A Workflow Utilizing R/Bioconductor, GSEA, and TIMER 2.0 to explore the role of the Vav Protein Family in Cutaneous Melanoma.

Avila A., Anselmino L., Menacho-Marquez A.

Vav proteins are RHO guanine nucleotide exchange factors (GEFs). This family consists of three members, which typically exhibit functional redundancy and are associated with proactive functions in cancer. However, their role in melanoma remains largely unexplored.

Our aim was to establish a systematic approach, utilizing bioinformatic techniques, to investigate the role of each member of the Vav family in melanoma.

Gene expression data from cutaneous melanoma patients were obtained from the 'Cancer Genome Atlas' database. Raw counts were subsequently normalized to counts per million (CPM) using the 'edgeR' package. The patient cohort (n=460) was stratified based on high or low expression levels of Vav1, Vav2, and Vav3. Survival plots were generated using the Kaplan-Meier estimator and ‘survminer’ package. The log-rank test revealed an association between high Vav2 expression and poorer prognosis, whereas elevated Vav1 and Vav3 expressions correlated with increased patient survival probability ($p<0.05$ in each case).

Gene set enrichment analysis was conducted for each comparison group using the GSEA software. To assess immune and stromal cell infiltration in tumor tissues, Immune Score and Microenvironment Score were calculated based on gene expression profiles of the tumor microenvironment, employing the ESTIMATE and xCell algorithms. Both Scores showed a strong and positive association with Vav1 and Vav3 expressions ($p<0.001$). Then, using eight different algorithms, with the 'estimate' package and the TIMER2.0 application, correlation with some cell types was evaluated. A robust positive correlation was identified between Vav1 expression and some types of immune cell signatures ($p<0.001$). Conversely, no significant correlation was observed between Vav2 or Vav3 expression and cell types.

Our findings suggest that a favorable prognosis in melanoma is linked to elevated expressions of Vav1 and Vav3, coupled with reduced Vav2 expression. This prognosis may arise from Vav1's impact on intercellular communication within the

tumor microenvironment, while heightened Vav3 expression could regulate the activation of tumor cell signaling pathways, thereby promoting greater immunogenicity.

Our study presents a comprehensive pipeline that could serve to explore the implications of other proteins in diverse disease contexts.

References

- Colaprico A., Silva T.C., Olsen C., Garofano L., Cava C., Garolini D., Sabedot T.S., Malta T.M.,Pagnotta S.M., Castiglioni I., Ceccarelli M., Bontempi G., Noushmehr H. 2015. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 44(8):e71. doi: 10.1093/nar/gkv1507
- Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., Ellis B., Gautier L.,Ge Y., Gentry J. 2004.Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 5(10):R80
- Kassambara, A. 2016. survminer R package: Survival Data Analysis and Visualization. Statistical tools for high-throughput data analysis
- Mounir M., Lucchetta M., Silva C.T., Olsen C., Bontempi G., Chen X., Noushmehr H., Colaprico A., Papaleo E. 2019. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS computational biology, 15(3), e1006701
- Robinson M.D. & Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology. 11, R25
- Silva C.T., Colaprico A., Olsen C., D'Angelo F., Bontempi G., Ceccarelli M., Noushmehr H. 2016. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. F1000Research, 5.
- Wickman H. & Grolemud G. 2018. Tidyverse package. https://www.tidyverse.org/