REUTERS/Jim Young

# METACORE –
# MAKING THE MOST OUT OF YOUR OMICS DATA

**Matthew Wampole, PhD.** – Solution Scientist

matthew.wampole@thomsonreuters.com

THOMSON REUTERS

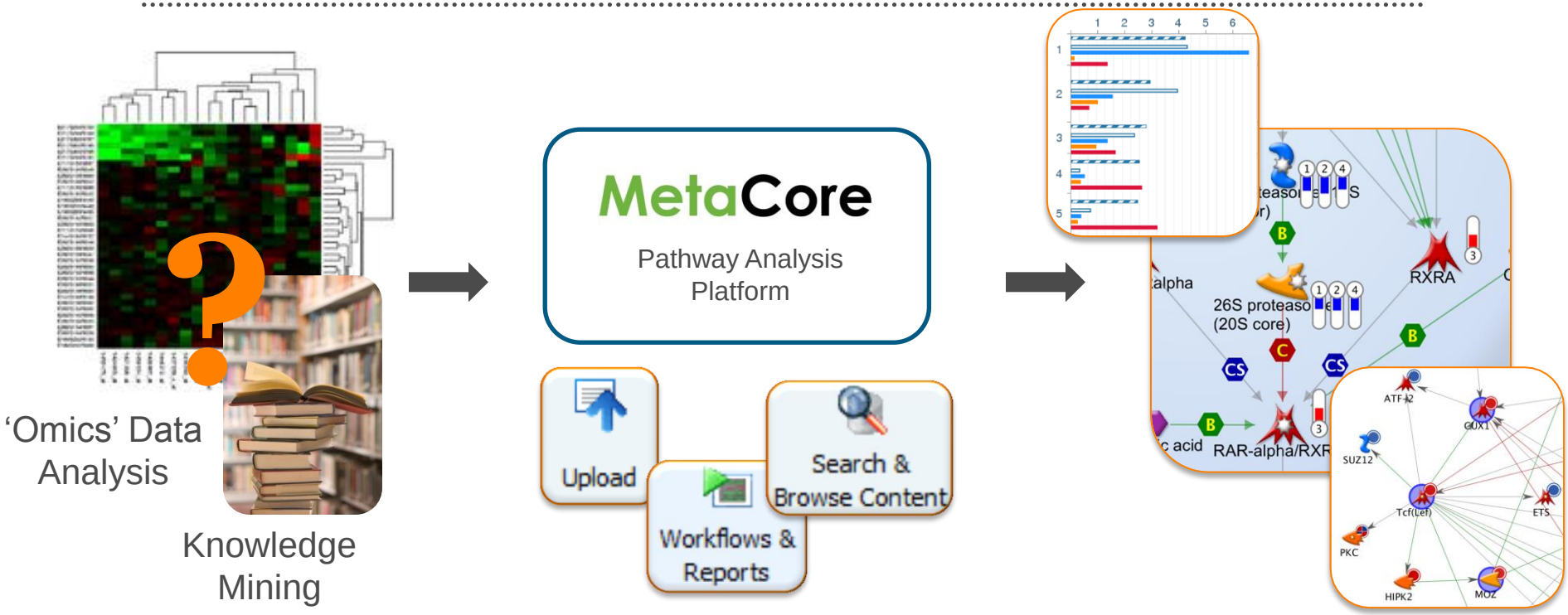
# AGENDA

- **Morning Session (Introductory Topics) – 9:30 AM to 12:30 PM**
  - 9:30-10:00 – General Overview: Thomson Reuters Systems Biology Solutions
  - 10:00-10:30 – Knowledge Mining: Explore the database and exporting
  - 10:30-11:00 – Uploading, filtering and setting a background
  - 11:00-11:45 – Running Functional Enrichments and exploring Pathway Maps
  - 11:45-12:30 – Running Workflows

- **Lunch Break – 12:30 PM to 1:30 PM**

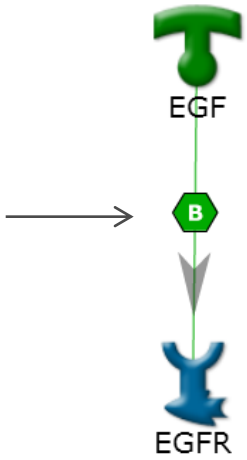
- **Afternoon Session (Advanced Topics) – 1:30 PM to 4:30 PM**
  - **Topics of Interest** (will discuss these or other topics of interest to attendees)
    - 30-45 min – Key Pathway Advisor – Hypothesizing key hubs using the causal reasoning algorithm
    - 45-60 min – Building Networks with MetaCore
    - 30-45 min – Using the Microarray repository for gene comparisons against public data
    - 30-45 min – Multi-omics analysis with miRNA & mRNA data
    - 45-60 min – Multi-omics analysis with Metabolomics & mRNA data

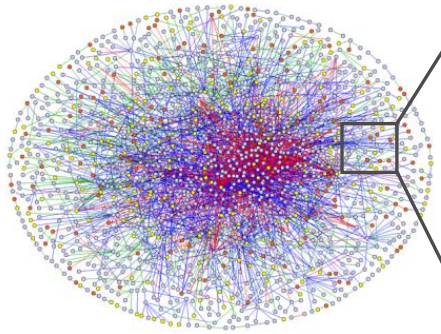
# METACORE:
# YOUR GPS IN PATHWAY ANALYSIS

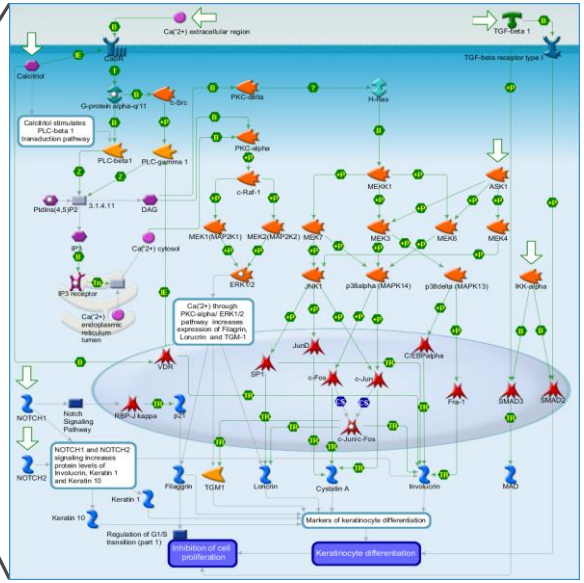



- **Gain molecular understanding of disease**
- **Analyze and understand experimental findings (Omics data) in the context of validated biological pathways.**
- **Generate and confirm hypotheses for novel biomarkers, targets, mechanisms of action**

# FROM PEER REVIEWED ARTICLES TO SIGNALING PATHWAYS





**PUBLICATIONS**
(195 for EGF-EGFR interaction)

**MOLECULAR INTERACTION**

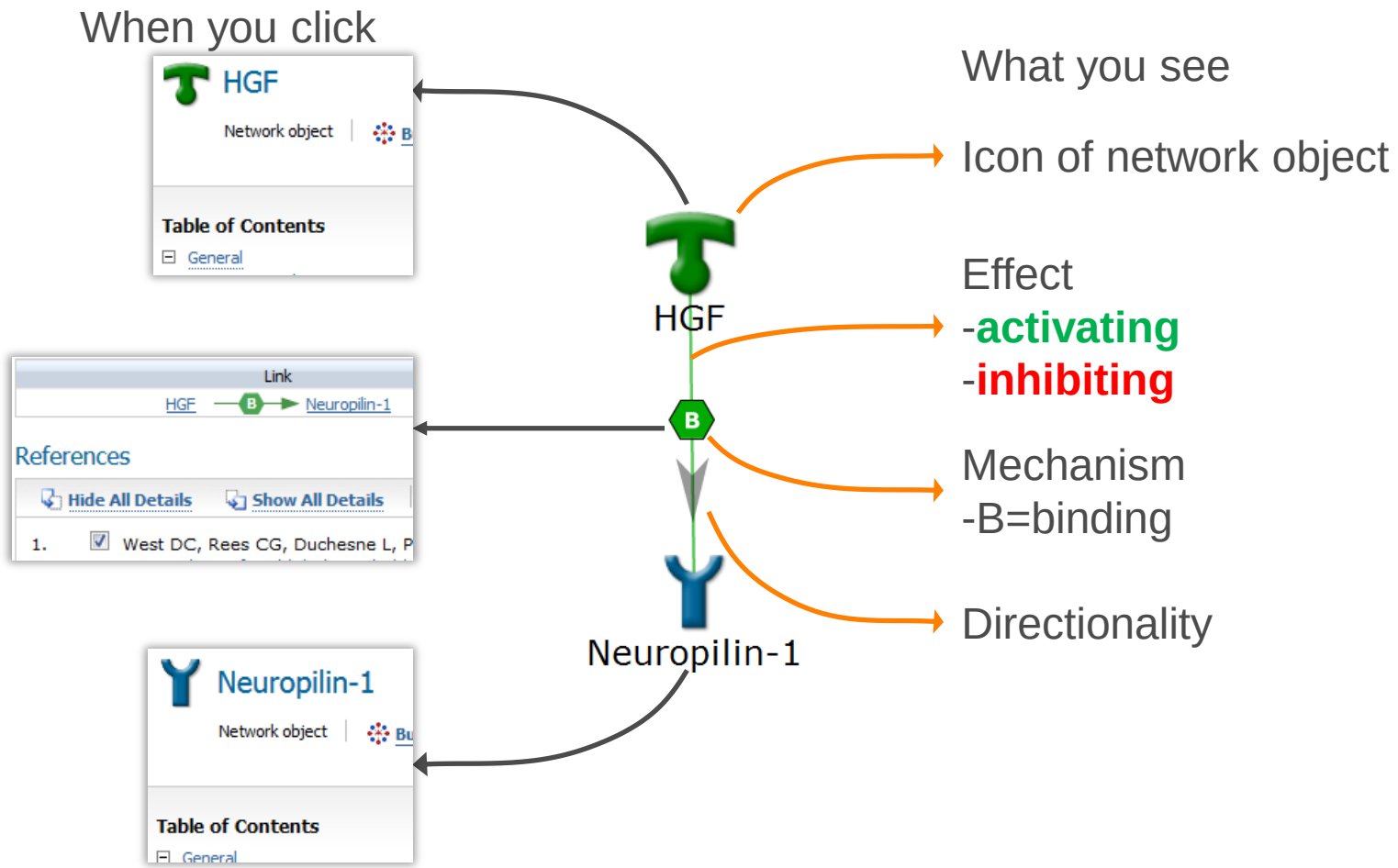
**GLOBAL NETWORK:**
*~ 1,600,000 molecular interactions*

***~ 3,000 canonical and disease signaling pathways***

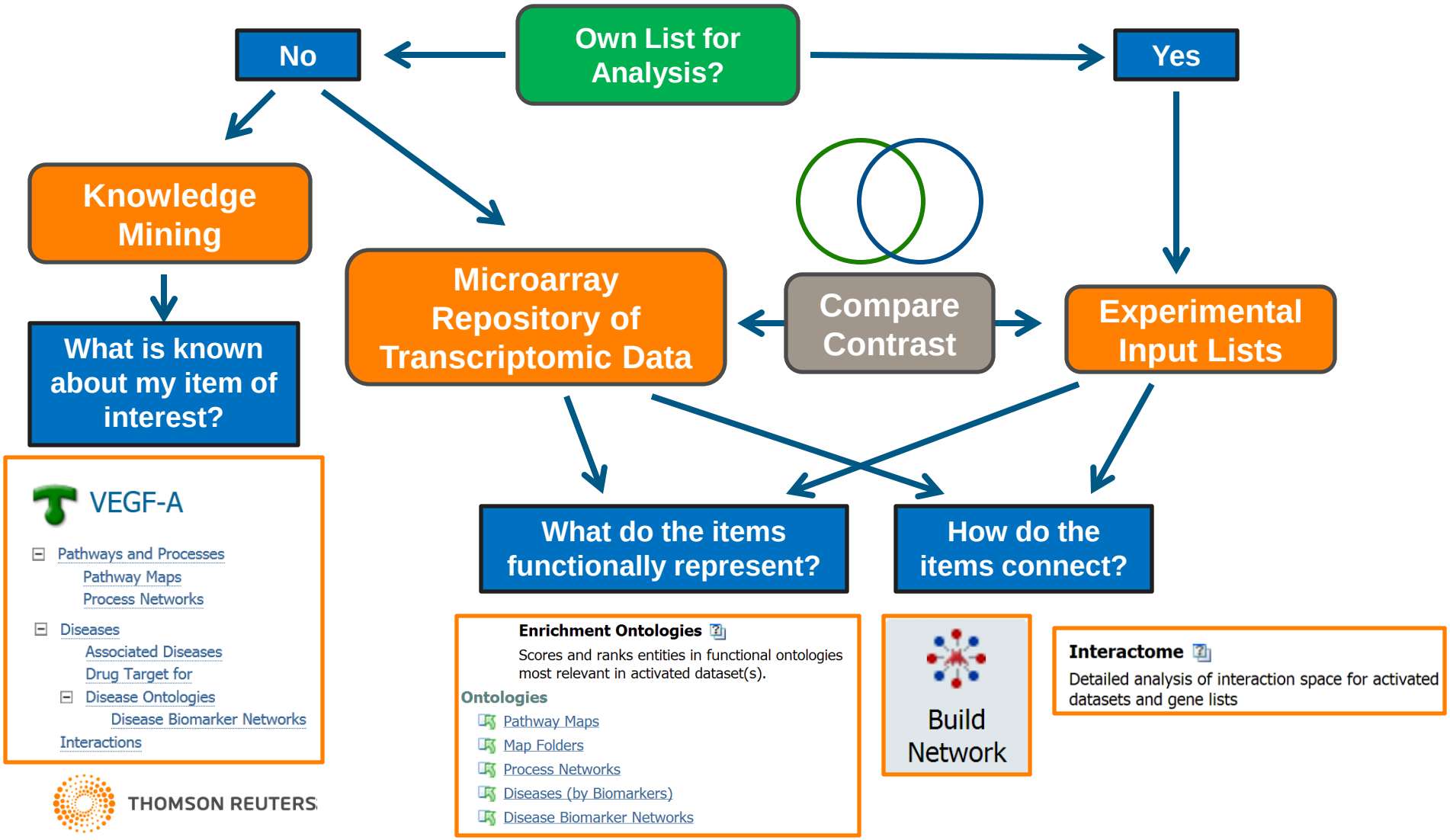
- Manual annotation from publications
- Team of PhDs, MDs
- More than 10 years

THOMSON REUTERS

# WHAT'S IN AN INTERACTION?

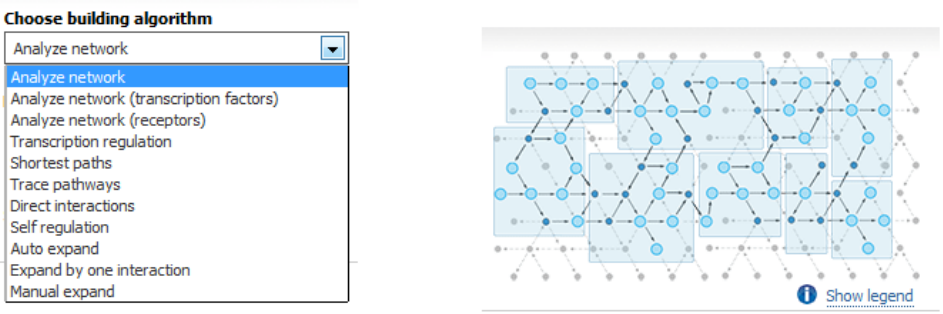


# WAYS TO APPROACH METACORE

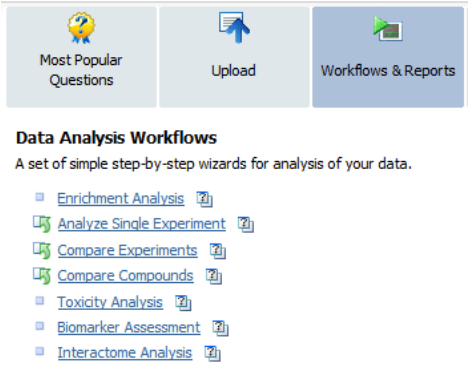


# FLEXIBILITY IN DATA ANALYSIS TOOLS

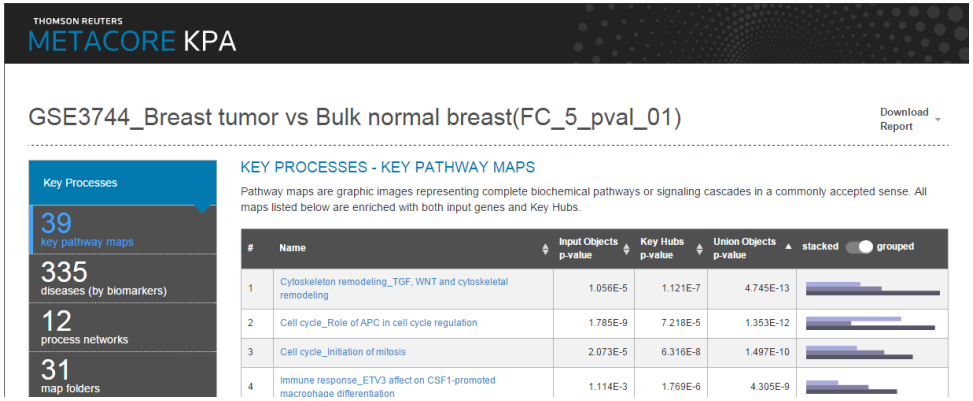
## 11 Different Network Building Algorithms, all with written and visual descriptions



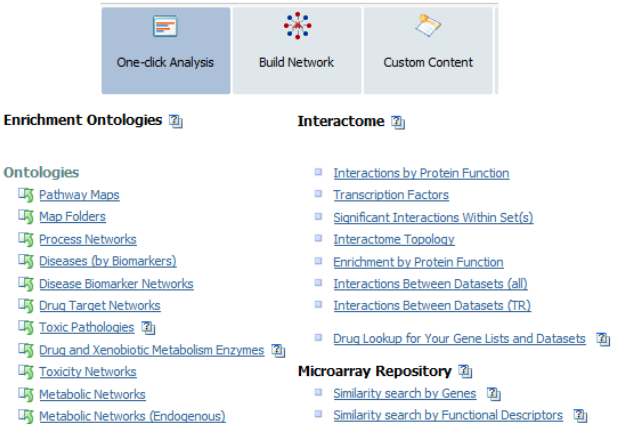

## Multiple automated Workflows to save, share and export




## Causal reasoning algorithm to find key hubs



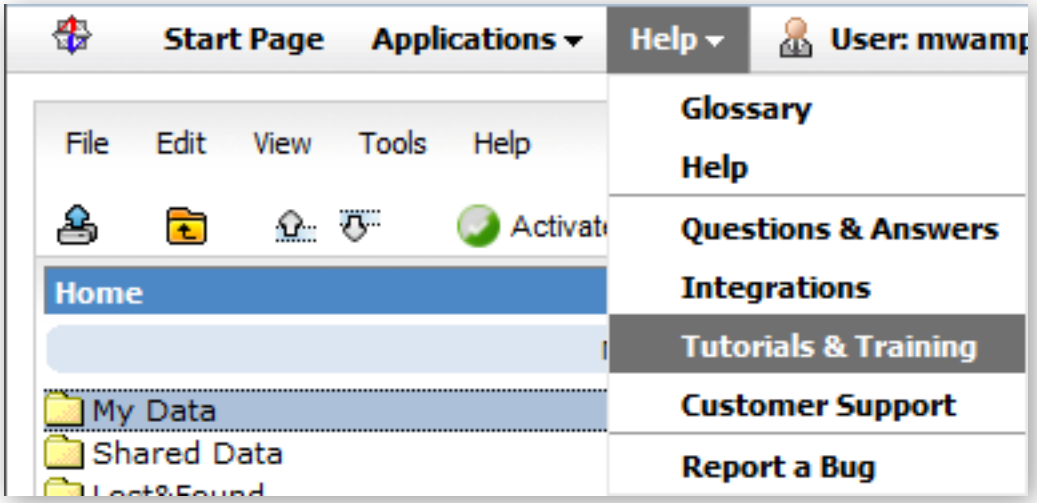

## But also One-Click analysis for instant answers

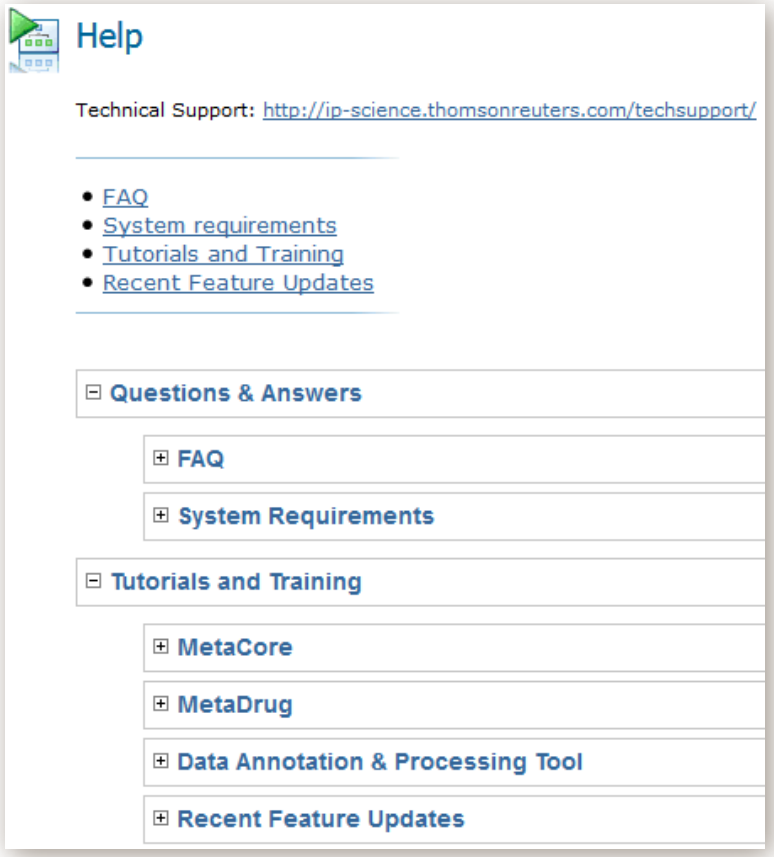


# GETTING SUPPORT

Technical Support & questions:
sbsupport@thomsonreuters.com

Monthly Training webinar recordings available at:
www.lsresearch.thomsonreuters.com
(in the Knowledge section)



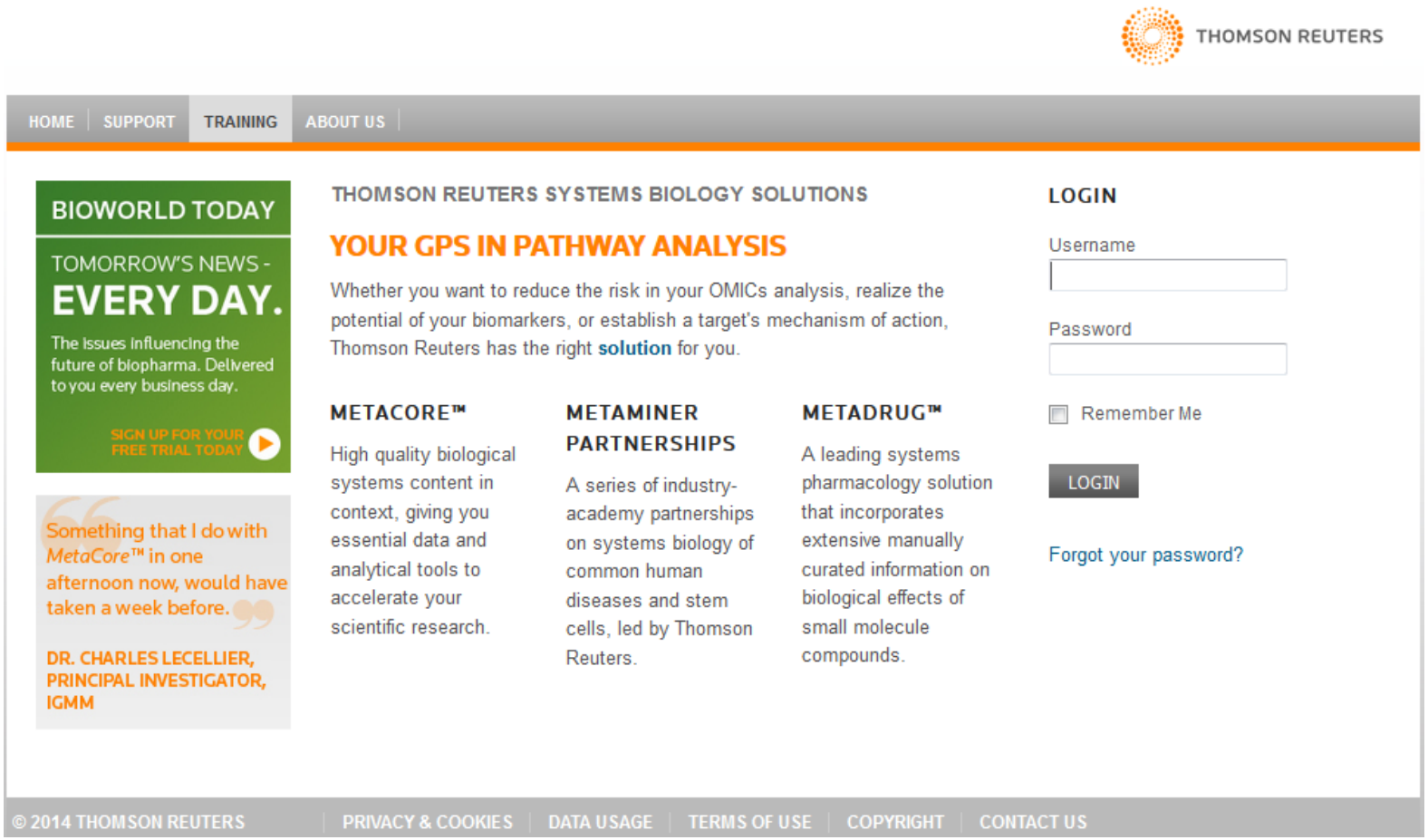
# PUBLISHING WITH METACORE

1. Make sure to cite “Thomson Reuters MetaCore”
2. Include the version and build found at the bottom left corner of MetaCore


   *important because the database is updated every quarter
3. Ensure reproducibility by detailing the steps you take:
   1. What background did you use?
   2. Did you apply some filter?
   3. Did you change the order of how you sorted the results?
4. You can export images and graphs to use in your papers and presentation!

# LOGIN TO METACORE

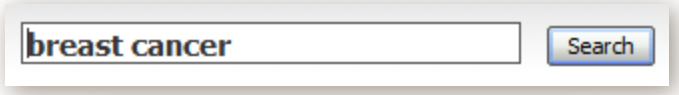
- Go to portal.genego.com

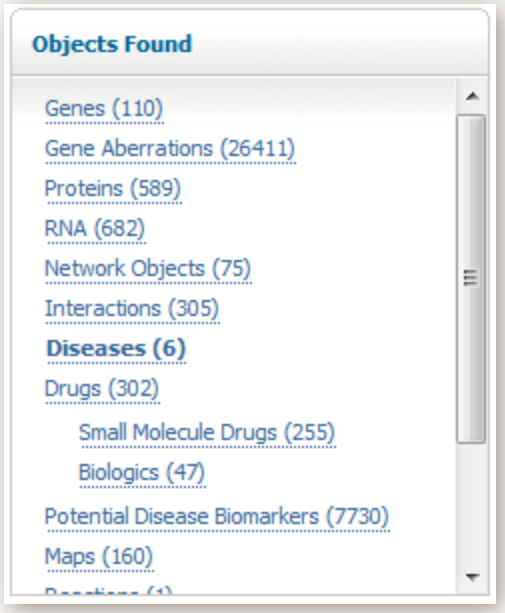


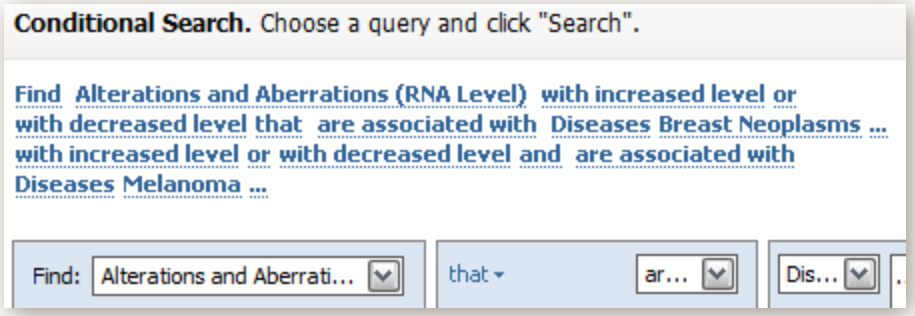
# KNOWLEDGE MINING MANUALLY CURATED CONTENT FROM PEER REVIEWED SOURCES

EZ search

Advanced Search

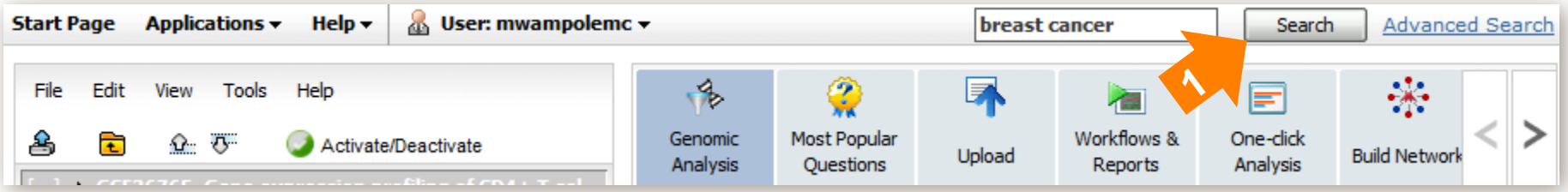


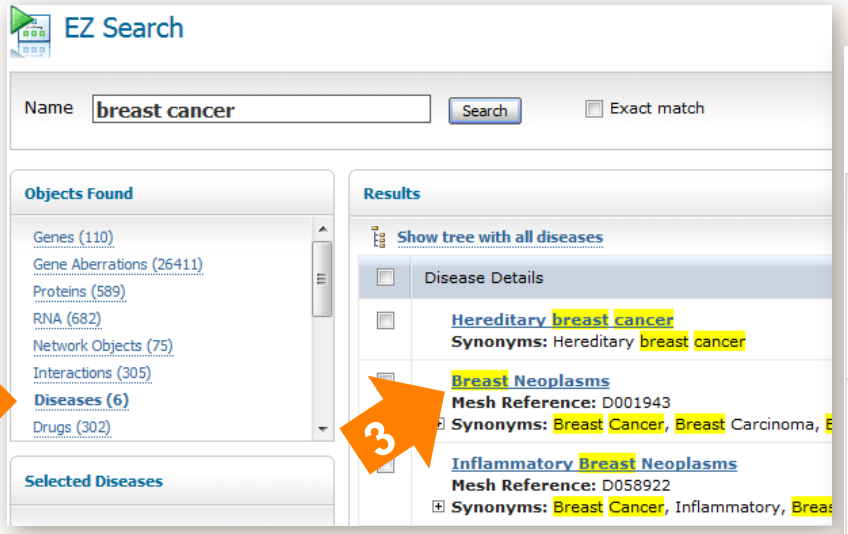



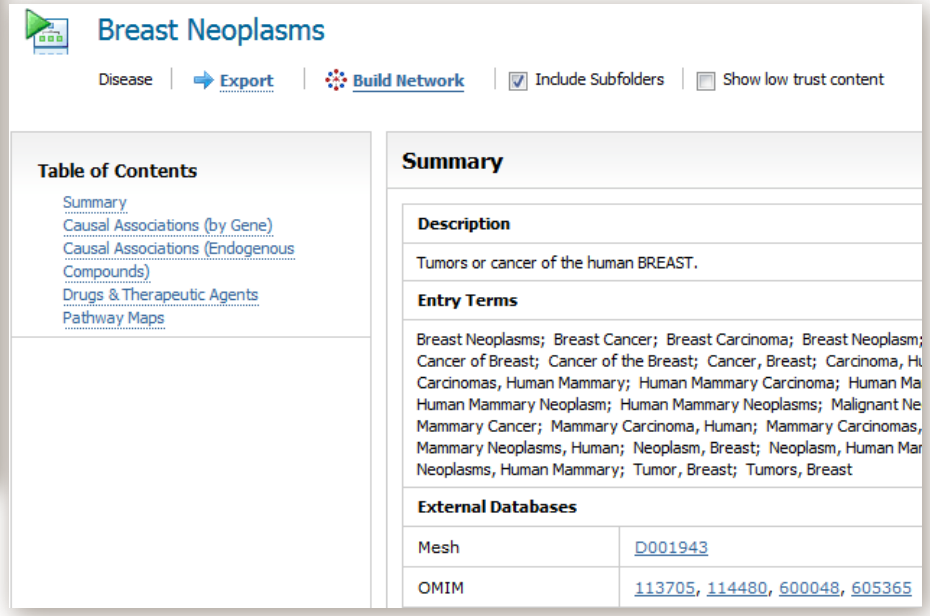




- Question:
  - What can I learn about genes being overly expressed in breast cancer?
  - What RNA has a change in abundance in both breast cancer and melanoma?

# EZ SEARCH FOR INFORMATION RELATED TO BREAST CANCER

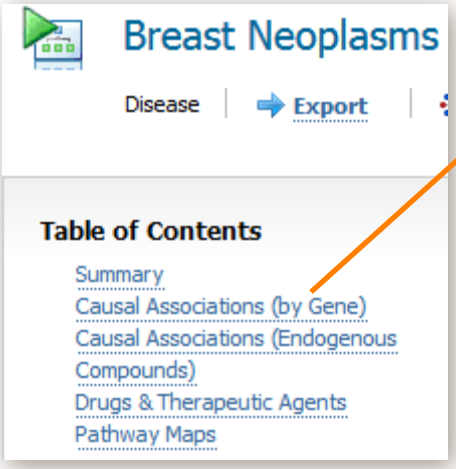



THOMSON REUTERS

# WHAT CAN I LEARN ABOUT GENES BEING OVERLY EXPRESSED IN BREAST CANCER?




**Causal Associations (by Gene)**

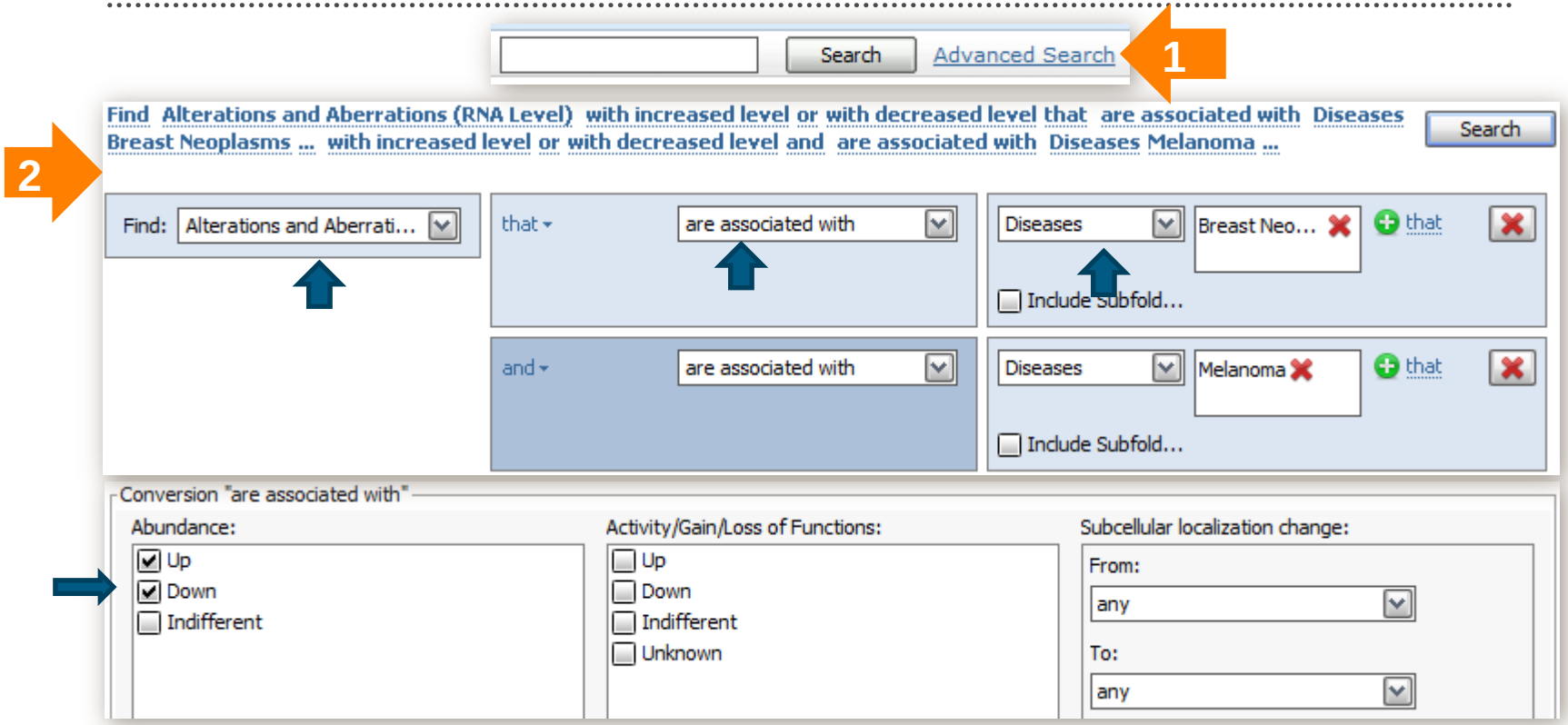
ctla 1/4

Result pages: 1 • 6 7 8 9 10 • 168 (Showing results 141 to 160 of 3344)

| # | Gene | Alteration Level | Alteration Type | Alteration Subtype | Details | Abundance | Activity/Gain/Loss of Function | Normal/Pathology Concentration | Subcellular Localization Change | Organ/Tissue Distribution | Disease | Info |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 156 | Rx ESR1 | Protein level | Protein | | ESR1_HUMAN | up, Indifferent, down | | 0.07074±0.06045 → 0.0856±0.06719 ug/L, 0.012±0.0083 → 0.0203±0.0116 ug/kg | | Serum, Skin, Lung, Bone and Bones, Prostate, Stromal Cells, Mammary Glands, Human, Breast, Mucous Membrane, Liver, Gastric Mucosa | Hereditary breast cancer, Inflammatory Breast Neoplasms, Carcinoma, Ductal, Breast, Breast Neoplasms, Male, Breast Neoplasms | |
| 157 | STAT3 | Protein level | Phosphorylation | | STAT3_(HUMAN)_Ph_Y705 | up | | | | Mammary Glands, Human | Breast Neoplasms | |
| 158 | CTLA4 | Protein level | Isoform | Alternative splicing | CTLA4_HUMAN_2 | up | | 7.69±4.04 → 23.53±14.96 ug/L | | Serum | Breast Neoplasms | |

- CTLA-4 expression changes have been seen in blood and tissue of patients
- CTLA-4 protein abundance has seen an increase in serum from patients

# USE ADVANCED SEARCH TO FIND GENES ASSOCIATED WITH TWO DISEASE




| # | Gene | Name | Type | Abundance | Activity/Gain/Loss of Funct... | Pathology |
|---|---|---|---|---|---|---|
| 29 | PROX1 | PROX1_(HUMAN)_transcript | Mature RNA | indifferent;down;u | down;up | Carcinoma, L<br>Neoplasms;C |
| 30 | ERBB3 | ERBB3_(HUMAN)_transcript | Mature RNA | indifferent;down;u | down;up | Glioblastoma<br>Neoplasms;M<br>Ductal;Endor<br>Breast;Menir |
| 31 | VEGFC | VEGFC_(HUMAN)_transcript | Mature RNA | indifferent;down;u | down;up | Glioblastoma<br>Acute;Neuro<br>Carcinoma, F |

# LIVE DEMO: DATASET

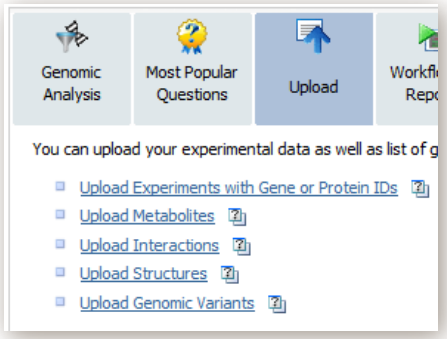
- GSE36765 – Gene expression profiling of CD4+ T cells infiltrating human breast cancer (Discovery Set).
  - Affy/HG-U133_Plus_2
  - Fold change: <= -5 or >= 5
  - FDR P-value: 0.05
  - Two groups of patients:
    - Breast Tumor CD4+ T-Lymph high infil vs. Patient PB CD4+ T-Lymph high infil. (730 DEGs)
    - Breast Tumor CD4+ T-Lymph low infil vs. Patient PB CD4+ T-Lymph low infil. (778 DEGs)

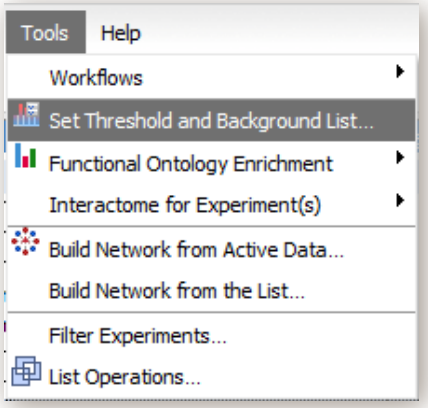
# CASE STUDY ON GENE EXPRESSION PROFILING OF CD4+ T CELLS INFILTRATING HUMAN BREAST CANCER

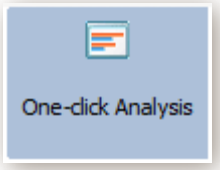
Upload dataset into MetaCore

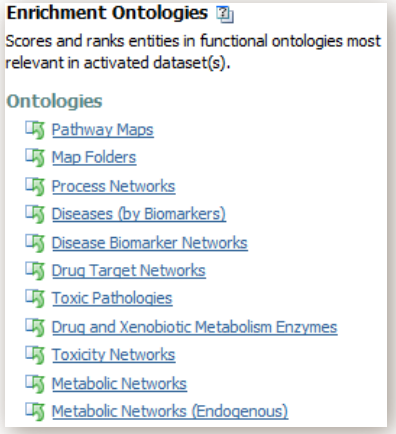
Set Threshold/ Background

Analyse Map Folders



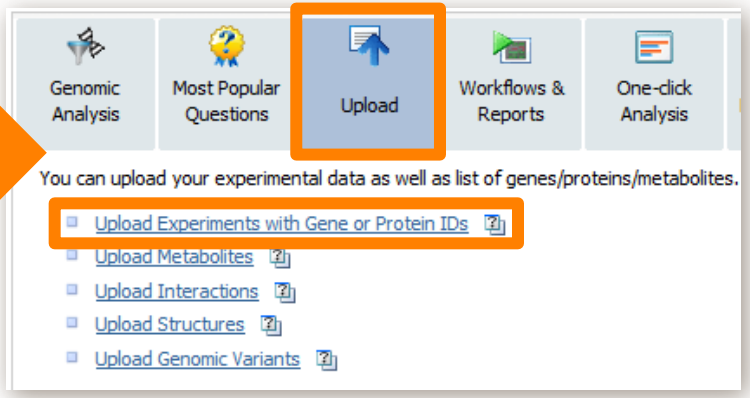



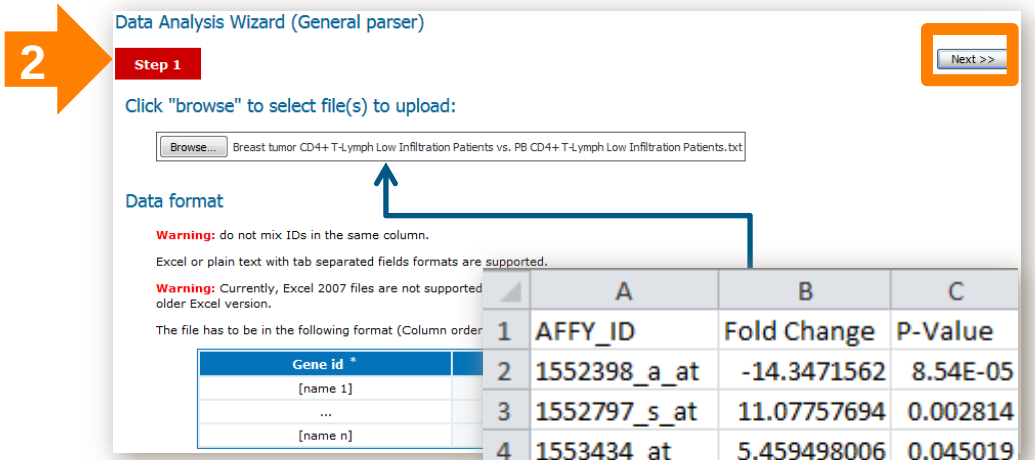



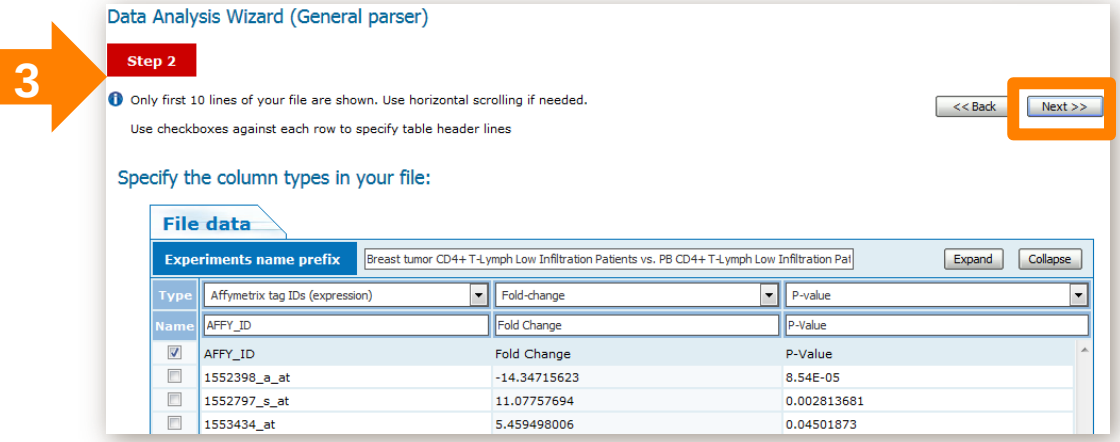

- Questions:
  - What overall process is impacted when comparing infiltrating vs. peripheral blood T-lymphocytes?
  - Are there particular pathways showing a difference between high and low infiltrating T-lymphocytes?

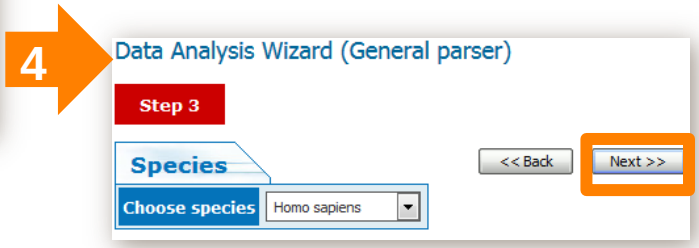
# UPLOAD TRANSCRIPTOMIC DATA INTO METACORE




| | A | B | C |
|---|---|---|---|
| 1 | AFFY_ID | Fold Change | P-Value |
| 2 | 1552398_a_at | -14.3471562 | 8.54E-05 |
| 3 | 1552797_s_at | 11.07757694 | 0.002814 |
| 4 | 1553434_at | 5.459498006 | 0.045019 |
| 5 | 1553602_at | 115.85537 | 0.025612 |
| 6 | 1553613_s_at | 19.86887452 | 0.000323 |
| 7 | 1553986_at | 14.02880884 | 0.000291 |
| 8 | 1554018_at | 16.6920076 | 0.008601 |
| 9 | 1554600_s_at | 5.360358331 | 0.003606 |
| 10 | 1555778_a_at | 22.14873322 | 0.013634 |

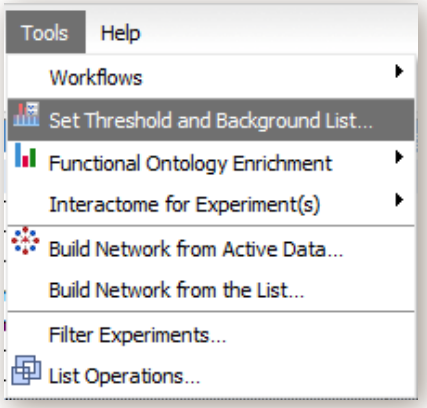


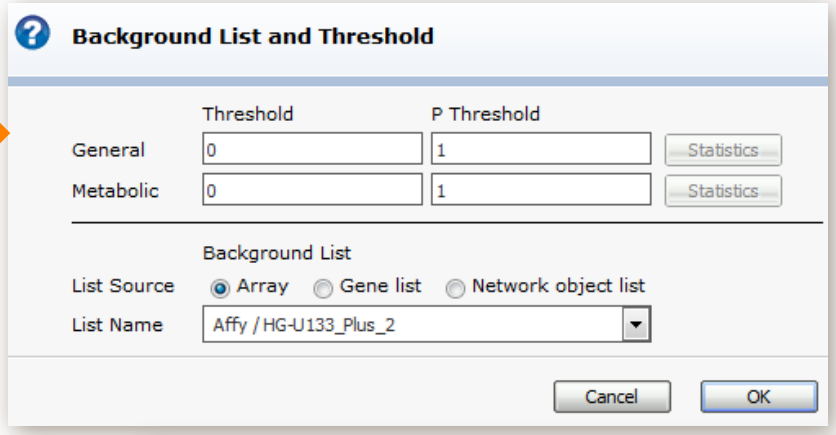

THOMSON REUTERS

# SET THRESHOLD/BACKGROUND AND RUN MAP FOLDER ANALYSIS

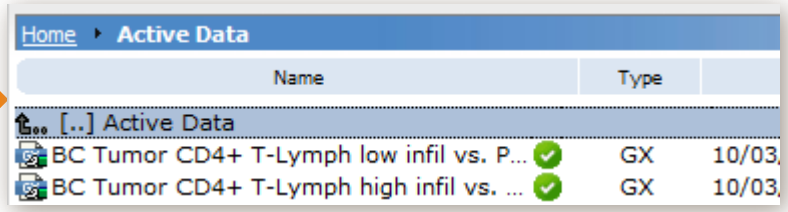
1



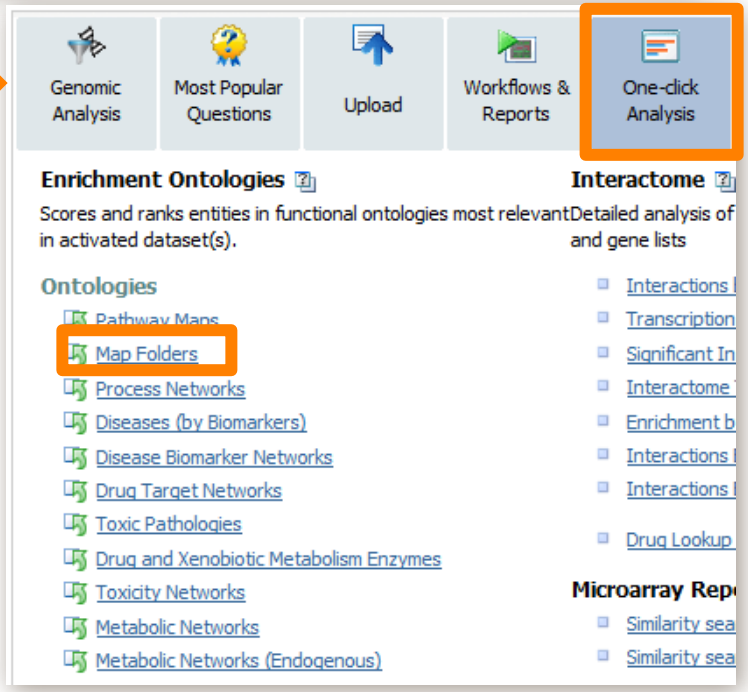

2




3




4




Map Folders are a collection of pathway maps grouped together under one overarching topic such as ‘Immune System Response’ or ‘Breast Neoplasms’

THOMSON REUTERS

# WHAT OVERALL PROCESS IS IMPACTED WHEN COMPARING INFILTRATING VS. PERIPHERAL BLOOD T-LYMPHOCYTES?

| | Experiment name | Species | Network Objects |
|---|---|---|---|
| ☑ | Breast tumor CD4+ T-Lymph High Infiltration Patients vs. PB CD4+ T-Lymph High Infiltration Patients | Homo sapiens | 692 |
| ☑ | Breast tumor CD4+ T-Lymph Low Infiltration Patients vs. PB CD4+ T-Lymph Low Infiltration Patients | Homo sapiens | 757 |

Ratio of differentially expressed genes from dataset (**GREEN**) over all network objects in folder (**RED**)

Graphical representation of –log(pValue)

| # | Map folders | 0 2 4 6 8 10 -log(pValue) | pValue | min(pValue) | FDR | Ratio |
|---|---|---|---|---|---|---|
| 1 | Systemic Lupus Erythematosus | | 1.311e-13<br>2.068e-6 | 1.311e-13 | 8.127e-12<br>1.165e-5 | 130/1267<br>115/1267 |
| 2 | Dermatitis, Allergic Contact | | 3.586e-13<br>7.333e-5 | 3.586e-13 | 1.112e-11<br>2.842e-4 | 107/958<br>87/958 |
| 3 | Neurofibromatoses | | 3.317e-9<br>7.715e-12 | 7.715e-12 | 3.427e-8<br>4.783e-10 | 110/1145<br>123/1145 |
| 4 | Immune system response | | 5.494e-11<br>2.849e-5 | 5.494e-11 | 1.135e-9<br>1.359e-4 | 113/1117<br>100/1117 |

Map folder name

Significance of overlap of differentially expressed genes in the folder

- Top 3/4 hits are related to immune disease or the immune system processes.
- Marked difference in the presence of differentially expressed genes between low and high infiltrating T-lymphocytes

THOMSON REUTERS

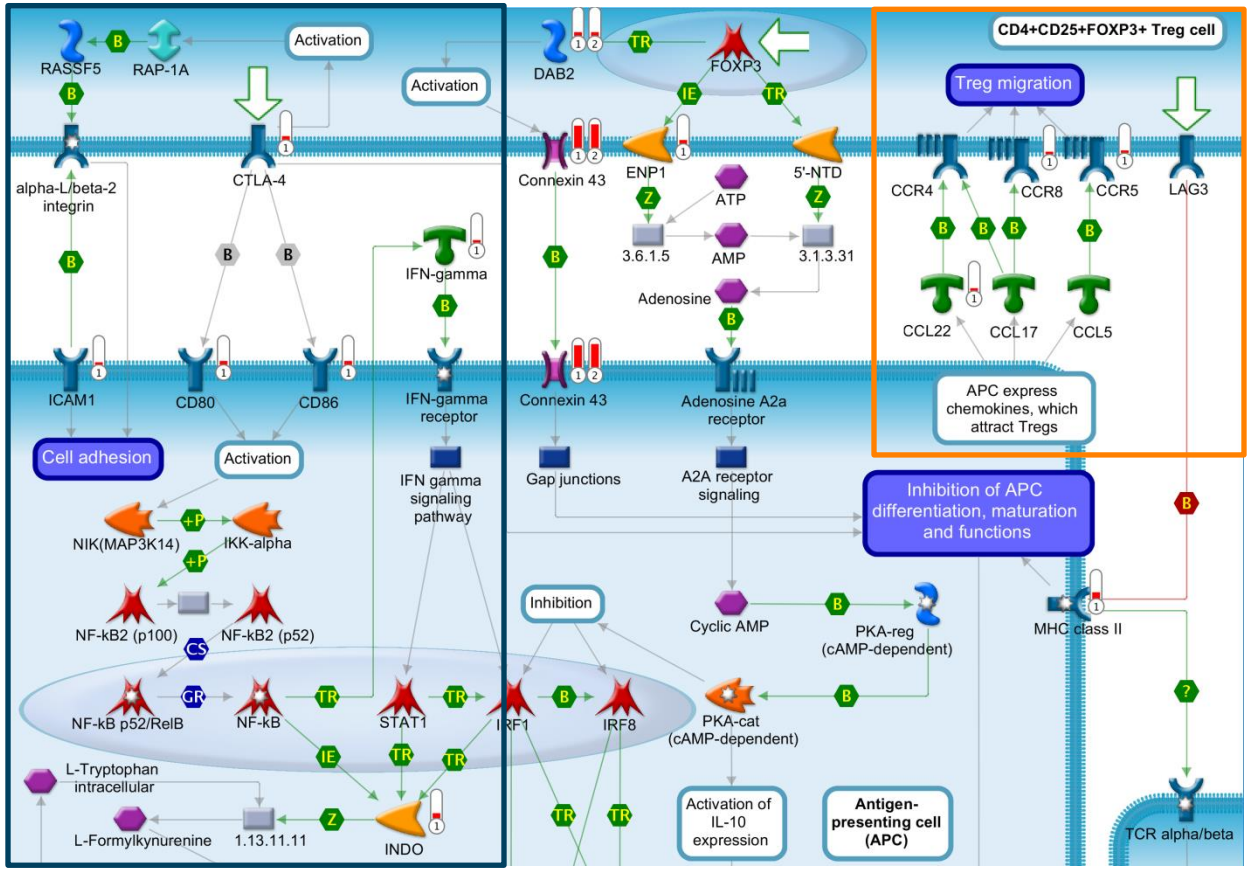
# ARE THERE PARTICULAR PATHWAYS SHOWING A DIFFERENCE BETWEEN HIGH AND LOW INFILTRATING T-LYMPHOCYTES?

| Experiment name | Species | Network Objects |
|---|---|---|
| Breast tumor CD4+ T-Lymph High Infiltration Patients vs. PB CD4+ T-Lymph High Infiltration Patients | Homo sapiens | 692 |
| Breast tumor CD4+ T-Lymph Low Infiltration Patients vs. PB CD4+ T-Lymph Low Infiltration Patients | Homo sapiens | 757 |

**Immune Response_ T regulatory cell-mediated modulation of antigen-presenting cell functions**

CTLA-4 activation of CD80 and CD86 can lead to the expression of IFN-gamma and INDO which can lead to Th1 cell apoptosis.

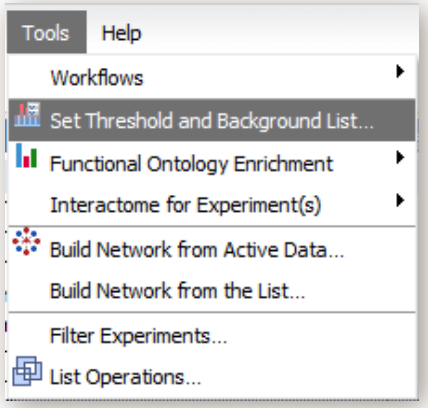
Increased expression of receptors (CCR8 & CCR5) and ligand (CCL22) for $T_{reg}$ migration is seen in high infiltrating T-lymphocytes.




THOMSON REUTERS

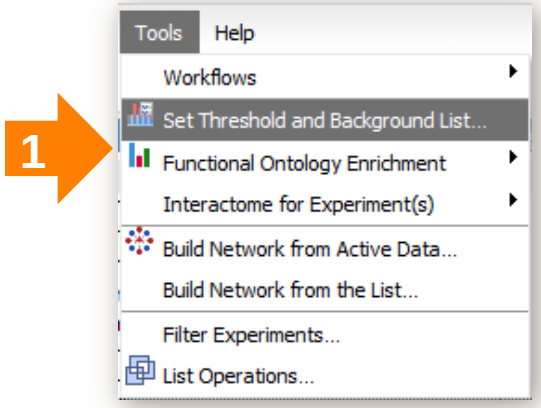
# COMPARING TWO DATASETS WITH COMPARE EXPERIMENTS

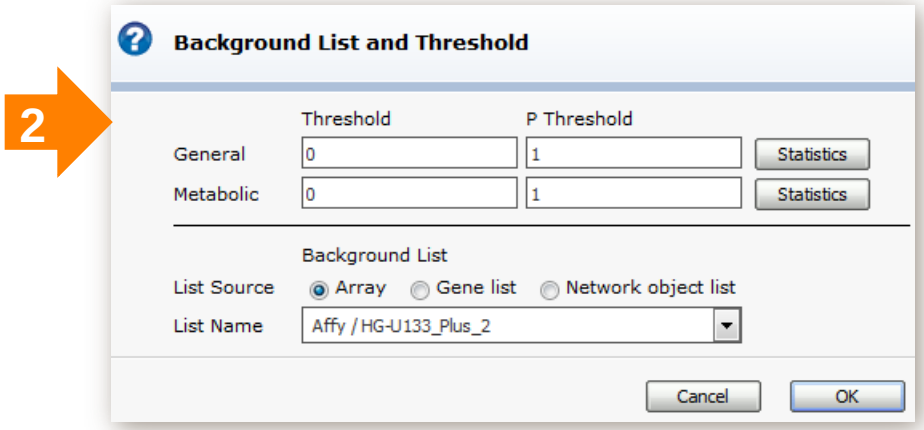
Set Threshold/ Background → Compare Experiments Workflow

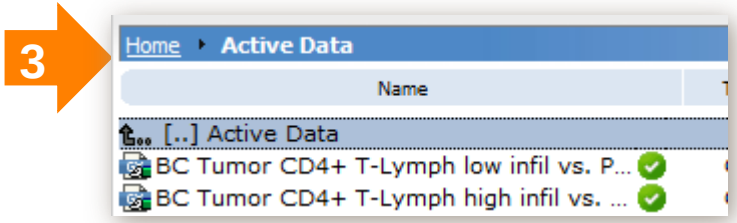


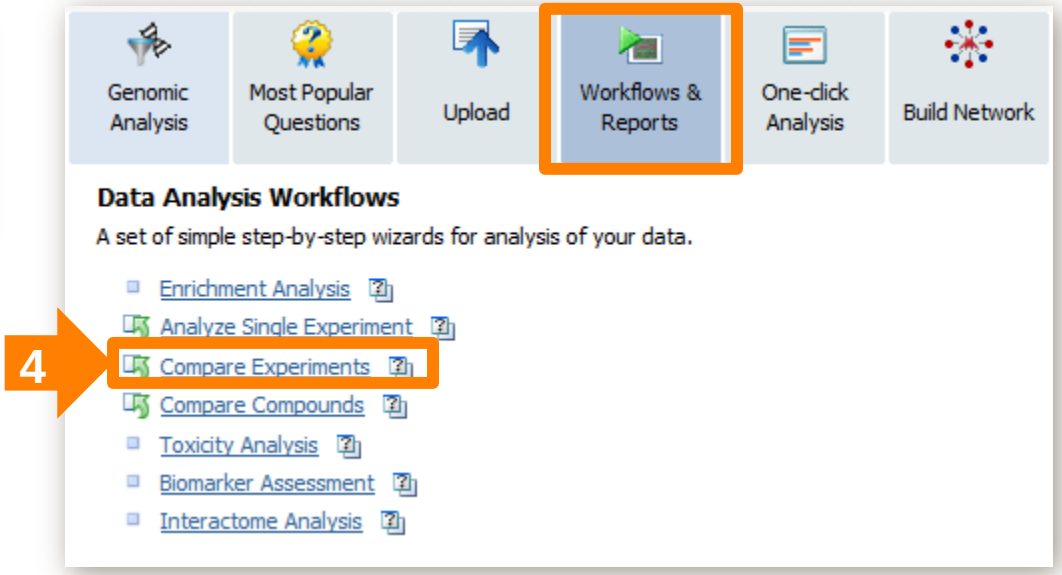
- Question:
  - What pathways and processes are commonly affected by high and low infiltrating T-lymphocytes?

# SET THRESHOLD/BACKGROUND AND RUN COMPARE EXPERIMENTS WORKFLOW



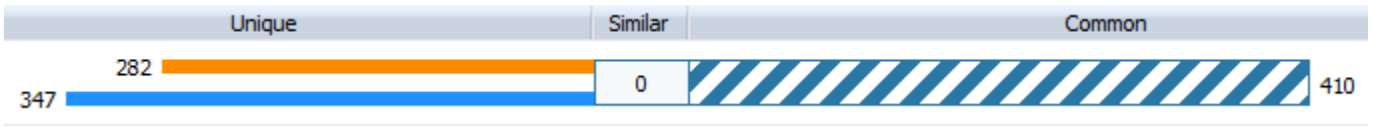



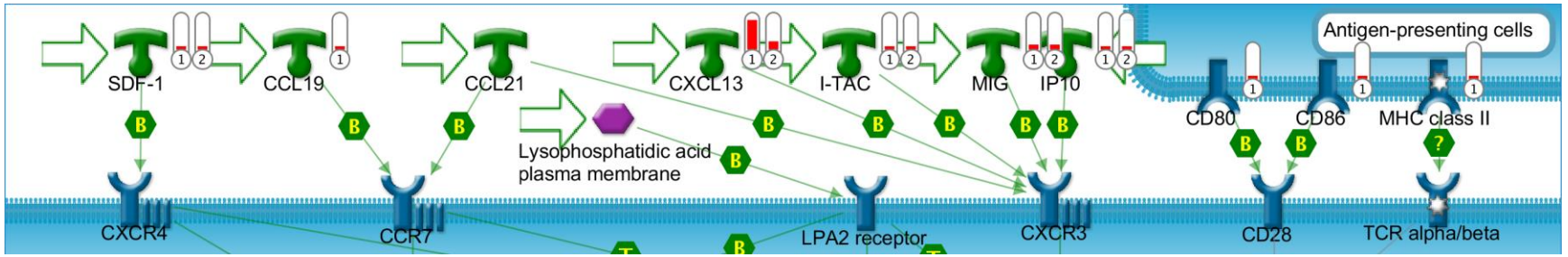



# WHAT PATHWAYS ARE COMMONLY AFFECTED BY HIGH AND LOW INFILTRATING T-LYMPHOCYTES?

| Experiment name | Species | Network Objects |
|---|---|---|
| Breast tumor CD4+ T-Lymph High Infiltration Patients vs. PB CD4+ T-Lymph High Infiltration Patients | Homo sapiens | 692 |
| Breast tumor CD4+ T-Lymph Low Infiltration Patients vs. PB CD4+ T-Lymph Low Infiltration Patients | Homo sapiens | 757 |




**Chemotaxis_Leukocyte chemotaxis**

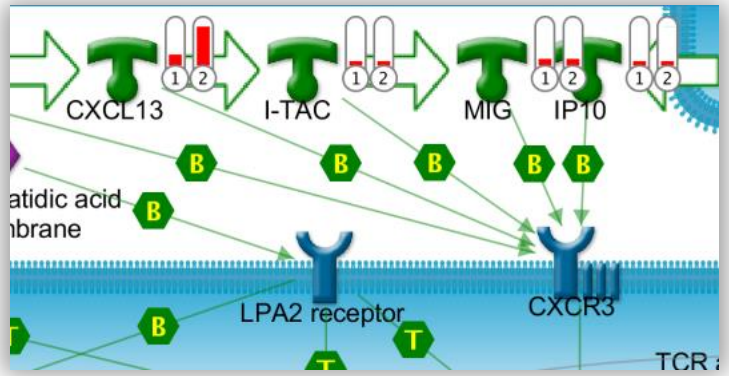


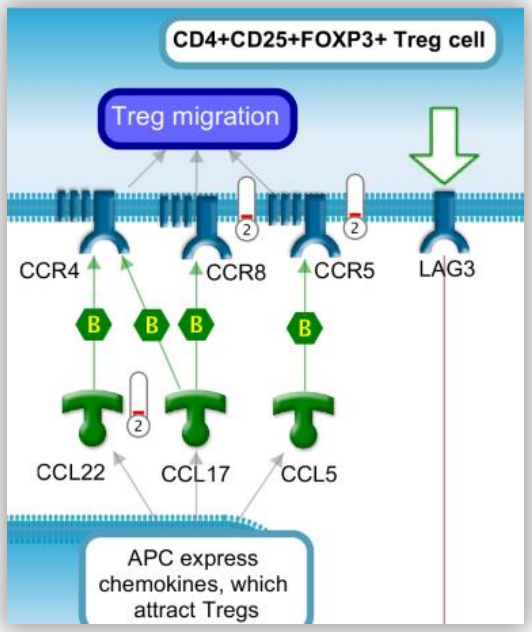

- Leukocyte chemotaxis had additional genes unique the highly infiltrating T-lymphocytes.
- CXCL13 has a far higher increase in expression related to highly infiltrating T-lymphocytes.

# SUMMARY

- CXCL13 if highly expressed by the CD4+ T-lymphocytes in the patients with high infiltration could be and interesting biomarker to look into.

- A number of receptors and ligands important for $T_{reg}$ migration have be up regulated in CD4+ T-lymphocytes highly infiltrating breast cancer tumors.



# LUNCH BREAK

# RETURN @ 1:30PM

# AGENDA

- **Morning Session (Introductory Topics) – 9:30 AM to 12:30 PM**
  - 9:30-10:00 – General Overview: Thomson Reuters Systems Biology Solutions
  - 10:00-10:30 – Knowledge Mining: Explore the database and exporting
  - 10:30-11:00 – Uploading, filtering and setting a background
  - 11:00-11:45 – Running Functional Enrichments and exploring Pathway Maps
  - 11:45-12:30 – Running Workflows

- **Lunch Break – 12:30 PM to 1:30 PM**

- **Afternoon Session (Advanced Topics) – 1:30 PM to 4:30 PM**
  - **Topics of Interest** (will discuss these or other topics of interest to attendees)
    - 30-45 min – Key Pathway Advisor – Hypothesizing key hubs using the causal reasoning algorithm
    - 30-45 min – Using the Microarray repository for gene comparisons against public data
    - 45-60 min – Building Networks with MetaCore
    - 30-45 min – Multi-omics analysis with miRNA & mRNA data
    - 45-60 min – Multi-omics analysis with Metabolomics & mRNA data

# KEY PATHWAY ADVISOR

Hypothesizing key hubs using the causal reasoning algorithm

# Key Pathway Advisor (KPA) Features

## Comprehensive Pathway Analysis

*Differential expression* is an *effect* rather than *cause* of disease. *Causal Network Analysis* identifies upstream molecules of DEGs to predict the causes of biological deregulation.

*Synergy Pathway Analysis* automatically defines processes and pathway maps equally enriched with both experimentally defined and predicted Key Hubs gene sets in order to show connected biologically relevant results.

## Integrated OMICs Data Analysis

KPA allows analyzing of gene expression analysis and associated gene variant data. Use *MetaCore Genomic Analysis Tools* to annotate your experimentally defined Gene Variant list and narrow it down to the list of the most promising protein function impact variants.

Resulting list could be uploaded into KPA with expression data in order to identify pathway impact and hypothesize about gain and loss of function effect.

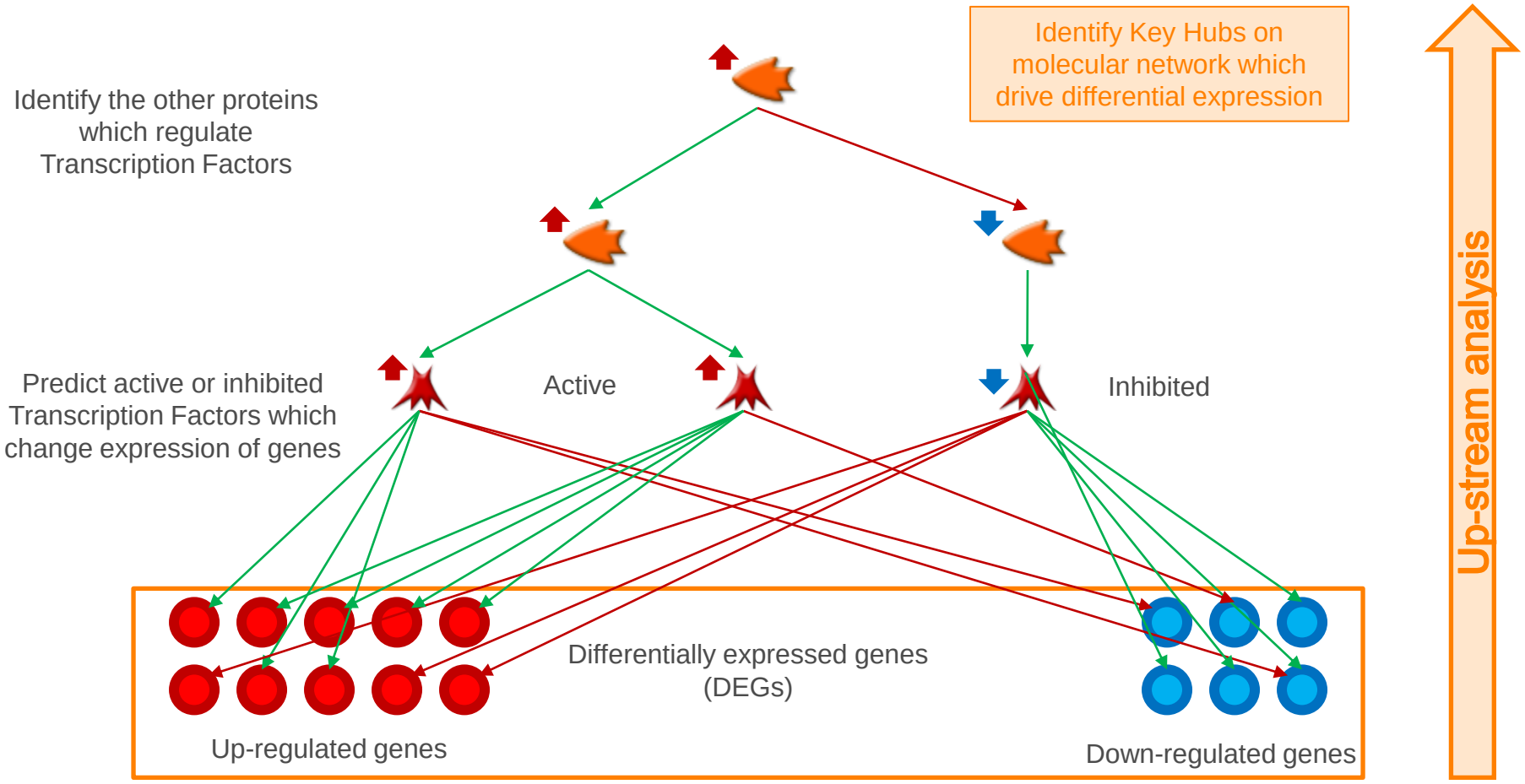
## One-Click Analysis

KPA has a simple and intuitive interface. Import your data into analysis and apply default analytical parameters – a whole comprehensive workflow will be calculated on the background showing all results in structured reports.

Study predicted Key Hubs: Identify driver genes and their activity change that regulate differential expression and/or have gene variants – candidate drug targets.

Get better biological understanding analyzing pathway maps enriched by DEGs, Gene Variants and predicted Key Hubs activity change.

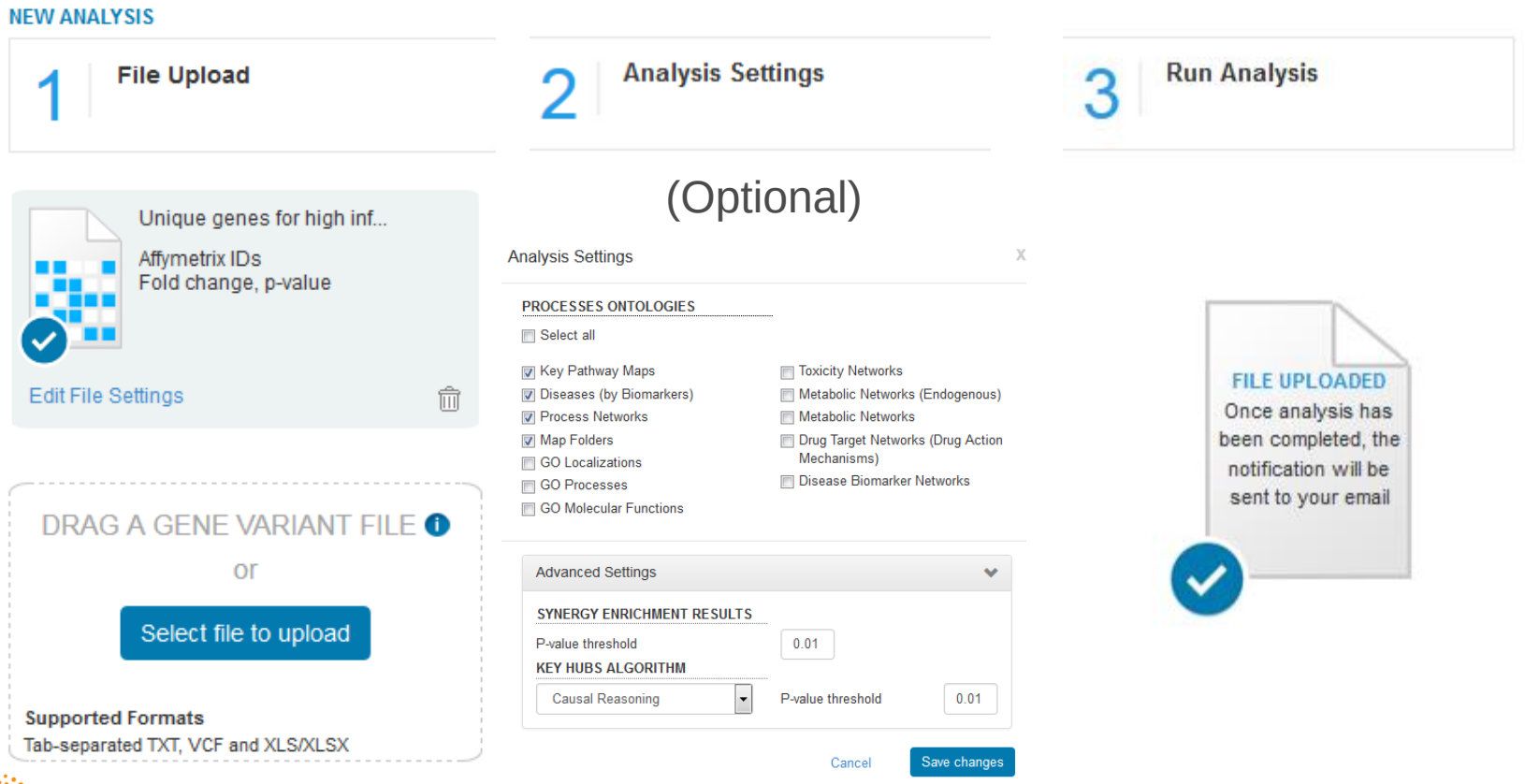
# CAUSAL REASONING EXPRESSION ANALYSIS




1. Chindelevitch L, Ziemek D, Enayetallah A, Randhawa R, Sidders B, et al. (2012) Causal reasoning on biological networks: interpreting transcriptional changes. Bioinformatics 28: 1114-1121.
2. Pollard J Jr, Butte AJ, Hoberman S, Joshi M, Levy J, Pappo J. (2005) A computational model to define the molecular causes of type 2 diabetes mellitus. Diabetes Technol Ther. 2005 Apr;7(2):323-36.

# HOW TO UPLOAD A DATASET INTO THE KEY PATHWAY ADVISOR

**Format and upload data** → **Choose settings** → **Analyse and explore results**




THOMSON REUTERS

# FORMATTING AND UPLOADING DATA

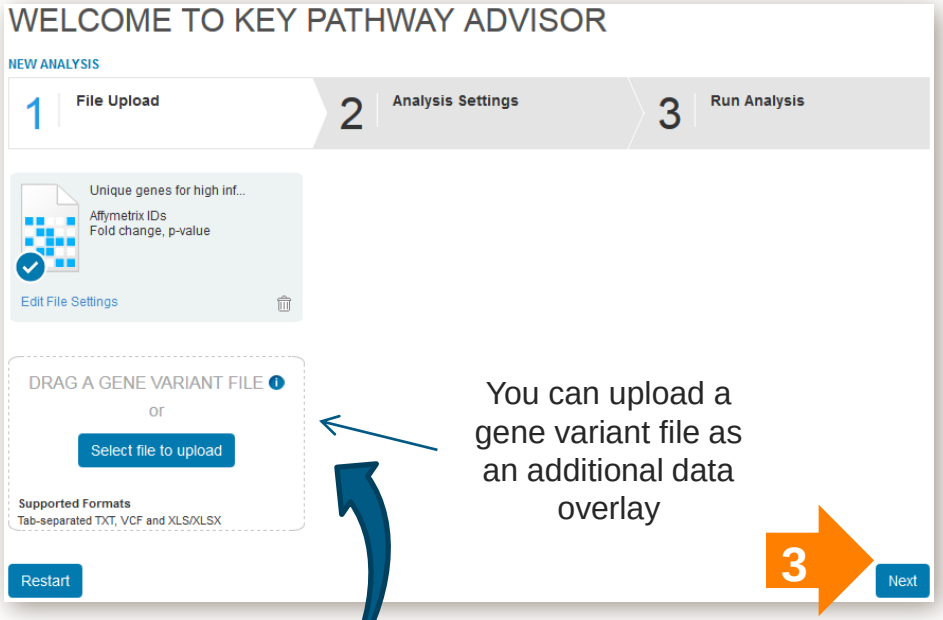
You can use the same identifiers as in MetaCore

Keep column titles simple

P-values can be included but are not required

1

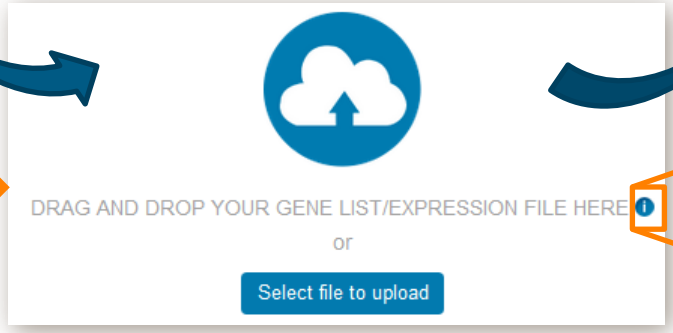
| | A | B | C |
|---|---|---|---|
| 1 | AFFY ID | fold change | p-value |
| 2 | 1405_i_at | -1.981575999 | 0.02476869 |
| 3 | 1552302_at | 2.118812101 | 0.01729475 |
| 4 | 1552303_a_at | 1.908469521 | 0.017252 |
| 5 | 1552309_a_at | 13.04226067 | 2.65E-06 |
| 6 | 1552362_a_at | -2.264977411 | 0.03711153 |
| 7 | 1552370_at | 2.488421551 | 0.00316702 |




You can upload a gene variant file as an additional data overlay

3

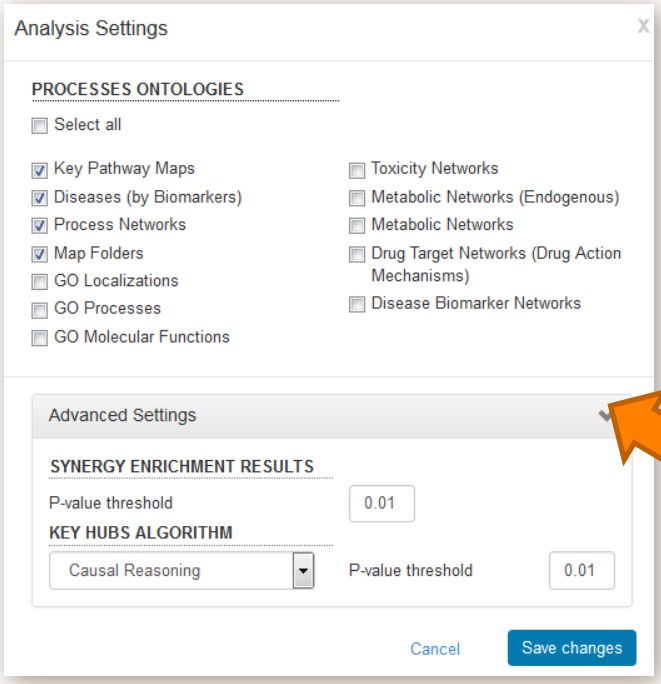
2

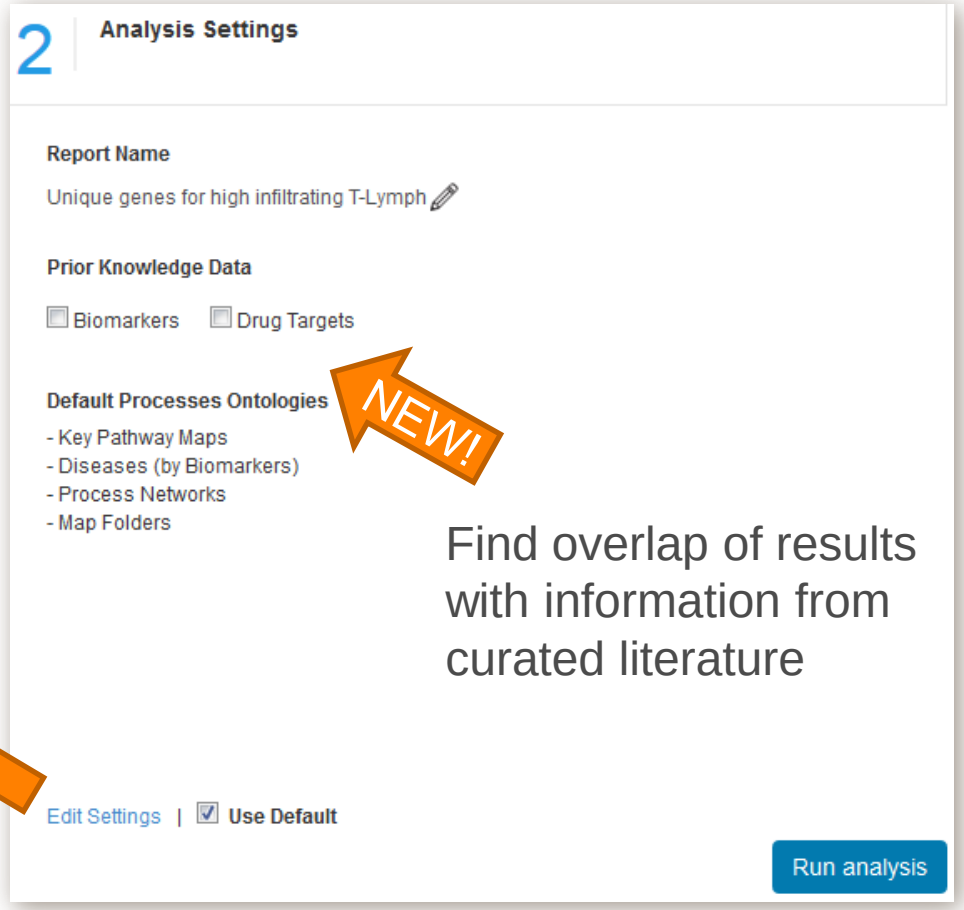



Click here to find additional information on uploading data

# ANALYSIS SETTINGS

Include additional ontologies for enrichment

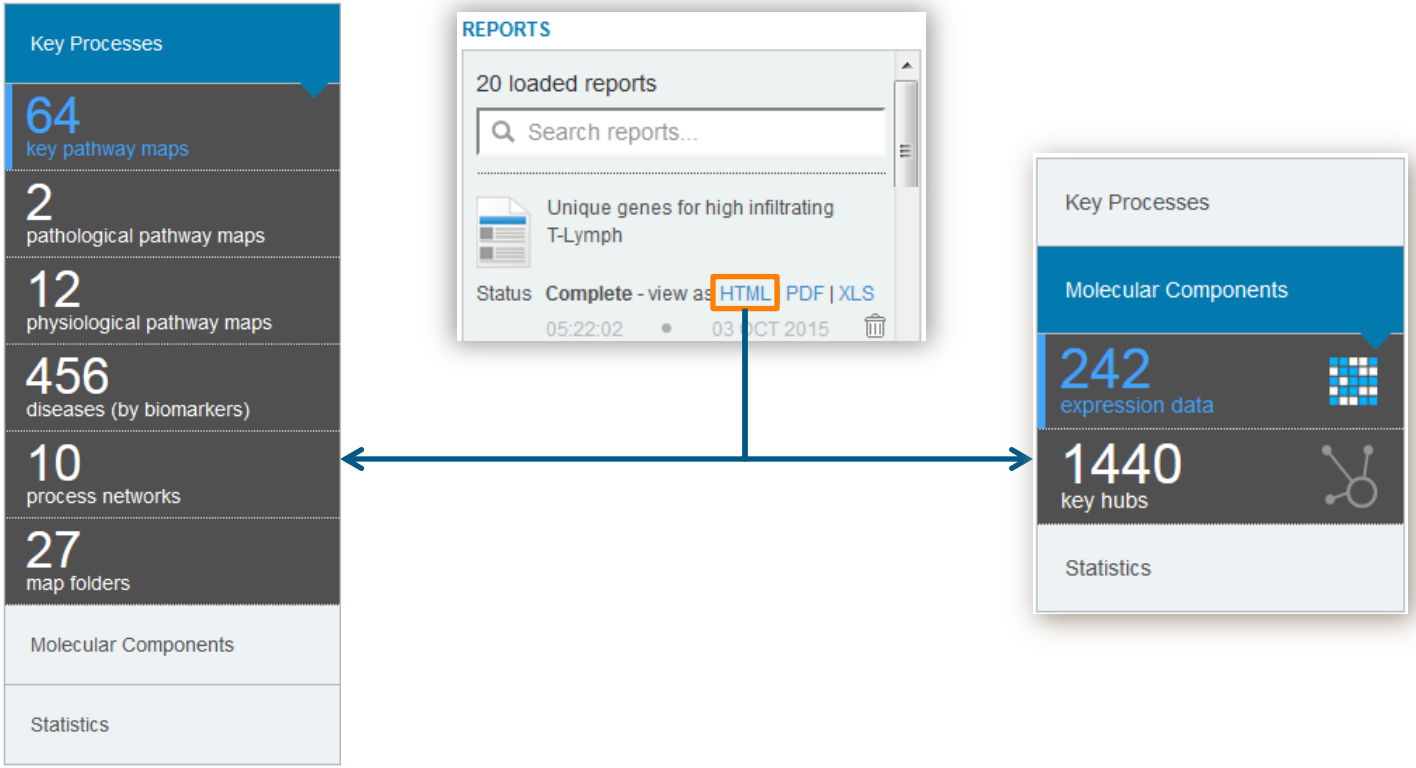



Find overlap of results with information from curated literature

THOMSON REUTERS

# USE HTML INTERFACE TO EXPLORE RESULTS




- Question:
  - What key pathways are linked with the unique differentially expressed genes from high infiltrating T-lymphocytes?
- Question:
  - What direct and indirect regulators could be responsible for the unique gene signature for high infiltrating T-lymphocytes?

# WHAT KEY PATHWAYS ARE LINKED WITH THE UNIQUE DIFFERENTIALLY EXPRESSED GENES IN HIGH INFILTRATING T-LYMPHOCYTES?

## Unique genes for high infiltrating T-Lymph

Download Report

**Key Processes**

- **64** key pathway maps
- **2** pathological pathway maps
- **12** physiological pathway maps
- **456** diseases (by biomarkers)
- **10** process networks
- **27** map folders

### KEY PATHWAY MAPS

Pathway maps are graphic images representing complete biochemical pathways or signaling cascades in a commonly accepted sense. All maps listed below are enriched with both input genes and Key Hubs. The enrichment analysis was done for whole set of pathway maps including physiological and pathological processes.

| # | Name | Input Objects p-value | Key Hubs p-value | Union Objects p-value | stacked grouped |
|---|---|---|---|---|---|
| 1 | Development_Cytokine-mediated regulation of megakaryopoiesis | 0.005264 | 1.363E-14 | 3.108E-16 | |
| 2 | Immune response_IL-22 signaling pathway | 7.681E-4 | 2.025E-13 | 5.072E-16 | |
| 3 | SLE genetic marker-specific pathways in T cells | 3.943E-6 | 5.004E-10 | 1.868E-15 | |
| 4 | Immune response_Role of PKR in stress-induced antiviral cell response | 0.005264 | 1.472E-13 | 3.726E-14 | |
| 5 | Immune response_Differentiation and clonal expansion of CD8+ T cells | 7.278E-6 | 5.429E-8 | 3.414E-13 | |
| 6 | Immune response_IL-12 signaling pathway | 0.00271 | 1.619E-10 | 2.683E-12 | |

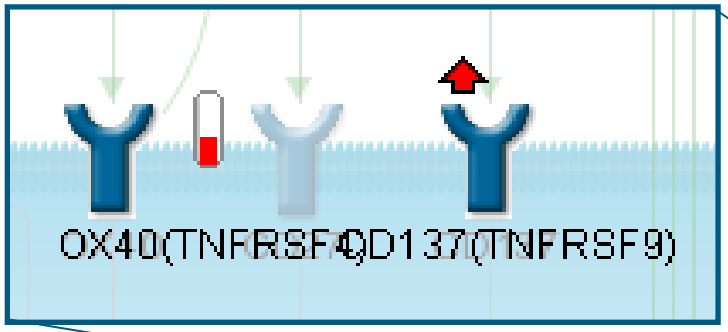
Enrichment p-value of differentially expressed genes

Enrichment p-value of key hubs

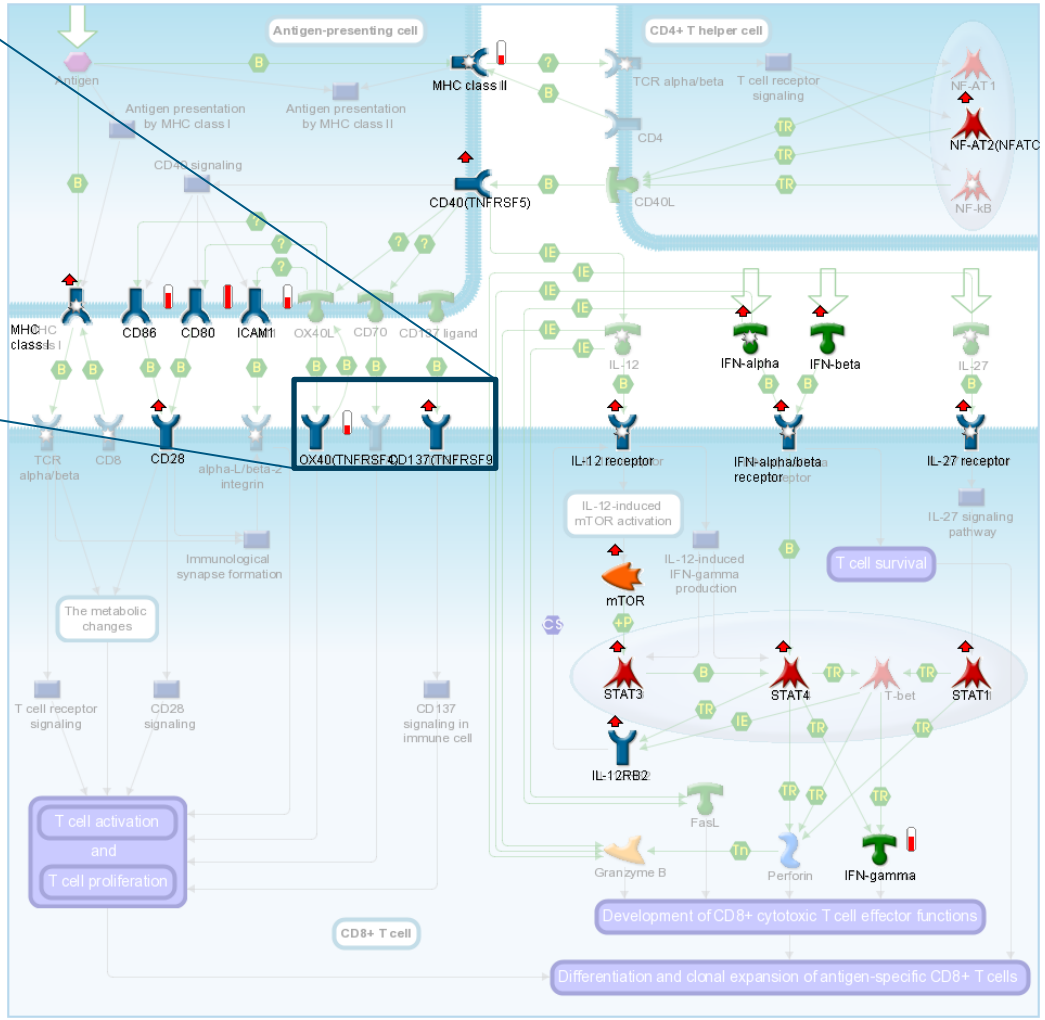
Enrichment p-value using union of key hubs and DEGs

THOMSON REUTERS

# HOW IS THIS SIGNATURE IMPACTING CD8+ DIFFERENTIATION AND EXPANSION?




**Immune Differentiation and clonal expansion of CD8+ T cells**




OX40 is uniquely over-expressed in the highly infiltrating T-lymphocytes.

CD137 is predicted to have an increase in activity due to the gene signature

# WHAT DIRECT OR INDIRECT REGULATORS COULD BE RESPONSIBLE FOR THE GENE SIGNATURE?

## Unique genes for high infiltrating T-Lymph

Download Report

Key Processes

Molecular Components

242 expression data

1446 key hubs

Statistics

### KEY HUBS

A list of network objects (proteins, protein complexes, miRNAs) that regulate expression of uploaded differentially expressed genes or molecules overconnected with the uploaded gene list.

| # | Network Object | Gene | Molecular Function | Activity | Correct/All prediction | p-value | Distance |
|---|---|---|---|---|---|---|---|
| 1 | MafK | MAFK | Transcription factor | ▲ | 90/101 | 7.118E-17 | 3 |
| 2 | MSK1 | RPS6KA5 | Protein kinase | ▲ | 85/97 | 5.216E-15 | 3 |
| 3 | LRP16 | MACROD1 | Generic binding protein | ▲ | 78/89 | 6.837E-14 | 3 |
| 4 | NIK(MAP3K14) | MAP3K14 | Protein kinase | ▲ | 80/93 | 3.112E-13 | 3 |
| | KIR2DL3 | KIR2DL3 | Generic receptor | ▼ | 29/33 | 5.464E-06 | 3 |
| | PD-1 | PDCD1 | Generic receptor | ▼ | 29/33 | 5.464E-06 | 3 |
| | CD137(TNFRSF9) | TNFRSF9 | Generic receptor | ▲ | 14/15 | 0.0004883 | 3 |

Indicates if a gene is being activated (+) or inhibited (-)

The correct number of DEGS at the specified distance from the key hub vs. all DEGs at that distance

significance of overlap between the correct and all predictions

Distance of key hub from DEGs

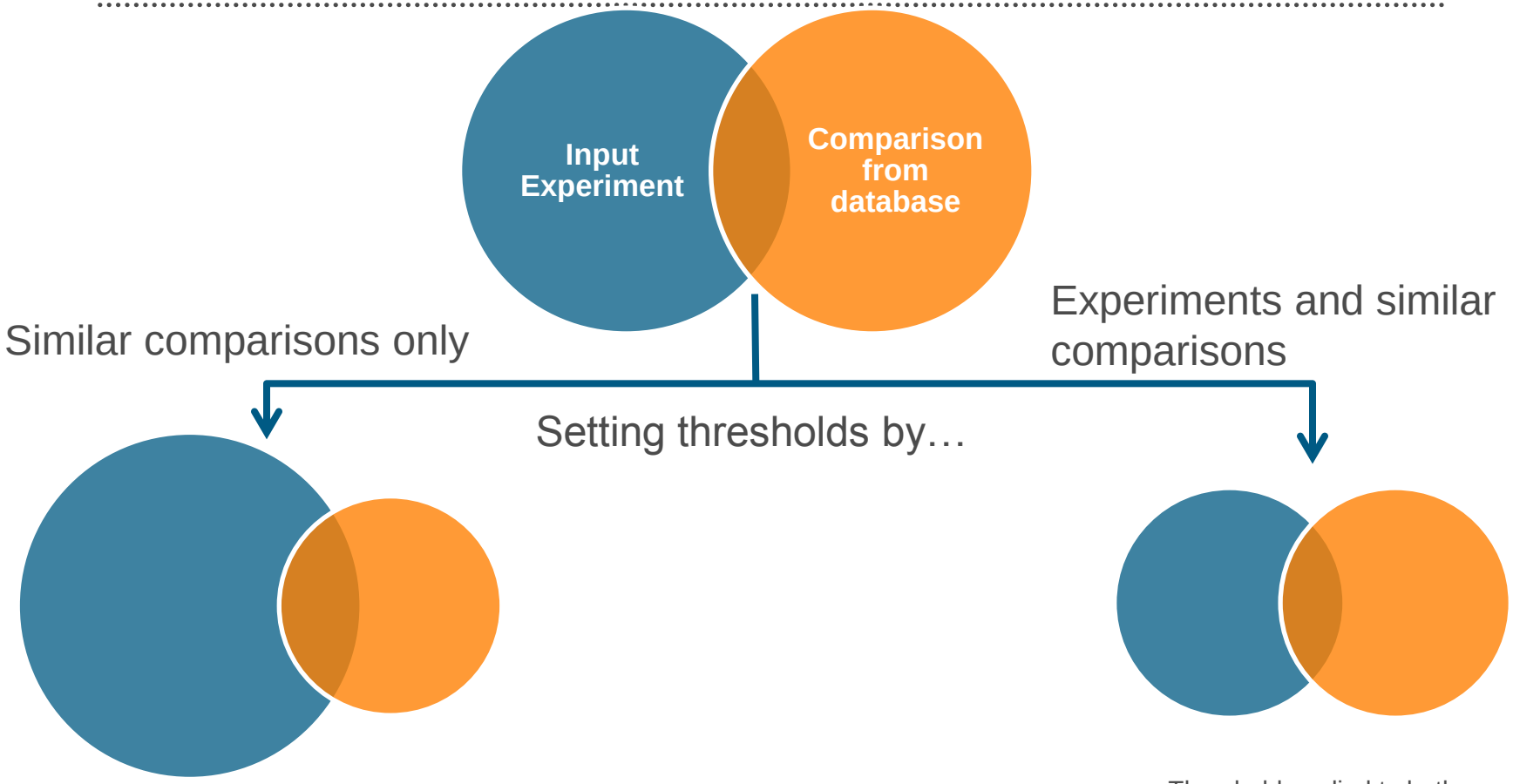
Indicates key hub is also a DEG

THOMSON REUTERS

# MICROARRAY REPOSITORY

Using the Microarray repository for gene comparisons against public data

# SIMILARITY SEARCH BY GENES



# DATASET BACKGROUND

- Patients were recruited at Hospital Santa Creu i Sant Pau (Barcelona, Spain)
  - 15 patients diagnosed with ulcerative colitis
  - 13 patients diagnosed with Crohn's disease
  - 7 health controls
- Samples were taken by endoscopic pinch biopsies of macroscopically un-inflamed mucosa from the ascending colon.

Identification of Novel Predictor Classifiers for Inflammatory Bowel Disease by Gene Expression Profiling

Trinidad Montero-Meléndez, Xavier Llor, Esther García-Planella, Mauro Perretti, Antonio Suárez

Published: October 14, 2013 • DOI: 10.1371/journal.pone.0076235

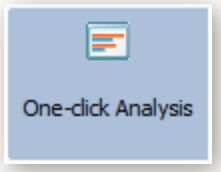
# TRAINING DATASET

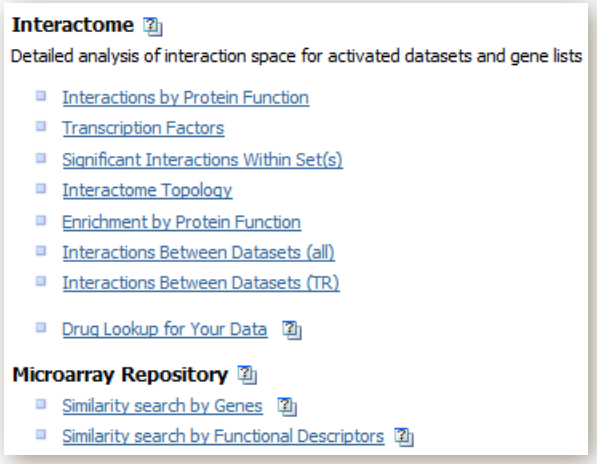
- GSE36807 – Genome-wide analysis of Crohn's disease and ulcerative colitis biopsy samples.
  - Affymetrix Human Genome U133 Plus 2.0
  - Thresholds
    - Fold change: <= -2 or >= 2
    - P-value: 0.05 FDR adjusted
  - Comparison:
    - Ulcerative Colitis vs. Normal (772 Network Objects)

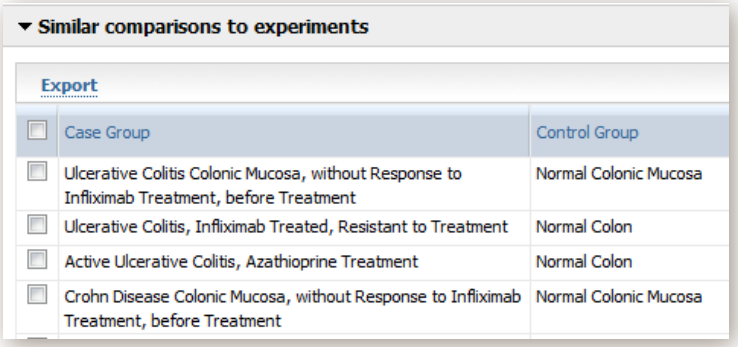
| **Series GSE36807** | Query DataSets for GSE36807 |
|---|---|
| Status | Public on Sep 23, 2013 |
| Title | Genome-wide analysis of Crohn's disease and ulcerative colitis biopsy samples. |
| Organism | Homo sapiens |
| Experiment type | Expression profiling by array |
| Summary | Gene expression patterns of Crohn's disease (CD) and ulcerative colitis (UC) colonic specimens were analyzed using whole-genome microarrays. Healthy control samples were included in order to detect gene expression changes associated with CD or UC. CD and UC samples were also compared in order to identify the molecular mechanisms that distinguish both fenotypes of inflammatory bowel disease. |

# USE SIMILARITY SEARCH BY GENES TO FIND DATASETS WITH SIMILAR GENE SIGNATURES

Similarity Search by Genes → Export Similar Comparisons



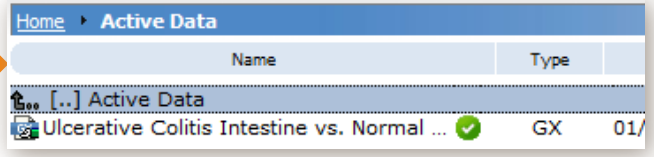



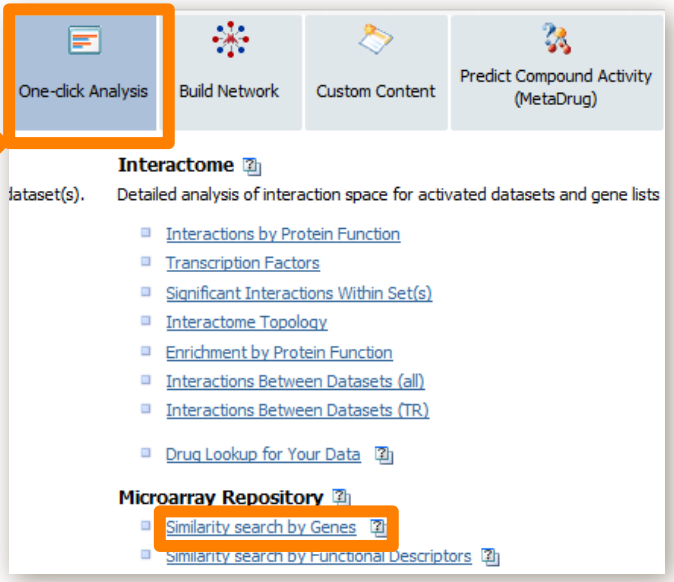

- Question:
  - What datasets in the Microarray Repository have a similar gene signature to ulcerative colitis?

# SET THRESHOLD/BACKGROUND AND RUN COMPARE EXPERIMENTS WORKFLOW

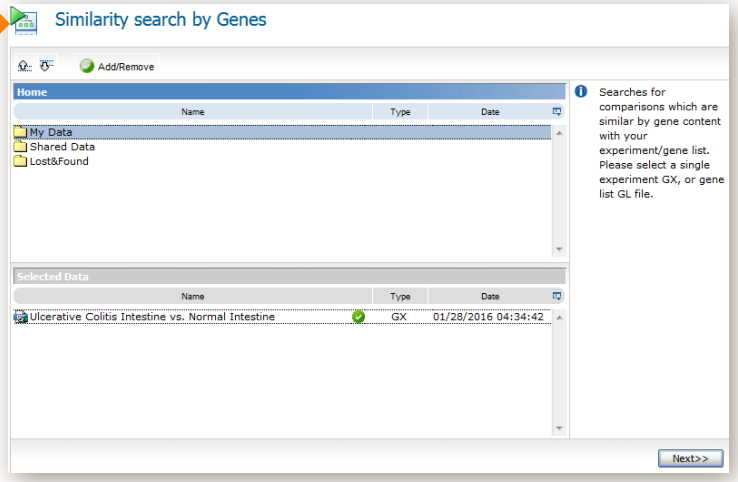
1



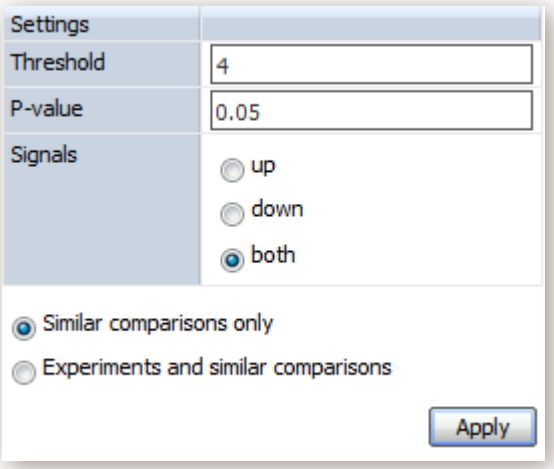

2




3




4




THOMSON REUTERS

# WHAT DATASETS IN THE MICROARRAY REPOSITORY HAVE A SIMILAR GENE SIGNATURE TO ULCERATIVE COLITIS?

**Similar comparisons to experiments**

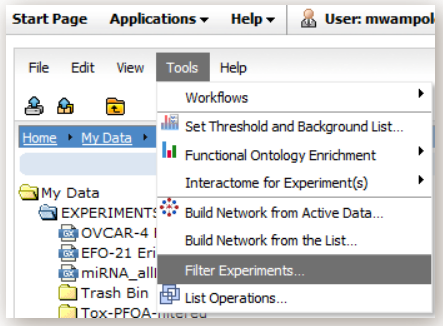
Export

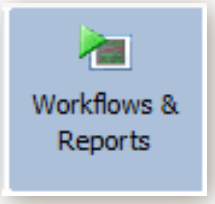
| Case Group | Control Group | Species | Gene Overlap | p-Value | View |
|---|---|---|---|---|---|
| Ulcerative Colitis Colonic Mucosa, without Response to Infliximab Treatment, before Treatment | Normal Colonic Mucosa | Homo sapiens | 304/785 | 0 | |
| Ulcerative Colitis, Infliximab Treated, Resistant to Treatment | Normal Colon | Homo sapiens | 295/775 | 0 | |
| Active Ulcerative Colitis, Azathioprine Treatment | Normal Colon | Homo sapiens | 284/570 | 0 | |
| Crohn Disease Colonic Mucosa, without Response to Infliximab Treatment, before Treatment | Normal Colonic Mucosa | Homo sapiens | 280/711 | 0 | |
| Active Ulcerative Colitis, Left-Sided Colitis | Normal Colon | Homo sapiens | 258/464 | 0 | |
| Active Ulcerative Colitis | Normal Colon | Homo sapiens | 252/450 | 0 | |
| Colorectal Cancer Leukocytes (CD45+ EpCAM-) | Colorectal Cancer Stromal Cells (CD45- EpCAM-) | Homo sapiens | 276/3022 | 2.842e-162 | |
| ✓ Stage II Colorectal Cancer | Normal Adjacent Colon to Stage II Colorectal Cancer | Homo sapiens | 163/667 | 7.15e-162 | |
| Anaplastic Thyroid Carcinoma | Nodular Goiter | Homo sapiens | 235/2065 | 4.854e-157 | |

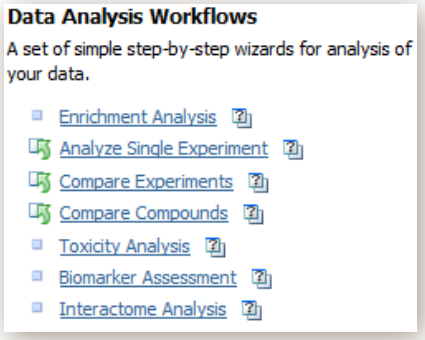
- Lots of similar gene signatures with other ulcerative colitis datasets.
- To see what genes are similar between Stage II colorectal cancer and ulcerative colitis, begin by exporting the CRC data to your metacore account.

# EXPORT RESULTS IN EXCEL FOR ANALYSIS OUTSIDE OF METACORE

**Filter dataset** → **Compare Experiments Workflow**

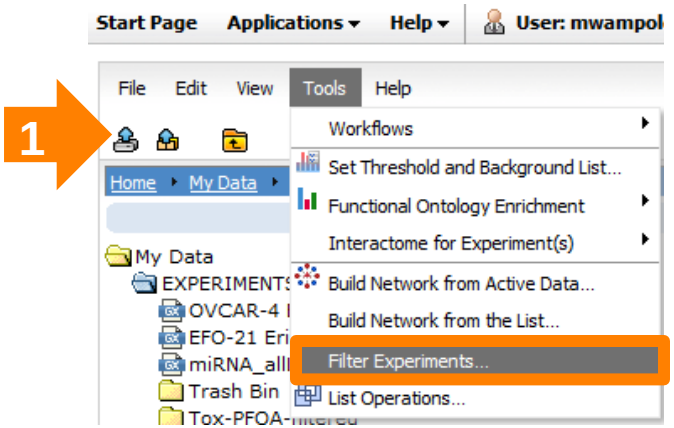


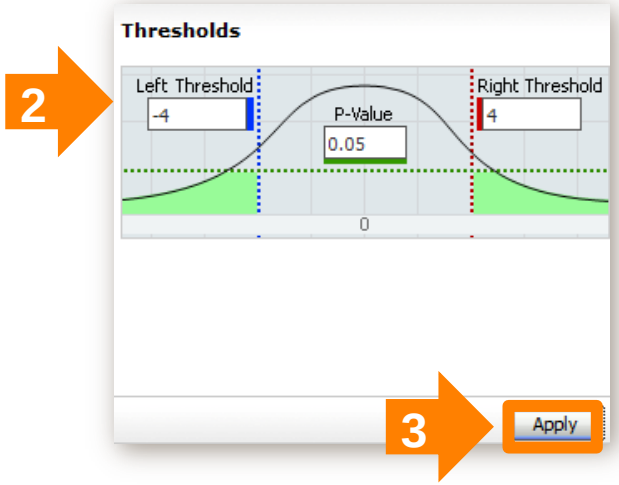



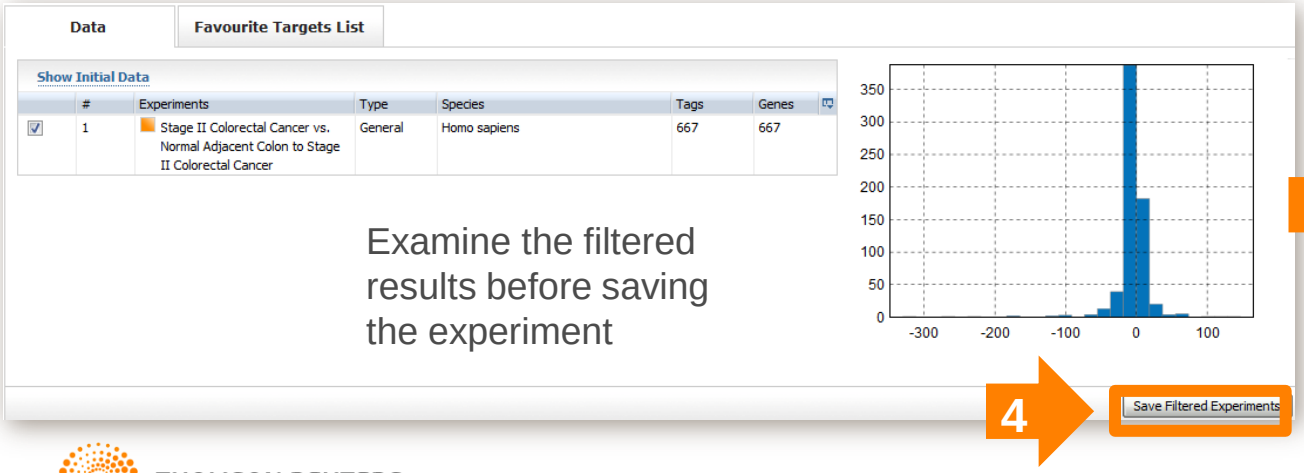

- Question:
  - What pathways are disrupted by both ulcerative colitis and colorectal cancer expression data?

# FILTER DATASET

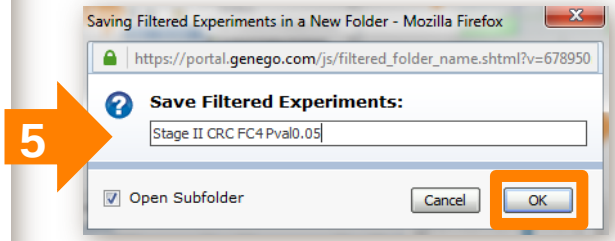



Set your filtering thresholds and click apply



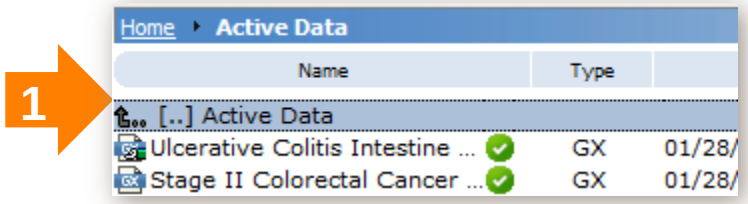

Examine the filtered results before saving the experiment

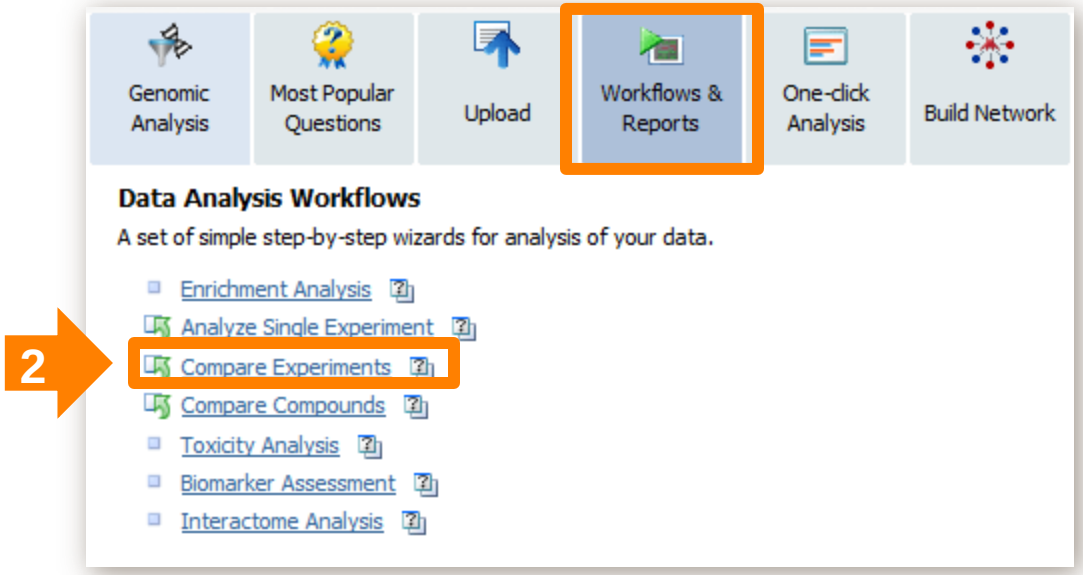



Choose a name for the folder and click ‘OK’

THOMSON REUTERS

# ACTIVATE DATASETS AND RUN COMPARE EXPERIMENTS WORKFLOW



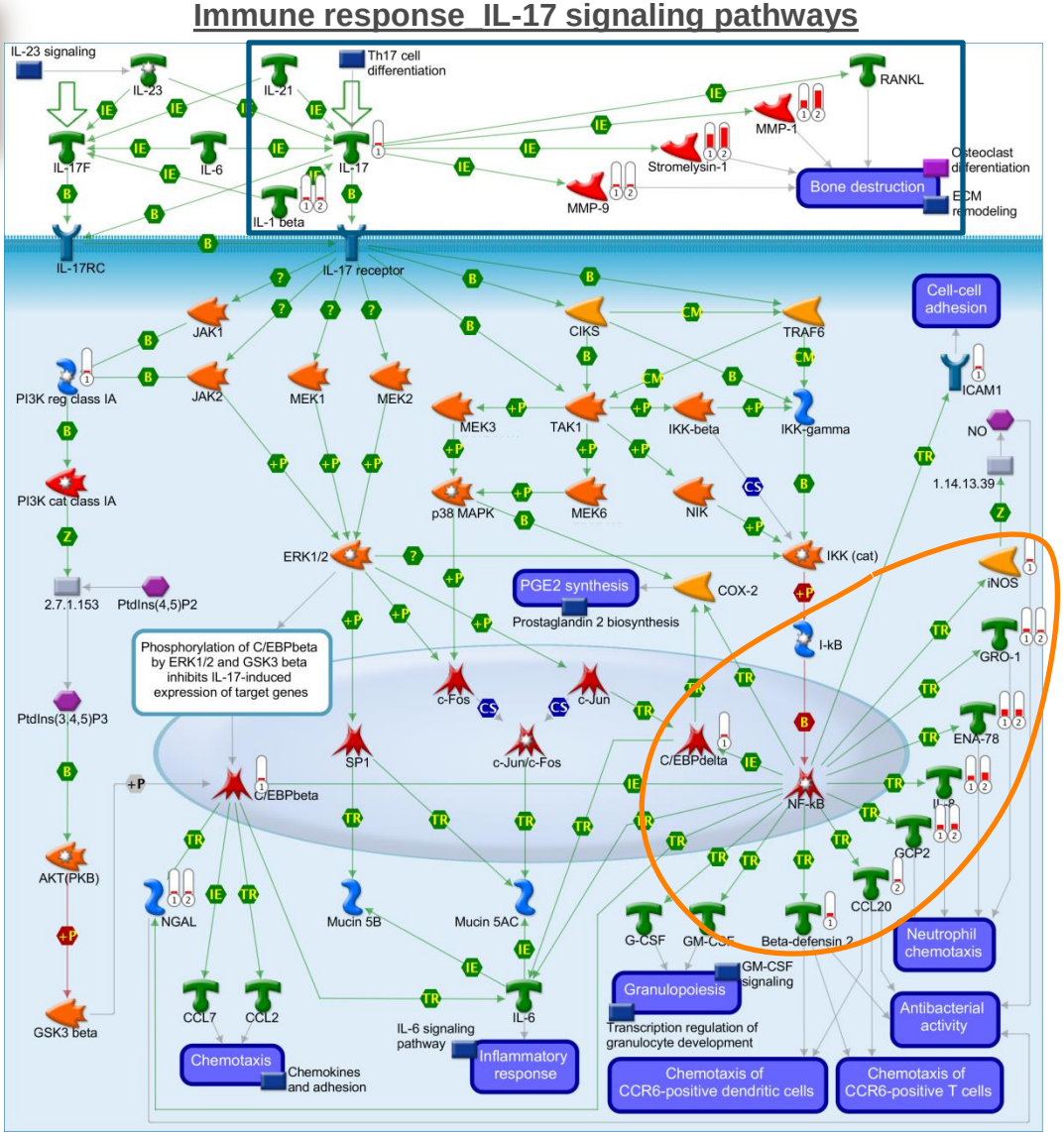
# WHAT PATHWAYS ARE DISRUPTED BY BOTH ULCERATIVE COLITIS AND COLORECTAL CANCER EXPRESSION DATA?

| Experiment name | Species | Network Objects |
|---|---|---|
| Ulcerative Colitis Intestine vs. Normal Intestine | Homo sapiens | 772 |
| Stage II Colorectal Cancer vs. Normal Adjacent Colon to Stage II Colorectal Cancer _FF | Homo sapiens | 721 |

Cell migration, adhesion, and chemotaxis are processes common to both gene signatures

Increased expression of many matrix metalloproteinases could be leading to the extracellular matrix remodeling in these diseases.

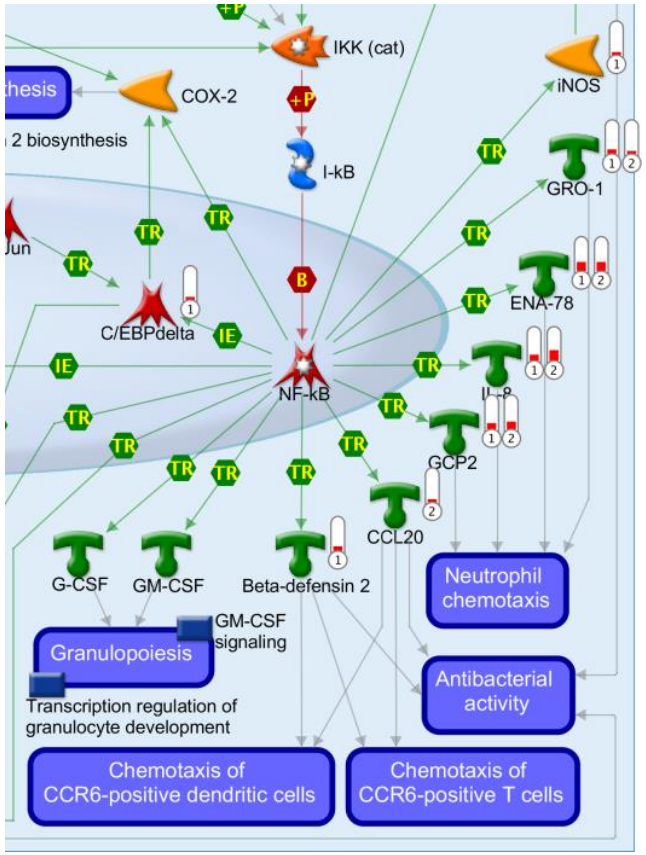
The common signature shares much in common with biomarkers from other gastrointestinal diseases.




THOMSON REUTERS

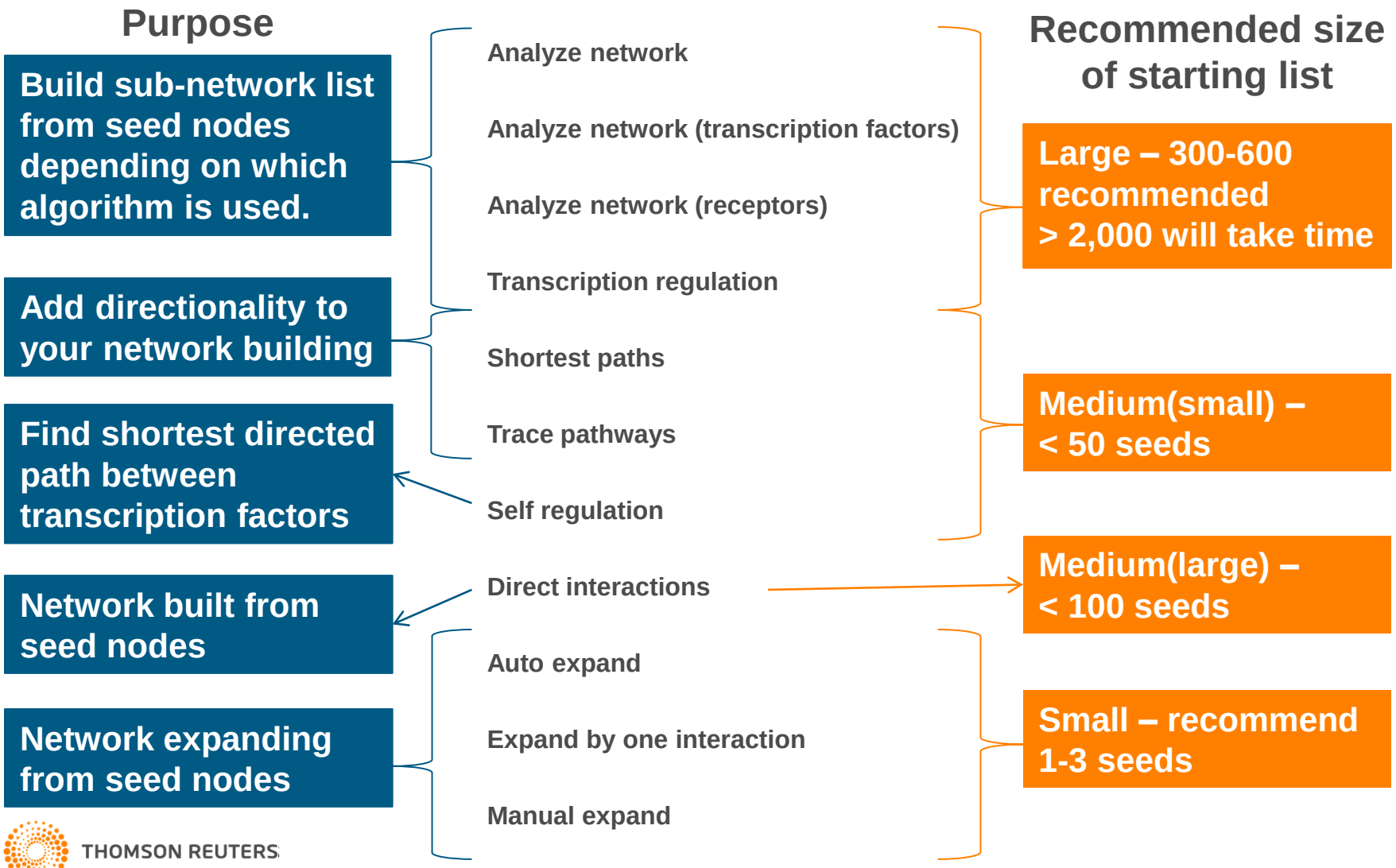
# SUMMARY

- The ulcerative colitis expression dataset was similar to other gastrointestinal diseases like Crohn's and colorectal cancer.
- The commonly expressed genes between UC and CRC point processes like cell migration and adhesion being impacted by both diseases.



# NETWORK BUILDING

# 11 NETWORK ALGORITHMS

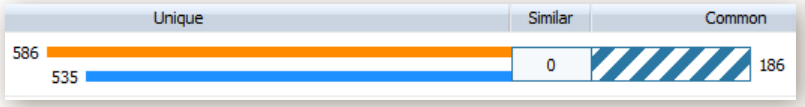


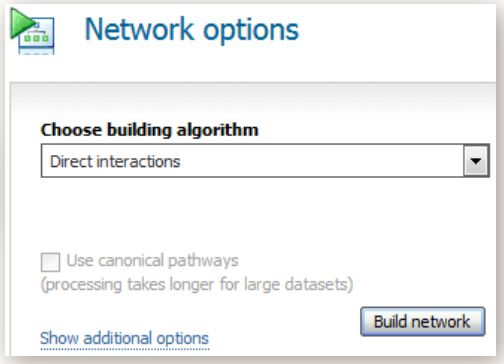

THOMSON REUTERS

# HOW TO USE THE DIRECT INTERACTIONS ALGORITHM TO BUILD A NETWORK USING COMPARE EXPERIMENTS RESULTS

Build network from Compare Experiment

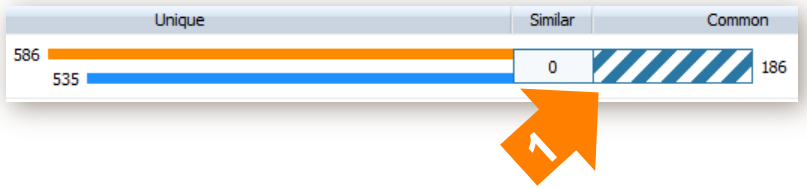
Set network building options



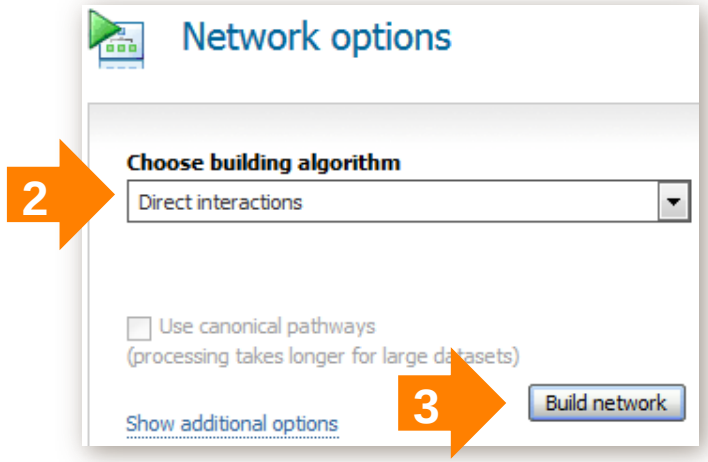

- Questions:
  - How might the differentially expressed genes from one of these signatures be directly interacting?

# BUILDING A NETWORK FROM COMPARE EXPERIMENTS RESULTS




Left click stripped area

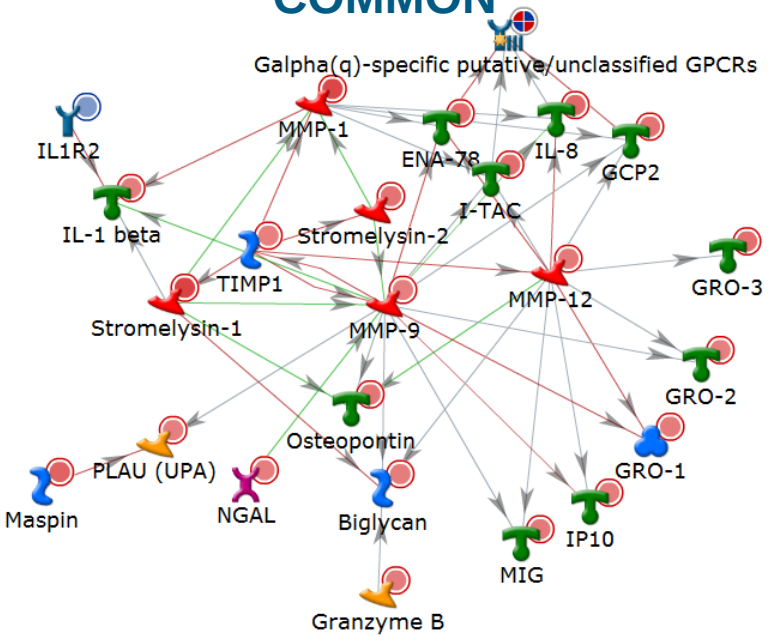
Select ‘Direct interactions’ algorithm




When finished click ‘Build network’

# WHAT DIFFERENTIALLY EXPRESSED GENES ARE DIRECTLY INTERACTING?

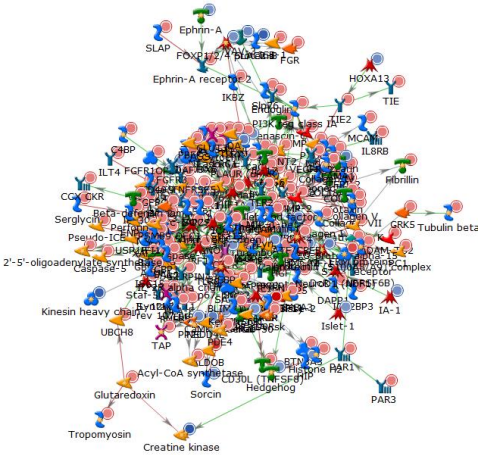
## COMMON




Relationship with cell migration and chemotaxis as well as a strong connection with Crohn's disease and IBD

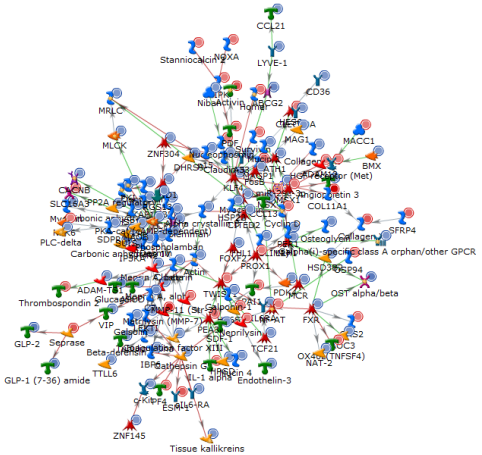
## UNIQUE UC

Relationship with immune response and autoimmune diseases



## UNIQUE CRC

Relationship with gastrointestinal diseases and cancers



# WORKING WITH MIRNA DATA

Multi-omics analysis with miRNA & mRNA data

# DATASET

Jernås *et al. BMC Immunology* 2013, **14**:32
http://www.biomedcentral.com/1471-2172/14/32

BMC Immunology

**RESEARCH ARTICLE** **Open Access**

## MicroRNA regulate immune pathways in T-cells in multiple sclerosis (MS)

Margareta Jernås[1*], Clas Malmeström[2], Markus Axelsson[2], Intawat Nookaew[5], Hans Wadenvik[3], Jan Lycke[2] and Bob Olsson[4]

- GSE43591 – MicroRNA regulate immune pathways in T-cells in multiple sclerosis (MS) mRNA.
  - Affy/HG-U133_Plus_2
  - Fold change: <= -1.3 or >= 1.3
  - Multiple Sclerosis T-Lymphocytes vs. Normal T-Lymphocytes

GEO Gene Expression Omnibus

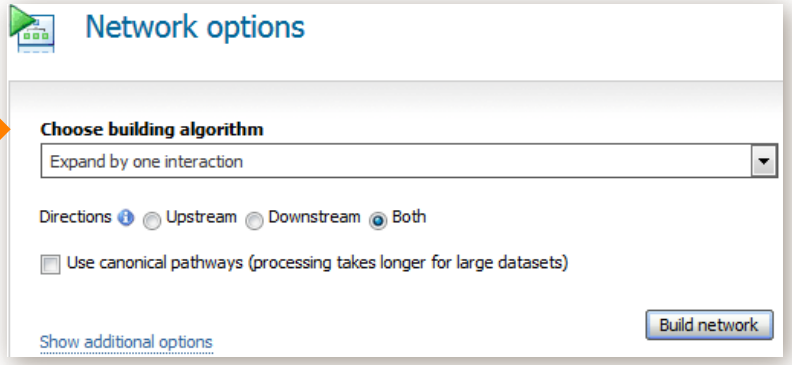
# UPLOADING MIRNA DATA

- Same procedure as gene expression, use Entrez ID or miRBase Accession ID (old style miRBase IDs are now under gene symbols)

# I WANT TO BUILD A NETWORK OF ALL THE MRNA MIR-494 COULD BIND WITH

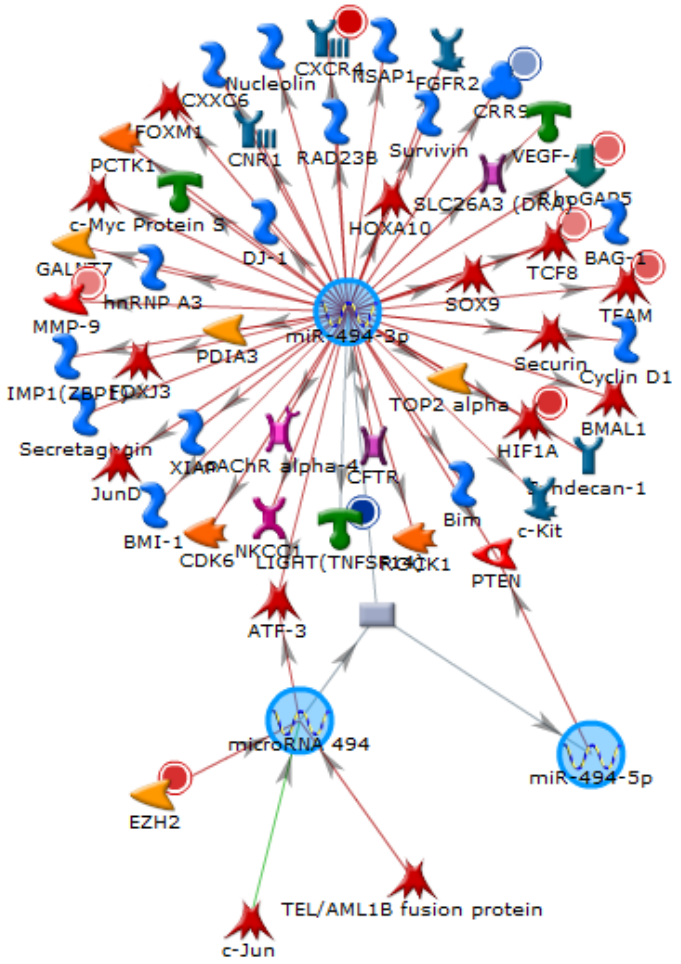
1

2




- Build a network using “expand by one interaction” algorithm for interactions with mir-494

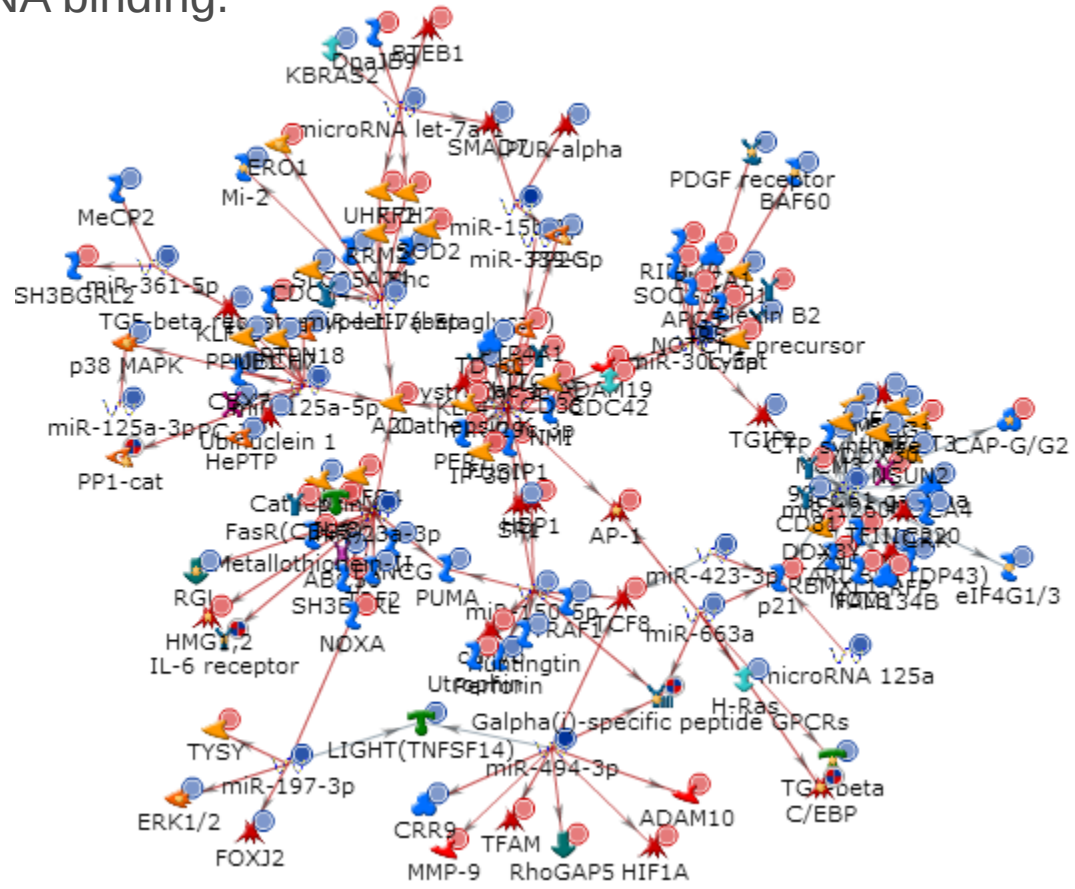
- 8 mRNA involved potentially binding with mir-494-3p have a change in expression including: CXCR4 and MMP-9



# WHAT IS THE RELATIONSHIP BETWEEN MIRNA AND MRNA EXPRESSION?

- Using the ‘direct interaction’ algorithm an interaction network between mRNA and miRNA expression can be built.
- Restrict interactions by using additional options to pre-filter the interactions for only miRNA binding.

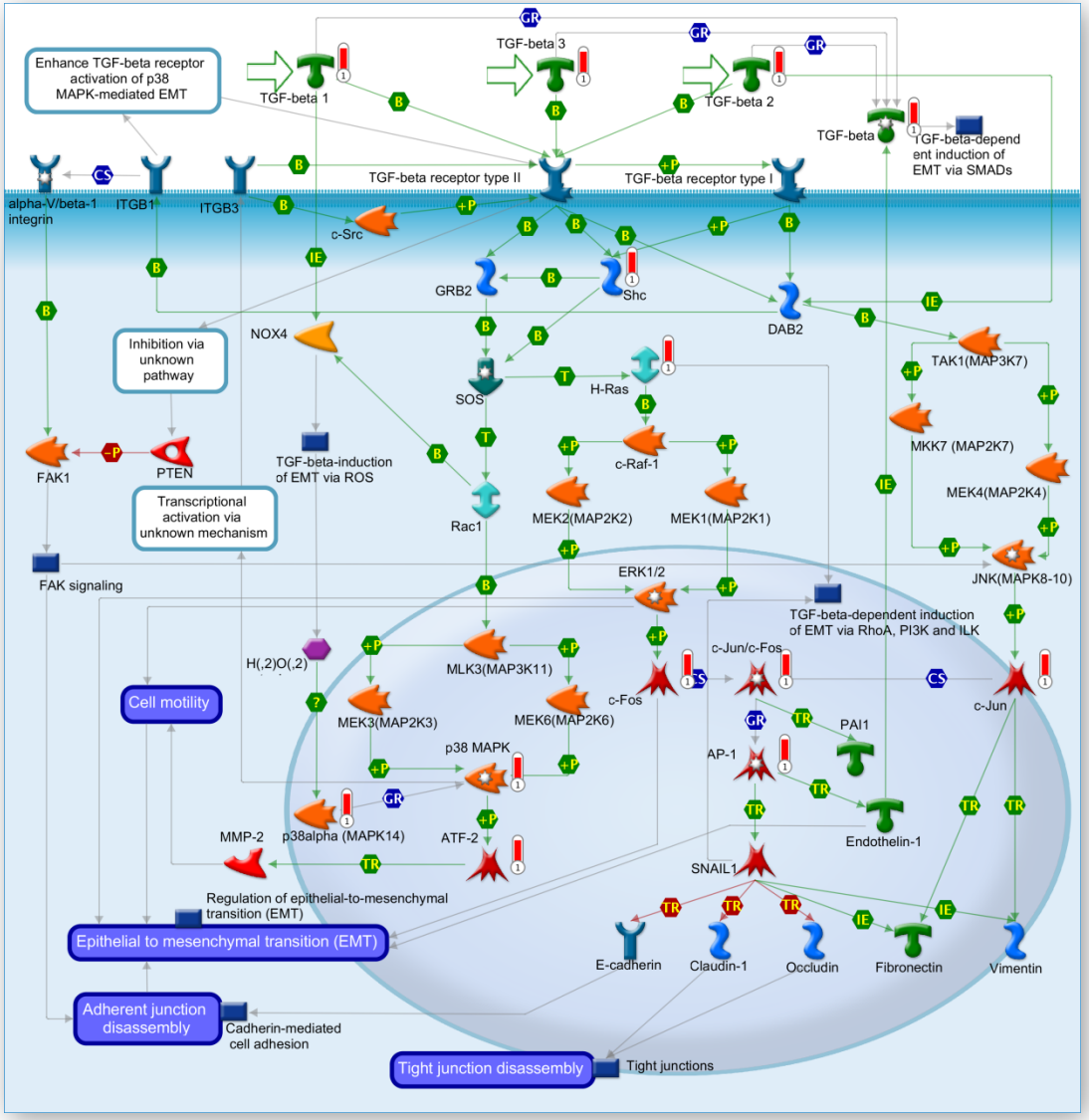
- This network identifies potential miRNA binding to their mRNA targets based on literature.
- exported as an experiment to run enrichments




THOMSON REUTERS

# WHAT PATHWAYS ARE IMPACTED BY MRNA FROM THE INTERACTION NETWORK?

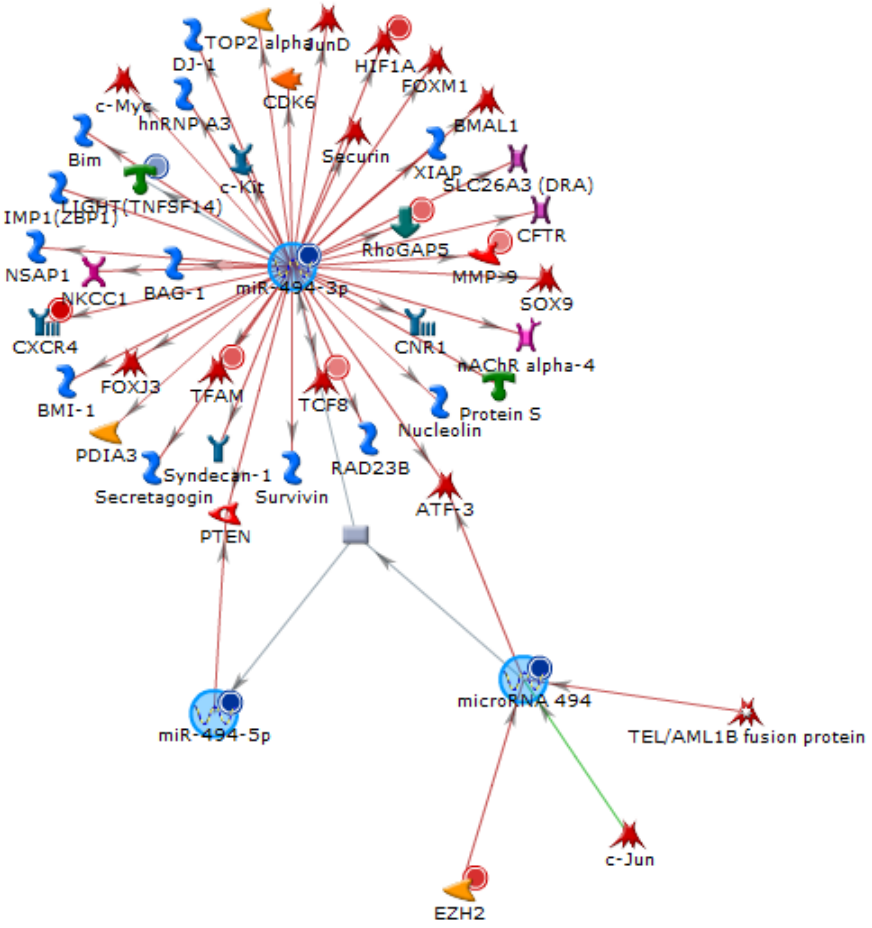
- mRNA from interaction network related to a number of immune response and development pathways.
- Interesting relationship between Multiple Sclerosis and mRNA interacting with miRNA.
- Could miRNA be impacting the expression of transcription factors related to AP-1 and it's target MMP-9 and IL-8?




THOMSON REUTERS

# SUMMARY

- mir-494 has known interactions with MMP-9 and CXCR4.
- Building an miRNA binding network between mRNA and miRNA reveals a large interconnected network.
- mRNA from this network can play a role on immune response and development.
- Expression of MMP-9, IL-8, CXCR4, and transcription factors related to AP-1 may have relations with miRNA.



# WORKING WITH METABOLITE DATA

## Multi-omics analysis with Metabolomics & mRNA data

# LIVE DEMO: DATASETS

**Gene Expression**

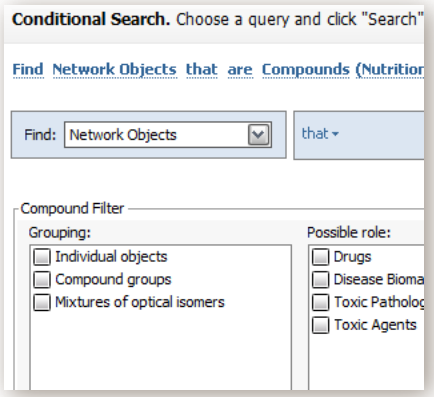
- GSE26886 – Gene expression profiling of Barrett's esophagus, adenocarcinoma, esophageal squamous epithelium and squamous cell carcinoma.
  - Affy/HG-U133_Plus_2
  - Fold change: ≤ -6 or ≥ 6 (P-value: ≤ 0.05)
  - Barrett Esophagus vs. Normal Esophagus
  - Esophageal Adenocarcinoma vs. Normal Esophagus

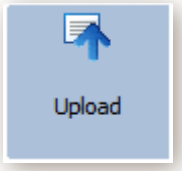
**Metabolite Data**

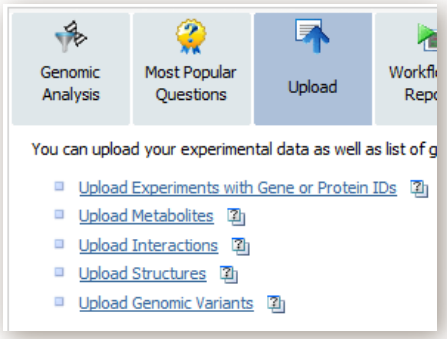
- Davis VW, Schiller DE, Eurich D, and Sawyer MB. (2012) Urinary metabolic signature of esophageal cancer and Barrett's esophagus. World Journal of Surgical Oncology. 10:271.
  - Validated urine biomarkers using two statistical methods.
  - Fold change: -1 decrease or 1 increase
  - Barrett Esophagus vs. Control
  - Esophageal Adenocarcinoma vs. Control

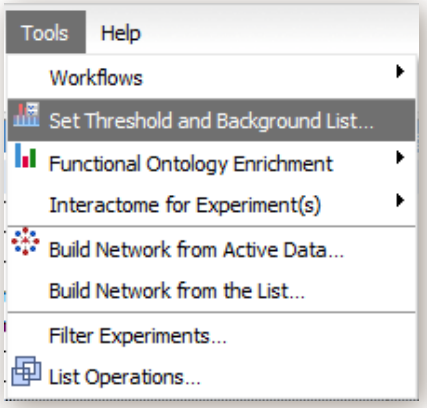
# CREATE A CUSTOM BACKGROUND LIST OF COMPOUNDS AND RUN ENRICHMENT ANALYSIS ON METABOLOMIC DATA

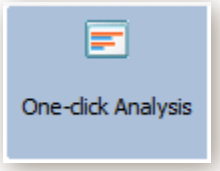
Create list of compounds → Upload metabolites in MetaCore → Set Threshold/ Background → Analyse Metabolic Networks

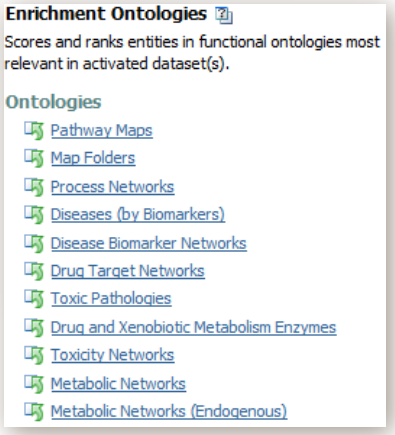


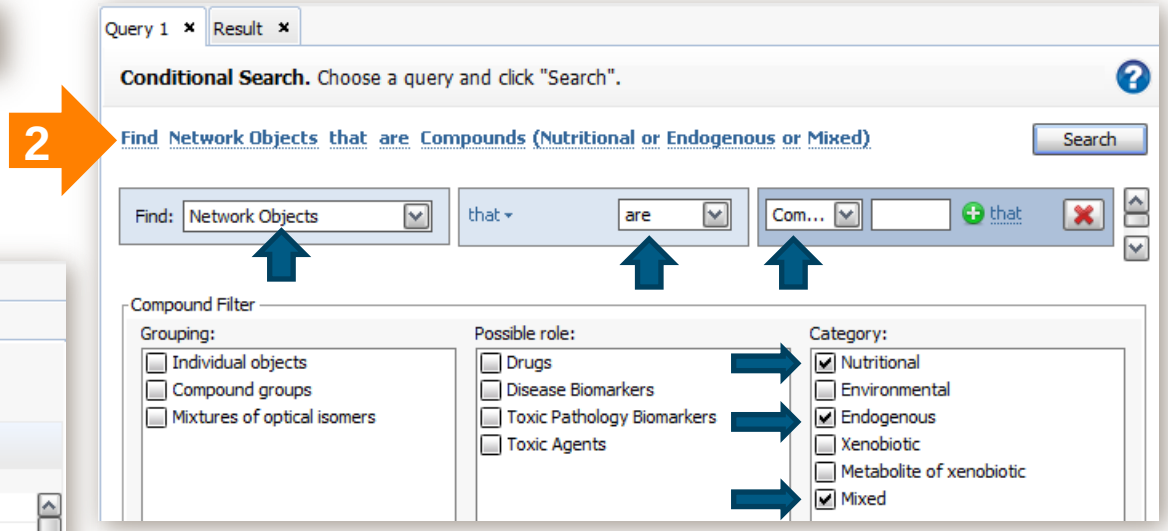



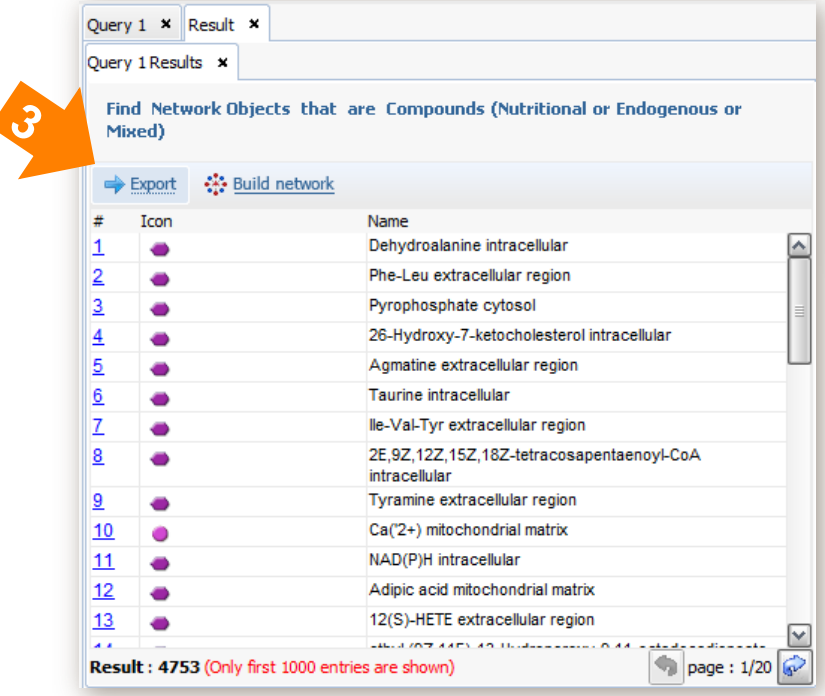



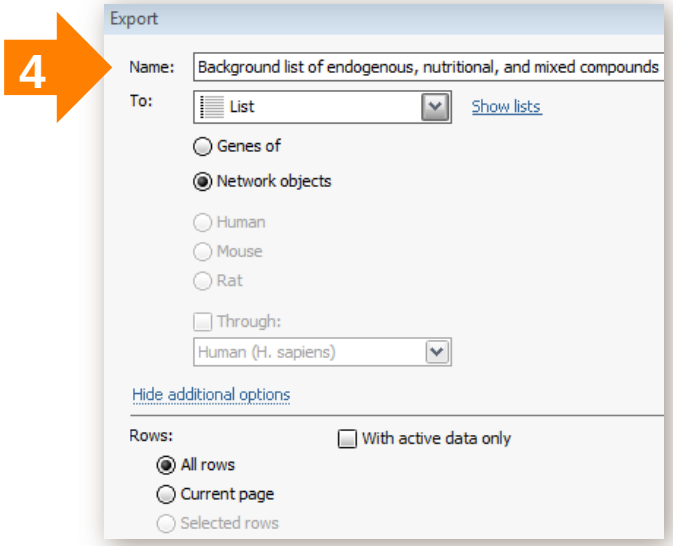

- Questions:
  - What are the differences between the metabolites from Barrett's Esophagus and Esophageal Adenocarcinoma?
  - Can overlaying transcriptomic data provide additional information on these differences?

THOMSON REUTERS

# USE ADVANCED SEARCH TO FIND AND SAVE A LIST OF COMPOUNDS

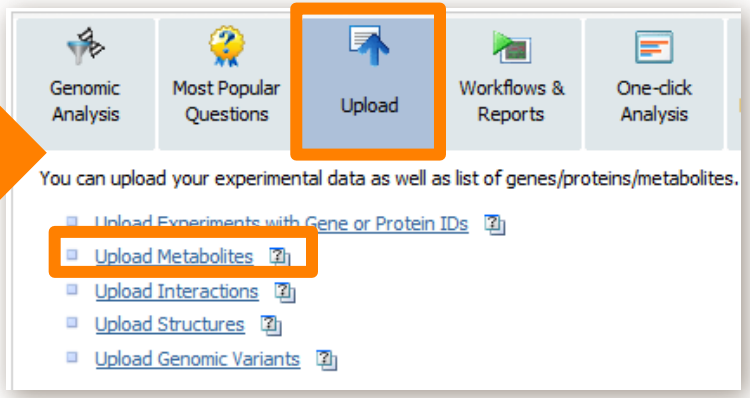


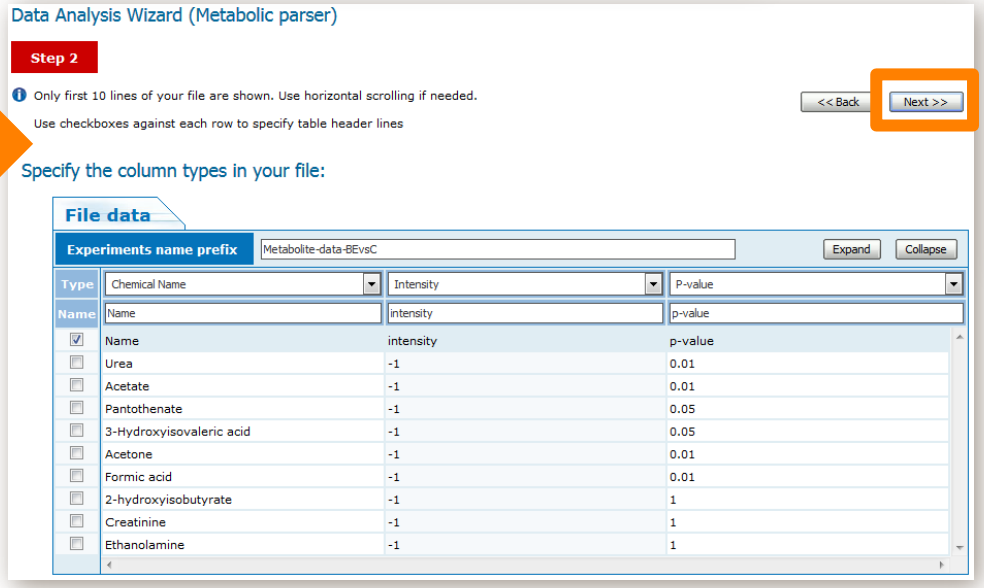




THOMSON REUTERS

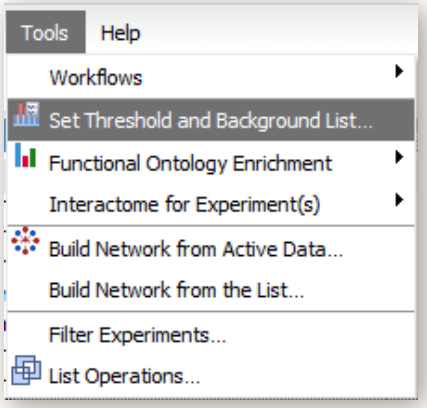
# UPLOAD METABOLOMIC DATA INTO METACORE



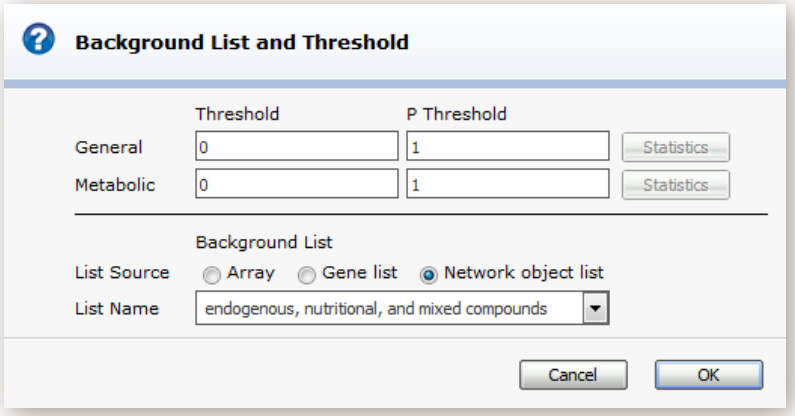


# SET THRESHOLD/BACKGROUND AND RUN METABOLOMIC NETWORK (ENDOGENOUS) ANALYSIS

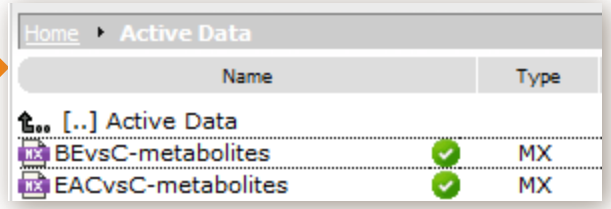
1



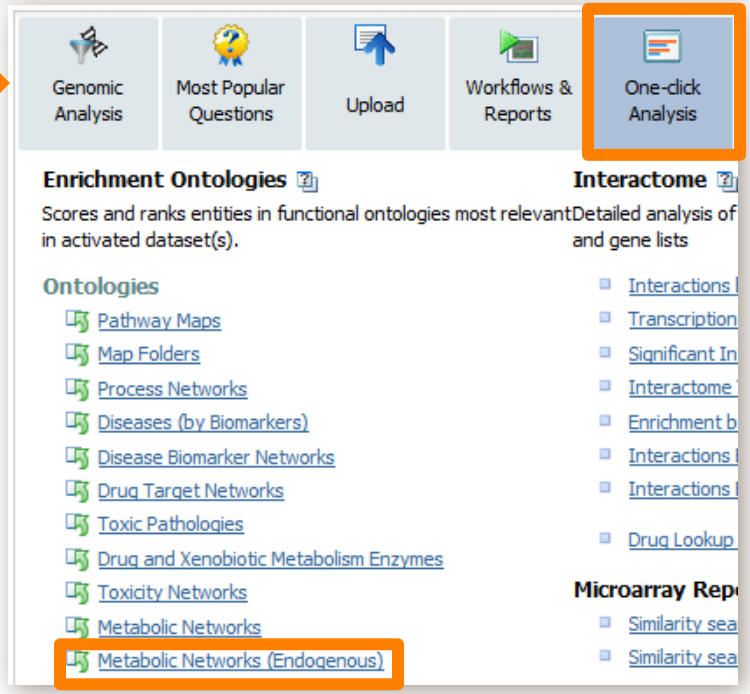

2




3

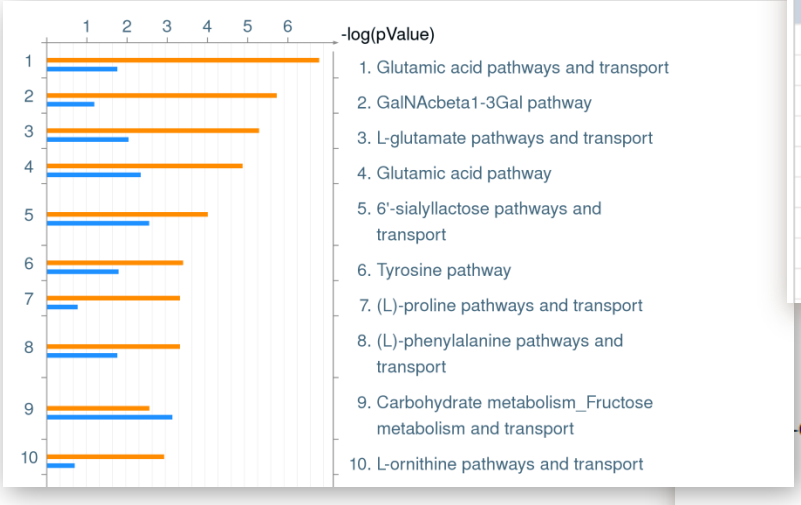



4

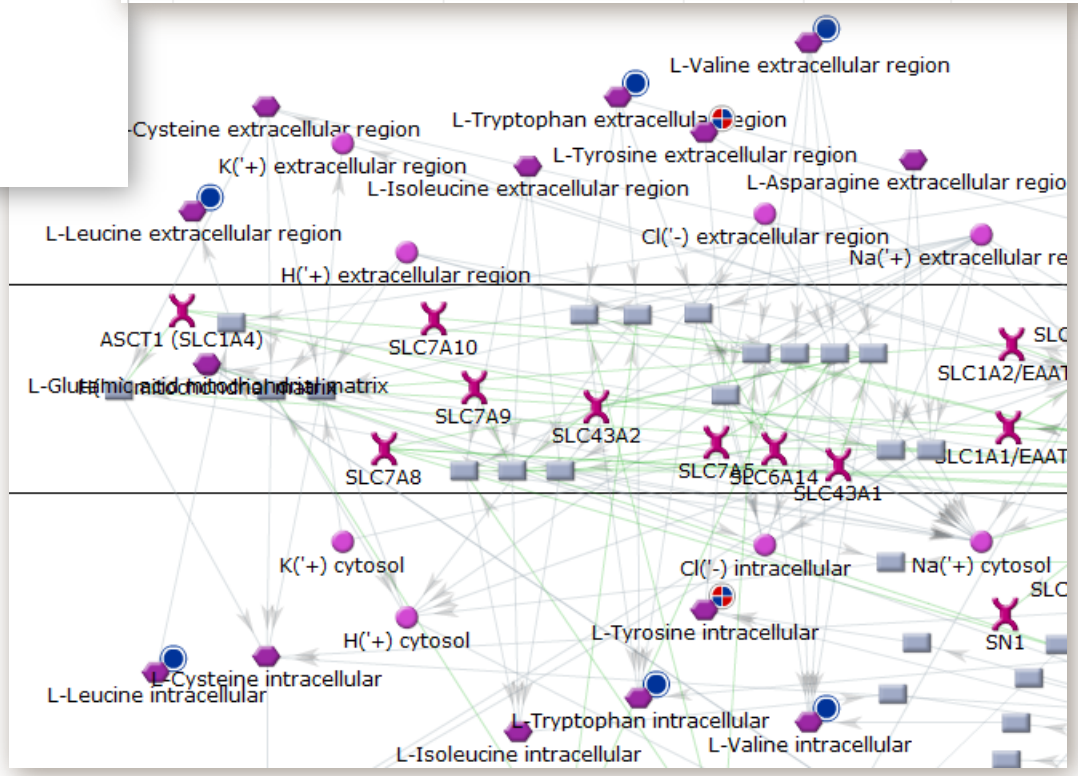



THOMSON REUTERS

# DIFFERENCES BETWEEN METABOLITES FROM BARRETT'S ESOPHAGUS AND ESOPHAGUL ADENOCARCINOMA

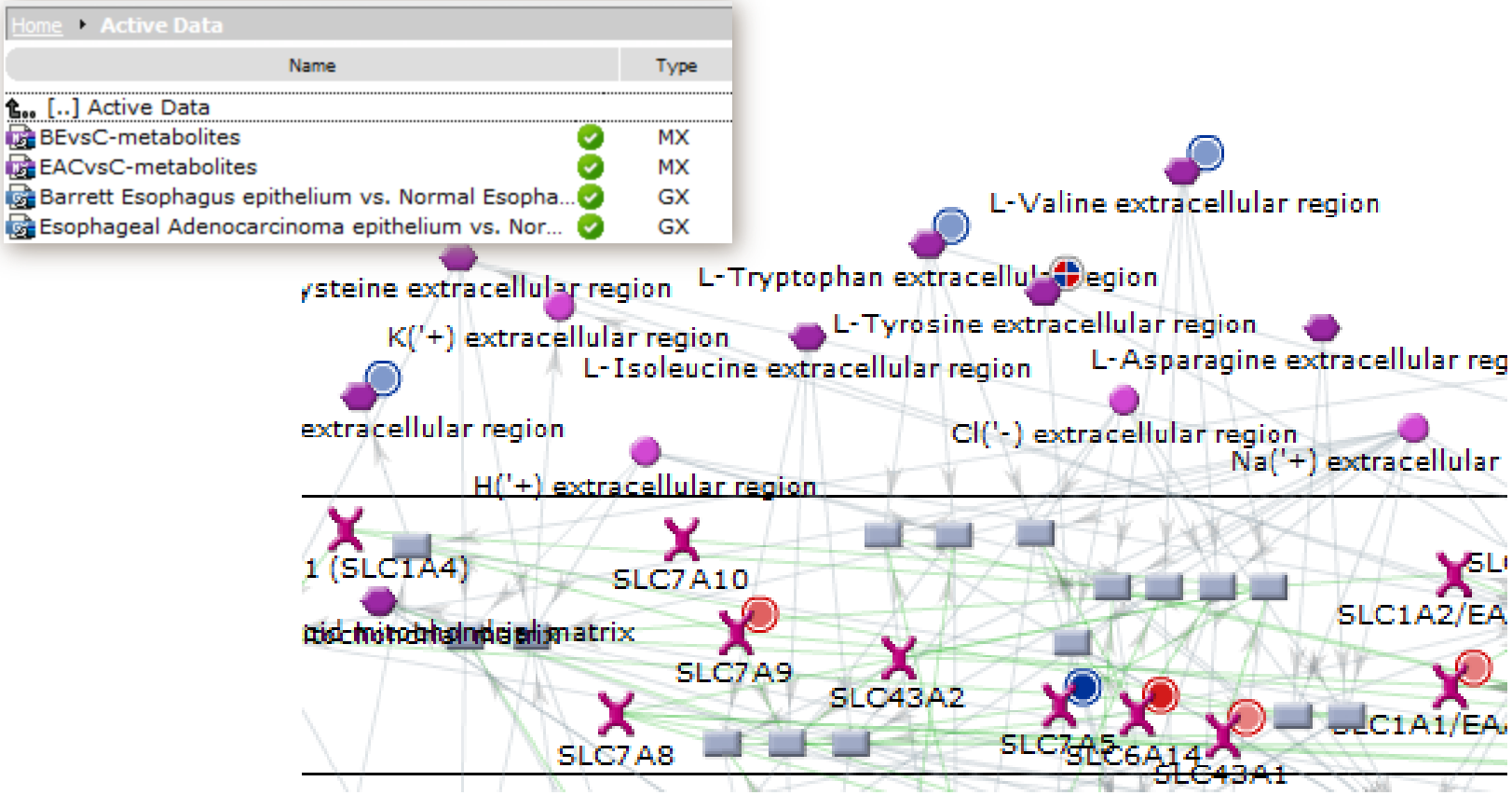



| Network Objects | # of Networks containing Network Object | Network Objects ranking | | |
|---|---|---|---|---|
| | | Overall | Metabolit... | Metabolit... |
| D-Glucose intracellular | 9 | 1 | 1 | |
| D-Glucose extracellular region | 9 | 1 | 1 | |
| L-Tryptophan intracellular | 8 | 2 | 2 | |
| L-Tryptophan extracellular region | 8 | 2 | 2 | |
| Urea intracellular | 8 | 2 | 2 | 1 |
| L-Tyrosine intracellular | 7 | 3 | 3 | 2 |
| Urea extracellular region | 6 | 4 | 4 | 3 |
| L-Tyrosine extracellular region | 5 | 5 | 5 | 4 |
| D-Sucrose extracellular region | 4 | 6 | | 5 |




Enrichment analysis of Metabolomic Networks (Endogenous) can quickly identify metabolites impact on processes.

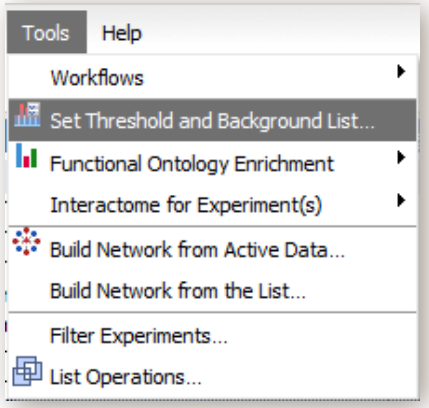
# OVERLAYING EXPRESSION AND METABOLIC DATA ONTO METABOLIC PROCESSES

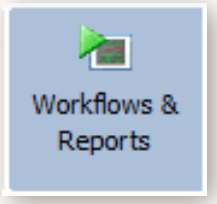


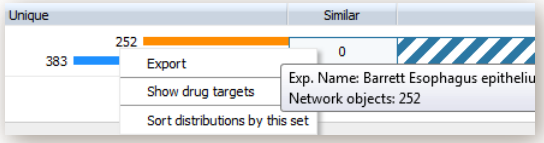

A number of channels are differentially expressed specifically in Barrett esophagus, but not esophageal adenocarcinoma.

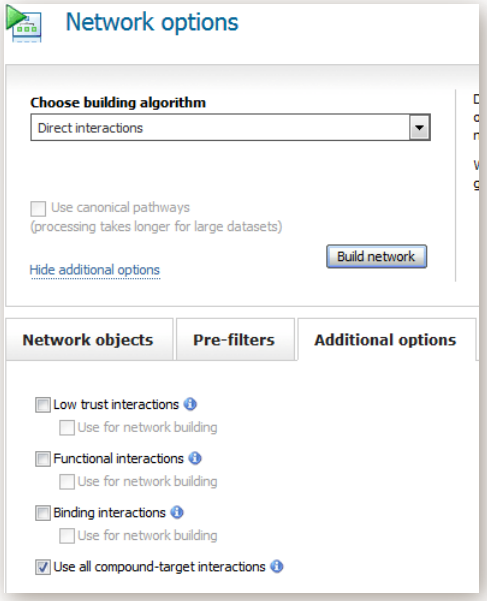
# BUILD A NETWORK USING METABOLOMIC AND TRANSCRIPTOMIC DATA

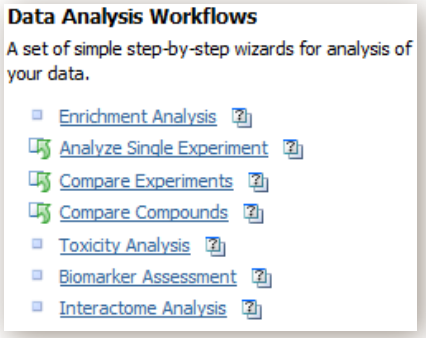
Set Threshold/ Background → Compare Experiments Workflow → Export unique network objects → Build Direct Interaction Network



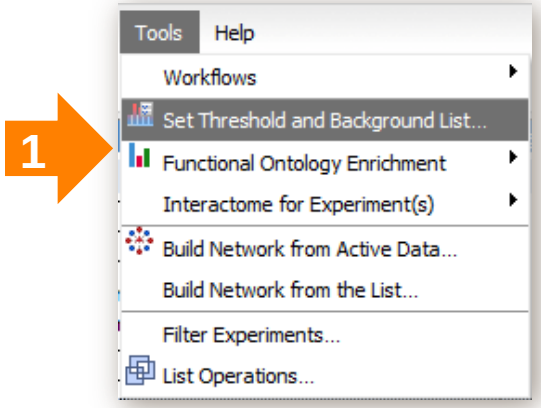



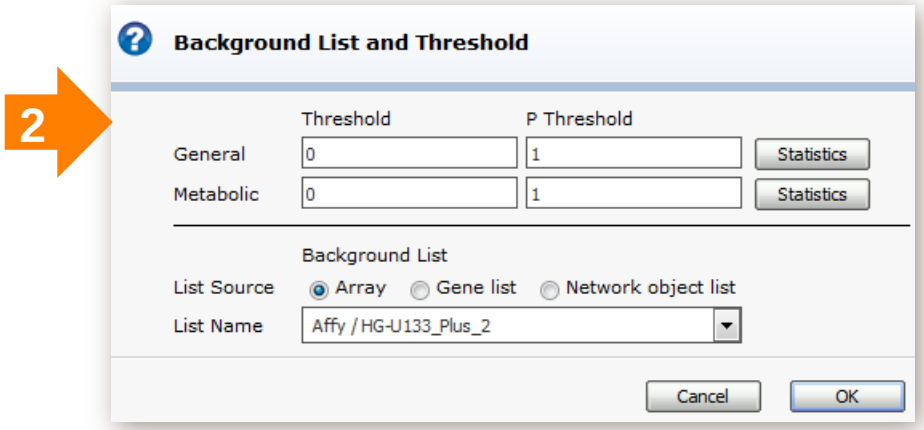



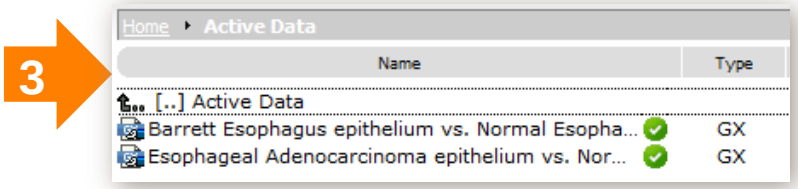



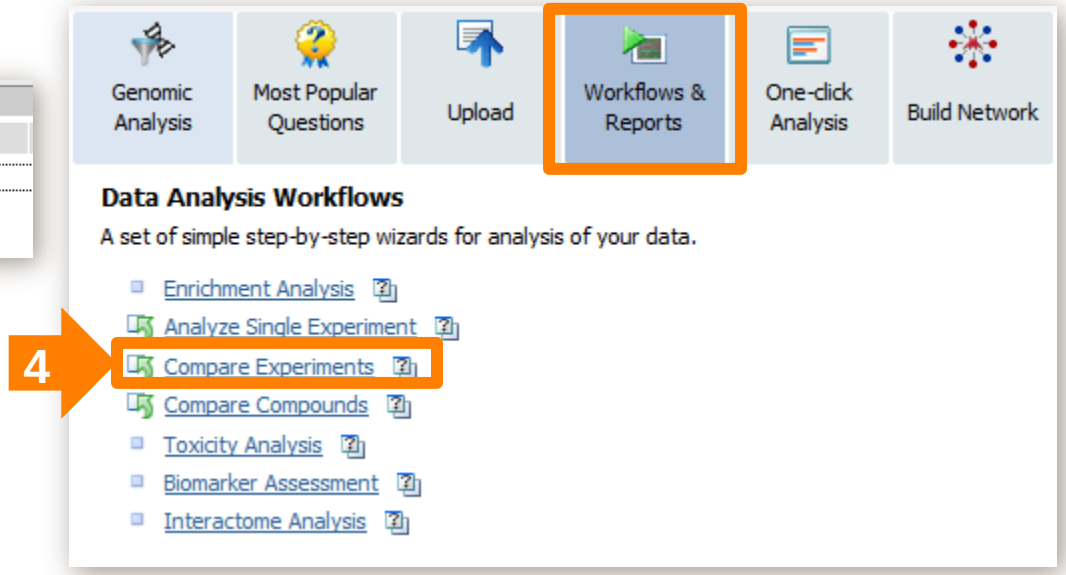
- Questions:
  - How are the differentially expressed genes common or unique to these two diseases related to the metabolomic data?

# SET THRESHOLD/BACKGROUND AND RUN COMPARE EXPERIMENTS WORKFLOW



# CAN DIFFERENCES IN GENE EXPRESSION HELP EXPLAIN THE METABOLITES?

| Experiment name | Species | Network Objects |
|---|---|---|
| Barrett Esophagus epithelium vs. Normal Esophagus epithelium_FF | Homo sapiens | 933 |
| Esophageal Adenocarcinoma epithelium vs. Normal Esophagus epithelium_FF | Homo sapiens | 1064 |

Common gene expression related to remodeling and stomach/gastrointestinal diseases

**Pathway Maps**

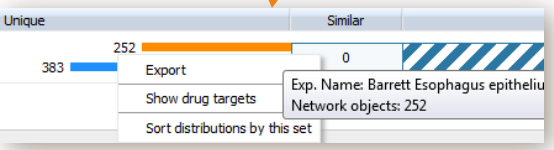
Export | Export to image | Sorting method: Particular maps | Total results: 10

| # | Maps | -log(pValue) | pValue | pValue ↑ | FDR | Rat |
|---|---|---|---|---|---|---|
| 1 | Cytoskeleton remodeling_Keratin filaments | | 1.000e+0 | 1.122e-10 | 1.000e+0 | 0/36 |
| | | | 1.122e-10 | | 9.978e-8 | 14/36 |
| | | | 1.000e+0 | | 1.000e+0 | 0/36 |
| | | | 7.043e-2 | | 6.043e-1 | 3/36 |

**Diseases (by Biomarkers)**

Export | Export to image | Sorting method: Particular diseases | Total results: 10

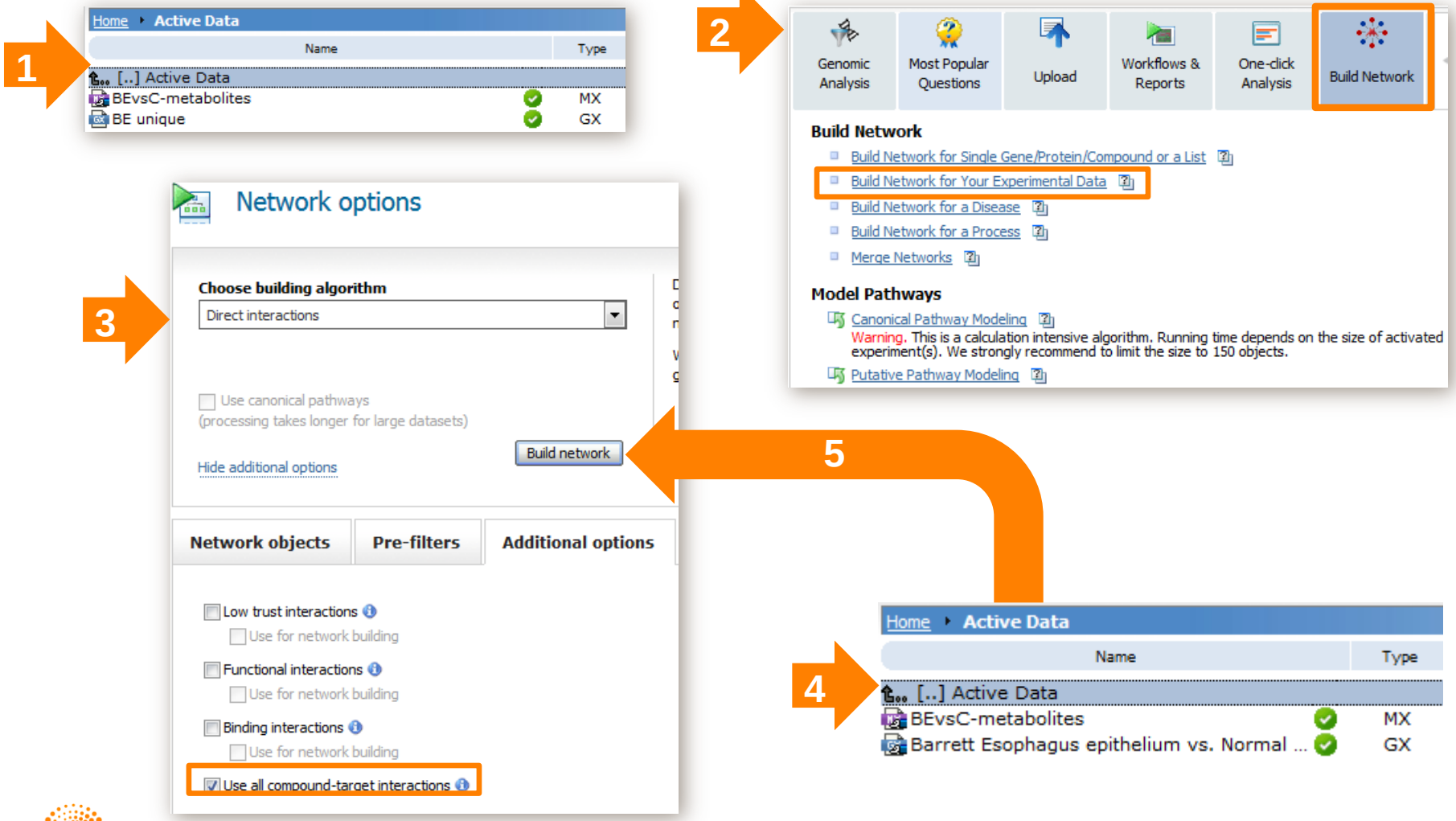
| # | Diseases | -log(pValue) | pValue | pValue ↑ | FDR | Rat |
|---|---|---|---|---|---|---|
| 1 | Stomach Diseases | | 1.000e+0 | 1.720e-21 | 1.000e+0 | 0/3358 |
| | | | 1.720e-21 | | 3.061e-18 | 221/3358 |
| | | | 7.911e-9 | | 2.832e-6 | 83/3358 |
| | | | 2.977e-14 | | 2.926e-12 | 128/3358 |
| 2 | Stomach Neoplasms | | 1.000e+0 | 8.470e-21 | 1.000e+0 | 0/3311 |
| | | | 8.470e-21 | | 7.539e-18 | 217/3311 |
| | | | 4.020e-9 | | 2.159e-6 | 83/3311 |
| | | | 9.978e-15 | | 1.159e-12 | 128/3311 |



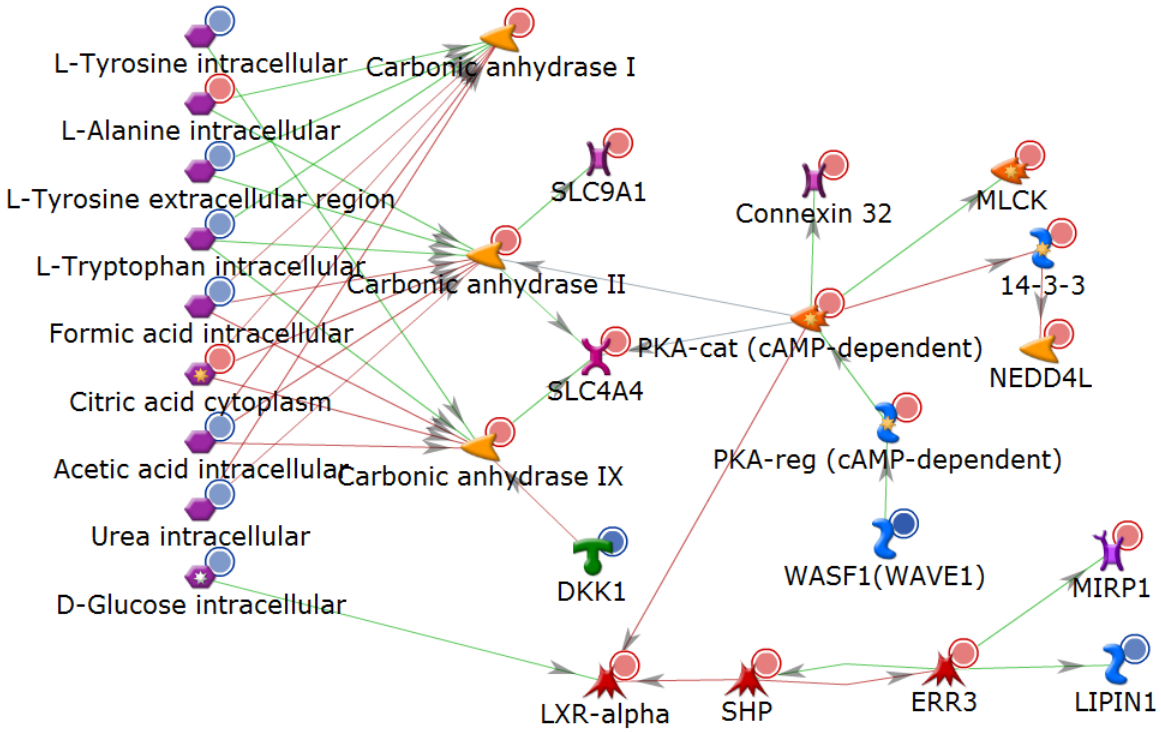

Export the unique gene expression for Barrett's Esophagus as a new experiment to build a network with it

THOMSON REUTERS

# BUILDING A NETWORK USING METABOLOMIC AND TRANSCRITOMIC DATA



# NETWORK OF UNIQUE GENE EXPRESSION AND METABOLITES FROM BARRETT'S ESOPGAHUS




## Diseases

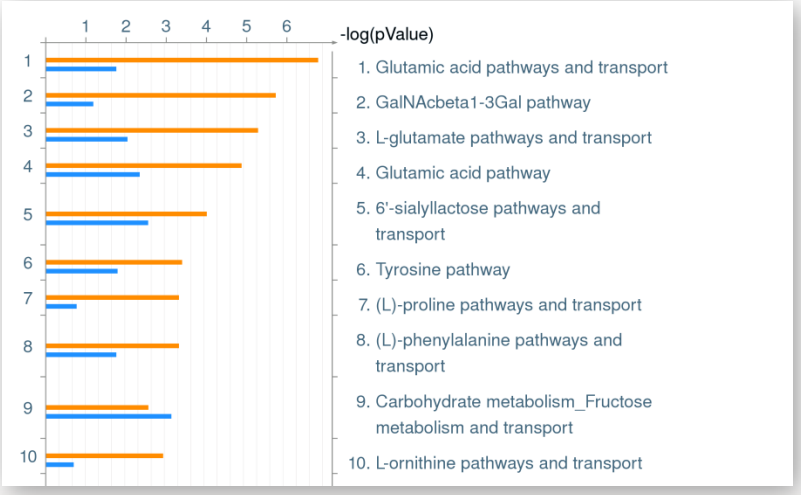
| Name | % | p-Value |
|---|---|---|
| Acidosis | 25.00% | 2.887e-7 |
| Acid-Base Imbalance | 25.00% | 3.852e-7 |
| Acidosis, Renal Tubular | 12.50% | 2.181e-5 |
| Pathological Conditions, Signs and ... | 68.75% | 2.530e-5 |
| Metabolic Diseases | 56.25% | 3.550e-5 |
| Bone Resorption | 18.75% | 6.484e-5 |
| Digestive System Diseases | 100.00% | 7.098e-5 |
| Nutritional and Metabolic Diseases | 56.25% | 8.921e-5 |

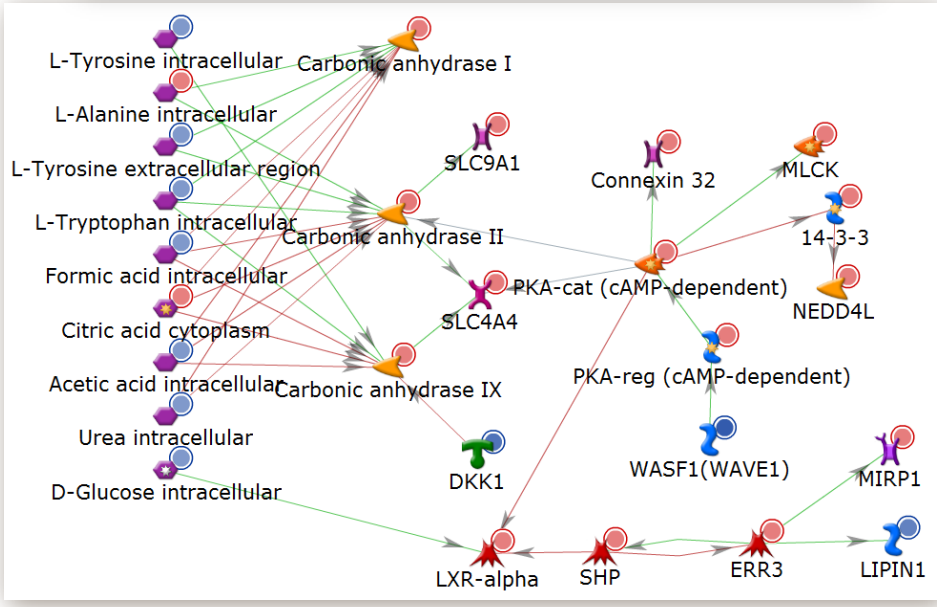
## GO Processes

| Name | % | p-Value |
|---|---|---|
| bicarbonate transport | 25.00% | 1.252e-8 |
| one-carbon metabolic process | 18.75% | 2.269e-6 |
| regulation of transcription from RN... | 56.25% | 7.966e-6 |
| ion transport | 50.00% | 8.808e-6 |
| regulation of intracellular pH | 18.75% | 1.425e-5 |
| regulation of cellular pH | 18.75% | 1.711e-5 |
| organic anion transport | 31.25% | 2.300e-5 |
| regulation of pH | 18.75% | 4.241e-5 |

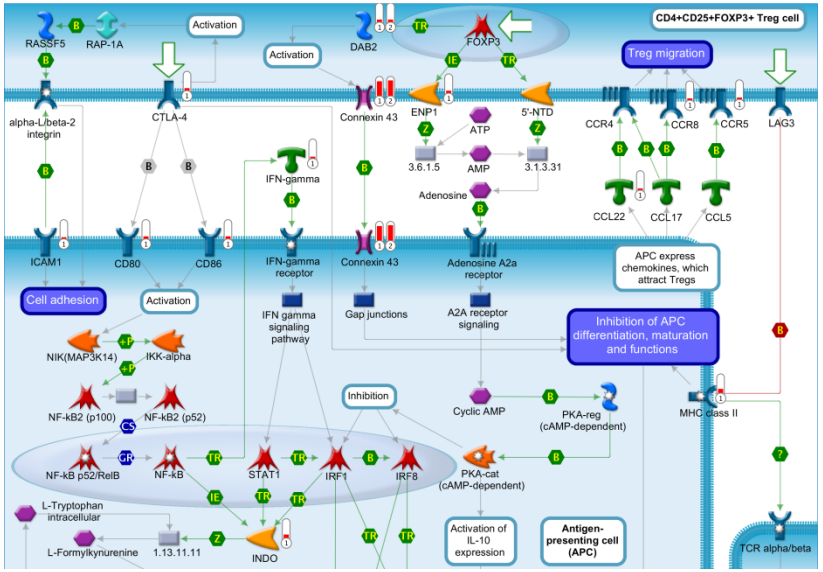
# SUMMARY

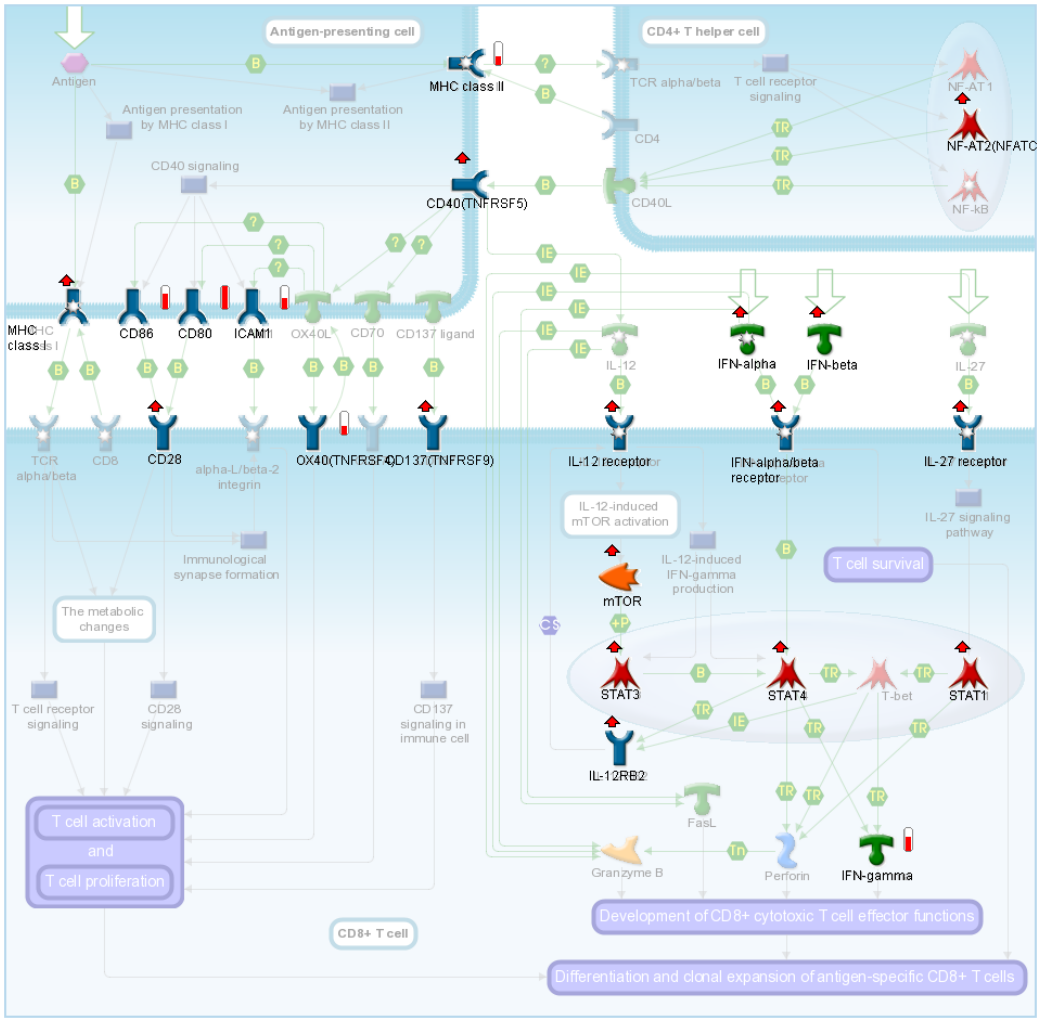
- Barrett's esophagus metabolites show significant changes in glutamic acid pathways.
- Overlaying expression shows expression changes of solute carriers specifically for Barrett's Esophagus.
- Some genes uniquely expressed by Barrett's esophagus are key in regulating acid/base balance and might be modulated by the metabolites.



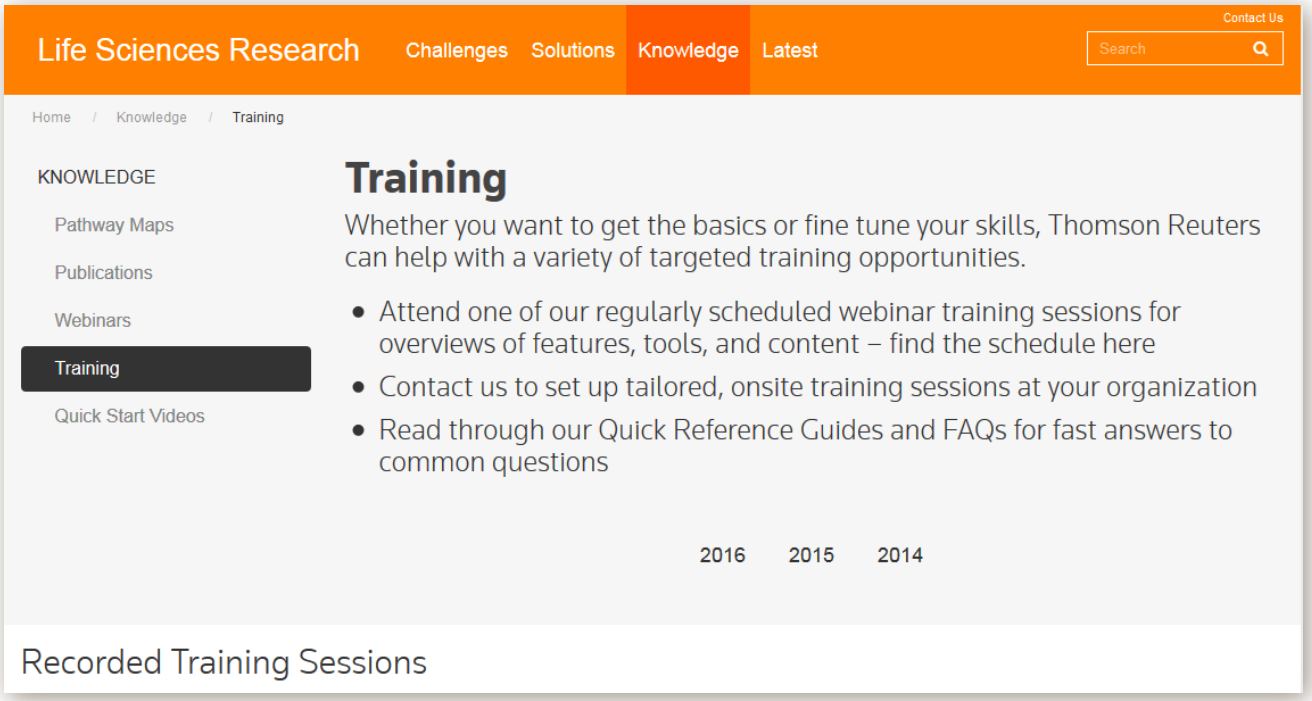
# ANY QUESTIONS?




THOMSON REUTERS

Contact:
matthew.wampole@thomsonreuters.com

# METACORE LIVE WEBINARS

- Follow Matthew Wampole on LinkedIn for information on the next Webinar LinkedIn profile
- You can find recordings of previously held sessions at http://lsresearch.thomsonreuters.com/knowledge/training-videos