XIII Argentine Congress of Bioinformatics and Computational Biology
XIII International Conference of the Iberoamerican Society of Bioinformatics
III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData

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

Avila Aylén[1], Anselmino Luciano[1, 2], Menacho-Marquez Mauricio [1, 2]

[1] Centro de Investigación y Producción de Reactivos Biológicos (CIPReB – FCM-UNR)

[2] Instituto de Inmunología Clínica y Experimental de Rosario (IDICER-CONICET-UNR)

aylen.avila@gmail.com

Melanoma and Vav proteins

Melanoma represents the most aggressive manifestation of skin cancer, arising from the malignant transformation of cutaneous melanocytes. Its global incidence is rising, and it stands out as one of the most highly metastatic cancer types with limited treatment options available.

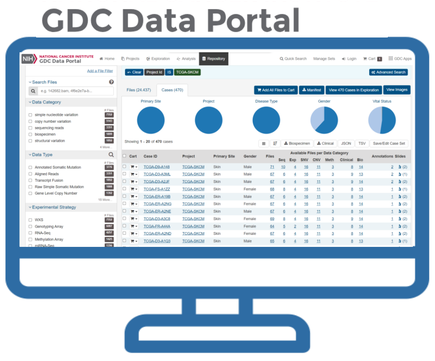
Within the intricate landscape of cancer biology, the Vav family of proteins assumes a significant role as activators of Rho GTPases, which are implicated in pro-oncogenic processes when their activity is deregulated.

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.

Search, Download, and Data Cleaning

NATIONAL CANCER INSTITUTE
GDC Data Portal



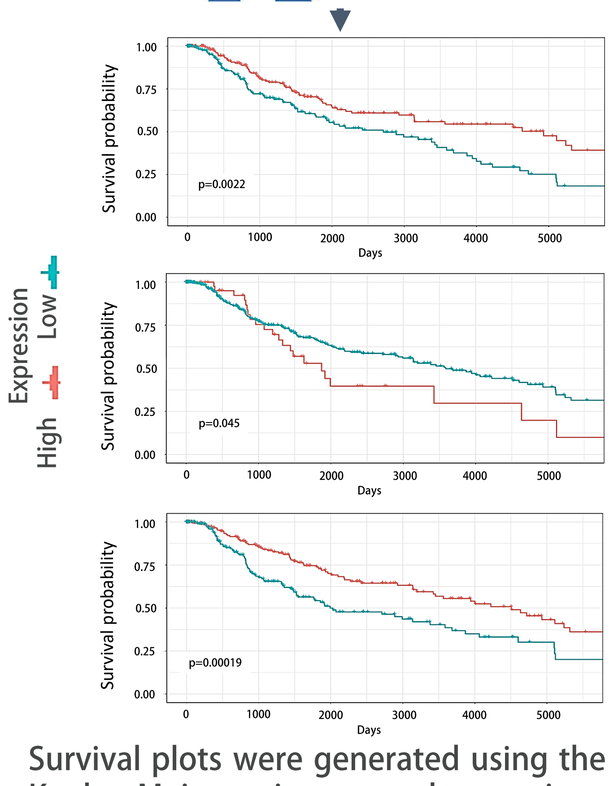
```
library("SummarizedExperiment")
library("TCGAbiolinks")
library("RTCGA.clinical")
library("edgeR")

query.exp <- GDCquery(
  project = "TCGA-SKCM",
  data.category = "Transcriptome Profiling",
  data.type = "Gene Expression Quantification",
  workflow.type = "STAR - Counts")

GDCdownload(
  query = query.exp,
  files.per.chunk = 100)

skcm.exp <- GDCprepare(
  query = query.exp,
  save = TRUE,
  save.filename = "skcmExp.rda")
```

	ID	times	status	VAV1	VAV2	VAV3	FenoVAV3	Fe
1	TCGA-3N-A9WB	518	1	0.7584649	15.448733	11.0975395	High	Lc
2	TCGA-3N-A9WC	1856	0	26.3088208	24.706554	51.3120911	High	Lc
3	TCGA-3N-A9WD	395	1	17.2641888	38.202245	7.7658981	Low	H
4	TCGA-BF-A1PU	387	0	0.7206094	30.368538	1.7353450	Low	H
5	TCGA-BF-A1PV	14	0	1.9076738	37.608427	0.8977289	Low	H
6	TCGA-BF-A1PX	282	1	23.3542221	22.089016	30.0374890	High	Lc
7	TCGA-BF-A1PZ	12	0	2.9778407	1.980264	1.6973692	Low	Lc
8	TCGA-BF-A1Q0	17	0	7.1820481	56.883170	4.9397655	Low	H
9	TCGA-BF-A3DJ	464	0	9.6267260	25.806781	5.1711129	Low	Lc
10	TCGA-BF-A3DL	28	0	2.1988719	7.789224	0.5404007	Low	Lc
11	TCGA-BF-A3DM	14	0	2.4255881	14.399794	5.6027668	Low	Lc
12	TCGA-BF-A3DN	32	0	2.8289981	10.858781	13.6877786	High	Lc
13	TCGA-BF-A5EO	338	0	6.9955220	13.714625	1.5947239	Low	Lc
14	TCGA-BF-A5EP	11	0	1.5407419	26.992998	1.7808576	Low	Lc
15	TCGA-BF-A5EQ	12	0	12.6369774	30.269510	4.1859988	Low	H
16	TCGA-BF-A5ER	12	0	1.1346008	30.890422	0.7320005	Low	H

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.

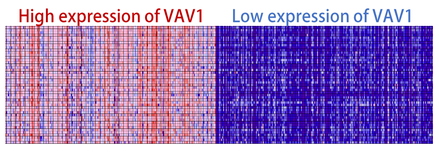
Analisys

'survminer'



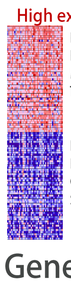
Survival plots were generated using the Kaplan-Meier estimator and 'survminer' package. The log-rank test revealed an association between high Vav2 expression and poorer prognosis, whereas elevated Vav1 and Vav3 expressions correlated with increased patient survival probability ($p<0.05$).

GSEA



ABI3, AOAH, APBB1IP, ARHGAP30, BIN2, C1QC, CCRL2, CD300A, CD300LF, CD84, CD86, CYBB, CYTH4, DOK2, GIMAP4, GIMAP6, GNGT2, HAVCR2, HCK, CLS1, IL10RA, IL15RA, IL21R, ITGB2, LAIR1, LAPTM5, LILRB1, LILRB2, MNDA, MS4A6A, MYO1F, NCF2, NCF4, NCKAP1L, P2RY6, PLEK, PTAFR, SELPLG, SIGLEC9, SIRPB2, SLA, SLA2, SLC7A7, SNX20, SPI1, STAC3, TNFAIP8L2, TNFRSF1B, VAV1

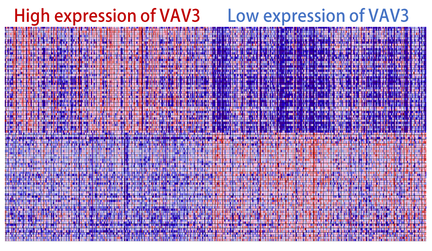
Genes positively related with the inhibition of tumor metastasis, reduction of cell motility, prevention of damage from inflammatory processes, activation of Ras for remodeling of the cytoskeleton and processes related to immune homeostasis



Positive relation: ANKS6, ARID3A, ATAD2B, BRD3, CAMSAP1, CCNJ, CDCA7, DDX31, EHMT1, EML4, EXOSC2, FAM117B, FBXW2, FUBP3, GPSM1, GTF3C4, HDAC7, HNRNPH3, IKBKAP, INPP5E, KHDRBS1, KLF12, KSR1, MAPK8, MSL1, NOL8, NUP188, PDCL, PLAGL2, PLEKHG2, PSMD5, QSOX2, RBMX, SET, SIRT1, SPIN1, STRBP, SURF6, TAF5, TEX10, TOR4A, TSC1, TTC21B, TTF1, VAV2, WDR5, XPA, ZBTB26, ZBTB34, ZBTB5

Negative relation: ADPRHL2, AGAP3, AGTRAP, ATP6V0A1, ATP6V0B, ATP6V0D1, ATP6V1B2, ATP6V1E1, BCAP31, BIN3, BRI3, CD63, COPG1, COX17, CYB5R1, DBNDD1, DGKG, DHRS1, DNAJC30, DNASE2, EMC3, FASTK, GPNMB, GRN, GSTZ1, LGALS3, LGALS3BP, MBP, MREG, MYO1D, NAPA, NR1H2, RETSAT, RRAGD, RTP4, S100B, SLC7A5, SNCA, SPECC1, SPSB1, STAP2, STYXL1, TMEM53, TMEM59, TYR, UPP1, WBP2, ZMYND12

Genes positively related to division cellular and transcription; and a negative relationship with processes related to increased cellular acidity, proton transport and vacuole formation, were found.



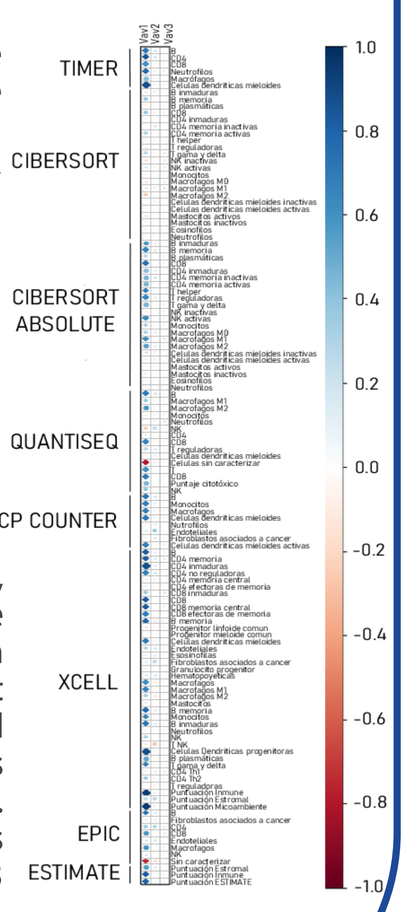
The gene set analisys did not suggest significative and characteristic differences for each phenotype.

ESTIMATE / TIMER2.0

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 ($p<0.001$).

Score		VAV1		VAV2		VAV3	
		Score	p value	Score	p value	Score	p value
Immune	xCell	0.923	2.2E-16	-0.035	0.5149	0.278	1.7E-09
	estimate	0.951	2.2E-16	-0.002	0.9696	0.275	2.6E-09
Microenvironment	xCell	0.935	2.2E-16	0.038	0.4229	0.288	3.9E-10
	estimate	0.931	2.2E-16	0.078	0.0937	0.289	2.5E-10

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.



Conclusions

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