Nodes

- A node is an item on a graph
- **EGAN** contains two types of nodes
  - Entrez Gene nodes
    - Represents a single gene with a backing Entrez Gene ID
  - Association nodes
    - Represents a semantically-related set of Entrez Gene nodes
Nodes are backed by web references

- Right-click on a gene node
  - Description/summary
  - Links to web references

This gene encodes tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. p53 protein is expressed at low level in normal cells and at a high level in a variety of transformed cell lines, where it's believed to contribute to...
Edges

- An edge connects two nodes (as a line)
- The node NRAS has an edge connecting it to MAPK1, BRAF and MAPK signaling pathway
  - All those nodes are “neighbors” of NRAS
- EGAN contains many types of edges
  - Edges between gene nodes
    - Protein-protein interactions (PPI)
      - BRAF has a PPI with MAPK1
    - PubMed co-occurrence
      - NRAS and BRAF are mentioned in the same article(s)
    - Chromosomal adjacency
      - Genes are adjacently located on the chromosome
  - Edges between gene and association nodes
    - Show which genes belong to which gene sets
    - All genes shown are members of the MAPK signaling pathway
Most edges are backed by literature

- You link out to each article and pre-defined search queries by right-clicking on each edge
- Reference counts can be displayed on edges

31 articles available that discuss NRAS and BRAF
How to use EGAN

- Load your experiment results using the Launch EGAN Wizard
- Your data must be in the proper 3-column format
  - ID, statistic (e.g. fold-change), p-value (or qvalue/FDR estimate)
- You should include all genes/proteins from your assay
  - i.e. don’t do a p/q value cutoff beforehand!
How to use **EGAN**

- Remember to specify the proper background of genes
  - Chip-based experiments
    - Keep all genes that were available on your chip
  - RNA-Seq experiments
    - Keep all genes that have transcript IDs
  - Proteomics experiments
    - Keep all genes with Protein IDs
  - Multiple experiments
    - Keep all genes that could have been discovered as significant by all experiments
How to use EGAN

- Find gene nodes of interest
  - Use the Entrez Gene Node Table

Click column header to sort, then click-and-drag to select top gene rows
How to use EGAN

- Show selected genes on the Network View
- Using information from the “focused” experiment
  - Gene node border color is relative to its statistic
  - Gene node border width is relative to the $-\log(p\text{-value})$
- Run layout algorithms, investigate gene information
How to use EGAN

• Calculate hypergeometric enrichment for association nodes
  – Lower p-values indicate association nodes that have a high degree of overlap with the set of visible gene nodes
• Selectively show enriched association nodes
How to use EGAN

- Think about how the gene nodes, edges and enriched association nodes relate to your experiment

- Remember to follow links to web references and literature

- Consider different gene sets from your experiment
  - Change the p-value cutoff and see how the network and enrichments change
  - Investigate the up-regulated genes, the down-regulated genes and the combined set

- Perform GSEA-like (AKA global, rank-based) enrichment
  - Must be specified beforehand in the Launch EGAN Wizard – see 10) SEED Enrichment
  - Note how the different enrichment algorithms compare/contrast

- Construct a module that adds non-significant connecting genes to the network

- Perform a combined analysis using the results of multiple experiments
  - Load multiple assay results
  - Compare your genes to gene lists from previous publications

- Remember to save your gene lists as groups in EGAN and save snapshots of interesting networks
More information

• See http://akt.ucsf.edu/EGAN/
  – Post questions to the discussion forum
  – Send an email to the developers