NCIBI Workshop

Module 2

Gene Set Enrichment Testing and Concept Mapping

National Center for Integrative Biomedical Informatics (NCIBI)
The University of Michigan Medical School
Palmer Commons Room 2017 Palmer Commons
100 Washtenaw Avenue Ann Arbor, MI 48109-2218
Phone: 734-615-5510 Fax: 734-615-6553

web site: http://www.ncibi.org/
email: ncibiinfo@umich.edu

Getting started with ConceptGen

What is ConceptGen?

ConceptGen is a *gene set enrichment* and *concept mapping tool* that can help you identify and visualize relationships and significant overlaps among sets of genes (concepts). ConceptGen is built on a repository of conceptual data drawn from diverse areas:

Functional annotations	GO Biological Process GO Molecular Function GO Cellular Component Protein Domains (Pfam) KEGG Pathways The Protein ANalysis THrough Evolutionary Relationships (Panther) database Biocarta Pathways
Literature derived	Medical Subject Headings (MeSH), Online Mendelian Inheritance in Man (OMIM)
Targets	Drug target sets (DrugBank) microRNA (miRBase) transcription factor binding targets (TransFac)
Experimental	Gene Expression datasets (GEO)
Interactions	Protein-interaction datasets (MiMI)
Other	Metabolites and Cytoband (chromosomal locations)

ConceptGen finds gene sets that are significantly over-represented from among the sub-categories (concepts) comprising each of the above data sources in another public or user-defined gene set. Significance of over-representation is measured by a modified Fisher's exact test (p-value) and is also shown by q-values. Q-values take into account the estimated proportion of false positives incurred (the false discovery rate) based on p-values.

What Can You Do with ConceptGen?

You can use ConceptGen in two ways, as follows:

- *For gene set enrichment testing*: Upload a list of genes of interest (larger lists tend to yield better results) and a descriptive name for the Concept. Query on your Concept to find other pre-defined concepts that are significantly enriched with genes in your uploaded set, and relationships among those concepts.

- *As a concept mapping tool*: Enter a term/topic of interest in the Search box. ConceptGen retrieves and displays predefined concepts that are semantically close to your entry. From the retrieved results, select one for your query. You can find other concepts which contain an overrepresented number of genes contained in the chosen Concept, and explore the network of relationships.

In this tutorial you will

► Login, upload a gene list, and give it a concept name (Bipolar-Smoking-MiMI)
► Query to find concepts related to your gene list (concept) and save them
► Find and save genes that overlap between your list and other concepts of interest
► Learn to filter by concept types of interest and/or by enrichment statistics
► Explore concept networks to see interrelations between concepts and their genes
► Save genes comprised in various concepts for later comparisons
► Link from ConceptGen to MiMI

You **will not** upload a background gene set, which should be used for microarray datasets or other any other dataset for which the complete human gene list was not assessed/measured. You may want to explore this and other additional features on your own.

Data Set and Case for Exploration

Data: Get the set of 89 candidate genes for this tutorial from http://portal.ncibi.org/gateway/virtual-workshop.html. Download it to your desktop to read into Conceptgen.

The 89 genes in this dataset were derived from five original genes that are known through experimentation to be associated with bipolar disorder and Tobacco Use Disorder. They are:

COMT, BDNF, MAOA, SLC6A4, and SLC6A3

None of the five has yet been proven to be associated with both disorders in a single manuscript, but all five have been studied in relation to both bipolar disorder and tobacco use disorder separately.

Complex diseases, including Bipolar Disorder (BD) are characterized by multiple genetic and environmental influences on susceptibility. Many genetic loci have been associated with BD, though replication of results has been challenging (Molecular genetics of bipolar disorder and depression, Kato T., Psychiatry Clin Neurosci. 2007 Feb;61(1):3-19, PMID: 17239033) and (Molecular genetics of bipolar disorder, Hayden EP, Nurnberger JI Jr., Genes Brain Behav. 2006 Feb;5(1):85-95, PMID: 16436192). Notably, BD is characterized by a high rate of co-morbid Tobacco Use Disorder (TUD), a condition that is also influenced by genetic variation.

Your exploration is directed toward uncovering genes that may be involved in both disorders. You have already done some exploratory analysis and have found 84 genes as the nearest neighbors of the five candidate genes (for a total of 89 genes). You now will use Conceptgen to refine this list of 89 genes based on conceptual associations. Conceptgen will also help you better understand the roles that these genes may play in the co-morbidity.

Login to Upload and Name a Gene List

Login

1. Access ConceptGen at: http://conceptgen.ncibi.org.
2. Click "Login" in the top right of the screen. A login box appears. If you have not yet registered, click on "Register" to first create a private account.
3. Type in your login email and password. Click the "Login" button.
   A screen appears listing gene lists you have already named as concepts. If you have not yet created a concept, the screen is blank.

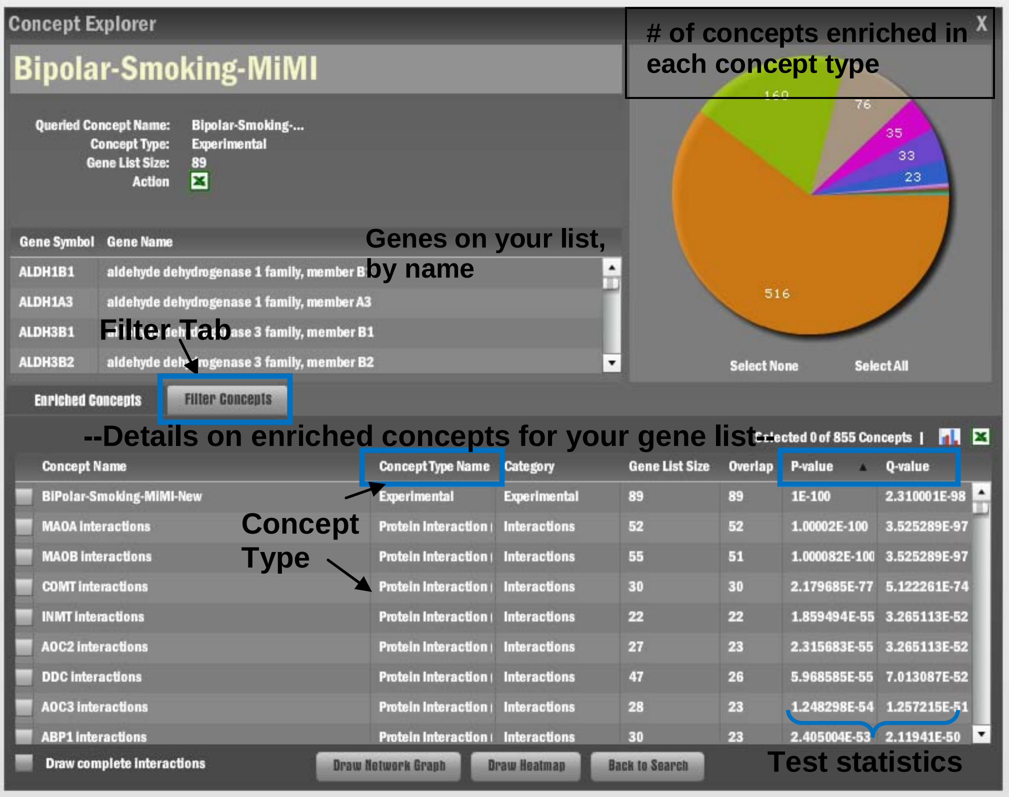
Upload a Gene List and View its Related Concepts

1. Click on "Upload Concept" or the green "+" button next to My Concepts.
2. Type in the new gene list name: Bipolar-Smoking-MiMI
3. Click the radio button Entrez Gene IDs because the gene list is in this format.
   Open the gene list file and copy and paste the IDs into the box. Click "Upload Gene List." The "My Concepts" page redisplays, now with your gene list included.
4. On "My Concepts," click the name of the list or the arrow at the end of the row. The Concept Explorer Results screen appears (see Figure 1).

The Explorer Window is the main results display from which you can do many tasks.

Query to Find/Save Concepts Related to Your Concept (Gene List)

1. Examine the Concept Explorer screen to find related concepts (Fig. 1).



Concept Explorer

Bipolar-Smoking-MiMI

Queried Concept Name:	Bipolar-Smoking-...
Concept Type:	Experimental
Gene List Size:	89
Action	☒

Gene Symbol	Gene Name
ALDH1B1	aldehyde dehydrogenase 1 family, member B1
ALDH1A3	aldehyde dehydrogenase 1 family, member A3
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
ALDH3B2	aldehyde dehydrogenase 3 family, member B2

Enriched Concepts | Filter Concepts

Selected 0 of 855 Concepts |

	Concept Name	Concept Type Name	Category	Gene List Size	Overlap	P-value ▲	Q-value
□	BiPolar-Smoking-MIMI-New	Experimental	Experimental	89	89	1E-100	2.310001E-98
□	MAOA interactions	Protein Interaction	Interactions	52	52	1.00002E-100	3.525289E-97
□	MAOB interactions	Protein Interaction	Interactions	55	51	1.000082E-100	3.525289E-97
□	COMT interactions	Protein Interaction	Interactions	30	30	2.179685E-77	5.122261E-74
□	INMT interactions	Protein Interaction	Interactions	22	22	1.859494E-55	3.265113E-52
□	AOC2 interactions	Protein Interaction	Interactions	27	23	2.315683E-55	3.265113E-52
□	DDC interactions	Protein Interaction	Interactions	47	26	5.968585E-55	7.013087E-52
□	AOC3 interactions	Protein Interaction	Interactions	28	23	1.248298E-54	1.257215E-51
□	ABP1 interactions	Protein Interaction	Interactions	30	23	2.405004E-53	2.11941E-50

□ Draw complete interactions

# of concepts enriched in each concept type

Genes on your list, by name

Filter Tab

--Details on enriched concepts for your gene list--

Concept Type

Test statistics

Figure 1: The Concept Explorer Results page.

2. Verify the genes contained in Bipolar-Smoking-MiMI by scrolling through the top left list.
3. See the sources from which the associated concepts are derived in the pie chart. Mouse over the pie chart to see the names of the concept types.

Familiarize yourself with the retrieved concepts by scrolling though the lower table. Click the column heading Concept Type Name to sort the concept types alphabetically. Concepts are retrieved and displayed in the Explorer if their q value is <0.05. For the MeSH concept type, you can find that Bipolar Disorder and Smoking are significantly enriched as you would expect from your prior research. You can re-sort the list by most significant p and q values by clicking on either column heading.

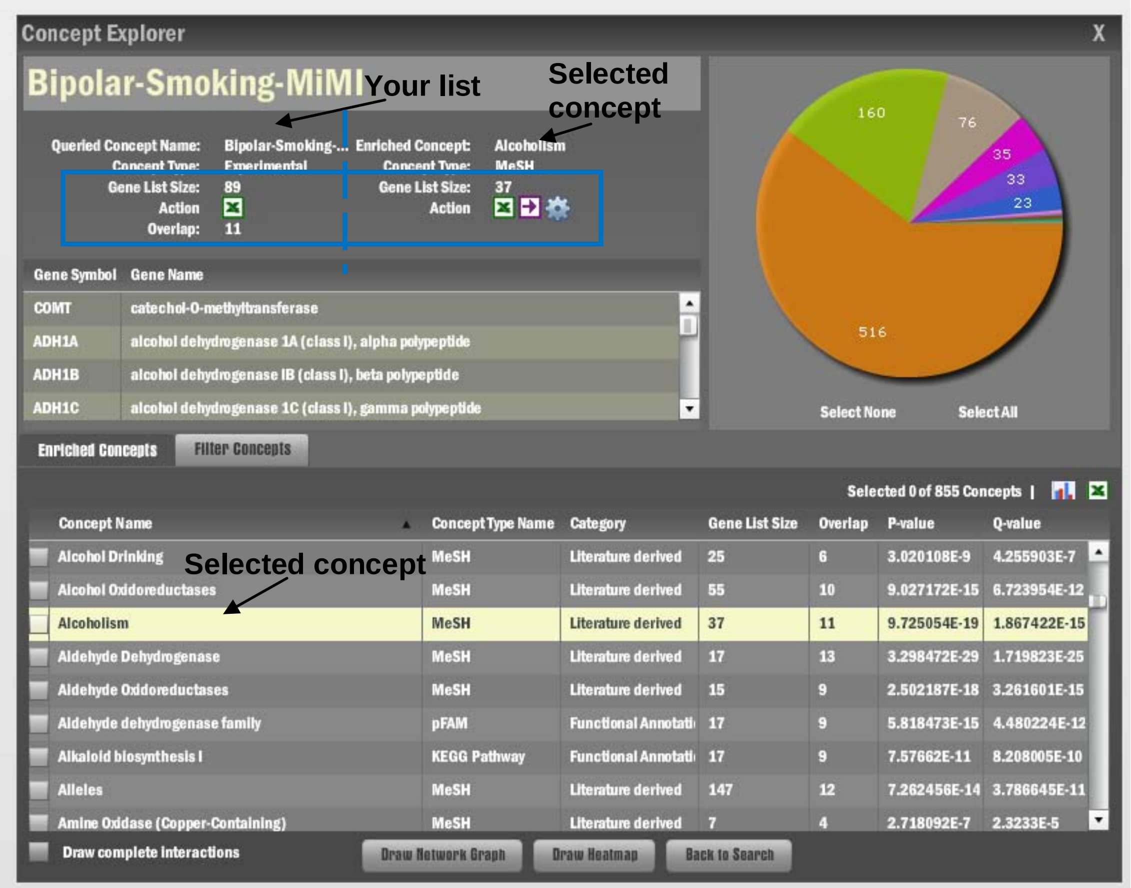
4. Save the table for future reference. Click the Excel icon on the far right(mid-screen). In the dialogue specify where to save it and then download it.

Find Genes Shared by Your List and Other Concepts

1. Look for GO annotations of interest by clicking on the Concept Type column header. The column sorts alphabetically. Scroll to GO Molecular Function.

2. Sort by “Concept name” alphabetically, and find the MeSH term “Alcoholism”. This is a known co-morbidity of Bipolar disorder and Tobacco use disorder. For “Alcoholism”, you see that 11 of the 37 genes annotated to this concept are from your list. Click on this row.

4. The top box that displays gene names now displays the genes that overlap between your list (the Bipolar-smoking-MiMI concept) and Alcoholism (Fig. 2) Save them by clicking the Excel icon next to Export Data (within the rectangle in Figure 2). Be sure to name it in a way that helps you remember that it has just the genes from your list of 89 that are also annotated for alcoholism.



Bipolar-Smoking-MiMI

Queried Concept Name:	Bipolar-Smoking-...	Enriched Concept:	Alcoholism
Concept Type:	Experimental	Concept Type:	MeSH
Gene List Size:	89	Gene List Size:	37
Action	☒	Action	☒ ➔ ⚙
Overlap:	11		

Gene Symbol	Gene Name
COMT	catechol-O-methyltransferase
ADH1A	alcohol dehydrogenase 1A (class I), alpha polypeptide
ADH1B	alcohol dehydrogenase IB (class I), beta polypeptide
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide

Selected 0 of 855 Concepts |

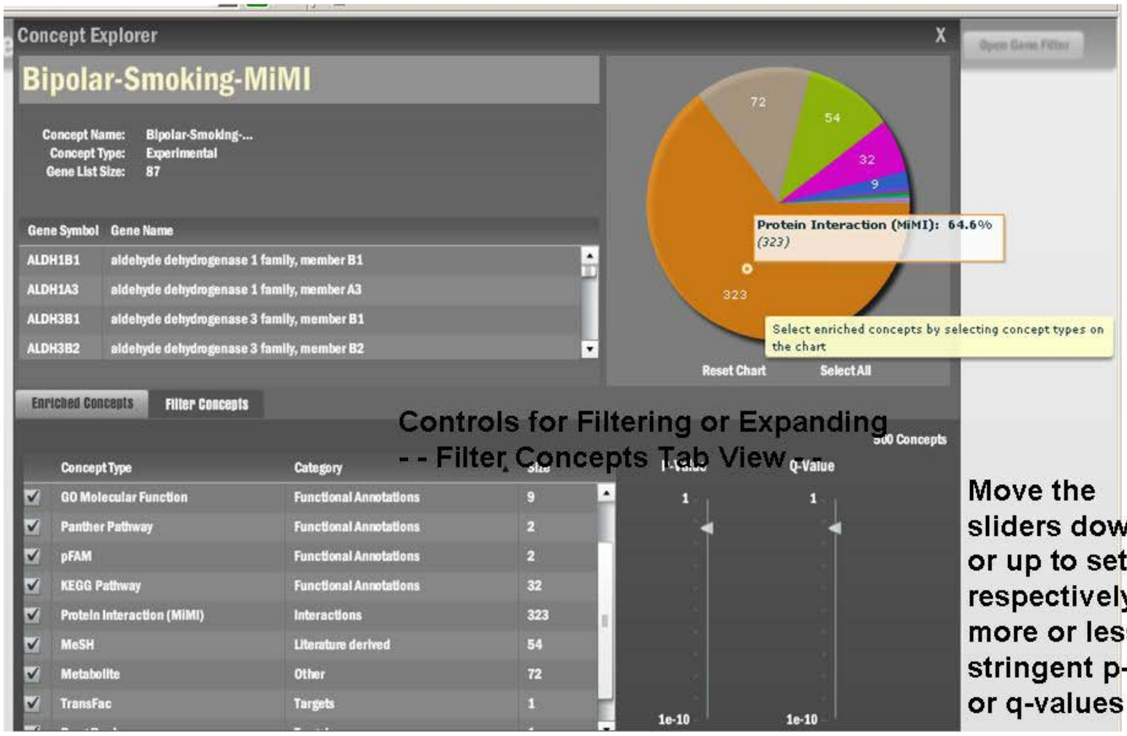
	Concept Name	Concept Type Name	Category	Gene List Size	Overlap	P-value	Q-value
□	Alcohol Drinking	MeSH	Literature derived	25	6	3.020108E-9	4.255903E-7
□	Alcohol Oxidoreductases	MeSH	Literature derived	55	10	9.027172E-15	6.723954E-12
□	Alcoholism	MeSH	Literature derived	37	11	9.725054E-19	1.867422E-15
□	Aldehyde Dehydrogenase	MeSH	Literature derived	17	13	3.298472E-29	1.719823E-25
□	Aldehyde Oxidoreductases	MeSH	Literature derived	15	9	2.502187E-18	3.261601E-15
□	Aldehyde dehydrogenase family	pFAM	Functional Annotati	17	9	5.818473E-15	4.480224E-12
□	Alkaloid biosynthesis I	KEGG Pathway	Functional Annotati	17	9	7.57662E-11	8.208005E-10
□	Alleles	MeSH	Literature derived	147	12	7.262456E-14	3.786645E-11
□	Amine Oxidase (Copper-Containing)	MeSH	Literature derived	7	4	2.718092E-7	2.3233E-5

□ Draw complete interactions

Figure 2. Viewing genes that overlap between your list (concept) and another.

Filter Based on Interest and Threshold Enrichment Statistics

1. Filter out rows that do not interest you by clicking on the "Filter Concepts" tab next to the "Enriched Concepts" tab (your current display). A new tab-based view appears, as pictured in Figure 3.



Concept Explorer

Bipolar-Smoking-MiMI

Concept Name: Bipolar-Smoking-...
Concept Type: Experimental
Gene List Size: 87

Gene Symbol	Gene Name
ALDH1B1	aldehyde dehydrogenase 1 family, member B1
ALDH1A3	aldehyde dehydrogenase 1 family, member A3
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
ALDH3B2	aldehyde dehydrogenase 3 family, member B2

Protein Interaction (MiMI): 64.6% (323)

Select enriched concepts by selecting concept types on the chart

500 Concepts

	Concept Type	Category	Size
☑	GO Molecular Function	Functional Annotations	9
☑	Panther Pathway	Functional Annotations	2
☑	pFAM	Functional Annotations	2
☑	KEGG Pathway	Functional Annotations	32
☑	Protein Interaction (MiMI)	Interactions	323
☑	MeSH	Literature derived	54
☑	Metabolite	Other	72
☑	TransFac	Targets	1

Controls for Filtering or Expanding
- - Filter Concepts Tab View - -

Move the sliders down or up to set, respectively, more or less stringent p- or q-values

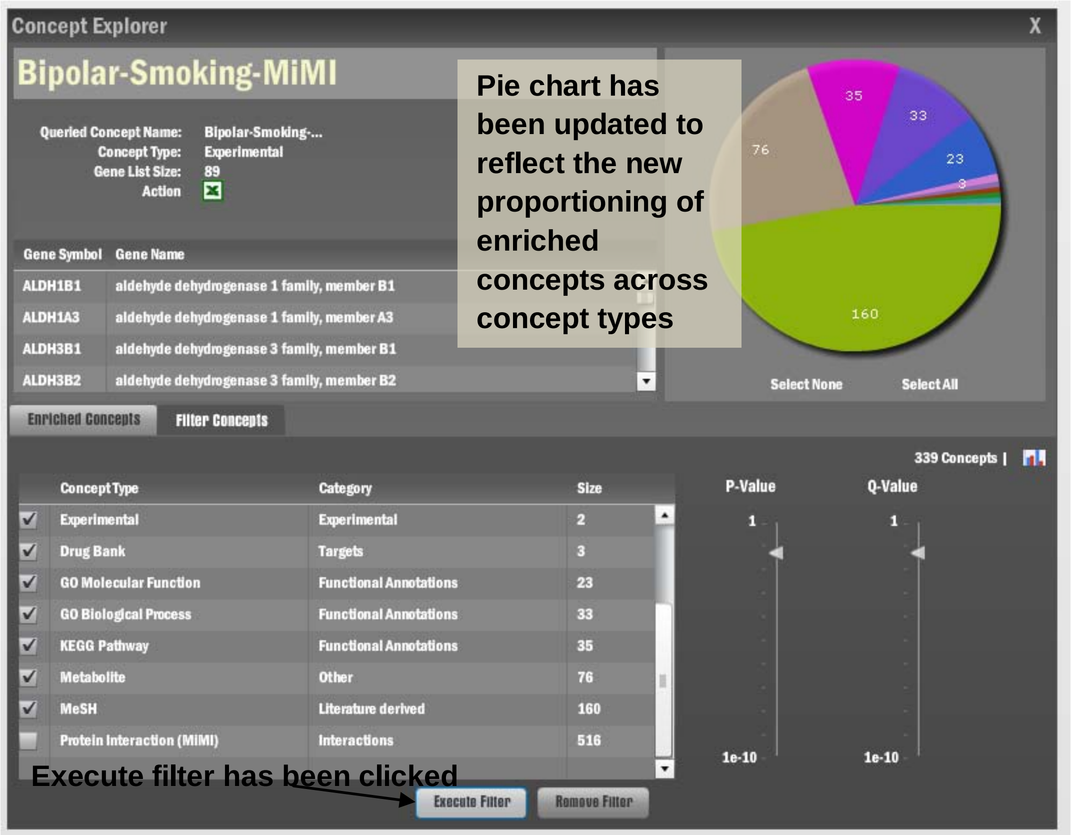
Uncheck concepts that you do not want displayed.

Activate the filters you create by clicking Execute Filter.

Figure 3. The Filter tab view.

2. In the Filter view, scroll down to “Protein Interaction (MiMI)” concept type, and then click to uncheck its box. Click “Execute Filter.” The MiMI rows will no longer appear in the Enriched Concepts view.

3. We could further winnow down the results by moving the slider under Q-value to the e-3 level, but for now we will leave it at 5e-02 (0.05). The pie chart numbers update to reflect the filtering (Fig. 4).



Concept Explorer

Bipolar-Smoking-MiMI

Queried Concept Name: Bipolar-Smoking-...
Concept Type: Experimental
Gene List Size: 89
Action

Gene Symbol	Gene Name
ALDH1B1	aldehyde dehydrogenase 1 family, member B1
ALDH1A3	aldehyde dehydrogenase 1 family, member A3
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
ALDH3B2	aldehyde dehydrogenase 3 family, member B2

Pie chart has been updated to reflect the new proportioning of enriched concepts across concept types

339 Concepts |

	Concept Type	Category	Size
☑	Experimental	Experimental	2
☑	Drug Bank	Targets	3
☑	GO Molecular Function	Functional Annotations	23
☑	GO Biological Process	Functional Annotations	33
☑	KEGG Pathway	Functional Annotations	35
☑	Metabolite	Other	76
☑	MeSH	Literature derived	160
□	Protein Interaction (MiMI)	Interactions	516

Execute filter has been clicked

Figure 4. Filter tab after protein interactions are filtered out

4. Click the Enriched Concepts tab to switch back to this view to select concepts of interest. You will display these and Bipolar-Smoking-MiMI as a concept network.

Select Conceptual Relationships of Interest for a Network View

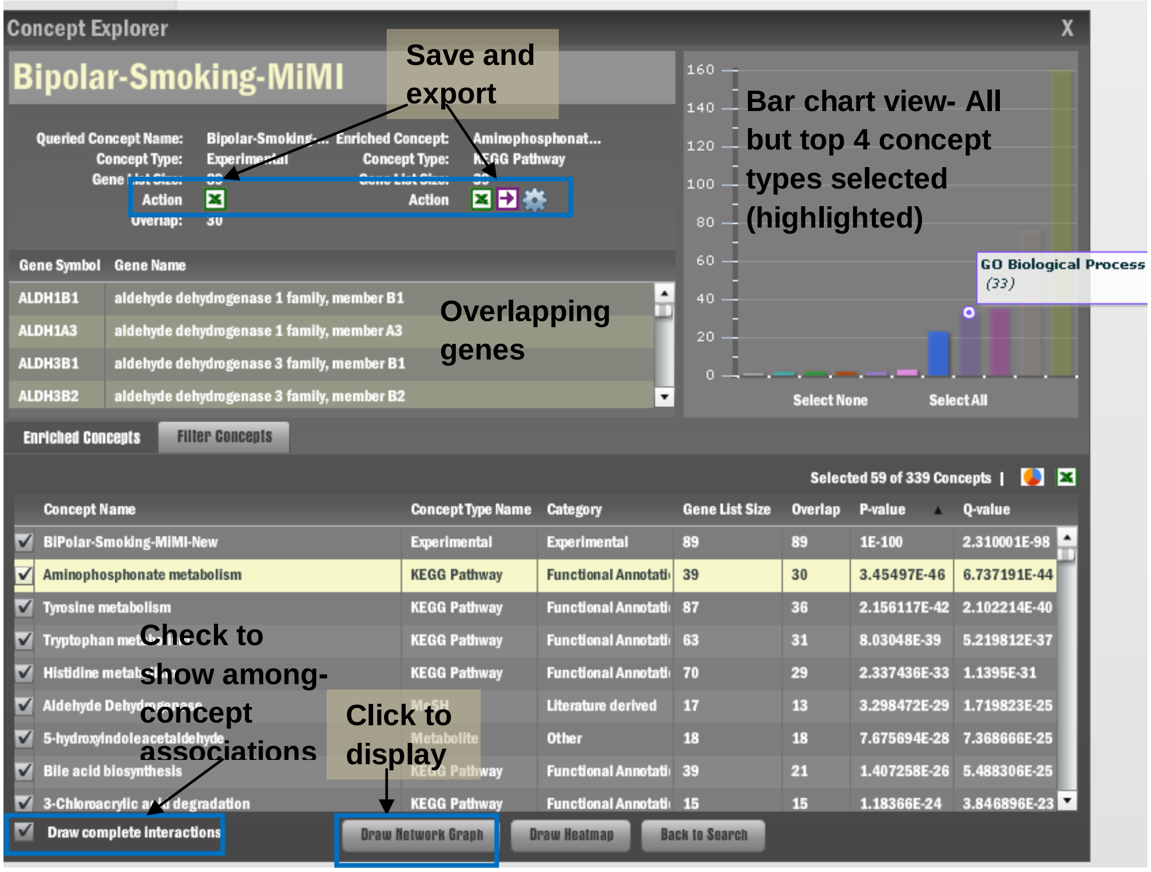
1. In the Enriched Concepts view, click on the bar plot icon next to the excel icon. Mouse over each bar to see the concept it represents. Protein interactions are no longer represented. The largest bars (from right to left) are MeSH, Metabolite, KEGG, and GO Biological Process. You will **not** create a concept association network displaying these categories of concepts because the network would be large and unwieldy. A better choice for seeing these concept relationships is to use the Heatmap View, which is discussed later in this tutorial.
2. In the bar chart, click on the bars for all concept types **except** MeSH, Metabolite, KEGG pathway, and GO Biological Process. This will provide a manageably sized and interesting set of concepts to view as a network graph. ConceptGen adds each successive concept type you click to the selection. Selections are indicated by a checked box in the Enriched Concept table in the bottom half of the screen.
3. You may also want to make sure that you display the concepts with the most significant q-values. To do this, click the q-value column heading of the Enriched Concept table to sort it by q-value. You see that the 3 top significant concepts are not checked. The top significant concept is “aminophosphonate metabolism” from the KEGG pathway, and it is not checked.

Check the boxes for these concepts to add them to the network view that you will display: Aminophosphsonate metabolism, tyrosine metabolism, and tryptophan metabolism.
These concepts are particularly relevant. Aminophosphonates are analogues of amino acids where the carboxyl group (COOH) has been replaced by a phosphonate group (POOH). Research shows that smoking impacts the production of aminophosphonates. You move on to see what other enriched concepts may be meaningful and to which genes they apply.

The neurotransmitters serotonin and dopamine are derived from tyrosine and tryptophan metabolism. Serotonin and dopamine are thought to be involved in both bipolar disorder and smoking and this result is consistent with that hypothesis. (Schildkraut 1974; Manji and Lenox 2000)

4. Check the box next to Draw Complete Interactions at the bottom left of the screen (see Fig. 5). This check assures a view of links among all selected concepts, not just between the concepts and Bipolar-Smoking-MiMI.

5. In the row of buttons under the table, click Draw Network Graph. A network view appears showing interactions among the concepts, including Bipolar-Smoking-MiMI (Fig. 6).



Queried Concept Name: Bipolar-Smoking-... Enriched Concept: Aminophosphonat...
Concept Type: Experimental Concept Type: KEGG Pathway
Gene List Size: 89 Gene List Size: 39
Action ☒ Action ☒
Overlap: 30

Gene Symbol	Gene Name
ALDH1B1	aldehyde dehydrogenase 1 family, member B1
ALDH1A3	aldehyde dehydrogenase 1 family, member A3
ALDH3B1	aldehyde dehydrogenase 3 family, member B1
ALDH3B2	aldehyde dehydrogenase 3 family, member B2

Selected 59 of 339 Concepts |

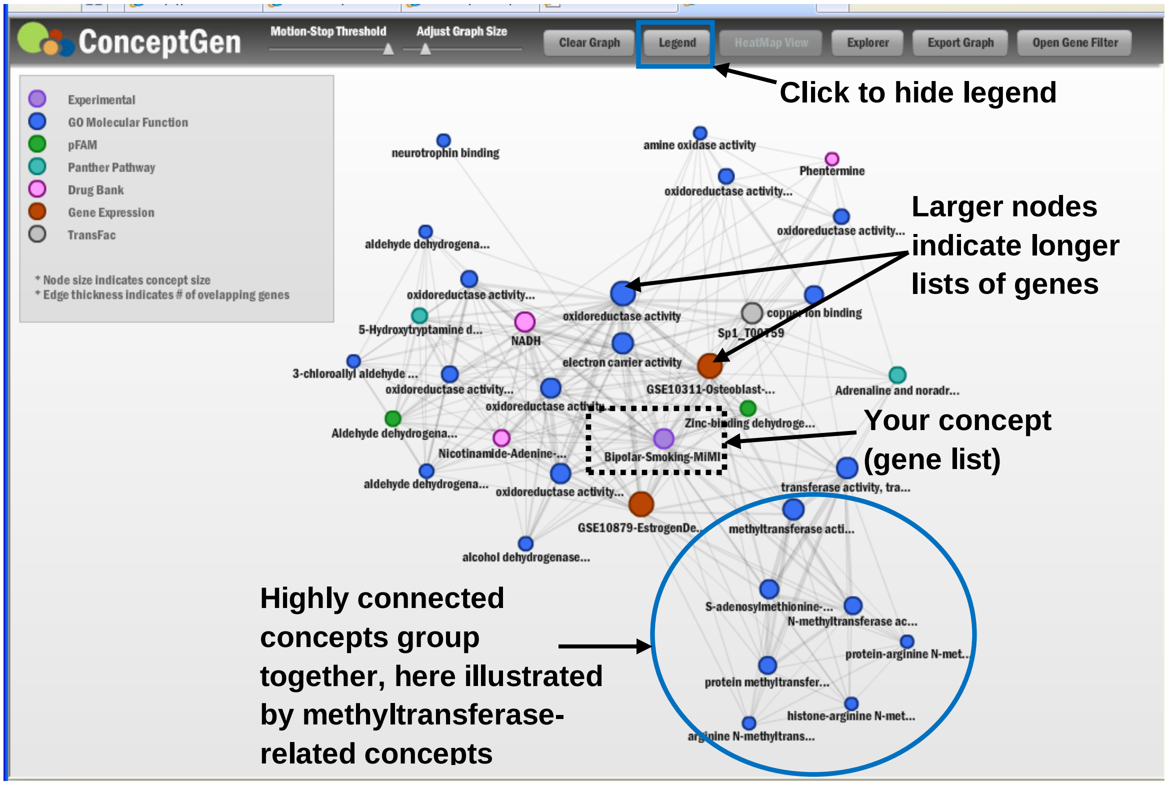
	Concept Name	Concept Type Name	Category	Gene List Size	Overlap	P-value ▲	Q-value
☑	BiPolar-Smoking-MiMI-New	Experimental	Experimental	89	89	1E-100	2.310001E-98
☑	Aminophosphonate metabolism	KEGG Pathway	Functional Annotati	39	30	3.45497E-46	6.737191E-44
☑	Tyrosine metabolism	KEGG Pathway	Functional Annotati	87	36	2.156117E-42	2.102214E-40
☑	Tryptophan metabolism	KEGG Pathway	Functional Annotati	63	31	8.03048E-39	5.219812E-37
☑	Histidine metabolism	KEGG Pathway	Functional Annotati	70	29	2.337436E-33	1.1395E-31
☑	Aldehyde Dehydrogenase	MeSH	Literature derived	17	13	3.298472E-29	1.719823E-25
☑	5-hydroxyindoleacetaldehyde	Metabolite	Other	18	18	7.675694E-28	7.368666E-25
☑	Bile acid biosynthesis	KEGG Pathway	Functional Annotati	39	21	1.407258E-26	5.488306E-25
☑	3-Chloroacrylic acid degradation	KEGG Pathway	Functional Annotati	15	15	1.18366E-24	3.846896E-23

☑ Draw complete interactions

Figure 5. Preparing for a network display of associations among selected concepts.

1. Examine and manipulate the network view.
   Node size represents the number of genes associated with a given concept.
   Node color (see the Legend) represents types of concepts.
   Edges (links) link concepts that contain a significant number of overlapping genes.
   Your concept – Bipolar-Smoking-MiMI- is in the center.

2. Adjust the size of the graph with the Adjust Graph Size slider on the top panel. Move individual nodes by clicking and dragging on them or the entire graph by clicking and dragging anywhere in the background. If desired, hide the legend of concept type colors by clicking "Legend" in the top panel.



Click to hide legend

Larger nodes indicate longer lists of genes

Your concept (gene list)

Highly connected concepts group together, here illustrated by methyltransferase-related concepts

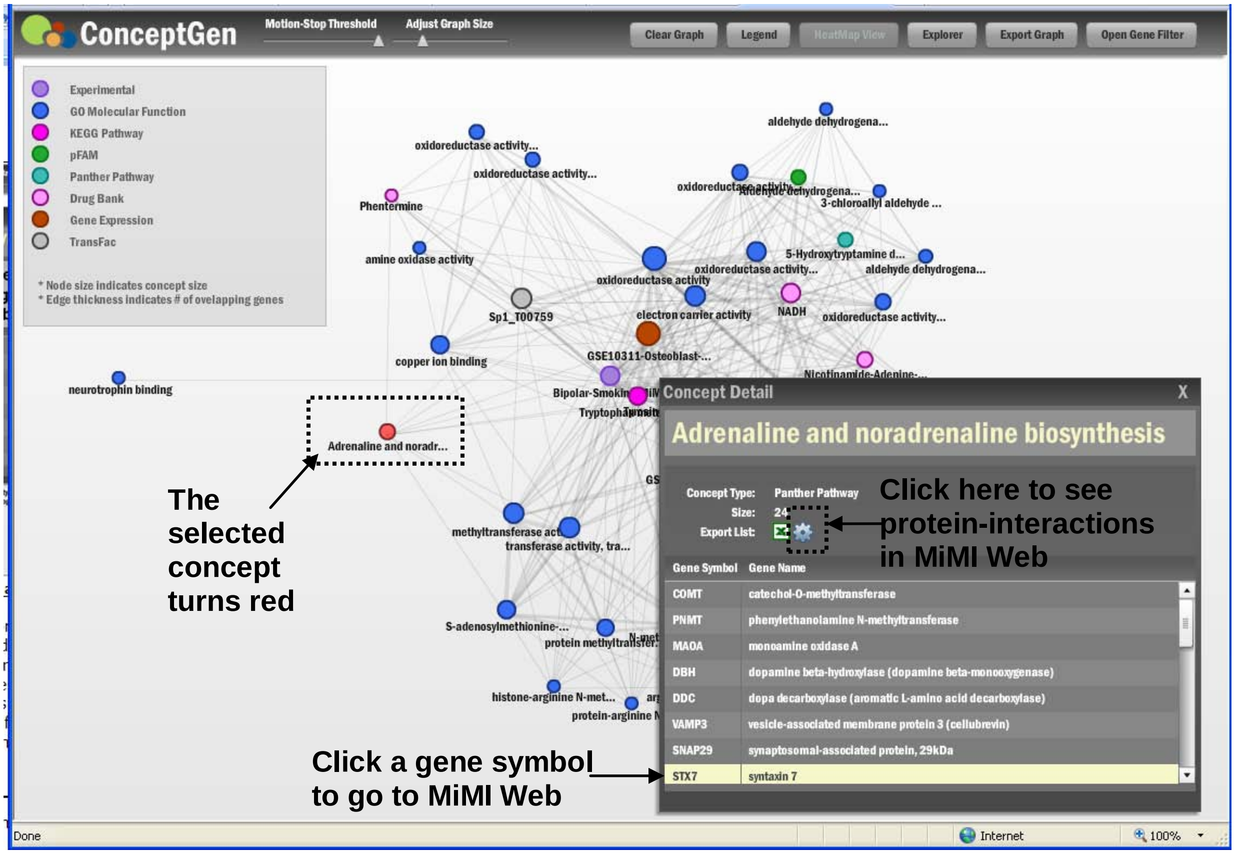
Figure 6: Graph Network View

3. You see that methyltransferase related activities form one cluster (Fig. 6). This is the result of the forced layout algorithm, which groups nodes that have high interconnectivity. Similarly, alcohol dehydrogenase, aldehyde dehydrogenase activity, and amine oxidase activity each form separate clusters, with Oxidoreductase Activity linking to several of these. Thus, the graphic itself helps you find related concepts because they tend to group together.

4. Click on any node (concept) to view the list of genes annotated for it. For any list that seems interesting, you can click on the spreadsheet icon to save it. Or you can click on the gear icon to view the gene list interactions in MiMI NetBrowser.

This list is different from the one you saved from the Explorer view. From Explorer, the genes overlapped between your list and the concept you selected. From the Network view, the genes are all those associated with a selected concept, not just those from your list. Once saved in Excel, you can compare concept lists against each other to find overlapping genes across two or more concepts.

5. Notice the Panther Pathway, “Adrenaline and noradrenaline biosynthesis” in the network. This is known to be a significant pathway affected by cigarette smoking, but is a novel hypothesis for involvement in Bipolar Disorder. Looking at the genes in this concept shows that three of the original five genes (COMT, MAOA, and SLC6A3) are in this pathway. We see STX7 is in this pathway; we do not know much about this gene, but find it interesting since syntaxins are involved in synaptic remodeling. Double click on this gene symbol to link directly to its MiMI gene info page (Fig. 7).
6. We can also click on an edge to view the edge details, for example the edge between “Adrenaline and noradrenaline biosynthesis” and our input gene set.



The selected concept turns red

Concept Detail

Adrenaline and noradrenaline biosynthesis

Concept Type: Panther Pathway
Size: 24
Export List:

Click here to see protein-interactions in MiMI Web

Gene Symbol	Gene Name
COMT	catechol-O-methyltransferase
PNMT	phenylethanolamine N-methyltransferase
MAOA	monoamine oxidase A
DBH	dopamine beta-hydroxylase (dopamine beta-monooxygenase)
DDC	dopa decarboxylase (aromatic L-amino acid decarboxylase)
VAMP3	vesicle-associated membrane protein 3 (cellubrevin)
SNAP29	synaptosomal-associated protein, 29kDa
STX7	syntaxin 7

Click a gene symbol to go to MiMI Web

Figure 7: Gene list associated with Adrenaline and noradrenaline biosynthesis (Panther Pathway).

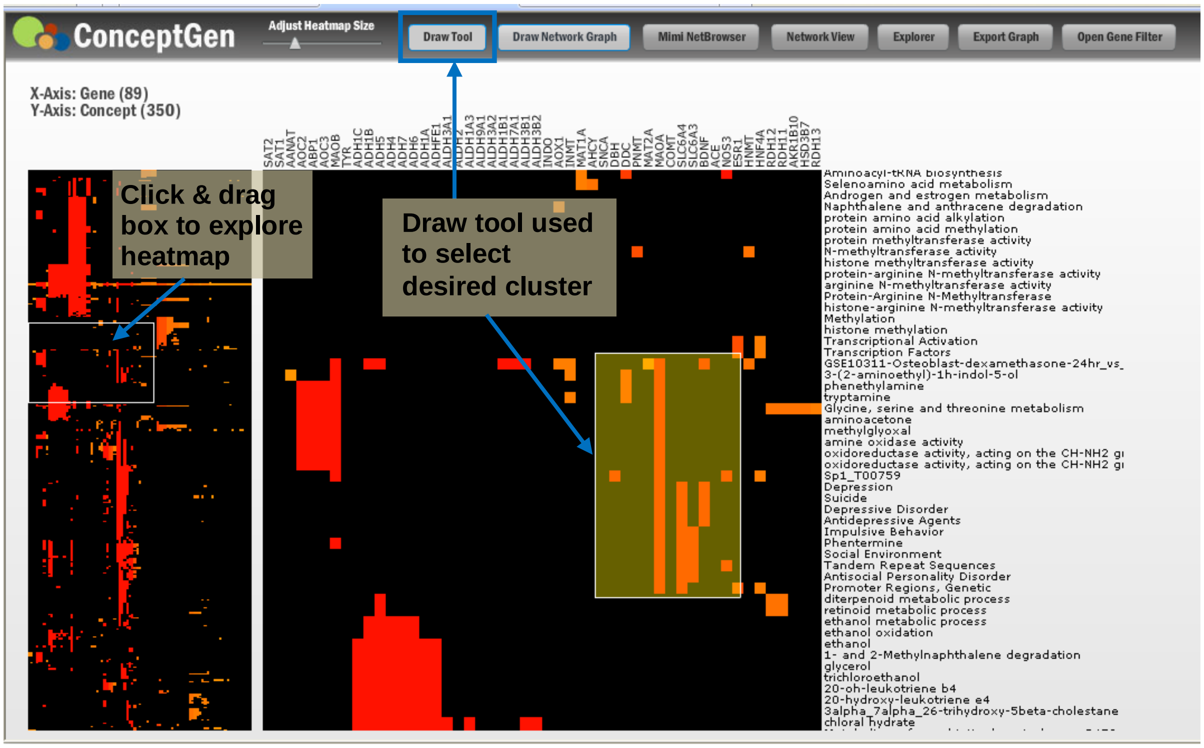
Once you have explored concepts as comprehensively as you like, you could follow-up on one or more of the novel hypotheses generated from ConceptGen analysis, for example testing the extent to which adrenaline levels are influenced by an interaction between Bipolar Disorder and smoking, particularly for individuals in either a manic or depressive episode. In addition, you can compare across saved overlapping lists and winnow down your original list to one that you believe is most relevant to building a story for your hypothesis. You can then use this new list and explore it further for protein interactions.

Heatmap Exploration of Results

The heatmap view can be used to explore high level interaction areas and their meaning for results involving large numbers of enriched concepts. It also allows you to see at a glance which genes are driving the most enrichments.

To view the heatmap with all concept types except MiMI:

1. From the Network View, click on “Explorer” to return to the Concept Explorer results window.
2. Under the chart, click on “Select All”
3. Click on Draw Heatmap on the bottom of the page. This will bring up the heatmap window of results.
4. For the heatmap, columns represent the genes in your gene list (89 columns for the 89 input genes) and rows represent enriched concepts.
5. On the left, you see an overview of the whole heatmap. We see six or so clusters of genes and enriched concepts.
6. We can adjust the heatmap size with the slider on the top of the screen
7. Looking through them, we see one large cluster of concepts involves mainly the aldehyde dehydrogenases and another mainly the alcohol dehydrogenases. We may notice that the input genes COMT, MAOA, SLC6A3, SLC6A4, and BDNF clustered together. We can look through the concepts in which these genes belong, and notice that “Attention deficit disorder with hyperactivity” and “Schizophrenia” are both among these.


ConceptGen
Adjust Heatmap Size
Draw Tool
Draw Network Graph
Mimi NetBrowser
Network View
Explorer
Export Graph
Open Gene Filter
X-Axis: Gene (89)
Y-Axis: Concept (350)
SAT2
SAT1
AANAT
AOC2
ABP1
AOC3
MAOB
TYR
ADH1C
ADH1B
ADH5
ADH4
ADH7
ADH6
ADH1A
ADHFE1
ALDH3A1
ALDH2
ALDH1A3
ALDH9A1
ALDH3A2
ALDH1B1
ALDH7A1
ALDH3B1
ALDH3B2
INDO
AOX1
INMT
MAT1A
AHCY
SNCA
DBH
DDC
PNMT
MAT2A
MAOA
COMT
SLC6A4
SLC6A3
BDNF
ACE
NOS3
ESR1
HNMT
HNF4A
RDH12
RDH11
AKR1B10
HSD3B7
RDH13
Aminoacyl-tRNA biosynthesis
Selenoamino acid metabolism
Androgen and estrogen metabolism
Naphthalene and anthracene degradation
protein amino acid alkylation
protein amino acid methylation
protein methyltransferase activity
N-methyltransferase activity
histone methyltransferase activity
protein-arginine N-methyltransferase activity
arginine N-methyltransferase activity
Protein-Arginine N-Methyltransferase
histone-arginine N-methyltransferase activity
Methylation
histone methylation
Transcriptional Activation
Transcription Factors
GSE10311-Osteoblast-dexamethasone-24hr_vs_
3-(2-aminoethyl)-1h-indol-5-ol
phenethylamine
tryptamine
Glycine, serine and threonine metabolism
aminoacetone
methylglyoxal
amine oxidase activity
oxidoreductase activity, acting on the CH-NH2 gı
oxidoreductase activity, acting on the CH-NH2 gı
Sp1_T00759
Depression
Suicide
Depressive Disorder
Antidepressive Agents
Impulsive Behavior
Phentermine
Social Environment
Tandem Repeat Sequences
Antisocial Personality Disorder
Promoter Regions, Genetic
diterpenoid metabolic process
retinoid metabolic process
ethanol metabolic process
ethanol oxidation
ethanol
1- and 2-Methylnaphthalene degradation
glycerol
trichloroethanol
20-oh-leukotriene b4
20-hydroxy-leukotriene e4
3alpha_7alpha_26-trihydroxy-5beta-cholestane
chloral hydrate

Click & drag box to explore heatmap

Draw tool used to select desired cluster

Figure 8: Heatmap view with section involving the input genes COMT, MAOA, and SLC6A4, and desired concepts selected using the Draw Tool

The genes are clustered by common occurrence in a concept, and the redness of color reflects the number of enriched concepts that include a specific gene.

We will next select a portion of the heatmap to view as a Network Graph.

1. Find the area that includes the gene expression dataset GSE10311, Depression, and antisocial personality disorder.
2. Click on the Draw Tool button
3. Click and drag in the heatmap to select a box covering the entire bright area including these concepts. (Fig. 8)
4. Click on Draw Network Graph (You may have to click Reset Graph first)
5. You can toggle between the Network and Heatmap views the appropriate buttons.
6. Notice that several other terms relating to depression are enriched, as well as the weight-loss drug Phentermine, which could lead to additional hypotheses.

Querying a Pre-built Concept in ConceptGen

1. Return to the main ConceptGen website from the Explorer window by clicking the “Back to Search” button
2. Type in a keyword or partial keyword of your choice, for example, “smok” and click Search Concepts
3. Concepts matching your search appear with the gene list size and the number of enriched concepts using q value < 0.05
4. You can sort the list by Name or another choice by using the drop-down menu
5. To link to the source information for any concept, click on the right arrow in the gray box at the right.
6. To go to the Explorer window to view results, click on a concept name.
7. Continue to explore as previously described by using the Explorer window and filtering options, network graph and heatmap view.

References

Brady, K. T. and R. B. Lydiard (1992). "Bipolar affective disorder and substance abuse." J Clin Psychopharmacol **12**(1 Suppl): 17S-22S.

Carmelli, D., G. E. Swan, et al. (1992). "Genetic influence on smoking--a study of male twins." N Engl J Med **327**(12): 829-33.

Daban, C., E. Vieta, et al. (2005). "Hypothalamic-pituitary-adrenal axis and bipolar disorder." Psychiatr Clin North Am **28**(2): 469-80.

Edvardsen, J., S. Torgersen, et al. (2008). "Heritability of bipolar spectrum disorders. Unity or heterogeneity?" J Affect Disord **106**(3): 229-40.

Galanter, C. A. and E. Leibenluft (2008). "Frontiers between attention deficit hyperactivity disorder and bipolar disorder." Child Adolesc Psychiatr Clin N Am **17**(2): 325-46, viii-ix.

Grant, B. F., F. S. Stinson, et al. (2005). "Prevalence, correlates, and comorbidity of bipolar I disorder and axis I and II disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions." J Clin Psychiatry **66**(10): 1205-15.

Keller, B. J. and R. C. McEachin (2009). "Identifying hypothetical genetic influences on complex disease phenotypes." BMC Bioinformatics **10 Suppl 2**: S13.

Manji, H. K. and R. H. Lenox (2000). "The nature of bipolar disorder." J Clin Psychiatry **61 Supp 13**: 42-57.

McEachin, R. C., B. J. Keller, et al. (2007). "Prioritizing Disease Genes by Analysis of Common Elements (PDG-ACE)." AMIA Annu Symp Proc: 1068.

McGuffin, P., F. Rijsdijk, et al. (2003). "The heritability of bipolar affective disorder and the genetic relationship to unipolar depression." Arch Gen Psychiatry **60**(5): 497-502.

Schildkraut, J. J. (1974). "Biogenic amines and affective disorders." Annu Rev Med **25**(0): 333-48.

Young, J. (2008). "Common comorbidities seen in adolescents with attention-deficit/hyperactivity disorder." Adolesc Med State Art Rev **19**(2): 216-28, vii.

Ziedonis, D., B. Hitsman, et al. (2008). "Tobacco use and cessation in psychiatric disorders: National Institute of Mental Health report." Nicotine Tob Res **10**(12): 1691-715.