



DrugMatrix[®] 8.0

Release Notes

March 2011

NIEHS/NTP

111 TW Alexander Drive Durham, NC 27709 <u>http://ntp.niehs.nih.gov/</u> http://www.niehs.nih.gov/

$DrugMatrix^{\tiny \tiny M}$

Table of Contents

Contents

New Features in DrugMatrix 7.0 Not in Manual	. 4
PLIER Normalization upgrade	. 4
New Features in DrugMatrix 6.0 Not in Manual	. 5
User Interface Improvements	. 5
Sort by Log Ratio	. 5
Data Export	. 5
Hypergeometric Analysis Tool	. 5
Pathway Impact Matrix Tool	. 5
Export Experiment Pathway Scores Tool	. 6
Connectivity Map Algorithm	. 6
Histopathology Domain	. 6
Experiment Names and Study Names	. 6
Staging System Improvements	. 7
Experiment Similarities	. 7
Postprocessing - Pathway Impact Calculations	. 7
Preloader	. 7
Staging Database Schema	. 7
HI Rat Focus Microarray	. /
New Features in Drugiviatrix 5.0 Not in Manual	. 8
	. 8
Support for Affymetrix Rat Genome 230 2.0 Microarray	. 9
User Interface Improvements	. 9
Pattern Tools	10
Changes to Features in Manual	13
Changes to Computational Methods	13
Miscellaneous Changes	13
Changes to the Pathway Impact Matrix Tool	14
Changes to the Pathway Visualization Tool	15
Drug Signature Posterior Probability Score:	16
Fixed Bugs	18
#292: Exporting an edited list does not export the edited name	18
Known Issues	19
#308: Advanced Search list pop-up menus do not sort alphabetically.	19
#344: Can not drag favorite or list to a File Management Tool after cancelling a previous attempt	19
P	

$DrugMatrix^{\scriptscriptstyle \mathrm{M}}$

#413: Compound names with apostrophes display unprintable characters in Favorites and Report Names.	19
#445: Gene Expression P-Values are not calculated for Singlet Uploaded	
Experiments.	19
#549: Expression Experiment Matrix displays multiple rows for some genes	20
#550: Matrix tools cannot operate with lists containing one item	20
#635: Hover over gene name is wrong in Study Report Time Course and Dose	
Response Windows	20
#774: Highlight of current experiment not working in Pathway Visualization tool.	20
#828: Double-clicking any Menu causes the application to hang	21
#898: The Back and Forward menus remain open	21
#1110: Expand All Folders function does not work in List Mode	21
#1017: Colors on Signature Histogram tool occasionally wrong	21
#1142: Incorrect mouse cursor in Experiment Dendrogram tab	21
#1145: Gene list displays multiple rows per gene	21



New Features in DrugMatrix 8.0 Not in Manual

PLIER Normalization upgrade

Affymetrix is moving away from MAS 5.0 normalization algorithm to PLIER. For this reason DrugMatrix has been updated to accept PLIER normalized CHP files and will no longer support MAS5.0 normalized data.

New Features in DrugMatrix 6.0 Not in Manual

User Interface Improvements

Sort by Log Ratio

Gene Expression changes now sort by log ratio by default. Induced genes sort by log ratio descending, so that the largest changes are at the top of the table. Repressed genes sort ascending so that the largest changes are at the top of the table.

Data Export

All tables in DrugMatrix reports now have an Export menu. This includes tables in the report areas as well as tables or graphics in the data visualizations. For example, with this feature, you can analyze a list of experiments with the Pathway Impact Matrix tool, and then export the pivoted data used to draw the heatmap as an Excel file that can be easily read into Spotfire or other analysis tools. This allows you to identify important data using the sophisticated data mining functionality of DrugMatrix, and then export the data to an advanced visualization tool like Spotfire for preparing figures.

Note that the Export feature in the Pathway Impact Matrix Tool enables a whole new set of features that were not implemented in the Spotfire Information Library. In the past, all pathway impacts were pre-calculated for Spotfire IL access using the Hypergeometric statistic. This was limited even though there are 3 different metrics that can be used in the Pathway Impact Matrix Tool. Providing the direct export from the Pathway Impact Matrix Tool has three major benefits:

- You can now use any of the 3 metrics and get your data into Spotfire.
- This new method is much faster than the Spotfire Information Library.
- This removes the need to compute the pathway impact scores in Staging, shortening the staging process by a day.

Hypergeometric Analysis Tool

The Hypergeometric Analysis tool has been integrated into the DrugMatrix user interface. The tool now runs with standard Experiment lists as input, and generates output as an Excel file.

Pathway Impact Matrix Tool

The Pathway Impact Matrix Tool has been extended with a new statistic, "Response," to complement the existing Hypergeometrix and % Changed options.



The pathway response is the sum of the absolute values of the log ratios of the pathway genes, and represents a measure of statistical strength rather than significance.

Export Experiment Pathway Scores Tool

The new Export Experiment Pathway Scores tool implements a method similar to what Iconix analysts commonly called Gene Response Analysis, with some additional optional variations. It supports the analysis of pathway impact by the three metrics available in the Pathway Impact Matrix tool (Response, Hypergeometric, and % Changed) for a list of Experiments on one Pathway. The results are returned as a tab-delimited text file, or piped directly into Spotfire for visualization as a scatter plot. The output data can optionally contain pathway impact scores for all of the DrugMatrix compounds in a selected tissue, for creating context. The output data have the experiments in your input Experiment list flagged, so that it is easy to identify them in a Spotfire visualization. In addition, the scores are ranked so that they can be presented as a smooth curve of scores across the whole dataset.

Connectivity Map Algorithm

The recently published connectivity map algorithm (Lamb et al., Science 313:1929-1935, 2006) was implemented as a replacement for the deprecated trinary algorithm in the pattern search tool. This algorithm is particularly appropriate for use with patterns derived from the literature, or from experiments on platforms dissimilar to the reference DrugMatrix experiments (e.g., QPCR, or cDNA microarrays). Pattern genes are specified as positive or negative, and are sought in the corresponding tail of the log ratio distribution of each experiment. We find that the connectivity map algorithm is less sensitive to gene expression amplitude than the Pearson's algorithm and it therefore achieves what we were trying to achieve with the trinary method, only better. This is now part of the Pattern Search tool.

Histopathology Domain

The Histopathology Lexicon in DrugMatrix has been implemented as a fully functional DrugMatrix Information domain. This means that you can search for a histopathology term in the Simple Search and Advanced Search tools; you can save lists of histopathology terms and you can view a domain report on a histopathology term, including full description of the lexicon, a list of experiments that show that histopathology as well as studies.

Experiment Names and Study Names

In DrugMatrix 6.0, Experiment names and Study names now have the study date appended to them. This allows one to differentiate between experiments performed on different dates, which was not possible in earlier versions.

Staging System Improvements

Experiment Similarities

The staging postprocessing step that calculates experiment similarities has a new input parameter that allows the configuration of how many of the top ranked similarities should be stored for visualization in the DrugMatrix Reports. This defaulted to 25 in older versions of DrugMatrix. The current default is 100, but this can easily be changed in warehouseConfig.xml in the Staging System. In addition, the code has been changed to not exclude the identity comparison, so now the DrugMatrix similarity table will contain the comparator experiment, as well as the similar experiments.

Postprocessing - Pathway Impact Calculations

With the new data export feature, available in all data visualization tools, including the Pathway Impact Matrix tool, the need for postprocessing calculations of pathway impacts has been deprecated. The new mechanism allows exporting data into Spotfire in a quicker and more flexible way. Therefore, we recommend not using the postprocessing calculations for pathway impact. These can be re-enabled if desired, but the calculations take a long time and are a major bottleneck in the turnaround time for data staging.

Preloader

The error reporting during the Preloader processing has been vastly improved to make it easier to determine what data is incorrect in Experiment Description Files.

Staging Database Schema

The staging database has been tightened up regarding data integrity to assure that only correct data are staged and the associates between compounds in staging and DrugMatrix are consistent.

HT Rat Focus Microarray

Support for the HT Rat Focus Microarray was added to the Staging System.

New Features in DrugMatrix 5.0 Not in Manual

Motifs

The Motif domain is included in DrugMatrix Online. Motifs represent a technology that clarifies and brings mechanistic understanding to Drug Signature analysis.

Iconix Drug Signature technology uses a sparse linear programming (SPLP) algorithm to identify and weight the genes to build the best classifier to identify a biological effect. The SPLP uses all genes and finds the most informative ones. It goes through a validation process that confirms that the gene expression data can create a valid classifier.

Motifs are generated in a very similar way. The main difference is that the genes provided to the SPLP are limited to genes in a specific pathway. When a Motif passes validation, it confirms that the genes in that pathway contain information that can classify the biological effect that the signature detects. It proves a relevance of that pathway to the biological mechanism of the biological effect.

The standard use-case is to first analyze a test compound with Drug Signatures. In this example, an Apoptosis signature is hit. To understand the mechanism, we note three motif hits, related to cell cycle, TGFbeta, and oxidative stress, indicating that apoptosis is occurring by these mechanisms.

Motifs also have the potential to identify pathway states. For example, the state of the various transcripts in the cell cycle pathway in the state of apoptosis vs. the state of cell proliferation. This is a future development for the Motif technology.

Motifs help explain the biological mechanisms underlying pathologies or other biological effects diagnosed by Drug Signatures.

The Motif domain contains the following informational tabs:

- Motif-to-Motif Similarity contains a correlation coefficient between the target motif and all other motifs in the data warehouse.
- Motif Genes contains the mean log ratio values of all genes in the specified motif across all of the experiments in the training set "+1" class.
- Motif Experiments contains SP score of the specified motif against all experiments in the data warehouse.
- Motif Details contains both the truth table and canonical pathway.
- Motif Class description

Several tools have been added to support the Motif domain.

- Search functionality (both simple and advanced).
- CanonicalPathwayVisualization tool that is run by clicking on the canonical pathway image in the Motif detail tab.
- Motif Translation tools that convert from pathway lists to motifs, and motif lists pathways.

Support for Affymetrix Rat Genome 230 2.0 Microarray

The Affymetrix GeneChip type "Rat Genome 230 2.0" (RG230-2) for has been added to the supported chip types. This GeneChip is comprised of a more than 31,000 probe sets which includes all of the GeneChip Rat Expression Set 230A (RAE230A). Users can run simple search, advanced search, GO data query, signature analysis as well as other tools in DrugMatrix[™] when analyzing data generated on the RG230-2 chip.

The data upload tool also includes support for the RG230-2 microarray.

User Interface Improvements

- The DrugMatrix[™] window is now expandable. The window can be maximized by using the maximize button in the top right hand corner, or it can be expanded manually by dragging the edges of the window.
- DrugMatrix[™] now supports multiple dataview windows open at the same time. The dataview tab has been removed and dataviews now appear in their own separate windows. Dataview windows are internal windows to the DrugMatrix[™] application. The can be moved and minimized to collapse into their title bar, similar to the minimize function on the old Mac OS (pre-X).
- 3. Data table and List Display menus now contain an EXPORT item. This allows the user to export any data currently displayed in the data table of view. All data from the table is exported (whether it is selected or not). In addition many tables also contain hidden data which is not directly displayed as a table column. This hidden data is also exported.
- 4. The Expression Experiment domain now contains a histopathology data table. This table contains the histopathology terms and scores associated with the animals in the respective expression experiment. This includes values for incidence and severity and the RIDIT score described in the DrugMatrix Calculations white paper found in the online help. Like other data tables, information in this table is queried using the advanced query tool.
- 5. The Signature domain has been extended to include additional data. Two new data tables were added, as well as several new signature attribute fields in the detail tab. The first new table displays all genes that make up the

signature, while the second table displays the expression experiments that were used to generate the signature label set.

6. The Display for Printing feature has been removed from DrugMatrix.

Pattern Tools

A set of tools for Pattern Searching has been added to the toolbox. Patterns are lists of Genes and amplitudes (at the expression, or cross-platform mapping level) that can be used to search across the Expression Experiment domain for experiments that have similar expressions. Patterns can be created in one of two ways (explained in more detail below), either by using the Pattern Creation tool, or by performing a series of steps using the Export and Import Pattern tools. Once created, a pattern can be used in conjunction with the Pattern Search tool to search across the expression experiment domain.

 Pattern Creator – This tool takes as input an experiment list and a gene list. It calculates either a mean or median log ratio value across those experiments for each gene specified in the gene list, and uses those as the amplitude values of the pattern. The calculation can be based on p-value-filtered, or unfiltered data.

PATTERN CREATION
EXPERIMENT LIST
N =
GENE LIST
N =
✓ Filter if p-value > 0.02
MEAN C MEDIAN
SELECT WHERE TO SAVE THE NEW LIST
Workspace
📰 RU1allsign
KU1experiment
🗶 exps
NAME
DESCRIPTION
X CANCEL D CREATE

 Export Pattern – This tool takes as input either a gene list or a preexisting pattern and converts that into a tab delimited file that can be easily edited using Excel. Users can create pattern lists by hand (without using the Pattern Creator tool) by

EXPORT PATTERN	— ×
GENE LIST OR PATTERN	
N =	
J	EXPORT

exporting a list of genes with this tool and then hand editing the amplitude values. Instructions on how to do this are included with the exported data and can be read directly from the excel file.

 Import Pattern – This tool takes a tab delimited file (of the format generated by the Export Pattern tool) and uses it to create a pattern list in the user's workspace. This is the second step in creating a pattern by hand. Simply save the file created with the Export Pattern tool and then use this tool to upload it.

IMPORT PATTERN
PATTERN FILE
BROWSE
SELECT WHERE TO SAVE THE NEW LIST
🔄 Workspace
👪 RU1allsign
RU1experiment
🗶 exps
X CANCEL MPORT

4. Pattern Search – This tool takes as input a single pattern list, and uses that pattern list to search across the entire DrugMatrix Expression Experiment domain. The results returned by this search is a list of experiments and a corresponding correlation coefficient score which describes the correlation between the expressed genes in the experiment and the weighted genes of the pattern.

PATTERN SEARCH		— ×
PATTERN LIST		
N =		
Return top	50	
C Return where >=		
Pearsons C		
O Irinary		
	X CANCEL	> SEARCH
1		

The search may be performed in Trinary mode. This sets both weighted values and expression data into a trinary state represented by 1, 0, or -1.

Changes to Features in Manual

Changes to Computational Methods

Posterior Probability (PPS) scores, for Drug Signature analysis, have been improved to consider the number of experimental and control arrays used when generating the expression experiment. When more arrays (i.e. more replicates, either biological or technical) are used are used to generate the log ratio expression profile for a single expression experiment, the measurement of average log ratio for each gene gets more precise. This higher level of precision for the genes used by the Drug Signatures, allows the numerical methods to make a better estimate of the how "real" a Drug Signature score is. A full description of this method can be found in the DrugMatrix Calculations White Paper in the online help page.

Miscellaneous Changes

- The Experiment domain has been modified to support the Motif domain. A single data table was added displaying all of the Motifs that hit this signature with a significant SP/PPS score. This table also contains links to the pathway and signature domains where appropriate.
- The Compound domain has been modified to support the Motif domain. A single data table was added displaying all of the Motifs that hit this signature with a significant SP/PPS score. This table also contains links to the pathway and signature domains where appropriate.
- 3. The Load Data tool has been modified to compute motif SP/PPS scores and generate a new data table based on those values.
- 4. The Dendrogram tool now creates dendrograms dynamically using the experiments found in the similarity tab. In previous versions the entire database wide dendrogram tree was generated during warehouse post processing and specified experiments were force fitted.
- 5. In the Signature truth table tab, total size now refers to the number of positive experiments plus the number of Negative experiments in the training set. Previously it referred to the total number of experiments, which included the zero class. This was changed because the zero class, which is excluded from the signature training, is not really relevant to the signature training and should not be represented in the total training size.

Changes to the Pathway Impact Matrix Tool

The functionality of the Pathway Impact Matrix tool has changed to allow the user to rank order pathways by impact scores. The user can choose to display the top N (number) of impacted pathways by maximum impact or by average impact. If the user chooses not to rank order pathways (by deselecting the "Display top N most impacted pathways" checkbox, the tool will function as in earlier releases.

PATHWAY IMPACT MATRIX
EXPRESSION EXPERIMENT LIST
N =
PATHWAY LIST
N =
DEFINE FILTER
P VALUE < 0.02
REGULATION
G BOTH
DISPLAY TOP 45 MOST IMPACTED PATHWAYS
C BY MAXIMUM % 1.3
BY AVERAGE % 1.3
DEFINE DISPLAY
SCORE TYPE % CHANGED
COLOR RANGE GREEN-WHITE-RED
MAX VALUE (%) 30

Changes to the Pathway Visualization Tool

The functionality of the Pathway Visualization tool has changed. The user must now enter both the value filter and p-value significance values. P-value filter determines the cutoff point for values to be displayed in the visualization. Values that do not make the cutoff will be displayed as white spots on the pathway image. The p-value significance filter determines if a value is significant or not. Significant values will be drawn as colored diamonds on the pathway image.

PATHWAY VISUALIZATION	
EXPRESSION EXPERIMENT LIST	p-
N =	
PATHWAY FAVORITE	
DEFINE FILTER 1.0	
P VALUE FILTER 0.05 P VALUE SIGNIFICANCE	
DEFINE PRESENTATION	
DISPLAY TYPE LOG RATIO	
COLOR RANGE GREEN-WHITE-RED 🔻	
MAX LOG RATIO 0.5 VALUE	
CANCEL DISPLAY	

Drug Signature Posterior Probability Score:

DrugMatrix contains a new statistical metric, a posterior probability value to supplement the scalar product used in signature scoring. The following slides describe this new metric:





ICON

Fixed Bugs

#292: Exporting an edited list does not export the edited name.

Description: After changing a list name with the Edit Item tool, exporting the list will not export it with the modified name information. Fixed in version 6.0.9.

Known Issues

#308: Advanced Search list pop-up menus do not sort alphabetically.

Description: In the advanced search, any pop-up menus displaying lists of list name (following the "in the named list…" clause) are presented in Workspace order, not alphabetically.

Work Around: Lists can be reorganized using the Organize Workspace feature which will change the order in the pop-up menu. Lists are displayed in the pop-up menu in the same order they appear in the workspace in Workspace Mode.

#344: Can not drag favorite or list to a File Management Tool after cancelling a previous attempt.

Description: After having executed a File Management Tool (Union, Subtract, or Intersect) with one type of list or favorite, and then cancelling the operation, the list tool remembers which type of list (gene, compound, etc.) was dragged and will not accept other types of lists for drag-and-drop onto the tool icon.

Work Around: Instead of dragging the list directly onto the icon, first click on the tool to open the tool dialog, and then you can drag any type of list into the dialog input area.

#413: Compound names with apostrophes display unprintable characters in Favorites and Report Names.

Description: Experiments, Studies or Compounds with an apostrophe in their name may not display correctly in the Favorites Menu or as to top-line name displayed in the report.

Work Around: None.

NOTE: The Favorites menu and list item names have been fixed, but the unprintable character still displays in the main display of the name in the domain report.

#445: Gene Expression P-Values are not calculated for Singlet Uploaded Experiments.

Description: When uploading expression data, if using only one microarray in either the treatment group or control group, the software cannot calculate a p-value for that group of microarrays.

Work Around: Use a minimum of two microarrays in each of the treatment and control groups. Run technical replicates if biological replicates are not available.



#549: Expression Experiment Matrix displays multiple rows for some genes.

Description: The underlying data model for expression data in DrugMatrix invokes a concept of "Expresson" – a set of probes from several microarray platforms that measure the same transcript. There can be several probes (only one per chip type) per Expresson, and then several Expressons per Gene. This design allows transparent cross-platform comparisons in DrugMatrix. The caveat is that Gene lists are stored at the Gene level and therefore include all Expresson transcriptional data for that gene, even if you are only interested in one of the possible Expressons. This issue becomes apparent when generating gene lists from the Expression -> Gene Translation tool or the Significant Gene Finder. These tools may identify specific Expressons into a Gene List. When that Gene List is used in a the Expression Experiment Matrix visualization tool, all of the Expressons belonging to that Gene are visualized, even though some of the system design. Please be aware of this when interpreting the Expression Experiment Matrix output.

Work Around: None

#550: Matrix tools cannot operate with lists containing one item.

Description: The Expression Experiment, Bio-Activity and Clinical Pathology Matrix tools require input lists with more than one item. Lists with only one item cause a failure in the clustering algorithm and a server error is returned. Please use lists with more than one item.

Work Around: None

#635: Hover over gene name is wrong in Study Report Time Course and Dose Response Windows.

Description: Hovers over truncated gene names should contain the non-truncated name, but in these windows, the hover only displays "true".

Work Around: None

#774: Highlight of current experiment not working in Pathway Visualization tool.

Description: The Pathway Visualization tool is supposed to highlight the current pathway, in the list at the left, with a light blue background. Currently this highlight appears and disappears as the mouse is moved over the list, and it is not working correctly.

Work Around: Passing the mouse over the list may make the highlight appear.



#828: Double-clicking any Menu causes the application to hang.

Description: Double-clicking, or clicking any menu multiple times in rapid succession, will cause the menu to lock in the open position and the user interface to freeze.

Work Around: The user interface will sometimes free up if you click on a different window and then return to the DrugMatrix window. This does not cause the menu to close. Ultimately, you will need to press the F5 button to refresh the window and log back into DrugMatrix. It is best to avoid double-clicking the menu.

#898: The Back and Forward menus remain open.

Description: After using the Back or Forward menus, the menus remain in the open state.

Work Around: Click anywhere besides the menu in the DrugMatrix window and the menu will close.

#1110: Expand All Folders function does not work in List Mode

Description: The "Expand all Folders" menu function does work in Workspace Mode, but not in List Mode

Work Around: None

#1017: Colors on Signature Histogram tool occasionally wrong.

Description: When the signatures hit scores are close to the boundary between numerical ranges in the color bar legend, occasionally a borderline score will appear to be the wrong color. This is due to a rounding error where the color is determined based on the non-rounded value, but the rounded value is displayed. For example, if the signature score is 0.65, it would be colored according to the 0.33 to 0.67 range, but would be displayed as 0.7.

Work Around: None

#1142: Incorrect mouse cursor in Experiment Dendrogram tab.

Description: The mouse cursor changes to a pointing finger anywhere in the dendrogram tab, whereas it should only change to this cursor when over a hyperlink. Work Around: None

#1145: Gene list displays multiple rows per gene

Description: When a gene list is displayed in the List Display area, there may be several rows with the same gene name, representing a replication of the gene name in several rows. This is due to there being multiple NCBI Unigene or LocusLink identifiers associated with that gene. For each identifier, a new row is created.

Work Around: This is operating as designed and the user just needs to be aware that this can happen.