

TRANSMART / METACORE INTEGRATION

VERSION 1.0-MC-130116

OPEN SOURCE EDITION

USER DOCUMENTATION

LIFE SCIENCES
INTELLECTUAL PROPERTY & SCIENCE
THOMSON REUTERS

JANUARY 17TH, 2013



THOMSON REUTERS

CONTENTS

CHAPTER 1 RELEASE NOTES	3
ABOUT TRANSMART WITH METACORE INTEGRATION	3
DOWNLOADING AND INSTALLING SOFTWARE	3
VERSION CHANGE LOG	3
CHAPTER 2 USING THE INTEGRATION	4
PATHWAY MAP ENRICHMENT IN DATASET EXPLORER	4
WHAT IS A PATHWAY MAP	4
WHAT IS ENRICHMENT ANALYSIS	4
HOW TO USE THE PLUGIN	5

CHAPTER 1 RELEASE NOTES

ABOUT TRANSMART WITH METACORE INTEGRATION

tranSMART with MetaCore Integration is a special build of tranSMART enabling a new functionality - enrichment analysis of gene lists in pathway maps of MetaCore, a data mining and pathway analysis software developed by Thomson Reuters.

It is based on GPL version of tranSMART and currently is available in both source and binary form at our GitHub repository. It is planned to incorporate functionality into 'master' branch of tranSMART in the nearest future.

DOWNLOADING AND INSTALLING SOFTWARE

A binary distribution can be downloaded from our GitHub repository at <https://github.com/eugene-rakhmatulin/transmartMetacoreBinaries>

It contains ready-to-deploy WAR file and additional files. Readme.md file contains installation instructions.

In addition, full source code is available at <https://github.com/eugene-rakhmatulin>

Please read Readme.md file in the root of transmartMetacoreBinaries repository before installing or compiling tranSMART.

This is a temporary location, and both source code and binaries will be migrated to the main tranSMART repository in the future.

VERSION CHANGE LOG

Version 1.0-MC-130116

Initial release. Based on pre-Postgres 1.0GA branch of tranSMART.

CHAPTER 2 USING THE INTEGRATION

PATHWAY MAP ENRICHMENT IN DATASET EXPLORER

For any questions related to Enrichment Analysis plugin functionality, for the tips on how to use the plugin, or for any scientific questions you may have regarding the analysis please feel free to contact:

Technical Questions:

Eugene Rakhmatulin, IT Director

eugene.rakhmatulin@thomsonreuters.com

Scientific Questions:

Tatiana Khasanova, Manager Translational Medicine

tatiana.khasanova@thomsonreuters.com

What is a pathway map

Pathway maps of MetaCore represent a tree-structured portfolio of over 1,300 canonical, disease-specific, and toxicity-specific pathway maps depicting signaling and metabolic cascades. Signaling cascades usually start on the cellular membrane with a ligand-receptor interaction and proceed via adaptors, second messengers, and kinase cascades to the transcription factors that regulate the expression of genes mediating cellular response. Each pathway map can include 3-5 signaling cascades involved in similar outcomes (we keep the number of objects on each pathway map to no more than 50).

Pathway maps are reconstructed based on findings about a particular signaling pathway that are well-accepted in scientific communities. Pathway maps are created based on the experimental evidence behind not only the specific protein-protein interactions but the whole cascades using, for example, reporter gene assays, etc. Often pathways are reconstructed based on the review articles that summarize experiments done in the original studies. Every pathway map is supported by 100-300 articles including the evidence for protein-protein interactions.

What is enrichment analysis

Briefly, enrichment analysis allows users to evaluate what biological processes are “enriched” with genes found in the gene list of interest (which can be generated in many different ways). The enrichment calculation uses the Fisher’s exact test or hypergeometric distribution to calculate the probability that the degree of overlap between the list of genes found in the list of interest and the genes represented in the functional ontology category, for example, in a pathway map, can happen by chance given an identical number of genes selected at random from the universe of genes annotated within the ontology. The p-value generated is used to rank order the categories within an ontology by their significance to the list of genes, thereby identifying maps or biological processes likely to be important for the genes of interest.

How to use the plugin

Enrichment analysis in MetaCore pathway maps is currently available in **two separate locations in Dataset Explorer**: as a separate tab in Dataset Explorer main menu, and as a part of Marker Selection workflow.

MetaCore Enrichment Analysis tab in the main menu allows you to run enrichment by pathway maps for a single cohort. To use the functionality, first select a desired cohort, then click the tab (Figure 1):

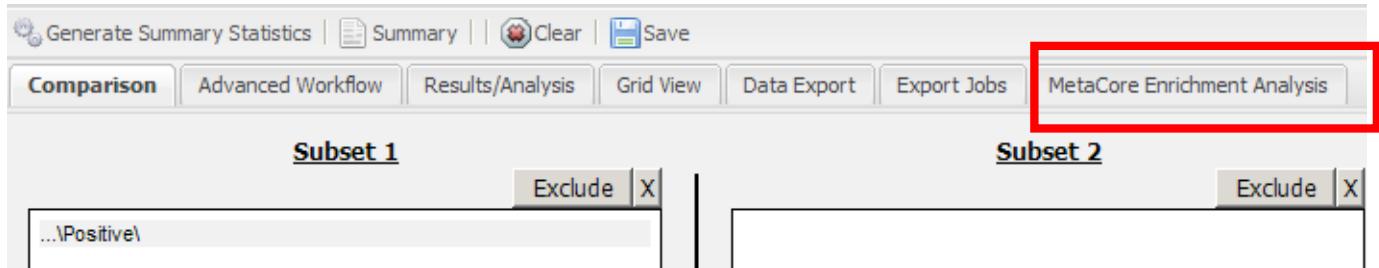


Figure 1: A fragment of transSMART interface showing MetaCore Enrichment Analysis tab.

A new window will open that allows you to drag gene expression data (found under Biomarker Data in the Dataset Explorer tree) in Data Preparation window (Figure 2). In the current version of transSMART, a "High Dimensional Data" tab has to be clicked to call up the data selection pop-up menu. All parameters can be left as default, no selection needs to be made in the pop-up menu for the enrichment analysis to run.

A screenshot of the Data Preparation window. At the top, it says 'Cohorts (Early Alpha version - only the first cohort will be used)' and 'Subset 1: (\Public Studies\Breast_Cancer_Minn_GSE5327\)'. Below this is a 'Variable Selection' section with a 'Data preparation' sub-section. It says 'Select a High Dimensional Data node from the Data Set Explorer Tree and drag it into the box.' A red box highlights the text input field where '...\\Breast\\' is typed. Below this is a 'High Dimensional Data' button. Further down, there are sections for 'Subset1' (Platform: mRNA, GPL Platform: Affymetrix Human Genome U133A Array, Sample: Breast, Tissue: ALL, Timepoint: ALL), 'Pathway:' (Marker Type: Gene Expression), and 'Specify a Z-Score threshold (optional):' (zscore |zscore| >= 2.4). A red box highlights the text input field for the z-score threshold. At the bottom are 'METACORE SETTINGS' and 'Run' buttons.

Figure 2: A fragment of transSMART interface showing settings of Data Preparation window.

Simply click button Run to perform the enrichment by all the genes for the cohort. You may wish to specify an optional z-score threshold to truncate the gene list.

Results of enrichment analysis will be displayed as a histogram as shown on Figure 3. The pathway maps will be prioritized based on the negative log(pValue) showing most significant pathway maps on the top. A bar corresponding to negative log(pValue), actual number of negative log(pValue), pValue, and a number of objects will be provided for each map. The number of objects is represented with two values: 1st number corresponds to the number of network objects on the map that are present in the gene list of interest; 2nd number corresponds to the total number of network objects on each map.

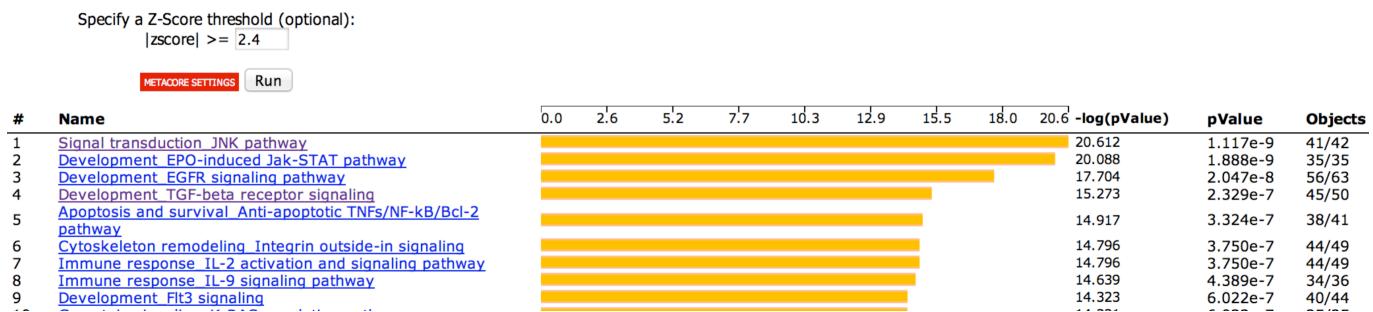


Figure 3: A histogram of enrichment analysis results in tranSMART interface.

Please note that this functionality currently works only for the first cohort. If the second subset is specified it will be ignored.

MetaCore Settings button defines which version of MetaCore should be used for the enrichment analysis (Figure 4). The settings are stored on per-user basis and persist between sessions.

METACORE SETTINGS

When you click on the button, you will be presented with the following options:

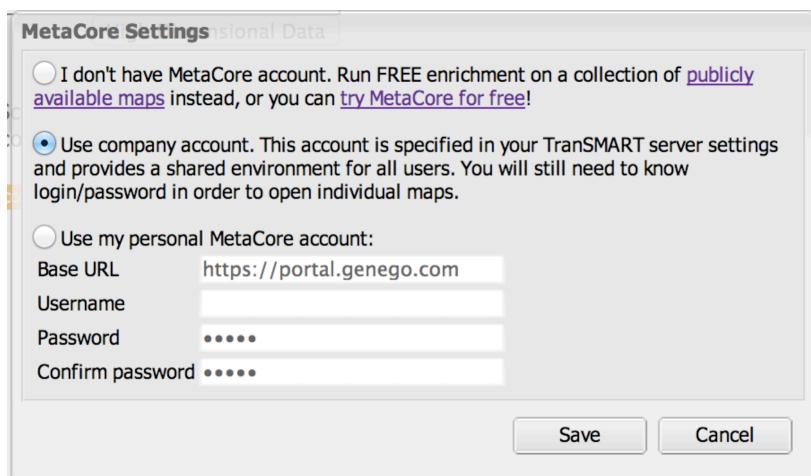


Figure 4: MetaCore settings dialog window in tranSMART interface.

The default is “Use company account” if one is specified in the server settings, or Free enrichment otherwise. Free enrichment uses a set of 260 (as of January 2013) pathway maps available at <http://pathwaymaps.com/maps/>

Metacore settings button color represents the current selection:

- **Red** if Free enrichment is selected
- **Orange** if company MetaCore account is being used
- **Green** if personal MetaCore account is being used

Please note that enrichment analysis will produce different results depending on the enrichment option selected, and also depending on the subset of pathway maps available on your personal MetaCore account if one is used. For instance, a user with a regular MetaCore license will get different results from a user with a MetaCore license enhanced with MetaCore Specialty Modules or MetaMiner access as additional pathway maps are available via many of MetaCore add-ons.

Another place where MetaCore enrichment is available is in **Marker Selection workflow** results. In this case, Marker Selection workflow will first calculate the set of top differentially expressed genes for your cohorts. Enrichment Analysis will then be performed on this set of differentially expressed genes to help you identify the important biological processes that distinguish your cohorts on the molecular level.

Table of top Markers METACORE SETTINGS **Run MetaCore Enrichment Analysis**

Gene Symbol	Probe ID	Raw p-value	Bonferroni	Holm	Hochberg	SidakSS	SidakSD	BH	BY	t	t (permutation)	Raw P (permutation)	Adjusted P (permut
ADCK2	222117_s_at	0.15908	1	1	0.99999	1	1	0.99999	1	1.408157	1.408157	0.2904040	1
ADO	212502_at	0.17689	1	1	0.99999	1	1	0.99999	1	1.350402	1.350402	0.2815657	1
APOA2	219465_at	0.15730	1	1	0.99999	1	1	0.99999	1	1.414216	1.414216	0.2777778	1
APOE	203381_s_at	0.13460	1	1	0.99999	1	1	0.99999	1	1.496190	1.496190	0.2803030	1
CCDC70	221167_s_at	0.16023	1	1	0.99999	1	1	0.99999	1	1.404307	1.404307	0.2878788	1
CHST8	221065_s_at	0.14513	1	1	0.99999	1	1	0.99999	1	1.456948	1.456948	0.2626263	1
CRYGA	207587_at	0.16975	1	1	0.99999	1	1	0.99999	1	1.373018	1.373018	0.2702020	1
CSPNBP2	221260_s_at	0.16281	1	1	0.99999	1	1	0.99999	1	-1.305600	-1.305600	0.2601010	1

Figure 5: Enrichment Analysis button in Marker Selection workflow of transSMART.

When you click on Run button the results will be displayed on top of the table with markers similar to their display in Figure 3.

When exploring the enrichment results, you can click on the pathway map name, and it will take you to the corresponding map page in MetaCore if you use have MetaCore license (Figure 6), or to the corresponding page on our public website pathwaymaps.com (Figure 7):

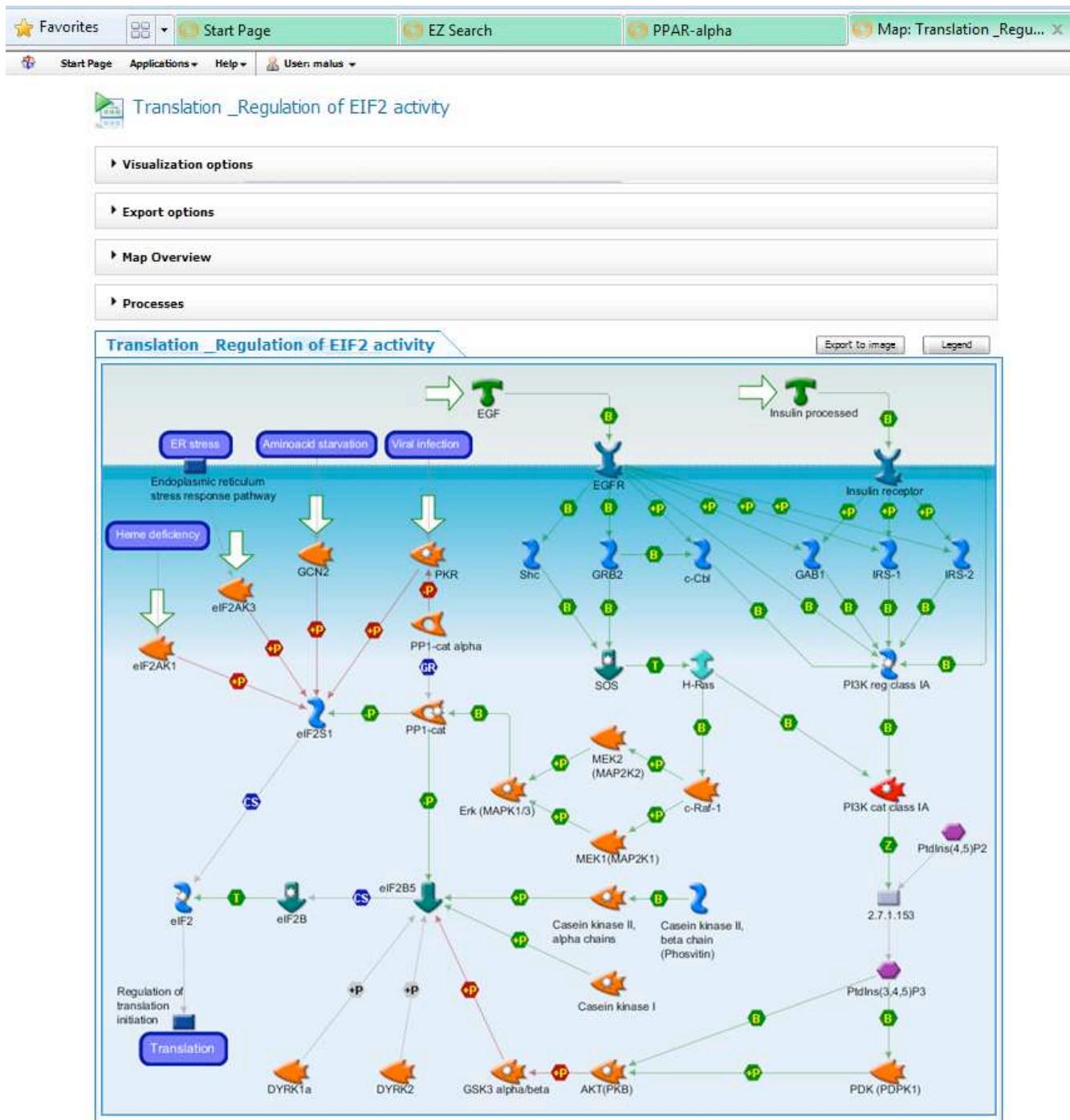


Figure 6: A pathway map displayed in MetaCore interface (available for MetaCore license holders).



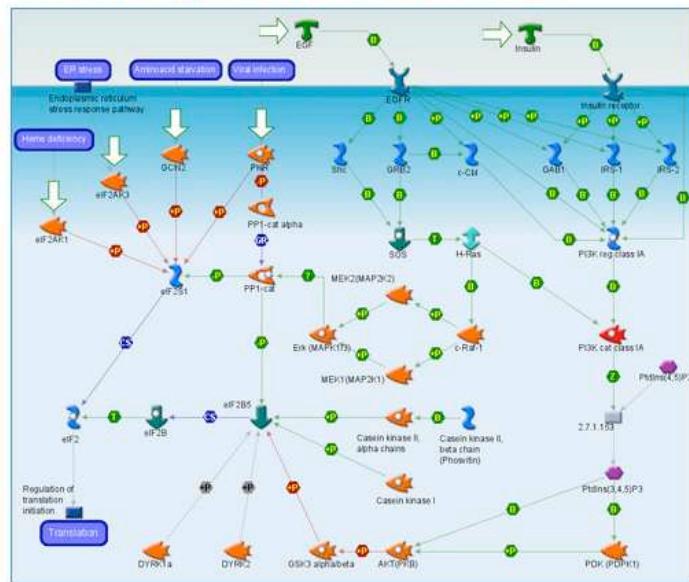
PATHWAY MAP DETAILS

SEARCH MAPS

BROWSE MAPS

FREE TRIAL

TRANSLATION _REGULATION ACTIVITY OF EIF2



[view in full size](#) | [open in MetaCore](#)

OBJECT LIST (LINKS OPEN IN METACORE):

c-Raf-1, Insulin, PP1-cat alpha, EGF, AKT(PKB), PP1-cat, MEK2(MAP2K2), Insulin receptor, c-Cbl, PI3K cat class IA, GAB1, PtdIns(3,4,5)P3, MEK1(MAP2K1), eIF2AK3, GRB2, EGFR, DYRK2, PKR, H-Ras, DYRK1a, IRS-1, PI3K reg class IA, eIF2S1, Casein kinase II, alpha chains, 2.7.1.153, Erk (MAPK1/3), eIF2AK1, SOS, IRS-2, eIF2, GSK3 alpha/beta, PDK (PDPK1), eIF2B5, eIF2B, GCN2, Shc, Casein kinase I, Casein kinase II, beta chain (Phosvitin), PtdIns(4,5)P2

Figure 7: A pathway map displayed in pathwaymaps.com website.

