WORKFLOWS TO FOLLOW UP ON RNA-SEQ ANALYSIS

HELENE R. MCMURRAY, PH.D.

EDWARD G. MINER LIBRARY

7 NOVEMBER 2016

TOPICS TO BE COVERED TODAY

- High-throughput sequencing basics
- Pattern recognition by hierarchical clustering
- Gene Ontology mapping
- Canonical pathway mapping

NEXT-GENERATION GAP

JOHN D. MCPHERSON
NATURE METHODS **6**, S2 - S5 (2009)

http://www.urmc.rochester.edu/libraries/miner/mdl.aspx?U=http://dx.doi.org/doi:10.1038/nmeth.f.268

- Discusses high-throughput sequencing platforms with references to more information on each technology
  - Roche 454
  - Illumina GAIIx
  - Applied Biosystems SOLID
  - Helicos HeliScope

SENSE FROM SEQUENCE READS: METHODS FOR ALIGNMENT AND ASSEMBLY

PAUL FLICEK & EWAN BIRNEY

NATURE METHODS **6**, S6 - S12 (2009)

http://www.urmc.rochester.edu/libraries/miner/mdl.aspx?U=http://dx.doi.org/doi:10.1038/nmeth.1376

- Discusses alignment methods
  - Hash-based methods
  - Burrows-Wheeler transform methods
- Explains assembly based on those alignments

COMPUTATION FOR CHIP-SEQ AND RNA-SEQ STUDIES

SHIRLEY PEPKE, BARBARA WOLD & ALI MORTAZAVI

NATURE METHODS **6**, S22 - S32 (2009)

http://www.urmc.rochester.edu/libraries/miner/mdl.aspx?U=http://dx.doi.org/doi:10.1038/nmeth.1371

- Provides overview of ChIP-seq and RNA-seq analyses
- ChIP-seq:
  - Classes of ChIP-seq signals
  - Peak-finders, regions, summits and sources
  - Publicly available ChIP-seq software
- RNA-seq:
  - Approaches to handle spliced reads
  - Quantifying gene expression
  - Publicly available RNA-seq software

VISUALIZATION OF OMICS DATA FOR SYSTEMS BIOLOGY

N. GEHLENBORG, S.I. O'DONOGHUE, N.S. BALIGA, A. GOESMANN, M.A. HIBBS, H. KITANO, O. KOHLBACHER, H. NEUWEGER, R. SCHNEIDER, D. TENENBAUM & A. GAVIN

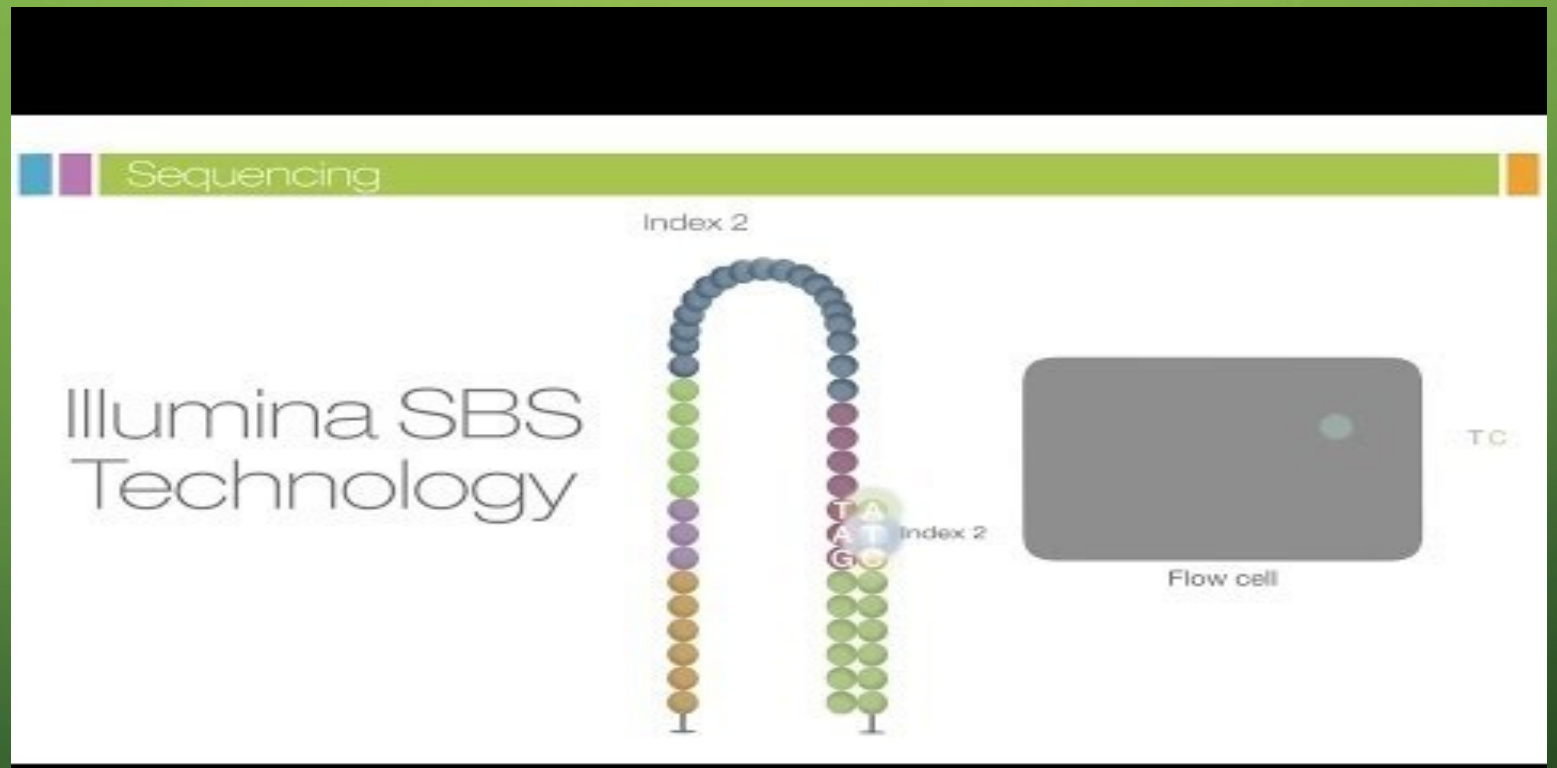
NATURE METHODS **7**, S56 - S68 (2010)

http://www.urmc.rochester.edu/libraries/miner/mdl.aspx?U=http://dx.doi.org/doi:10.1038/nmeth.1436

- Visualization based on networks vs. pathways
  - Protein interaction networks
  - Spatial information
  - Expression profile data
  - Multivariate –omics analysis
  - Metabolic network visualization
  - Pathway editing

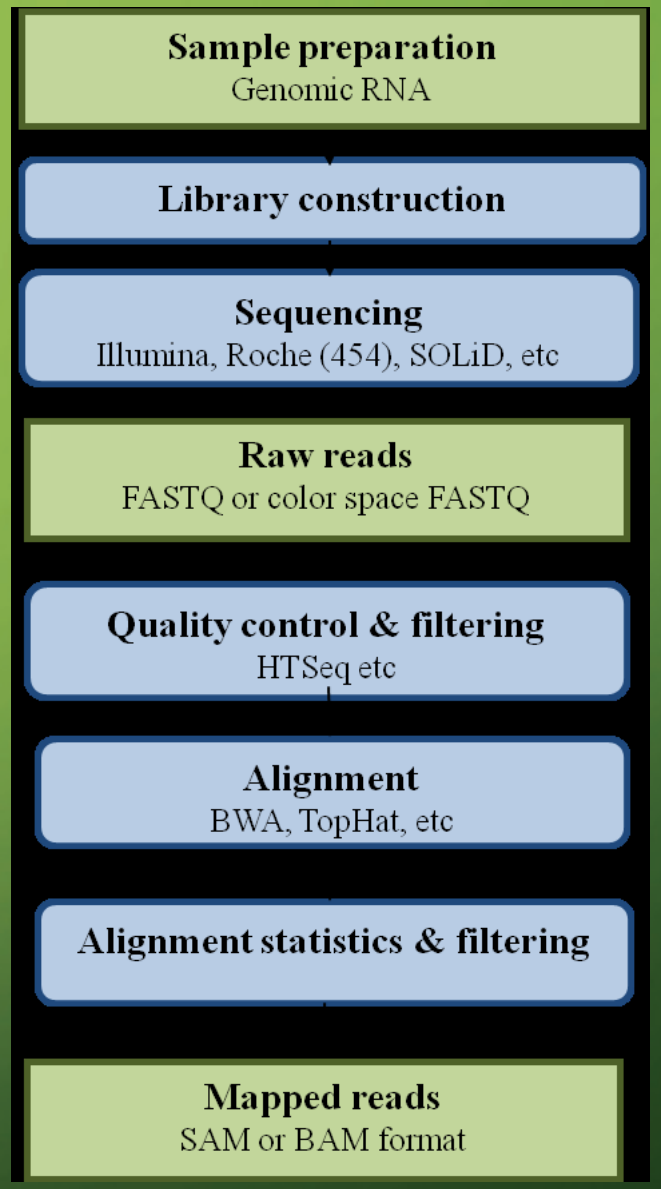
HIGH-THROUGHPUT SEQUENCING WORKFLOW

HOW DOES ILLUMINA SEQUENCING BY SYNTHESIS WORK?

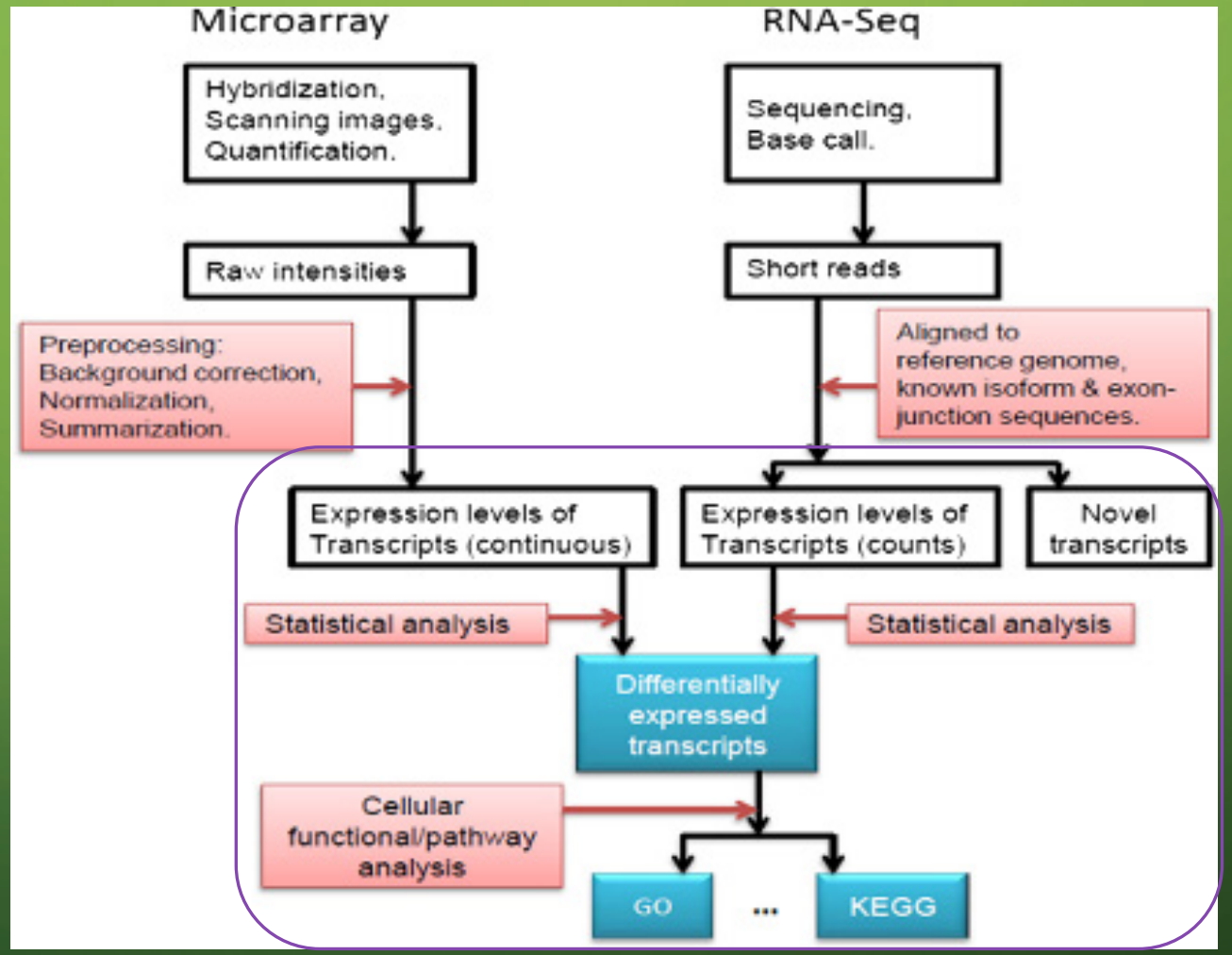


BASIC HIGH-THROUGHPUT SEQUENCING WORKFLOW: UPSTREAM

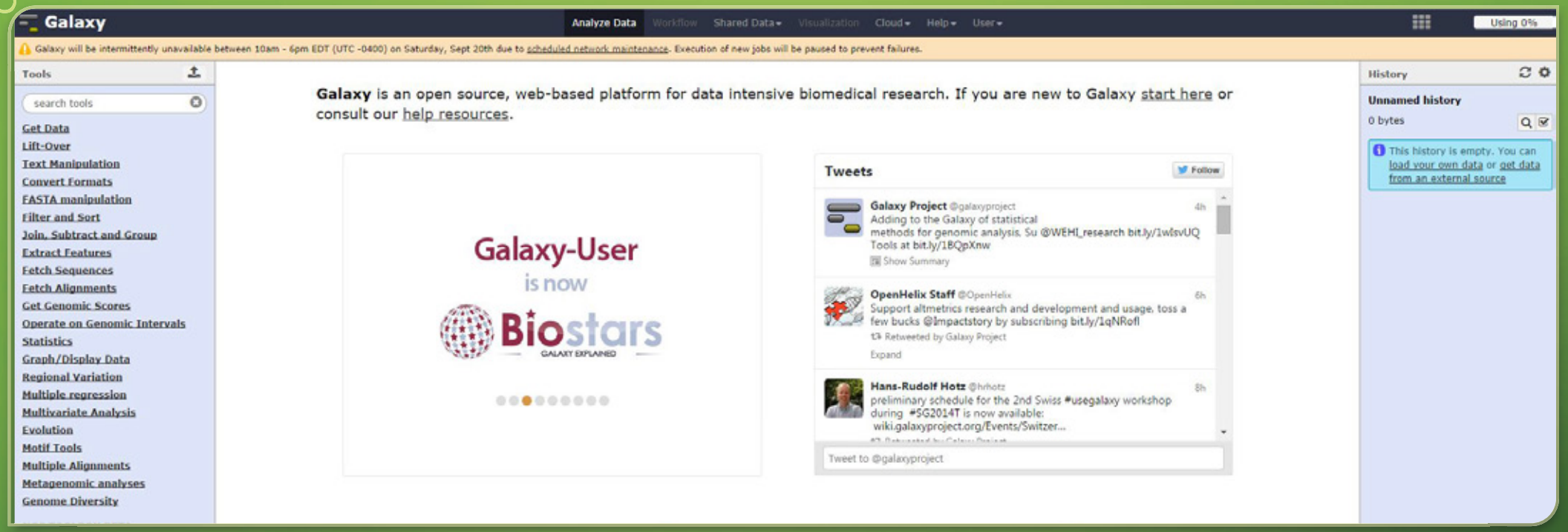
http://en.wikipedia.org/wiki/List_of_sequence_alignment_software



BASIC HIGH-THROUGHPUT SEQUENCING WORKFLOW: DOWNSTREAM

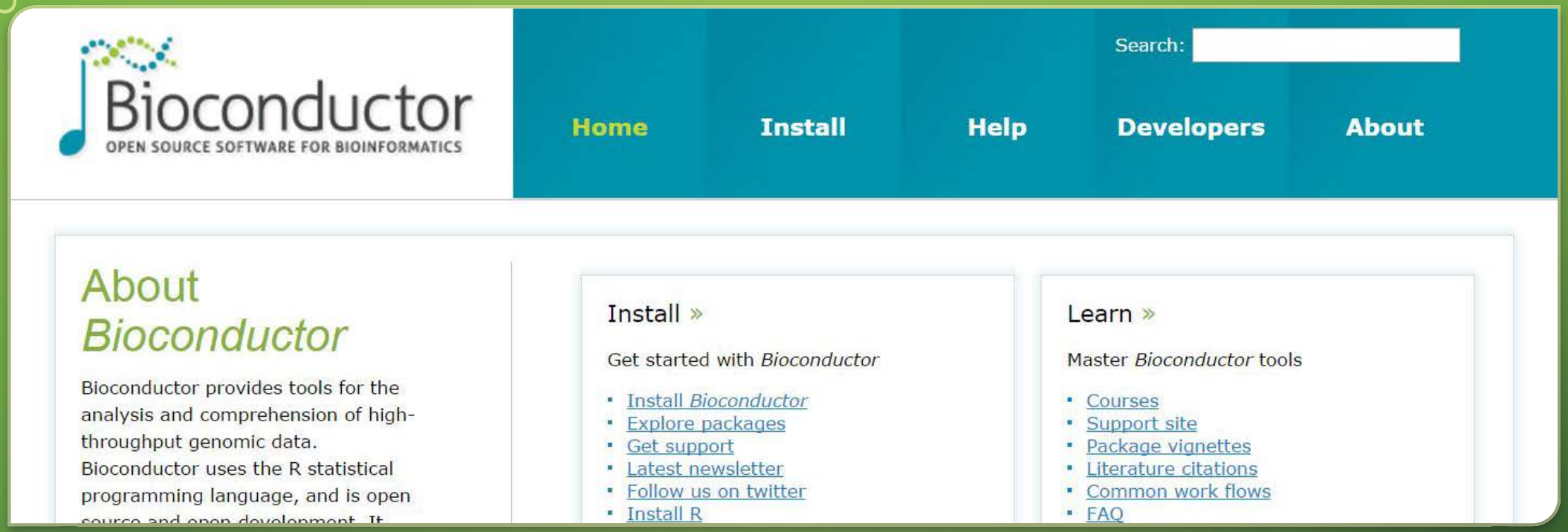


From: Fang Z, et al. Cell Biosci. 2012;2:26



GALAXY SUITE OF TOOLS

http://galaxyproject.org/



BIOCONDUCTOR AND R

https://www.bioconductor.org/

https://www.r-project.org/

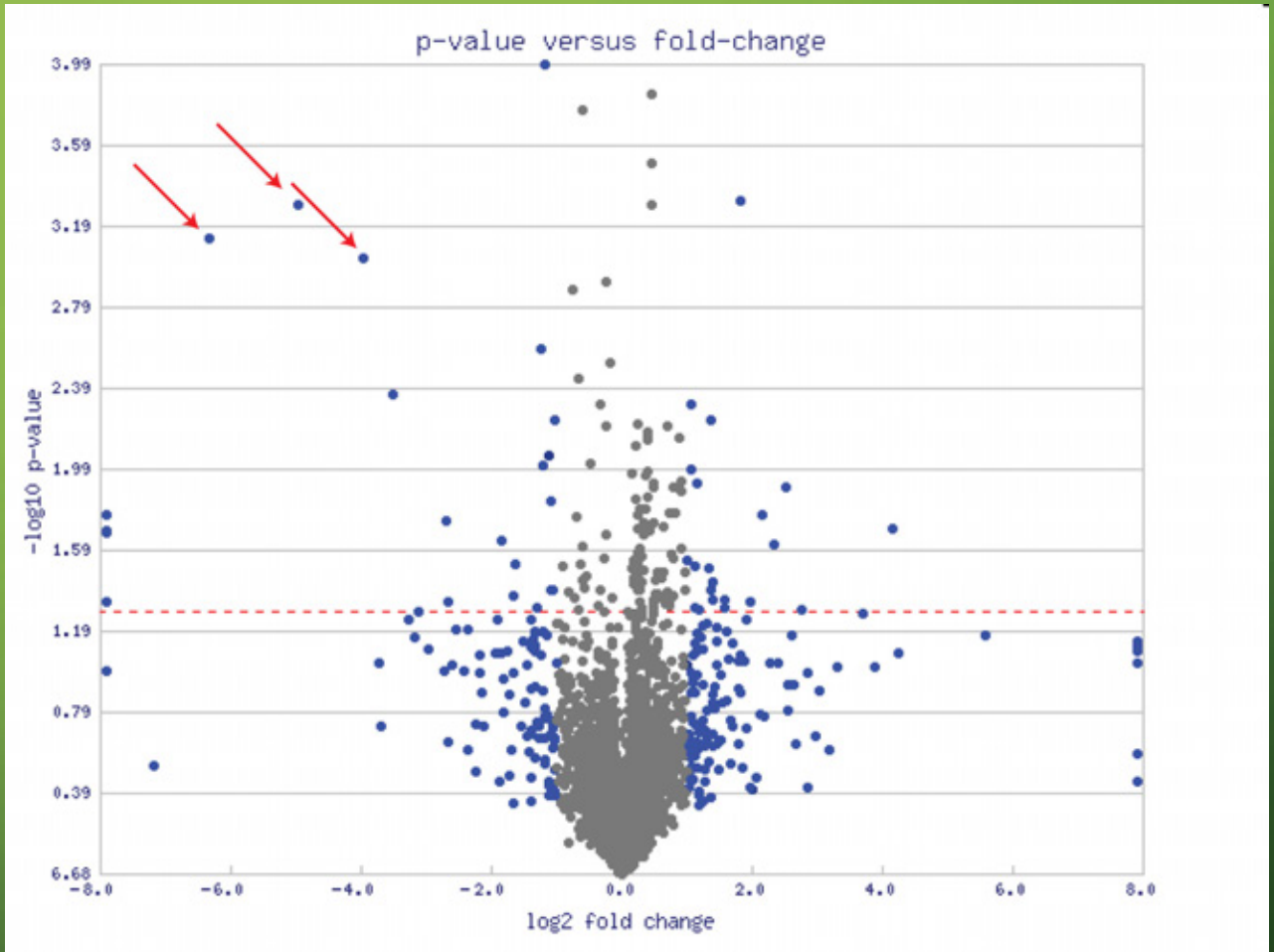
IDENTIFICATION OF PATTERNS IN THE DATA

COMMON WAYS TO MAKE SENSE OF –OMICS DATA

- Identification of patterns in the data
  - Advantage: Unbiased; Disadvantage: Can be hard to interpret, easy to over-interpret
  - Hierarchical clustering
  - Principle Components Analysis
  - Other graphical representations
- Mapping new information on to prior knowledge
  - Gene Ontology mapping
  - Canonical pathway mapping
  - Set-level analysis (GSEA, e.g.)
  - Protein interaction analysis
  - Common functions of compounds

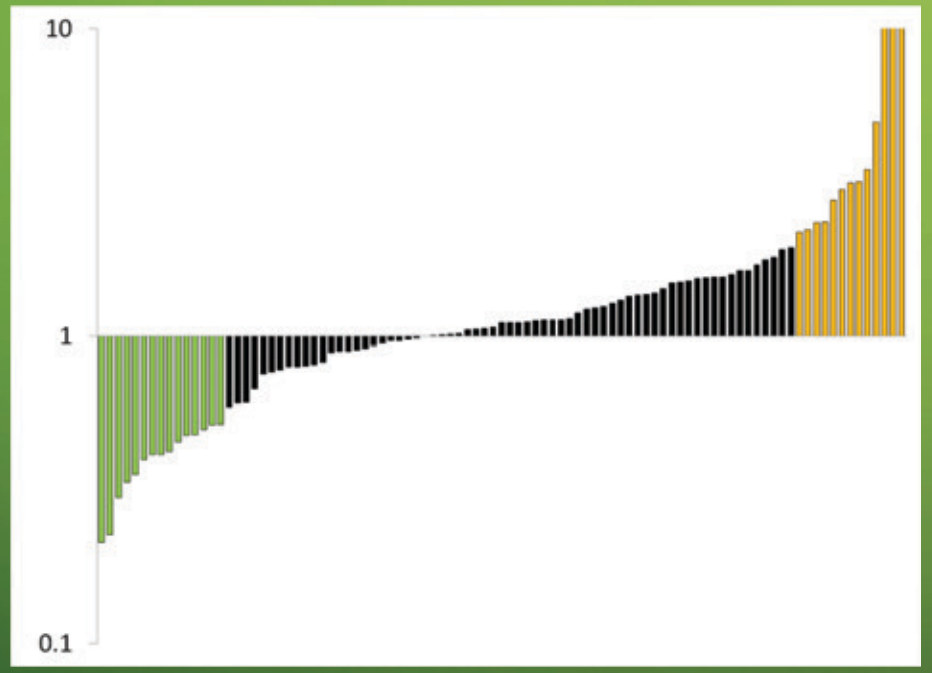
VOLCANO PLOT

By Roadnottaken - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=8901192

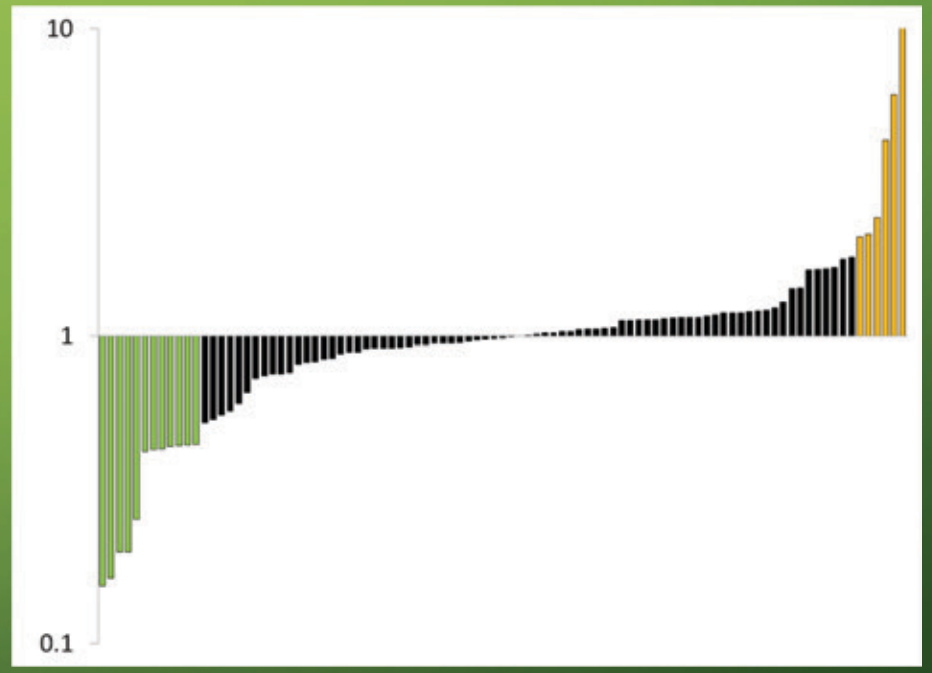


WATERFALL PLOTS

Cell Type 1



Cell Type 2



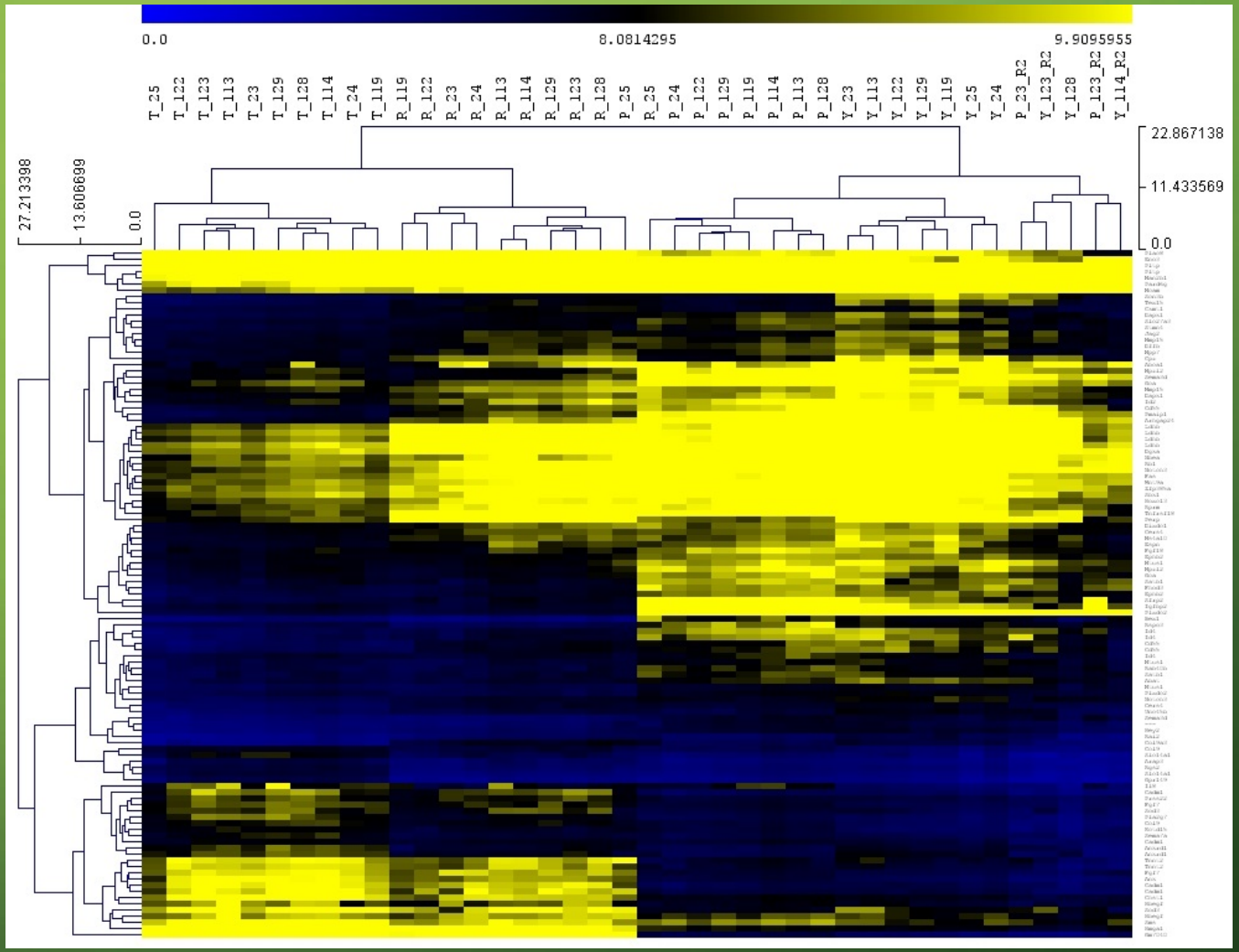
Gene order on these graphs is different!

HIERARCHICAL CLUSTERING

- Hierarchical clustering relates the samples or elements to one another in hierarchical fashion
  - Based on various metrics (correlation, ranking or distance between samples)
- There are other types of clustering, used less commonly but may be important in your project
  - k-means clustering, self-organizing maps (SOM), etc.

HIERARCHICAL CLUSTERING

Clustering and visualization via TM4 MeV software, Version 4.9.0



New in JMP® Pro 12

As the advanced analytics version of our software, JMP Pro contains everything users know and love about JMP – and more. With the release of this latest version – JMP Pro 12 – users will enjoy new capabilities and see a leap in performance in almost all of its earlier platforms.

› Learn how to get JMP Pro 12

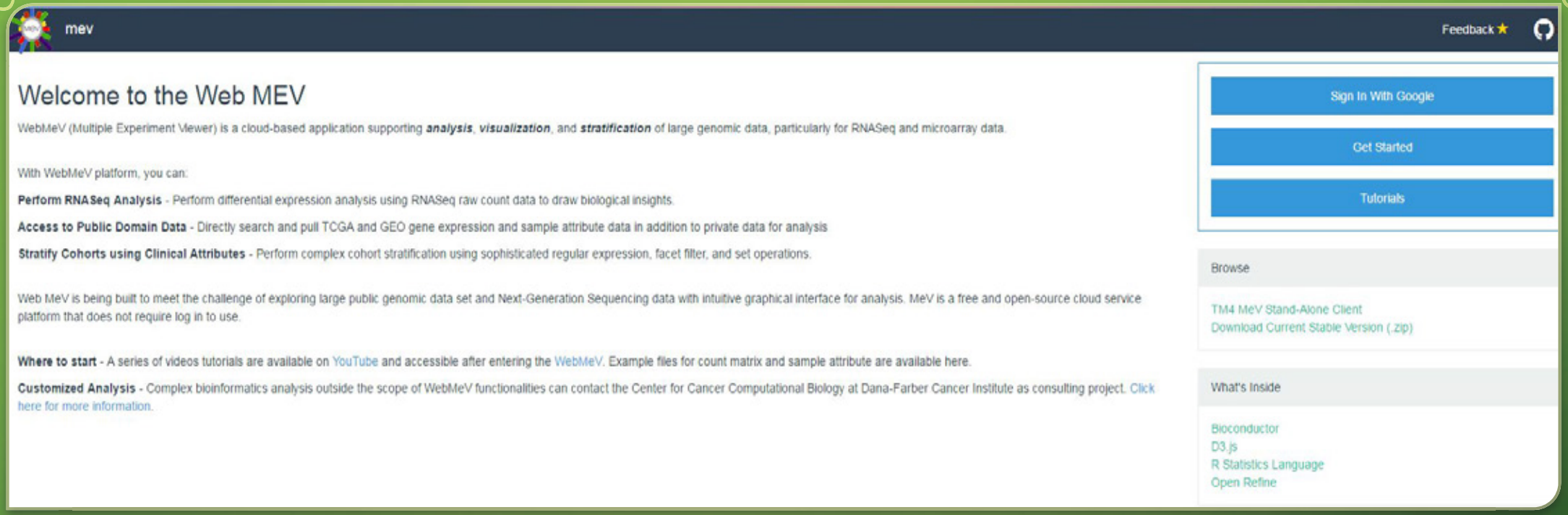
JMP: BETTER STATISTICAL ANALYSIS

http://www.jmp.com/en_us/home.html

http://tech.rochester.edu/services/software-site-licensing/

MORPHEUS

https://software.broadinstitute.org/morpheus/



MULTI-EXPERIMENT VIEWER (MEV)

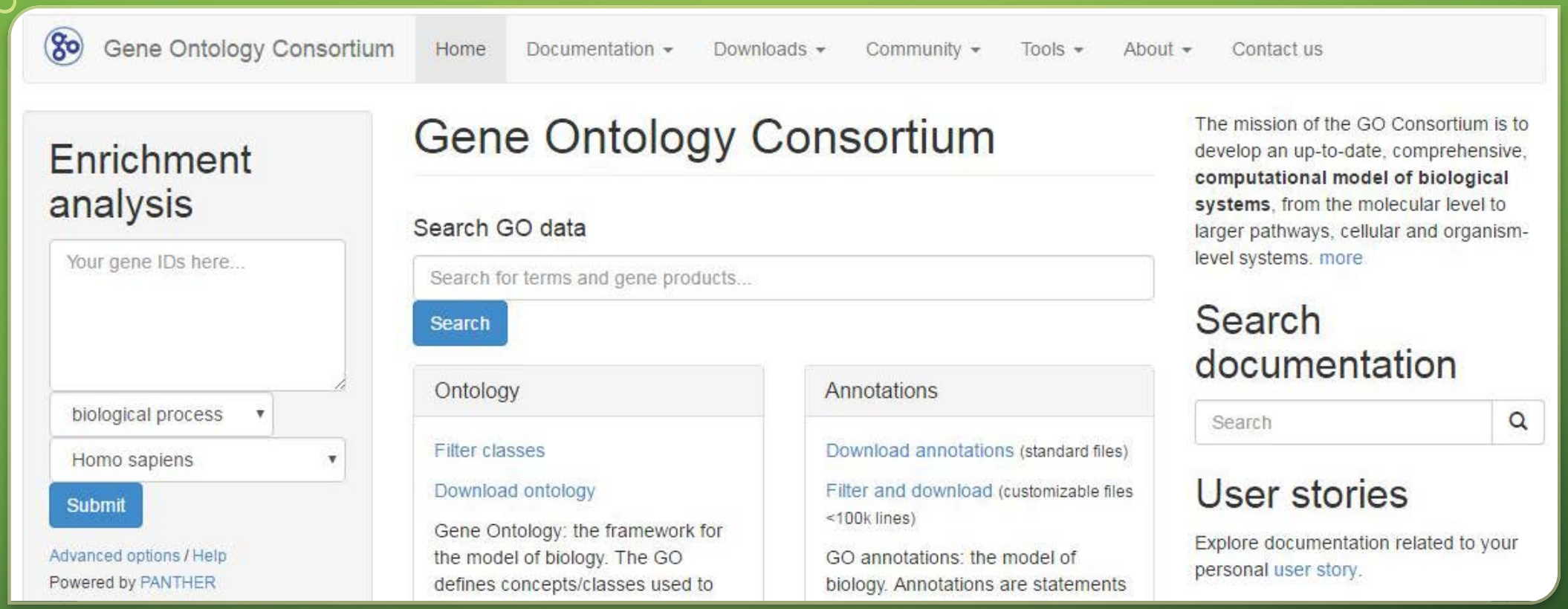
http://www.tm4.org/#/welcome

MAPPING NEW INFORMATION ONTO PRIOR KNOWLEDGE

SET-LEVEL / ENRICHMENT ANALYSIS

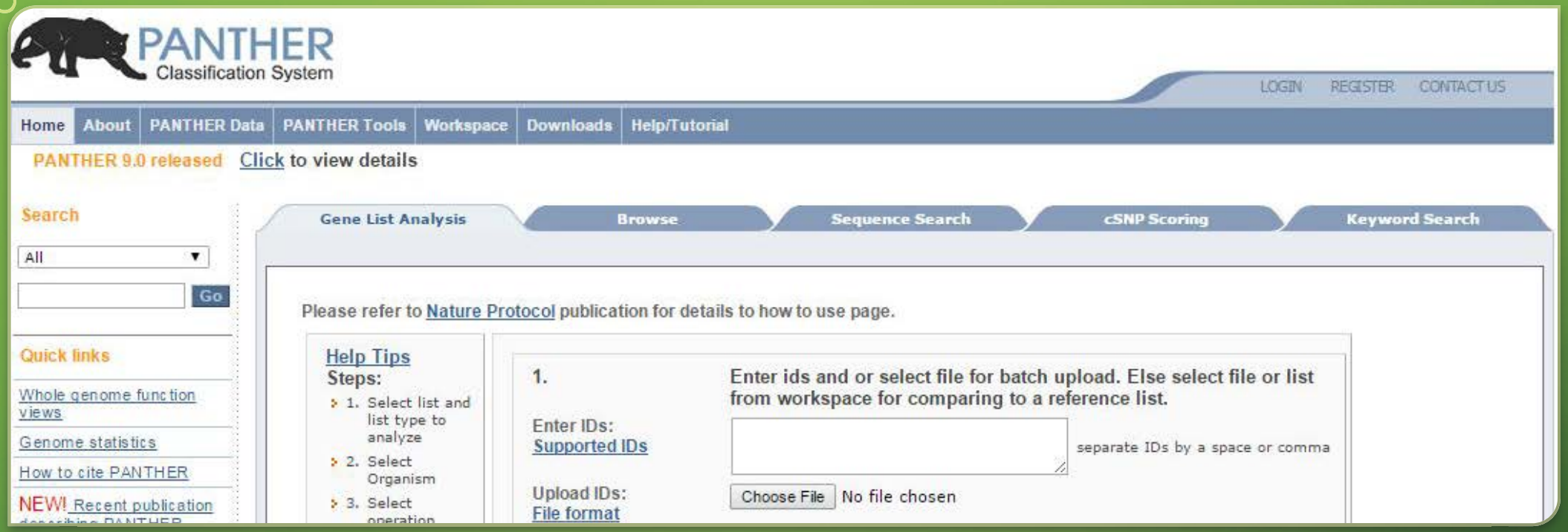
- Tests for over-representation of elements of a given set in a larger set
  - Gene ontology analysis
    - Hierarchical system of categorizing genes
    - Controlled vocabulary
    - Assess over- or under-represented ontology terms
  - Pathway analysis
    - Allows integration of novel experimental data with known biological function
      - Relies on accurate mapping of pathways in public domain
      - Assumes that pathways are the same in all contexts

GENE ONTOLOGY ANALYSIS



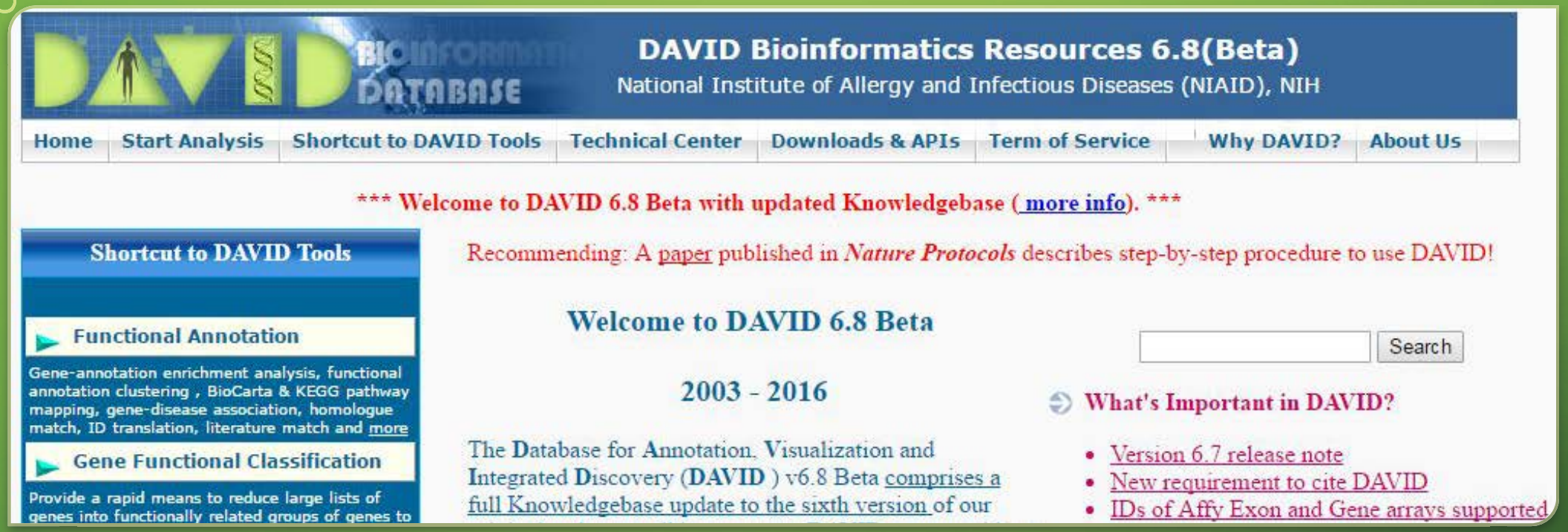
THE GENE ONTOLOGY CONSORTIUM

http://geneontology.org/



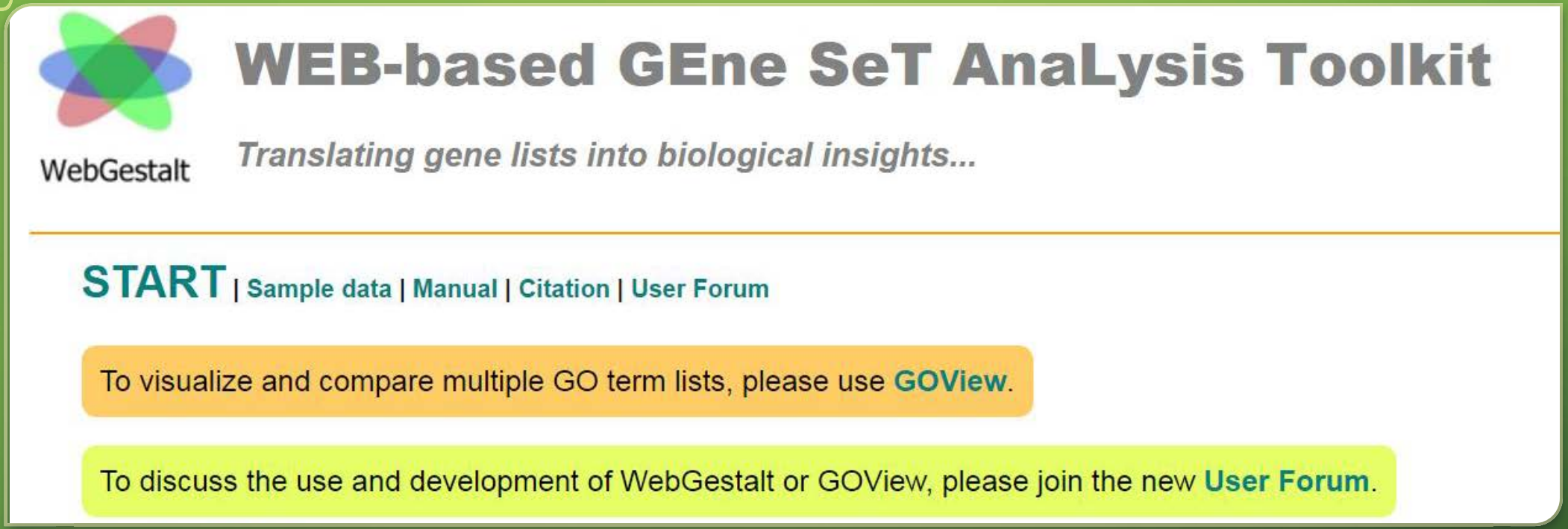
PROTEIN ANALYSIS THROUGH EVOLUTIONARY RELATIONSHIPS (PANTHER)

http://www.pantherdb.org/



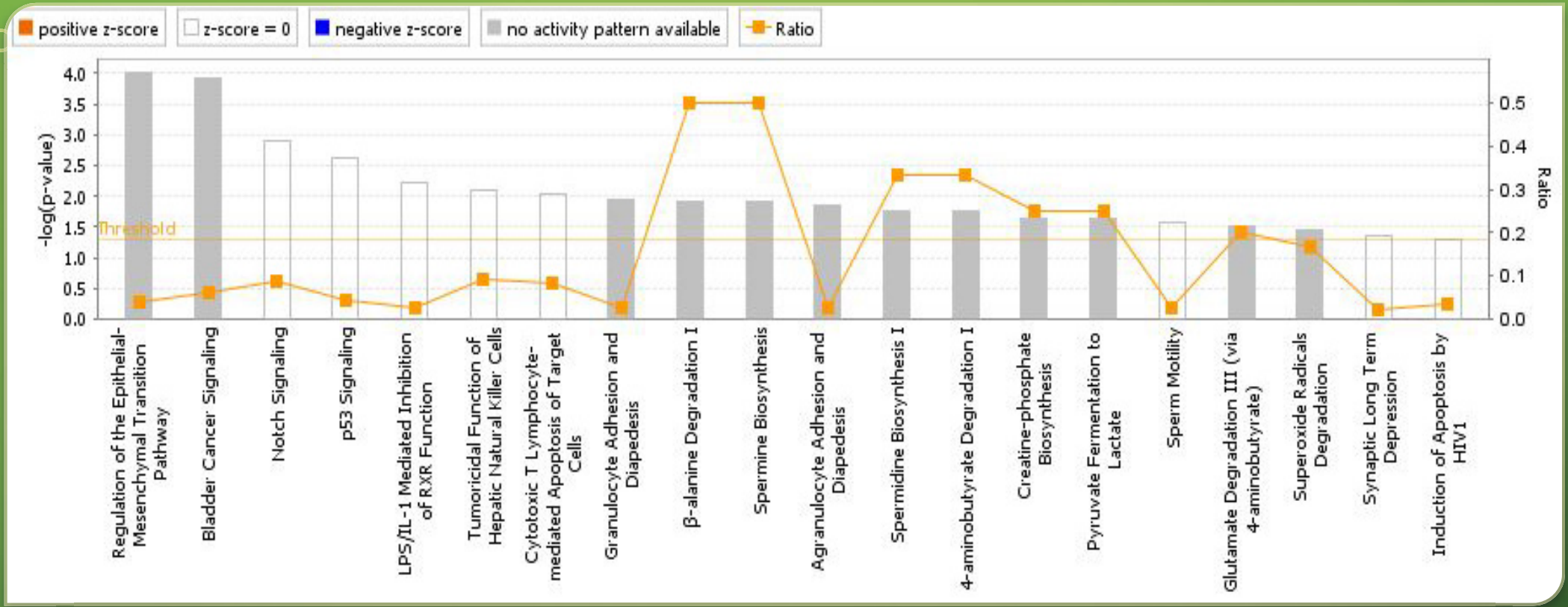
DATABASE FOR ANNOTATION, VISUALIZATION AND INTEGRATED DISCOVERY (DAVID)

https://david.ncifcrf.gov/



WEB-BASED GENE SET ANALYSIS TOOLKIT (WEBGESTALT)

http://www.webgestalt.org/

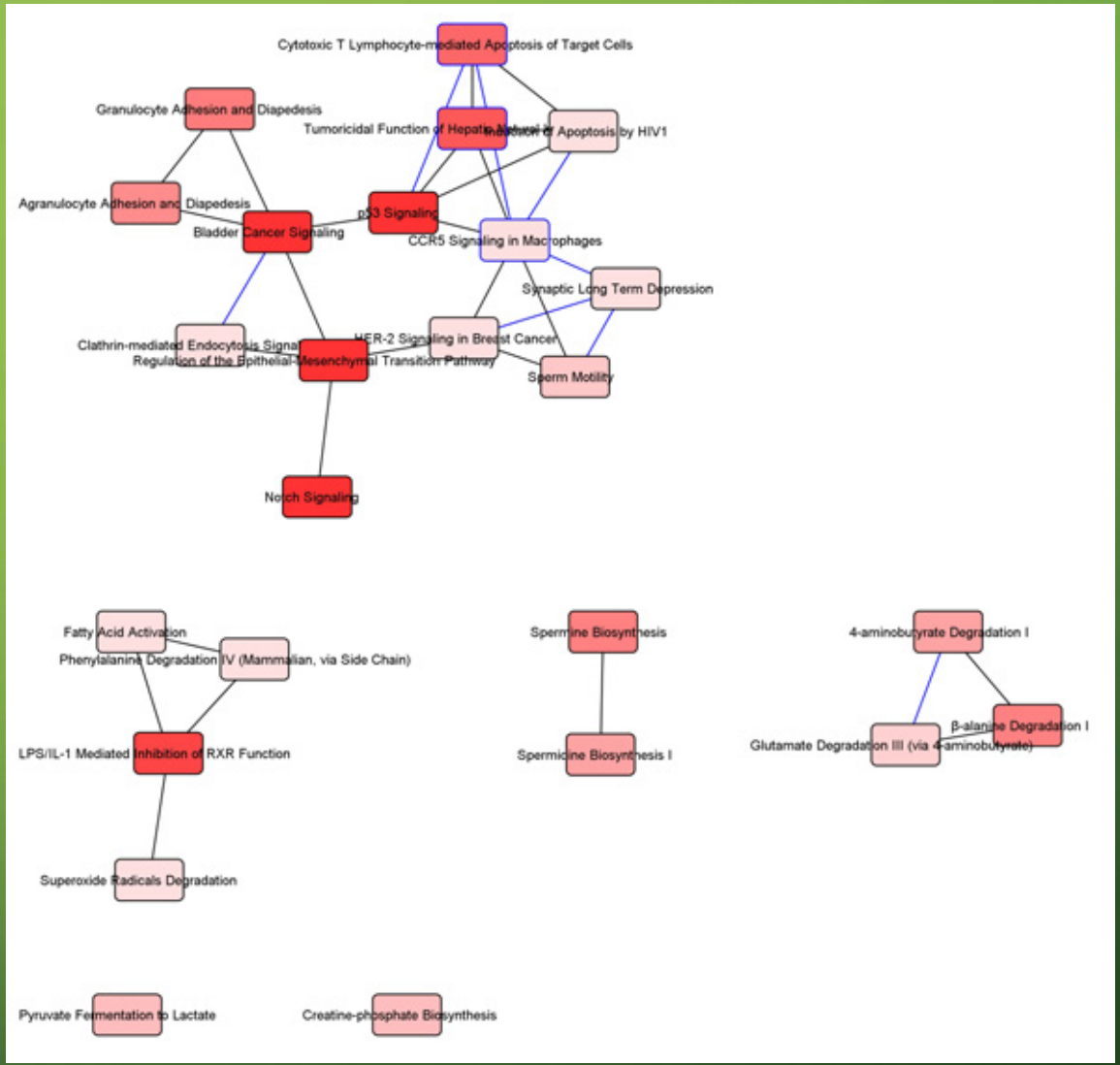


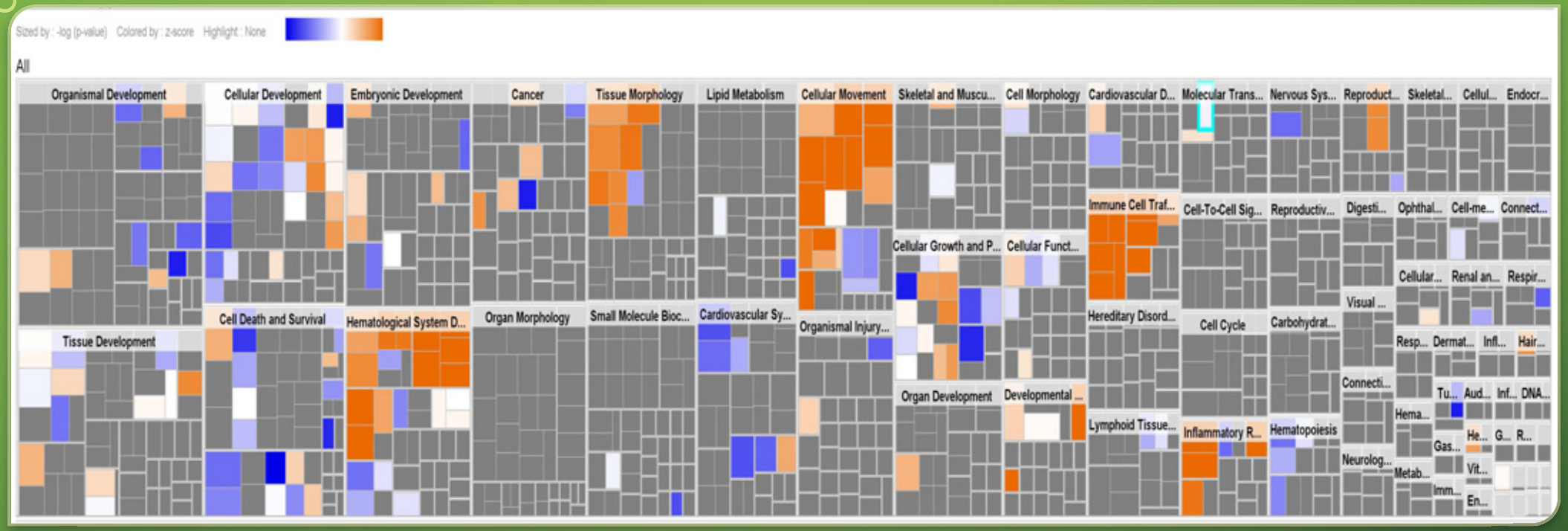
GENE ONTOLOGY ANALYSIS

Created with Ingenuity Pathway Analysis software, March 2016

GENE ONTOLOGY ANALYSIS

Created with Ingenuity Pathway Analysis software, March 2016





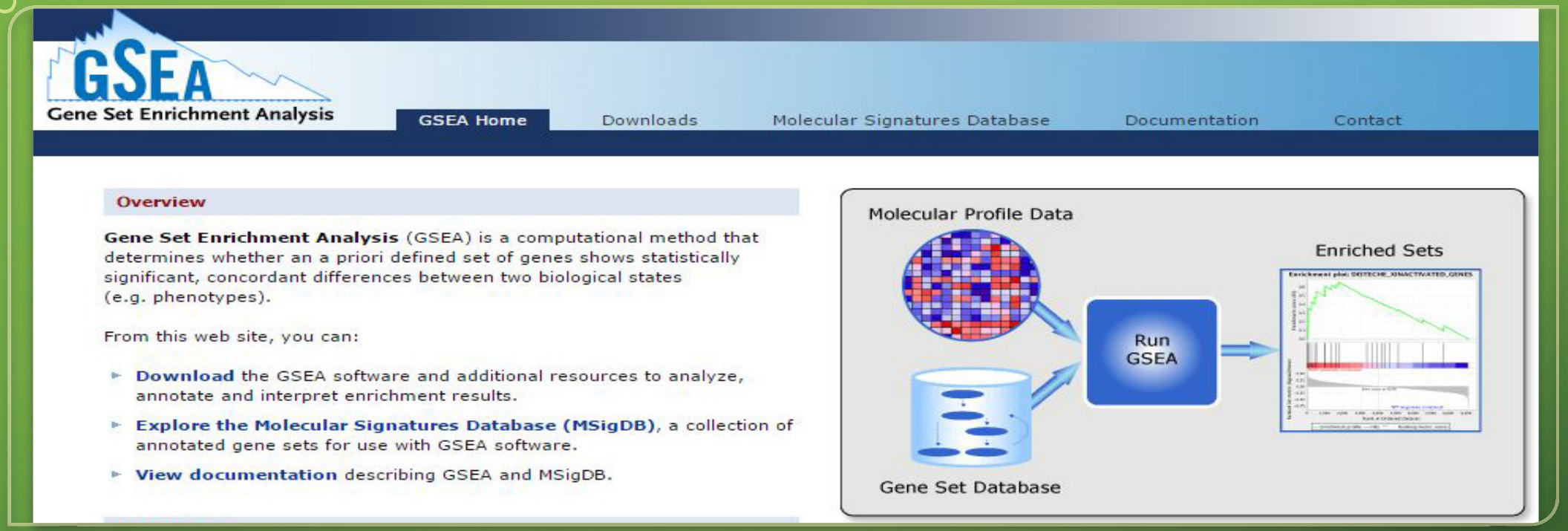
GENE ONTOLOGY ANALYSIS

Created with Ingenuity Pathway Analysis software, March 2016

PATHWAY ANALYSIS

THE REACTOME PROJECT

http://www.reactome.org/



GENE SET ENRICHMENT ANALYSIS (GSEA) & THE MOLECULAR SIGNATURES DATABASE (MSIGDB)

http://www.broadinstitute.org/gsea/index.jsp

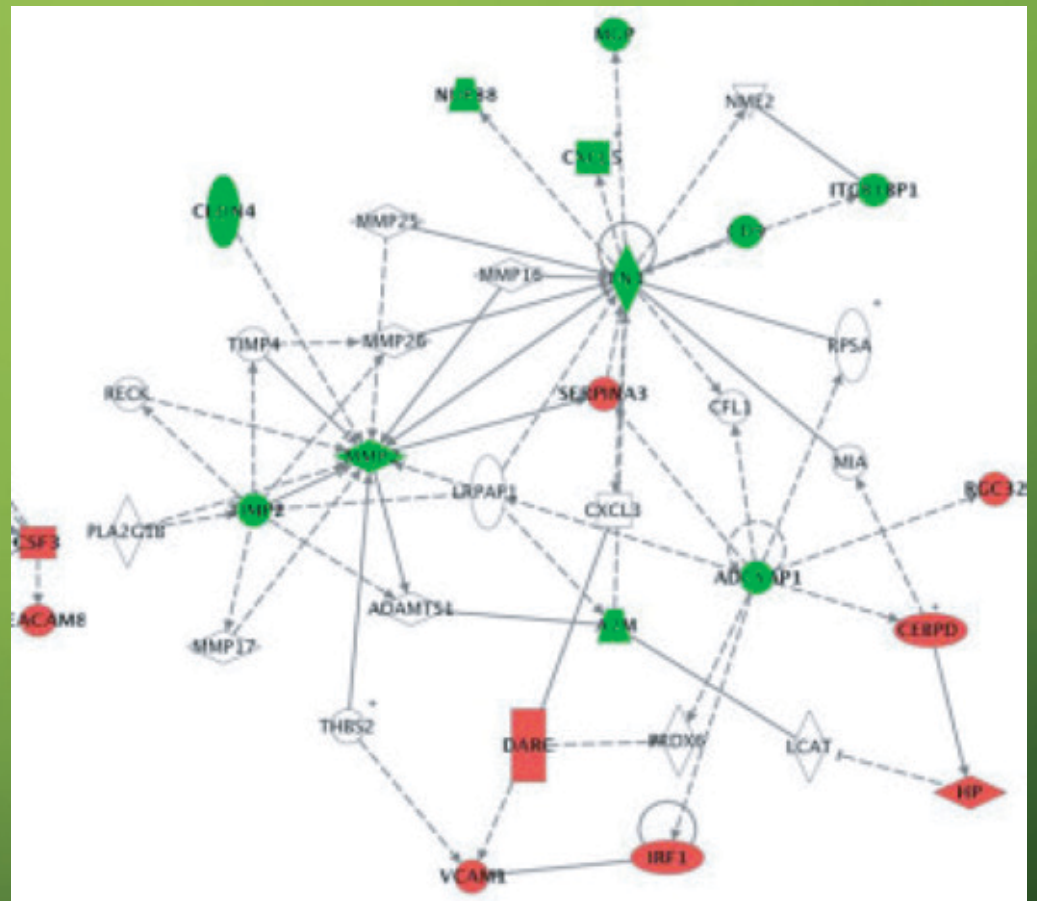
INGENUITY PATHWAY ANALYSIS

http://www.ingenuity.com/products/ipa

https://www.urmc.rochester.edu/libraries/Miner/research/MolecularBiologyTools.cfm

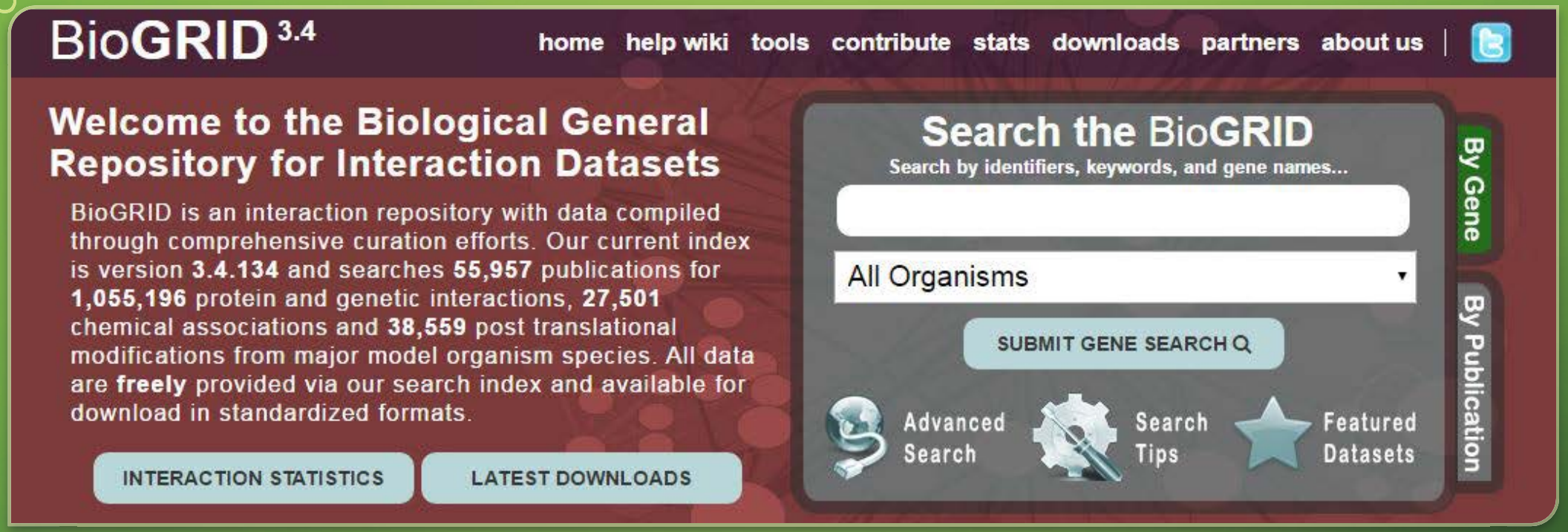
INGENUITY PATHWAY ANALYSIS OUTPUT

Hypothetical “network” generated by Ingenuity Pathway Analysis



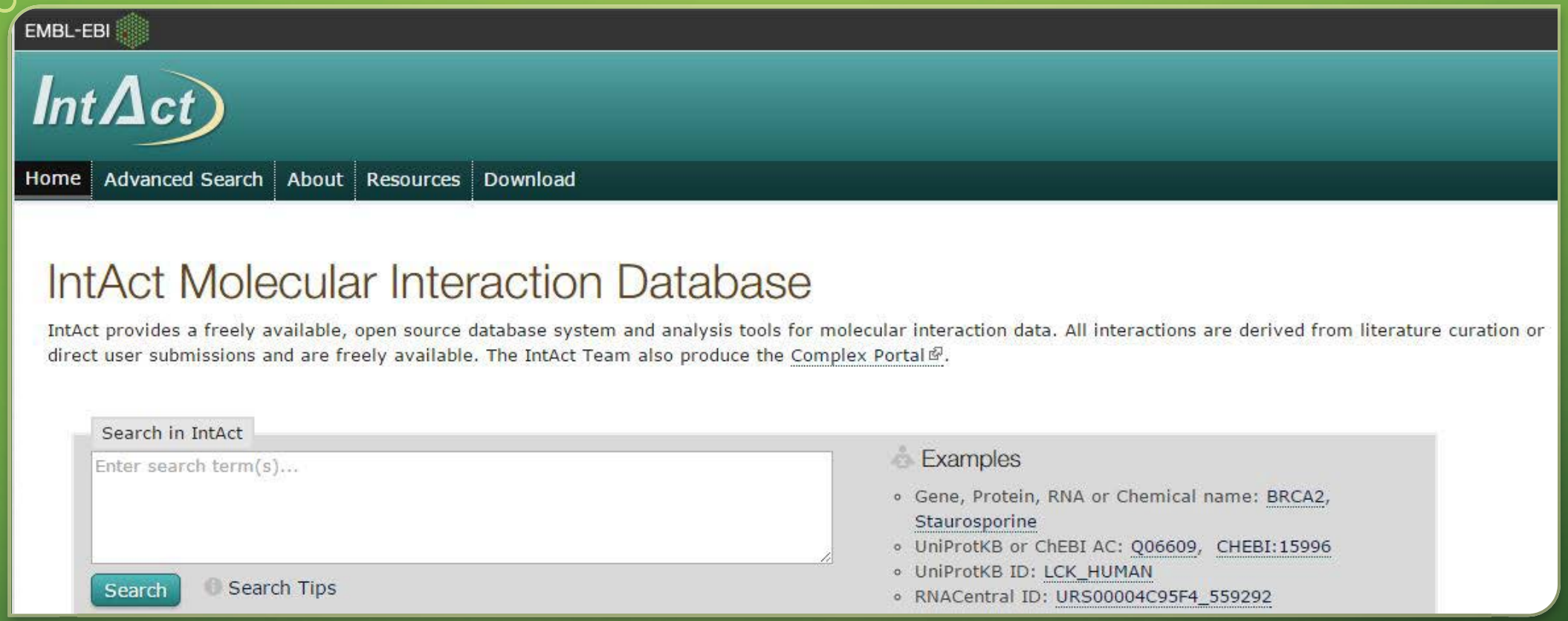
b: 8 hrs; IPA Network 1

PROTEIN INTERACTION ANALYSIS



BIOLOGICAL GENERAL REPOSITORY FOR INTERACTION DATASETS (BIOGRID)

http://thebiogrid.org/



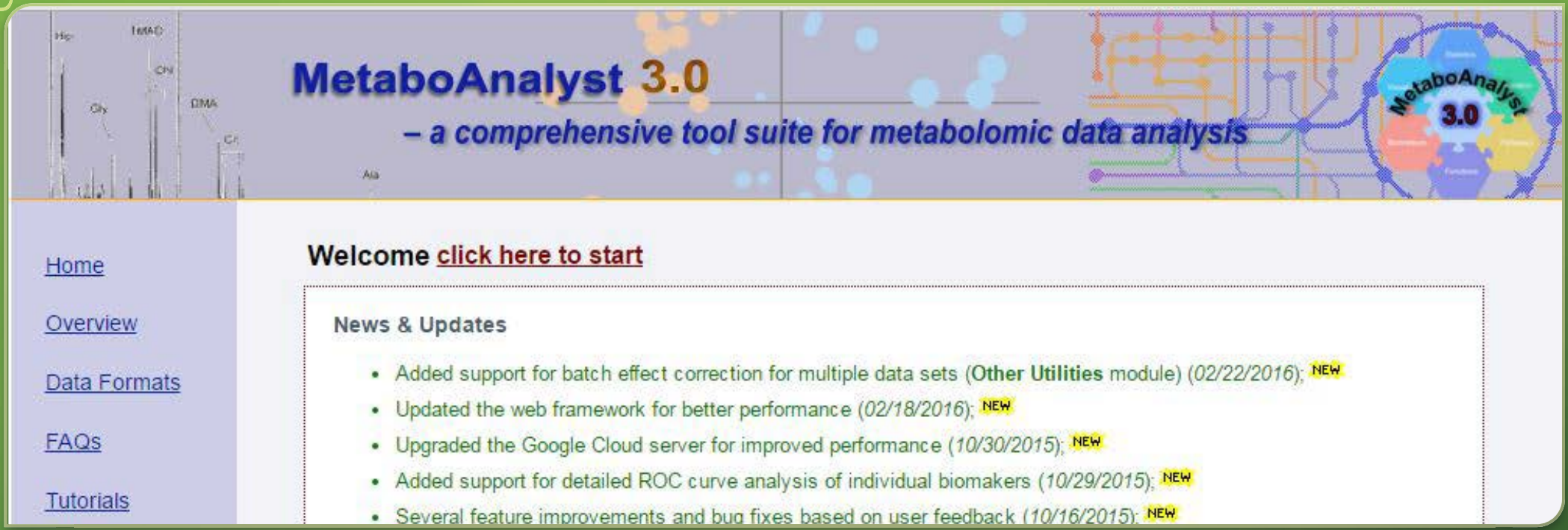
INTACT MOLECULAR INTERACTION DATABASE

http://www.ebi.ac.uk/intact/

SEARCH TOOL FOR THE RETRIEVAL OF INTERACTING GENES/PROTEINS (STRING)

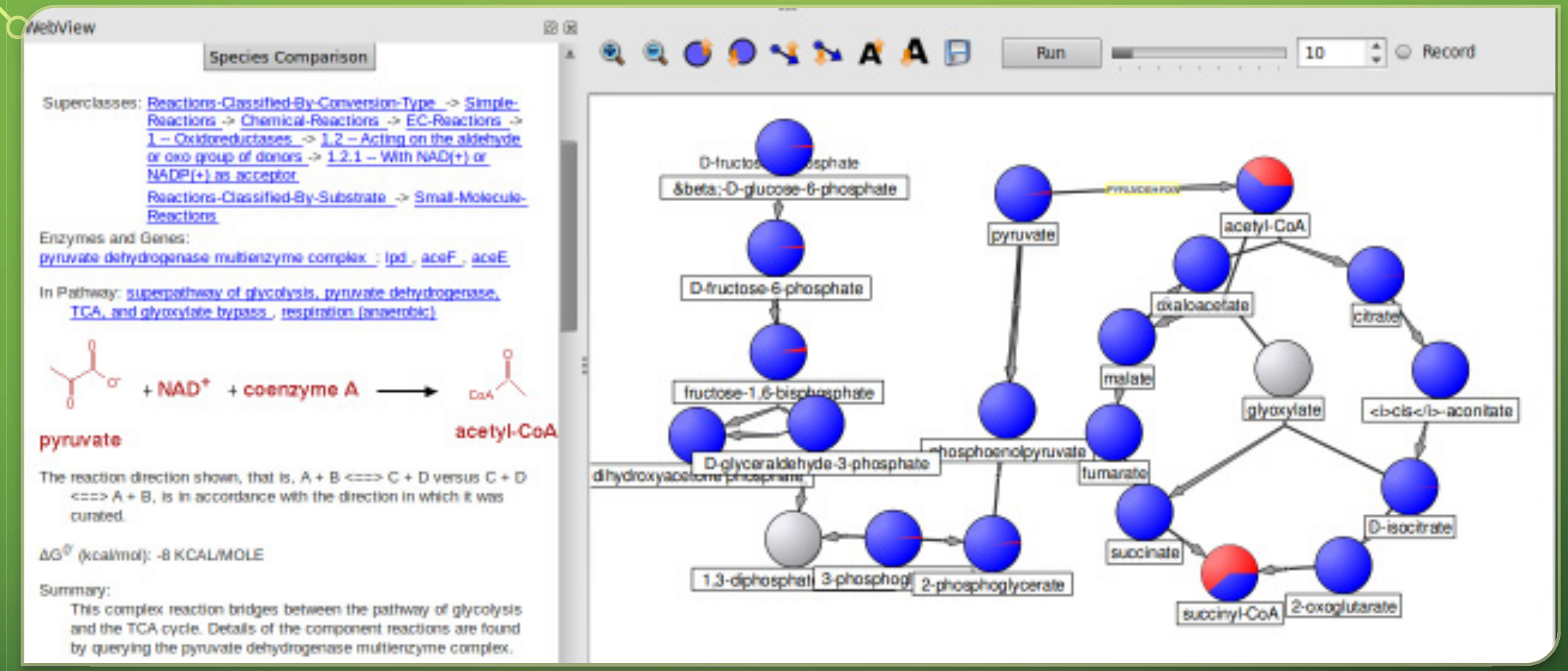
http://string-db.org/

METABOLOMICS DATA ANALYSIS



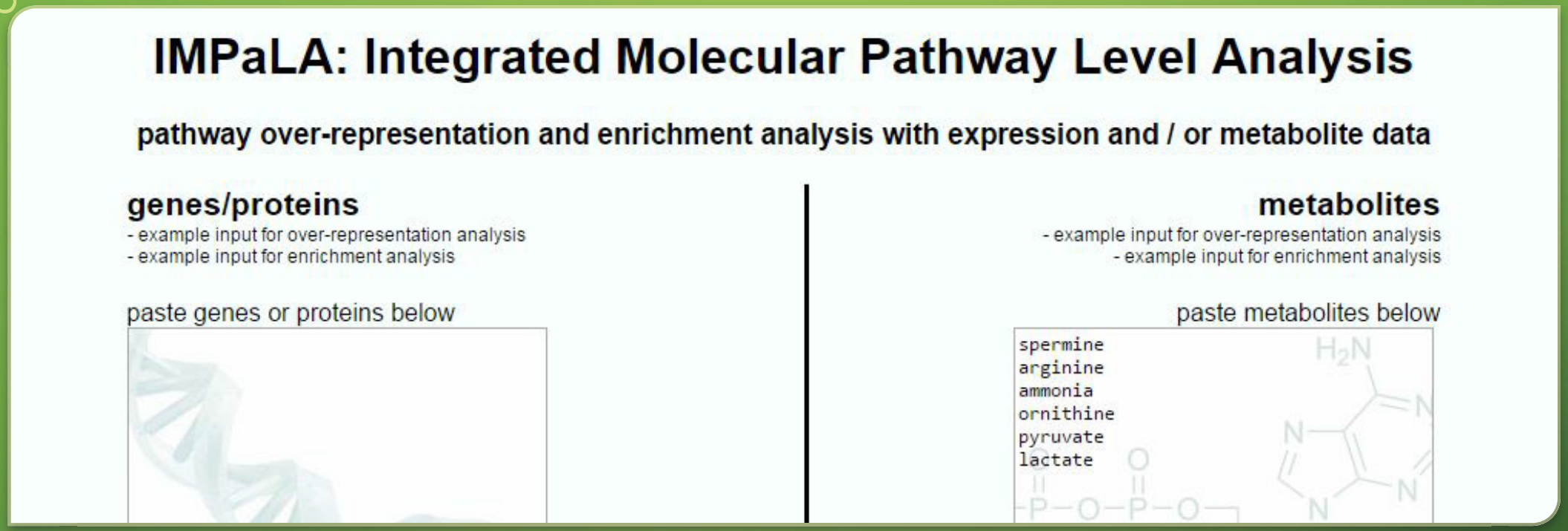
METABOANALYST 3.0

http://www.metaboanalyst.ca/



MAVEN: OPEN-SOURCE METABOLOMICS DATA ANALYZER

http://genomics-pubs.princeton.edu/mzroll/index.php?show=index



INTEGRATED MOLECULAR PATHWAY LEVEL ANALYSIS (IMPALA)

http://impala.molgen.mpg.de/

HIGH-THROUGHPUT CHEMICAL SCREENING

PUBCHEM

https://pubchem.ncbi.nlm.nih.gov/