

Availability of data and materials

Our data was from two sources:

1. GEPIA: <http://gepia.cancer-pku.cn/>

All the correlation analysis in our research are from GEPIA. GEPIA is an open-access public dataset. Everyone can log in the website without an account.

2. TCGA datasets

We used a paid bioinformatic website (<https://www.xiantao.love/databases>) for bioinformatics analysis of expression, prognosis, ROC curve of SPP1. Our accession numbers is 15821960178. And anyone is allowed to access the site's database with their accession number. The bioinformatics analysis of expression, prognosis, ROC curve of SPP1 is mainly from TCGA datasets.

2.1 The expression of SPP1 in GBM and LGG

- Software: R (version 3.6.3) (statistical analysis and visualization)
- R package: mainly ggplot2 [version 3.3.3] (for visualization)
- Molecule: SPP1[ENSG00000118785]
- Disease: Multiform glioblastoma and Low grade glioma
- Data: UCSC XENA (<https://xenabrowser.net/datapages/>) uniformly processed RNAseq data in TPM format of TCGA and GTEx through the Toil process (Vivian J et al., 2017). Extract the corresponding normal tissue data from TCGA's GBM or LGG and GTEx.
- Data conversion: RNAseq data in TPM (transcripts per million reads) format and log2 conversion for expression comparison between samples
- Data filtering: None
- Significance mark: ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$

2.2 The prognostic survival of SPP1 in GBM and LGG

Software: R (version 3.6.3) (statistical analysis and visualization)

- R package: survminer package [0.4.9 version] (for visualization), survival package [3.2-10 version] (for statistical analysis of survival data)
- Molecule: SPP1[ENSG00000118785]
- Grouping: Group with the smallest p value
- Prognosis type: Overall Survival
- Disease: Multiform glioblastoma and Low grade glioma
- Data: TCGA (<https://portal.gdc.cancer.gov/>) GBM and LGG project in level 3 HTSeq-FPKM format RNAseq data
- Data conversion: RNAseq data in FPKM (Fragments Per Kilobase per Million) format is converted into TPM (transcripts per million reads) format and log2 conversion is performed
- Supplementary data: The prognostic data comes from a Cell article (LIU, Jianfang, et al., 2018)
- Data filtering: remove control/normal (not all items have control/normal) + retain clinical information

2.3 ROC curve of SPP1 in GBM and LGG

- Software: R (version 3.6.3) (statistical analysis and visualization)
- R package: pROC package [1.17.0.1 version] (for analysis) || ggplot2 package [3.3.3 version] (for visualization)
- Molecule: SPP1[ENSG00000118785]
- Clinical variables: Tumor vs Normal
- Disease: Multiform glioblastoma and Low grade glioma
- Data: UCSC XENA (<https://xenabrowser.net/datapages/>) uniformly processed RNAseq data in TPM format of TCGA and GTEx through the Toil process (Vivian J et al., 2017). Extract the corresponding normal tissue data in GBM or LGG of TCGA and GTEx.
- Data conversion: RNAseq data in TPM (transcripts per million reads) format and log2

conversion for expression comparison between samples

- Data filtering: None
- ROC result interpretation: the abscissa is the false positive rate (FPR), the ordinate is the true positive rate (True Positive Rate, TPR).