings imply that suppressing GSTP1 could successfully mitigate the cisplatin resistance in cervical cancer cells triggered by VPS13C deficiency.

In the later figures (Figure 6) in the paper authors presents their findings on repeated drug experiment in vivo (in mice) and these findings strongly suggest that the GSTP1 inhibitor NBDHEX has the capacity to successfully counteract cisplatin resistance induced by VPS13C deficiency in cervical cancer cells.

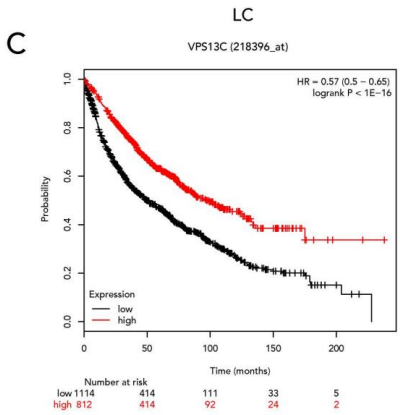
Our Results

KM Plots

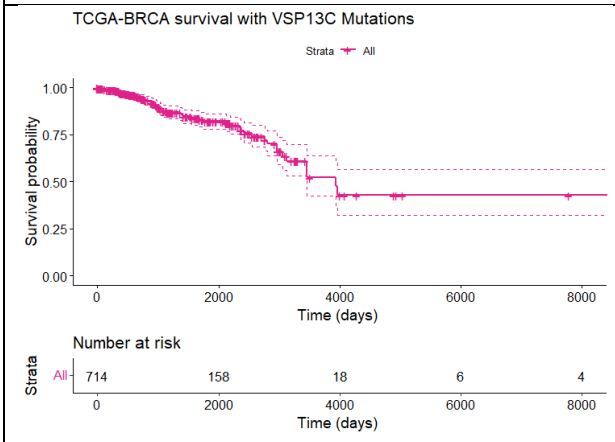
In the original analyses, Kaplan–Meier plotter (http://kmplot.com/) is used to assess survival between low and high expression of VPS13C (using the Affymetrix ID 218396_at) in breast, lung cancer (squamous cell carcinoma and adenocarcinoma) and gastric cancer. This website uses databases from GEO, EGA, and TCGA . Since the original publication in August 2023, data from these databases has been updated. We repeated this analysis on December 9th, 2023. Additionally, Kaplan Meier survival plots were also recreated in R (version 4.3.1) from the GDC data portal using TCGA data from breast cancer (BRCA), lung adenocarcinoma (LUAC), lung squamous cell carcinoma (LUSC), gastric cancer (STAD), and cervical cancer (CESC). However, in the TCGA data we only looked at observations with a mutation in VPS13C, so there was not a comparison of low and high expression of the gene. Comparison of patients with normal expression to patients with the mutation in VPS13C would remedy this.



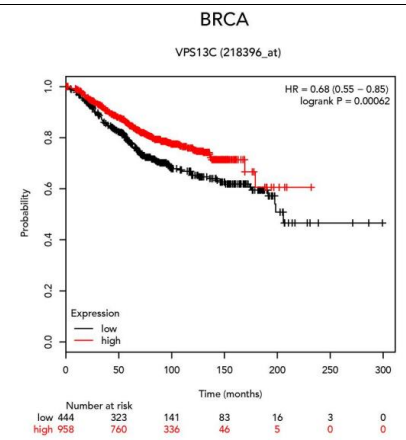
Our KM Plot – Lung Cancer



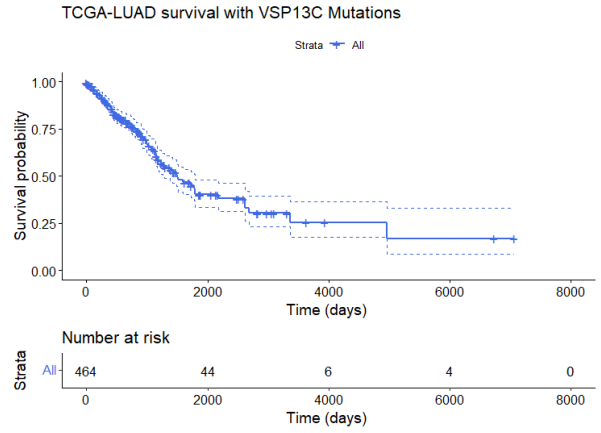
Original KM Plot – Lung Cancer



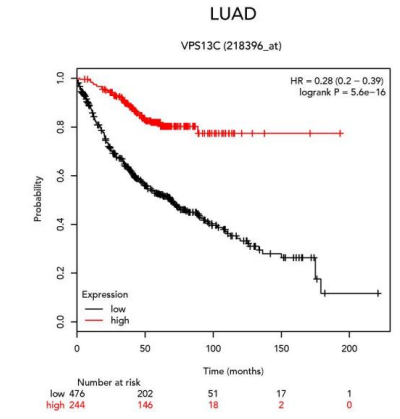
Our KM Plot – Breast Cancer



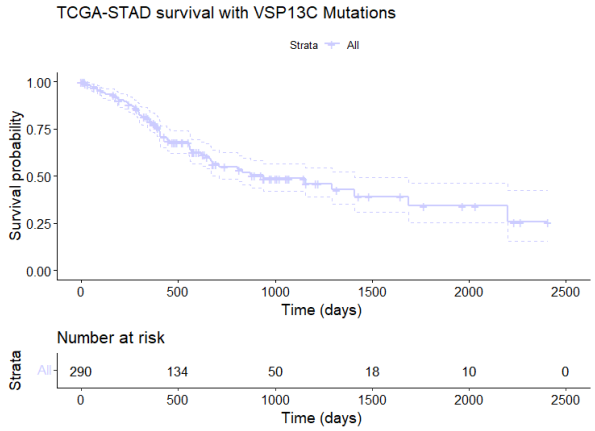
Original KM Plot – Breast Cancer



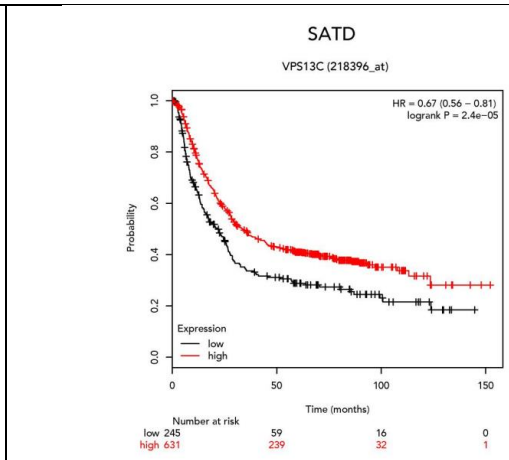
Our KM Plot - Lung Adenocarcinoma



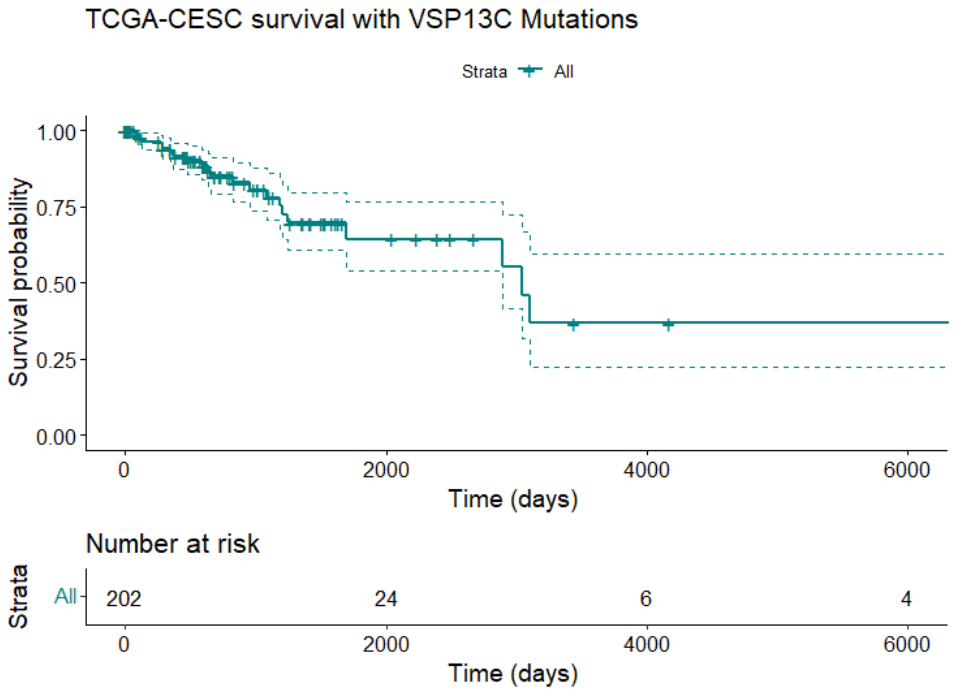
Original KM Plots - Lung Adenocarcinoma



Our KM Plot - Stomach Adenocarcinoma



Original KM Plots - Stomach Adenocarcinoma



Our KM Plot – Cervical Squamous Cell Carcinoma

Figure 9: Our KM plots via R with updated data vs. Original plots from kmplot.com

PCA Results

Gene expression data was downloaded from Gene Expression Omnibus (GEO) via excel sheets. Both cell lines in experimental expression data [GSE235466 (SiHa), GSE235366(ME180)] were combined and genes with mostly missing values or 0s were imputed by KNN (K=10) then log2 transformed. As expected, SiHa and ME80 cell line cells can be distinguished as separate groups (PC1=47.57, PC2=26.48) however cell lines by treatment and individual samples cannot, which is expected for gene expression data. Although both cells can have a decent between group variance (47.57%) as opposed to within group variance (26.48%) there is still some improvement that could be made to minimize the variance within the groups (cell lines).

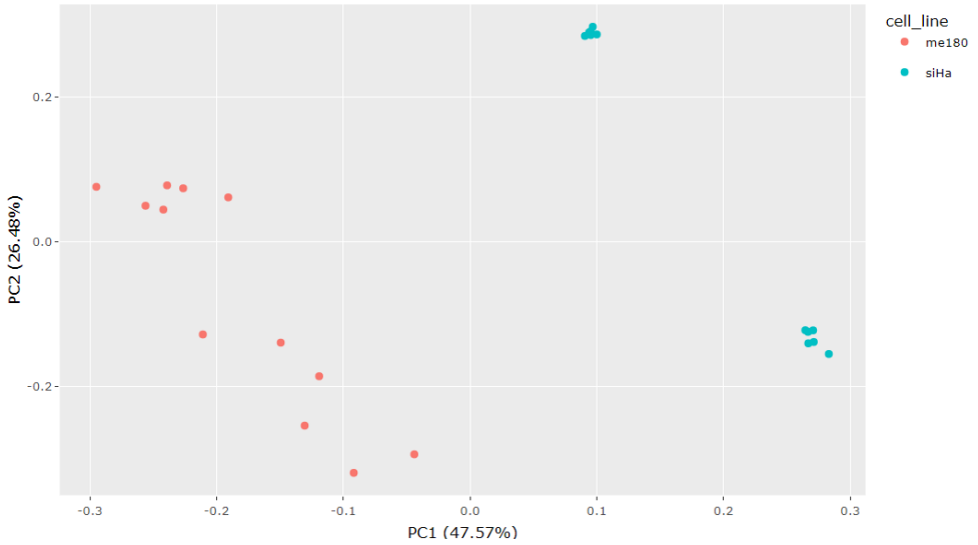


Figure 10: PCA by Cell Line

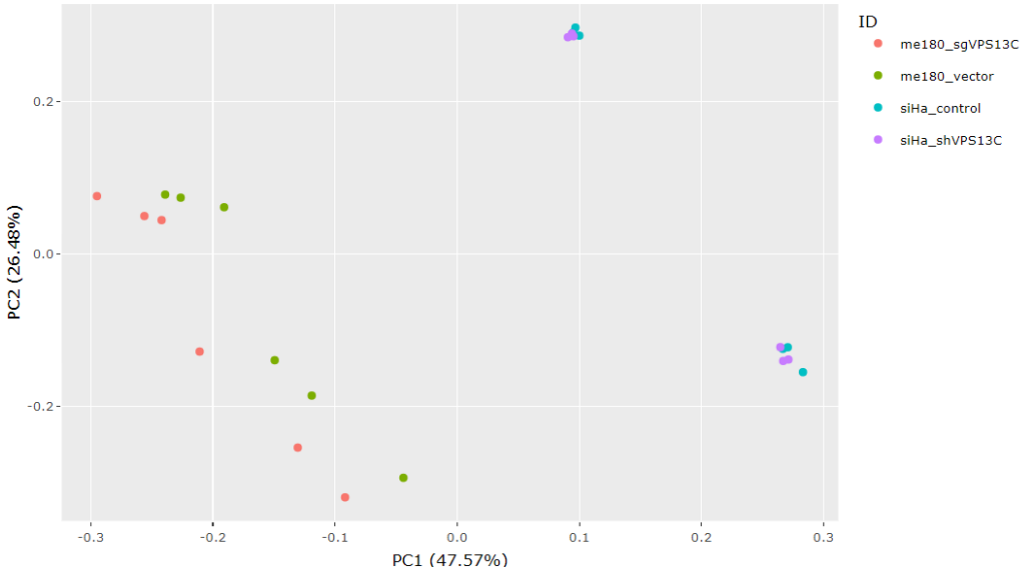


Figure 11: PCA by Cell Line Treatment

Conclusion

In this paper, the author explores the important role of VPS13C in cisplatin resistance of cervical cancer. Though there is still need of conducting a molecular mechanistic study in future to reconfirm the influences of VPS13C on GSTP1 expression. We analyzed TCGA data to imitate KM plot to assess survival between low and high expression of VPS13C but due to the update of the databases we could not find the same hazard ratio, but it was close. In addition, we showed the PCA result of distinguishing SiHa and ME180 cell line with higher group variance than within group variance.

Due to time limitations, we did not do further analysis on exploring top 3-5 DE genes and correlation to VPS13C which might be our future addition with this result. In the future, investigation into other top differentially expressed genes identified in Figure 2B such as Nucleoside Diphosphate Kinase 2 (NME2), Uridine Phosphorylase 1 (UPP1), Inosine Triphosphatase (ITPA), and Uridine-Cytidine Kinase 2 (UCK2) would be conducted. These genes have been associated with various cancers including cervical cancer (Hudelist et. Al, 2005; Qu et. Al, 2021; Lee et. Al, 2022; Fu et. Al, 2022). Additionally, future studies into other genes associated with chemoresistance in cervical cancer and other cancer types warrant further investigation for potential genetic therapeutic targets as demonstrated by this publication.

References

Tan X, Wang X, Liao X, Wang X, Jiang Z, Liang W, Cao C, Gong D, Hu Z, Tian X. Downregulation of VPS13C promotes cisplatin resistance in cervical cancer by upregulating GSTP1. iScience. 2023 Jul 10;26(8):107315. doi: 10.1016/j.isci.2023.107315. PMID: 37520723; PMCID: PMC10372835.

Okunade KS. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020 Jul;40(5):602-608. doi: 10.1080/01443615.2019.1634030. Epub 2019 Sep 10. Erratum in: J Obstet Gynaecol. 2020 May;40(4):590. PMID: 31500479; PMCID: PMC7062568.

Tian X., Wang X., Cui Z., Liu J., Huang X., Shi C., Zhang M., Liu T., Du X., Li R., et al. A Fifteen-Gene Classifier to Predict Neoadjuvant Chemotherapy Responses in Patients with Stage IB to IIB Squamous Cervical Cancer. *Adv. Sci.* 2021;8:2001978. doi: 10.1002/advs.202001978.

Hudelist G, Czerwenka K, Singer C, Pischinger K, Kubista E, Manavi M. cDNA array analysis of cytobrush-collected normal and malignant cervical epithelial cells: a feasibility study. Cancer Genet Cytogenet. 2005 Apr1;158(1):35-42. doi: 10.1016/j.cancergencyto.2004.08.007. PMID: 15771902.

Qu X, Shi Z, Guo J, Guo C, Qiu J, Hua K. Identification of a novel six-gene signature with potential prognostic and therapeutic value in cervical cancer. Cancer Med. 2021 Oct;10(19):6881-6896. doi: 10.1002/cam4.4054. Epub 2021 Sep 8. PMID: 34498424; PMCID: PMC8495282.

Lee Y, Jang EJ, Yoon HY, Yee J, Gwak HS. Effect of ITPA Polymorphism on Adverse Drug Reactions of 6-Mercaptopurine in Pediatric Patients with Acute Lymphoblastic Leukemia: A Systematic Review and Meta-Analysis. Pharmaceuticals (Basel). 2022 Mar 29;15(4):416. doi: 10.3390/ph15040416. PMID: 35455413; PMCID: PMC9027773.

Fu Y, Wei XD, Guo L, Wu K, Le J, Ma Y, Kong X, Tong Y, Wu H. The Metabolic and Non-Metabolic Roles of UCK2 in Tumor Progression. Front Oncol. 2022 May 20;12:904887. doi: 10.3389/fonc.2022.904887. PMID:35669416; PMCID: PMC9163393.