CORALE User Manual

Comprehensive Oral Cancer Explorer (CORALE):
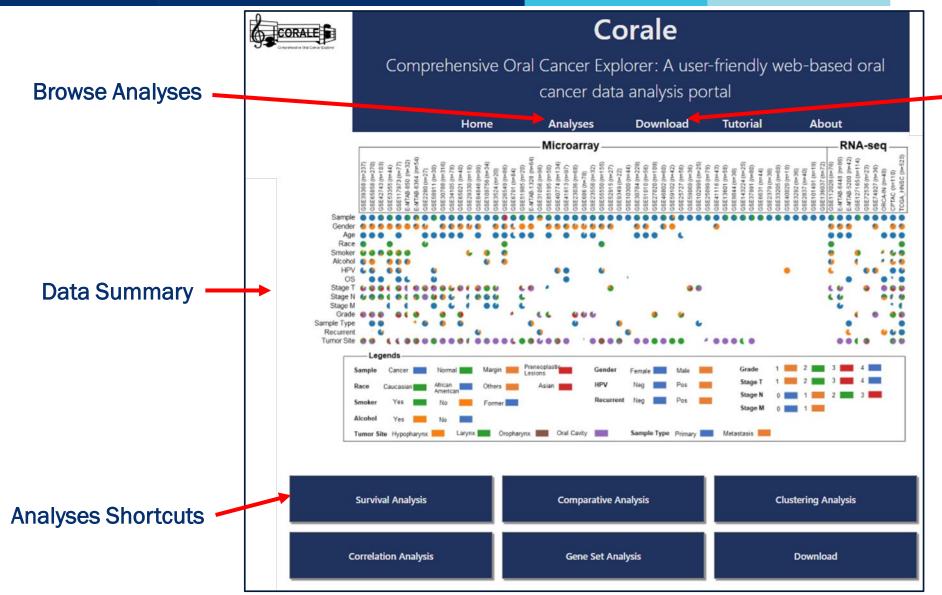A user-friendly web-based oral cancer data analysis portal

Introduction

- CORALE is a useful resource to perform secondary-analyses in harmonized multi-omics oral cancer data sets.

- CORALE is publicly available at http://corale.moffitt.org

- CORALE provides
  - Survival Analysis
  - Comparative Analysis
  - Clustering Analysis
  - Correlation Analysis
  - Gene Set Enrichment Analysis
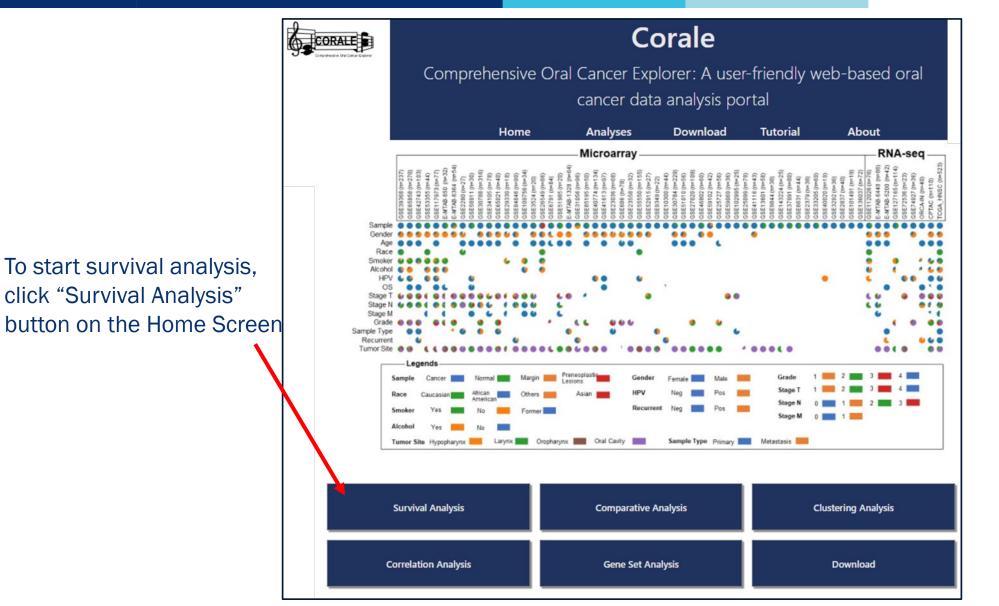 - Harmonized Dataset Download

CORALE Home Page

Browse Analyses

Download CORALE Datasets

Data Summary

Analyses Shortcuts

Survival Analysis

Survival Analysis

To start survival analysis, click "Survival Analysis" button on the Home Screen

Survival Analysis – Data selection

You will see this Survival Analysis screen.

Select any dataset that you want to use for Survival Analysis

Survival Analysis

Perform overall survival analysis of a selected gene or tumor immune cell types in a selected cohort.
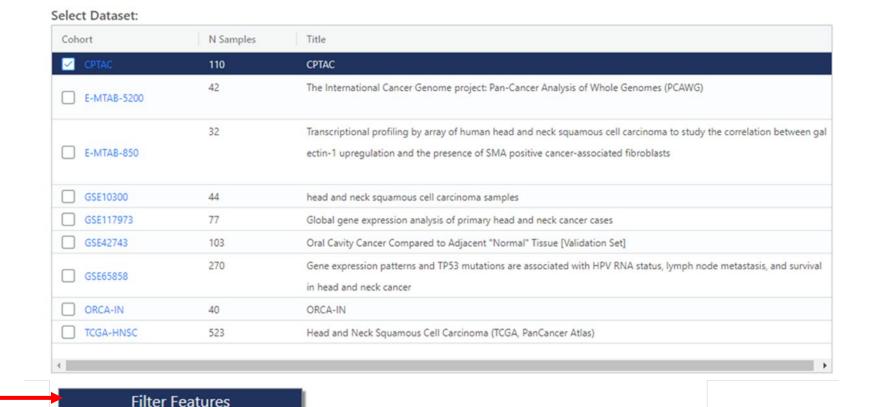
Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)
☐ E-MTAB-850	32	Transcriptional profiling by array of human head and neck squamous cell carcinoma to study the correlation between galectin-1 upregulation and the presence of SMA positive cancer-associated fibroblasts
☐ GSE10300	44	head and neck squamous cell carcinoma samples
☐ GSE117973	77	Global gene expression analysis of primary head and neck cancer cases
☐ GSE42743	103	Oral Cavity Cancer Compared to Adjacent "Normal" Tissue [Validation Set]
☐ GSE65858	270	Gene expression patterns and TP53 mutations are associated with HPV RNA status, lymph node metastasis, and survival in head and neck cancer
☐ ORCA-IN	40	ORCA-IN
☐ TCGA-HNSC	523	Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)

Survival Analysis – Filter Features (optional)

To select subset of samples in the dataset, scroll down the screen and click "Filter Features" to open features list.

Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)
☐ E-MTAB-850	32	Transcriptional profiling by array of human head and neck squamous cell carcinoma to study the correlation between gal ectin-1 upregulation and the presence of SMA positive cancer-associated fibroblasts
☐ GSE10300	44	head and neck squamous cell carcinoma samples
☐ GSE117973	77	Global gene expression analysis of primary head and neck cancer cases
☐ GSE42743	103	Oral Cavity Cancer Compared to Adjacent "Normal" Tissue [Validation Set]
☐ GSE65858	270	Gene expression patterns and TP53 mutations are associated with HPV RNA status, lymph node metastasis, and survival in head and neck cancer
☐ ORCA-IN	40	ORCA-IN
☐ TCGA-HNSC	523	Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)

Filter Features

Grouping Features:

Clinical Features: *Expand Selections* ⌄
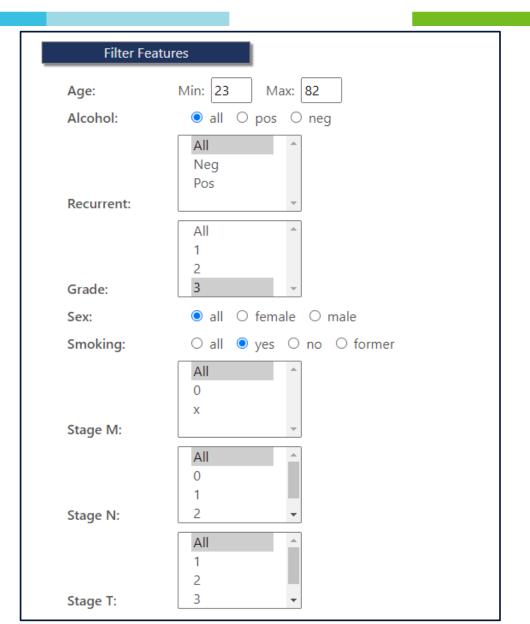
7

Survival Analysis – Filter Features (optional)

Choose sample selection criteria to use the subset of the samples in the dataset

This is an example of selecting only the samples with
1) Samples with the patients of 50 or older
and
2) Samples with Grade 3
and
4) Samples from Current smokers

Survival Analysis – Grouping Features

For the comparative survival analysis, grouping the samples into groups is necessary. Select any feature you want to use for separating the samples into groups.
In this example, "Sex" is selected as the grouping feature, and the survival analysis will be performed for Male and Female group

Grouping Features:

Clinical Features: *Collapse Selections* ⌃

	Yes	No
Alcohol:	○	●
Recurrent:	○	●
Grade:	○	●
Sex:	●	○
Smoking:	○	●
Stage M:	○	●
Stage N:	○	●
Stage T:	○	●

- Or -

Select Gene: [   ]

- Or -

TME Method: [Choose TME method ▼]

[Submit]

Survival Analysis – Grouping Features

Instead of selecting a clinical feature, you can also input a gene name or select a TME method.

In this case, the samples in the dataset will be divided into high/low group based on the value of the selected feature (Gene or TME), using the median value as the cut point.

Grouping Features:

Clinical Features: *Collapse Selections* ⌃

Alcohol:	○ Yes	◉ No
Recurrent:	○ Yes	◉ No
Grade:	○ Yes	◉ No
Sex:	○ Yes	◉ No
Smoking:	○ Yes	◉ No
Stage M:	○ Yes	◉ No
Stage N:	○ Yes	◉ No
Stage T:	○ Yes	◉ No

- Or -

Select Gene: EGFR

EGFR
EGFR-AS1

- Or -

TME Method: Choose TME method ⌄

Some of the samples might be removed because the essential values for this analysis are missing.

Submit

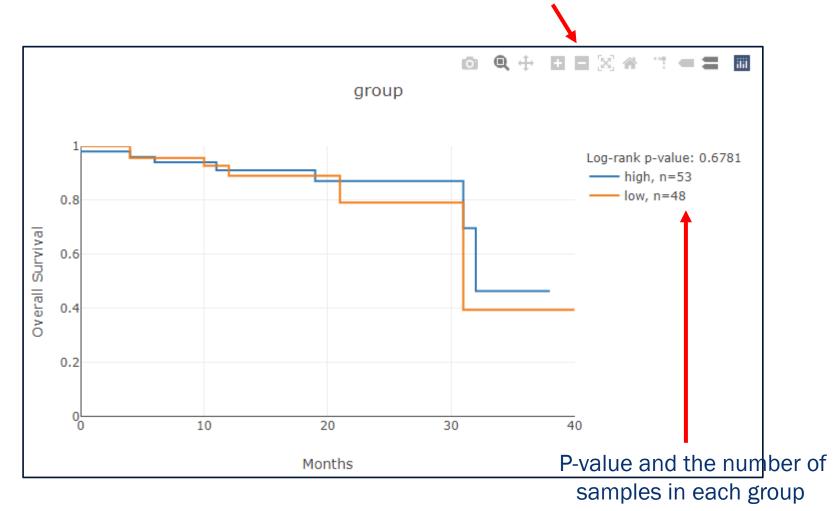# Survival Analysis – Grouping Features

When you are done selecting the features, click "Submit" to run Survival Analysis

Survival Analysis – Results

Image control panel (e.g., save the image, zoom in/out)

You will see the survival analysis plot on the bottom of the page in several seconds.

P-value and the number of samples in each group

In the Survival Analysis, some of the samples might be removed because the essential values for this analysis are missing.

Comparative Analysis

Comparative Analysis

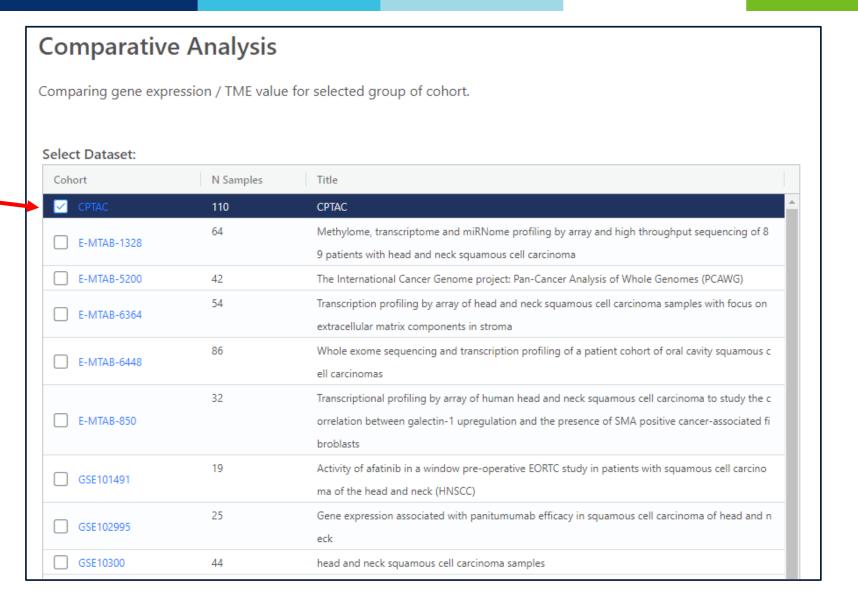To start comparative analysis, click "Comparative Analysis" button on the Home Screen
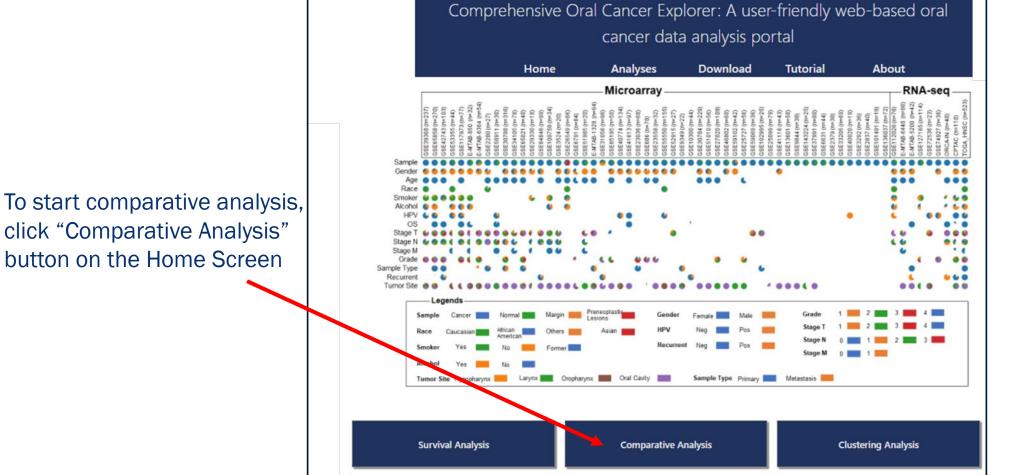
Comparative Analysis – Data selection

You will see this Comparative Analysis screen.

Select any dataset that you want to use for Comparative Analysis

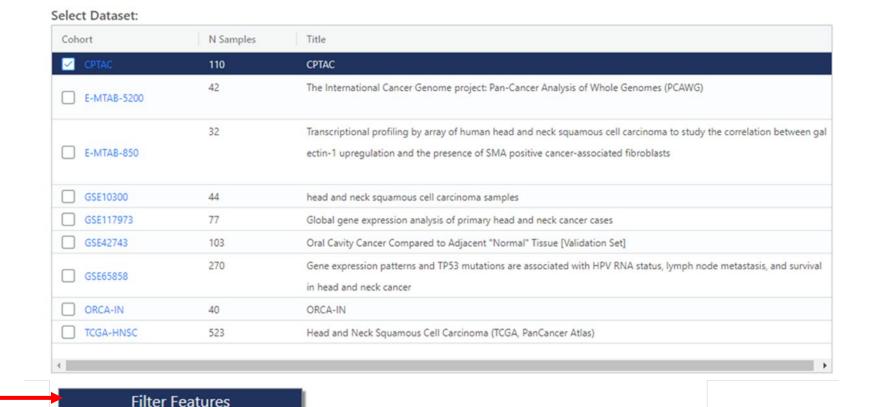# Comparative Analysis – Filter Features (optional)

Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)
☐ E-MTAB-850	32	Transcriptional profiling by array of human head and neck squamous cell carcinoma to study the correlation between gal ectin-1 upregulation and the presence of SMA positive cancer-associated fibroblasts
☐ GSE10300	44	head and neck squamous cell carcinoma samples
☐ GSE117973	77	Global gene expression analysis of primary head and neck cancer cases
☐ GSE42743	103	Oral Cavity Cancer Compared to Adjacent "Normal" Tissue [Validation Set]
☐ GSE65858	270	Gene expression patterns and TP53 mutations are associated with HPV RNA status, lymph node metastasis, and survival in head and neck cancer
☐ ORCA-IN	40	ORCA-IN
☐ TCGA-HNSC	523	Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)

To select subset of samples in the dataset, scroll down the screen and click "Filter Features" to open features list.

Filter Features

Grouping Features:

Clinical Features: *Expand Selections* ⌄

Comparative Analysis – Filter Features (optional)

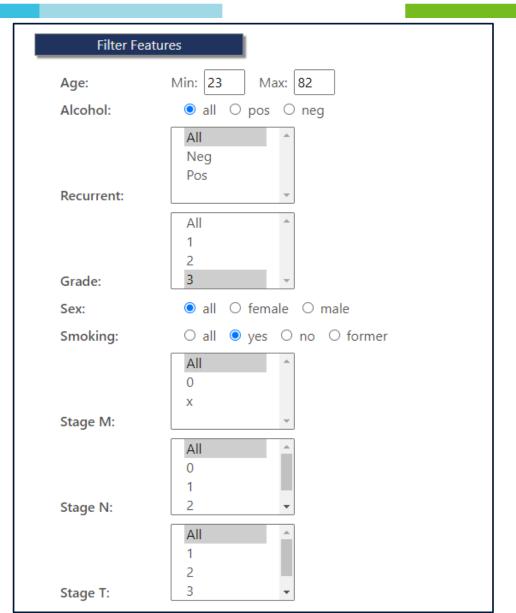Choose sample selection criteria to use the subset
of the samples in the dataset

This is an example of selecting only the samples
with
1) Samples with the patients of 50 or older
and
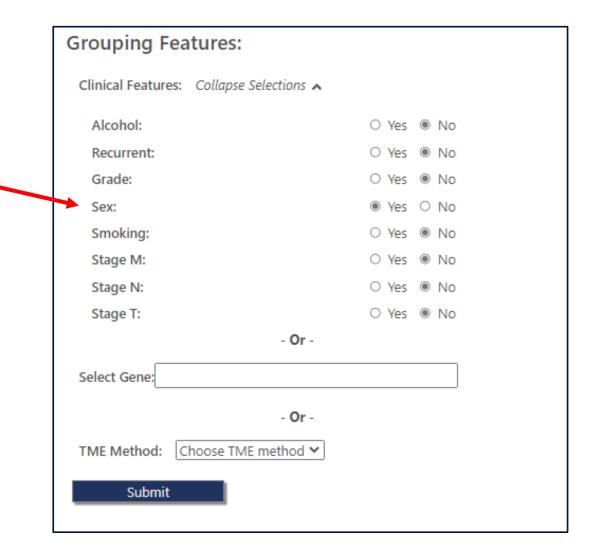2) Samples with Grade 3
and
4) Samples from Current smokers

Comparative Analysis – Grouping Features

For the comparative analysis, grouping the samples into groups is necessary. Select any feature you want to use for separating the samples into groups.

In this example, "Sex" is selected as the grouping feature, and the analysis will be performed for Male and Female group

Grouping Features:

Clinical Features: *Collapse Selections* ∧

Alcohol:	○ Yes	● No
Recurrent:	○ Yes	● No
Grade:	○ Yes	● No
Sex:	● Yes	○ No
Smoking:	○ Yes	● No
Stage M:	○ Yes	● No
Stage N:	○ Yes	● No
Stage T:	○ Yes	● No

- Or -

Select Gene: [   ]

- Or -

TME Method: [Choose TME method ▾]

[**Submit**]

Comparative Analysis – Grouping Features

Instead of selecting a clinical feature, you can also input a gene name or select a TME method.

In this case, the samples in the dataset will be divided into high/low group based on the value of the selected feature (Gene or TME), using the median value as the cut point.

Grouping Features:

Clinical Features: *Collapse Selections* ⌃

Alcohol:	○ Yes	◉ No
Recurrent:	○ Yes	◉ No
Grade:	○ Yes	◉ No
Sex:	○ Yes	◉ No
Smoking:	○ Yes	◉ No
Stage M:	○ Yes	◉ No
Stage N:	○ Yes	◉ No
Stage T:	○ Yes	◉ No

- Or -

Select Gene: EGFR

EGFR
EGFR-AS1

- Or -

TME Method: Choose TME method ⌄

Some of the samples might be removed because the essential values for this analysis are missing.

Submit

Comparative Analysis – Target Features
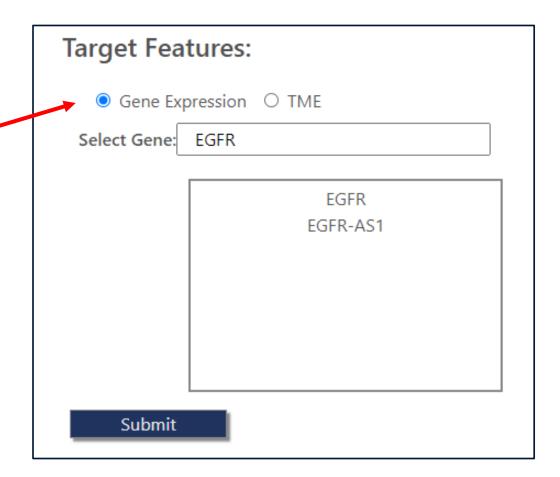
Scroll down to the page and you will see "Target Features" section

For Comparative Analysis, selecting "Target Features" is necessary
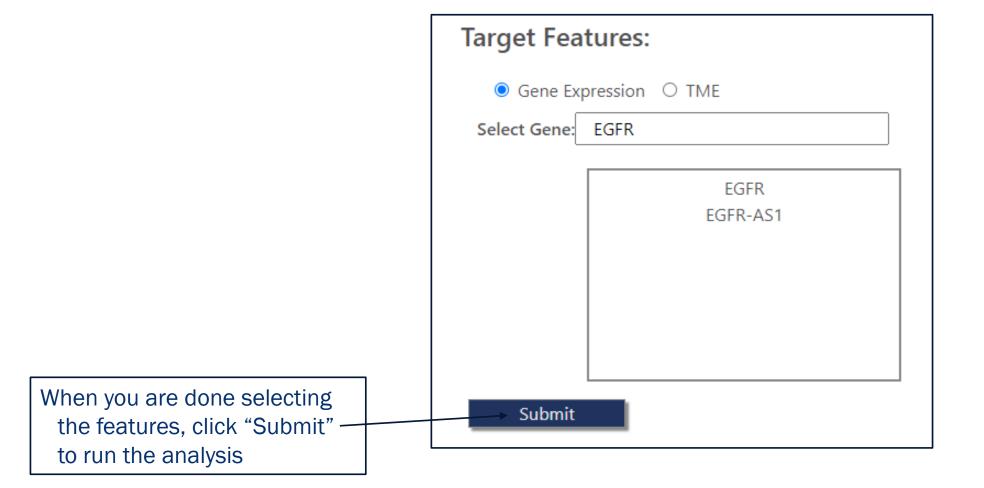
You can select any gene names for compare gene expression level or you can also select TME method + cell type to see different cell type ratio between the groups,

Comparative Analysis – Target Features

Target Features:

◉ Gene Expression   ○ TME

Select Gene: EGFR

EGFR
EGFR-AS1

Submit

When you are done selecting the features, click "Submit" to run the analysis

Comparative Analysis – Results

Image control panel (e.g., save the image, zoom in/out)

You will see the comparative analysis plot on the bottom of the page in several seconds.

Press 'Reset' to perform the survival analysis again with different settings

You can switch it to violin plot by selecting "violin' here

P-value and the number of samples in each group
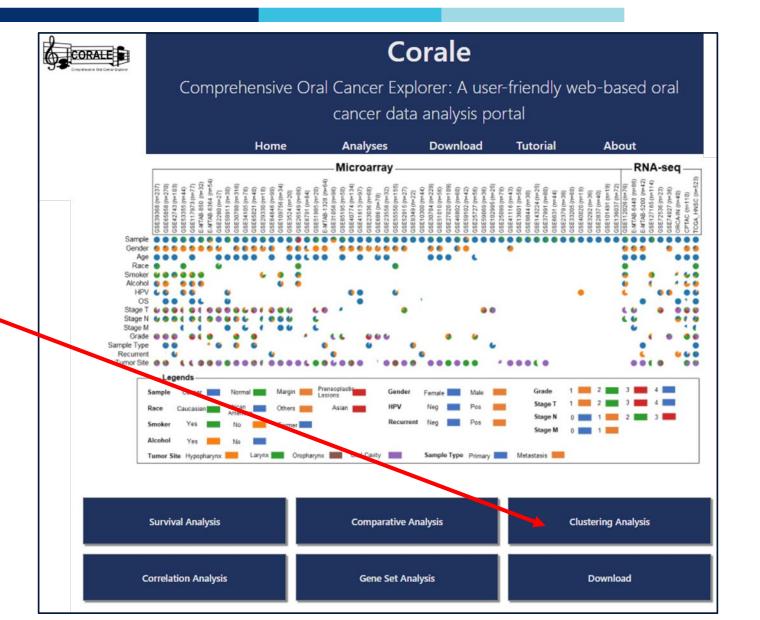
Clustering Analysis

Clustering Analysis

To run clustering analysis, click "Clustering Analysis" button on the Home Screen

Clustering Analysis – Gene Entry
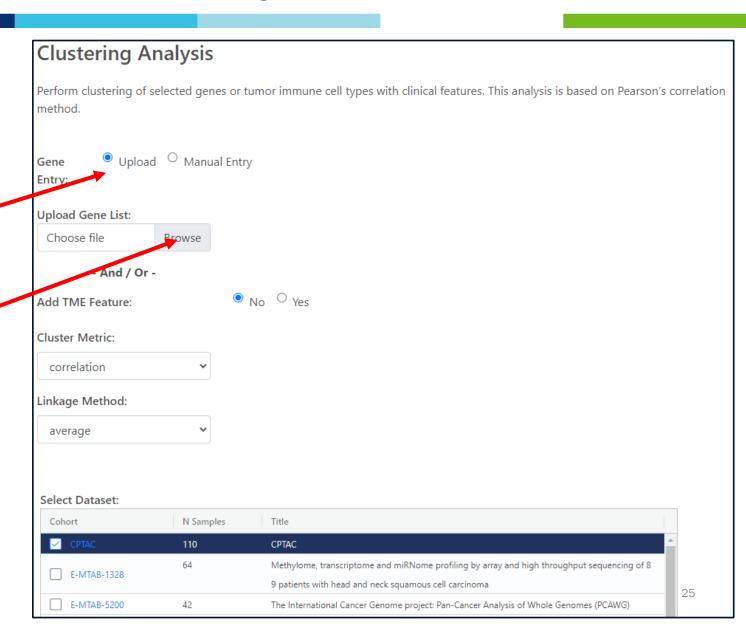
You will see this clustering analysis input screen.
There are two ways to input gene list for the analysis:

Uploading a gene list file:
To upload a text file with a gene list, select "Upload" at "Gene Entry" option.

Click "Browse" button and select the gene list file from your local machine. The gene file must be a text file with one gene symbol in each row.

Clustering Analysis

Perform clustering of selected genes or tumor immune cell types with clinical features. This analysis is based on Pearson's correlation method.

Gene Entry: ● Upload ○ Manual Entry

Upload Gene List:

| Choose file | Browse |

- And / Or -

Add TME Feature: ● No ○ Yes

Cluster Metric:

| correlation ⌄ |

Linkage Method:

| average ⌄ |

Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-1328	64	Methylome, transcriptome and miRNome profiling by array and high throughput sequencing of 8 9 patients with head and neck squamous cell carcinoma
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)

25

Clustering Analysis – Gene Entry

Another way to input gene list is manually typing the gene symbols.

Manually input gene list:
To manually input a gene list, select "Manual Entry" at "Gene Entry" option.

Input gene symbols in the gene names input window. The gene names must be separated by a semicolon (";").

Clustering Analysis

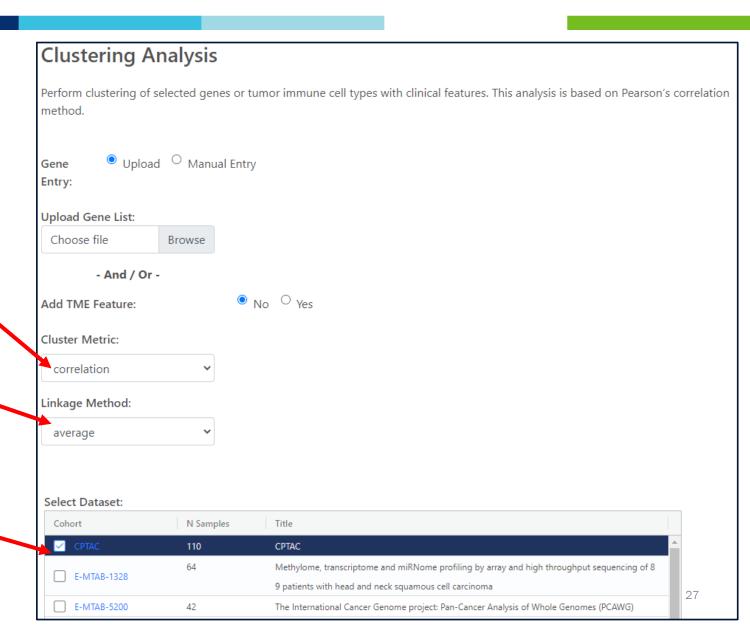Perform clustering of selected genes or tumor immune cell types with clinical features. This analysis is based on Pearson's correlation method.

Gene Entry: ○ Upload ● Manual Entry

Enter Gene Names:
Separate multiple genes with a semicolon

EGFR;TP53;KRAS;BRAF

- And / Or -

Add TME Feature: ● No ○ Yes

Cluster Metric:

correlation

Linkage Method:

average

Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-1328	64	Methylome, transcriptome and miRNome profiling by array and high throughput sequencing of 8 9 patients with head and neck squamous cell carcinoma
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)

Clustering Analysis – Data selection

The metric for clustering method can be changed here.
The default option is "Correlation".

Linkage method can be changed here, the default option is "average".

Select dataset for the analysis

Clustering Analysis

Perform clustering of selected genes or tumor immune cell types with clinical features. This analysis is based on Pearson's correlation method.

Gene Entry: ● Upload ○ Manual Entry

Upload Gene List:

| Choose file | Browse |

- And / Or -

Add TME Feature: ● No ○ Yes

Cluster Metric:

| correlation ▾ |

Linkage Method:

| average ▾ |

Select Dataset:

Cohort	N Samples	Title
☑ CPTAC	110	CPTAC
☐ E-MTAB-1328	64	Methylome, transcriptome and miRNome profiling by array and high throughput sequencing of 8 9 patients with head and neck squamous cell carcinoma
☐ E-MTAB-5200	42	The International Cancer Genome project: Pan-Cancer Analysis of Whole Genomes (PCAWG)

27

Clustering Analysis – Filter Features (optional)

To select subset of samples in the dataset, scroll down the screen and click "Filter Features" to open features list.
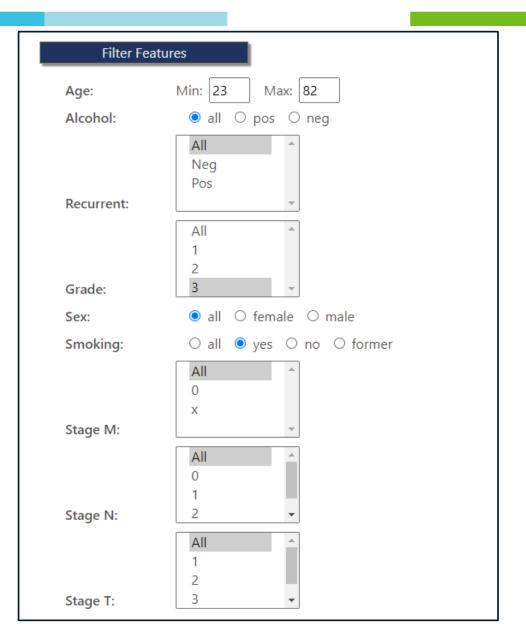
Clustering Analysis – Filter Features (optional)
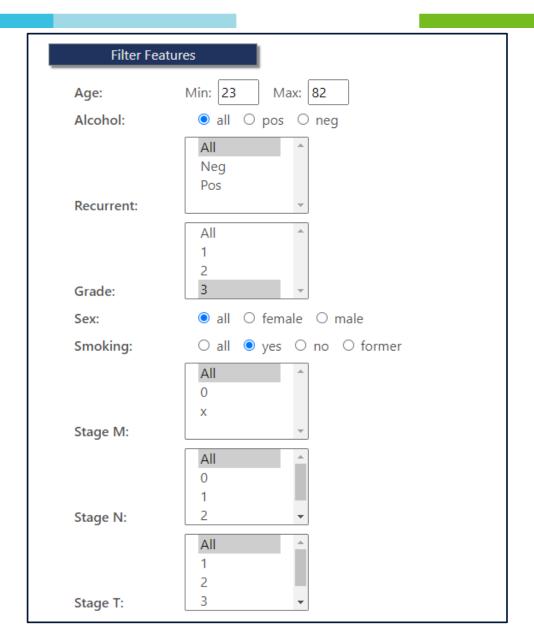
Choose sample selection criteria to use the subset of the samples in the dataset

This is an example of selecting only the samples with
1) Samples with the patients of 50 or older
and
2) Samples with Grade 3
and
4) Samples from Current smokers

Clustering Analysis – Additional Features (optional)

In the clustering analysis results, additional clinical feature can be displayed together.

In this example, "Smoking" is selected, and the smoking status of each samples will be shown in the analysis results.
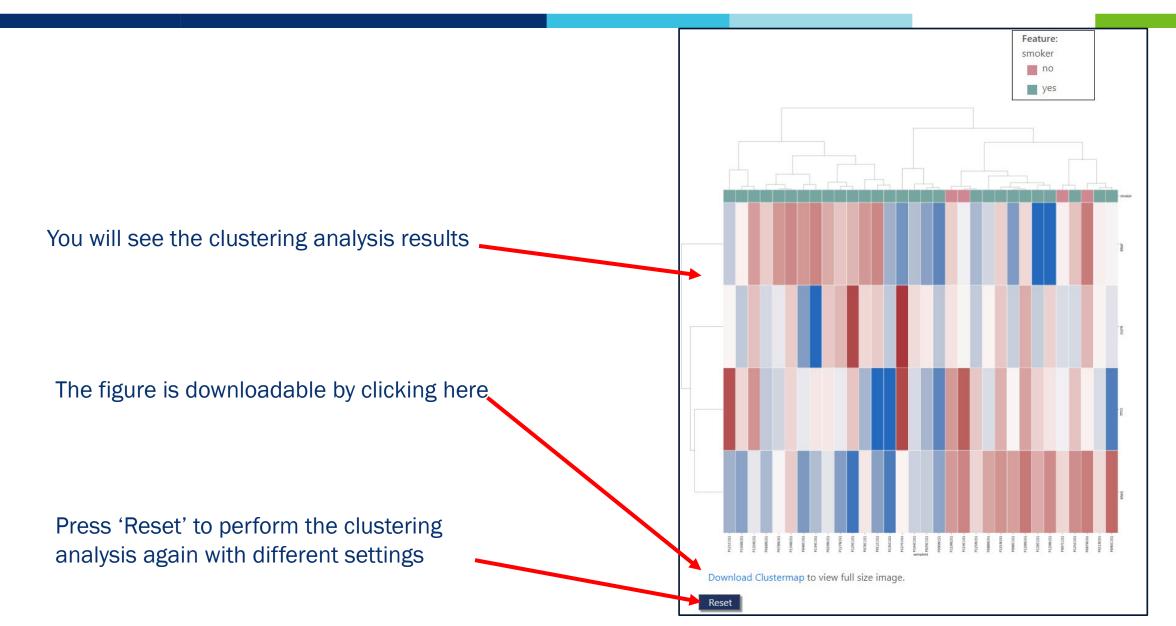
Press "Submit" button to plot the clustering analysis results.

Filter Features

Display additional clinical feature: *Collapse Selections* ⌃

Alcohol:	○ Yes	● No
Recurrent:	○ Yes	● No
Grade:	○ Yes	● No
Sex:	○ Yes	● No
Smoking:	● Yes	○ No
Stage M:	○ Yes	● No
Stage N:	○ Yes	● No
Stage T:	○ Yes	● No

Submit

Clustering Analysis – Results

You will see the clustering analysis results

The figure is downloadable by clicking here

Press 'Reset' to perform the clustering analysis again with different settings

Correlation Analysis

Correlation Analysis
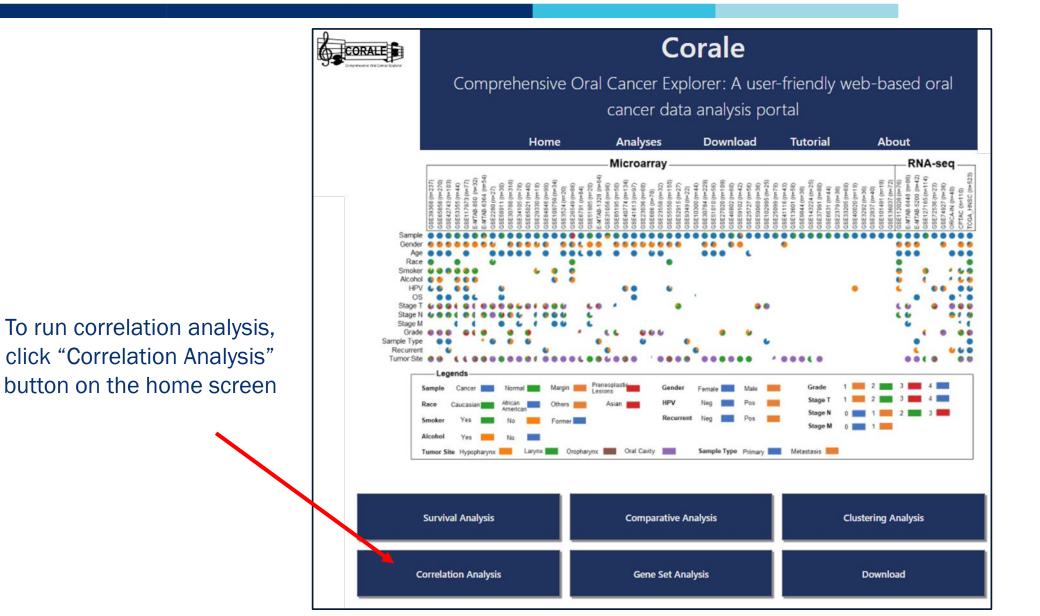
To run correlation analysis, click "Correlation Analysis" button on the home screen

Correlation Analysis – Data selection

You will see this Correlation Analysis screen.

Select any dataset that you want to use for Correlation Analysis

Correlation Analysis – Filter Features (optional)

To select subset of samples in the dataset, scroll down the screen and click "Filter Features" to open features list.

Correlation Analysis – Filter Features (optional)

Choose sample selection criteria to use the subset of the samples in the dataset

This is an example of selecting only the samples with

1) Samples with the patients of 50 or older
and
2) Samples with Grade 3
and
4) Samples from Current smokers

Correlation Analysis – Selecting Features
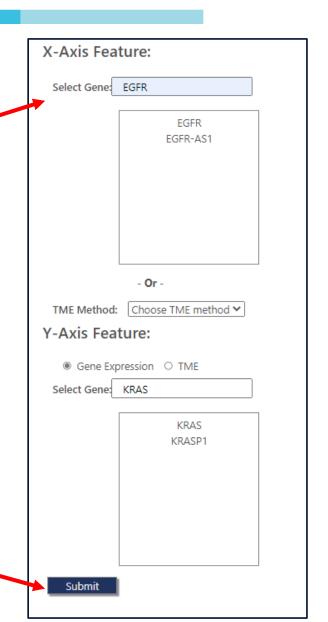
For performing correlation analysis, it is required to select two features from gene expression values or TME values.

In this example, "EGFR" gene expression value is selected as X-axis feature, and "KRAS" as Y-axis feature.

Press Submit to start the analysis.

Correlation Analysis – Results

Image control panel (e.g., save the image, zoom in/out)

You will see the scatter plot on the bottom of the page in several seconds.

P-value and the number of samples in each group

p-value: 0.0263
r: 0.2118

sample, n=110

Press 'Reset' to perform the survival analysis again with different settings

Reset
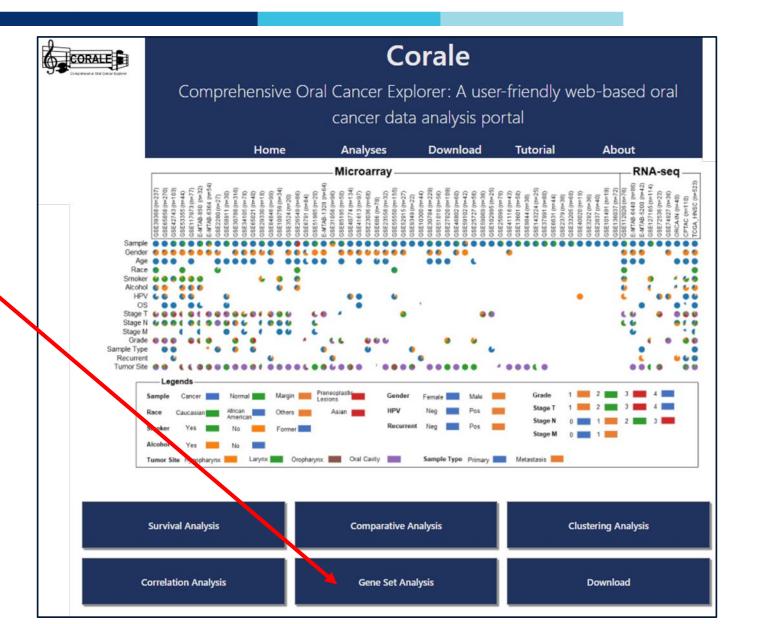
Gene Set Analysis

Gene Set Analysis

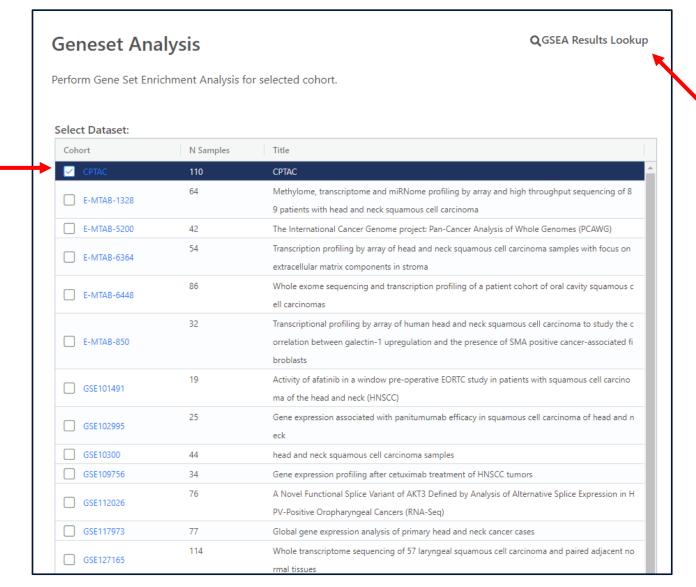To start gene set analysis, click "Gene set analysis" on the home screen.

Gene Set Analysis – Data selection

You will see this gene set analysis screen.

Select any dataset that you want to use for Survival Analysis

You can retrieve your previous gene set analysis results through this link

Gene Set Analysis – Filter Features (optional)

To select subset of samples in the dataset, scroll down the screen and click "Filter Features" to open features list.
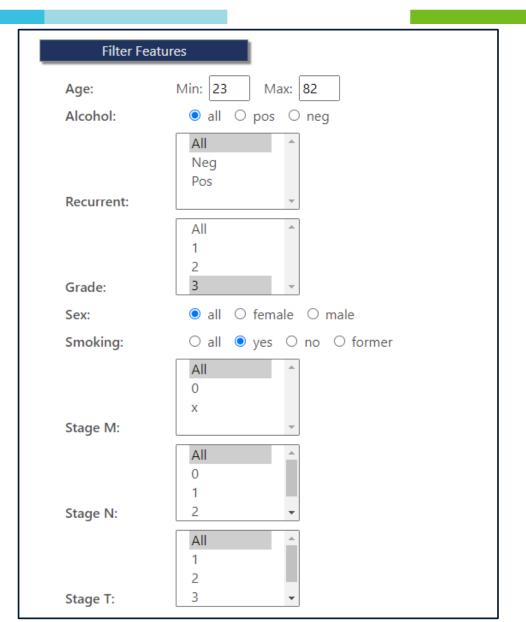
Gene Set Analysis – Filter Features (optional)

Choose sample selection criteria to use the subset of the samples in the dataset

This is an example of selecting only the samples with
1) Samples with the patients of 50 or older
and
2) Samples with Grade 3
and
4) Samples from Current smokers

Gene Set Analysis – Grouping Features

For the gene set analysis, grouping the samples into groups is necessary. Select any feature you want to use for separating the samples into groups.

In this example, "Sex" is selected as the grouping feature, and the survival analysis will be performed for Male and Female group

Grouping Features:

Clinical Features: *Collapse Selections* ∧

Alcohol:	○ Yes	● No
Recurrent:	○ Yes	● No
Grade:	○ Yes	● No
Sex:	● Yes	○ No
Smoking:	○ Yes	● No
Stage M:	○ Yes	● No
Stage N:	○ Yes	● No
Stage T:	○ Yes	● No

- Or -

Select Gene: [   ]

- Or -

TME Method: [Choose TME method ▼]

[Submit]

Gene Set Analysis – Grouping Features
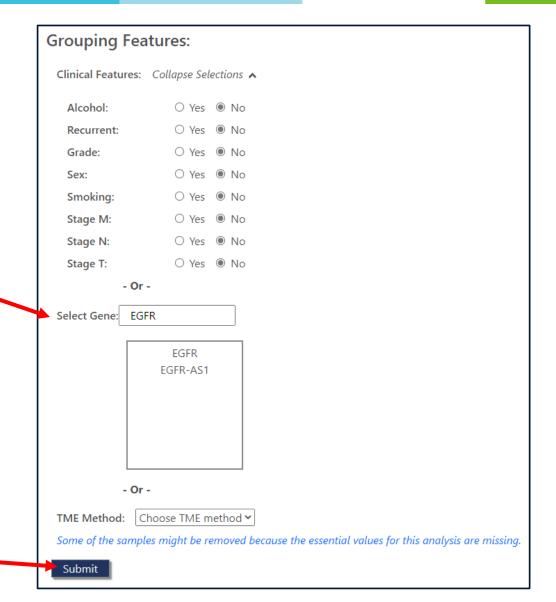
Instead of selecting a clinical feature, you can also input a gene name or select a TME method.

In this case, the samples in the dataset will be divided into high/low group based on the value of the selected feature (Gene or TME), using the median value as the cut point.

Hallmark or KEGG pathway gene sets can be selected as gene set type. For the details, please visit MSigDB

Press 'Submit' to start GSEA

Grouping Features:

Clinical Features: *Collapse Selections* ⌃

Alcohol:	○ Yes	◉ No
Recurrent:	○ Yes	◉ No
Grade:	○ Yes	◉ No
Sex:	○ Yes	◉ No
Smoking:	○ Yes	◉ No
Stage M:	○ Yes	◉ No
Stage N:	○ Yes	◉ No
Stage T:	○ Yes	◉ No

- Or -

Select Gene: EGFR

EGFR
EGFR-AS1

- Or -

TME Method: Choose TME method ⌄

Some of the samples might be removed because the essential values for this analysis are missing.

Submit

Gene Set Analysis – Running GSEA
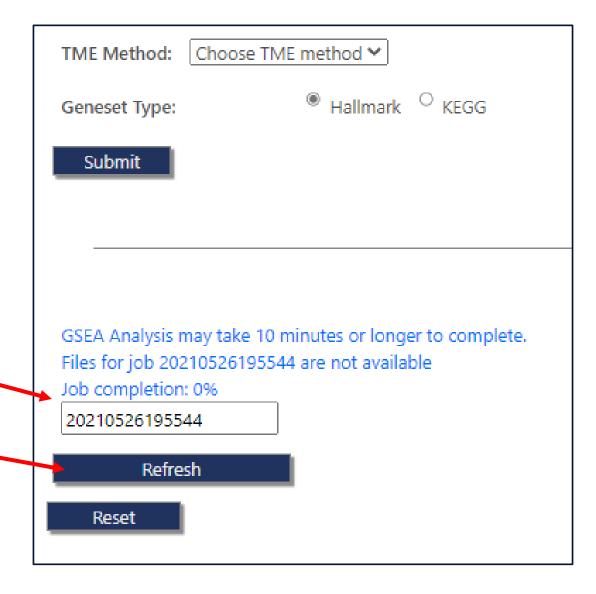
After you press 'Submit' button, you will see the progress of the submitted job with its Job ID.

If you want to retrieve the results later, this Job ID can be used.

Press 'Refresh' to see the current job status.

TME Method: Choose TME method ▼

Geneset Type: ⦿ Hallmark ◯ KEGG

Submit

GSEA Analysis may take 10 minutes or longer to complete.
Files for job 20210526195544 are not available
Job completion: 0%

20210526195544

Refresh

Reset

Gene Set Analysis – Results

When the analysis is done, the results will be displayed after pressing 'Refresh' button.

Download GSEA Archive

Feature Positive Class
Class: high

Feature Negative Class
Class: low

Positive NES:

| term	es	nes	pval	fdr	geneset_size
HALLMARK_E2F_TARGETS	0.5362	1.7823	0.0000	0.0042	200
HALLMARK_G2M_CHECKPOINT	0.5087	1.7013	0.0000	0.0105	200
HALLMARK_MYC_TARGETS_V1	0.4183	1.4164	0.0135	0.1549	200
HALLMARK_MYC_TARGETS_V2	0.4763	1.3378	0.0746	0.2444	58
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITI...	0.3890	1.2869	0.0308	0.2943	200
HALLMARK_MITOTIC_SPINDLE	0.3605	1.2126	0.0669	0.4428	199
HALLMARK_HYPOXIA	0.2688	0.9121	0.7269	1.0000	200
HALLMARK_MTORC1_SIGNALING	0.2465	0.8361	0.9310	1.0000	200
HALLMARK_INTERFERON_ALPHA_RESPONSE	0.2699	0.8244	0.8553	1.0000	97
HALLMARK_ANGIOGENESIS	0.3106	0.7806	0.8171	1.0000	36
HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.2253	0.7645	0.9955	1.0000	200
HALLMARK_DNA_REPAIR	0.2315	0.7577	0.9796	1.0000	150
HALLMARK_APICAL_JUNCTION	0.2109	0.6959	1.0000	1.0000	200
HALLMARK_GLYCOLYSIS	0.1964	0.6625	1.0000	1.0000	200
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	0.2010	0.6258	1.0000	1.0000	113
HALLMARK_NOTCH_SIGNALING	0.2286	0.5766	0.9878	1.0000	32
HALLMARK_INFLAMMATORY_RESPONSE	0.1571	0.5320	1.0000	1.0000	200
HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.1945	0.5276	1.0000	1.0000	42
HALLMARK_TGF_BETA_SIGNALING	0.1804	0.4993	1.0000	1.0000	54
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.1669	0.4964	1.0000	1.0000	87
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.1397	0.4684	1.0000	1.0000	200
HALLMARK_APOPTOSIS	0.1349	0.4399	1.0000	1.0000	161
HALLMARK_PROTEIN_SECRETION	0.1284	0.3960	1.0000	1.0000	96
HALLMARK_UV_RESPONSE_UP	0.1194	0.3888	1.0000	1.0000	158
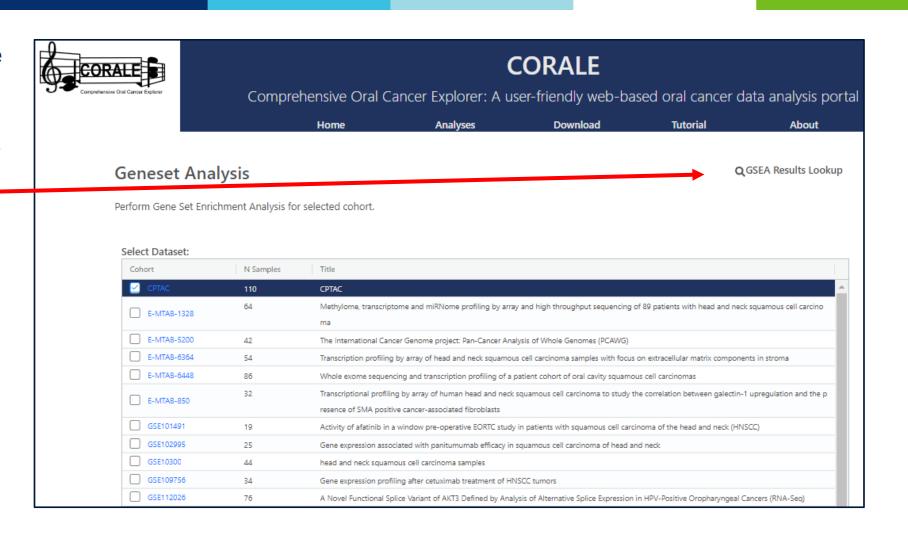
Gene Set Analysis – Retrieving Results

There is another way to retrieve the GSEA results using Job ID

On the main screen of Geneset Analysis, press "GSEA Results Lookup" button.
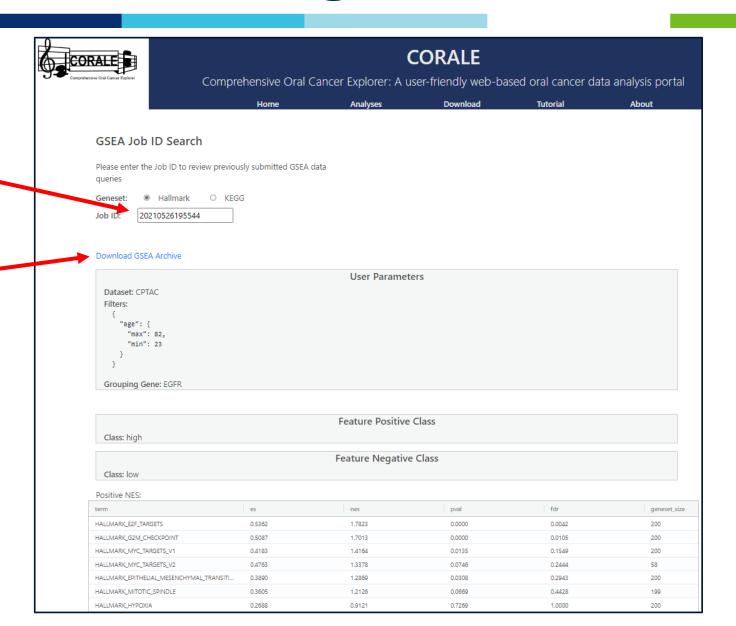
Gene Set Analysis – Retrieving Results

As soon as typing the Job ID in the input window, the geneset analysis results will be displayed.

The raw results files can be downloaded by clicking this "Download GSEA Archive" button
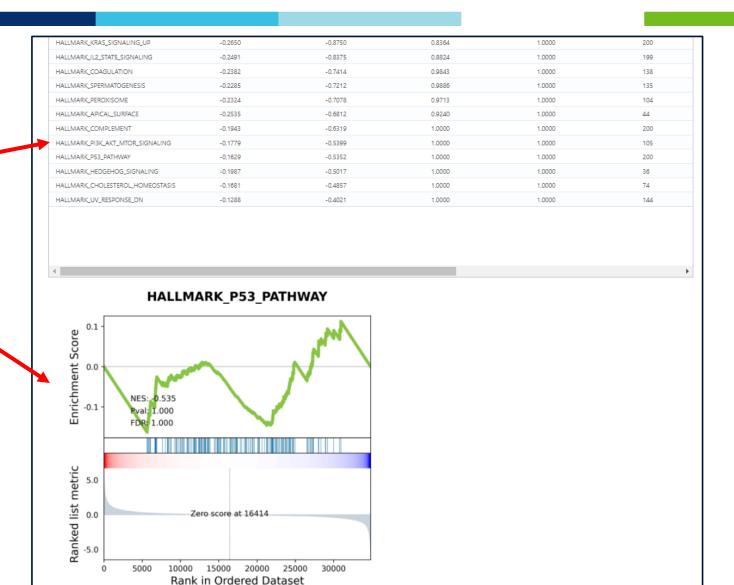
Gene Set Analysis – Retrieving Results

By clicking the result you are interested in,

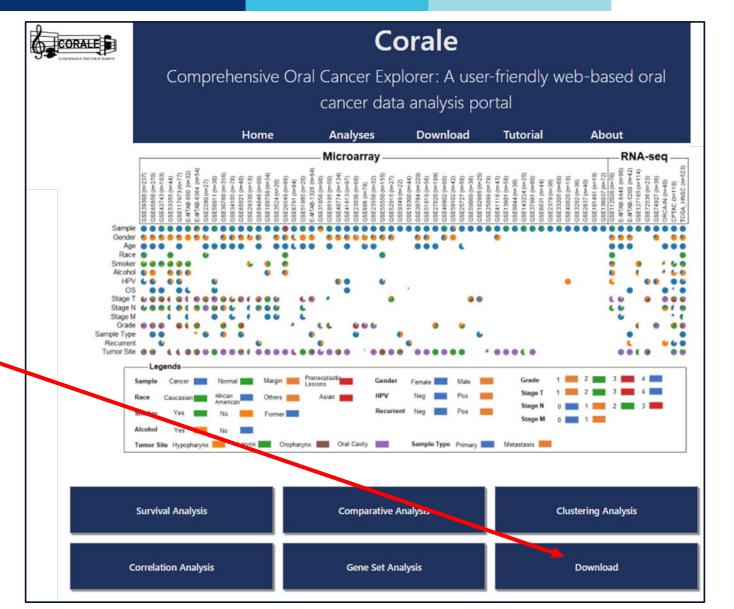you can find the analysis plot with the details on the bottom of the page.

HALLMARK_KRAS_SIGNALING_UP	-0.2650	-0.8750	0.8364	1.0000	200
HALLMARK_IL2_STAT5_SIGNALING	-0.2491	-0.8375	0.8824	1.0000	199
HALLMARK_COAGULATION	-0.2382	-0.7414	0.9843	1.0000	138
HALLMARK_SPERMATOGENESIS	-0.2285	-0.7212	0.9886	1.0000	135
HALLMARK_PEROXISOME	-0.2324	-0.7078	0.9713	1.0000	104
HALLMARK_APICAL_SURFACE	-0.2535	-0.6812	0.9240	1.0000	44
HALLMARK_COMPLEMENT	-0.1943	-0.6319	1.0000	1.0000	200
HALLMARK_PI3K_AKT_MTOR_SIGNALING	-0.1779	-0.5399	1.0000	1.0000	105
HALLMARK_P53_PATHWAY	-0.1629	-0.5352	1.0000	1.0000	200
HALLMARK_HEDGEHOG_SIGNALING	-0.1987	-0.5017	1.0000	1.0000	36
HALLMARK_CHOLESTEROL_HOMEOSTASIS	-0.1681	-0.4857	1.0000	1.0000	74
HALLMARK_UV_RESPONSE_DN	-0.1288	-0.4021	1.0000	1.0000	144

HALLMARK_P53_PATHWAY

NES: -0.535
Pval: 1.000
FDR: 1.000

Zero score at 16414

Download datasets /
Retrieve the details of the datasets

Download datasets

To see the datasets details and download the datasets, click "Download" button on the home screen
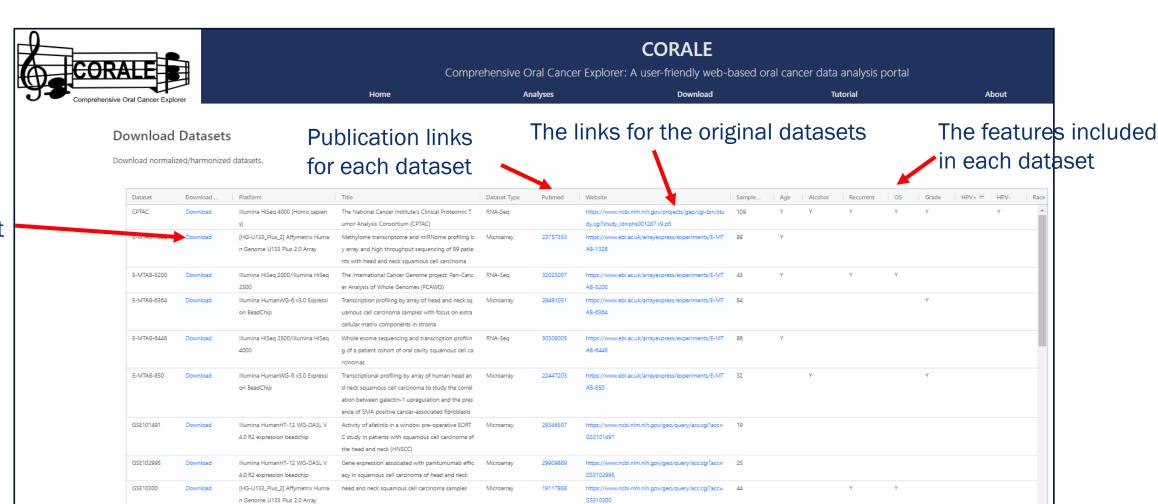
Download datasets

This page shows all the available datasets in CORALE and its detailed information.

Publication links for each dataset

The links for the original datasets

The features included in each dataset

Download a dataset

OTHER INFO

OTHER INFO

- ## Please cite CORALE
  - Kyubum Lee, Mengyu Xie, Scott D. Cukras, John H. Lockhart, Rodrigo Carvajal, Elsa R. Flores, Christine H. Chung, and Aik-Choon Tan, **Comprehensive Oral Cancer Explorer (CORALE): A user-friendly web-based oral cancer data analysis portal.** *[Under Review] (2021)*

- ## Please contact us
  - Aik Choon Tan, PhD: AikChoon.Tan@moffitt.org
  - Kyubum Lee, PhD: Kyubum.Lee@moffitt.org

- ## Thank you for using CORALE