Using Bioinformatics for the identification of novel Epithelial to Mesenchymal transition regulators

Karim Kazi Neha
Molecular Biology and Biotechnology (Intensive)
School of Biological Sciences, The University of Hong Kong

Summer Research Fellowship (SRF) 2020 for Science Students
Poster Number: F7
Name: Karim Kazi Neha
University Number: 3035603631
Professor: Alice Wong
Major: Molecular Biology and Biotechnology (Intensive)

Abstract

As Epithelial to Mesenchymal transition (EMT) is the major cause of most of cancers, it is crucial to find novel epithelial markers to regulate EMT and stop cancer from progressing. 16,182 overlapping genes were found by our collaborator through a master algorithm and bioinformatics. However, not all of them are responsible for EMT, and thus, does not lead to ovarian cancer significantly. In order to find new EMT markers, the expression level of genes in high metastatic (HM) and non-metastatic (NM) environment is found, as well as the overall survival. Besides confirming whether they are indeed membrane proteins, correlation expression of the said genes with CDH1, CDH2, EpCAM and VIM were also checked. After rounds of elimination the 16,182 genes, we have found that UNC5B matches all the criteria. Taken together, UNC5B has a possibility of becoming a potential EMT marker. However, this requires further investigation.

Introduction

Epithelial-to-mesenchymal transition (EMT) is crucial for wound healing and in the development of the mesoderm and neural tube, however, it is also a major factor causing metastasis and tumor progression, thus associated with cell plasticity and cell-cell adhesion. By permitting the epithelial cells to differentiate into mesenchymal cells causing the loss of cell polarity, EMT increases cell mobility and invasiveness. Hence, regulation of EMT is imperative as these regulators are the switches that induce cancer stem-cell phenotype, block oncogene-induced senescence and suppress the host immune surveillance system. Besides the canonical markers of EMT which includes transcription factors (ZEB1, ZEB2, SNAI1, SLUG and TWIST1), epithelial markers (CDH1 and EpCAM) and mesenchymal markers (CDH2, VIM and FN1); novel cell surface markers are needed to increase the early detection of carcinoma or possible carcinoma candidate.

UNC5B (Unc-5 Netrin Receptor B) is a netrin receptor found on plasma membrane that functions by negatively regulating the branching of blood vessels in angiogenesis, besides performing its role in axon guidance. It also acts as a dependence receptor necessary for inducing apoptosis when not connected to netrin ligands, thus, playing an important role in cancer progression. They also act as a guide by influencing cellular size, formation of outgrowths and cellular movement – direction and rate. As a result, the functoin of UNc5B makes it a potential marker of EMT.

Logrank P value and Location

To find novel EMT cell markers, PubMed was used to ensure that nothing has been published regarding EMT and the genes in the gene list. A second list containing NM & HM RNA sequence data of the metastatic cancer cell HEYA8 was used to obtain the log2fold change (HM/NM) and Padj value. The log2fold change is the ratio of the expression of genes in HM to the NM; and is encouraged to have a value of 1 while the Padj value should be <0.05 to indicate significance in expression levels. Then The Human Protein Atlas was utilized to narrow down the genes responsible for the translation of cell markers, that is, cell surface proteins or transmembrane proteins. Based on these data, UNC5B was selected for further analysis.

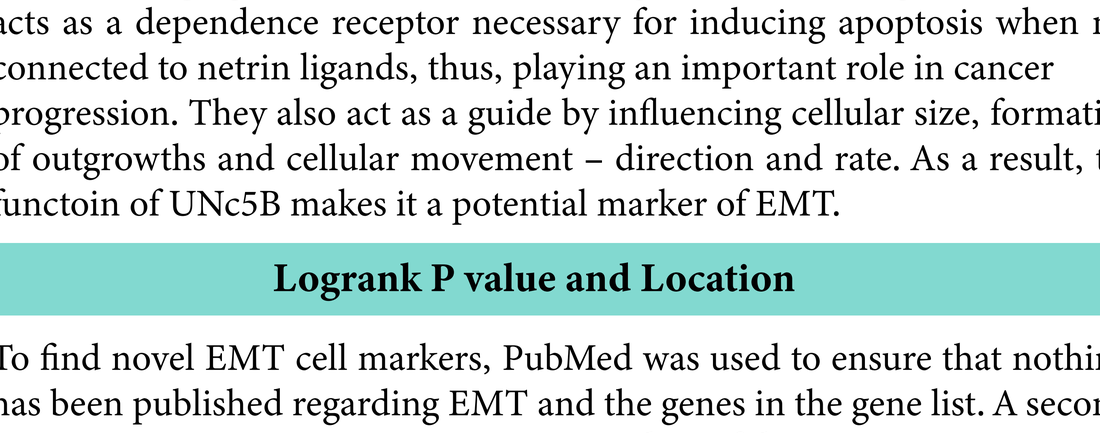

Figure 1: UNC5B expression level in HEYA8 NM and HM RNA sequence data (A); UNC5B location in cells (B)

From Figure 1 – (A), it can be said that UNC5B has a log2fold change of 1.432 and a Padj value of 0.0397 which indicates it is significantly expressed more in HM than in NM. Moreover, Figure 1–(B) shows the placement of the proteins translated by UNC5B inside the cell. The diagram enforces the idea of UNC5B being responsible of a membrane protein.

Correlation Plot

After the above-mentioned series of elimination and selection, Pearson correlation between the selected genes and canonical EMT markers such E-cadherin (CDH1), N-cadherin (CDH2), epithelial cell adhesion molecule (EpCAM) and vimentin (VIM) were found in the case of 33 pan-cancers using The Cancer Genome Atlas (TCGA) dataset from Gene Expression Profiling Interactive Analysis (GEPIA) website. The correlation plot proved whether the gene correlates with EMT by utilizing the R-value and p-value of the shortlisted genes. If the genes had a positive R-value with CDH1 and EpCAM; and negative R-value with CDH2 and VIM, it indicated a epithelial marker. On the contrary, a negative R-value with CDH1 and EpCAM; and positive R-value with CDH2 and VIM, it indicated a mesenchymal marker.

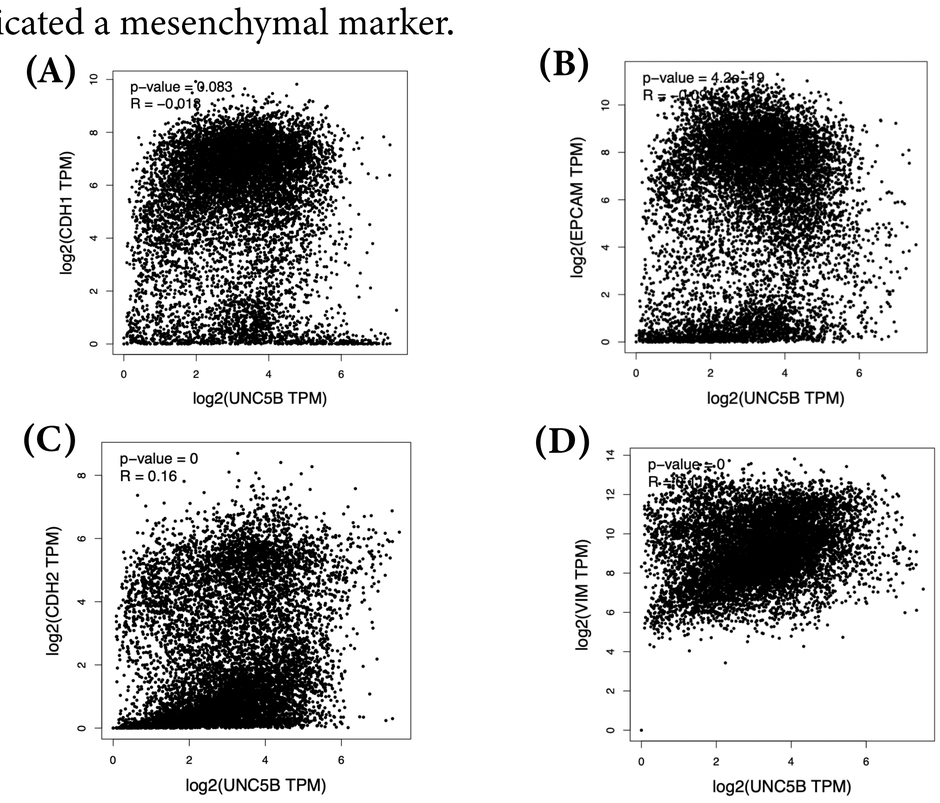

Figure 2: Correlation between UNC5B and CDH1(A); EpCAM (B); CDH2 (C); and VIM (D) in ovarian cancer from TCGA dataset

Figure 2 – (A) and (B) illustrates that UNC5B has negative correlation with CDH1 and EpCAM with R-values -0.081 and -0.091 respectively. Figure 2 – (C) and (D) on the other hand, depicts the positive correlation between CDH2 and VIM with R-values 0.16 and 0.11 respectively. These figures justify UNC5B being an essential mesenchymal marker for EMT.

Survival Analysis

Overall survival of UNC5B was obtained obased on the TCGA dataset. After substantiating that the overall survival is significant by having a log-rank p value of < 0.05, the association of a high or low expression level of the gene with poor overall survival time can be determined. This says whether the gene plays an enhanced or suppressed role on ovarian cancer.

As shown in Figure 3, the logrank p-value of UNC5B is 0.025 which indicates significant overall survival. This in turn shows us that higher expression level of UNC5B has poor overall survival.

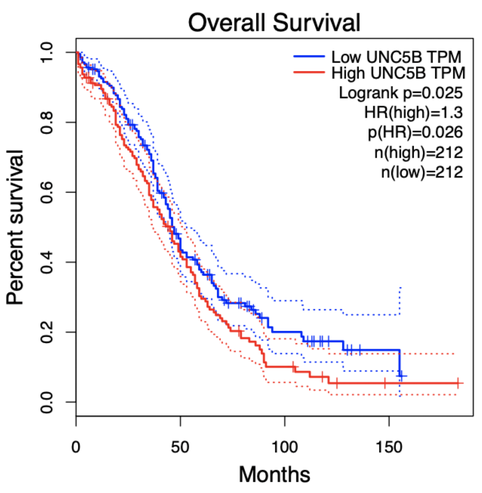

Figure 3: Overall Survival of UNC5B expression in ovarian cancer from TCGA dataset

Conclusion

The final gene after many eliminations is UNC5B. UNC5B meets all the criteria as it novel, has log2fold change >1, is a cell surface marker, correlated negatively with CDH1 and EpCAM and positively with CDH2 and VIM and. The survival plot demonstrates that poor overall survival correlates with higher expression level of UNC5B, making it a potential EMT marker. However, the effect of UNC5B still needs to be verified for it to be an EMT marker in the future.

References

1. Zheng, H., & Kang, Y. (2013, April 22). Multilayer control of the EMT master regulators.
2. UniProt Consortium European Bioinformatics Institute Protein Information Resource SIB Swiss Institute of Bioinformatics. (2020, August 12). Netrin receptor UNC5B.