

---

---

# Diabetic Neuropathy Microarray Knowledge-Base (DNMKB)

## User's Manual

(Released on 1/2/2014)

<http://jdrf.neurology.med.umich.edu/DNMKB/>

*Copyright 2013 Program for Neurology Research and Discovery.*

*All rights reserved.*

Feldman Laboratory

Program for Neurology Research and Discovery

University of Michigan

Ann Arbor, MI 48109, USA

Email: [DNMKB-help@umich.edu](mailto:DNMKB-help@umich.edu)

---

---

# Table of Contents

- Introducing DNMKB .....1**
  - Statistics..... 3
  - Accessing DNMKB..... 4
  - Features..... 4
  
- Starting DNMKB.....6**
  - Login ..... 6
  - Select Options..... 6
  
- Main Menu .....7**
  - Browse Menu ..... 7
  - Search Menu..... 8
  - Analysis Menu..... 9
  
- Understanding Results .....11**

---

---

## Introducing DNMKB

Diabetic neuropathy (DN) is the most common and debilitating complication of diabetes, but the pathogenesis is not fully understood despite extensive research. Recently, the DN research community employed high-throughput technologies to examine DN-associated transcriptomic changes in human and animal models. To comprehensively understand the complex systems associated with DN, it is critical to have a disease-specific data storage and analysis system to facilitate effective mining and seamless integration of the enormous amount of data.

Here, we present the Diabetic Neuropathy Microarray Knowledge-Base (DNMKB), a centralized repository and analysis portal of diabetic neuropathy (DN)-related transcriptomics data. DNMKB has been developed to facilitate the efficient storage and exploration of the high-volume microarray data. Table 1 lists the current data sets (as of 12/19/2013), including both published and unpublished data. While access to unpublished data is currently limited to laboratory members, it will also be made publicly available once the associated studies are published.

Table 1. Overall statistics

Total Number	
Number of experiments	13
Number of DEG sets	52
Species	Human, mouse, and rat

## Statistics

DNMKB currently contains 52 differentially expressed gene (DEG) sets from 13 DN-related microarray data sets, including transcriptomic profiles in peripheral (dorsal root ganglia, sciatic nerve, and sural nerve) and central (hippocampus) nervous tissues from several mouse models (db/db, BTBR ob/ob, high-fat diet, and Streptozotocin-induced) and human subjects. The details of the current transcriptomics data sets are summarized in Table 2. Four microarray data sets from other investigators, identified from a public microarray database ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>), were processed by our in-house analysis pipeline and included to maximize the data comparability.

---

---

## Accessing DNMKB

DNMKB is accessible at <http://jdrf.neurology.med.umich.edu/DNMKB/>. Public users can access any published microarray data, while the Feldman Lab members or collaborators have unrestricted access to the database. Member login ID and password are given by the system administrator and non-members can use the following login:

Email: **public@access**  
Password: **public**

Table 2. Summary of current transcriptomics data in DNMKB

Data Set Name	Species	Type	Genotype/BG	Age	Tissue	# of DEG sets	Treatment	Published
db/db	mouse	2	BKS db/db	8~24 wks	SCN, DRG	10		Mostly
db/db autonomic	mouse	2	BKS db/db	24 wks	AG	1		No
HighFat	mouse	2	C57BL/6	36 wks	SCN, DRG	2		No
DBA2	mouse	1	DBA2J	34 wks	SCN	3	Rosiglitazone	Yes
PIO	mouse	1 & 2	BKS db/db	16 wks	SCN, DRG	8	Pioglitazone	No
obob-male	mouse	2	BTBR ob/ob	5~13 wks	SCN	6		No
obob-female	mouse	2	BTBR ob/ob	26 wks	SCN	1		No
SOD1	mouse	NA	C57BL/6	2~30m	SCN	7		Yes
Human DN	human	1 & 2	NA	NA	Sural	2		Yes
Pub1-Rat	rat	1	Goto-Kakizaki	10 wks	Hippocampus, Cortex	2		Yes
Pub2-Rat	rat	1	Sprague-Dawley	6~8 wks	DRG	3		Yes
Pub3-Rat	rat	1	Wistar	6~13 wks	DRG	5		Yes
Pub4-Ins2Akita	mouse	1	Ins2_Akita/+	20~26 days	SCN, DRG	2		Yes

\* DRG: dorsal root ganglia, SCN: sciatic nerve, AG: autonomic ganglia

## Features

DNMKB allows users to explore the compendia of genes and biological functions (pathways) perturbed in the neuronal tissue by diabetes or drug treatment. Users can easily identify the most frequently and highly regulated genes in either all or selected datasets (across different animal models, tissues, and ages). Users can search the database using various criteria such as gene name, Entrez ID, gene symbol (or synonym), and associated biological function (Gene Ontology term and KEGG pathway name or their respective ID) and explore gene regulation across studies. All the results are downloadable in Excel spread sheets to facilitate the users' downstream analyses.

---

---

---

---

### ***Browse mode***

Users can browse the compendia of the transcriptomics datasets. The full list of differentially expressed genes (DEGs) will be available for any selected dataset using user-defined significance level and fold-change criteria.

### ***Search mode***

Users can search the database using various criteria such as gene name, Entrez ID, gene symbol (or synonym), and associated biological function (Gene Ontology term and KEGG pathway name or their respective ID).

### ***Analysis mode***

Users can perform various analyses on the DEG sets. Currently supported features include functional enrichment analysis for identifying enriched biological functions among the DEGs, gene set analysis for identifying the gene-level overlap among selected DEG sets, and transcriptional network analysis for network-level comparison of two selected DEG sets.

---

## Starting DNMKB

### Login

Click the 'LOGIN' button on the front page of DNMKB and proceed with registered or public ID and password. Public users can enter 'public@access' as email and 'public' as password, or simply click the 'Public Login' button.



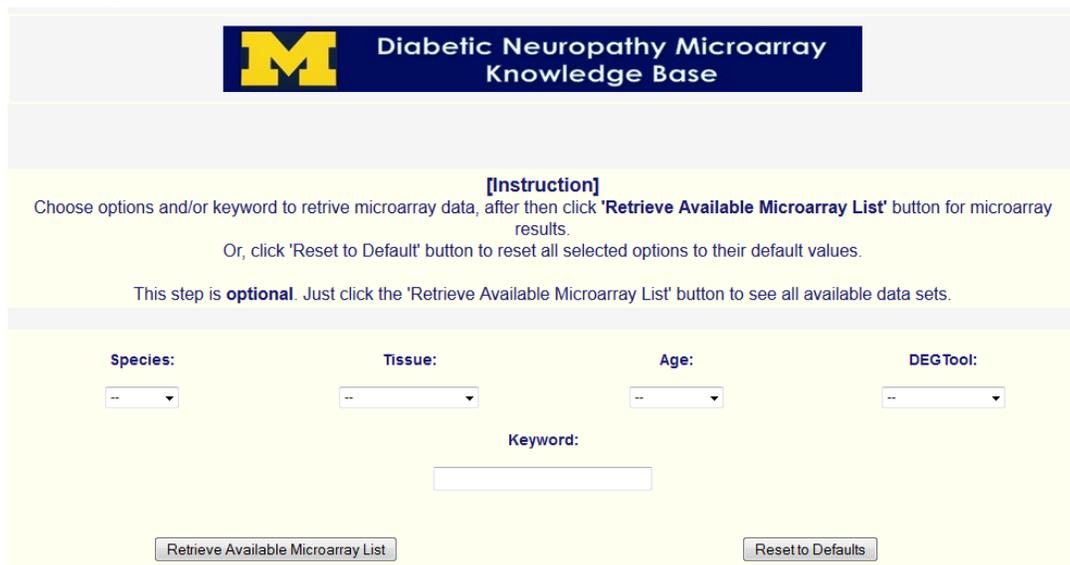
If you are a registered user, enter your email address and password.  
**Public users** can use 'public@access' as email and 'public' as password.

If you forgot your password, contact your [system administrator](#).

E-mail:   
Password:

Figure 1. Login

### Select Options



**M** Diabetic Neuropathy Microarray Knowledge Base

**[Instruction]**  
Choose options and/or keyword to retrieve microarray data, after then click '**Retrieve Available Microarray List**' button for microarray results.  
Or, click 'Reset to Default' button to reset all selected options to their default values.  
This step is **optional**. Just click the 'Retrieve Available Microarray List' button to see all available data sets.

Species:  Tissue:  Age:  DEGTool:

Keyword:

Figure 2. Select options

The first step after login is to retrieve the available microarray data sets and filter them using four major criteria and/or simple keywords.

---

- Drop-down menu: Select options for 'Species', 'Tissue', 'Age' and 'DEG tool' or type search keywords into the textbox to retrieve microarray dataset. Then click '**Retrieve Available Microarray List**' button. It should be noted that this filtering step is optional, but users must click the retrieve button to proceed.
- Reset: Click '**Reset to Defaults**' button to reset all selected options to their default values.

## Main Menu

### Browse Menu

'BROWSE' provides users with an efficient way to retrieve all DEGs with their fold-change information. The results can be sorted by different criteria such as the number of experiments having each gene as a DEG or by the fold-changes in a specific DEG set. DEGs from different species will be automatically mapped across different species using the NCBI HomoloGene database (<http://www.ncbi.nlm.nih.gov/homologene>).

Users should first select datasets to browse. Optionally, users can adjust the sorting and handling species options before retrieving the results. Users can also specify the significance value or minimum fold-change cutoff to limit the results to highly significant DEGs. If these values are not specified, the default values (each DEG set have their own default criteria) will be automatically used.

Once all options are chosen, users need to click the '**Generate Matrix**' button to proceed.

The screenshot shows the 'Diabetic Neuropathy Microarray Knowledge Base' interface. At the top, there is a logo with a yellow 'M' on a blue background. Below the logo, there are three tabs: 'BROWSE', 'SEARCH', and 'ANALYSIS'. The 'BROWSE' tab is active. The main area contains a form with the following elements:

- A dropdown menu for 'Sort by: # of experiments with DEGs'.
- A dropdown menu for 'Combine multiple species: use mouse gene as base' with a note: 'Note: Combining multiple species works with two species.'
- Buttons for 'Generate Matrix' and 'Reset'.
- Input fields for 'Significance Cut-off <=' and 'Minimum Fold-Change'.
- A section titled '----- List of available Microarray Experiments -----' containing a table.

Select	Display Order	Experiment name	Species	Tissue	Age(wks)	DEG Tool	Cutoff	Note	PubMed ID	Lab PI	Published
<input type="checkbox"/>	1	dbdb*DRG*24wk*db+ vs dbdb	Mouse	DRG	24	IBMT	0.050	db+ vs dbdb in 24wk DRG		Eva Feldman at U of Michigan	Unpublished
<input type="checkbox"/>	2	dbdb*DRG*8wk*db+ vs dbdb	Mouse	DRG	8	IBMT	0.050	db+ vs dbdb in 8wk DRG		Eva Feldman at U of Michigan	Unpublished

Figure 3. Browse menu

### Sort by

This option specifies how the retrieved DEGs in the result page are ordered. The default is '# of experiments with DEGs', putting the most frequently perturbed DEGs across multiple conditions on the top list. If a specific dataset is selected, then the genes will be sorted by the fold-change values in the selected dataset. The list of sorting options in the drop-down

---

---

menu is different based on the selected datasets resulting by the previous step.

### ***Combine multiple species***

This option specifies how genes from multiple species are handled and displayed. The default is **'use mouse gene as base'** as the majority of the datasets are using mouse.

## **Search Menu**

'SEARCH' provides users with search flexibilities to retrieve specific DEGs of interest. As in Browse menu, users can provide custom significance and fold-change cutoffs. If these values are not specified, the default values for each DEG set will be used.

Once the keywords or significance and fold-change cutoff values are typed, users click the **'Generate Matrix'** button.

### ***Search criteria***

DNMKB supports seven types of search criteria. Only one search criterion should be used for each query, although multiple keywords are allowed in selected criteria (noted as MULTI below). Allowed separators include 'semicolon', 'comma', 'tab', 'space', 'newline'.

- **Gene IDs:** Entrez gene IDs **[MULTI]**
- **Gene Symbols:** Entrez gene symbols (either official or synonyms) **[MULTI]**
- **Gene Names:** Entrez gene name (either complete or partial names)
- **GO IDs:** Gene Ontology IDs **[MULTI]**
- **GO Term:** Gene Ontology term (either complete or partial terms)
- **KEGG IDs:** KEGG Pathway IDs **[MULTI]**
- **KEGG Pathway Name:** KEGG pathway (either complete or partial names)

### ***Show non-DEGs option***

This option specifies if the result matrix will include any non-DEGs. This feature is useful in case the genes of users' interests do not show up in the matrix and users want to make sure if the genes are included in the array platform. The default is 'Do NOT show any non-DEGs'. If 'Show any non-DEGs' is selected, the following colors will be used to represent different DEG types:

- Green: included in the array and a DEG
- White: included in the array but not a DEG
- Gray: not included in the array

All other features in the 'SEARCH' menu are identical to those in the 'BROWSE' menu.

---

---



# Diabetic Neuropathy Microarray Knowledge Base

---

BROWSE
SEARCH
ANALYSIS

Search by gene symbol, name, ID, GO, KEGG pathways (accepted separators: 'semicolon', 'comma', 'tab', 'space', 'newline' for multiple terms/IDs)

[HIDE detailed search options!](#)

Search by	term/IDs
Gene IDs (examples: "20555, 20556, 20557")	
Gene Symbols (examples: "Sod1, Sod2, Sod3")	
Gene Names (example: "superoxide dismutase")	
GO IDs (examples: "GO:0010875, GO:0016787")	
GO Term (example: "oxidative stress")	
KEGG IDs (examples: "hsa01100, hsa04930")	
KEGG Pathway Name (example: "diabetes mellitus")	

\* Show non-DEGs  Do NOT show any non DEGs

\* Sort by: # of experiments with DEGs

\* Combine multiple species:

Note: Combining multiple species works with two species.

\* Significance Cut-off <=  \* Minimum Fold-Change

----- List of available Microarray Experiments -----

Select	Display Order	Experiment name	Species	Tissue	Age(wks)	DEG Tool	Cutoff	Note	PubMed ID	Lab PI	Published
<input type="checkbox"/>	1	dbdb*DRG*2wk*db+ vs dbdb	Mouse	DRG	2k	IBMT	0.050	db+ vs dbdb in 2wk DRG		Eva Feldman at U of Michigan	Unpublished
<input type="checkbox"/>	2	dbdb*DRG*8wk*db+ vs dbdb	Mouse	DRG	8k	IBMT	0.050	db+ vs dbdb in 8wk DRG		Eva Feldman at U of Michigan	Unpublished

Figure 4. Search Menu

## Analysis Menu

'ANALYSIS' provides users with further analysis tools to identify meaningful information from selected DEG sets. Three different analysis methods are currently supported in DNMBK; 'Functional Enrichment Analysis', 'Gene Set Analysis' and 'Transcriptional Network Analysis'.



# Diabetic Neuropathy Microarray Knowledge Base

---

BROWSE
SEARCH
ANALYSIS

\* Functional Enrichment Analysis

- Cellular Component
- Biological Process
- Molecular Function
- KEGG Pathway

\* Gene Set Analysis

\* Transcriptional Network Analysis

\* Significance Cut-off <=  \* Minimum Fold-Change

----- List of available Microarray Experiments -----

Select	Toggle	Display Order	Experiment name	Species	Tissue	Age(wks)	DEG Tool	Cutoff	Note	PubMed ID	Lab PI	Published
<input type="checkbox"/>		1	dbdb*DRG*2wk*db+ vs dbdb	Mouse	DRG	2wk	IBMT	0.050	db+ vs dbdb in 2wk DRG		Eva Feldman at U of Michigan	Unpublished
<input type="checkbox"/>		2	dbdb*DRG*8wk*db+ vs dbdb	Mouse	DRG	8wk	IBMT	0.050	db+ vs dbdb in 8wk DRG		Eva Feldman at U of Michigan	Unpublished

Figure 5. Analysis menu

---

---

### ***Functional Enrichment Analysis (FEA)***

Gene Ontology (GO; <http://www.geneontology.org/>) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/>) pathway information are used in FEA. Basically, GO is classified by three groups; Cellular Component, Biological Process and Molecular Function. Therefore, DNMKB provides four different categories to perform the functional enrichment analysis.

Users can select as many DEG sets as they want and then click the '**Generate Matrix**' button. FEA generates heat maps for selected DEG sets. Users can click heat map icons to see the bigger images for the maps. If users want to download all the information generated by FEA, click 'Download the complete results ([FuncEnrichment.zip](#))'

### ***Gene Set Analysis (GSA)***

GSA is done by clicking the 'Generate Venn-Diagram' button after selecting between two and 5 DEG sets. GSA provides a Venn-diagram showing the number of overlapped gene sets between them.

### ***Transcriptional Network Analysis (TNA)***

TNA identifies conserved transcriptional networks between two DEG sets based on the gene-gene co-citation data. Sentence- and abstract-level gene-gene co-citation information was mined by using SciMiner on the complete PubMed abstracts (over 21 millions). Currently, DNMKB performs the sentence-level analysis by default. Once gene-gene co-citation networks are generated for the selected DEG sets, a graphical analysis tool TALE identifies sub-networks shared by two DEG networks. Since it takes long time to generate a final graph using TALE, TNA will provide users with the URL where the results, once ready, will be displayed.

# Understanding Result Tables

The figure below illustrates an example of the matrix.

## Browse Results

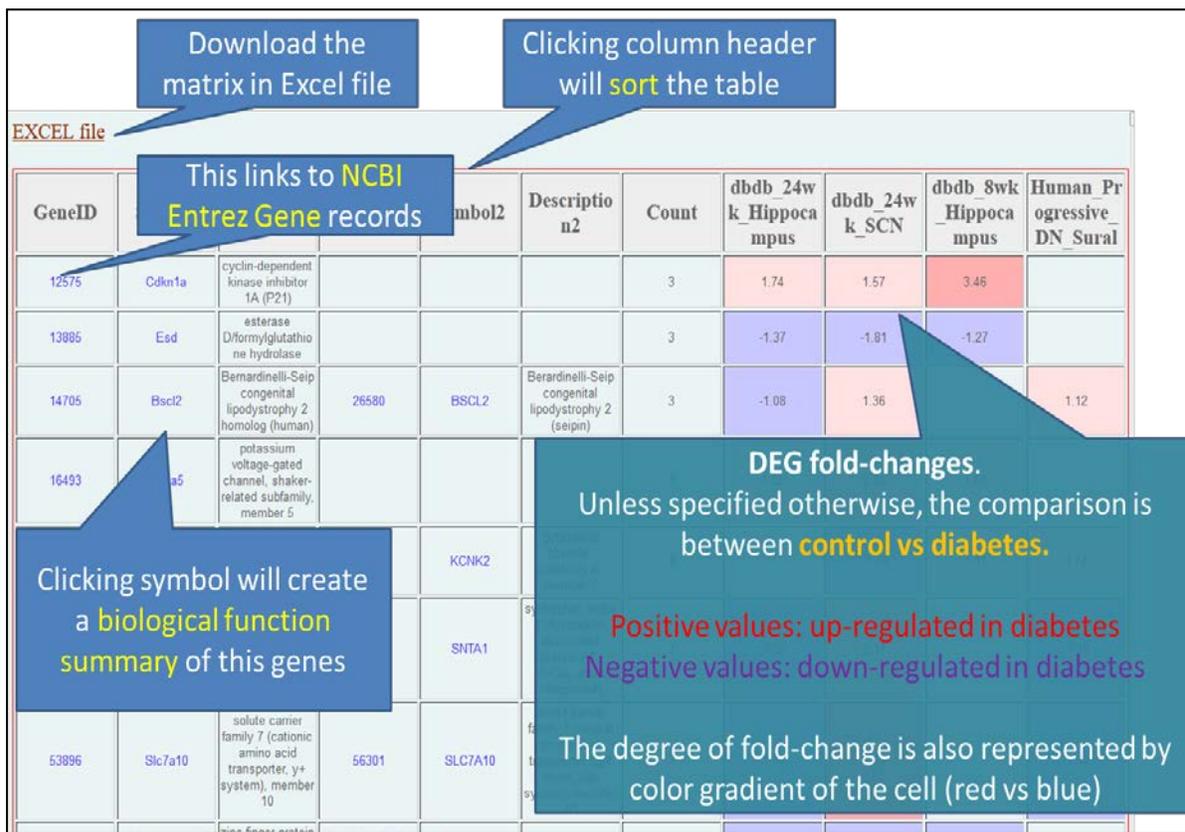


Figure 6. Browse Menu Result Table

- The matrix is downloadable in Excel file.
- Clicking the column headers will sort the table.
- Clicking gene IDs will show the detailed gene information (NCBI Entrez Gene database)
- Clicking symbols will create a summary page of biological functions (GO and KEGG pathway) associated with the selected gene.
- The values correspond to the fold-changes between control and diabetes, unless specified otherwise. **Positive values: up-regulated in diabetes** and **negative values: down-regulated in diabetes**. The degree of fold-change is also represented by color gradient of the cell (red vs blue)

As shown above, clicking symbols will create a summary page for biological functions in terms of GO and KEGG pathway associated with the selected gene (in a new window). Depending on the number of associated function, the loading time of this page may take

up to a minute. So, be patient.

The current DNMKB displays not only those explicitly assigned GO terms but also those implicitly assigned GO terms as well, which can be inferred from the explicitly assigned GO terms and the hierarchical GO structure. Future version will allow users to select which sets of GO terms to use (explicitly assigned terms are less in number, thus taking much less time to load).

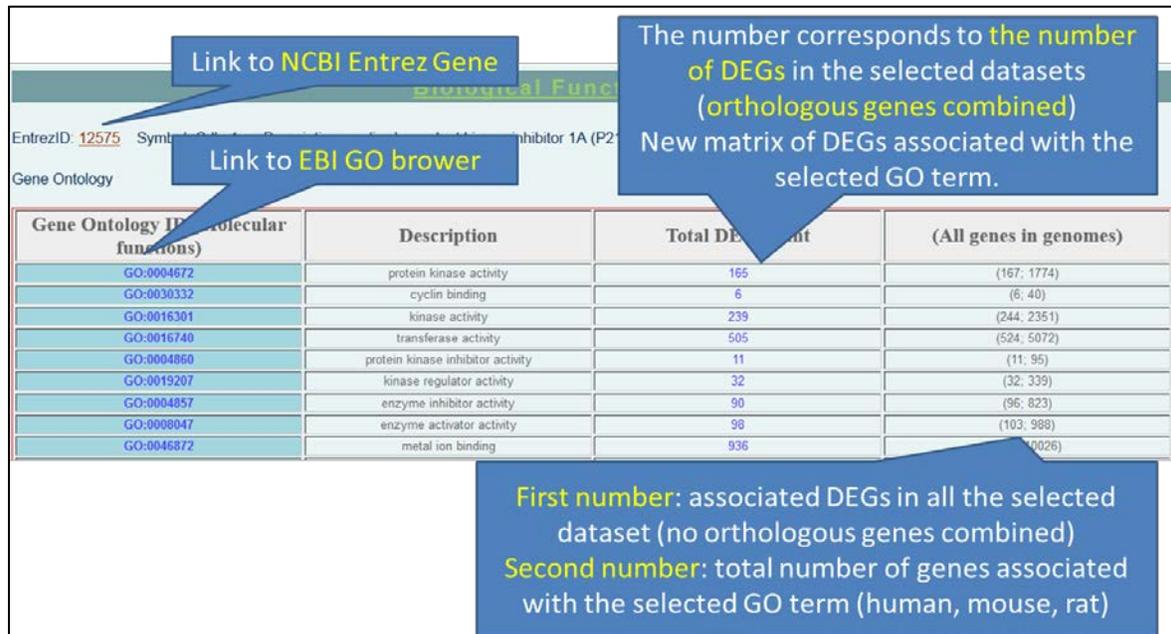


Figure 6. Summary Table for Biological Functions

## Search Results

The 'SEARCH' result table is similar to the 'BROWSE' menu except that in the 'SEARCH' menu users can specifically search for the DEGs in the database using various criteria.

## Analysis Results

### Functional Enrichment Analysis

The results of FEA are gene annotation information of the enrichment analysis and clustered heat-map images of top functions. DNMKB provides the analysis results both in text format as well as Excel format, facilitating users to perform additional down-stream analyses of the DEGs using other tools. By default, biological functions in terms of GO terms and KEGG pathways with a Benjamini-Hochberg (BH) corrected P-value < 0.05 are deemed significant and will be included in the heat-map. The heat-map will include the top 10 most over-represented biological functions in each DEG set, clustered based on the significance values (log-transformed BH-corrected P-values), to visually represent overall

similarity and difference between the DEG sets.

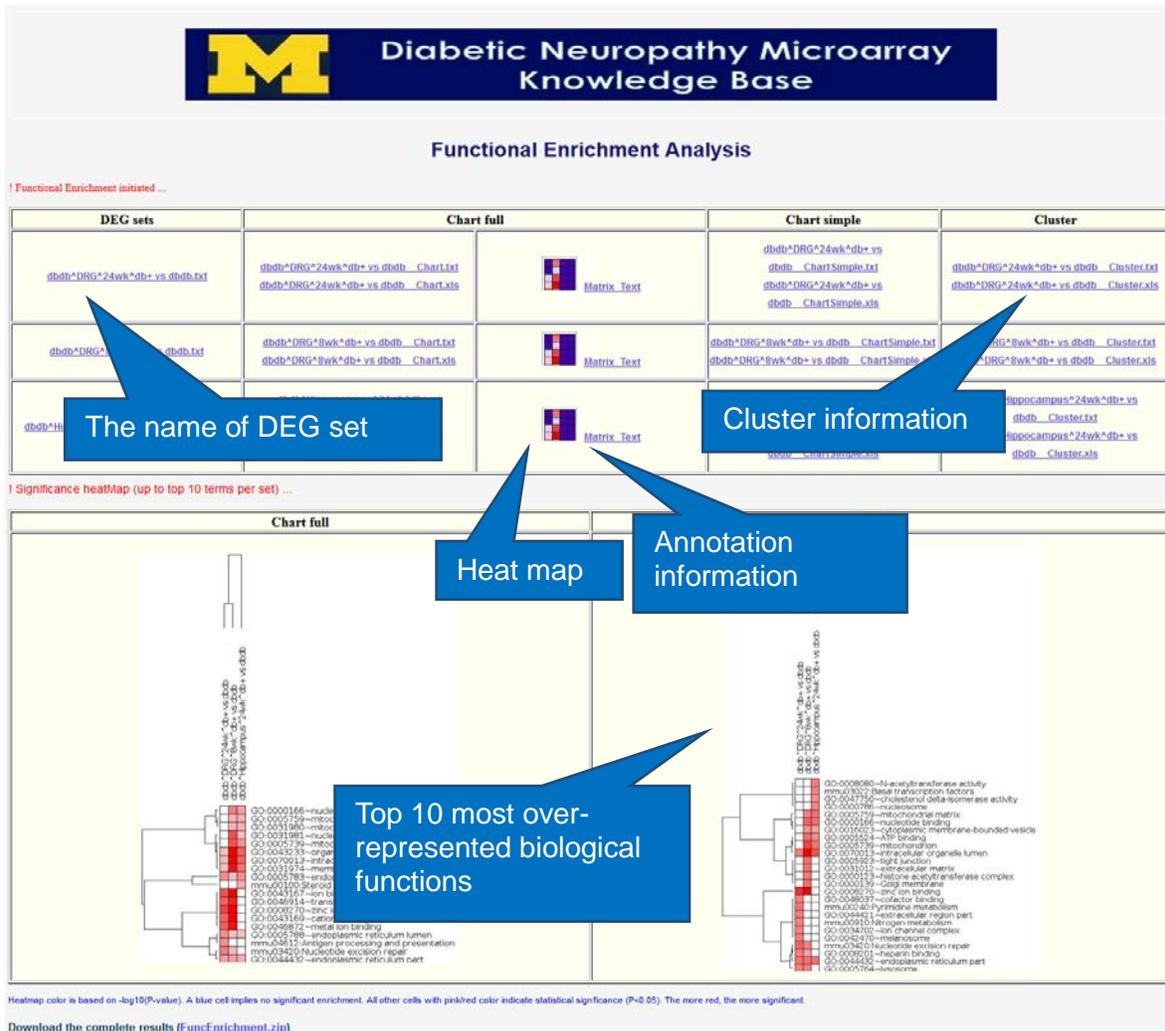
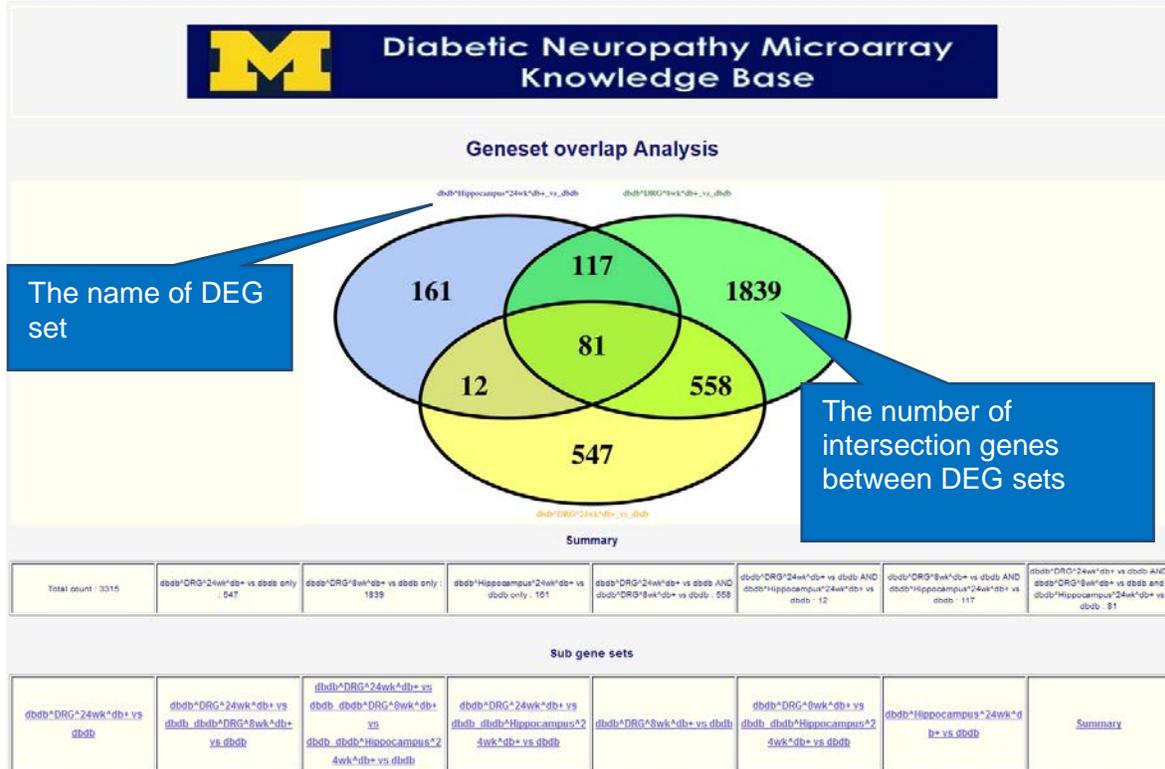


Figure 7. Functional Enrichment Analysis Results

### Gene Set Analysis (GSA)

The result of GSA is a Venn-diagram showing the overlap between DEG sets. DNMBK also provides the list of overlapping genes, available for download.



#### Summary

Total count : 3315	dbdb*DRG*24wk*db+ vs dbdb only : 547	dbdb*DRG*8wk*db+ vs dbdb only : 1839	dbdb*Hippocampus*24wk*db+ vs dbdb only : 101	dbdb*DRG*24wk*db+ vs dbdb AND dbdb*DRG*8wk*db+ vs dbdb : 558	dbdb*DRG*24wk*db+ vs dbdb AND dbdb*Hippocampus*24wk*db+ vs dbdb : 12	dbdb*DRG*8wk*db+ vs dbdb AND dbdb*Hippocampus*24wk*db+ vs dbdb : 117	dbdb*DRG*24wk*db+ vs dbdb AND dbdb*DRG*8wk*db+ vs dbdb AND dbdb*Hippocampus*24wk*db+ vs dbdb : 81
--------------------	--------------------------------------	--------------------------------------	--	--	--	--	---

#### Sub gene sets

dbdb*DRG*24wk*db+ vs dbdb	dbdb*DRG*24wk*db+ vs dbdb dbdb*DRG*8wk*db+ vs dbdb	dbdb*DRG*24wk*db+ vs dbdb dbdb*DRG*8wk*db+ vs dbdb dbdb*Hippocampus*24wk*db+ vs dbdb	dbdb*DRG*24wk*db+ vs dbdb dbdb*Hippocampus*24wk*db+ vs dbdb	dbdb*DRG*8wk*db+ vs dbdb	dbdb*DRG*8wk*db+ vs dbdb dbdb*Hippocampus*24wk*db+ vs dbdb	dbdb*Hippocampus*24wk*db+ vs dbdb	Summary
---------------------------	--	--	---	--------------------------	--	-----------------------------------	---------

Figure 8. Gene Set Analysis Results

### **Transcriptional Network Analysis (TNA)**

TAN will be performed between two selected DEG sets. DNMKB will use 6 levels of allowed mismatch (a.k.a. approximate parameter in TALE) ranging from 0 (perfect match) to 0.5 (allowing up to 50% mismatch). Summary of the shared sub-networks is given at the bottom of the result page along with links to the network format file (.gml), which can be loaded into Cytoscape (<http://www.cytoscape.org/>), an open-source network visualization tool. A simple image will be generated for those networks with less than 300 nodes. The list of genes (as entrez gene IDs) will be available for download as well.

## DNMKB - Co-Citation Network Analysis

Congratulation! Your co-citation network analysis has been successfully completed. Your selected differentially expressed sets were

Details of selected DEG sets

[[HIDE analysis option summary](#)]

Larger DEG set	DNMKB exp ID	39
	Number of DEGs	4017
Shared sub-networks identified by TALE	Name	dbdb*SCN*24wk*db+ vs dbdb.txt
	DNMKB exp ID	5
	Number of DEGs	4938
HumanDN-CI*Sural*NA*NonProg vs Prog.txt		

Network file (.gml) can be loaded onto Cytoscape

Shared sub-networks

Network file (.gml) can be loaded into [Cytoscape](#), a open-source network visualization tool. Please, network images not generated on the server for a network with over 300 nodes.

Shared gene lists

Network-number	% of allowed mismatch	% of seed nodes	# of nodes in shared network	# of nodes in shared network	Network file (.gml)	Network images	Gene annotation
1	0%	10%	1	1	<a href="#">Download network file (.gml)</a>		<a href="#">Genes</a>
2	10%	10%	153	2489	<a href="#">Download network file (.gml)</a>		<a href="#">Genes</a>
3	20%	10%	404	5591	<a href="#">Download network file (.gml)</a>	Network is too big to show	<a href="#">Genes</a>
4	30%	10%	595	6774	<a href="#">Download network file (.gml)</a>	Network is too big to show	<a href="#">Genes</a>
5	40%	10%	742	7260	<a href="#">Download network file (.gml)</a>	Network is too big to show	<a href="#">Genes</a>
6	50%	10%	797	7382	<a href="#">Download network file (.gml)</a>	Network is too big to show	<a href="#">Genes</a>

**END OF THE USER'S MANUAL**