ARMT

Description

ARMT is auto RNA-seq data minning tool.
The aim of ARMT is :
i) integrate and facilitate the downstream analysis of RNA-seq data,
ii) provide the way to analyze genes set according to GSVA,
iii) explore the correlation of gene expression and mutation.
Author:
Guanda Huang —— 202010108432@mail.scut.edu.cn
Hongli Du —— hldu@scut.edu.cn
Year: 2021
License: GPL(>=3)

Function

Provide TCGA clinical data
Create .gmt file for arbitrary gene sets
Normalization of counts matrix (TPM)
Gene set variant analysis
Survival analysis
Differential analysis
Correlation analysis
Analysis for multiple sets of data
Plot the mutation information

Feature

Easy to use
Automation
Visualization
Comprehensiveness
Integrating GSVA score analysis

Catalog

1. The structure of ARMT

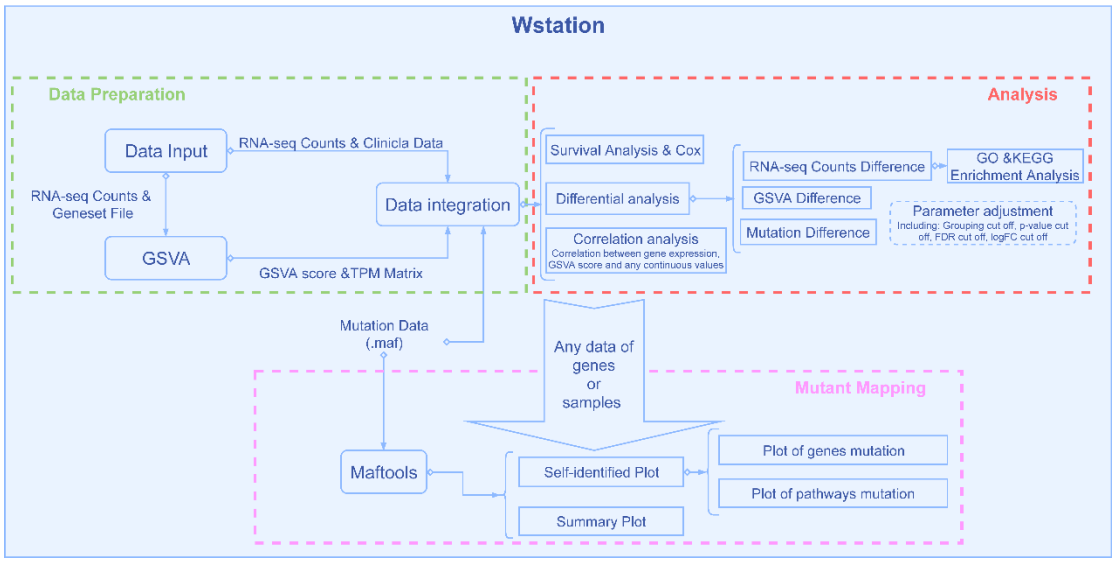


Figure 1. The workflow and structure of ARMT

2. Dependencies

R >=4.0.2, Rstudio, R packages("devtools" or "remotes")

The other dependent R packages are automatically installed, including:
maftools
GSVA
GSVAdata
limma
GSEAbase
org.Hs.eg.db
edgeR
survival
clusterProfiler
DOSE
...
By default, outdated dependencies are automatically upgraded. In interactive sessions you can select a subset of the dependencies to upgrade.

3. Installation

To install this package from Github, please, use the code below.

```
if (!requireNamespace("devtools", quietly = TRUE))

install.packages("devtools")

devtools::install_github("Dulab2020/ARMT")
```

4. Quick Star

The following commands should be used to start the graphical user interface (GUI).

```
ARMT::run_app()
```

5. Graphical User Interface (GUI) of ARMT

The GUI is developed based on 'shiny' package, and has four pages including: **Data, Normalization&GSVA, Integration&Analysis, Mutant mapping**

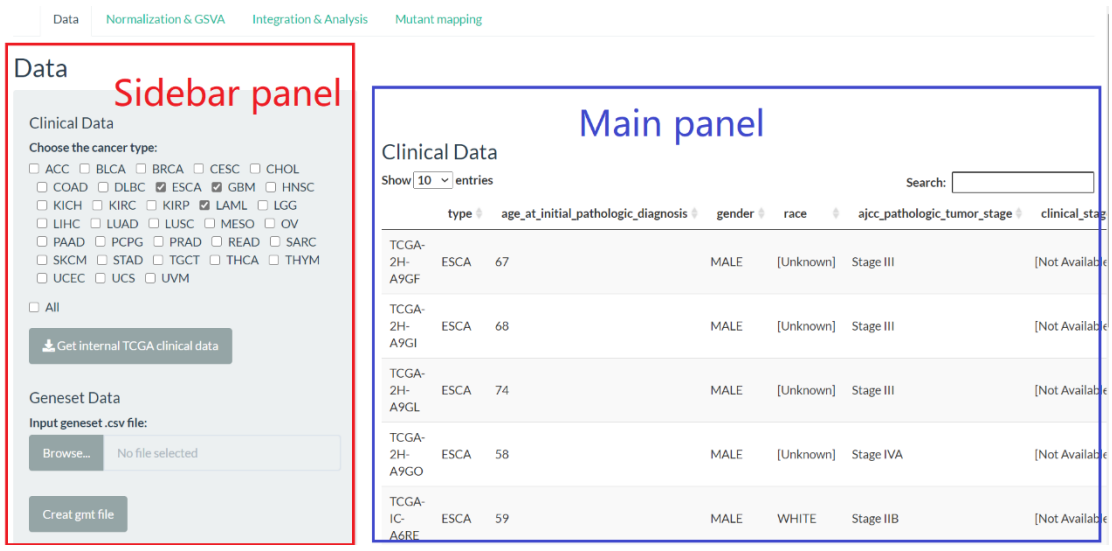


Figure 2. GUI of ARMT: The GUI consists of two parts, sidebar panel at left and main panel at right.

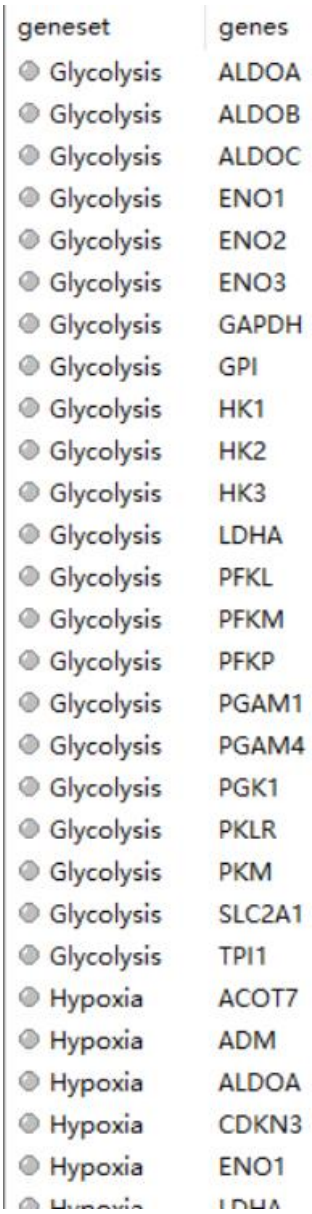
The sidebar panel is used to input data and adjust parameters; the main panel is used to demonstrate and save the results.

6. The page of ARMT

.6.1. Data

This page of ARMT provides TCGA clinical data and builds .gmt file of arbitrary gene sets. Through the internal datasets of ARMT, you can get TCGA clinical data of 33 cancer types. (*Figure2*)

To build a .gmt file, a .csv file should be provided, and it must contain two column as shown in *Figure3*.



geneset	genes
Glycolysis	ALDOA
Glycolysis	ALDOB
Glycolysis	ALDOC
Glycolysis	ENO1
Glycolysis	ENO2
Glycolysis	ENO3
Glycolysis	GAPDH
Glycolysis	GPI
Glycolysis	HK1
Glycolysis	HK2
Glycolysis	HK3
Glycolysis	LDHA
Glycolysis	PFKL
Glycolysis	PFKM
Glycolysis	PFKP
Glycolysis	PGAM1
Glycolysis	PGAM4
Glycolysis	PGK1
Glycolysis	PKLR
Glycolysis	PKM
Glycolysis	SLC2A1
Glycolysis	TPI1
Hypoxia	ACOT7
Hypoxia	ADM
Hypoxia	ALDOA
Hypoxia	CDKN3
Hypoxia	ENO1

Figure 3. The .csv file to create .gmt file

This file in *Figure3* has two columns. The first column declares gene sets. The second one contains genes in corresponding gene set.

6.2. Normalization&GSVA

In this page, ARMT automatically normalize the expression counts matrix to TPM matrix, and use log2(TPM) to process gene set variant analysis (GSVA). The input data must be gene expression matrix with **Ensembl** ID, and the ID can be transformed to **Symbol** through normalization.

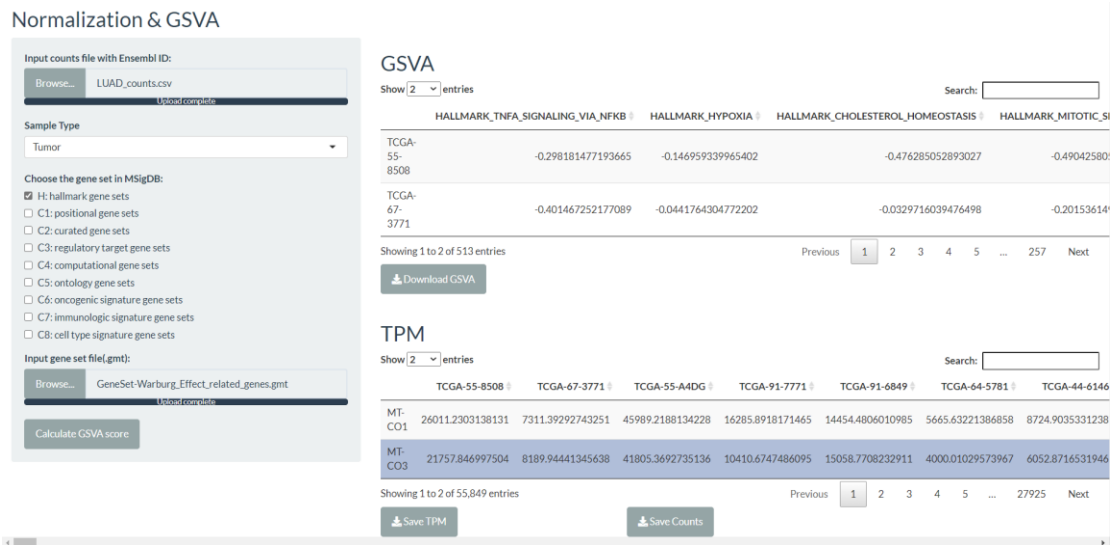


Figure 4. Normalization&GSVA

There are 9 gene sets from MSigDb in ARMT internal datasets available for direct selection to process GSVA, and it is also supported to input a .gmt file of arbitrary gene set from 'Data' page. If the input data is from TCGA, you can select normal, tumor or all samples to process normalize and GSVA (default: tumor).

6.3. Integration&Analysis

ARMT can integrate clinical, GSVA, gene expression (TPM), gene mutation data together, and use these data to carry out analysis including: survival analysis&cox proportional hazards regression analysis, differential analysis, enrichment analysis and correlation analysis.

6.3.1. Data Integration

The integrated data should contain the common samples. You should input the clinical, GSVA, TPM data by .csv file which can be produced by 'Data' and 'Integration&Analysis' page, and the mutation data should be entered by .maf file. The next analysis is carried out according to these integrated data. You can choose a column in integrated data table as the 'Group Column' to separate the data into multiple groups to analyze respectively. Also, you can input your own data in the same format.

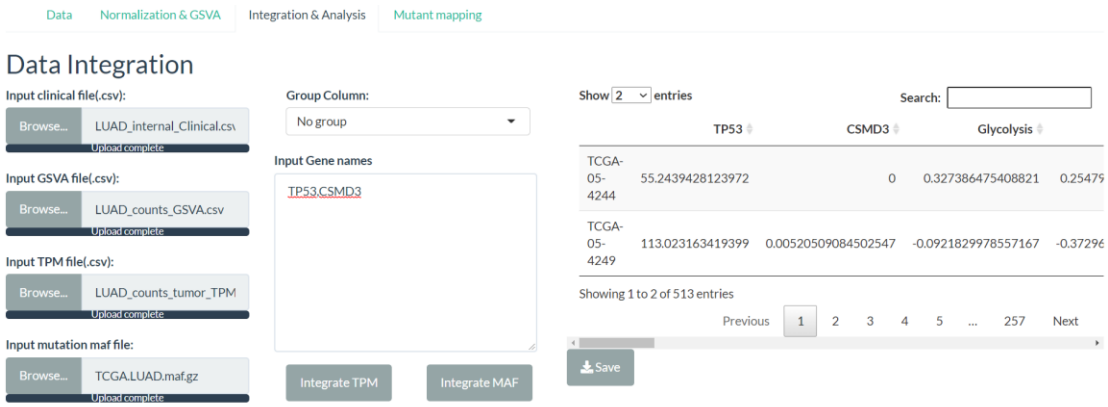


Figure 5. Data Integration

In this page, clinical data and GSVA data will be merged automatically, and you can enter your interested gene name in your TPM and mutation data to integrate their expression and mutation information by clicking 'Integrate TPM' and 'Integrate Maf' button. The integrated data is showed in main panel.

6.3.2. Survival analysis&Cox proportional hazards regression analysis

In this page, you should choose two columns in integrated data table as the survival time and status of samples. Each column of integrated data table can be seen as a factor of samples to carried out survival analysis and cox proportional hazards regression analysis.

Figure 6. Survival Analysis

If the survival factor is a continuous variable, the samples will be grouped into two parts (high&low). You can choose multiple factors to obtain multiple analysis results (K-M curve).

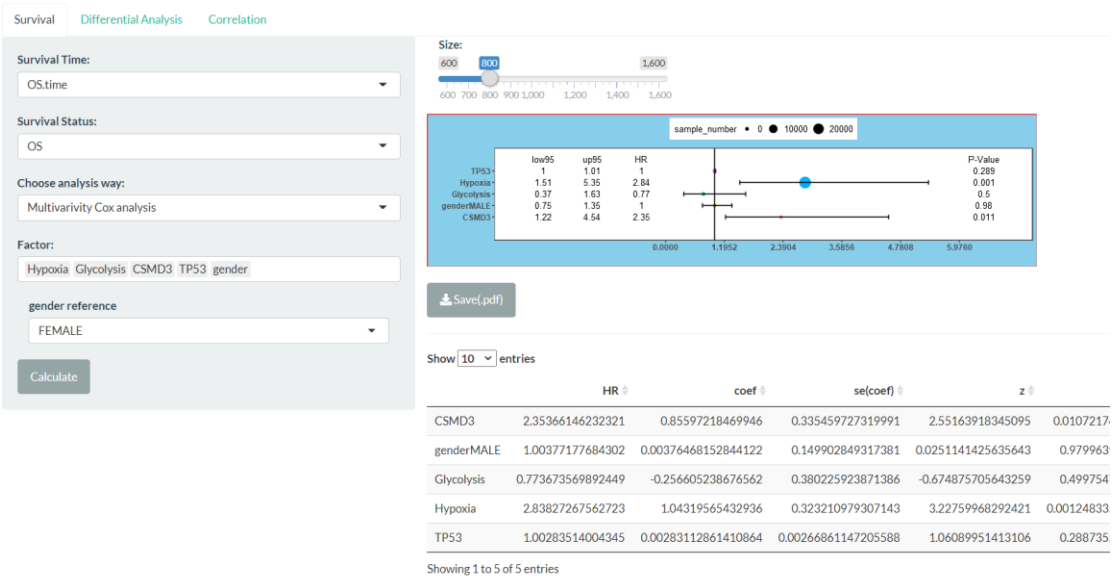


Figure 7. Cox Proportional Hazards Regression Analysis

The cox analysis can focus on single-variable or multiple-variable. If the factor in cox analysis is not numeric variable, one value of this variable should be set as a reference.

6.3.3 Differential analysis

ARMT can carry out differential analysis to GSVA score, gene mutation and expression. The difference factor should be selected in the integrated data, and the samples will be grouped by it. If the difference factor is a numeric variable, the samples will be grouped into high and low, and the low part will be set as control group. The group cut off value 't' is an adjustable parameter in ARMT(upper and lower t*100%). If the difference factor is not a numeric variable, you should select two values used to group the samples into experiment group and control group. You can filter the result by p-value, logFC and FDR (adj.p) after analysis and visualize it in main panel.

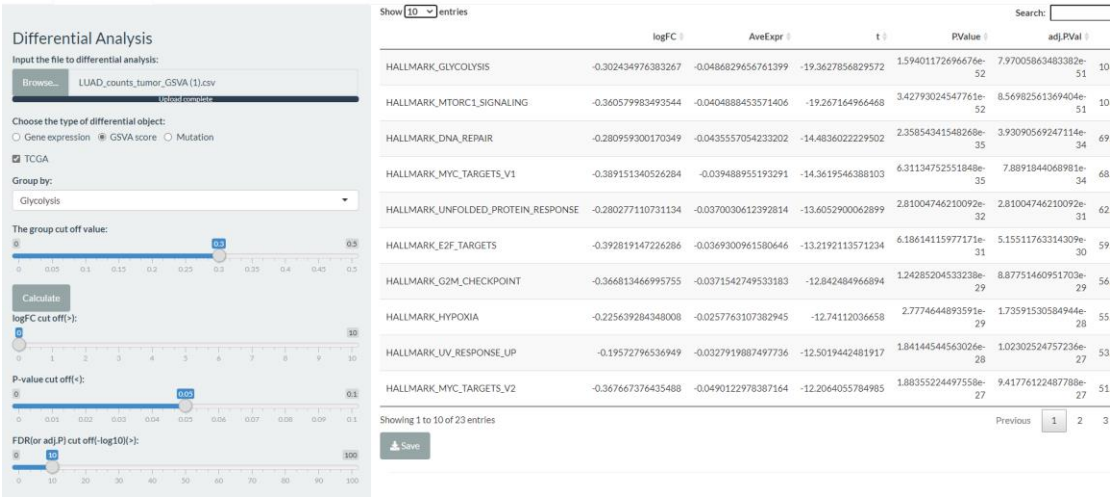


Figure 8. The differential analysis of GSVA score

To analyze difference of GSVA score, the GSVA result of samples in integrated data should be entered through .csv file, and you can get this file in 'Normalization&GSVA' page.

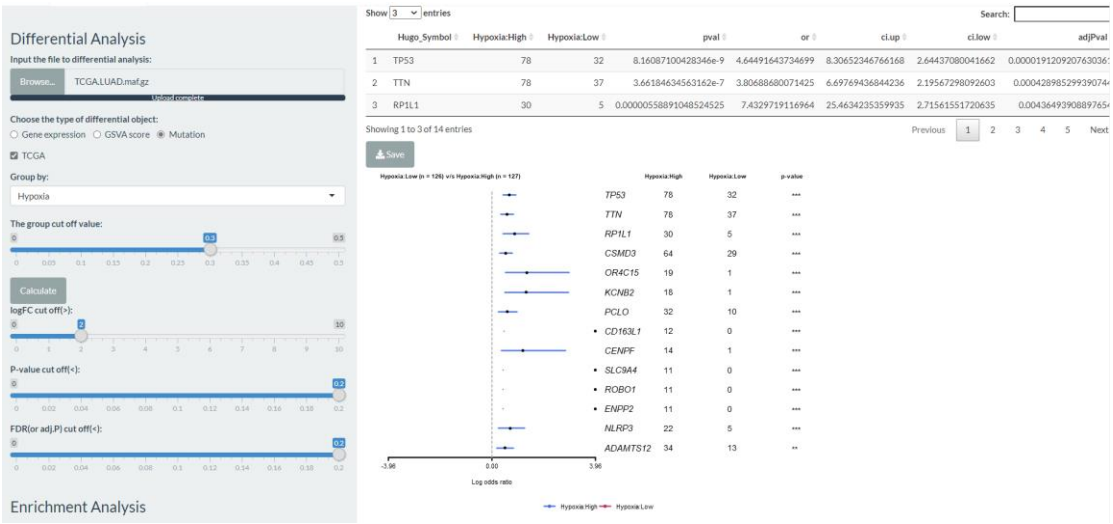


Figure 9. The differential analysis of gene mutation

To analyze difference of gene mutation, the mutation information of samples in integrated data should be entered through .maf file. The result is demonstrated with forest plot in main panel.

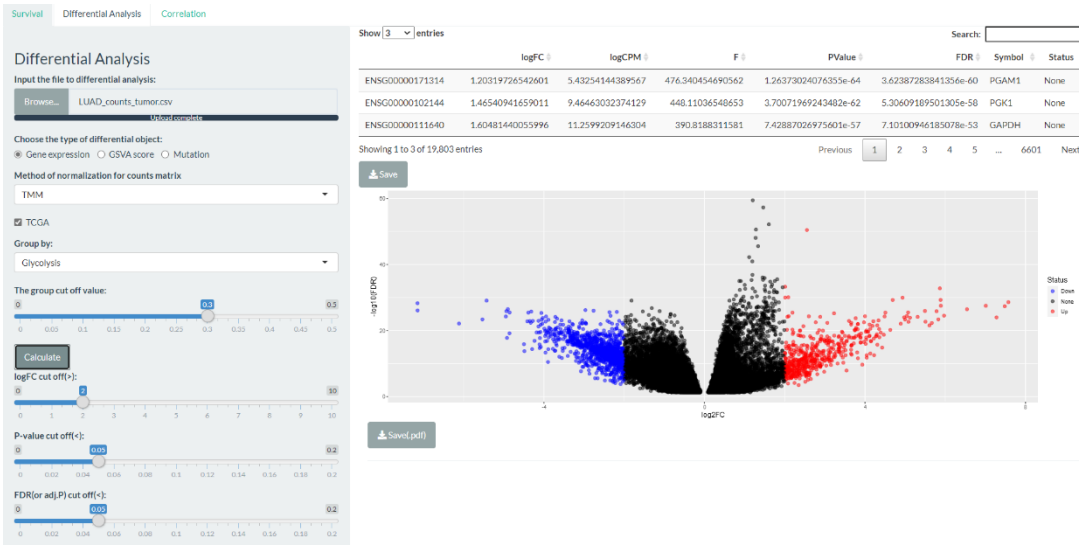


Figure 10. The differential analysis of gene expression

To analyze difference of gene expression (DEG), the counts matrix of samples in integraated data should be entered through .csv file. ARMT provides four methods to normalize counts matrix: TMM, TMMwsp, RLE, upperquartile. The result is demonstrated with volcano plot in main panel.

6.3.4. Enrichment analysis

The enrichment analysis ways in ARMT include GO and KEGG. It can be processed for differential expression genes (DEG) from above differential analysis or for arbitrary gene list input by user.

Figure 11. Enrichment Analysis

You should set up the p-value and q-value cut off before enrichment analysis, and you can select interested ontology (MF, CC, BP) of GO enrichment. The enriched genes are up

regulation and down regulation in DEG, and their enrichment results (up regulation, down regulation and all DEG) are showed in different pages of main panel. These results are demonstrated with bar plot and dot plot.

6.3.5. Correlation analysis

ARMT can calculate the correlation between any continuous variable factors of integration data, such as TPM and GSVA score. The result is demonstrated in main panel with correlation coefficient matrix and heat map, and it can be filtered by p-value and correlation coefficient (r, Spearman or Pearson).

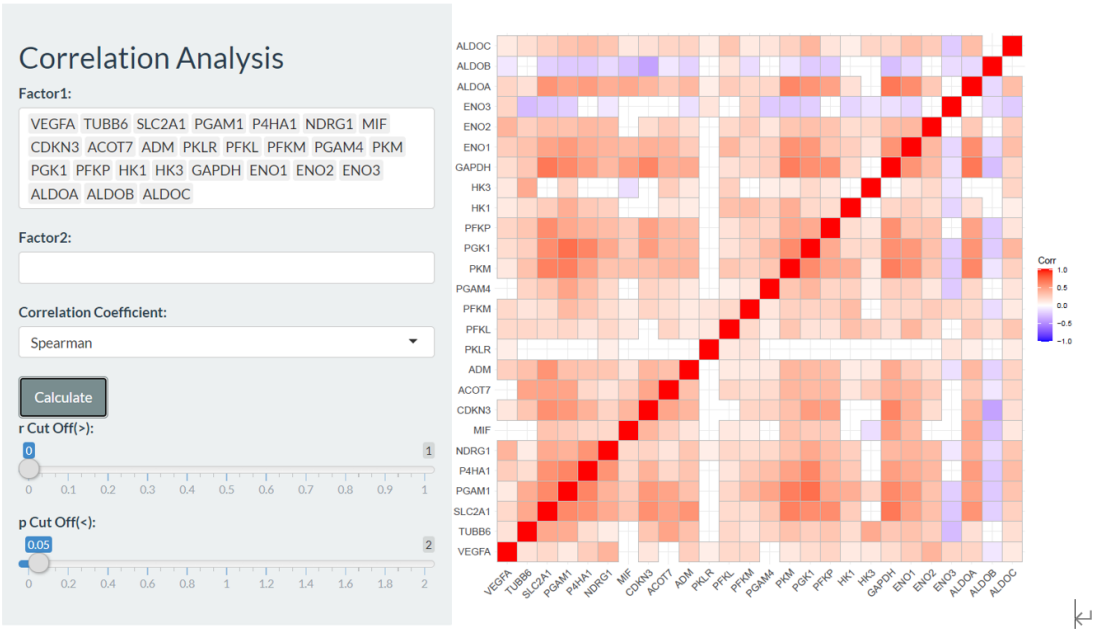


Figure 12. ARMT can calculate correlation coefficient of each pair-factors.

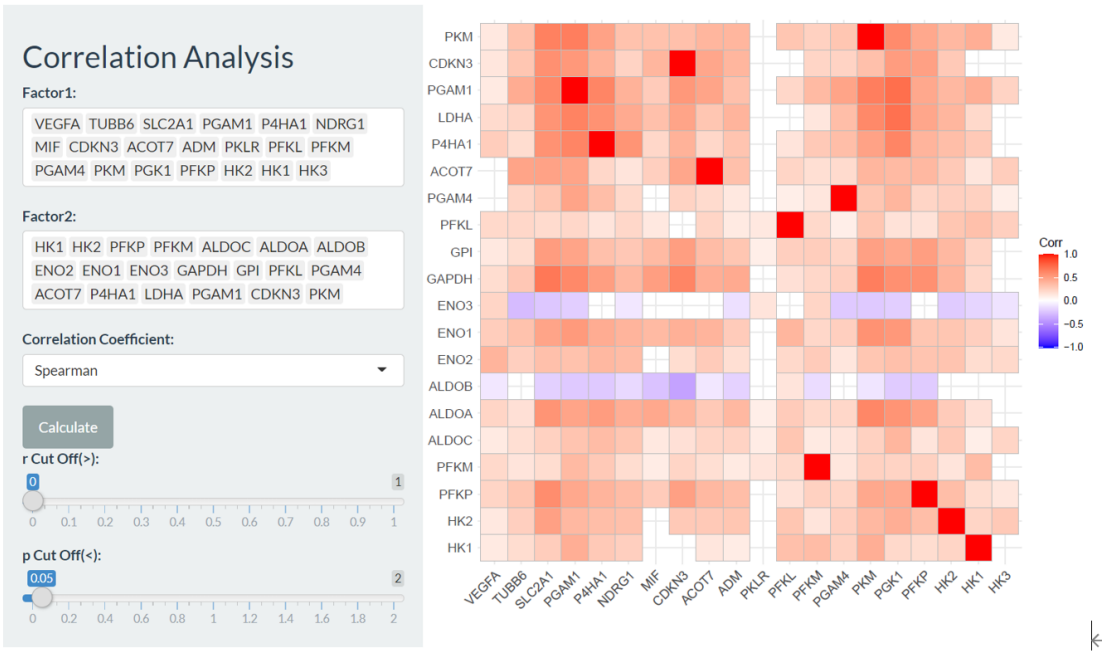


Figure 13. ARMT can calculate correlation coefficient between two list of factors.

If the integrated data is grouped by 'Group Column', ARMT can calculate correlation in each group of data by using single factor or all factors in the list 'Factor1'

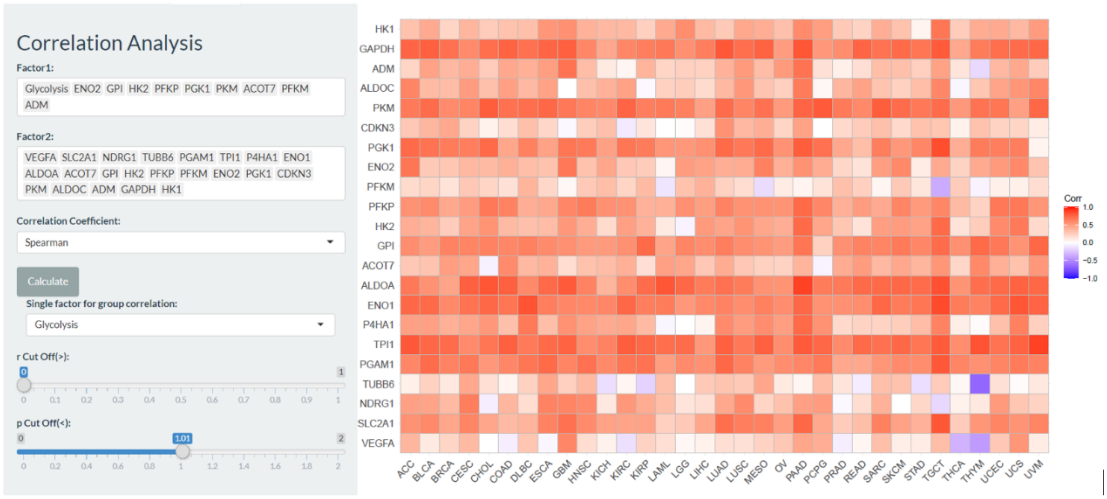


Figure 14. For each group of data, the correlation coefficient between single factor and 'Factor2' list is put together in one result, and the single factor name is replaced by group name in correlation matrix. It is a case of pan-cancer correlation analysis.

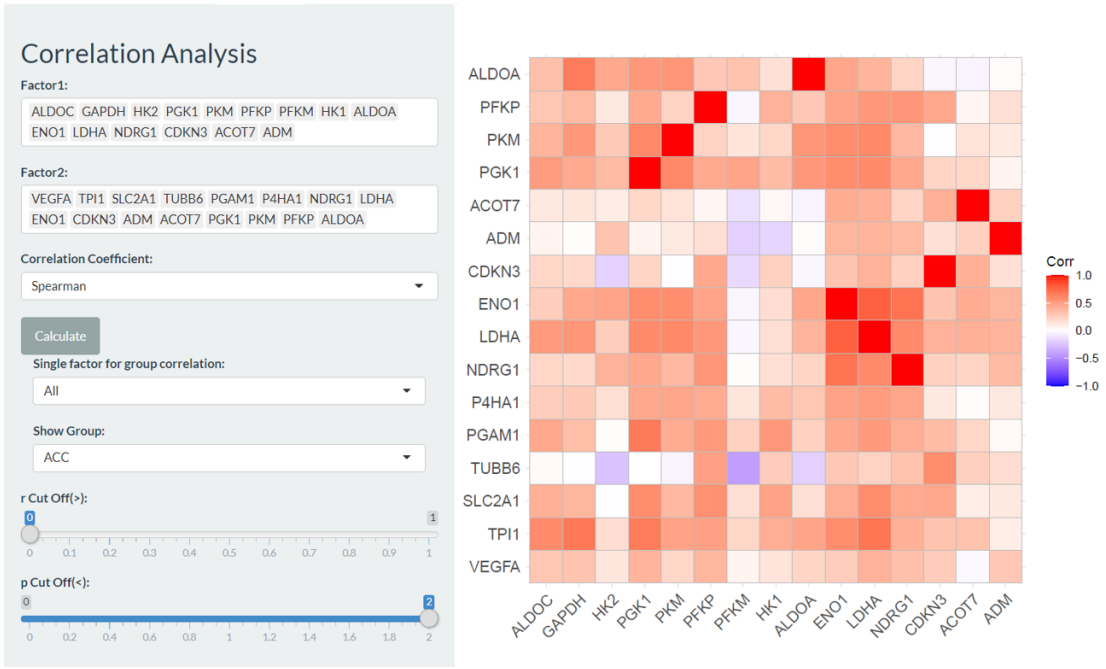


Figure 15. When all factors in 'Factor1' is demonstrated, you can choose one result of multiple groups in sidebar panel to show in main panel.

6.4. Mutant mapping

In corporation with 'maftools', ARMT can visualize the gene mutation in .maf file. There are two plot modes: Maftools Summary and Self-defined. The number of top mutant genes plot out can be set in sidebar panel.

Figure 16. Maftools Summary

ARMT can plot the maf summary and illustrate the enrichment of known Oncogenic Signaling Pathways in TCGA cohorts. It is also supported to draw the oncoplot of interested pathway completely.

In Self-defined mode, ARMT can plot the map for any genes or gene sets, and the mutation types (deletion, insert, SNV) can be specified.

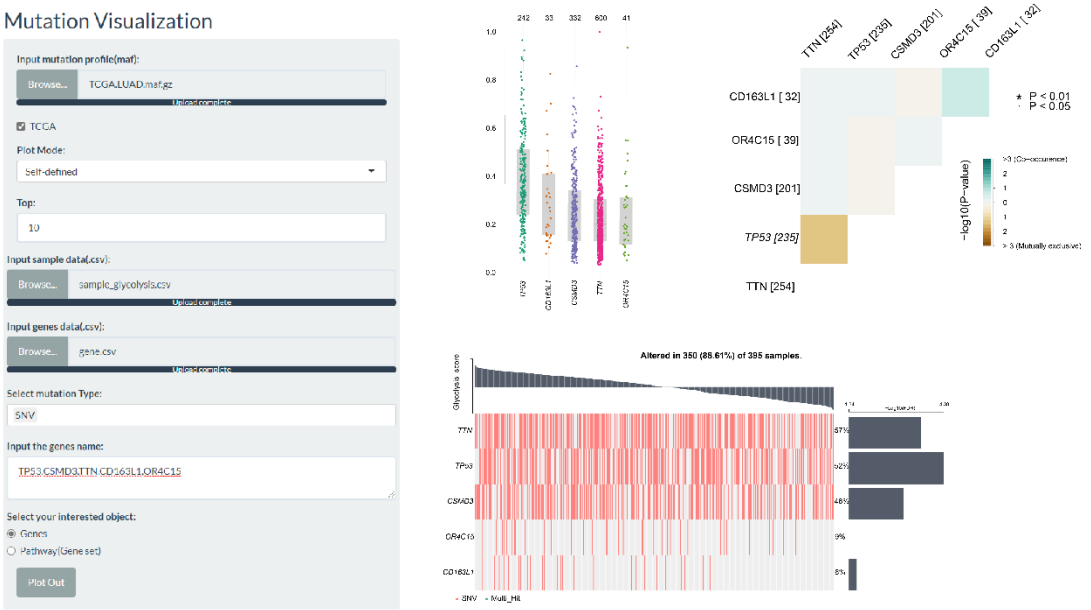


Figure 17. Mutant map plot for genes

About the specific genes, ARMT can plot out the VAF (Variant Allele Frequencies) as a boxplot, detect genes mutually exclusive or co-occurring, and produce oncoplot with any gene

information and sample data.

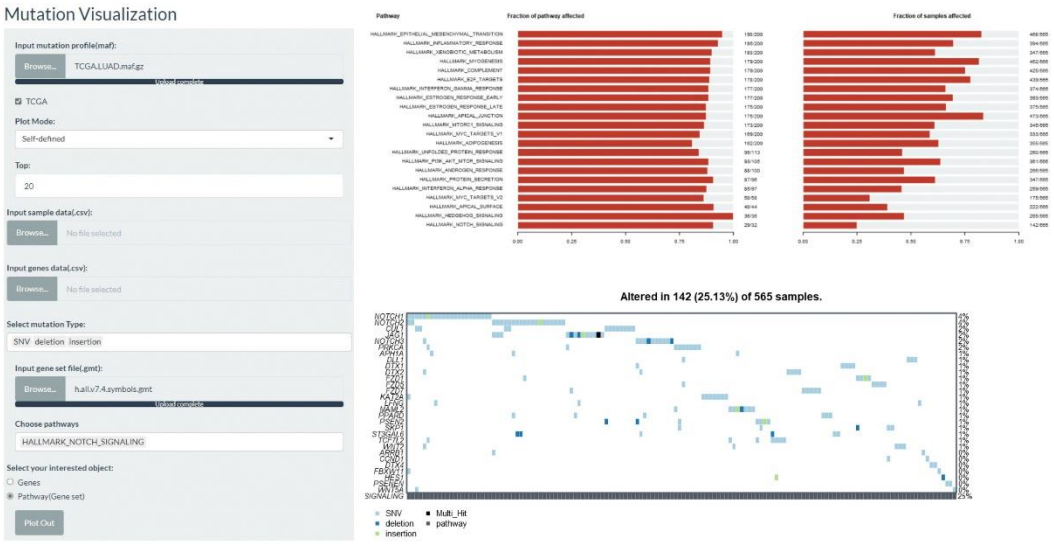


Figure 18. Mutant map plot for gene sets

Input the gmt file, ARMT shows the enrichment of mutation, and can plot a oncoplot of specific gene sets.

7. Format of Input File

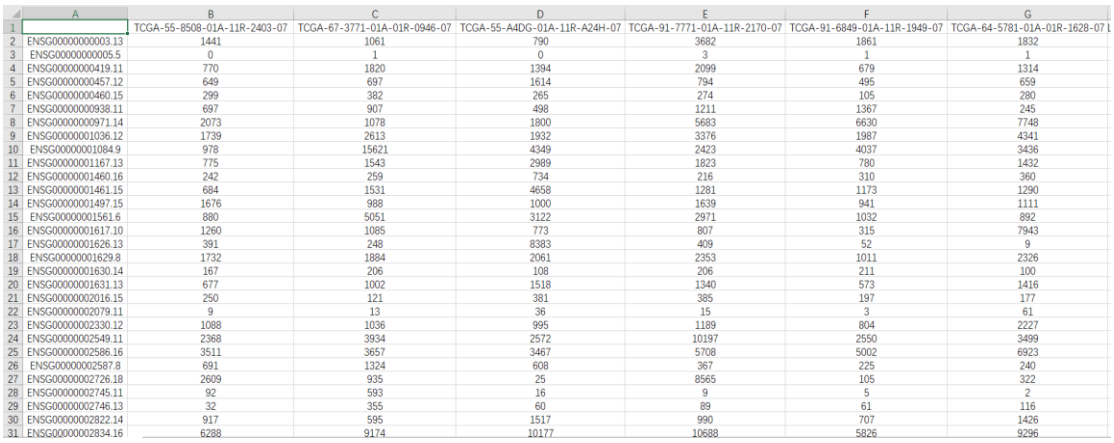
The input data of ARMT include gene sets, gene expression matrix, gene mutation, GSVA score matrix and clinical information.

7.1. Gene sets

In 'Data' page, the gene set file (.csv, Figure3) is used to create .gmt file, and the .gmt file is the main format of gene sets input.,

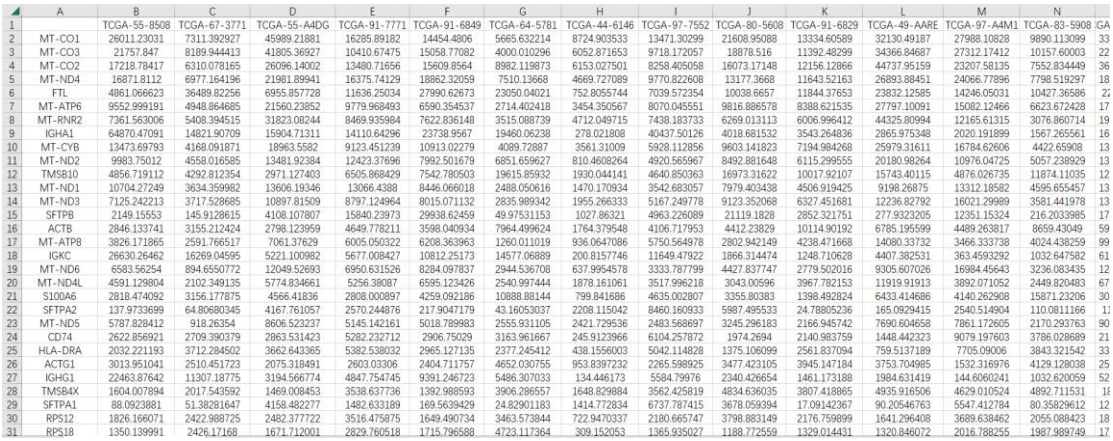
7.2. Gene expression matrix

For gene expression, the counts matrix can be normalized to TPM matrix in ARMT.



	A	B	C	D	E	F	G
1		TCGA-55-8508-01A-11R-2403-07	TCGA-67-3771-01A-01R-0946-07	TCGA-55-A4DG-01A-11R-A24H-07	TCGA-91-7771-01A-11R-2170-07	TCGA-91-6849-01A-11R-1949-07	TCGA-64-5781-01A-01R-1628-07
2	ENSG00000000003.13	1441	1061	790	3682	1861	1832
3	ENSG00000000005.5	0	1	0	3	1	1
4	ENSG00000000419.11	770	1820	1394	2099	679	1314
5	ENSG00000000457.12	649	697	1614	794	495	659
6	ENSG00000000460.15	299	382	265	274	105	280
7	ENSG00000000938.11	697	907	498	1211	1367	245
8	ENSG00000000971.14	2073	1078	1800	5683	6630	7748
9	ENSG00000001036.12	1739	2613	1932	3376	1987	4341
10	ENSG00000001084.9	978	15621	4349	2423	4037	3436
11	ENSG00000001167.13	775	1543	2989	1823	780	1432
12	ENSG00000001460.16	242	259	734	216	310	360
13	ENSG00000001461.15	684	1531	4658	1281	1173	1290
14	ENSG00000001497.15	1676	988	1000	1639	941	1111
15	ENSG00000001561.6	880	5051	3122	2971	1032	892
16	ENSG00000001617.10	1260	1085	773	807	315	7943
17	ENSG00000001626.13	391	248	8383	409	52	9
18	ENSG00000001629.8	1732	1884	2061	2353	1011	2326
19	ENSG00000001630.14	167	206	108	206	211	100
20	ENSG00000001631.13	677	1002	1518	1340	573	1416
21	ENSG00000002016.15	250	121	381	385	197	177
22	ENSG00000002079.11	9	13	36	15	3	61
23	ENSG00000002330.12	1088	1036	995	1189	804	2227
24	ENSG00000002549.11	2368	3934	2572	10197	2550	3499
25	ENSG00000002586.16	3511	3657	3467	5708	5002	6923
26	ENSG00000002587.8	691	1324	608	367	225	240
27	ENSG00000002726.18	2609	935	25	8565	105	322
28	ENSG00000002745.11	92	593	16	9	5	2
29	ENSG00000002746.13	32	355	60	89	61	116
30	ENSG00000002822.14	917	595	1517	990	707	1426
31	ENSG00000002834.16	6288	9174	10177	10688	5826	9296

Figure 19. Counts matrix must be in a .csv file. The row name is gene Ensembl ID and the column name is the sample ID. It is used to normalization and differential analysis.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		TCGA-55-8508	TCGA-67-3771	TCGA-55-A4DG	TCGA-91-7771	TCGA-91-6849	TCGA-64-5781	TCGA-44-6146	TCGA-97-7552	TCGA-80-5608	TCGA-91-6829	TCGA-49-AARE	TCGA-97-A4M1	TCGA-83-5908
2	MT-CO1	26011.23031	7311.392927	45989.21881	16285.89182	14454.4806	5665.632214	8724.903533	13471.30299	21608.95088	13334.60589	32130.49187	27988.10828	9890.113099
3	MT-CO3	21757.847	8189.944413	41805.36927	10410.67475	15058.77082	4000.010296	6052.871653	9718.172057	18878.516	11392.48299	34366.84687	27312.17412	10157.60003
4	MT-CO2	17218.78417	6310.078165	26096.14002	13480.71656	15609.8564	8982.119873	6153.027501	8258.405058	16073.17148	12156.12866	44737.95159	23207.58135	7552.834449
5	MT-ND4	16871.8112	6977.164196	21981.89941	16375.74129	18862.32059	7510.13668	4669.727089	9770.822608	13177.3668	11643.52163	26893.88451	24066.77896	7798.519297
6	FTL	4861.066623	36489.82256	6955.857728	11636.25034	27990.62673	23050.04021	752.8055744	7039.572354	10038.6657	11844.37653	23832.12585	14246.05031	10427.36586
7	MT-ATP6	9552.999191	4948.864685	21560.23852	9779.968493	6590.354537	2714.402418	3454.350567	8070.045551	9816.886578	8388.621535	27797.10091	15082.12466	6623.672428
8	MT-RNR2	7361.563006	5408.394515	31823.08244	8469.935984	7622.836148	3515.088739	4712.049715	7438.183733	6269.013113	6006.996412	44325.80994	12165.61315	3076.860714
9	IGHA1	64870.47091	14821.90709	15904.71311	14110.64296	23738.9567	19460.06238	278.021808	40437.50126	4018.681532	3543.264836	2865.975348	2020.191899	1567.265561
10	MT-CYB	13473.69793	4168.091871	18963.5582	9123.451239	10913.02279	4089.72887	3561.31009	5928.112856	9603.141823	7194.984268	25979.31611	16784.62606	4422.65908
11	MT-ND2	9983.75012	4558.016585	13481.92384	12423.37696	7992.501679	6851.659627	810.4608264	4920.565967	8492.881648	6115.299555	20180.98264	10976.04725	5057.238929
12	TMSB10	4856.719112	4292.812354	2971.127403	6505.868429	7542.780503	19615.85932	1930.044141	4640.850363	16973.31622	10017.92107	15743.40115	4876.026735	11874.11035
13	MT-ND1	10704.27249	3634.359982	13606.19346	13066.4388	8446.066018	2488.050616	1470.170934	3542.683057	7979.403438	4506.919425	9198.26875	13312.18582	4595.655457
14	MT-ND3	7125.242213	3717.528685	10897.81509	8797.124964	8015.071132	2835.989342	1955.266333	5167.249778	9123.352068	6327.451681	12236.82792	16021.29989	3581.441978
15	SFTPB	2149.15553	145.9128615	4108.107807	15840.23973	29938.62459	49.97531153	1027.86321	4963.226089	21119.1828	2852.321751	277.9323205	12351.15324	216.2033985
16	ACTB	2846.133741	3155.212424	2798.123959	4649.778211	3598.040934	7964.499624	1764.379548	4106.717953	4412.23829	10114.90192	6785.195599	4489.263817	8659.43049
17	MT-ATP8	3826.171865	2591.766517	7061.37629	6005.050322	6208.363963	1260.011019	936.0647086	5750.564978	2802.942149	4238.471668	14080.33732	3466.333738	4024.438259
18	IGKC	26630.26462	16269.04595	5221.100982	5677.008427	10812.25173	14577.06889	200.8157746	11649.47922	1866.314474	1248.710628	4407.382531	363.4593292	1032.647582
19	MT-ND6	6583.56254	894.6550772	12049.52693	6950.631526	8284.097837	2944.536708	637.9954578	3333.787799	4427.837747	2779.502016	9305.607026	16984.45643	3236.083435
20	MT-ND4L	4591.129804	2102.349135	5774.834661	5256.38087	6595.123426	2540.997444	1878.161061	3517.996218	3043.00596	3967.782153	11919.91913	3892.071052	2449.820483
21	S100A6	2818.474092	3156.177875	4566.41836	2808.000897	4259.092186	10888.88144	799.841686	4635.002807	3355.80383	1398.492824	6433.414686	4140.262908	15871.23206
22	SFTPA2	137.9733699	64.80680345	4167.761057	2570.244876	217.9047179	43.16053037	2208.115042	8460.160933	5987.495533	24.78805236	165.0929415	2540.514904	110.0811166
23	MT-ND5	5787.828412	918.26354	8606.523237	5145.142161	5018.789983	2555.931105	2421.729536	2483.568697	3245.296183	2166.945742	7690.604658	7861.172605	2170.293763
24	CD74	2622.856921	2709.390379	2863.531423	5282.232712	2906.75029	3163.961667	245.9123966	6104.257872	1974.2694	2140.983759	1448.442323	9079.197603	3786.028689
25	HLA-DRA	2032.221193	3712.284502	3662.643365	5382.538032	2965.127135	2377.245412	438.1556003	5042.114828	1375.106099	2561.837094	759.5137189	7705.09006	3843.321542
26	ACTG1	3013.951041	2510.451723	2075.318491	2603.03306	2404.711757	4652.030755	953.8397232	2265.598925	3477.423105	3945.147184	3753.704985	1532.316976	4129.128038
27	IGHG1	22463.87642	11307.18775	3194.566774	4847.754745	9391.246723	5486.307033	134.446173	5584.79976	2340.426654	1461.173188	1984.631419	144.6060241	1032.620059
28	TMSB4X	1604.007894	2017.543592	1469.008453	3538.637736	1392.988593	3906.286557	1648.829884	3562.425819	4834.636035	3807.418865	4935.916506	4629.010524	4892.711531
29	SFTPA1	88.0923881	51.38281647	4158.482277	1482.633189	169.5639429	24.82901183	1414.772834	6737.787415	3678.059394	17.09142367	90.20546763	5547.412784	80.35829612
30	RPS12	1826.166071	2422.988725	2482.377722	3516.475875	1649.490734	3463.573844	722.9470337	2180.665747	3798.883149	2176.759899	1641.296408	3689.638462	2055.088423
31	RPS18	1350.139991	2426.17168	1671.712001	2829.760518	1715.796588	4723.117364	309.152053	1365.935027	1188.772559	1329.014431	1320.846072	2016.788255	1987.989749

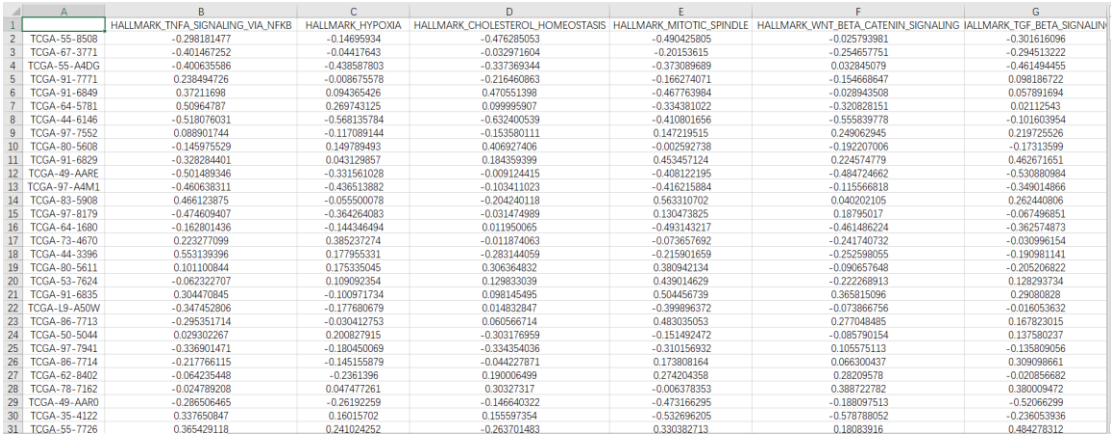
Figure 20. TPM matrix must be in a .csv file. The row name is gene symbol ID and the column name is the sample ID. It can be integrated with other data.

7.3. Gene mutation

To get the mutation information, ARMT requires standard .maf file like the mutation annotation format file in TCGA.

7.4. GSVA score matrix

The GSVA score matrix can be obtained in 'Normalization&GSVA' page.

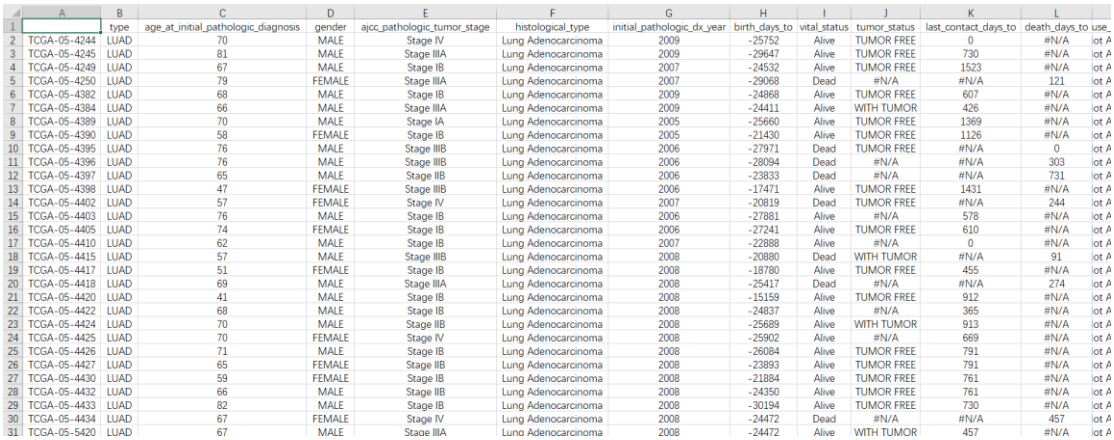


	A	B	C	D	E	F	G
1		HALLMARK_TNFA_SIGNALING_VIA_NFKB	HALLMARK_HYPOXIA	HALLMARK_CHOLESTEROL_HOMEOSTASIS	HALLMARK_MITOTIC_SPINDLE	HALLMARK_WNT_BETA_CATENIN_SIGNALING	HALLMARK_TGF_BETA_SIGNALIN
2	TCGA-55-8508	-0.298181477	-0.14695934	-0.476285053	-0.490425805	-0.025793981	-0.301616096
3	TCGA-67-3771	-0.401467252	-0.04417643	-0.032971604	-0.20153615	-0.254657751	-0.294513222
4	TCGA-55-A4DG	-0.400635586	-0.438587803	-0.337369344	-0.373089689	0.032845079	-0.461494455
5	TCGA-91-7771	0.238494726	-0.008675578	-0.216460863	-0.166274071	-0.154668647	0.098186722
6	TCGA-91-6849	0.37211698	0.094365426	0.470551398	-0.467763984	-0.028943508	0.057891694
7	TCGA-64-5781	0.50964787	0.269743125	0.099995907	-0.334381022	-0.320828151	0.02112543
8	TCGA-44-6146	-0.518076031	-0.568135784	-0.632400539	-0.410801656	-0.555839778	-0.101603954
9	TCGA-97-7552	0.088901744	-0.117089144	-0.153580111	0.147219515	0.249062945	0.219725526
10	TCGA-80-5608	-0.145975529	0.149789493	0.406927406	-0.002592738	-0.192207006	-0.17313599
11	TCGA-91-6829	-0.328284401	0.043129857	0.184359399	0.453457124	0.224574779	0.462671651
12	TCGA-49-AARE	-0.501489346	-0.331561028	-0.009124415	-0.408122195	-0.484724662	-0.530880984
13	TCGA-97-A4M1	-0.460638311	-0.436513882	-0.103411023	-0.416215884	-0.115566818	-0.349014866
14	TCGA-83-5908	0.466123875	-0.055500078	-0.204240118	0.563310702	0.040202105	0.262440806
15	TCGA-97-8179	-0.474609407	-0.364264083	-0.031479989	0.130473825	0.18795017	-0.067496851
16	TCGA-64-1680	-0.162801436	-0.144346494	0.011950065	-0.493143217	-0.461486224	-0.362574873
17	TCGA-73-4670	0.223277099	0.385237274	-0.011874063	-0.073657692	-0.241740732	-0.030996154
18	TCGA-44-3396	0.553139396	0.177955331	-0.283144059	-0.215901659	-0.252598055	-0.190981141
19	TCGA-80-5611	0.101100844	0.175335045	0.306364832	0.380942134	-0.090657648	-0.205206822
20	TCGA-53-7624	-0.062322707	0.109092354	0.129833039	0.439014629	-0.222268913	0.128293734
21	TCGA-91-6835	0.304470845	-0.100971734	0.098145495	0.504456739	0.365815096	0.29080828
22	TCGA-L9-A50W	-0.347452806	-0.177680679	0.014832847	-0.399896372	-0.073866756	-0.016053632
23	TCGA-86-7713	-0.295351714	-0.030412753	0.060566714	0.483035053	0.277048485	0.167823015
24	TCGA-50-5044	0.029302267	0.200827915	-0.303176959	-0.151492472	-0.085790154	0.137580237
25	TCGA-97-7941	-0.336901471	-0.180450069	-0.334354036	-0.310156932	0.105575113	-0.135809056
26	TCGA-86-7714	-0.217766115	-0.145155879	-0.044227871	0.173808164	0.066300437	0.309098661
27	TCGA-62-8402	-0.064235448	-0.2361396	0.190006499	0.274204358	0.28209578	-0.020856682
28	TCGA-78-7162	-0.024789208	0.047477261	0.30327317	-0.006378353	0.388722782	0.380009472
29	TCGA-49-AAR0	-0.286506465	-0.26192259	-0.146640322	-0.473166295	-0.188097513	-0.52066299
30	TCGA-35-4122	0.337650847	0.16015702	0.155597354	-0.532696205	-0.578788052	-0.236053936
31	TCGA-55-7726	0.365429118	0.241024252	-0.263701483	0.330382713	0.18083916	0.484278312

Figure 21. The GSVA score is saved in a matrix of .csv file. The row name is the sample ID, and the column name is gene set name.

7.5. Clinical information

The TCGA clinical information can be obtained in 'Data' page.



	A	B	C	D	E	F	G	H	I	J	K	L	
1		type	age_at_initial_pathologic_diagnosis	gender	ajcc_pathologic_tumor_stage	histological_type	initial_pathologic_dx_year	birth_days_to	vital_status	tumor_status	last_contact_days_to	death_days_to	use_
2	TCGA-05-4244	LUAD	70	MALE	Stage IV	Lung Adenocarcinoma	2009	-25752	Alive	TUMOR FREE	0	#N/A	ot A
3	TCGA-05-4245	LUAD	81	MALE	Stage IIIA	Lung Adenocarcinoma	2009	-29647	Alive	TUMOR FREE	730	#N/A	ot A
4	TCGA-05-4249	LUAD	67	MALE	Stage IB	Lung Adenocarcinoma	2007	-24532	Alive	TUMOR FREE	1523	#N/A	ot A
5	TCGA-05-4250	LUAD	79	FEMALE	Stage IIIA	Lung Adenocarcinoma	2007	-29068	Dead	#N/A	#N/A	121	ot A
6	TCGA-05-4382	LUAD	68	MALE	Stage IB	Lung Adenocarcinoma	2009	-24868	Alive	TUMOR FREE	607	#N/A	ot A
7	TCGA-05-4384	LUAD	66	MALE	Stage IIIA	Lung Adenocarcinoma	2009	-24411	Alive	WITH TUMOR	426	#N/A	ot A
8	TCGA-05-4389	LUAD	70	MALE	Stage IA	Lung Adenocarcinoma	2005	-25660	Alive	TUMOR FREE	1369	#N/A	ot A
9	TCGA-05-4390	LUAD	58	FEMALE	Stage IB	Lung Adenocarcinoma	2005	-21430	Alive	TUMOR FREE	1126	#N/A	ot A
10	TCGA-05-4395	LUAD	76	MALE	Stage IIIB	Lung Adenocarcinoma	2006	-27971	Dead	TUMOR FREE	#N/A	0	ot A
11	TCGA-05-4396	LUAD	76	MALE	Stage IIIB	Lung Adenocarcinoma	2006	-28094	Dead	#N/A	#N/A	303	ot A
12	TCGA-05-4397	LUAD	65	MALE	Stage IIB	Lung Adenocarcinoma	2006	-23833	Dead	#N/A	#N/A	731	ot A
13	TCGA-05-4398	LUAD	47	FEMALE	Stage IIIB	Lung Adenocarcinoma	2006	-17471	Alive	TUMOR FREE	1431	#N/A	ot A
14	TCGA-05-4402	LUAD	57	FEMALE	Stage IV	Lung Adenocarcinoma	2007	-20819	Dead	TUMOR FREE	#N/A	244	ot A
15	TCGA-05-4403	LUAD	76	MALE	Stage IB	Lung Adenocarcinoma	2006	-27881	Alive	#N/A	578	#N/A	ot A
16	TCGA-05-4405	LUAD	74	FEMALE	Stage IB	Lung Adenocarcinoma	2006	-27241	Alive	TUMOR FREE	610	#N/A	ot A
17	TCGA-05-4410	LUAD	62	MALE	Stage IB	Lung Adenocarcinoma	2007	-22888	Alive	#N/A	0	#N/A	ot A
18	TCGA-05-4415	LUAD	57	MALE	Stage IIIB	Lung Adenocarcinoma	2008	-20880	Dead	WITH TUMOR	#N/A	91	ot A
19	TCGA-05-4417	LUAD	51	FEMALE	Stage IB	Lung Adenocarcinoma	2008	-18780	Alive	TUMOR FREE	455	#N/A	ot A
20	TCGA-05-4418	LUAD	69	MALE	Stage IIIA	Lung Adenocarcinoma	2008	-25417	Dead	#N/A	#N/A	274	ot A
21	TCGA-05-4420	LUAD	41	MALE	Stage IB	Lung Adenocarcinoma	2008	-15159	Alive	TUMOR FREE	912	#N/A	ot A
22	TCGA-05-4422	LUAD	68	MALE	Stage IB	Lung Adenocarcinoma	2008	-24837	Alive	#N/A	365	#N/A	ot A
23	TCGA-05-4424	LUAD	70	MALE	Stage IIB	Lung Adenocarcinoma	2008	-25689	Alive	WITH TUMOR	913	#N/A	ot A
24	TCGA-05-4425	LUAD	70	FEMALE	Stage IV	Lung Adenocarcinoma	2008	-25902	Alive	#N/A	669	#N/A	ot A
25	TCGA-05-4426	LUAD	71	MALE	Stage IB	Lung Adenocarcinoma	2008	-26084	Alive	TUMOR FREE	791	#N/A	ot A
26	TCGA-05-4427	LUAD	65	FEMALE	Stage IIB	Lung Adenocarcinoma	2008	-23893	Alive	TUMOR FREE	791	#N/A	ot A
27	TCGA-05-4430	LUAD	59	FEMALE	Stage IB	Lung Adenocarcinoma	2008	-21884	Alive	TUMOR FREE	761	#N/A	ot A
28	TCGA-05-4432	LUAD	66	MALE	Stage IIB	Lung Adenocarcinoma	2008	-24350	Alive	TUMOR FREE	761	#N/A	ot A
29	TCGA-05-4433	LUAD	82	MALE	Stage IB	Lung Adenocarcinoma	2008	-30194	Alive	TUMOR FREE	730	#N/A	ot A
30	TCGA-05-4434	LUAD	67	FEMALE	Stage IV	Lung Adenocarcinoma	2008	-24472	Dead	#N/A	#N/A	457	ot A
31	TCGA-05-5420	LUAD	67	MALE	Stage IIIA	Lung Adenocarcinoma	2008	-24472	Alive	WITH TUMOR	457	#N/A	ot A

Figure 22. The clinical data must be in a .csv file, and the row name is sample ID and column name is sample characteristic. Any information about samples in this format file can be input as clinical data.

The example data has been uploaded to https://github.com/Dulab2020/ARMT.