

Exvar: A gene expression and genetic variation data analysis and visualization R package

Supplementary material: Pipeline validation Example

Hiba Ben Aribi¹, Imraan Dixon², Najla Abassi³, and Olaitan I. Awe^{4,5}

¹ Faculty of Sciences of Tunis, University of Tunis El Manar, Tunis, Tunisia

² Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

³ Higher Institute of Biotechnology Sidi Thabet, Manouba University, Tunisia

⁴ Department of Computer Science, University of Ibadan, Ibadan, Oyo State, Nigeria

⁵ African Society for Bioinformatics and Computational Biology, Cape Town, South Africa

Methodology

A public RNA-seq dataset from the SRA database SRP074425 (Wang et al., 2016) was analyzed to validate the *Homo Sapiens* specific pipeline. The dataset corresponds to RNA sequencing data from the glioma cells from human patients and healthy brain cells (Wang et al., 2016). The dataset consists of 20 individuals, including 9 patients with a primary tumor, 6 patients with a recurrent tumor, and 5 control individuals. The sex ratio for patients with primary tumors is predominantly male with a ratio of 7:2, while patients with recurrent tumors have a sex ratio of 4:2. The controls have a sex ratio of 3:2.

The FASTQ files were first pre-processed using the `processfastq()` function. The gene count was determined using the `counts()` function and the output count CSV file was visualized using the `vizexp()` function to identify the DEGs and the associated ontologies. The variants were called using the `callsnp()`, `callindel()`, and `callcnv()` functions, and the output VCF files were visualized using the `vizsnp()` and `vizcnv()` functions.

Results

(i) Gene expression and ontology analysis:

The p-value was set to less than 0.05 and the logFC value was set to $|\text{LogFC}| > 2$, to define the differentially expressed genes (DEG). The expression analysis identified 2247 DEG, of which 2245 were upregulated and 2 were downregulated genes, as represented in Figure 1.

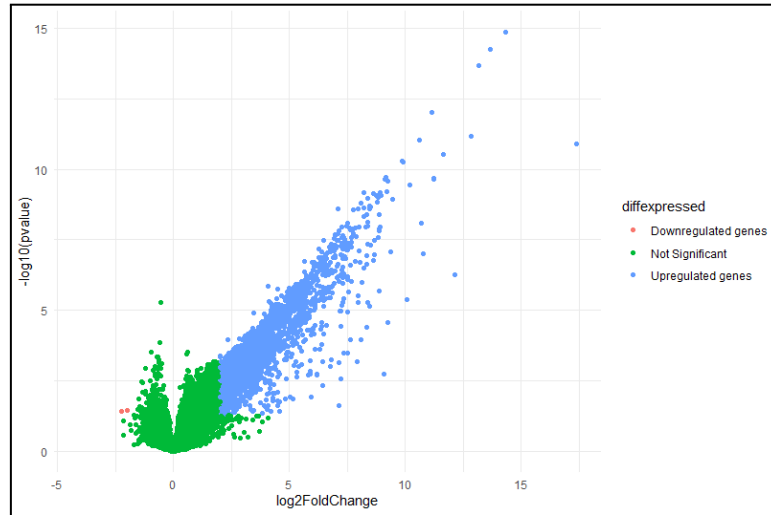
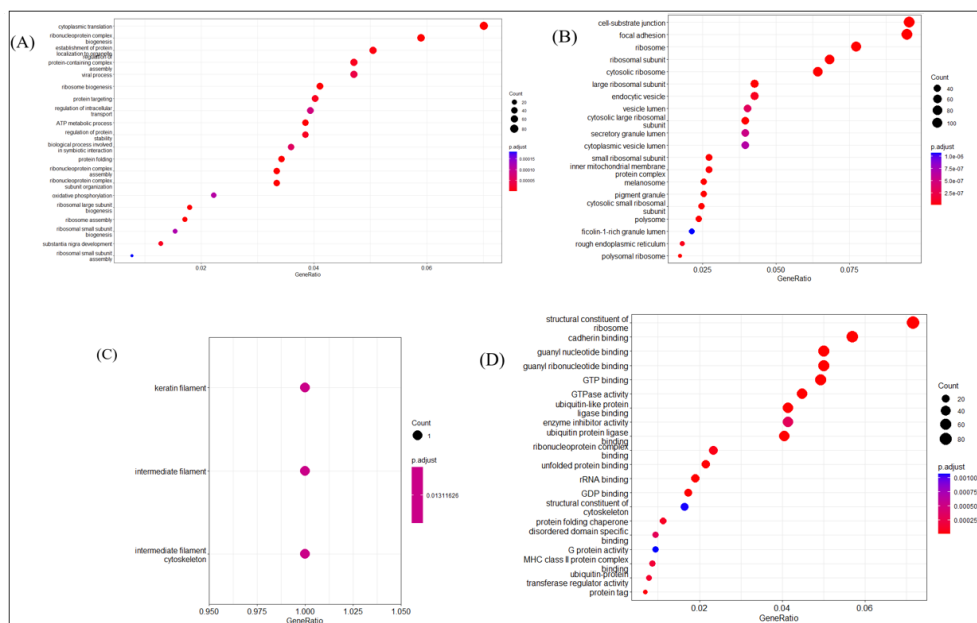


Figure 1. Volcano plot illustrating the differentially expressed genes

The top 20 statistically significant Biological Processes and Cellular Components associated with the upregulated genes are represented respectively in Figure 9 (A) and (B). Three Cellular Components are associated with the downregulated genes, as represented in Figure 9 (C). The top 20 statistically significant Molecular Functions associated with the upregulated genes are represented in Figure 9 (D). The ontology analysis did not associate any statistically significant Biological Process and Molecular Function with the downregulated genes.



(ii) SNP data analysis

The analysis shows that 7 upregulated genes and no downregulated genes have SNPs (Table 1).

(iii) CNV data analysis

The p-value was set to less than 0.01 to define the significance threshold to filter recurrent CNVs. The target gene biotype was set to “protein-coding”. Overlap analysis identified CNVs in 15 DEGs, all upregulated genes (Table 1).

Table 1. Human pipeline validation dataset analysis result

Genes	Expression	SNPs	CNVs
ESR1, FANCA, FBN1, GFRA1, GHR, GRB10, NRG1, KIFC3, NEK1, SLC13A1, PLCH2, PTPRT, ACOT7, NINL, GRIP1, OSBP2, TENM2, CDH22, TRPM8, KATNAL2, FRMD5, SHANK3, COL23A1, SYT12, PROSER3, PEBP4, EXD1, DNAH12, EPHA6, COL28A1, GXYLT2	Upregulated	yes	yes
BSG, SSH2, TNFRSF13C, ASB8	Downregulated	yes	yes
APOB, BDNF, BMP1, CAPN5, CDH12, CENPE, CEACAM3, DSG2, EGR3, FGF10, GAD2, GRIK3, GRM1, OTX1, PLA2G2A, RRM2, SCN7A, SFTPC, MED22, SYN2, PXDN, ST8SIA2, TNK1, BRSK2, MAGI2, TRDN, ABCA7, PRMT5, SLC24A2, TMEM59L, SEZ6L2, PLEK2, DAZAP1, FAM184B, SOX8, HES2, CYP2W1, CRTAC1, LRRC36, HR, PRMT8, BARHL1, CHST8, LMF1, CENPM, GALNT14, ESPN, SLC25A18, CORO6, SLX9, ZIC5, PLXNA4, FHAD1, SLC16A10, LYSMD4, TEK5, REM2, STKLD1, GLIS3, ANKRD24, PROSER2, MDGA1, PIWIL3	Upregulated	no	yes
HSPA1B, ASS1, SLC15A2, TIAM1, ARL6IP5, NXT1, OTULINL, APBB1IP, BHLHE41, REEP4	Downregulated	no	yes
AOAH, C3, C5AR1, CAPZB, TNFSF8, CD53, CD74, CD81, CR1, CSF1R, CSF3R, DDX6, DOCK2, EPB41, ICAM1, IL6R, IL13RA1, IL16, ITPR2, LEPR, ABLIM1, P2RX7, PAK1, PGM5, PTAFR, PTPRC, PYGL, RPS6KA3, SELL, SOAT1, SORL1, TBXAS1, H2AC6, RGS5, TM9SF2, ABCG2, LPIN2, TOMM70, JOSD1, DHRS9, LHFPL2, HPSE, LILRB1, NUDT21, PLAAT3, SLC02B1, CD300A, TFEC, EXPH5, CYFIP1, SSBP2, GCA, COQ2, PYCARD, NDUFA13, GPN3, AIG1, BIN2, PCYOX1, PCBP3, CPVL, GDAP2, LPCAT2, LPAR5, SLC39A8, AHNK, CPED1, DOCK8, AIF1L, RASSF5, LRRC8C, LZIC, PLXDC2, PAQR8, NAV3, MGMT1, NIBAN1, FOXP2, SIGLEC11, TAGAP, ZNF816, ZNF573, TATDN3, SIRPA, SYT6, SLC16A14, SLC2A12, TMEM154, SLC9A9, SRGAP2D	Downregulated	yes	no
ADCY8, ADORA1, ALK, ANK3, ZFHX3, ATP2A1, ATP2B2, ATP4B, BLM, CXCR5, BMP8B, CA5A, CA12, CACNA1C, CAMK2B, CASR, CDH1, CDH4, CHGB, CHRNA1, CHRNA2, CHRNA2, CHRNE, COL4A6, CPS1, DCC, DPEP1, SARDH, DSPP, ERCC2, EZH2, FGD1, FGF13, FGFR2, AFF2, GABBR1, NR6A1, GFRA3, GHRHR, GLI2, GNAT1, GRIK2, GRIK4, GRIN2A, HRG, HSPG2, HTR2C, IDUA, INSR, KCNC1, KCNC2, KCNJ1, KCNN1, KCNN3, KRT82, LAMA5, LHCGR, LOXL2, LRP2, LRP4, CHST6, MEIS2, TRPM1, MPL, MSRA, MUC6, MYH7, MYO1A, NPY6R, NRCAM, P2RY2, PAPP, PBX3, PDE6A, PIGF, PKP2, PLXNA2, PTK7, PTPRR, RAD9A, MOK, ROBO1, ROBO2, RTN2, RYR2, SCN8A, SGCD, ST3GAL3, SLC5A1, SLC8A3, SLC9A5, SOX4, TBX5, NR2F2, TG, TP73, TRPV1, ZAN, MAFK, ADAM12, HMGA2, DPF1, ITGA8, GCM1, STC2, ADAM19, ADAM23, SCEL, CACNA1I, KALRN, NFS1, PKMYT1, CBFA2T2, NEURL1, DGKI, LARGE1, MAGI1, NRXN3, NRXN2, TSIX, ADAMTS4, ADAMTS2, VPS9D1, CELSR1, CROCC, PCDHGA8, IFT140, PCDHA9, LZTS3, DLGAP5, FAM30A, PTPRU, SGK2, PRG4, MIR9-1HG, CARM1, IGF2BP3, PNMA2, POLQ, ALDH1L1, ME3, PRSS21, SDS, ADAMTS7, ADAMTS6, GPR176, EXOC3, RPH3A, LAMB2P1, TTC39A, PLCH1, UNC13A, MYO16, HECW1, FSTL4, ATP10B, COLGALT2, PHF8, MAPK8IP3, FBXL7, PLCB1, ASTN2, KAZN, MTUS2, IQCE, SSBP3, NGEF, POU2F3, KRT23, L3MBTL1, GPSM1, TTLL3, TIAM2, CPNE7, STK36, TMOD4, TRPM5, NPC1L1, HOOK2, ATP6V0A4, ARHGEF4, IMPG2, PDE11A, LRP1B, C21orf58, TSPEAR, KCNK10, SDK2, UGT1A8, SLC6A20, LAX1, MCM10, SVOP, NLRP2, TENM3, CFAP44, PCDHGC5, PCDHGB2, PCDHAC1, PCDHA13, PCDHA11, PCDHA7, PCDHA6, TEX14, RNF17, KIAA1217, WDR93, KNL1, ANO2, NDRG3, PLEKHG5, STRIP2, CRACD, HECW2, LRRC7, NLGN2, MARCHF, CCDC191, KCNT1, SHROOM3, CAMSAP3, CPNE5, SLC7A14, SPTBN4, CREBZF, NTN4, CACNG6, NPAS3, SMOC1, CIAO3, PCDH15, TMEM108, LILRP2, MMEL1, PAGR1, CLIP4, GRTP1, EHMT1, JHY, CNTNAP3, GRHL2, CCDC33, SPSB1,	Upregulated	yes	no

<p>OPA3, CFAP43, SCUBE1, LPAL2, ARHGAP39, TSEN2, ASXL3, KCNH6, COL21A1, ST6GALNAC5, PMFBP1, NUF2, CDCA7, BRIP1, ADGRV1, ARMC2, DRC7, THOC3, ZBED3, MEGF10, FNDC1, GTPBP3, LNX1, HDGFL2, PLPP7, ARHGEF39, AOPEP, FAM222A, FIBCD1, TNS4, CCNB3, ZFHX2, TRIM15, GLYATL1, DRC1, PRDM6, EGFEM1P, RP1L1, CSMD2, TMEM132B, PSMG3-AS1, KIR3DL3, ZNF837, TOP1MT, IP6K3, TMC2, MORN4, MMP21, LARGE2, SLC24A4, TMEM266, AGBL1, RSKR, LOXHD1, CCDC159, NLRP13, NLRP8, TDRD10, DNAI3, MYOM3, DCST2, SYT2, MIR1-1HG, CLHC1, CPNE4, ZDHHC19, COL6A6, PNCK, ADAMTS14, CNBD2, DNHD1, BEST3, ISM2, BEAN1, MGAT5B, RBFOX3, DNAH2, CCBE1, KLC3, SPC24, SIX5, CILP2, ACTG1P25, CNIH3, CFAP57, ENTHD1, IGSF11, SLC25A48, VKORC1L1, AMZ1, RIBC1, OTUD7A, ANKFN1, MARCHF10, TOGARAM2, IDO2, LINC00858, PLPP4, DNAH10, PIANP, GRAMD2A, PATL2, VWA3B, SLC51A, CENPV, ZBTB7C, C8orf74, TDRD6, GPC2, CDHR3, NYAP1, CASC2, COL24A1, COL6A5, SYCE2, SCML4, SUN3, C2orf72, HS6ST3, KSR2, TMPRSS12, NPW, HCCAT5, TTLL6, SLC26A11, WDR62, NWD1, FAM182A, MIR663AHG, C20orf203, SLC9C1, SH3PXD2B, LINC00174, FAM83H, LINC02209, LINC01289, LIN9, NOP14-AS1, NLRP14, C1QL4, CIBAR2, CCDC39, H1-10-AS1, ZC3H12B, PKD1L3, CDKL4, LRRIQ4, ATP13A5, EYS, SERINC2, LNP1, TEX9, KIF7, STUM, LHFPL4, LHFPL3, ILDR2, TMEM233, TRABD2B, GTF2IRD2B, MROH5, GRID2IP, MEX3D, EML6, LINC01122, ANKHD1-EIF4EBP3, DENND6B, GPR179, SMG1P5, GOLGA6FP, FAM227A, LINC00598, GUSBP1, ENTPD1-AS1, ANKDD1B, FHIP1A, SHISA9, LINC01004, ARHGEF33, EOLA2-DT, HYDIN2, SLC25A25-AS1, GNG12-AS1, MSH5-SAPCD1, PYCARD-AS1, KIAA2012, NALCN-AS1, LINC01993</p>			
---	--	--	--

Reference

Wang J, Cazzato E, Ladewig E, Frattini V, Rosenbloom DI, Zairis S, Abate F, Liu Z, et al. Clonal evolution of glioblastoma under therapy. *Nat Genet.* 2016 Jul;48(7):768-76. Doi: <https://doi.org/10.1038/ng.3590>. Epub 2016 Jun 6.