

Index

1. RNA-Seq

01. DEG analysis for single factor · · · · ·	03
02. DEG Visualization · · · · ·	04
03. Gene Ontology Enrichment Analysis · · · ·	06
04. DEG analysis for multiple factor · · · · ·	09
05. DEG visualization & Gene Ontology Enrichment for pre-analyzed data · · · · ·	11
06. Input files · · · · ·	12

2. GWAS

01. Upload Data · · · · ·	14
02. Manhattan and QQ Plot · · · · ·	15
03. LD heatmap Visualization · · · · ·	16
04. Input files · · · · ·	17

3. Genome Browser

01. Launching Docker Image	20
02. Upload Data - Build	21
03. Upload Data - Upload	22
04. Upload Data – Build_folder	23
05. Remove Data - Track	23
06. Remove Data - Dataset	24
07. RNA-Seq & GWAS	24
08. RNA-Seq	24
09. GWAS	26

<1.1 DEG analysis for single factor>

(A): Upload Count table. A file named 'Ara_root_rev_count.txt' is selected from a dropdown menu. The 'Upload complete' button is highlighted.

(B): Control table Meta information. Fields include 'Control start column number' (1), 'Control end column number' (3), and a 'submit' button.

(C): Normalization. Fields include 'CPM cutoff' (2), 'Number of sample to contain above CPM cutoff' (3), and a 'submit' button.

(D): DEG analysis results table. It shows logFC, logCPM, LR, PValue, and FDR for 10,000 entries. Buttons at the bottom include 'Start analysis', 'Download', 'Visualization', and 'GO analysis'.

<Figure 1. DEG single factor analysis page >

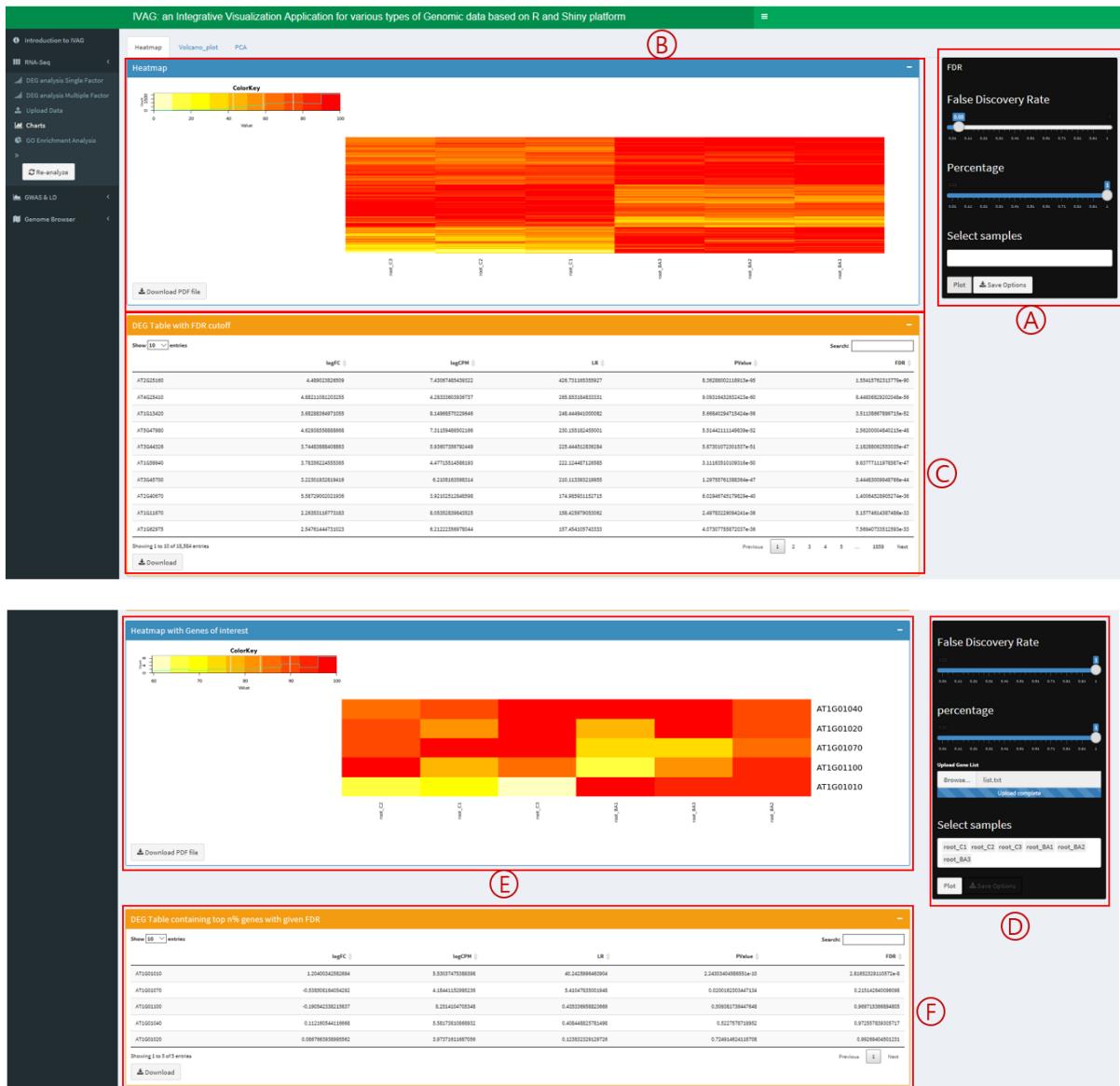
Ⓐ: Upload count table generated using htseq-count or similar software after mapping RNA-Seq data to reference Genome.

Ⓑ: 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

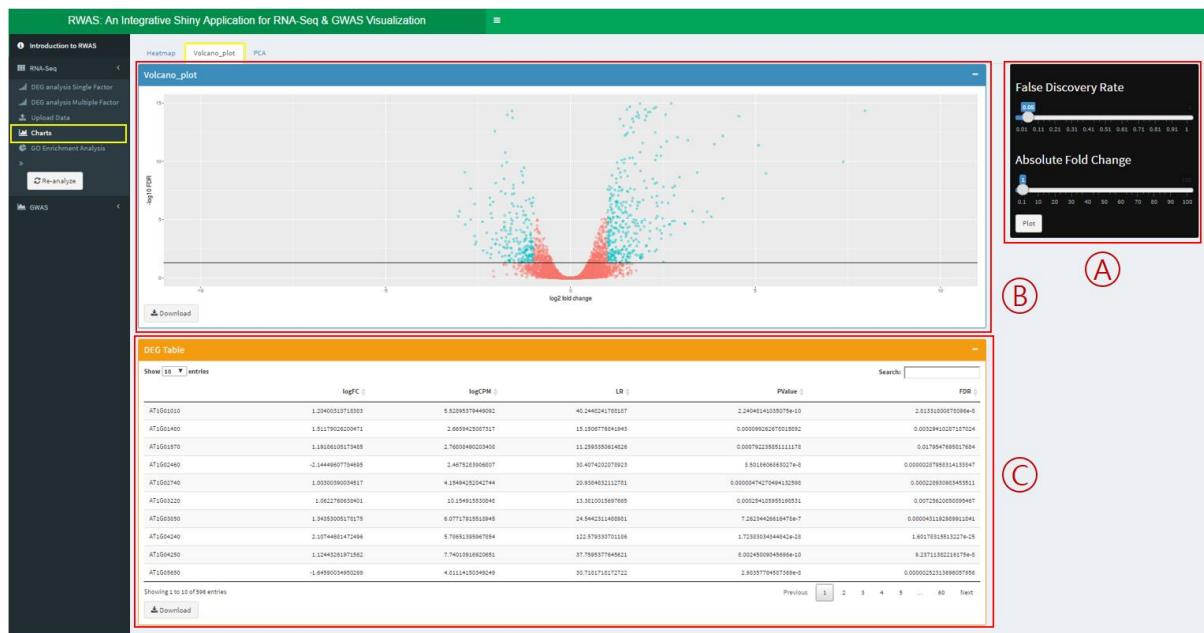
Ⓒ: Shows DEG analysis table of genes used to create heatmap. Click Download button to download

table in text file format.

④: 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.

⑤: Generated heatmap.

⑥: Shows DEG analysis table of genes used to create heatmap. Click Download button to download table in text file format.

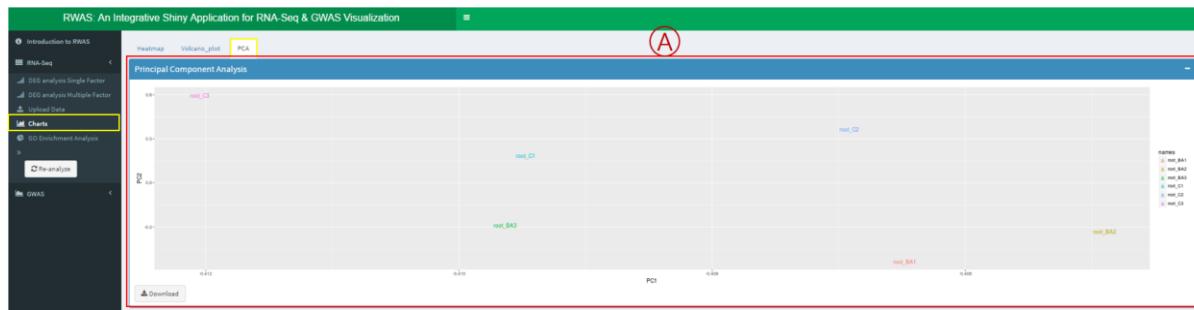


<Figure 3. Volcano plot visualization>

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

Ⓑ: Generates volcano plot by plotting logf fold change versus $-\log_{10}(\text{False Discovery Rate})$. User can download generated plot in PDF format by clicking the Download button.

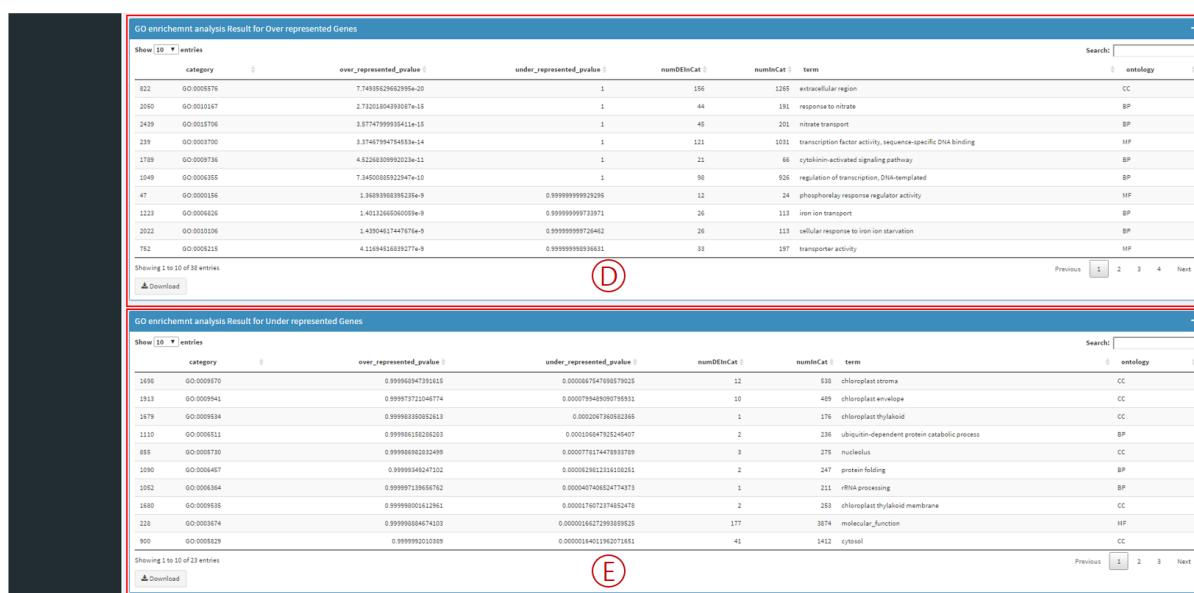
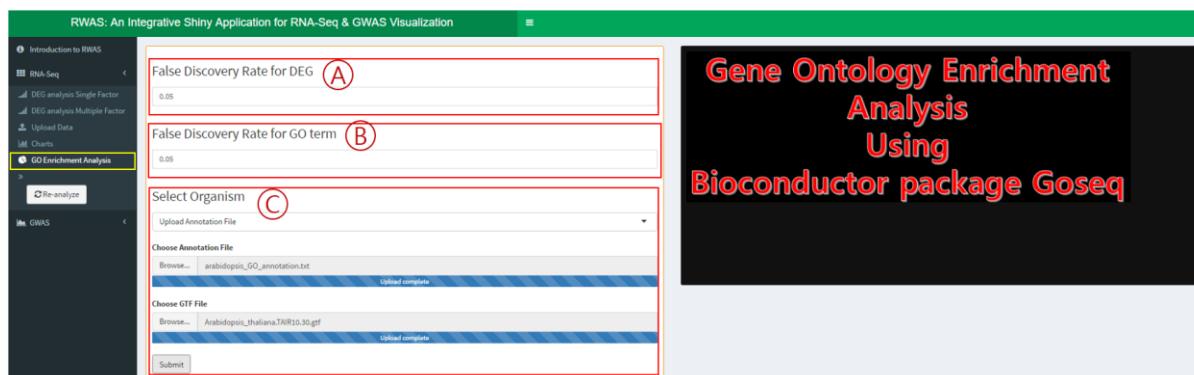
Ⓒ: DEG analysis result table for genes used to plot volcano plot is shown. This table can be downloaded in text format by clicking the Download button.

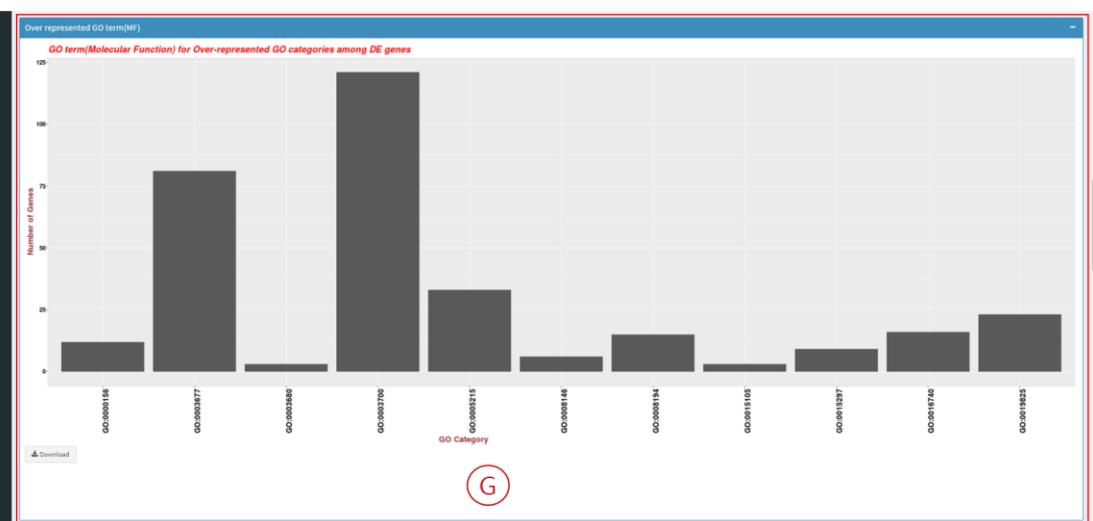
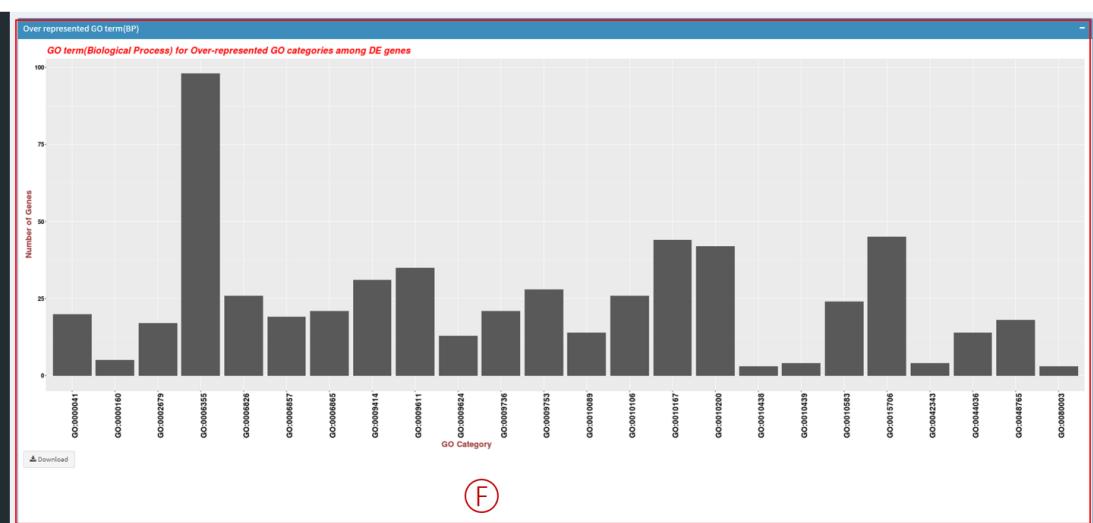


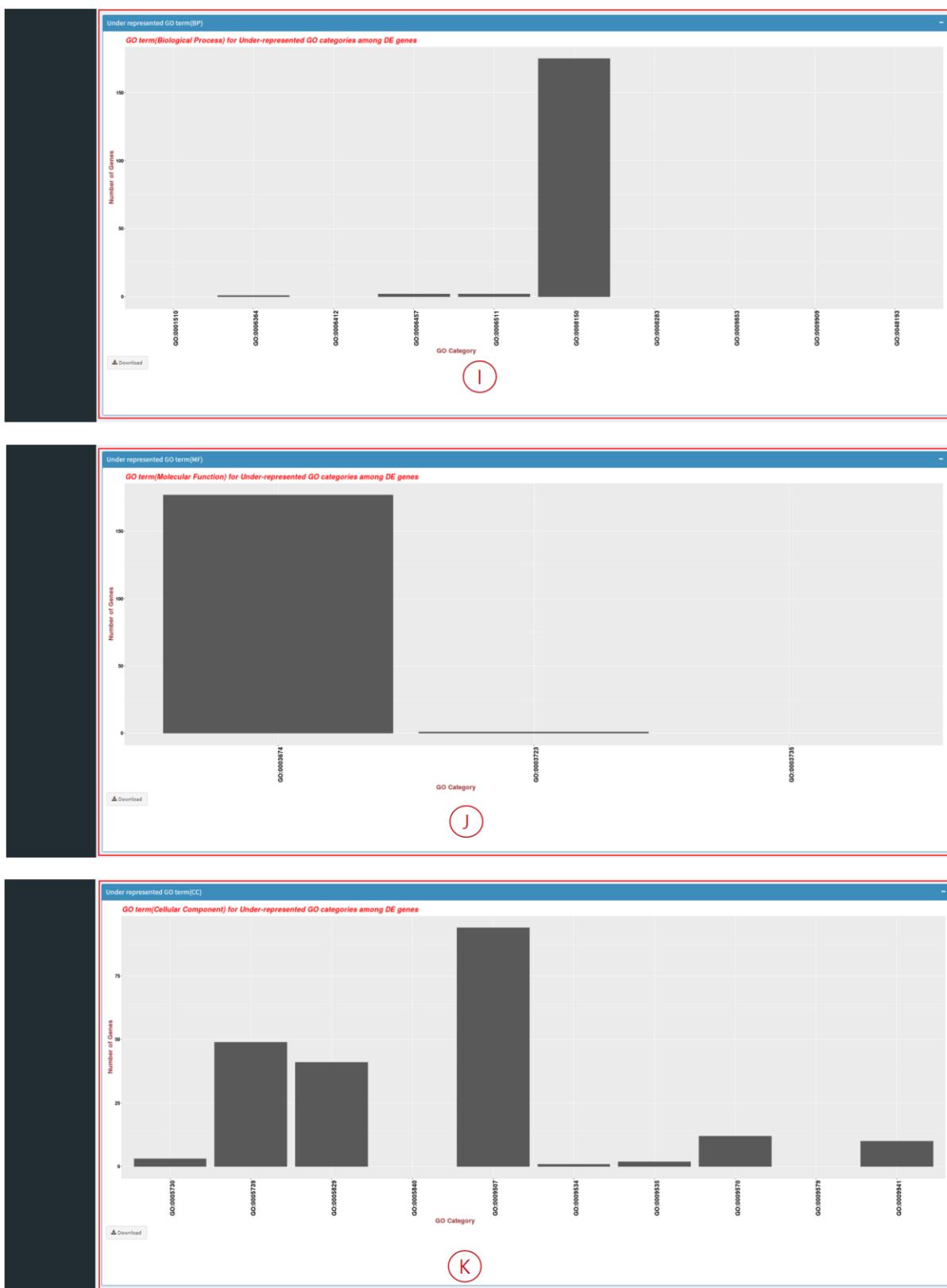
<Figure 4. PCA plot visualization>

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

<1.3 Gene Ontology Enrichment Analysis>







<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.

④: Specify False Discovery Rate to be used in Gene Ontology Enrichment analysis.

⑤: Upload Gene Ontology annotation file and GTF file.

⑥: 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.

⑦: 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.

⑧ ~ ⑩: 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>

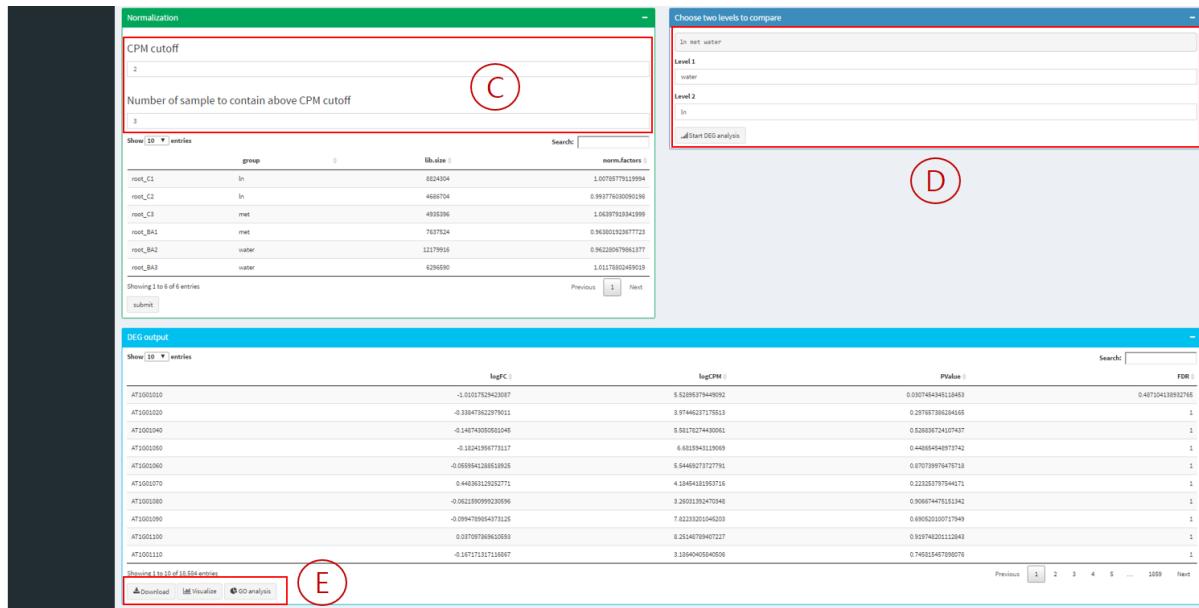
The screenshot displays the RWAS (RNA-Seq & GWAS Visualization) application interface. On the left, a sidebar menu includes options like 'Introduction to RWAS', 'RNA-Seq', 'DEG analysis Single Factor', 'DEG analysis Multiple Factor' (which is selected), 'Open Data', 'Charts', and 'Enrichment Analysis'. A 'Re-analyze' button is also present. The main area has two main sections:

- Upload:** This section contains two input fields: 'Choose Raw count data' (with a file named 'Airp_root_raw_count.txt' selected) and 'Choose Meta data' (with a file named 'meta_data.txt' selected). Both fields show 'Upload complete'. A red circle labeled 'A' highlights this section.
- Design Matrix:** This section shows a table titled 'Design Matrix' with columns 'In', 'met', and 'water'. The data rows are:

	In	met	water
root_C1	1	0	0
root_C2	1	0	0
root_C3	0	1	0
root_BA1	0	1	0
root_BA2	0	0	1
root_BA3	0	0	1

A red circle labeled 'B' highlights this section.

At the bottom of the main area, there are buttons for 'Previous', 'Next', and 'Submit'.



<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>

The screenshot shows the 'RWAS: An Integrative Shiny Application for RNA-Seq & GWAS Visualization' interface. On the left, a sidebar menu includes 'Introduction to RWAS', 'RNA-Seq' (selected), 'DEG analysis Single Factor', 'DEG analysis Multiple Factor', 'Upload Data' (highlighted with a red circle A), 'Charts', 'GO Enrichment Analysis', 'Re-analyze' (button), and 'GWAS'. The main area is titled 'Upload data' and contains three sections: 'Upload Count table' (with 'Choose Raw count data' and 'Browse...' button), 'Choose DEG output' (with 'Browse...' button), and a bottom row with 'Visualize' and 'GO analysis' buttons (both highlighted with red circles B and C).

<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>

```

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

```

```

1 ln
2 ln
3 met
4 met
5 water
6 water

```

(B)

```

1 "logFC" "logCPM" "LR" "PValue" "FDR"
2 "AT1G01010" 1.19688737736099 5.52688808335964 39.3089106372121 3.61786523882999e-10 4.80222720383886e-08
3 "AT1G01020" 0.0822804977656206 3.97165716089378 0.111434656713485 0.738516674183035 0.992950029705138
4 "AT1G01030" -0.162468882715308 0.0823715495057947 0.0610995702734676 0.804766390210944 0.994580412175897
5 "AT1G01040" 0.106377858075599 5.57910426454308 0.37042230935759 0.542774163320339 0.974600223816041
6 "AT1G01050" 0.0153574773648363 6.67932177116587 0.00814645735631103 0.92808241163436 0.995983980196374
7 "AT1G01060" 0.0456180046924779 5.54156373113906 0.019399476207294 0.889227311214195 0.995931204225289
8 "AT1G01070" -0.544012070552485 4.18302675004016 5.48539400852262 0.0191759975507977 0.211398693538396
9 "AT1G01080" 0.174667523566418 3.26203485385487 0.299539868880617 0.584171025511136 0.978544418926473
10 "AT1G01090" 0.0770064556487513 7.82001219000772 0.206235076914759 0.649734453606224 0.987057549001829
11 "AT1G01100" -0.194374421491345 8.24999782635235 0.450922121063243 0.501897378775529 0.967135588725176
12 "AT1G01110" -0.0356166922910766 3.18109685561797 0.011510637873462 0.914560789141767 0.995983980196374
13 "AT1G01120" 0.641291331519332 6.3387621639768 11.0961455503092 0.000865073303584205 0.0196921958851816
14 "AT1G01130" 0.467864553003074 1.76456938585317 0.977622941476459 0.322786369058292 0.904971916676201
15 "AT1G01140" -0.542815616073306 2.63788851959664 2.13563384343845 0.143910966721881 0.68944588214333
16 "AT1G01160" -0.0296144956218807 5.15484537653471 0.0203745510116882 0.88649595793983 0.995931204225289
17 "AT1G01170" 0.09595143656512992 3.64734295039377 0.15817628576344 0.69084114621376 0.987813981559841
18 "AT1G01180" 0.24596413317094 3.7881988269758 0.650354637870404 0.419985915804817 0.950470048855293
19 "AT1G01190" -0.375672462290263 3.86317903951983 0.635708343635596 0.425269154293829 0.951393075845317
20 "AT1G01200" 0.246507464818891 3.01178043499523 0.779877708445121 0.377178543497332 0.933754390534284

```

(C)

```

1 AT1G01010 GO:0006888
2 AT1G01010 GO:0007275
3 AT1G01010 GO:0043090
4 AT1G01020 GO:0006665
5 AT1G01020 GO:0006665
6 AT1G01020 GO:0016125
7 AT1G01020 GO:0016125
8 AT1G01030 GO:0006355
9 AT1G01030 GO:0006355
10 AT1G01030 GO:0006355
11 AT1G01030 GO:0006355
12 AT1G01030 GO:0009908
13 AT1G01030 GO:0048366
14 AT1G01040 GO:0000226
15 AT1G01040 GO:0000226
16 AT1G01040 GO:0000278
17 AT1G01040 GO:0000911
18 AT1G01040 GO:0006306
19 AT1G01040 GO:0006306
20 AT1G01040 GO:0006342
21 AT1G01040 GO:0006342
22 AT1G01040 GO:0006342
23 AT1G01040 GO:0006396
24 AT1G01040 GO:0006396
25 AT1G01040 GO:0006396
26 AT1G01040 GO:0006396
27 AT1G01040 GO:0007267
28 AT1G01040 GO:0009616
29 AT1G01040 GO:0009616
30 AT1G01040 GO:0009616
31 AT1G01040 GO:0009616
32 AT1G01040 GO:0009630
33 AT1G01040 GO:0009880
34 AT1G01040 GO:0009908

```

D

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 >

IVAG: an Integrative Visualization Application for various types of Genomic data based on R and Shiny platform

A

Upload Data

Choose GWAS result data

Choose the type of input file:

- unzipped.tsv
- (g)zipped.tsv

Visualize Manhattan Plot Visualize QQ Plot Add into Genome Browser

Choose LD matrix data

Choose the type of input file:

- unzipped.tsv
- (g)zipped.tsv

Visualize LD-heatmap Add into Genome Browser

B

Annotate Gene ID

Choose GTF file

Up & Downstream flanking size(bp): 3000

Choose the type of Gene nomenclature:

- Gene ID(example: ENSG00000139618.10)
- Gene Name(example: BRCA2)

Start Annotation Download Visualize Manhattan Plot Visualize QQ Plot

C

Make LD matrix

Choose your VCF file to make LD(r2) matrix: ALL.chr6.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf

Choose the type of input file:

- unzipped.vcf
- gzipped.vcf

D

VCF subsetting options

CHR: 1 From(bp): 1 To(bp): 10000

E

SNP-pruning options

Window size in SNPs: 50 Number of SNPs to shift the window at each step: 5 R2 threshold to keep: 0.5

Start LD Analysis Show analysis log

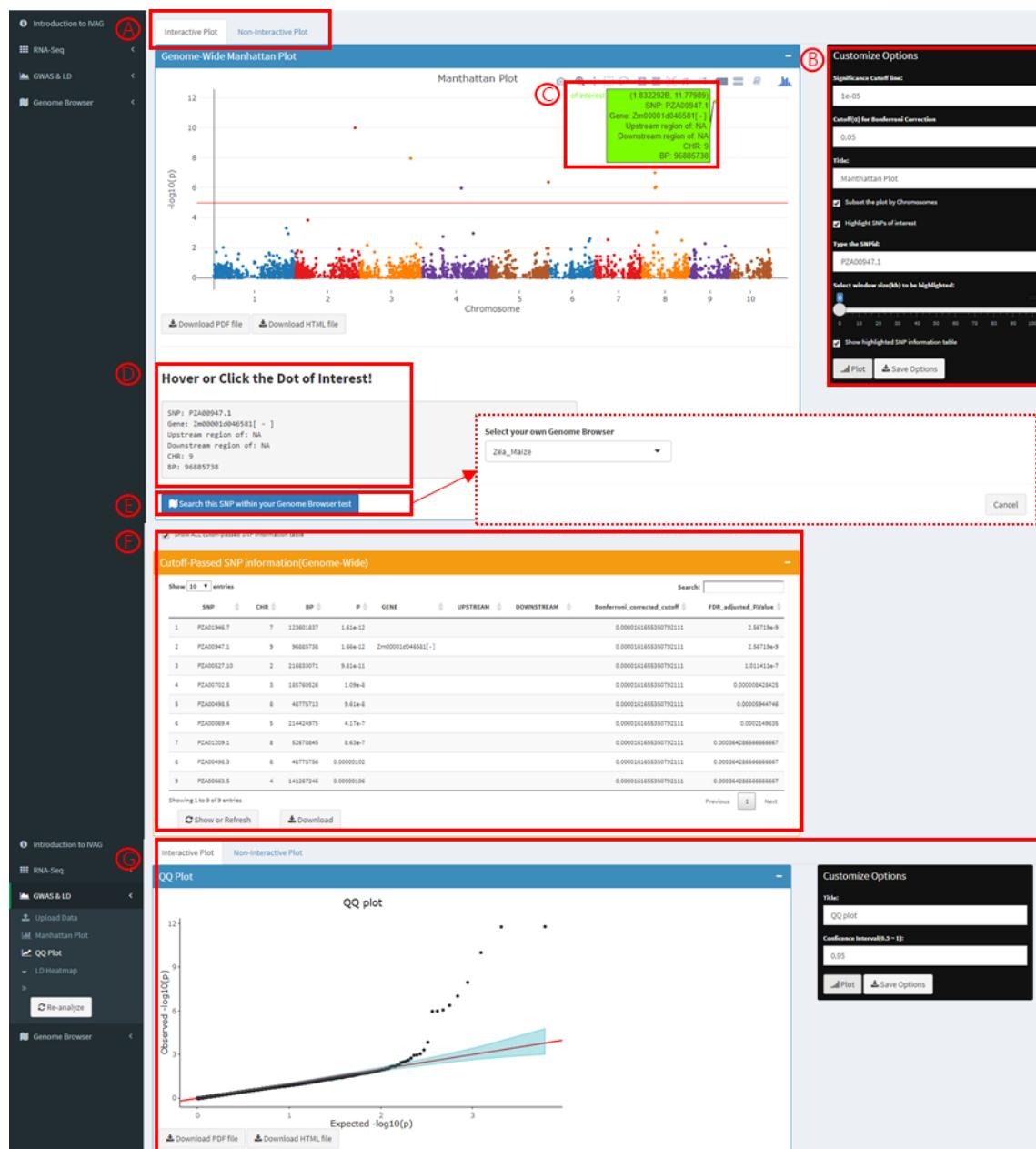
Download LD-matrix(Please wait!!) Download marker information Visualize LD-heatmap

<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.

- ④: Users can annotate gene information into all markers in the GWAS result file.
- ⑤: Choose a VCF file you want to use in the LD analysis.
- ⑥: Subset the region of your interest by assigning CHR, Start, and END parameters.
- ⑦: 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>.

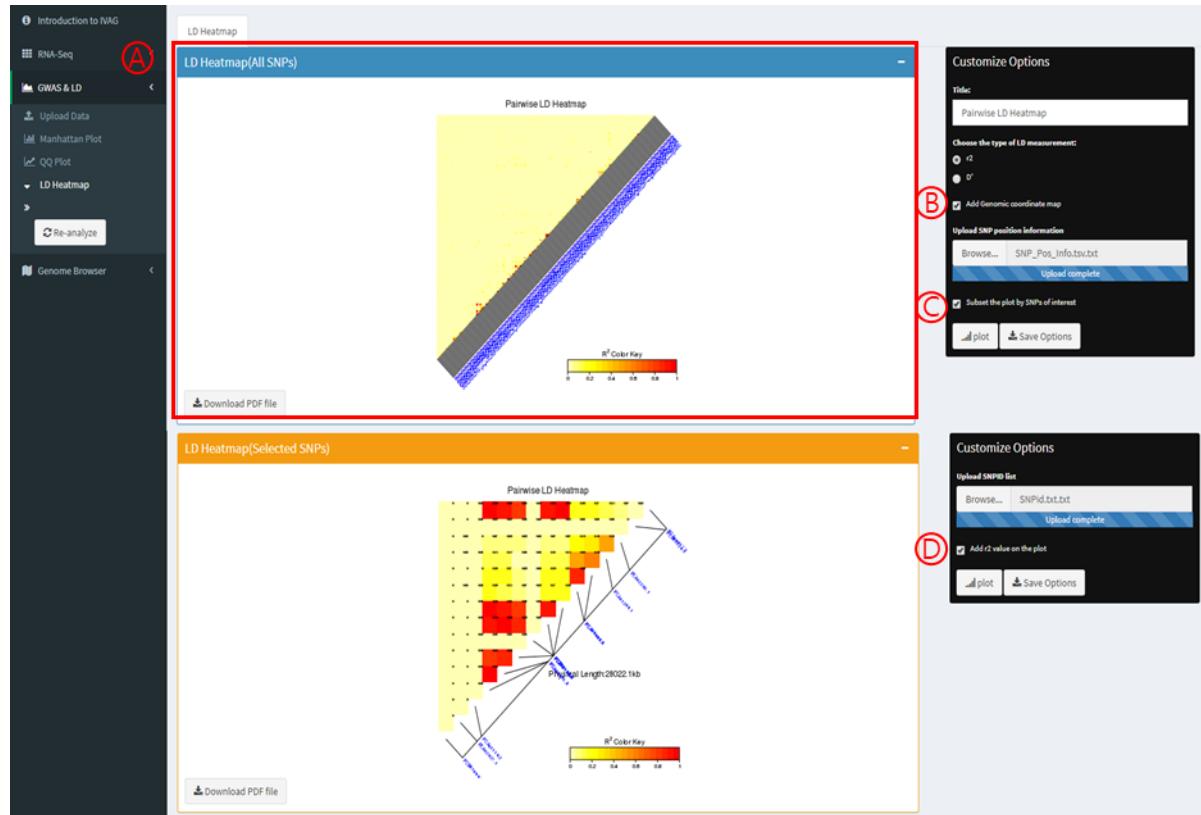
<2.2 Manhattan and QQ Plot Visualization >



<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

④: 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 >

- 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 >

- 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	PZA03316.3	PZA00368.17	PZA00368.1	PZA00058.1	PZA00058.5	PZA00632.1	PZA00416.4	PZA00416.2	PZA01601.1	PZA03178.3	PZA03178.1	
2	PZA02174.2	NA	0.003131425	0.001322904	1.89E-05	0.010741308	0.001236276	6.80E-05	0.004041285	0.00030022	0.012297859	0.011686865	0.007863762	1.12E-06	0.004849854
3	PZA02174.2	NA	NA	0.0004783	0.021465195	0.001231729	0.001992238	0.024481744	0.010871922	0.00101117	0.002377562	0.016651214	0.029450244	0.005942213	0.005134748
4	PZA03316.4	NA	NA	NA	0.06055998	0.003111189	-7.1E-05	4.94E-05	0.000890745	0.007548125	0.002149237	0.005553662	0.000882867	0.006944157	0.000186331
5	PZA03316.1	NA	NA	NA	NA	0.031768088	0.00528988	0.004006694	0.030093192	0.00026265	0.000316417	0.008766957	0.020765621	0.009565184	0.000657431
6	PZA03316.3	NA	NA	NA	NA	NA	0.025351076	0.000510142	0.014710731	0.00710129	0.00223847	0.000121821	0.000939886	0.002376762	0.000639039
7	PZA00368.17	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	0.022382311	0.001951984	0.00375188	0.001289014	2.77E-05	8.82E-08	0.000300585	0.005470765
9	PZA00058.1	NA	NA	NA	NA	NA	NA	NA	0.079778481	0.0092099	0.010593625	0.017851142	0.00487344	8.59E-05	0.025033994
10	PZA00058.5	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	0.0106222	0.012087334	1.87E-05
12	PZA00416.4	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	0.028323356	0.012447245	0.00960156
14	PZA01601.1	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	0.150240822
16	PZA03178.1	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 SNPs IDs 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

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

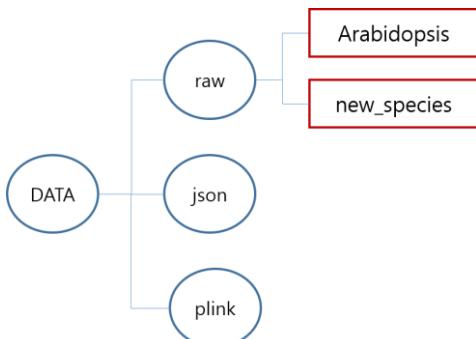
	A	B	C
1	SNP	Chromosome	Position
2	PZA02388.1	8	169137
3	PZA02174.2	8	4101256
4	PZA03316.4	8	4766593
5	PZA03316.1	8	4766694
6	PZA03316.3	8	4766801
7	PZA00368.17	8	5632196
8	PZA00368.1	8	5632308
9	PZA00058.1	8	5966657
10	PZA00058.5	8	5966698
11	PZA00632.1	8	6017018
12	PZA00416.4	8	8098163
13	PZA00416.2	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 Launching Docker Image >



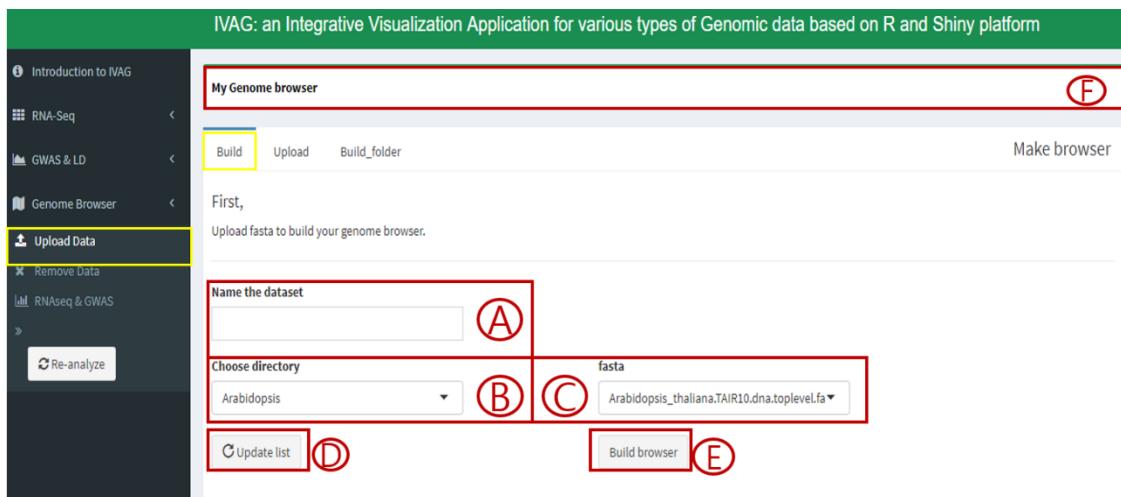
<Figure 17. 도커에 마운트하는 폴더 예시>

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 command	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 >

<3.2 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.

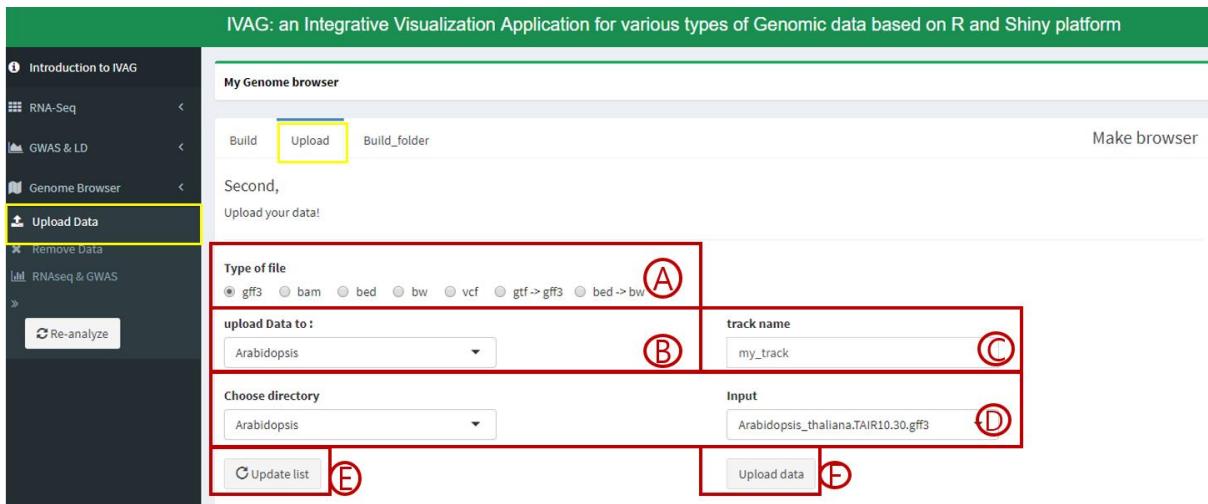
Ⓒ: Shows files of subdirectory chosen from Ⓑ. Choose fasta file.

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

Ⓔ: Click Build browser button to construct genome browser.

Ⓕ: Leads to JBrowse page

<3.3 Upload Data – Upload >



<Figure 19. Upload data to browser page>

(A): 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.

(B): Select genome browser to upload.

(C): Specify track name.

(D): Specify data to upload.

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

(F): Upload to genome browser.

<3.4 Upload Data – Build_folder >

The screenshot shows the IVAG application interface. The left sidebar has a dark theme with white text and icons. The 'Upload Data' option is highlighted with a yellow box. The main content area has a light gray background. At the top, there's a green header bar with the text 'IVAG: an Integrative Visualization Application for various types of Genomic data based on R and Shiny platform'. Below the header, the title 'My Genome browser' is displayed. In the center, there are three tabs: 'Build', 'Upload', and 'Build_folder', with 'Build_folder' highlighted by a yellow box. To the right of the tabs is a button 'Make browser'. Below the tabs, instructions say 'Make new dataset, upload all files in folder.' On the left, there's a dropdown menu labeled 'Choose directory' with 'Arabidopsis' selected. To its right is a text input field labeled 'Name the dataset' with a red circle around it. Below these are two buttons: 'Update list' with a red circle around it and 'Build browser' with a red circle around it.

<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.

<3.5 Remove Data - Track >

The screenshot shows the IVAG application interface. The left sidebar has a dark theme with white text and icons. The 'Upload Data' option is highlighted with a yellow box. The main content area has a light gray background. At the top, there's a green header bar with the text 'IVAG: an Integrative Visualization Application for various types of Genomic data based on R and Shiny platform'. Below the header, the title 'My Genome browser' is displayed. In the center, there are two tabs: 'Track' and 'Dataset', with 'Track' highlighted by a yellow box. To the right of the tabs is a 'Remove' button. Below the tabs, instructions say 'Remove track'. On the left, there's a dropdown menu labeled 'Dataset:' with 'Arabidopsis' selected. To its right is a text input field labeled 'track name' with a red circle around it. Below these are two buttons: 'Remove track' with a red circle around it and another 'Remove track' button.

<Figure 21. Remove Data - Track page>

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

<3.6 Remove Data - Dataset >

The screenshot shows the 'My Genome browser' interface. On the left, a sidebar lists various options: 'Introduction to IVAG', 'RNA-Seq', 'GWAS & LD', 'Genome Browser', 'Upload Data', 'Remove Data' (which is highlighted with a yellow box), and 'RNaseq & GWAS'. Below these are 'Re-analyze' and '»'. The main content area has two tabs at the top: 'Track' and 'Dataset' (the latter is highlighted with a yellow box). Below the tabs is a 'Remove dataset' button. A red box labeled 'A' encloses a dropdown menu titled 'Choose dataset' with 'Arabidopsis' selected. Another red box labeled 'B' encloses the 'Remove dataset' button.

<Figure 22. Remove Data - Dataset page>

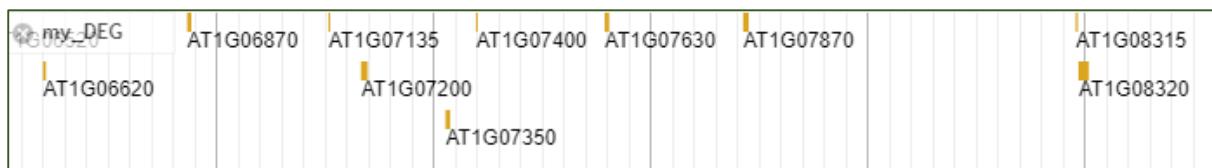
Ⓐ: Select genome browser to be removed.

Ⓑ: Remove genome browser.

<3.7 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.

<3.8 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.

The screenshot shows the IVAG application interface. On the left, there is a sidebar with various options: Introduction to IVAG, RNA-Seq, GWAS & LD, Genome Browser, Upload Data, Remove Data, and RNAseq & GWAS (which is highlighted). Below these are Re-analyze and Help buttons. The main area is titled "My Genome browser" and contains a "Rna_seq" section. This section has several input fields and dropdown menus:

- "Upload Rna_seq data" with a "Browse..." button and a message "No file selected".
- "Type of track:" with radio buttons for "peak", "peak with logFC", and "peak with logCPM" (the last one is selected).
- "Unit of analysis:" with radio buttons for "gene" (selected) and "transcript".
- "Option:" with a checkbox for "with cutoff".
- "Upload Data to:" set to "Arabidopsis".
- "track name" set to "DEG_RESULT".
- "Choose directory" dropdown set to "Arabidopsis" and "gtf or gff" dropdown set to "Arabidopsis_thaliana.TAIR10.30.gff3".
- "Upload data" button.
- "Refresh input file & list" button.

Red circles labeled A through H are overlaid on the interface to point to specific elements: A points to the "Upload Rna_seq data" field; B points to the "Type of track:" dropdown; C points to the "Unit of analysis:" dropdown; D points to the "Option:" checkbox; E points to the "Upload Data to:" dropdown; F points to the "track name" input field; G points to the "gtf or gff" dropdown; and H points to the "Upload data" button.

<Figure 25. RNAseq&GWAS - RNAseq >

Ⓐ: Upload DEG analysis result table.

Ⓑ: Specify track type.

Ⓒ: Specify analysis level(gene, transcript).

Ⓓ: Can specify FDR cutoff for DEG.

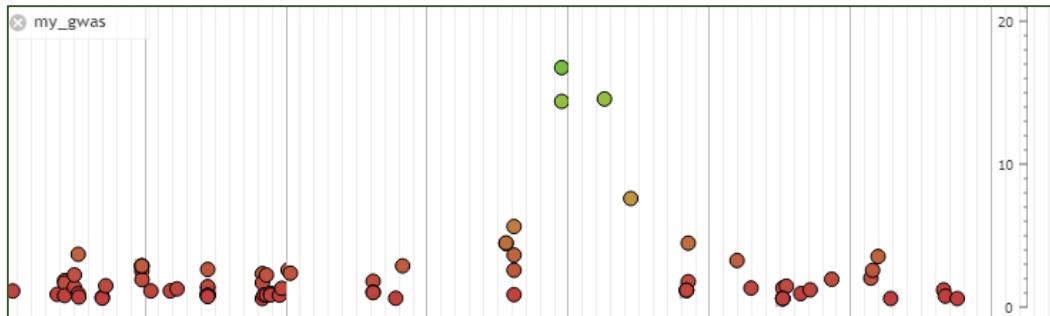
Ⓔ: Select genome browser to be uploaded.

Ⓕ: Set track name.

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

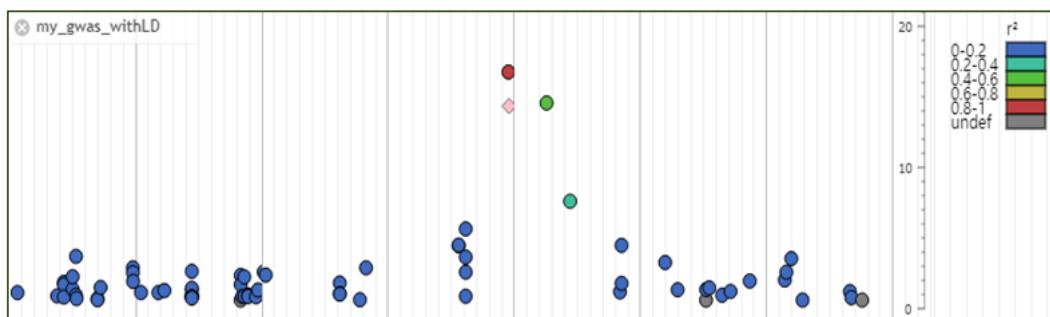
Ⓗ: Create RNA-Seq track.

<3.9 RNA-Seq & GWAS - GWAS >



<Figure 26. JBrowse GWAS track >

JBrowse GWAS track. X axis shows Genome coordinate and t axis shows $-\log(p\text{-value})$. 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.

The screenshot shows the "GWAS" upload interface. The top section has a red box around the "Upload GWAS data" field, which contains a "Browse..." button, a "No file selected" message, and a radio button labeled "A" for "with LD data". To the right is a checkbox labeled "B". Below this is a "Upload Data to :" dropdown menu with "Arabidopsis" selected, circled with "C". The next row has a "track name" input field with "GWAS_RESULT" and a "max Height (minimum p value)" input field with "1e-20", both circled with "D" and "E" respectively. At the bottom left is an "Upload data" button circled with "F", and at the bottom right is a "Refresh input file & list" button.

<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 $-\log(p\text{-value})$. Specify the maximum value for y axis.
- Ⓕ: Upload GWAS track.

The screenshot shows a user interface for LD analysis. At the top, there is a header labeled "LD". Below it, a section titled "Upload LD data" contains a "Browse..." button and a text input field showing "No file selected". A red circle labeled "Ⓐ" is drawn around this area. Below this is a section titled "Upload Data to :" with two dropdown menus: "Arabidopsis" and "GFF", both with red circles labeled "Ⓑ" drawn around them. At the bottom left is a red-bordered "Upload data" button with a red circle labeled "Ⓒ" drawn around it. To its right is a "Refresh input file & list" button.

<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.