Pharma SUG China 2025 - Paper DV- 156

Pathway in RNA-seq Analysis Using R Shiny App: From Design to Visualization with clinical endpoint

Zhanyun Hu, Sanofi Corporation.

ABSTRACT

With the advancement of precision medicine, biomarker analysis plays a pivotal role in exploratory analysis and clinical decision making (e.g., drug mechanism-of-action, dose recommendation, patient selection). This paper explores biomarker analysis using an R Shiny APP, covering the workflow of RNA-seq pathway result visualization. R Shiny, an interactive web framework, facilitates dynamic visualization of RNA-seq pathway analysis. Users can explore functional relationships between clinical endpoints and RNA-seq pathways through interactive plots, including heatmaps, box plots, scatter plots, and bar plots, etc that are provided in the APP. Our work offers a scalable and user-friendly solution for biomarker analysis to efficiently explore data and derive reliable insights.

Key words: RNA-seq, Pathway, R Shiny APP, heatmaps, box plots, scatter plots, bar plots

BACKGROUND

Pathway refers to a collection of genes that work together in biological processes. RNA-Seq pathway analysis helps researchers interpret large-scale gene expression data by organizing it into the context of biological pathways. Instead of focusing on individual genes, this analysis identifies groups of genes involved in specific cellular processes that show significant changes in activity. This perspective provides a more meaningful understanding of the molecular events that underlie various biological states and diseases.

GSEA (Gene Set Enrichment Analysis) one of important pathway analysis methods. GSEA is used to identify collections of genes whose expression patterns undergo overall changes under two biological states, such as disease vs. health. It does not require setting the differential expression threshold in advance. Instead, it analyzes the enrichment of the gene set at the top/bottom of the sequencing gene list and calculates the statistical significance through permutation test .

The main output metrics include standardized enrichment score (NES), and false detection rate (FDR).

Pathway analysis serves as the core bridge for RNA-seq to transition from gene lists to biological significance, revealing the mechanisms of phenotypic changes through functional associations and providing target directions for subsequent experiments.

INTRODUCTION

This paper will introduce the Design and Visualization of Pathway of RNA-seq and clinical endpoint in R shiny APP. The dummy clinical endpoint data and model between clinical endpoint and pathway results generated by different methods such as GSEA or other methods will be prepared in advance for visualization.

Here is the Structure of Shiny APP:

Key UI:

- Choose clinical endpoint:
- Choose visit for clinical endpoint:
- Choose effect:
- Analysis Method:
- Pathway list choice:

Server:

- Plot: Scatter plot, box plot, bar plot
- Summary table.
- Heatmap

DESIGN OF SHINY APP

DEFINE UI PANEL

Figure 1: UI Panel with Drop down list:①②③④⑤

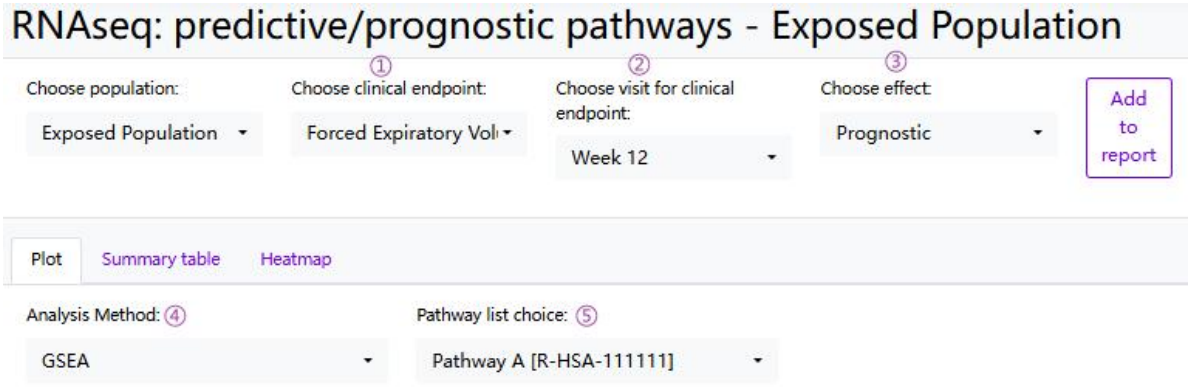


Figure 1 shows different dropdown list for user selection. User can choose different values in dropdown list to explore the relation between clinical endpoint and pathway with different method by effect.

① Continuous or Binary/Categorical kind endpoint can be selected for different plots

② Visit from clinical end point, in the pathway model data

③ Effect from pathway model data.

④ Method from pathway data

⑤ Pathway type from pathway data

Example Code of one dropdown list:

```
pickerInput(
          ns("endpoint_selected"),
          label = "Choose clinical endpoint:",
          choices = NULL
        ) %>% column(width = 2)
```

DATA PREPARATION

Three datasets are required in the APP:

rnaseq_pathway_expression.rds, clinical_endpoint.rds, rnaseq_pathway_predprog.rds

1) rnaseq_pathway_expression.rds :

   a) Contain the Pathway expression data generated by different methods such as GSEA or other method.

   b) In data of each method, the USUBJID, AVISIT, ARM and each pathway value is necessary.

   c) In attributes of the data, we can get the names of each method that can be retrieved in dropdown list.

```
method <- names(attributes(rnaseq_pathway_expression))
    updatePickerInput(session,
      inputId = "path_method",
      choices = method,
      selected = method[1]
    )
```

   d) When selecting one method from the dropdown list, the plot from corresponding data will be displayed in Shiny server panel.

rnaseq_pathway_expression

Show Attributes

Name	Type	Value
rnaseq_pathway_expre...	list [2]	List of length 2
GSVA	list [559 x 13] (S3: data.frame	A data.frame with 559 rows and 13 columns
GSEA	list [559 x 13] (S3: data.frame	A data.frame with 559 rows and 13 columns

GSEA

Filter

	BARCODE	USUBJID	AVISIT	ARM	ARMN	AVISITN	Pathway A [R-HSA-111111]	Pathway B [R-HSA-222222]	Pathway C [R-HSA-333333]
1	TED12345-9001	12345-056-0001-03018	Baseline	DrugX 3 mg Q2W	2	0	1.6290112	5.286609	5.812152
2	TED12345-9002	12345-124-0001-03006	Week 4	SAR123456 1 mg Q2W + DrugX 3 mg Q2W	3	4	2.0277510	4.169999	5.674448
3	TED12345-9004	12345-124-0001-03008	Baseline	SAR123456 1 mg Q2W + DrugX 3 mg Q2W	3	0	2.9045864	3.873285	4.972352
4	TED12345-9006	12345-124-0001-03013	Week 12	SAR123456 1 mg Q2W	1	12	2.6597120	4.411546	5.267371
5	TED12345-9008	12345-124-0002-04004	Week 4	SAR123456 1 mg Q2W	1	4	2.4465198	4.380670	5.754897
6	TED12345-9009	12345-124-0002-04012	Baseline	SAR123456 1 mg Q2W + DrugX 3 mg Q2W	3	0	2.3096482	5.025051	5.634216
7	TED12345-9011	12345-124-0002-06004	Baseline	DrugX 3 mg Q2W	2	0	3.0072267	3.773110	5.030232
8	TED12345-9012	12345-124-0003-03002	Week 4	DrugX 3 mg Q2W	2	4	1.9325439	6.013123	6.055858
9	TED12345-9013	12345-124-0003-03011	Baseline	DrugX 3 mg Q2W	2	0	3.0987771	3.611756	5.750331
10	TED12345-9015	12345-158-0002-08008	Baseline	DrugX 3 mg Q2W	2	0	2.3067121	4.358531	6.018814

```
> attributes(rnaseq_pathway_expression)
$names
[1] "GSVA" "GSEA"
```

2) rnaseq_pathway_predprog.rds :

a) This model data is derived from expression data per model statistician specified.

b) Name of Pathway, contrast, endpoint, visit and other column from model is required in this data.

c) The column name of contract and visit is predefined in the attributes $contrast and $visit. Shiny APP retrieves the attributes to get the column name.

rnaseq_pathway_predprog

Filter

	Description	setSize	NES	pvalue	adjusted_pvalue	population	contrast	endpoint	visit
1	Pathway A [R-HSA-111111]	30/50	-1.150	0.0773055	0.0081000	EXPOSEFL	Prognostic	FEV1POST	Week 12
2	Pathway B [R-HSA-222222]	224/300	-2.080	0.8934055	0.9901000	EXPOSEFL	Prognostic	FEV1POST	Week 12
3	Pathway C [R-HSA-333333]	121/400	-0.900	0.0002055	0.0001000	EXPOSEFL	Prognostic	FEV1POST	Week 12
4	Pathway D [R-HSA-444444]	359/800	-2.260	0.7899055	0.9901000	EXPOSEFL	Prognostic	FEV1POST	Week 12
5	Pathway E [R-HSA-555555]	21/400	-1.040	0.0045055	0.0010000	EXPOSEFL	Prognostic	FEV1POST	Week 12
6	Pathway F [R-HSA-666666]	99/150	-0.480	0.0005055	0.0010000	EXPOSEFL	Prognostic	FEV1POST	Week 12
7	Pathway G [R-HSA-777777]	110/300	-0.910	0.6589055	0.6666000	EXPOSEFL	Prognostic	FEV1POST	Week 12
8	Pathway A [R-HSA-111111]	30/50	-2.150	0.0773005	0.0081000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
9	Pathway B [R-HSA-222222]	224/300	-3.080	0.8934005	0.9901000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
10	Pathway C [R-HSA-333333]	121/400	-1.900	0.0002005	0.0001000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
11	Pathway D [R-HSA-444444]	359/800	-3.260	0.7899005	0.9901000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
12	Pathway E [R-HSA-555555]	21/400	-2.040	0.0045005	0.0010000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
13	Pathway F [R-HSA-666666]	99/150	-1.480	0.0005005	0.0010000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12
14	Pathway G [R-HSA-777777]	110/300	-1.910	0.6589005	0.6666000	EXPOSEFL	Predictive: DrugX 3 mg Q2W - Placebo	FEV1POST	Week 12

```
> attributes(rnaseq_pathway_predprog)
$names
[1] "Description"     "setSize"         "NES"
"pvalue"          "adjusted_pvalue" "population"
"contrast"        "endpoint"        "visit"

$class
[1] "data.frame"

$contrast
[1] "contrast"

$visit
[1] "visit"
```

3) clinical_endpoint.rds:

a) Dataset contains the clinical endpoint data. USUBJID, treatment ARM, AVISIT and clinical endpoint value is required.

b) The continuous or categorical kind of the endpoint are predefined in the attributes $endpoint according to PARAM. It's used for defining the dropdown list.

clinical_endpoint

Filter

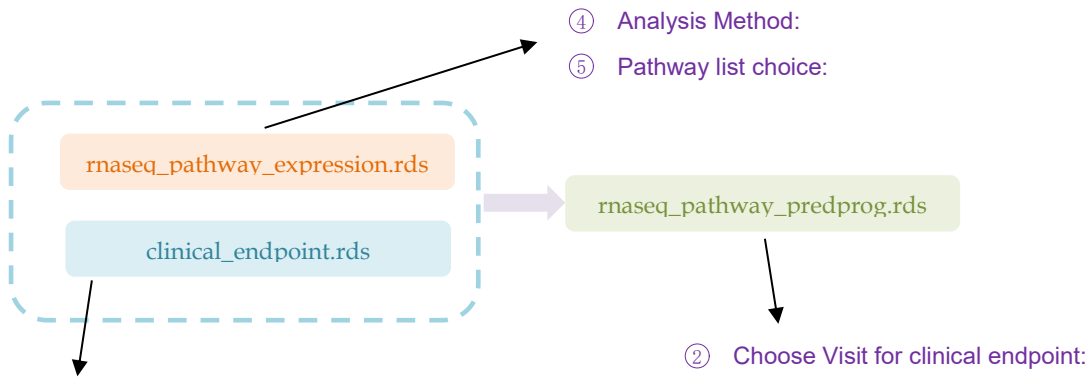
	STUDYID	USUBJID	ARM	PARAM	PARAMCD	AVAL	AVALC	BASE	CHG	PCHG	AVISIT	AVISITN
53	TED12345	12345-124-0001-03008	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Pre Bronchodilator	FEV1PRE	3.46	NA	3.60	-0.14	-3.8888889	Week 10	12
54	TED12345	12345-124-0001-03008	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Post Bronchodilator	FEV1POST	3.62	NA	3.91	-0.29	-7.4168798	Week 10	12
55	TED12345	12345-124-0001-03008	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Pre Bronchodilator	FEV1PRE	3.30	NA	3.60	-0.30	-8.3333333	Week 11	13
56	TED12345	12345-124-0001-03008	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Pre Bronchodilator	FEV1PRE	3.34	NA	3.60	-0.26	-7.2222222	Week 12	14
57	TED12345	12345-124-0001-03008	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Post Bronchodilator	FEV1POST	3.47	NA	3.91	-0.44	-11.2531969	Week 12	14
60	TED12345	12345-124-0001-03013	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Pre Bronchodilator	FEV1PRE	1.73	NA	1.73	0.00	0.0000000	Baseline	0
61	TED12345	12345-124-0001-03013	DrugX 3 mg Q2W	Forced Expiratory Volume in 1 Second (L) Post Bronchodilator	FEV1POST	2.11	NA	2.11	0.00	0.0000000	Baseline	0

```
> attributes(clinical_endpoint)
$names
 [1] "STUDYID" "USUBJID" "ARM"      "PARAM"
"PARAMCD" "AVAL"    "AVALC"    "BASE"     "CHG"
"PCHG"    "AVISIT"
[12] "AVISITN"

$endpoint
   PARAMCD   PARAM_cat
1  FEV1PRE  continuous
2 FEV1POST  continuous
3     LOAC categorical
4     RESP categorical
5   RESPSR categorical
6  RESPCAT categorical
```

DATA SOURCE TO UI PANEL

Figure 2: source of each dropdown list



a. Continuous

b. Binary/Categorical

Figure 2 shows different dropdown list from different source rds data

① clinical_endpoint$PARAM

② rnaseq_pathway_predprog$visit

③ rnaseq_pathway_predprog$contrast

④ names(rnaseq_pathway_expression)

⑤ pathway column in naseq_pathway_expression

SERVER PANEL

In the server part, different report can be available in different Tabs. Here is the summary of each Tabs.

Tabs	Plot Type	Comment	resource data
Plot	scatter plot	Continuous Clinical endpoint vs Prognostic effect	clinical_endpoint rnaseq_pathway_expression
	box plot by treatment	Continuous Clinical endpoint vs Predictive effect	
	box plot by categorical endpoint	Binary/Categorical Clinical endpoint vs Prognostic effect	
	bar plot	Binary/Categorical Clinical endpoint vs Predictive effect	
Summary Table	listing of the model result		rnaseq_pathway_predprog
Heatmap	Heat map		clinical_endpoint rnaseq_pathway_expression rnaseq_pathway_predprog

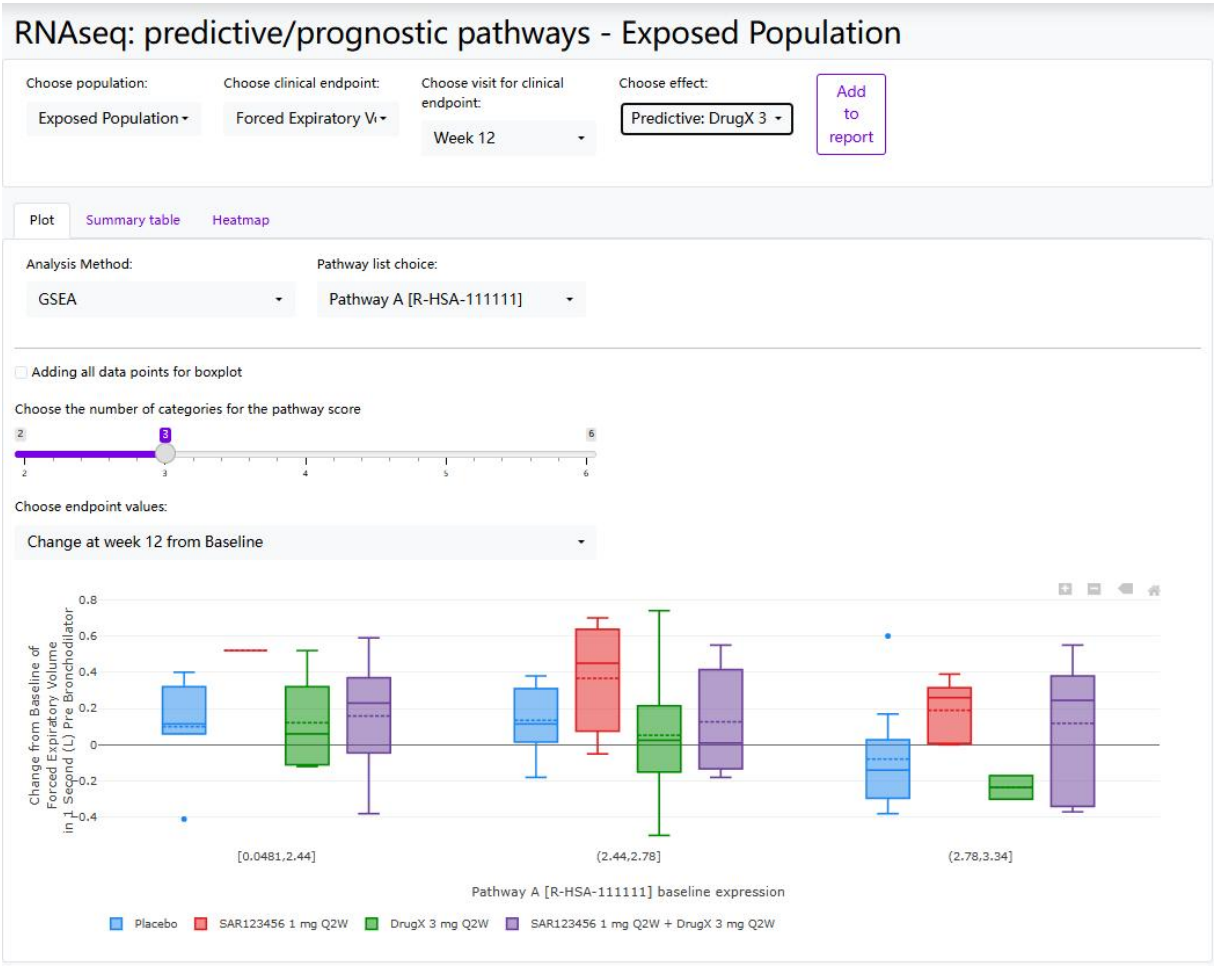
Tabs – plot

The different type of plot is displayed if user selects different kind of ②Clinical endpoint or ③effect

a) Figure 3: Continuous Clinical endpoint vs Prognostic effect

In Figure 3: In this scatter plot, X axis is the numeric Pathway value. Y axis is the numeric clinical endpoint. User can know if there is any correlation trend between numeric clinical endpoint and pathway.

b) Figure 4: Continuous Clinical endpoint vs Predictive effect

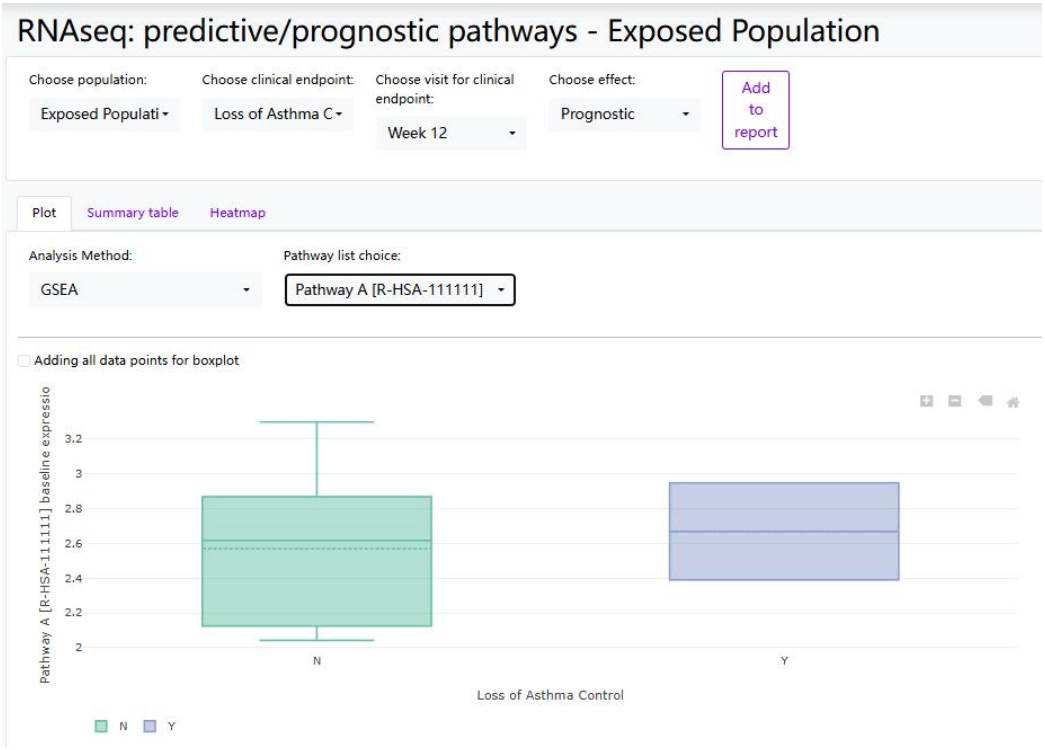


In Figure 4: The box plot displayed the value of end point grouped by treatment and the baseline pathway value.

The pathway value can be divided into different parts per quantile.

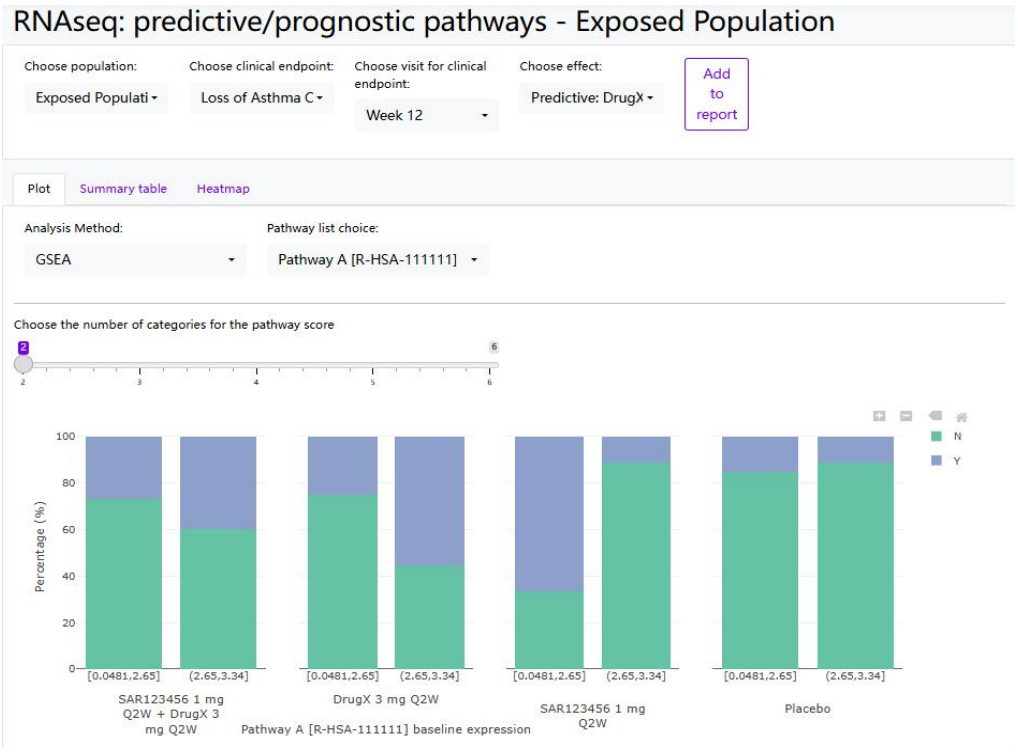
```
protgrp = cut(!!as.name(bm), include.lowest = T, breaks = quantile(!!as.name(bm), probs = seq(0, 1, by = (1 / protcat)), na.rm = TRUE))
```

c) Figure 5: Binary/Categorical Clinical endpoint vs Prognostic effect



In Figure 5: The box plot displays the pathway value grouped by the Binary/Categorical Clinical endpoint.

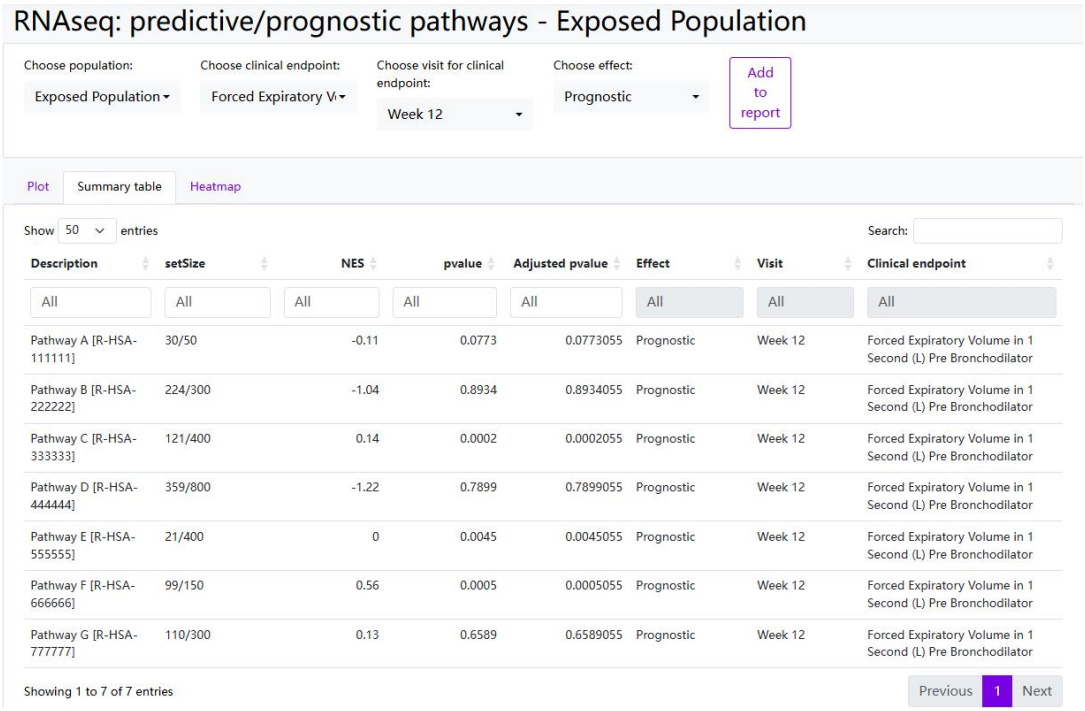
d) Figure 6: Binary/Categorical Clinical endpoint vs Predictive effect



In Figure 6: The bar plot displays percentage of each endpoint grouped by treatment and the pathway value.

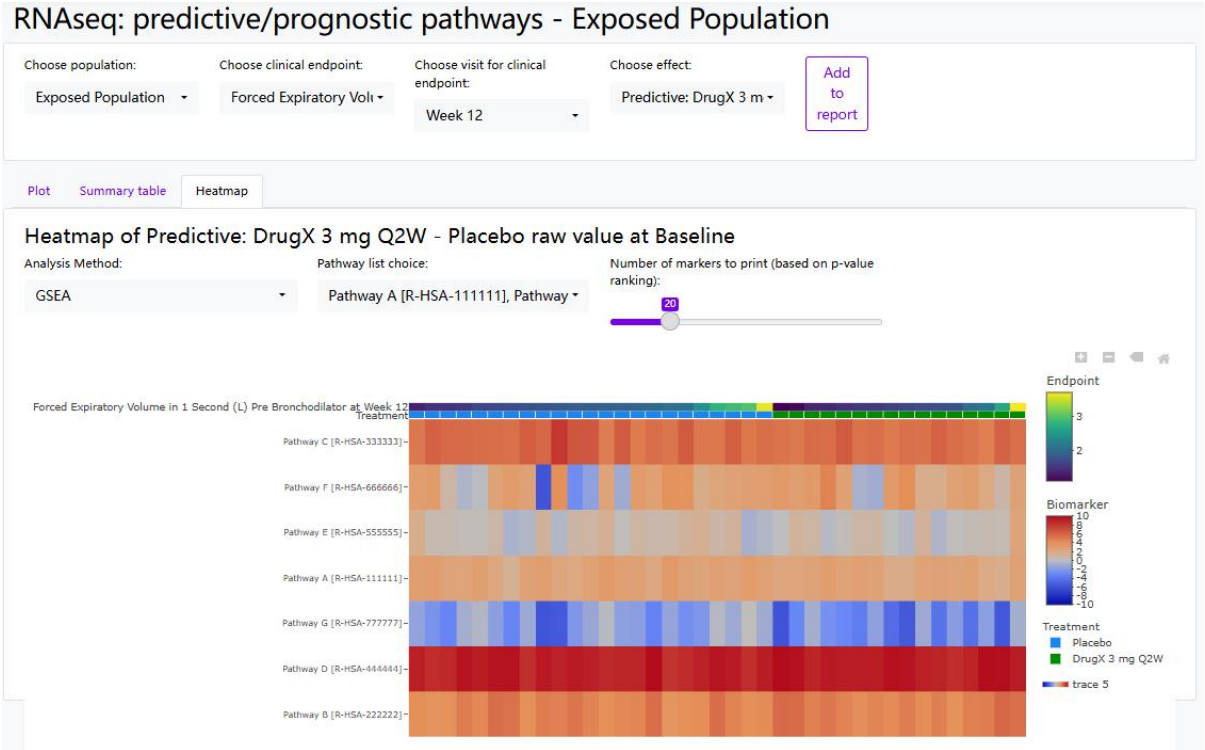
Tabs – Summary table

Figure 7: NES result.



Tabs – Heatmap

Figure 8: Heatmap of by different treatment and endpoint



In Figure 8: The heatmap displays the different pathway value of each subject at baseline grouped by treatment and endpoint. The order of pathway is ranked by the p value from of the model data.

CONCLUSION

User prepares rnaseq_pathway_expression.rds, clinical_endpoint.rds, rnaseq_pathway_predprog.rds data. Then the Shiny APP can load the data automatically and provide different kinds of interactive plots that allows user quicklier exploring pathway RNA-seq data to detect the potential relationship between pathway and clinical endpoint.

REFERENCES

Accessed June 15, 2025 What Is RNA-Seq Pathway Analysis and How Is It Used? - Biology Insights

ACKNOWLEDGMENTS

The author would like to thank Emmanuel Bonnet, Gawain Gao, Jason Deng, Nils Ternes, Antoine Brisacier, Jenny Hou for their development support or advice.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Zhanyun Hu

Zhanyun.Hu@sanofi.com