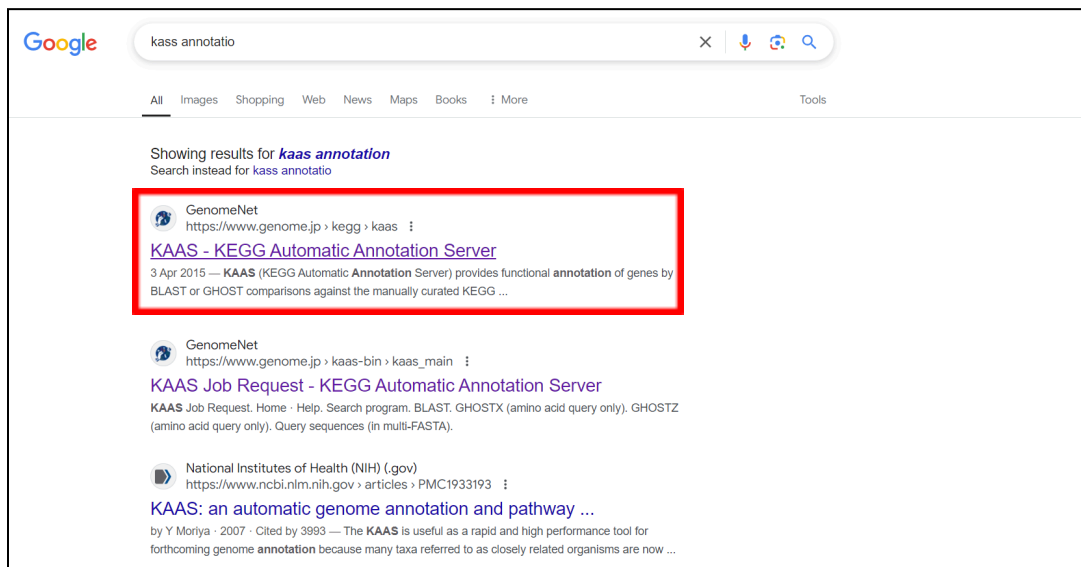


Step-By-Step Method To Do KEGG Pathway Enrichment Analysis

* All the sample input files used in this tutorial are available in the github repository
https://github.com/ajigib/IV_Workshop

* You can clone the this repository using git clone
`git clone https://github.com/ajigib/IV_Workshop.git`

1. KEGG Annotation Of Protein Sequences:



- Open your favorite web browser and search for the KAAS annotation tool.
- Click on the first link of google result to open the KAAS annotation tool.

genome.jp/kaas-bir/kaas_org

File upload (☐ Nucleotide)
 xp_protein_... uences fasta

Query name

E-mail address

GENES data set (javascript button or text box)
 Representative set

Manual selection
 (popup window)

Selected organisms (up to 40 organisms)

 (Example) eco, bsu, syn

Assignment method
☒ BBH (bi-directional best hit)
☐ SBH (single-directional best hit)

| | |
|------|---|
| hgt | Heterocephalus glaber (naked mole-rat) |
| ccan | Castor canadensis (American beaver) |
| ocu | Oryctolagus cuniculus (rabbit) |
| tup | Tupaia chinensis (Chinese tree shrew) |
| cfa | Canis lupus familiaris (dog) |
| vvp | Vulpes vulpes (red fox) |
| aml | Ailuropoda melanoleuca (giant panda) |
| umr | Ursus maritimus (polar bear) |
| uah | Ursus arctos (brown bear) |
| elk | Enhydra lutris kenyoni (northern sea otter) |
| oro | Odobenus rosmarus divergens (Pacific walrus) |
| fca | Felis catus (domestic cat) |
| ptg | Panthera tigris altaica (Amur tiger) |
| ppad | Panthera pardus (leopard) |
| bta | Bos taurus (cow) |
| bom | Bos mutus (wild yak) |
| blu | Bos indicus (zebu cattle) |
| bubu | Bubalus bubalis (water buffalo) |
| chx | Capra hircus (goat) |
| oar | Ovis aries (sheep) |
| ssc | Sus scrofa (pig) |
| cfr | Camelus ferus (Wild Bactrian camel) |
| cdk | Camelus dromedarius (Arabian camel) |
| bacu | Balaenoptera acutorostrata scammoni (minke whale) |
| lve | Lipotes vexillifer (Yangtze River dolphin) |
| nor | Orcinus orca (killer whale) |

- For the gene data set you can select any representative set or manual selection, but in this case we will select manual selection.
- Click on the organism list button, a new small window will pop-up, click on **chx** (*Capra hircus*).

E-mail address

GENES data set (javascript button or text box)
 Representative set

Manual selection
 (popup window)

Selected organisms (up to 40 organisms)

 (Example) eco, bsu, syn

Assignment method
☒ BBH (bi-directional best hit)
☐ SBH (single-directional best hit)

Automatic Annotation Server Ver. 2.1

- KEGG organism code will appear in the white box, you can add related species for annotation according to your input species.

Query name

query

E-mail address

ajaybhatia.igib@gmail.com

GENES data set (javascript button or text box)

Representative set

for GENES for Eukaryotes for Prokaryotes

Manual selection

Organism list (popup window)

Selected organisms (up to 40 organisms)

ctx

(Example) eco, bsu, syn

clear

Show the number of sequences in selected organisms.

Assignment method

☒ BBH (bi-directional best hit)


☐ SBH (single-directional best hit)

Compute Clear

Automatic Annotation Server Ver. 2.1

- Fill your email address in a dedicated text box.
- Keep the BBH (bi-directional best hit) selected in the assignment method section.
- Now click on the compute button on the lower right side.

KAAS - Job request accepted Inbox x

 **noreply@genome.jp**
to me ▼

KEGG Automatic Annotation Server Ver. 2.1
Your request ID is 1729055972.

Please click on the link below to submit or cancel your job.

(Submit)
https://www.genome.jp/kaas-bin/kaas_main?mode=submit&id=1729055972&key=cRQ8lwYP

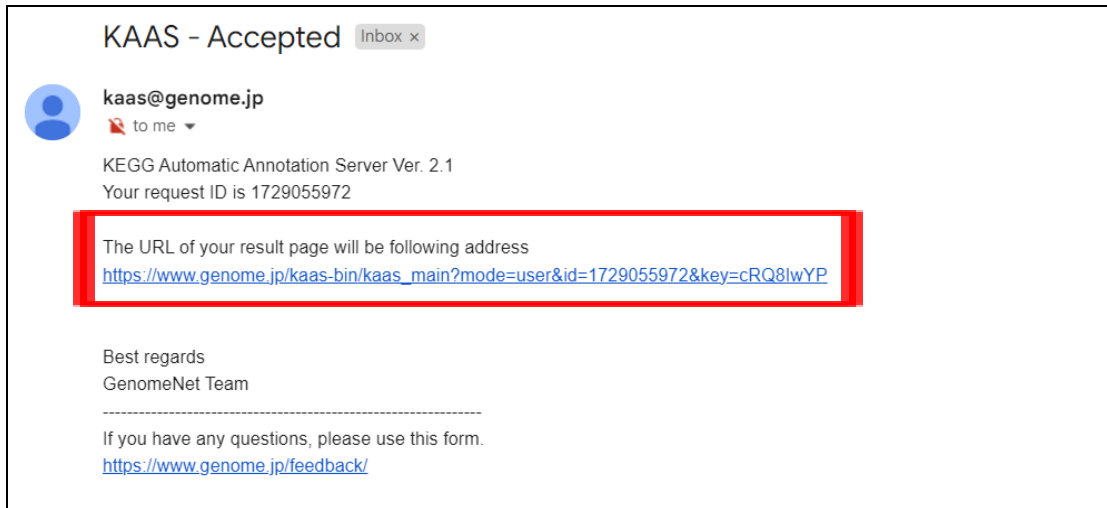
(Cancel)
https://www.genome.jp/kaas-bin/kaas_main?mode=cancel&id=1729055972&key=cRQ8lwYP

If no action is taken within 7 days, your request will be deleted.

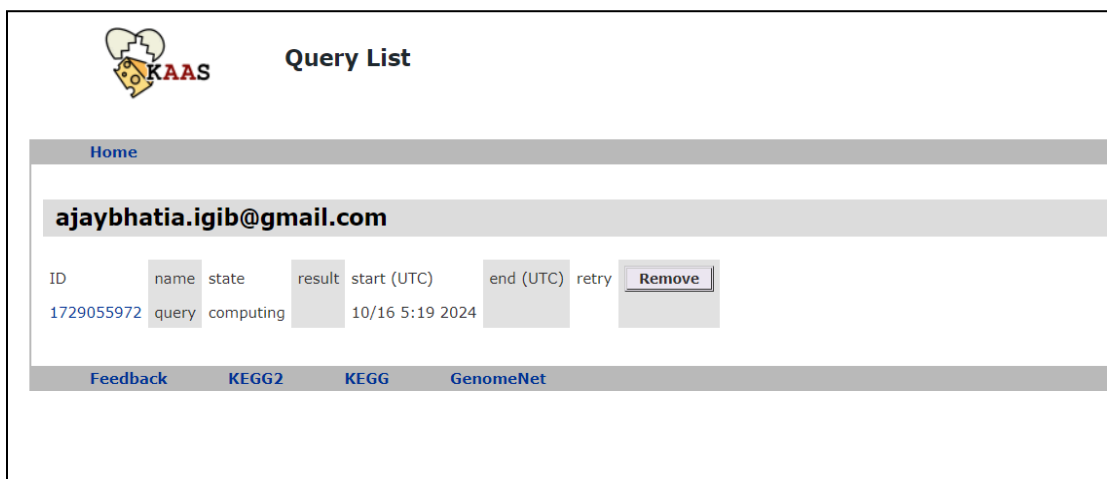
Best regards
GenomeNet Team

If you have any questions, please use this form.
<https://www.genome.jp/feedback/>

- You will receive an email from KASS after few minutes, open the email and click on Submit link

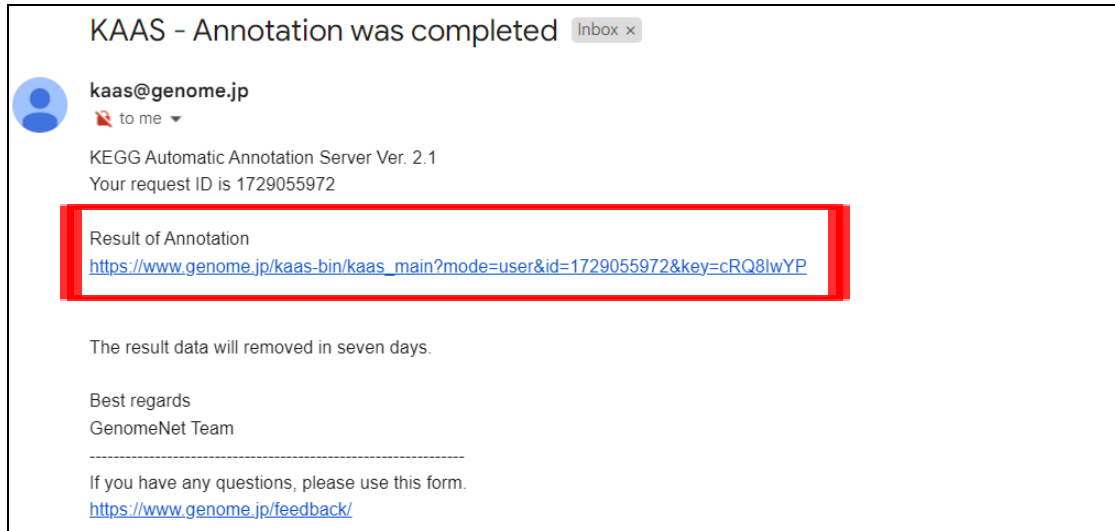


- You will receive the next email from KASS, click the provided link in this email and you will get result page of your submitted job for annotation.

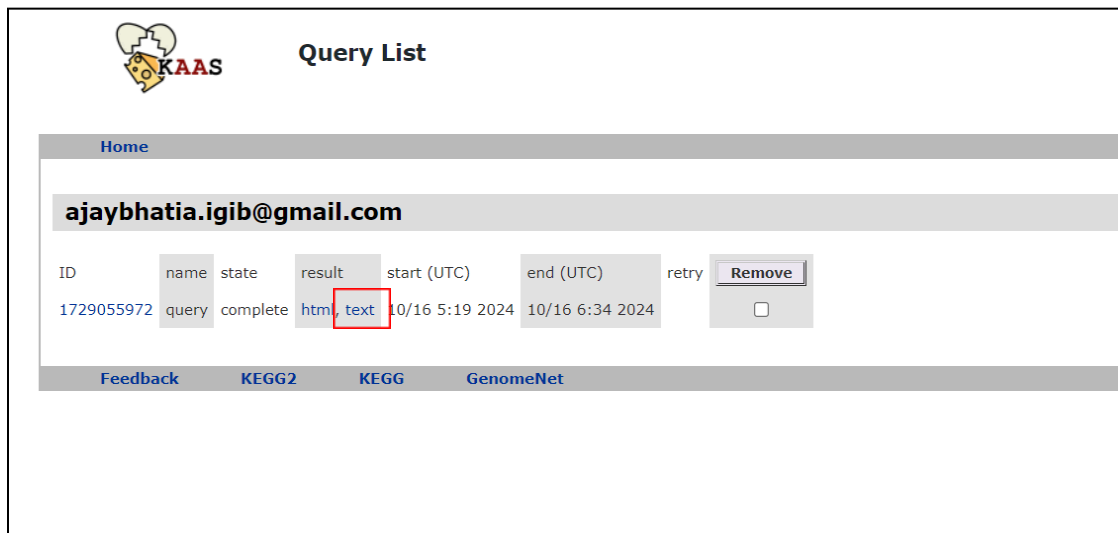


- Your result page will look like this.
- It can take a few hours to complete.
- You will receive another email from KASS when it is completed.

* I already upload the result of this KASS annotation job in github with name query.ko, you can access this from repository https://github.com/ajigib/IV_Workshop



- You will receive a final email from KASS when the KEGG annotation is completed.
- Click on the provided link in the email to get the annotation result.

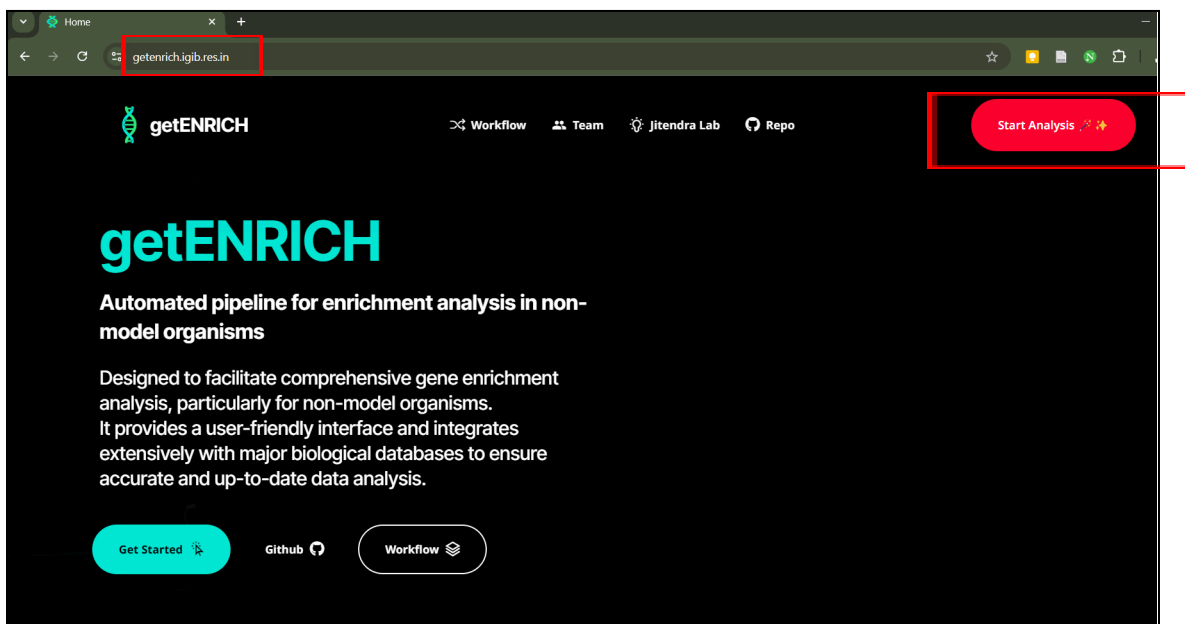


- In the KAAS annotation result page, you will get two links, one is for HTML file and other is for text file.
- Click on the text file, your result will be downloaded with name query.ko in the Download folder.
- You can move the query.ko file from the Download folder to your desired folder in which you want to do all the analysis.

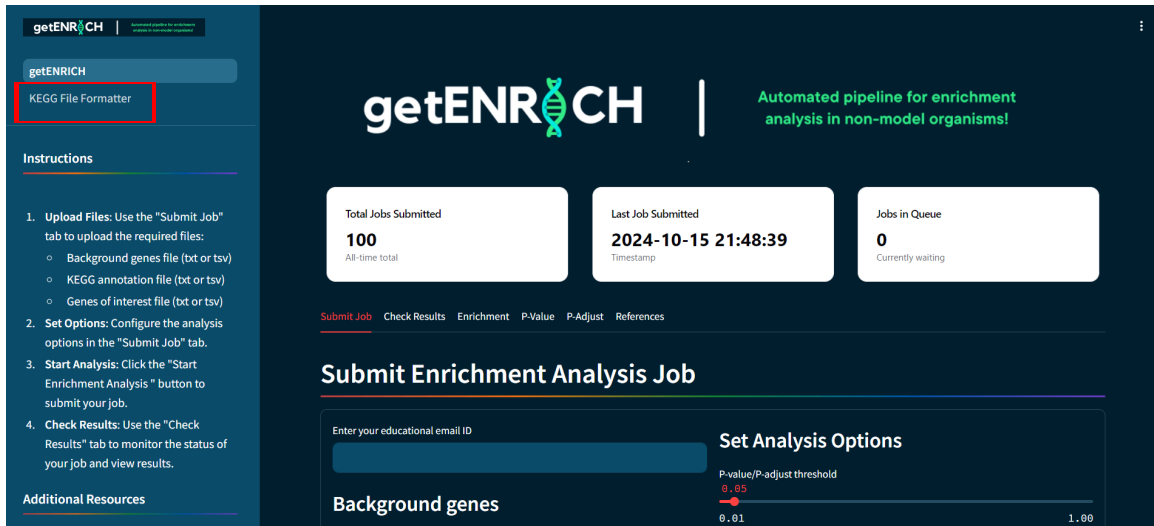
2. Making Of Input Files For getENRICH Tool:

- The getENRICH tool takes three input files: list of foreground genes, list of background genes and KEGG annotation file to all the genes (in getENRICH specific file format).
- You can make getENRICH specific KEGG annotation format in the getENRICH tool by simply uploading the query.ko file.
- You can select the list of foreground genes and background genes according to your research requirements.

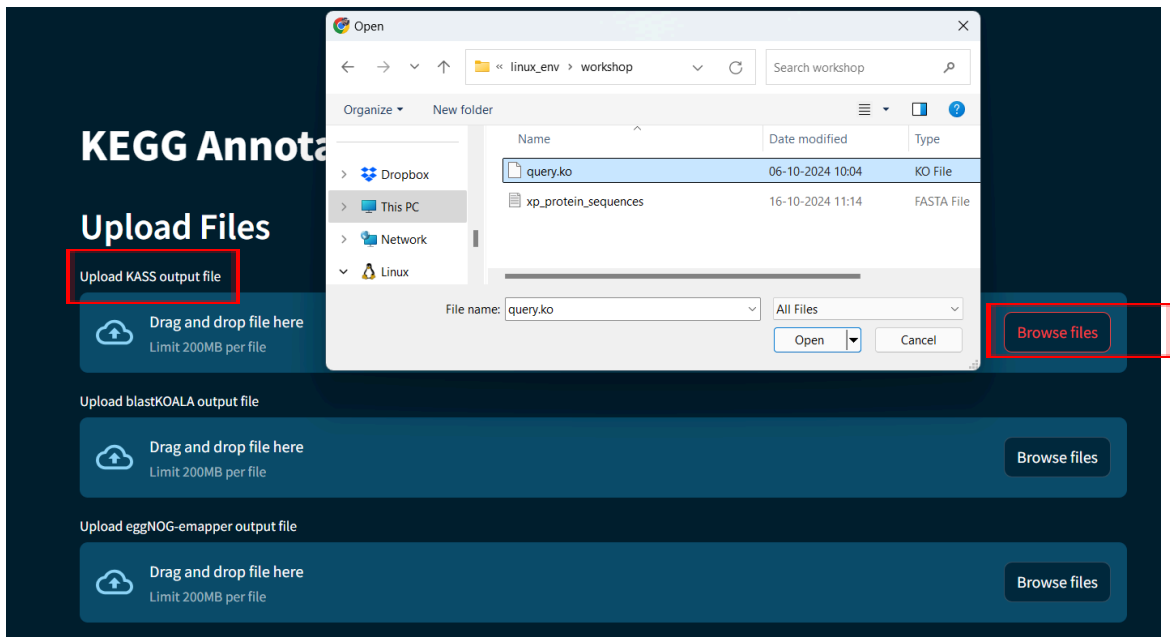
* I upload the ready-made three input files for enrichment analysis in github with name [background.txt](#), [foreground.txt](#) and [kegg_annotationTOgenes.txt](#) you can access this from repository https://github.com/ajigib/IV_Workshop



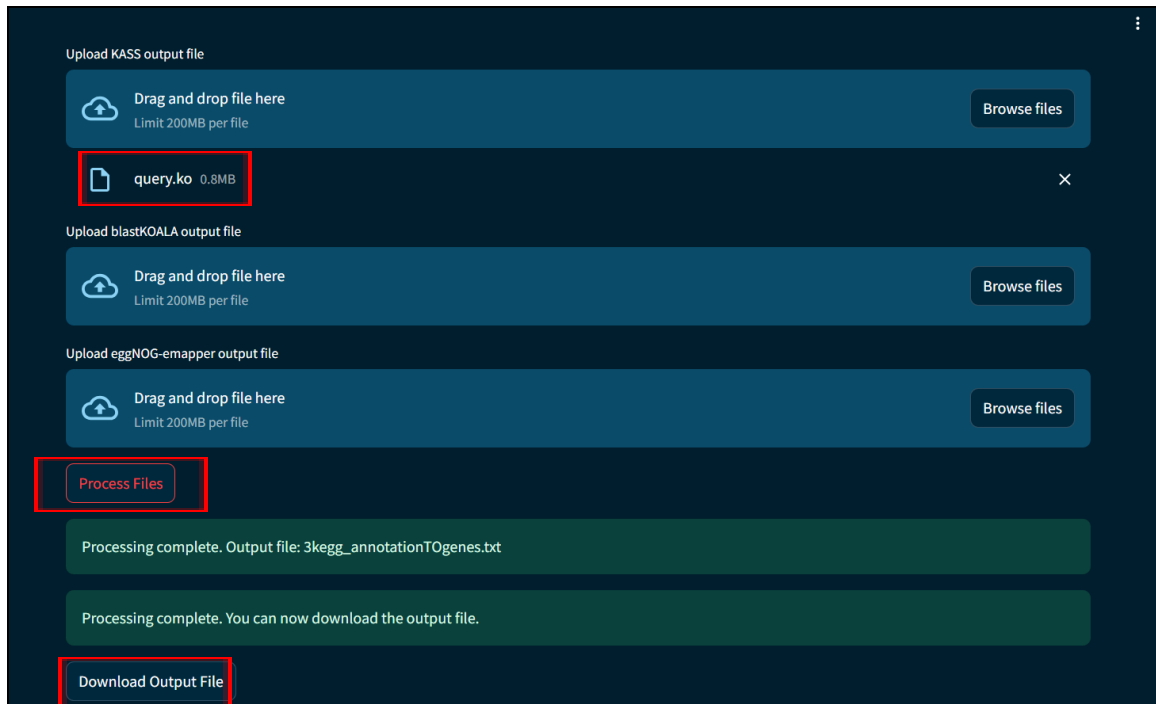
- Type getenrich.igib.res.in web address bar or click on link <https://getenrich.igib.res.in>
- The home page to getENRICH tool will open like this.
- Click on the Start Analysis button in red color.



- This page of the getENRICH tool will open.
- Click on KEGG File Formatter to convert the query.ko file into getENRICH specific format.



- Here you can convert the annotation files of KAAS, blastKOALA and eggNOG-emapper into the getENRICH specific file format.
- We have the output of KASS, so click on Browse files in Upload KASS output file section and select the query.ko file from your file explorer.
- The query.ko file is output of KAAS annotation job, if you don't have the query.ko file, you can access [query.ko](https://github.com/ajigib/IV_Workshop) from github repository https://github.com/ajigib/IV_Workshop



- After upload the query.ko file click on Process files button,
- Your file will convert into getENRICH annotation file format.
- A new button (Download Output File) will appear in a few seconds.
- Click on the Download Output File button, and your getENRICH annotation file will download into your Download folder with the name 3kegg_annotationTOgenes.txt.
- You can move the annotation file from the Download folder to the folder where you want to do enrichment analysis.

3. Start KEGG Pathway Enrichment Analysis:

- Now you have the KEGG annotation file for the getENRICH tool.
- You select the list of foreground genes and background genes according to your research needs.

* I upload the ready-made three input files for enrichment analysis in github with name [background.txt](#), [foreground.txt](#) and [kegg_annotationTOgenes.txt](#) you can access this from repository https://github.com/ajigib/IV_Workshop

Background genes

Drag and drop file here
Limit 200MB per file • TXT, TSV

background.txt 396.5KB

KEGG annotation

Drag and drop file here
Limit 200MB per file • TXT, TSV

kegg_annotationTOgenes.txt 0.6MB

Genes of interest

Drag and drop file here
Limit 200MB per file • TXT, TSV

foreground.txt 11.7KB

Your analysis results will be stored on our servers for up to 7 days. Please download any important data within this timeframe.

Start Enrichment Analysis

Prepare KEGG Annotation

Job Submitted ID to the server: **ba66cb00-c497-4e46-94ad-47f043088a37**

You may copy the ID to check the progress in Check Results tab

Select Analysis Option

- ☒ Default: Basic analysis report without additional visualizations.
- ☐ P-Value Focus: Emphasizes p-value-based significant pathways.
- ☐ P-Adjust Focus: Prioritizes adjusted p-value significant pathways
- ☐ Comprehensive

P-Value Options

- ☐ Generate pubMed trends plot
- ☒ Generate heatmap
- ☒ Generate upset plot
- ☒ Generate tree plot
- ☒ Generate KEGG pathway diagrams

P-Adjust Options

- ☐ Generate pubMed trends plot
- ☐ Generate heatmap
- ☐ Generate upset plot
- ☐ Generate tree plot
- ☐ Generate KEGG pathway diagrams

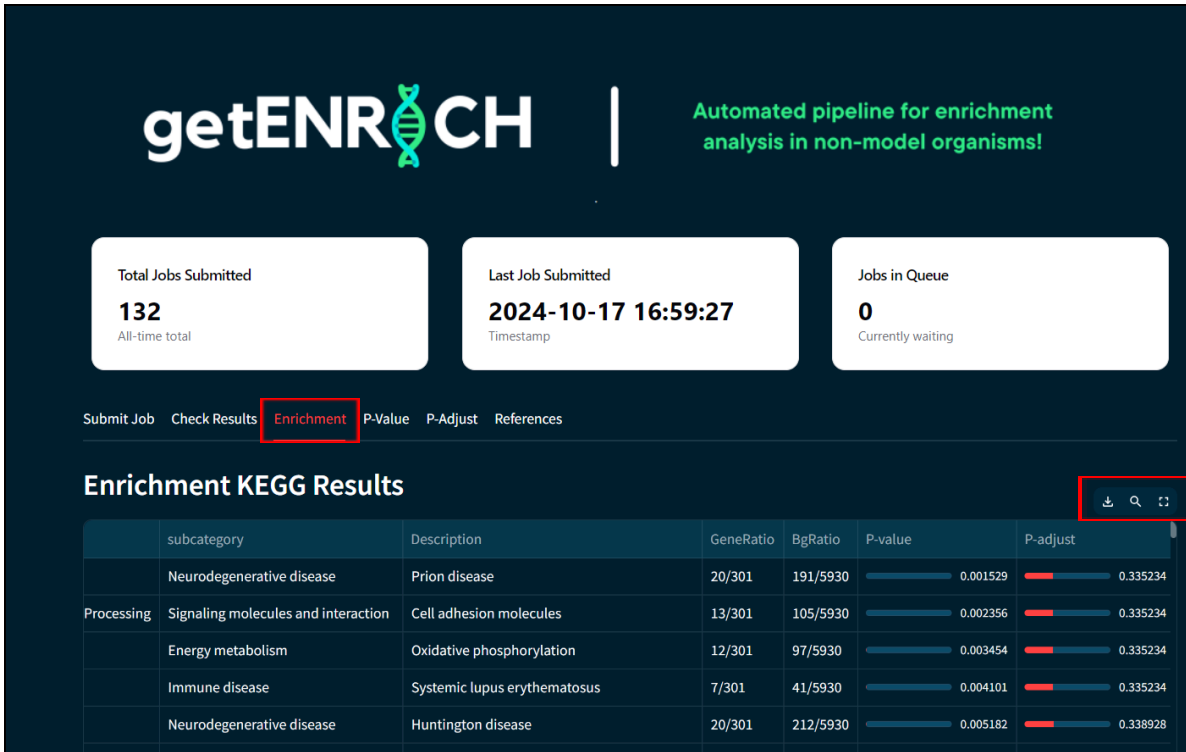
- Upload background.txt, kegg_annotationTOgene.txt and foreground.txt file in dedicated upload boxes.
- Default value of significance is 0.05, which can be increased by sliding the red line to the right word (not recommended).
- In Select Analysis Option, you have four options, selected according to your need. To save time, in this tutorial we select the Default option with heatmap, upset plot, tree plot and KEGG pathway diagrams in P-Value Option.
- The pubMed tree plot is taking very much time to form because it fetches the information from all the research papers of the past 10 years in the pubMed database. So use this option only when you need it.
- Now click on Start Enrichment Analysis.
- In a few seconds, a job ID will be generated, this is a unique id for your enrichment analysis job (written after Job Submitted ID to the server).
- Save this ID for future reference.

The screenshot shows the getENRICH web application interface. At the top, the logo 'getENRICH' is displayed next to the tagline 'Automated pipeline for enrichment analysis in non-model organisms!'. Below this, two white boxes provide statistics: 'Total Jobs Submitted: 132 (All-time total)' and 'Last Job Submitted: 2024-10-17 16:59:27 (Timestamp)'. A navigation bar contains tabs: 'Submit Job', 'Check Results' (highlighted with a red box), 'Enrichment', 'P-Value', 'P-Adjust', and 'References'. The main section is titled 'Check Results'. It features a text input field labeled 'Enter Previously Submitted Job ID' containing the ID 'ba66cb00-c497-4e46-94ad-47f043088a37' (highlighted with a red box). Below the input is a red button labeled 'Check Job Status and View Results' (also highlighted with a red box). At the bottom, a status message in a blue box reads: 'Job is in progress: [2024-10-17 17:01:04] Generating Treeplot of p-value Significant Enriched Pathways' (highlighted with a red box).

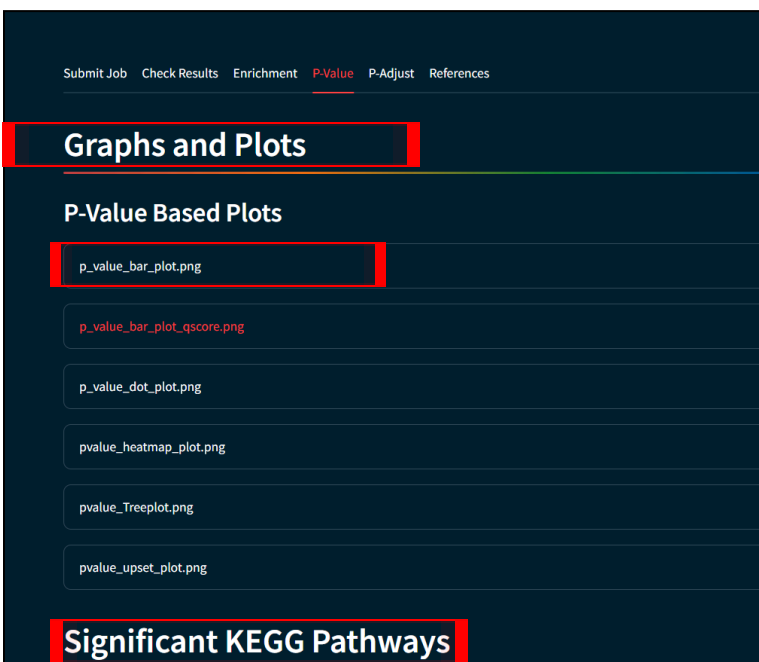
- Click on the Check Result tab.
- Paste the copied job ID inside the box.
- Click on Check Job Status and View Results, a message will pop-up below the red button, this will update you the real time status of your submitted job when you click this button.

This screenshot shows the 'Check Results' section of the getENRICH application after a successful job completion. The 'Enter Previously Submitted Job ID' field still contains the ID 'ba66cb00-c497-4e46-94ad-47f043088a37'. The red 'Check Job Status and View Results' button is present. Below it, a green success message box states: 'Job completed successfully! Kindly verify report in Enrichment, P-Adjust, P-Value tabs above 🍌'. At the bottom, there is a button labeled 'Download Report Data'.

- Job status message will turn into Job completed successfully when your job is completed. But don't click on Download Report Data before exploring the Enrichment tab.



- In Enrichment tab, you can review the result of enriched pathway in tabular format
- You can check the result from here before downloading this or you can download this table in csv or search from the top right side.



- In the next tab (p-value), you can explore the plots and diagrams of enriched pathways by clicking drop-down options.

getENRICH | Automated pipeline for enrichment analysis in non-model organisms!

Total Jobs Submitted
132
All-time total

Last Job Submitted
2024-10-17 16:59:27
Timestamp

Submit Job Check Results Enrichment P-Value P-Adjust References

Check Results

Enter Previously Submitted Job ID

ba66cb00-c497-4e46-94ad-47f043088a37

Check Job Status and View Results

✓ Job completed successfully! Kindly verify report in Enrichment, P-Adjust, P-Value tabs above 🙌

Download Report Data

- Now you can download the result of getENRICH by clicking the button Download Report Data.
- A zipped file (enrichment_output) will be downloaded into your Download folder.
- Extract the zipped file to get the KEGG pathway enrichment result.

* In case you don't get the enrichment_output file, i upload the result of same input data in [enrichment_output](#) folder of github repository https://github.com/ajigib/IV_Workshop