SUPPLEMENTARY INFORMATION

IVAG: An Integrative Visualization Application for Various Types of Genomic Data Based on R-Shiny and the Docker Platform

Tae-Rim Lee[§], Jin Mo Ahn[§], Gyuhee Kim[§], Sangsoo Kim*

Department of Bioinformatics and Life Science, Soongsil University, Seoul 06978, Korea

http://www.genominfo.org/src/sm/gni-15-178-s001.pdf

Supplementary Table 1. R packages used in IVAG

Name	Reference
shiny	Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2017). shiny: Web Application Framework for R. R package version 1.0.5. https://CRAN.R-project.org/package=shiny
shinydashboard	Chang W, Ribeiro BB (2017). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.6.1. https://CRAN.Rproject.org/package=shinydashboard
shinycssloaders	Sali A (2017). shinycssloaders: add CSS loading animations to 'shiny' outputs. R package version 0.2.0. https://CRAN.R-project.org/package=shinycssloaders
shinyjs	Attali D (2017). shinyjs: easily improve the user experience of your shiny apps in seconds. R package version 0.9.1. https://CRAN.R-project.org/package=shinyjs
RColorBrewer	Neuwirth E (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer
gplots	Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WH, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2016). gplots: various R programming tools for plotting data. R package version 3.0.1. https://CRAN.R-project.org/package=gplots
ggplot2	Wickham H. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, 2009
easyGgplot2	Kassambara A (2014). easyGgplot2: perform and customize easily a plot with ggplot2. R package version 1.0.0.9000. http://www.sthda.com
rmarkdown	Allaire JJ, Cheng J, Xie Y, McPherson J, Chang W, Allen J, Wickham H, Atkins A, Hyndman R, Arslan R (2017). rmarkdown: dynamic documents for R. R package version 1.6. https://CRAN.R-project.org/package=rmarkdown
rtracklayer	Lawrence M, Gentleman R, Carey V. "rtracklayer: an {R} package for interfacing with genome browsers". *Bioinformatics* 2009;25:1841-1842.
DT	Xie Y (2016). DT: a wrapper of the JavaScript library 'DataTables'. R package version 0.2. https://CRAN.R-project.org/package=DT
data.table	Dowle M, Srinivasan A (2017). data.table: extension of `data.frame`. R package version 1.10.4. https://CRAN.R-project.org/package=data.table
gridExtra	Auguie B (2017). gridExtra: miscellaneous functions for "Grid" graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
GenomicFeatures	Lawrence M, Huber W, Pagés H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9:e1003118
edgeR	Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140
goseq	Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 2010;11:R14
pheatmap	Kolde R (2015). pheatmap: Pretty Heatmaps. R package version 1.0.8. https://CRAN.R-project.org/package=pheatmap
plotly	Sievert C, Parmer C, Hocking T, Chamberlain S, Ram K, Corvellec M, Despouy P (2017). plotly: create interactive web graphics via 'plotly.js'. R package version 4.7.1. https://CRAN.R-project.org/package=plotly
manhattanly	Bhatnagar S (2016). manhattanly: interactive Q-Q and Manhattan plots using 'plotly.js'. R package version 0.2.0. https://CRAN.R-project.org/package=manhattanly
LDheatmap	Shin JH, Blay S, McNeney B, Graham J (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw* 2006;16:Code Snippet 3
GWASTools	Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, et al. GWASTools: an R/Bioconductor package for quality control and analysis of

	genome-wide association studies. Bioinformatics 2012;28:3329-3331
genetics	Warnes G, with contributions from Gregor Gorjanc, Leisch F, Man M (2013). Genetics: population genetics. R package version 1.3.8.1. https://CRAN.R-project.org/package=genetics
org.At.tair.db	Carlson M (2015). org.At.tair.db: Genome wide annotation for Arabidopsis. R package version 3.2.3.

Supplementary Table 2. Plugins attached in JBrowse

Name	Reference
gccontent	https://github.com/elsiklab/gccontent
narrowpeak	https://github.com/cmdcolin/narrowpeak
gwasviewer	https://github.com/elsiklab/gwasviewer
bedGraph to BigWig	https://github.com/ENCODE-DCC/kentUtils

Supplementary Table 3. Example data source

Name	Reference
RNA-seq single factor	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4298/
RNA-seq multiple factor	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4243/
GWAS and LD analysis	Tutorial data of R package GAPIT http://zzlab.net/GAPIT/GAPIT_Tutorial_Data.zip numeric to vcf format conversion was carried out using TASSEL

<Supplementary material 2>

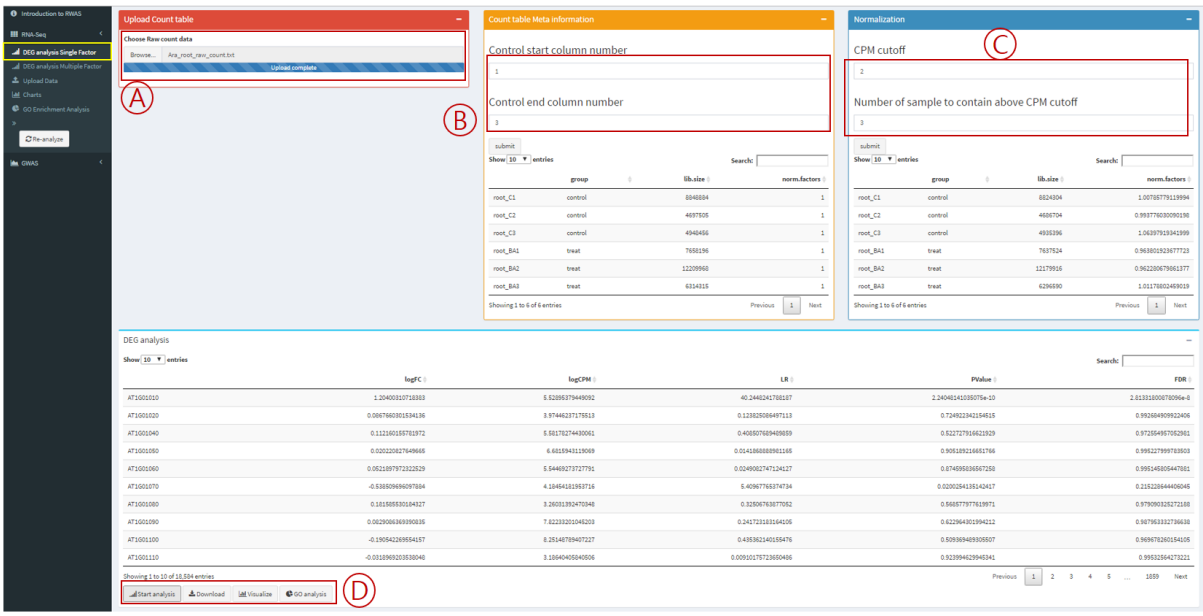
Index

1. RNA-Seq

2. GWAS

3. Genome Browser

<1.1 DEG analysis for single factor>



<Figure 1. DEG single factor analysis page >

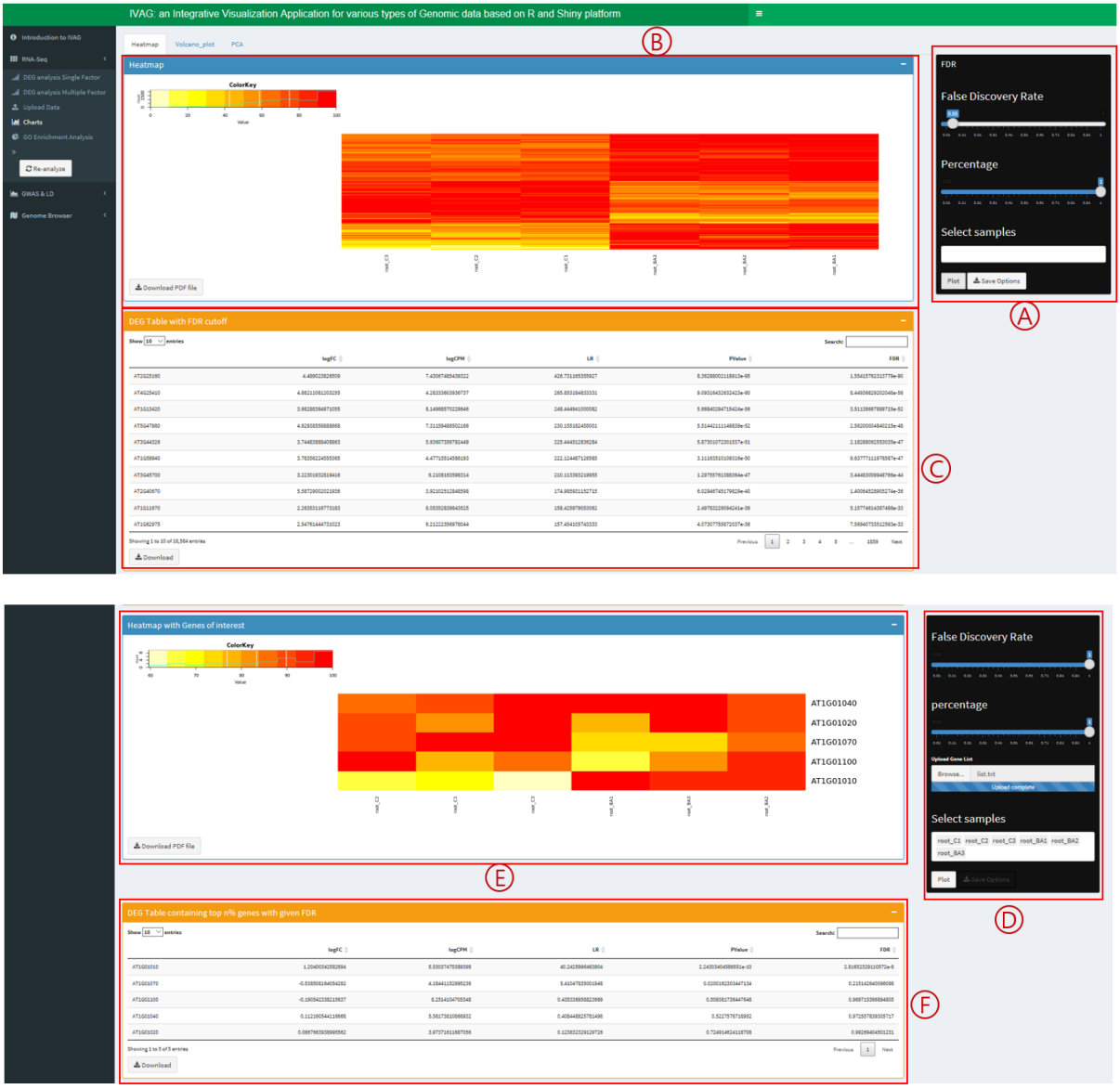
Ⓐ: Upload count table generated using htseq-count or similar software after mapping RNA-Seq data to reference Genome.

Ⓑ: Specify the range of control columns.

Ⓒ: In order to analyze differentially expressed genes, each genes need at least 6~10 counts. However, in raw count table, there are lots of genes with 0 counts which need to be filtered out. We use count per million to filter out genes with low or no counts. CPM cutoff criteria filters gene out using the smallest library size. For example, if the smallest library size was 4,000,000, CPM cutoff 2 would filter out genes with less than 8 counts. Number of samples to contain above CPM cutoff criteria can be specified as to how many samples must satisfy above filtering criteria. For example, raw count table with samples having 3 replicates each, if we specify Number of sample to contain above CPM as 3, at least 3 samples must meet the filtering criteria in order to proceed.

Ⓓ: Clicking the Start analysis button will run DEG analysis and produce the results as a table. Users can click Download button to get the results as text file format. Visualization button will lead the users to a page where they can generate Heatmap, Volcano plot, PCA analysis plot with the DEG analysis result. Clicking the GO analysis button will lead to a page where Gene Ontology Enrichment analysis can be done.

<1.2 DEG Visualization>



<Figure 2. Heatmap visualization>

Ⓐ: Specify False Discovery Rate to filter out genes to be used to draw heatmap. Percentage can be set to draw heatmap with top n% of filtered genes. User can specify samples to be drawn on heatmap.

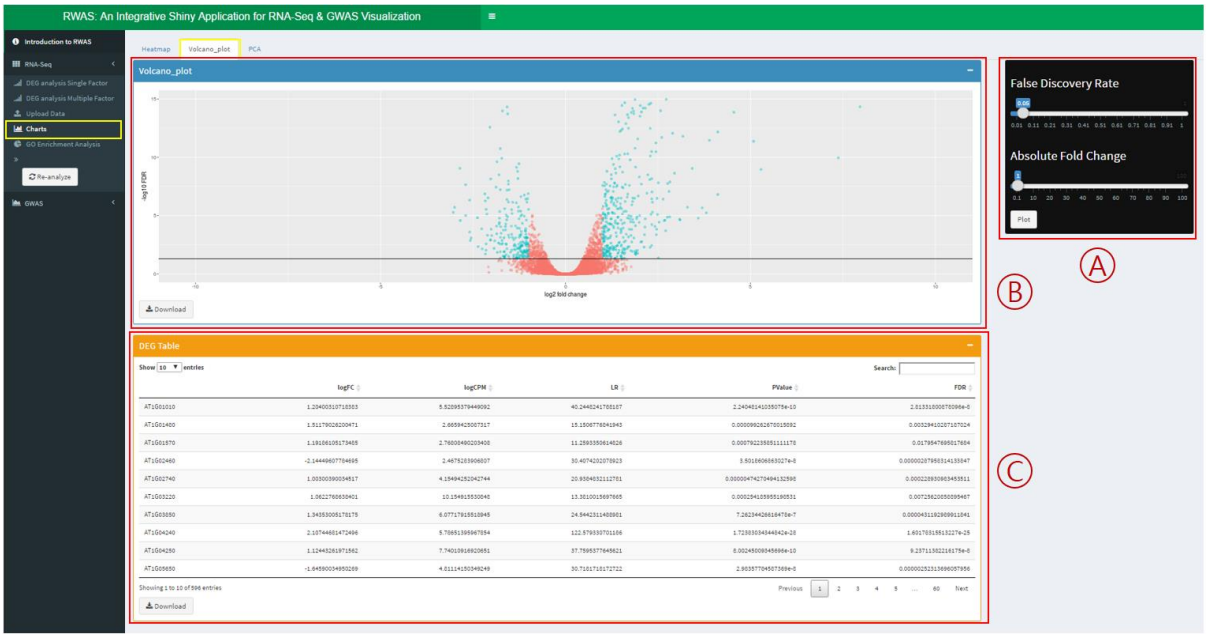
Ⓑ: Generated heatmap

Ⓒ: Shows DEG analysis table of genes used to create heatmap. Click Download button to download table in text file format.

Ⓓ: Specify False Discovery Rate to filter out genes to be used to draw heatmap. Percentage can be set to draw heatmap with top n% of filtered genes. User can specify samples to be drawn on heatmap. Gene list can be uploaded to draw heatmap with genes of interest.

Ⓔ: Generated heatmap.

Ⓕ: Shows DEG analysis table of genes used to create heatmap. Click Download button to download table in text file format.

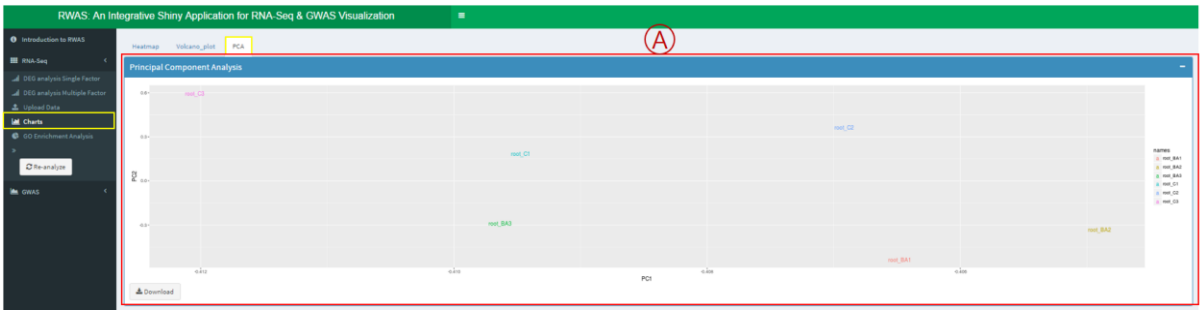


<Figure 3. Volcano plot visualization>

Ⓐ: User can specify False Discovery Rate or Absolute Fold Change to generate Volcano plot.

Ⓑ: Generates volcano plot by plotting logf fold change versus –log10(False Discovery Rate). User can download generated plot in PDF format by clicking the Download button.

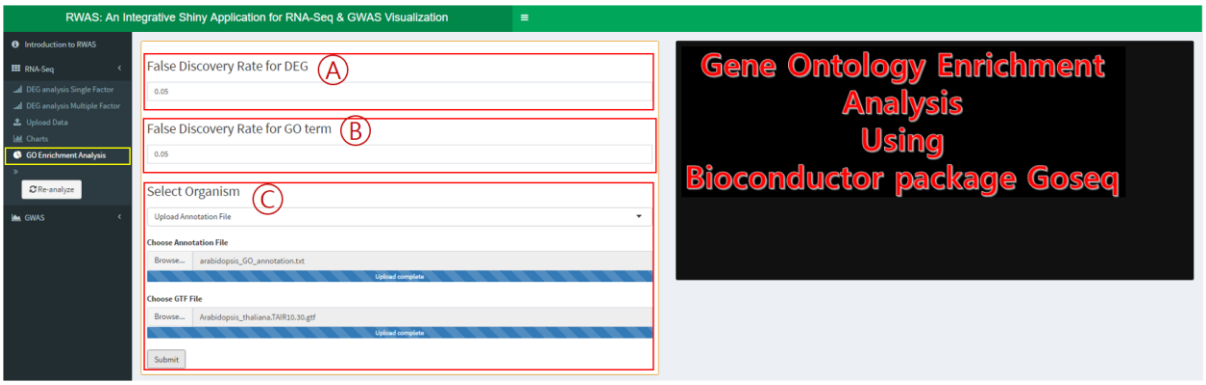
Ⓒ: DEG analysis result table for genes used to plot volcano plot is shown. This table can be downloaded in text format by clicking the Download button.

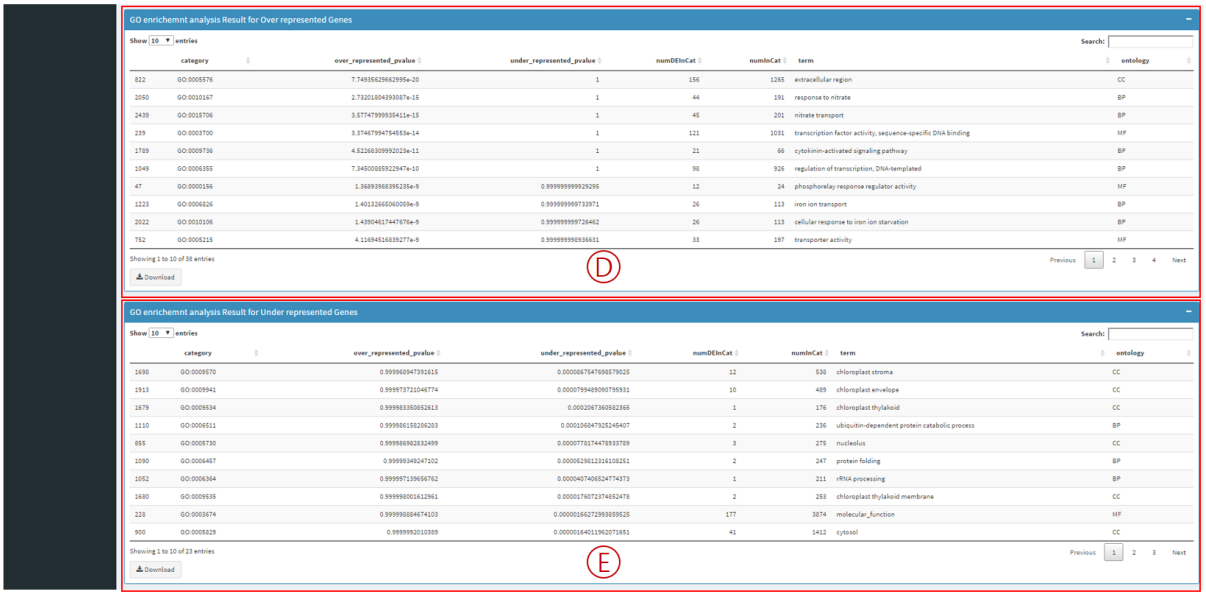


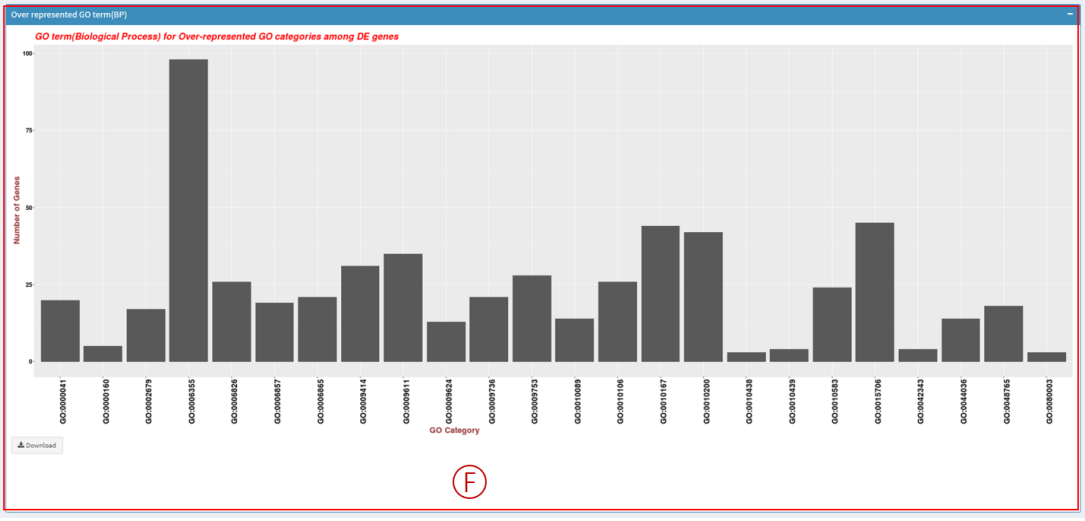
<Figure 4. PCA plot visualization>

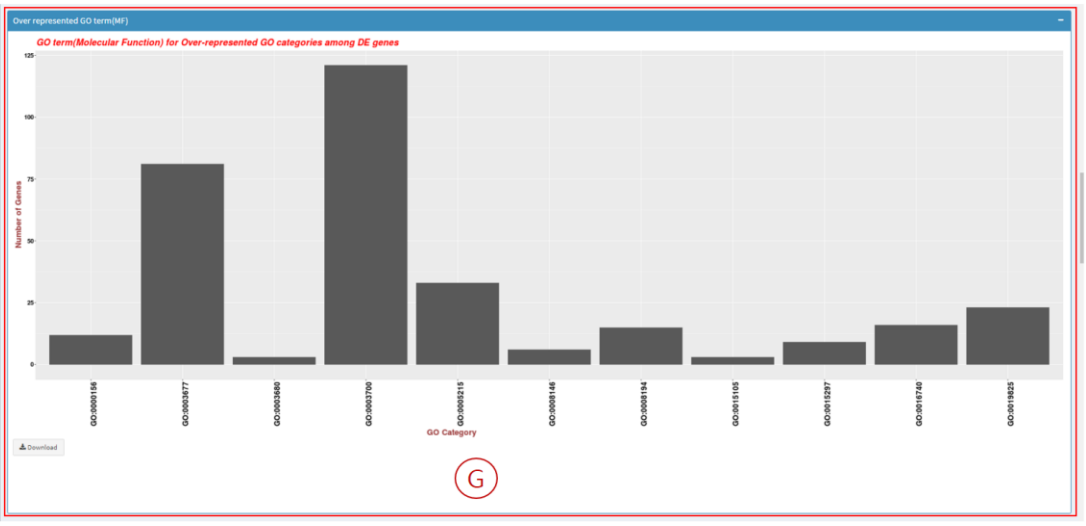
Ⓐ: Generates Principal analysis plot. This can be downloaded in PDF format by clicking the Download button.

<1.3 Gene Ontology Enrichment Analysis>

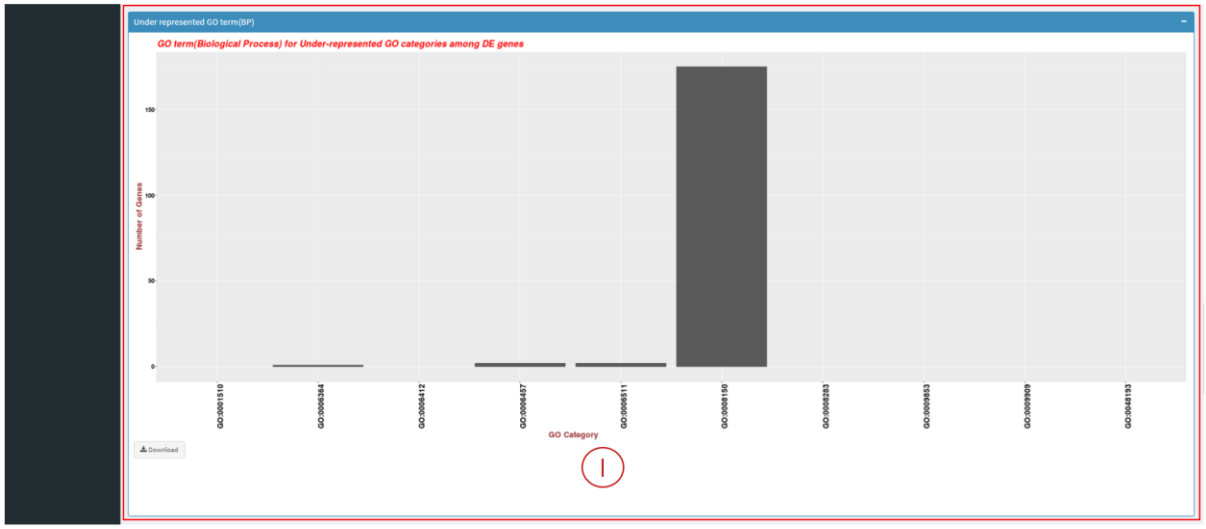


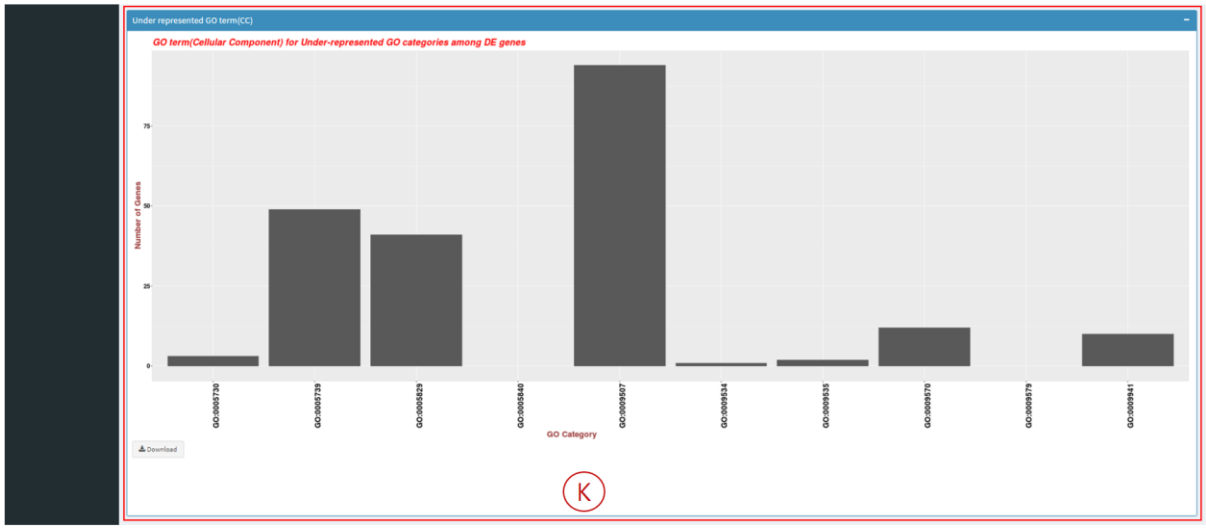



Over represented GO term(BP)
GO term(Biological Process) for Over-represented GO categories among DE genes
Number of Genes
GO Category
Download
F


Over represented GO term(MF)
GO term(Molecular Function) for Over-represented GO categories among DE genes
Number of Genes
GO Category
Download
G


Over represented GO term(CC)
GO term(Cellular Component) for Over-represented GO categories among DE genes
Number of Genes
GO Category
Download
H





<Figure 5. Gene Ontology Enrichment analysis page>

Ⓐ: Specify filtering criteria(False Discovery Rate) for differentially expressed genes which will go through Gene Ontology Enrichment analysis.

Ⓑ: Specify False Discovery Rate to be used in Gene Ontology Enrichment analysis.

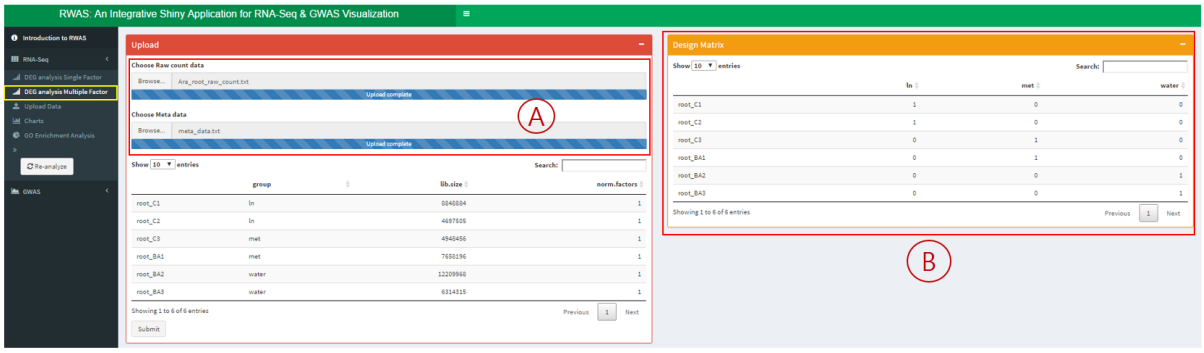
Ⓒ: Upload Gene Ontology annotation file and GTF file.

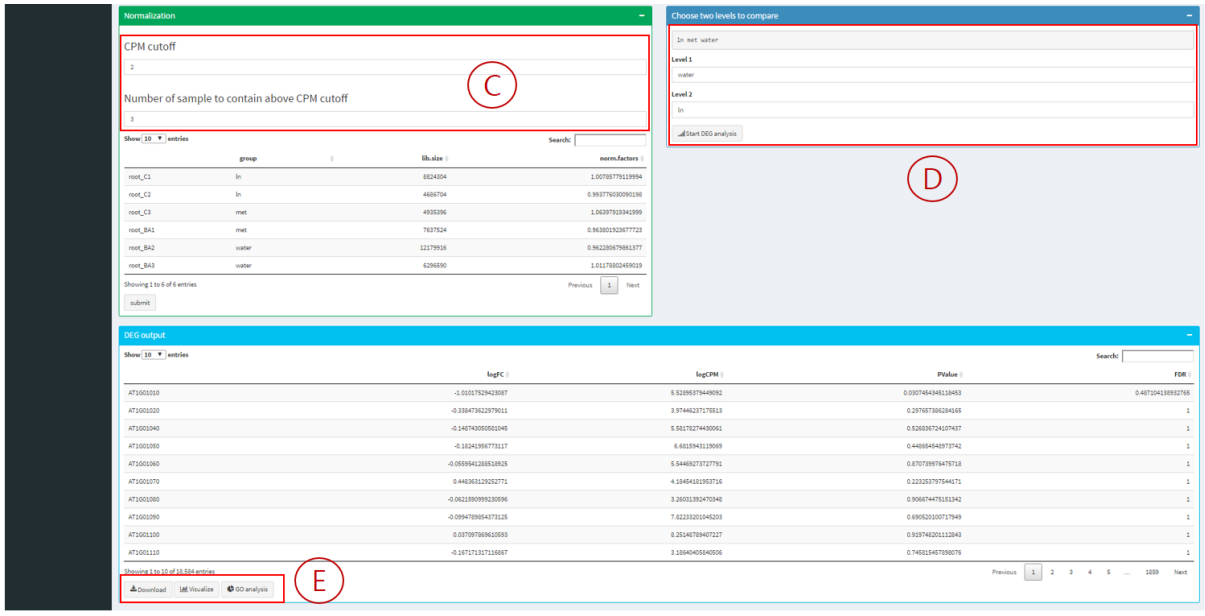
Ⓓ: Shows over represented Gene Ontology Enrichment analysis result among differentially expressed genes. 5th column(numInCat) shows how many genes are allocated to the specific category and 4th column(numDEInCat) shows among those genes, how many of them are differentially expressed. Users can click Download button to retrieve the results in text file format.

Ⓔ: Shows under represented Gene Ontology Enrichment analysis result among differentially expressed genes. 5th column(numInCat) shows how many genes are allocated to the specific category and 4th column(numDEInCat) shows among those genes, how many of them are differentially expressed. Users can click Download button to retrieve the results in text file format.

Ⓕ ~ Ⓚ: Shows the histogram of differentially expressed genes with specific Gene Ontology term and ontology such as biological process, cellular component and molecular function. These graphs can be downloaded in PDF format by clicking the Download button.

<1.4 DEG analysis for multiple factor>





<Figure 6. Multi factor DEG analysis multi factor page>

Ⓐ: Upload count table generated using htseq-count or similar software after mapping RNA-Seq data to reference Genome with meta information file.

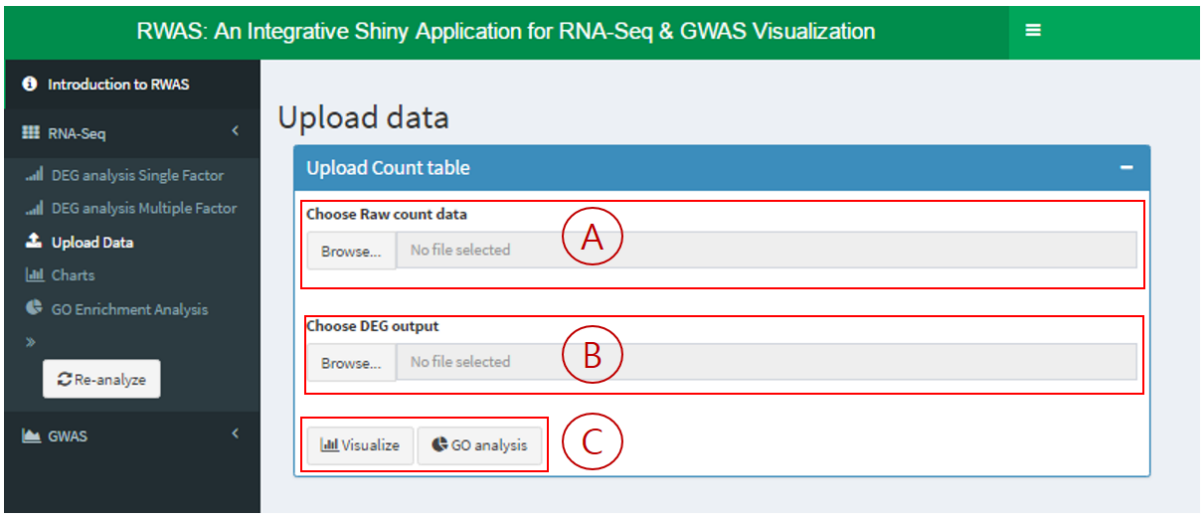
Ⓑ: Generates design matrix when all the files are uploaded.

Ⓒ: In order to analyze differentially expressed genes, each genes need at least 6~10 counts. However, in raw count table, there are lots of genes with 0 counts which need to be filtered out. We use count per million to filter out genes with low or no counts. CPM cutoff criteria filters gene out using the smallest library size. For example, if the smallest library size was 4,000,000, CPM cutoff 2 would filter out genes with less than 8 counts. Number of samples to contain above CPM cutoff criteria can be specified as to how many samples must satisfy above filtering criteria. For example, raw count table with samples having 3 replicates each, if we specify Number of sample to contain above CPM as 3, at least 3 samples must meet the filtering criteria in order to proceed.

Ⓓ: Among multiple samples, choose two samples to proceed DEG analysis.

Ⓔ: Clicking the Start analysis button will run DEG analysis and produce the results as a table. Users can click Download button to get the results as text file format. Visualization button will lead the users to a page where they can generate Heatmap, Volcano plot, PCA analysis plot with the DEG analysis result. Clicking the GO analysis button will lead to a page where Gene Ontology Enrichment analysis can be done.

<1.5 DEG visualization and Gene Ontology Enrichment analysis for pre-analyzed data>



<Figure 7. Pre-analyzed data upload page>

Ⓐ: Upload count table generated using htseq-count or similar software after mapping RNA-Seq data to reference Genome.

Ⓑ: Upload DEG analysis result table.

Ⓒ: Visualization button will lead the users to a page where they can generate Heatmap, Volcano plot, PCA analysis plot with the DEG analysis result. Clicking the GO analysis button will lead to a page where Gene Ontology Enrichment analysis can be done.

<1.6 Input files>

```
ln1 ln2 meth1   meth2   water1  water2
AT1G01010   440 366 419 625 334 482
AT1G01020   1095    1030    856 1447    759 1183
AT1G01030   0   0   1   0   0   2
AT1G01040   2455    1706    2672    4204    1834    3098
AT1G01046   11  10  20  52  17  30
AT1G01050   11180   11030   10013   16152   8180    11682
AT1G01060   652 473 800 1410    642 996
AT1G01070   109 97  91  119 58  113
AT1G01073   0   0   0   0   0   0
AT1G01080   2209    1951    2116    3061    1655    2390
AT1G01090   8647    8519    7271    10956   5500    8672
AT1G01100   15938   13942   10564   16053   8958    12852
AT1G01110   0   0   0   0   0   0
AT1G01115   0   0   0   0   0   0
AT1G01120   8   1   15  17  6   19
AT1G01130   68  65  51  62  43  60
AT1G01140   478 463 624 1003    529 829
AT1G01150   8   1   21  19  12  6
AT1G01160   3483    2737    3300    4585    2468    3396
AT1G01170   5226    4862    5953    7873    4204    5819
AT1G01180   5   3   15  34  20  19
AT1G01183   0   0   0   0   0   0
AT1G01190   0   3   4   3   0   1
AT1G01200   26  24  11  5   2   8
AT1G01210   604 628 539 898 441 645
AT1G01220   3277    2761    4580    7004    3652    5094
AT1G01225   779 686 810 1294    754 1085
AT1G01230   2199    2120    1979    2881    1673    2318
AT1G01240   310 304 1098    1665    922 1442
AT1G01250   2   0   0   1   0   0
AT1G01260   1340    1136    1258    1943    1035    1632
AT1G01270   0   0   0   0   0   0
AT1G01280   0   0   0   0   0   0
```

A

```
ln
ln
met
met
water
water
```

B

```
"logFC" "logCPM" "LR" "PValue" "FDR"
"AT1G01010" 1.19688737736099 5.52688808335964 39.3089106372121 3.61786523882999e-10 4.80222720383886e-08
"AT1G01020" 0.0822804977656206 3.97165716089378 0.111434656713485 0.738516674183035 0.992950029705138
"AT1G01030" -0.162468882715308 0.0823715495057947 0.0610995702734676 0.804766390210944 0.994580412175897
"AT1G01040" 0.106377858075599 5.57910426454308 0.37042230935759 0.542774163320339 0.974600223816041
"AT1G01050" 0.0153574773648363 6.67932177116587 0.00814645735631103 0.92808241163436 0.995983980196374
"AT1G01060" 0.0456180046924779 5.54156373113906 0.019399476207294 0.889227311214195 0.995931204225289
"AT1G01070" -0.544012070552485 4.18302675004016 5.48539400852262 0.0191759975507977 0.211398693538396
"AT1G01080" 0.174667523566418 3.26203485385487 0.299539868880617 0.584171025511136 0.978544418926473
"AT1G01090" 0.0770064556487513 7.82001219000772 0.206235076914759 0.649734453606224 0.987057549001829
"AT1G01100" -0.194374421491345 8.24999782635235 0.450922121063243 0.501897378775529 0.967135588725176
"AT1G01110" -0.0356166922910766 3.18109685561797 0.011510637873462 0.914560789141767 0.995983980196374
"AT1G01120" 0.641291331519332 6.3387621639768 11.0961455503092 0.000865073303584205 0.0196921958851816
"AT1G01130" 0.467864553000374 1.76456938585317 0.977622941476459 0.322786369058292 0.904971916676201
"AT1G01140" -0.542815616073306 2.63788851959664 2.13563384343845 0.143910966721881 0.68944588214333
"AT1G01160" -0.0296144956218807 5.15484537653471 0.0203745510116882 0.88649595793983 0.995931204225289
"AT1G01170" 0.0959514365512992 3.64734295039377 0.15817628576344 0.69084114621376 0.987813981559841
"AT1G01180" 0.24596413317094 3.7881988269758 0.650354637870404 0.419985915804817 0.950470048855293
"AT1G01190" -0.375672462290263 3.86317903951983 0.635708343635596 0.425269154293829 0.951393075845317
"AT1G01200" 0.246507464818891 3.01178043499523 0.779877708445121 0.377178543497332 0.933754390534284
```

C

```
AT1G01010	GO:0006888
AT1G01010	GO:0007275
AT1G01010	GO:0043090
AT1G01020	GO:0006665
AT1G01020	GO:0006665
AT1G01020	GO:0016125
AT1G01020	GO:0016125
AT1G01030	GO:0006355
AT1G01030	GO:0006355
AT1G01030	GO:0006355
AT1G01030	GO:0009908
AT1G01030	GO:0048366
AT1G01040	GO:0000226
AT1G01040	GO:0000226
AT1G01040	GO:0000278
AT1G01040	GO:0000911
AT1G01040	GO:0006306
AT1G01040	GO:0006306
AT1G01040	GO:0006306
AT1G01040	GO:0006342
AT1G01040	GO:0006342
AT1G01040	GO:0006342
AT1G01040	GO:0006396
AT1G01040	GO:0006396
AT1G01040	GO:0006396
AT1G01040	GO:0006396
AT1G01040	GO:0007267
AT1G01040	GO:0009616
AT1G01040	GO:0009616
AT1G01040	GO:0009616
AT1G01040	GO:0009616
AT1G01040	GO:0009630
AT1G01040	GO:0009880
AT1G01040	GO:0009908
```

D

```
#!genome-build TAIR10
#!genome-version TAIR10
#!genome-date 2010-09
#!genome-build-accession GCA_000001735.1
#!genebuild-last-updated 2010-09
1	tair	gene	3631	5899	.	+	.	gene_id "AT1G01010"; gene_version "1"; gene_name "NAC001"; gene_source "ta
1	tair	transcript	3631	5899	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; tr
1	tair	exon	3631	3913	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 3760	3913	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	start_codon 3760	3762	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; tr
1	tair	exon	3996	4276	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 3996	4276	.	+	2	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	exon	4486	4605	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 4486	4605	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	exon	4706	5095	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 4706	5095	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	exon	5174	5326	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 5174	5326	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	exon	5439	5899	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transc
1	tair	CDS 5439	5627	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; transcript
1	tair	stop_codon 5628	5630	.	+	0	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"; tr
1	tair	five_prime_utr 3631	3759	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"
1	tair	three_prime_utr 5631	5899	.	+	.	gene_id "AT1G01010"; gene_version "1"; transcript_id "AT1G01010.1"
1	tair	gene	5928	8737	.	-	.	gene_id "AT1G01020"; gene_version "1"; gene_name "ARV1"; gene_source "tair
1	tair	transcript	5928	8737	.	-	.	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; tr
1	tair	exon	8571	8737	.	-	.	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transc
1	tair	CDS 8571	8666	.	-	0	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transcript
1	tair	start_codon 8664	8666	.	-	0	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; tr
1	tair	exon	8417	8464	.	-	.	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transc
1	tair	CDS 8417	8464	.	-	0	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transcript
1	tair	exon	8236	8325	.	-	.	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transc
1	tair	CDS 8236	8325	.	-	0	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transcript
1	tair	exon	7942	7987	.	-	.	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transc
1	tair	CDS 7942	7987	.	-	0	gene_id "AT1G01020"; gene_version "1"; transcript_id "AT1G01020.1"; transcript
```

E

<Figure 8. >

Ⓐ: Raw count table / Used in DEG analysis

Ⓑ: Meta information specifying which column belongs to which group in raw count table / Used in DEG analysis multiple factor.

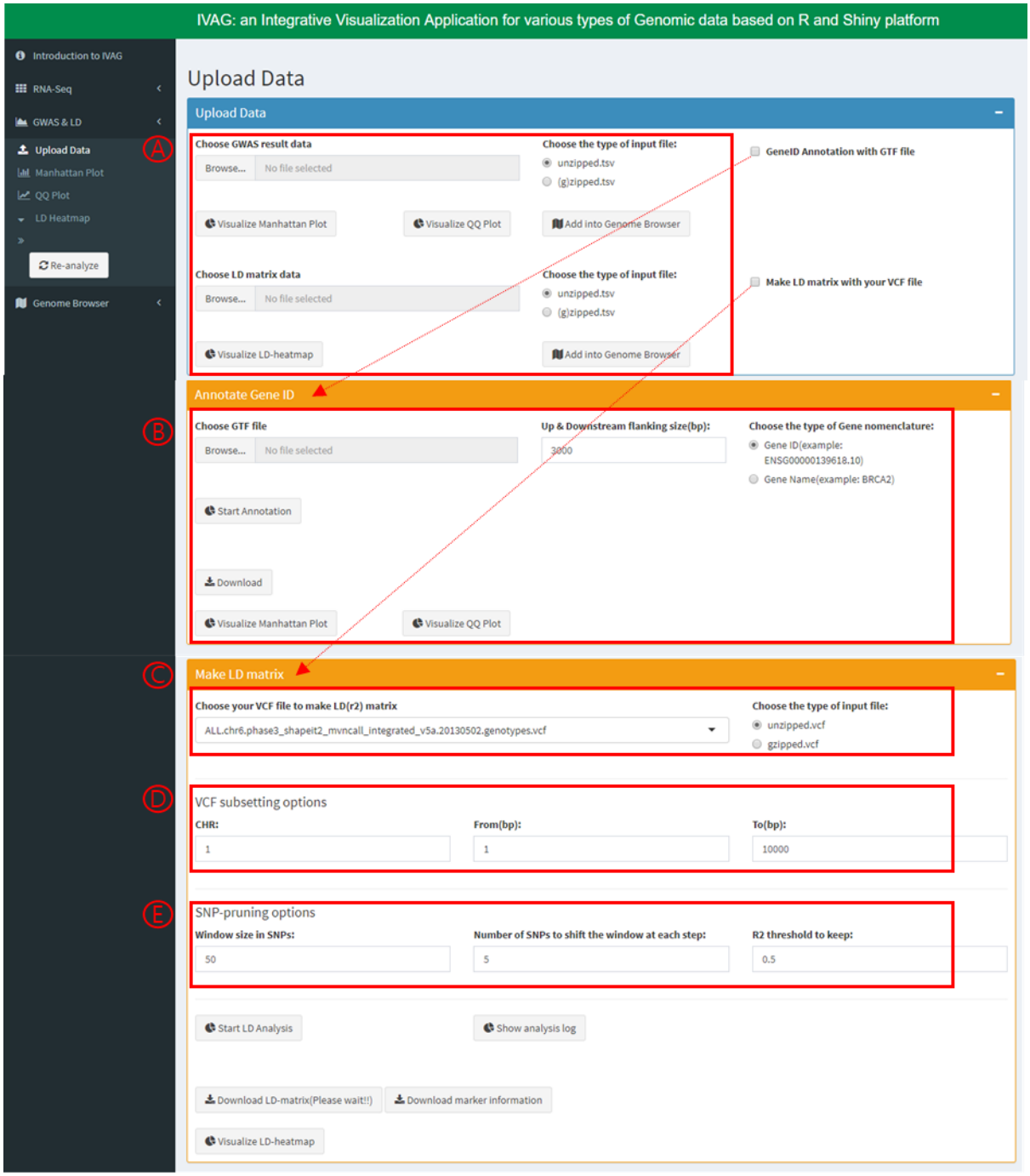
Ⓒ: DEG analysis result / pre analyzed DEG result table can be uploaded. Column header must be the same as above figure but, 'logCPM' and 'LR' can be omitted.

Ⓓ: Gene Ontology annotation file / Text file containing Gene ID and corresponding Gene Ontology category. Gene ID must be the same as the ones used in raw count table.

Ⓔ: GTF file / Used in Gene Ontology Enrichment analysis to calculate gene length.

2. GWAS & LD heatmap

<2.1 Upload Data >



<Fig. 9. GWAS & LD Upload and Analysis page>

Ⓐ: GWAS result and LD matrix files (TSV format) can be uploaded. Users can move into plotting page by clicking Visualize buttons. Also, uploaded files can be added into pre-built genome browser.

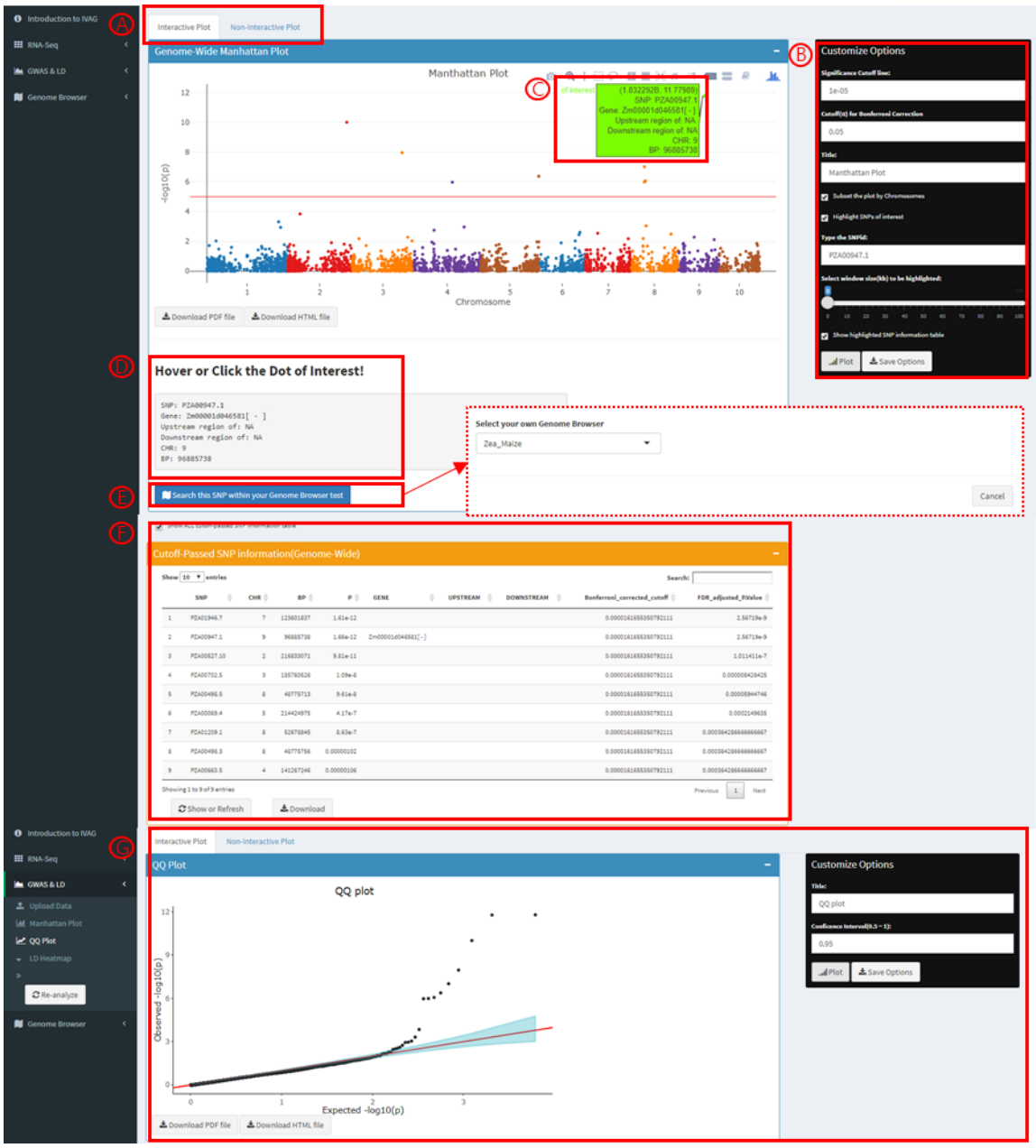
Ⓑ: Users can annotate gene information into all markers in the GWAS result file.

Ⓒ: Choose a VCF file you want to use in the LD analysis.

Ⓓ: Subset the region of your interest by assigning CHR, Start, and END parameters.

Ⓔ: All markers will be pruned using PLINK analysis option "--indep-pairwise" to reduce the number of markers to be used in LD calculation. Detailed information for this analysis can be found at http://zzz.bwh.harvard.edu/plink/summary.shtml#prune.

<2.2 Manhattan and QQ Plot Visualization >



<Fig. 10. Manhattan and QQ plot visualization page>

Ⓐ: Users can select what type of plot to draw; Interactive or Non-Interactive.

Ⓑ: Adjust specific parameters to customize your plot. You can subset the plot by chromosome and highlight the SNPs of your interest.

Ⓒ: Detailed information will be shown when the cursor is hovered over a specific point of interest.

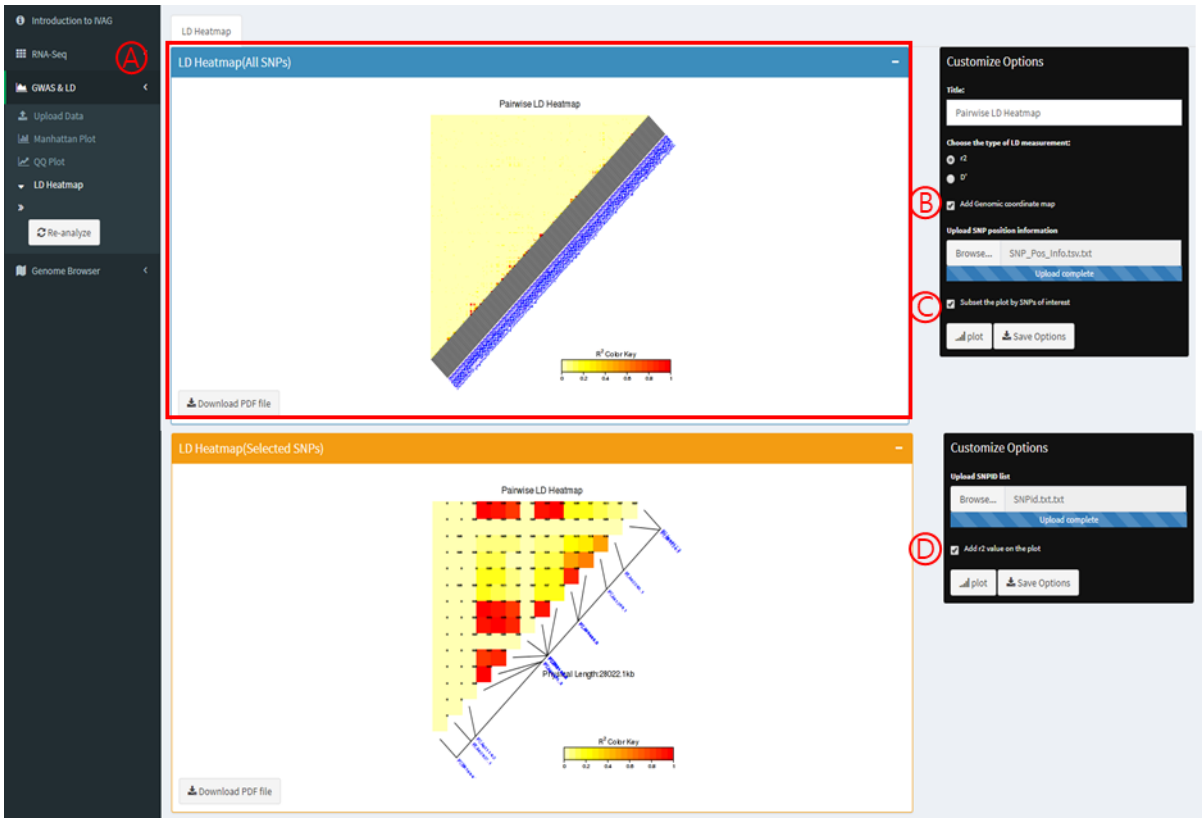
Ⓓ: If you click a specific dot, all information for that dot will be recoded here.

Ⓔ: User can move directly into the genome browser to search the SNP selected at the previous step. To use this function, genome browser construction should be done in advance.

Ⓕ: Information of all markers that passed the cutoff will be shown here.

Ⓖ: QQ plot can also be drawn. Light blue shade indicates the confidence interval for the null hypothesis that assumes there is no association between SNP and a trait.

<2.3 LD heatmap Visualization>



<Fig. 12. LD heatmap visualization page>

Ⓐ: All markers in the LD matrix (uploaded or generated) will be plotted on the upper panel.

Ⓑ: Genomic coordinate map can be added when genomic coordinate information is given.

Ⓒ: The LD heatmap can be subset by uploading a list of markers of your interest

Ⓓ: r2 value for each SNP-SNP pair can be added on the plot.

<2.4 Input Files >

GWAS

SNP	Chromosome	Position	P.value
PZA01946.7	7	123601837	1.61E-12
PZA00947.1	9	96885738	1.66E-12
PZA00527.10	2	216833071	9.81E-11
PZA00702.5	3	185760526	1.09E-08
PZA00498.5	8	48775713	9.61E-08
PZA00069.4	5	214424975	4.17E-07
PZA01209.1	8	52678845	8.63E-07
PZA00498.3	8	48775756	1.02E-06
PZA00663.5	4	141267246	1.06E-06
PZA02808.12	2	44606596	0.000143238
tb1.11	1	264848126	0.000483664

<Fig. 13. GWAS_Result.tsv -1 >

1. IVAG intakes GWAS result summary statistics file that comprise Marker ID, Chromosome, position, and p-value columns in order.

SNP	Chromosome	Position	P.value	GENE	UPSTREAM	DOWNSTREAM
PZA01946.7	7	123601837	1.61E-12	NA	NA	NA
PZA00947.1	9	96885738	1.66E-12	Zm00001d046581[-]	NA	NA
PZA00527.10	2	216833071	9.81E-11	NA	NA	NA
PZA00702.5	3	185760526	1.09E-08	NA	NA	NA
PZA00498.5	8	48775713	9.61E-08	NA	NA	NA
PZA00069.4	5	214424975	4.17E-07	NA	NA	NA
PZA01209.1	8	52678845	8.63E-07	NA	NA	NA
PZA00498.3	8	48775756	1.02E-06	NA	NA	NA
PZA00663.5	4	141267246	1.06E-06	NA	NA	NA
PZA02808.12	2	44606596	0.000143238	NA	NA	NA
tb1.11	1	264848126	0.000483664	NA	NA	NA

<Fig. 13. GWAS_Result.tsv - 2 >

2. Another version of the input file contains additional three columns which could be annotated from IVAG Gene ID annotation function.

LD heatmap -1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		PZA02388.1	PZA02174.2	PZA03316.4	PZA03316.1	PZA03316.3	PZA00368.17	PZA00368.1	PZA00058.1	PZA00058.5	PZA00632.1	PZA00416.4	PZA00416.2	PZA01601.1	PZA03178.3	PZA03178.1
2	PZA02388.1	NA	0.003131425	0.001322904	1.89E-05	0.010741308	0.00[illegible]236276	6.80E-05	0.004041285	0.00030022	0.012297859	0.011686865	0.007863762	1.12E-06	0.004849854	0.001331938
3	PZA02174.2	NA	NA	0.0004783	0.021465195	0.001231729	0.00[illegible]992238	0.024481744	0.010871922	0.00101117	0.002377562	0.016651214	0.029450244	0.055942213	0.005134748	0.002694006
4	PZA03316.4	NA	NA	NA	0.06055998	0.003111189	[illegible].71E-05	4.94E-05	0.000890745	0.007548125	0.002149237	0.005553662	0.000882867	0.006944157	0.000186331	0.001078003
5	PZA03316.1	NA	NA	NA	NA	0.031768088	0.00[illegible]528988	0.004006694	0.030093192	0.00026265	0.000316417	0.008766957	0.020765621	0.009565184	0.000657431	0.015585438
6	PZA03316.3	NA	NA	NA	NA	NA	0.025351076	0.000510142	0.014710731	0.007110129	0.00223847	0.000121821	0.000939886	0.002376762	0.000639039	0.001826396
7	PZA00368.17	NA	NA	NA	NA	NA	NA	0.23464068	0.015698233	0.006223222	0.000859469	0.006282812	1.15E-05	0.002566236	5.83E-06	1.23E-05
8	PZA00368.1	NA	NA	NA	NA	NA	NA	NA	0.022382311	0.001951984	0.000375188	0.001289014	2.77E-05	8.82E-08	0.000300585	0.005470765
9	PZA00058.1	NA	NA	NA	NA	NA	NA	NA	NA	0.079778481	0.0092099	0.010593625	0.017851142	0.000487344	8.59E-05	0.025033994
10	PZA00058.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.007402666	0.029095655	0.010838279	0.003178828	0.003044483	8.38E-05
11	PZA00632.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.000169487	0.01006222	0.012087334	1.87E-05	0.001650081
12	PZA00416.4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.510533148	0.037531322	0.015276814	0.004631683
13	PZA00416.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.028323356	0.012447245	0.000960156
14	PZA01601.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.021422767	0.016685048
15	PZA03178.3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0.150240822
16	PZA03178.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

<Fig. 14. LD_Matrix_Rsquare.tsv >

1. LD matrix file is tab-separated and has r2 values for all SNP-SNP pairs in an upper triangular matrix form.

2. The first row (header) is SNPIDs sorted by their position.

3. The first column, colored in blue, should not be included in the input LD matrix file, but I just showed it to help you better understand this format. SNP IDs in column header and row header are ordered equally and you can interpret this LD matrix like r2 value between PZA00368.17 and PZA03316.3 is 0.025351076 as described in green.

4. Note that your input file **SHOULD** look like the red boxed one.

LD heatmap - 2

chr8_SNPid - 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

```
PZB01094.3
PZA02927.1
PZA03114.2
PZA03381.2
PZA03381.1
PZB02114.1
PZB02114.2
PZB02114.3
PZB00145.1
PZA00498.5
PZA00498.3
PZA01209.1
PZA02203.1
PZA03126.1
PZA00417.3
```

<Fig. 15. SNPid.txt >

1. This text file has the list of SNP ids of your interest. You can extract a subset of LD heatmap that contains only SNPs specified with this file.

LD heatmap - 3

	A	B	C
1	SNP	Chromosome	Position
2	PZA02388.1	8	169137
3	PZA02174.2	8	4101256
4	PZA03316.4	8	4766593
5	PZA03316.1	8	4766694
6	PZA03316.3	8	4766801
7	PZA00368.17	8	5632196
8	PZA00368.1	8	5632308
9	PZA00058.1	8	5966657
10	PZA00058.5	8	5966698
11	PZA00632.1	8	6017018
12	PZA00416.4	8	8098163
13	PZA00416.2	8	8098271
14	PZA01601.1	8	8404207
15	PZA03178.3	8	11602192

<Fig. 16. SNP_Pos_Info.tsv >

1. genomic coordinate information file requires three columns; SNP, Chromosome, and position.

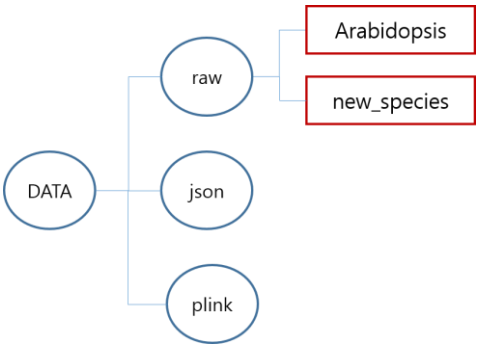
3. JBrowse

<3.1 Docker installation >

Install the docker before running IVAG. Docker is available on multiple platforms, on cloud and on-premises. Choose the best installation path for you and install it from the URL link below.

(https://docs.docker.com/engine/installation/)

<3.2 Launching Docker Image >



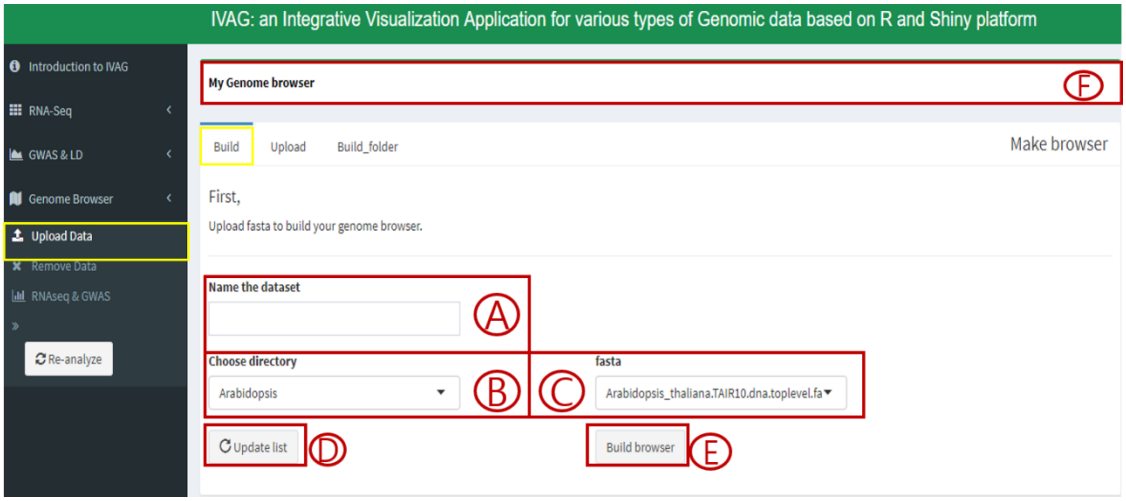
<Figure 17. Example folder mounted on a docker >

1. Create DATA directory for mounting docker.
2. Create subdirectory raw, json, plink on Data directory.
   - raw : Upload file for JBrowse
   - json : Information about data uploaded on JBrowse
   - plink : Save tmp, log files generated from IVAG LD analysis
3. Create subdirectory for each species under raw directory and put each data. (fasta, bam, gtf, gff3, bed, bw, vcf)
4. Specify DATA directory when launching docker image

DATA directory	C:₩Users₩USER₩Desktop₩DATA
Docker image launch commend	docker run -ti -v C:₩Users₩USER₩Desktop₩DATA:/jbrowse/my_data -p 8080:80 -p 8383:3838 leetaerim/ivag:v1 /bin/bash -c "Rscript load.R"

<Table 1. Example of launching docker image >

<3.3 Upload Data – Build >



<Figure 18. Build Genome Browser page>

Ⓐ: Specify genome browser name

Ⓑ: Shows subdirectory of raw directory mounted on docker. Among the list, select subdirectory to construct genome browser.

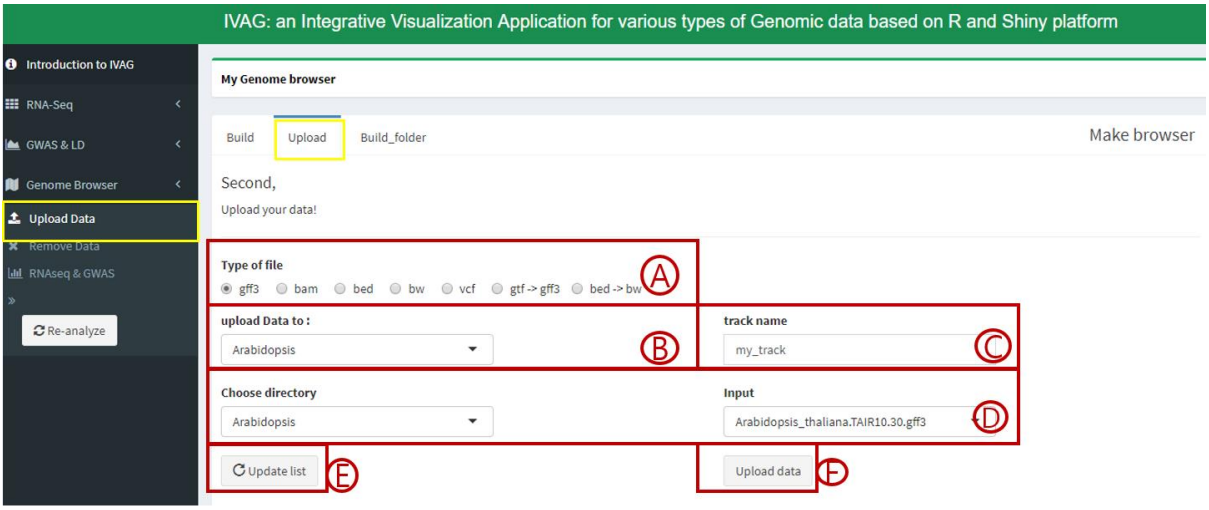
Ⓒ: Shows files of subdirectory chosen from Ⓑ. Choose fasta file.

Ⓓ: Click Update list button to refresh list on Ⓑ.

Ⓔ: Click Build browser button to construct genome browser.

Ⓕ: Leads to JBrowse page

<3.4 Upload Data – Upload >



<Figure 19. Upload data to browser page>

Ⓐ: Specify file type to upload.

- ➔ vcf : Takes long time to index data for ID search. So asks whether user will index or not.
- ➔ bam : If bam file size is large, asks whether to divide file into pieces to upload.

Ⓑ: Select genome browser to upload.

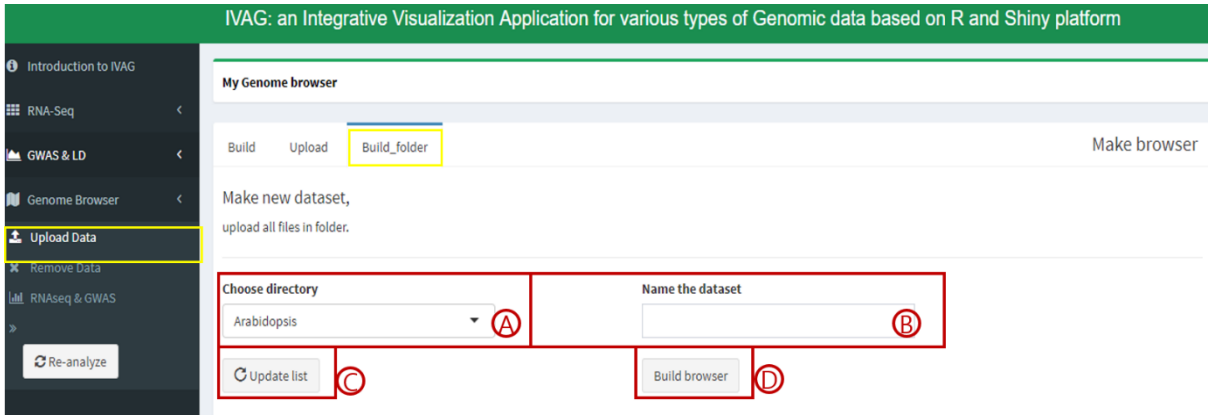
Ⓒ: Specify track name.

Ⓓ: Specify data to upload.

Ⓔ: Click Update list button to refresh list on Ⓑ,Ⓓ.

Ⓕ: Upload to genome browser.

<3.5 Upload Data – Build_folder >



<Figure 20. Build and Upload folder to browser page>

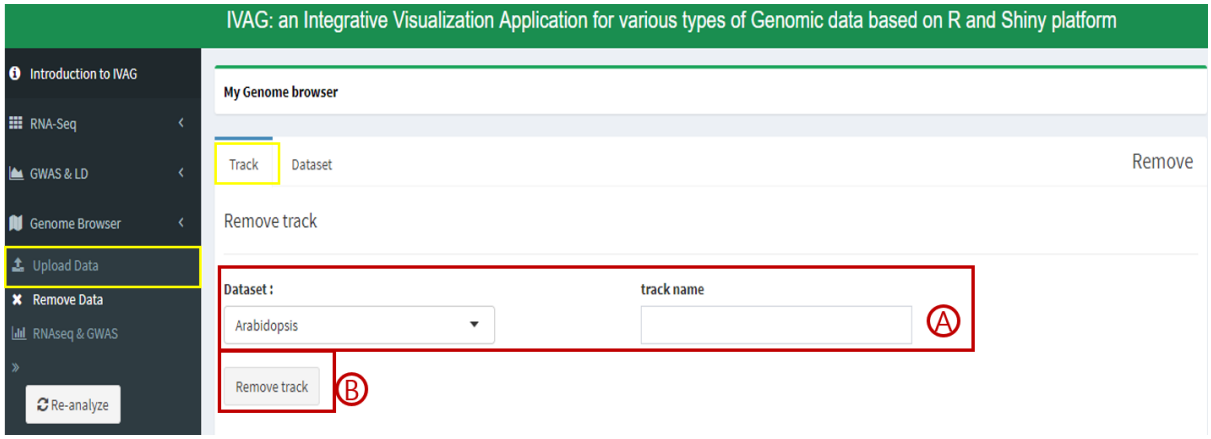
Ⓐ: Build genome browser using fasta file in specified directory and upload all the file in the directory automatically.

Ⓑ: Specify genome browser name.

Ⓒ: Click Update list button to refresh list on Ⓐ.

Ⓓ: Build genome browser.

<3.6 Remove Data - Track >

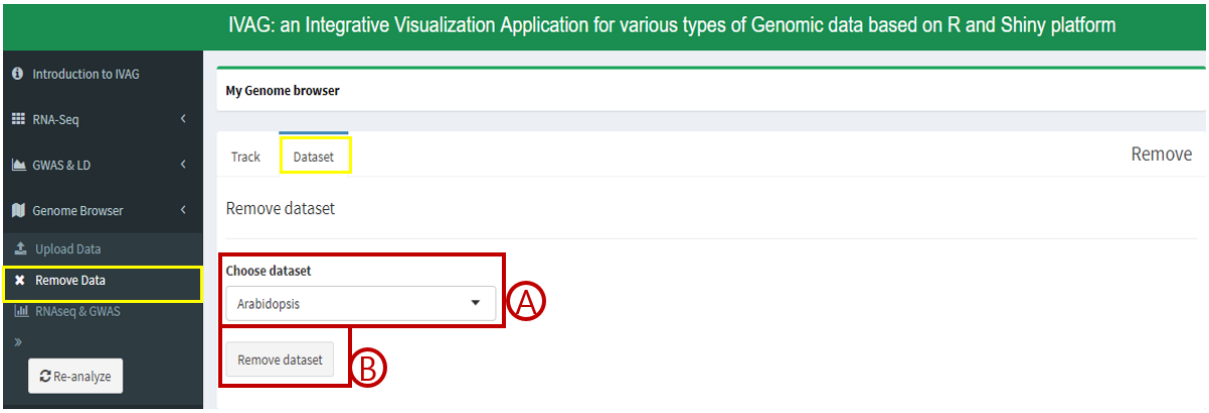


<Figure 21. Remove Data - Track page>

Ⓐ: Specify genome browser name and type track name to be deleted.

Ⓑ: Remove track.

<3.7 Remove Data - Dataset >



<Figure 22. Remove Data - Dataset page>

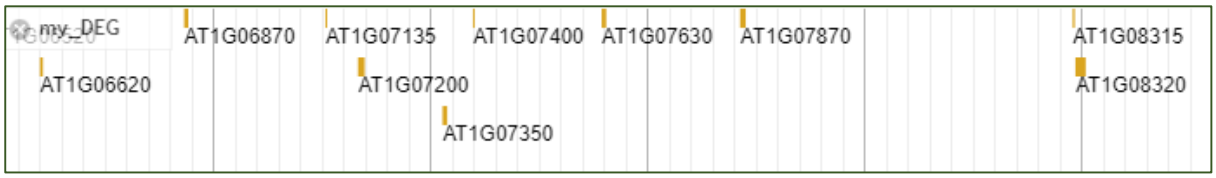
Ⓐ: Select genome browser to be removed.

Ⓑ: Remove genome browser.

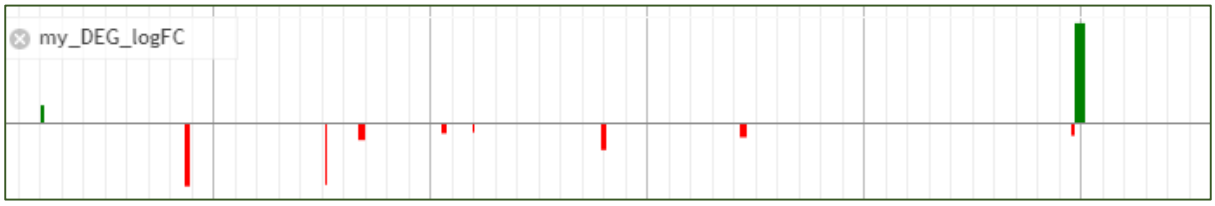
<3.8 RNA-Seq & GWAS >

IVAG RNA-Seq and GWAS analysis result can be uploaded to IVAG genome browser. Results can be uploaded via file but can be omitted by clicking "Add into Genome Browser" button on analysis page.

<3.9 RNA-Seq & GWAS – RNA-Seq >



<Figure 23. JBrowse RNAseq peak track >



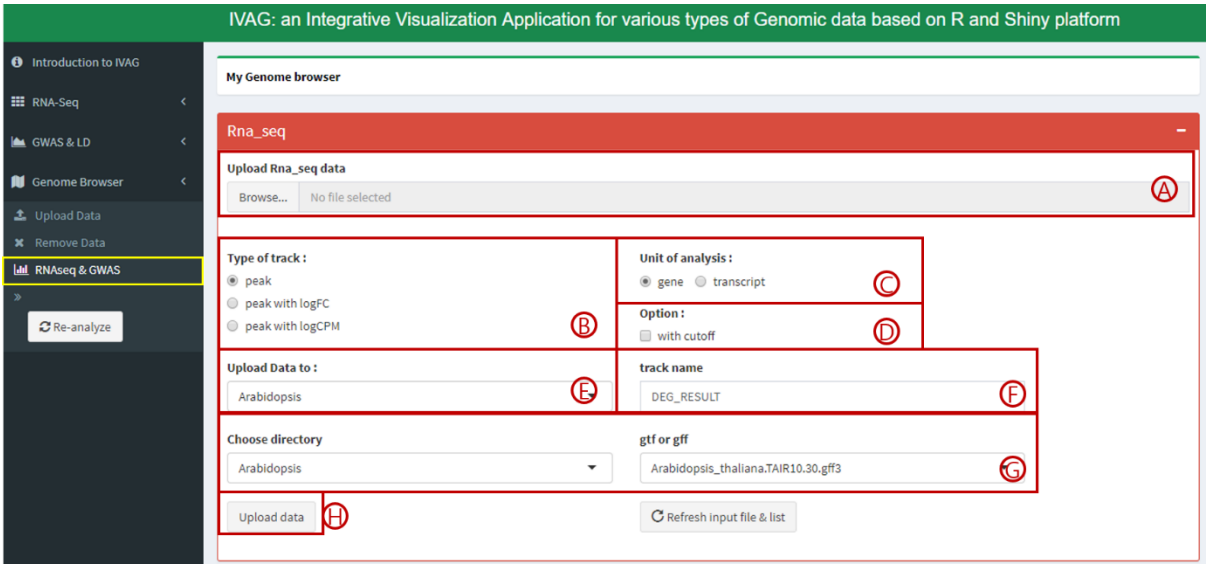
<Figure 24. JBrowse RNAseq peak with logFC track >

IVAG RNA-Seq analysis result can be uploaded in 3 track types.

① peak : DEG list is uploaded on JBrowse in bed file format (Figure 23)

② peak with logFC : Uploaded to JBrowse in BigWig format. (Figure 24)

③ peak with logCPM : Uploaded to JBrowse in BigWig format.



<Figure 25. RNAseq&GWAS - RNAseq >

Ⓐ: Upload DEG analysis result table.

Ⓑ: Specify track type.

Ⓒ: Specify analysis level(gene, transcript).

Ⓓ: Can specify FDR cutoff for DEG.

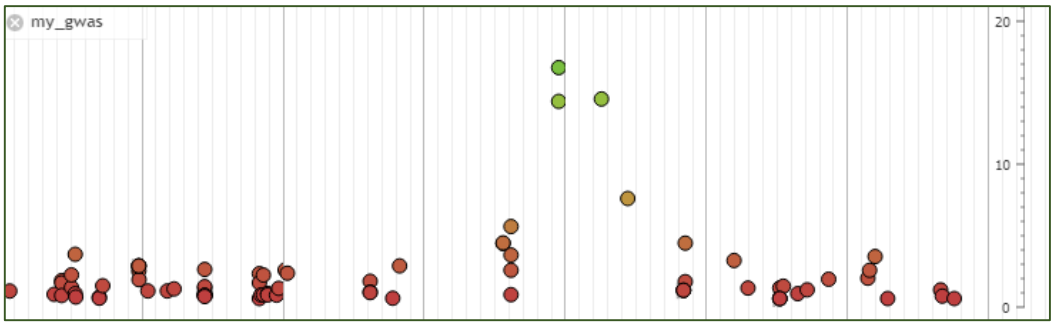
Ⓔ: Select genome browser to be uploaded.

Ⓕ: Set track name.

Ⓖ: Genome browser needs gene positions in order to process DEG list. Therefore GTF file needs to be specified.

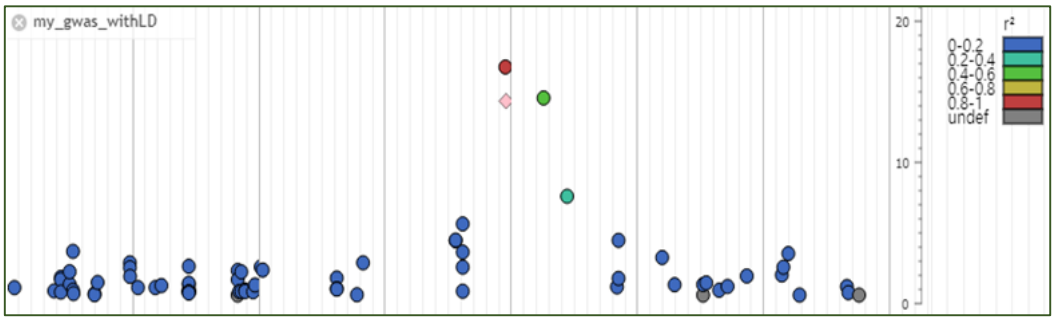
Ⓗ: Create RNA-Seq track.

<3.10 RNA-Seq & GWAS - GWAS >



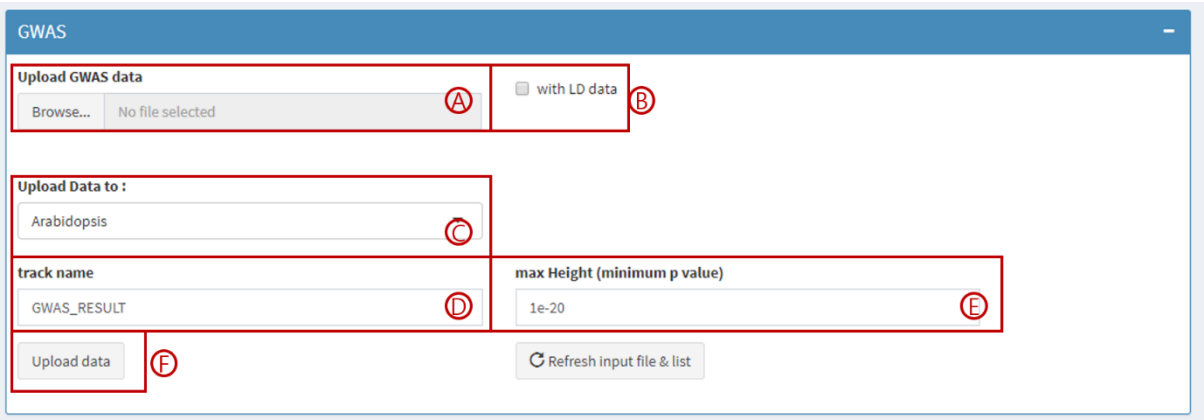
<Figure 26. JBrowse GWAS track >

JBrowse GWAS track. X axis shows Genome coordinate and t axis shows –log(p-value). Each dot represents SNP. Double click the dot to see the ID and coordinate.



<Figure 27. JBrowse GWAS with LD track >

Uploading GWAS data with LD data will produce default GWAS track (Figure 27). Dot represents SNP. Double clicking the dot will show R^2 value between other dots (Figure 27). Double clicked dot will be shown as pink diamond.



<Figure 28. RNAseq&GWAS - GWAS >

Ⓐ: Upload IVAG GWAS analysis file.

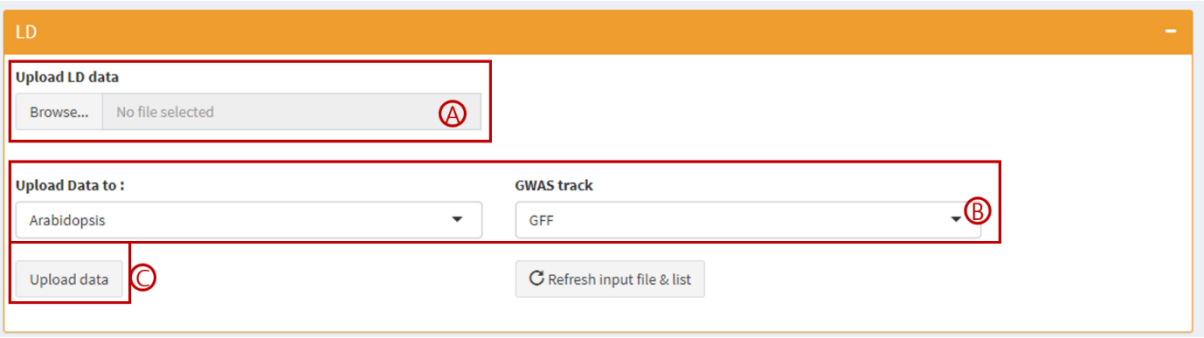
Ⓑ: LD analysis file can be uploaded as well.

Ⓒ: Select genome browser to be uploaded.

Ⓓ: Set track name.

Ⓔ: JBrowse GWAS track's y axis represents –log(p-value). Specify the maximum value for y axis.

Ⓕ: Upload GWAS track.



<Figure 29. RNAseq&GWAS – LD >

Ⓐ: Upload IVAG LD analysis result file.

Ⓑ: Select GWAS track to be uploaded.

Ⓒ: Click Upload data to add LD information to the specified GWAS track.