

Transcriptomics Part 2: Differential Expression DESeq2

Bridging the Bench-Machine Learning Gap

Dr. Emily A. Beck

Dr. Jake Searcy

In the interest of time let's check out a ShinyApp developed for handling DE analyses

iDEP.93 Load Data Pre-Process Heatmap k-Means PCA DEG1 DEG2 Pathway Genome Biocluster Network R

[Click here to load demo data](#)
and just click the tabs for some magic!

Reset

1. Select or search for your species.
Best matching species

2. Choose data type
 Read counts data (recommended)
 Normalized expression values (RNA-seq FPKM, microarray, etc.)
 Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)
Browse... No file selected

Analyze public RNA-seq datasets for 9 species
Optional: Upload an experiment design file(CSV or text)
Browse... No file selected

Check this out if you want example of our gene ids, or download gene mapping.
Optional: Gene ID Examples

Ready to load data files.

We updated instruction for local installation [here](#). The most recent database file is now publically available, free of charge for non-profit organizations.

Check out the 50,000+ datasets of uniformly processed public RNA-seq data [here!](#)

We have switched back to Ensembl Release 103, which is used in iDEP 0.93. The previous database issue only affects pathway analysis in the mouse. So we replaced the mouse pathway data with that derived from Release 100. This is a temporary fix. Old versions of iDEP can still be used. See the R tab.

NEW 9/3/21: Live in testing mode [iDEP v0.94](#) based on Ensembl Release 104 and STRING-db v11.
If your gene IDs are not recognized, please let us know. We might be able to add customized gene mappings to Ensembl gene IDs.

New version 0.93 released on 5/23/2021 includes upgrades to R 4.05, Bioconductor 3.12, larger database (5000+ species) from Ensembl Release 103 and STRING-db v11. Massive, manually-collected pathway database for 20 model organisms. Fixed KEGG pathway chart and gene plot.

We recently hired Jenny Qi for database updates and user support. [Email Jenny](#) for questions.
iDEP has not been thoroughly tested. Please let us know if you find any issue/bug.
We will be happy to help prepare your data for iDEP.

?

```
graph TD; A[Read counts or FPKMs] --> B[Remove lowly expressed genes]; B --> C[Convert gene IDs]; C --> D[DESeq2 / limma]; C --> E[Transform data (log, vst)]; D --> F[Fold changes]; D --> G[Gene lists]; E --> H[K-means clustering]; E --> I[PCA]; E --> J[Hierarchical clustering]; F --> K[Pathway analysis]; G --> L[Enrichment analysis]; H --> M[Biclustering & Co-expression networks]; K --> N[Visualize on genomes]; L --> O[GO terms, KEGG pathways, TF target genes, miRNA target genes]; M --> O;
```

iDEP workflow

<http://bioinformatics.sdsstate.edu/idep93/>

In the interest of time let's check out a ShinyApp developed for handling DE analyses

iDEP.93 Load Data Pre-Process Heatmap k-Means PCA DEG1 DEG2 Pathway Genome Biocluster Network **R**

[Click here to load demo data](#)
and just click the tabs for some magic!

Reset

1. Select or search for your species.
Best matching species

2. Choose data type
 Read counts data (recommended)
 Normalized expression values (RNA-seq FPKM, microarray, etc.)
 Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)
Browse... No file selected

Analyze public RNA-seq datasets for 9 species
Optional: Upload an experiment design file(CSV or text)
Browse... No file selected

Check this out if you want example of our gene ids, or download gene mapping.
Optional: Gene ID Examples

Ready to load data files.

We updated instruction for local installation [here](#). The most recent database file is now publically available, free of charge for non-profit organizations.

Check out the 50,000+ datasets of uniformly processed public RNA-seq data [here!](#)

We have switched back to Ensembl Release 103, which is used in iDEP 0.93. The previous database issue only affects pathway analysis in the mouse. So we replaced the mouse pathway data with that derived from Release 100. This is a temporary fix. Old versions of iDEP can still be used. See the R tab.

NEW 9/3/21: Live in testing mode iDEP v0.94 based on Ensembl Release 104 and STRING-db v11.

If your gene IDs are not recognized, please let us know. We might be able to add customized gene mappings to Ensembl gene IDs.

New version 0.93 released on 5/23/2021 includes upgrades to R 4.05, Bioconductor 3.12, larger database (5000+ species) from Ensembl Release 103 and STRING-db v11. Massive, manually-collected pathway database for 20 model organisms. Fixed KEGG pathway chart and gene plot.

We recently hired Jenny Qi for database updates and user support. Email Jenny for questions.

iDEP has not been thoroughly tested. Please let us know if you find any issue/bug.

We will be happy to help prepare your data for iDEP.

```
graph TD; A[Read counts or FPKMs] --> B[Remove lowly expressed genes]; B --> C[Convert gene IDs]; C --> D[DESeq2 / limma]; C --> E[Transform data (log, vst)]; D --> F[Fold changes]; D --> G[Gene lists]; E --> H[K-means clustering]; E --> I[PCA]; E --> J[Hierarchical clustering]; F --> K[Pathway analysis]; G --> L[Enrichment analysis]; H --> M[Bioclustering & Co-expression networks]; K --> N[Visualize on genomes]; L --> N; M --> O[GO terms, KEGG pathways, TF target genes, miRNA target genes];
```

iDEP workflow

<http://bioinformatics.sdsstate.edu/idep93/>

What files do I need to perform DE analysis?

X	T100	T101	T102	T103	T104	T105	T106	T107	T108	T109	T300	T304	T305	T306	T307	T308	T309	T311	T312	T313
ENSMUG0000005947	37	59	64	80	71	71	45	45	57	84	56	44	32	58	49	37	72	64	66	32
ENSMUG0000015800	408	661	674	946	881	843	1114	683	786	958	562	360	672	335	550	222	500	577	393	456
ENSMUG0000015802	19	52	37	73	37	35	51	37	57	54	24	15	19	24	33	21	29	20	13	26
ENSMUG0000015804	1	2	3	17	1	11	5	4	3	7	1	1	1	1	3	2	1	2	1	0
ENSMUG0000022525	3203	3974	4798	6165	4863	5746	5193	3376	4966	8047	2368	3156	3476	2890	4457	2257	3786	3623	3591	2823
ENSMUG0000008350	0	0	1	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0
ENSMUG0000001817	40	47	43	61	57	59	55	38	64	71	21	31	25	27	35	20	46	39	24	28
ENSMUG0000045549	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	1
ENSMUG0000001819	6	1	1	3	2	8	5	4	5	5	2	5	5	4	9	7	3	5	1	3
ENSMUG0000000838	96	185	161	204	171	308	281	198	195	216	80	91	128	146	92	65	91	162	76	69
ENSMUG0000032293	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ENSMUG0000018699	9	24	27	36	26	33	28	23	28	27	27	4	11	21	18	10	21	24	5	18
ENSMUG0000040640	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG0000032712	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG0000033002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG0000042733	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG0000045769	0	0	1	2	4	4	4	5	1	2	4	1	3	4	0	0	0	3	1	4
ENSMUG0000014205	0	0	0	0	0	1	0	0	1	0	0	0	0	1	3	0	0	2	0	0
ENSMUG0000019518	919	997	1053	1178	1160	1015	1226	1032	1200	1121	956	710	1107	920	1173	561	896	853	738	713
ENSMUG0000019519	17	46	19	38	38	23	36	19	30	29	43	16	26	31	34	15	18	27	29	34

Count file

What files do I need to perform DE analysis?

Study_design	T100	T101	T102	T103	T104	T105	T106	T107	T108	T109	T300	T304	T305	T306	T307	T308	T309	T311	T312	T313
Group	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	
M_Diet	CTR	HFD																		
Sex	M	M	F	F	F	F	M	F	M	F	M	M	M	M	M	F	F	M	M	

Metadata file

Problem: DE packages are not good at handling “redundant” data

In the interest of time let's check out a ShinyApp developed for handling DE analyses

iDEP.93 Load Data Pre-Process Heatmap k-Means PCA DEG1 DEG2 Pathway Genome Biocluster Network R

[Click here to load demo data](#)

and just click the tabs for some magic!

1. Select or search for your species.

Best matching species

2. Choose data type

Read counts data (recommended)

Normalized expression values (RNA-seq FPKM, microarray, etc.)

Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)

Browse... No file selected

Analyze public RNA-seq datasets for 9 species

Optional: Upload an experiment design file(CSV or text)

Browse... No file selected

Check this out if you want example of our gene ids, or download gene mapping.

Optional: Gene ID Examples

Ready to load data files.

We updated instruction for local installation [here](#). The most recent database file is now publically available, free of charge for non-profit organizations.

Check out the 50,000+ datasets of uniformly processed public RNA-seq data [here!](#)

We have switched back to Ensembl Release 103, which is used in iDEP 0.93. The previous database issue only affects pathway analysis in the mouse. So we replaced the mouse pathway data with that derived from Release 100. This is a temporary fix. Old versions of iDEP can still be used. See the R tab.

NEW 9/3/21: Live in testing mode iDEP v0.94 based on Ensembl Release 104 and STRING-db v11.

If your gene IDs are not recognized, please let us know. We might be able to add customized gene mappings to Ensembl gene IDs.

New version 0.93 released on 5/23/2021 includes upgrades to R 4.05, Bioconductor 3.12, larger database (5000+ species) from Ensembl Release 103 and STRING-db v11. Massive, manually-collected pathway database for 20 model organisms. Fixed KEGG pathway chart and gene plot.

We recently hired Jenny Qi for database updates and user support. Email Jenny for questions.

iDEP has not been thoroughly tested. Please let us know if you find any issue/bug.

We will be happy to help prepare your data for iDEP.

```
graph TD; A[Read counts or FPKMs] --> B[Remove lowly expressed genes]; B --> C[Convert gene IDs]; C --> D[DESeq2 / limma]; C --> E[Transform data (log, vst)]; D --> F[Fold changes]; D --> G[Gene lists]; F --> H[Pathway analysis]; G --> I[Enrichment analysis]; E --> J[K-means clustering]; E --> K[PCA]; E --> L[Hierarchical clustering]; J --> M[Bioclustering & Co-expression networks]; M --> N[Visualize on genomes]; M --> O[GO terms, KEGG pathways, TF target genes, miRNA target genes];
```

iDEP workflow

<http://bioinformatics.sdsstate.edu/idep93/>

Start by ingesting our data

Click here to load demo data

and just click the tabs for some magic!

Reset

1. Select or search for your species.

Best matching species ▾

2. Choose data type

Read counts data (recommended)
 Normalized expression values (RNA-seq FPKM, microarray, etc.)
 Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)

Browse... No file selected

Analyze public RNA-seq datasets for 9 species

Optional: Upload an experiment design file(CSV or text)

Browse... No file selected

Check this out if you want example of our gene ids, or download gene mapping.

Optional: Gene ID Examples

Study_design	T100	T101	T102	T103	T104	T105	T106	T107	T108	T109	T300	T304	T305	T306	T307	T308	T309	T311	T312	T313
Group	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	
M_Diet	CTR	HFD																		
Sex	M	M	F	F	F	M	F	M	F	M	M	M	M	M	M	F	F	M	M	

X	T100	T101	T102	T103	T104	T105	T106	T107	T108	T109	T300	T304	T305	T306	T307	T308	T309	T311	T312	T313
ENSMUG00000005947	37	59	64	80	71	71	45	45	57	84	56	44	32	58	49	37	72	64	66	32
ENSMUG00000015800	408	661	674	946	881	843	1114	683	786	958	562	360	672	335	550	222	500	577	393	456
ENSMUG00000015802	19	52	37	73	37	35	51	37	57	54	24	15	19	24	33	21	29	20	13	26
ENSMUG00000015804	1	2	3	17	1	11	5	4	3	7	1	1	1	1	3	2	1	2	1	0
ENSMUG00000022525	3203	3974	4798	6165	4863	5746	5193	3376	4966	8047	2368	3156	3476	2890	4457	2257	3786	3623	3591	2823
ENSMUG00000008350	0	0	1	1	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0
ENSMUG00000001817	40	47	43	61	57	59	55	38	64	71	21	31	25	27	35	20	46	39	24	28
ENSMUG00000045549	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	1
ENSMUG00000001819	6	1	1	3	2	8	5	4	5	5	2	5	5	4	9	7	3	5	1	3
ENSMUG0000000838	96	185	161	204	171	308	281	198	195	216	80	91	128	146	92	65	91	162	76	69
ENSMUG00000032293	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ENSMUG00000018699	9	24	27	36	26	33	28	23	28	27	27	4	11	21	18	10	21	24	5	18
ENSMUG00000040640	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG00000032712	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG00000033002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG00000042733	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENSMUG00000045769	0	0	1	2	4	4	4	5	1	2	4	1	3	4	0	0	0	3	1	4
ENSMUG00000014205	0	0	0	0	0	1	0	0	1	0	0	0	0	1	3	0	0	2	0	0
ENSMUG00000019518	919	997	1053	1178	1160	1015	1226	1032	1200	1121	956	710	1107	920	1173	561	896	853	738	713
ENSMUG00000019519	17	46	19	38	38	23	36	19	30	29	43	16	26	31	34	15	18	27	29	34

We will go through one group together then
work on your own