

NGS: Downstream Analysis & Processing

•1

The Plan...

DOWNSTREAM ANALYSIS

Gene sets

- Transcriptome studies
- Cistrome / Epigenome studies
 - Function enrichment
 - GSEA, DAVID, IPA
 - STRING, Cytoscape

DOWNSTREAM PROCESSING

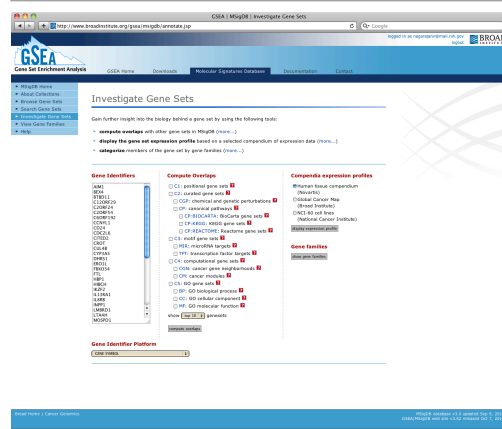
Genome coordinate sets

- NGS alignment files (reads/tags)
- ChIP-Seq peaks
 - BEDTools
 - Galaxy
 - UCSC Table browser

•2

Gene Set Enrichment Analysis (GSEA)

<http://software.broadinstitute.org/gsea/index.jsp>



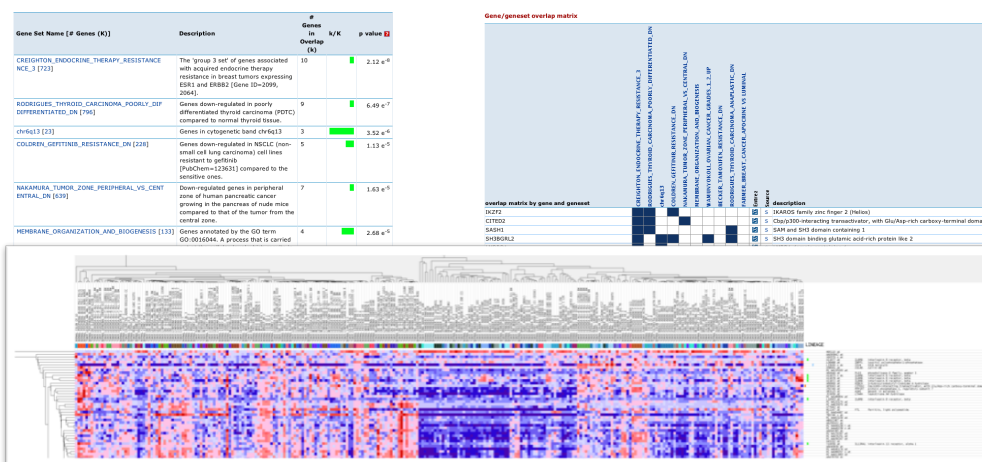
Gene Set Enrichment Analysis (GSEA)

- Broad Institute
- “A computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states”
- Download** the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- Explore the Molecular Signatures Database (MSigDB)**, a collection of annotated gene sets for use with GSEA software.

•3

Gene Set Enrichment Analysis (GSEA)

<http://software.broadinstitute.org/gsea/index.jsp>



4

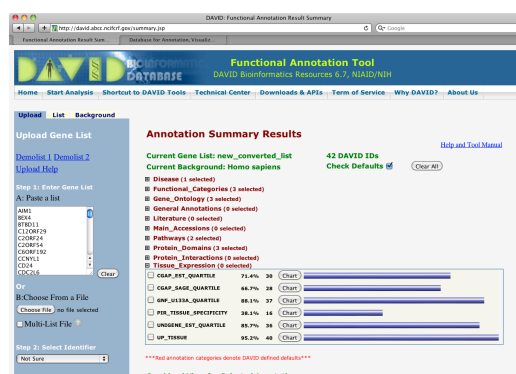
Database for Annotation, Visualization and Integrated Discovery (DAVID)

<https://david.ncicrf.gov/>

DAVID provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list



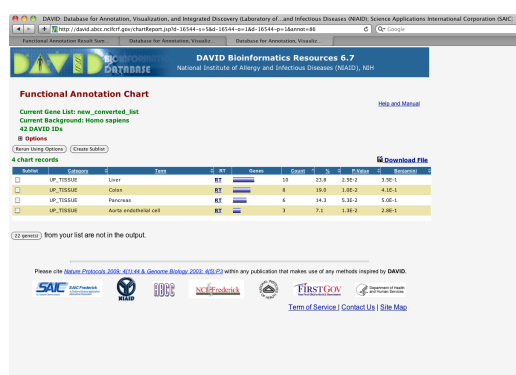
•5

Database for Annotation, Visualization and Integrated Discovery (DAVID)

<https://david.ncicrf.gov/>

Cont.

- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another.

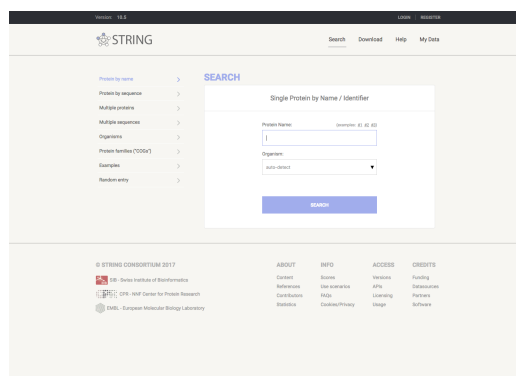


6

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)

<https://string-db.org>

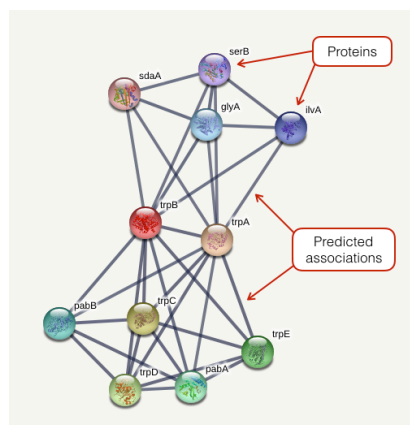
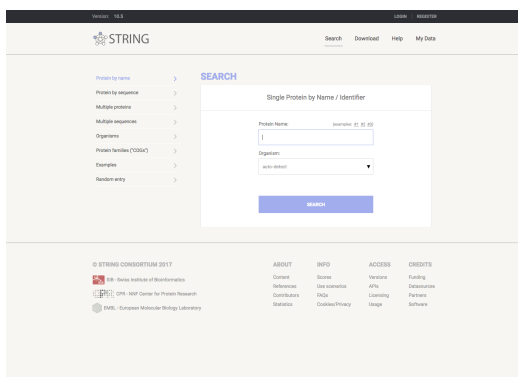
STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a biological database and web resource of known and predicted protein-protein interactions.



•7

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)

<https://string-db.org>



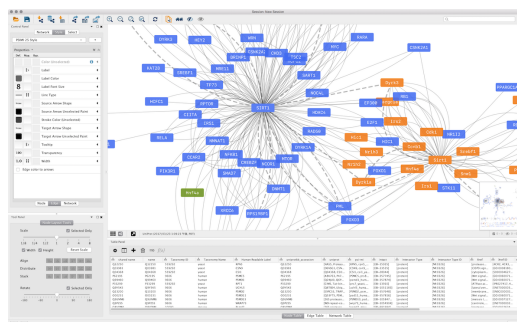
8

Cytoscape

<http://www.cytoscape.org/>

Software

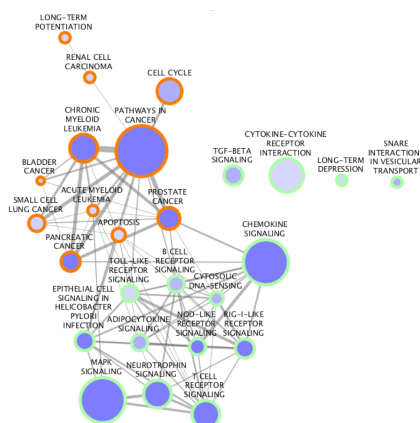
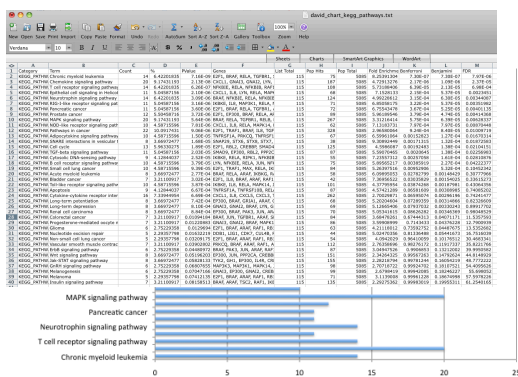
Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data.



•9

Cytoscape

<http://www.cytoscape.org/>



•10

BED-Tools

<http://bedtools.readthedocs.io/en/latest/>

Command line

Open source

A variety of input formats (BED, BEDPE, SAM/BAM, GFF, VCF)

Things we could do...

- Intersect, Union
- Merge
- Coverage
- Subtract
- Convert
- Closest
- Shuffle
- Group, Sort

•11

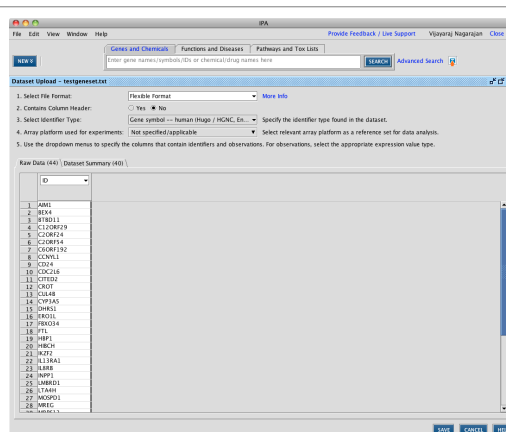
Commercial Tools

Ingenuity Pathway Analysis (IPA)

<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>

Ingenuity Pathway Analysis is a powerful analysis and search tool that uncovers the significance of 'omics data and identifies new targets or candidate biomarkers within the context of biological systems.

- Canonical Pathway
- Molecule Activity Predictor (MAP)
- Upstream Regulator Analysis
- Mechanistic Networks
- Downstream Effects Analysis
- Regulator Effects
- Network Analysis
- ...



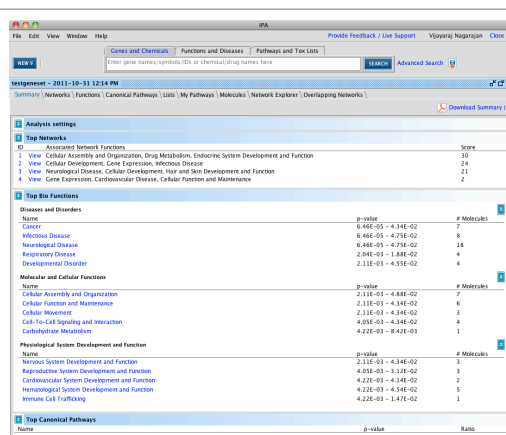
13

Ingenuity Pathway Analysis (IPA)

<https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>

Cont...

- Comparison Analysis
- MicroRNA Target Filter
- Isoform View
- Disease View
- Tox Lists and Tox Functions
- Gene and ChemView
- Interactive Disease and Functions Nodes
- Biomarker Filter
- Path Designer



14