Microarray Interactive Network Exploration and Representation

Acknowledgments

- *Harshana Randeni, Ray Leung, Jane Sivieng* - implementation
- *Julia Sellmeier* – user interface design and usability testing
- *Mike Bain* – machine learning
- *Dan Catchpoole* – data and evaluation

The problem

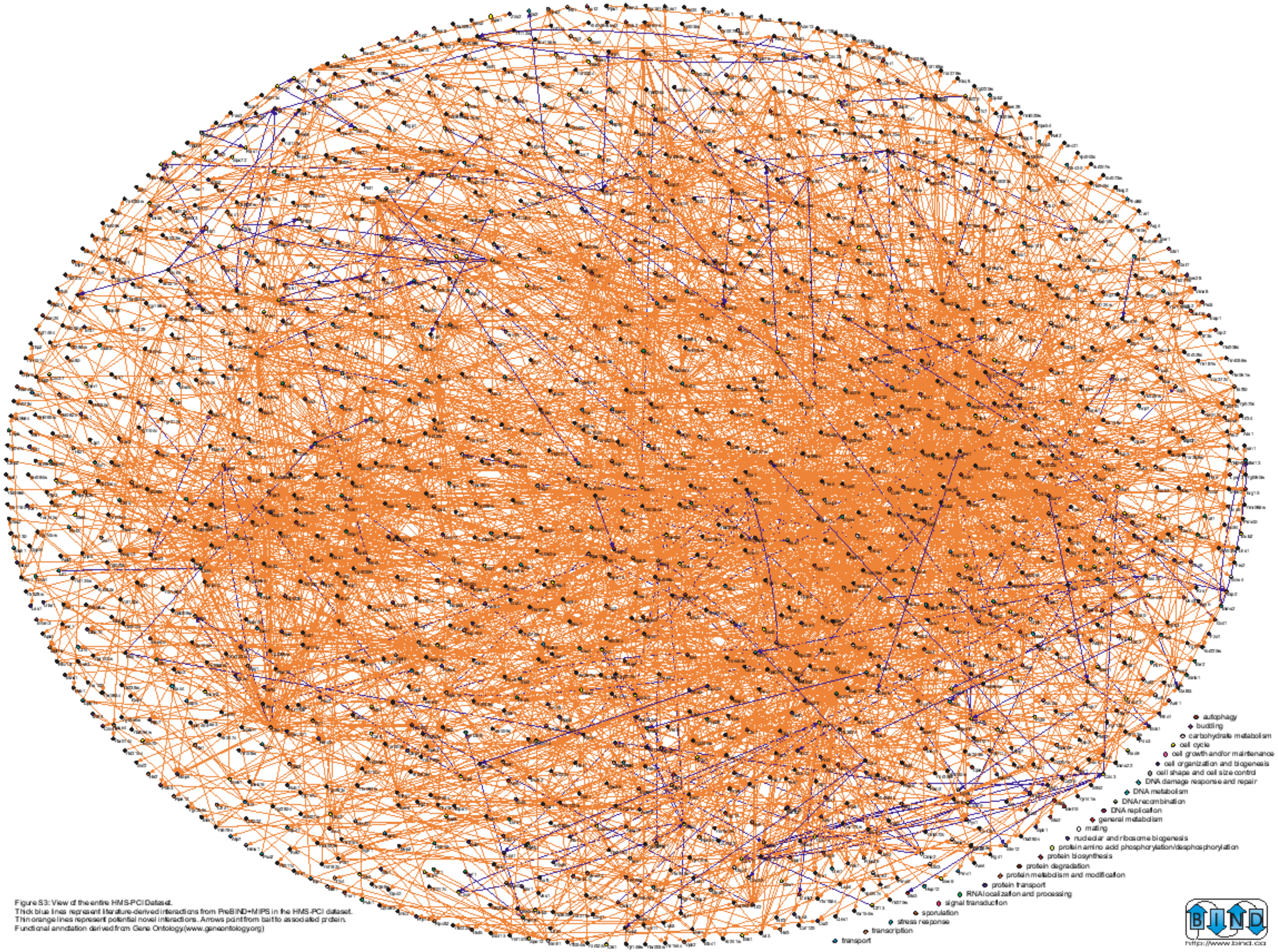


Figure S3: View of the entire HMS-PCI Dataset.
Thick blue lines represent literature-derived interactions from PreBIND+MIPS in the HMS-PCI dataset.
Thin orange lines represent potential novel interactions. Arrows point from bait to associated protein.
Functional annotation derived from Gene Ontology (www.geneontology.org)

http://www.bind.ca

Dealing with data overload

- Simple co-expression analysis/clustering
- “One-shot” network inference
- Many biologists still prefer a “one gene at a time” way of doing things

Learning tree-structured models from microarray data

- *Supervised* learning approach (classification instead of clustering)
- Allows the use of efficient algorithms
- eg: regression or model trees for numerical prediction
  - no pre-discretisation required
  - inspectable – not a “black-box”
- Combine multiple classifiers as “agents”
  - “Communication” through causal effects
  - Combine trees into networks

Learning trees on expression data

- General Idea

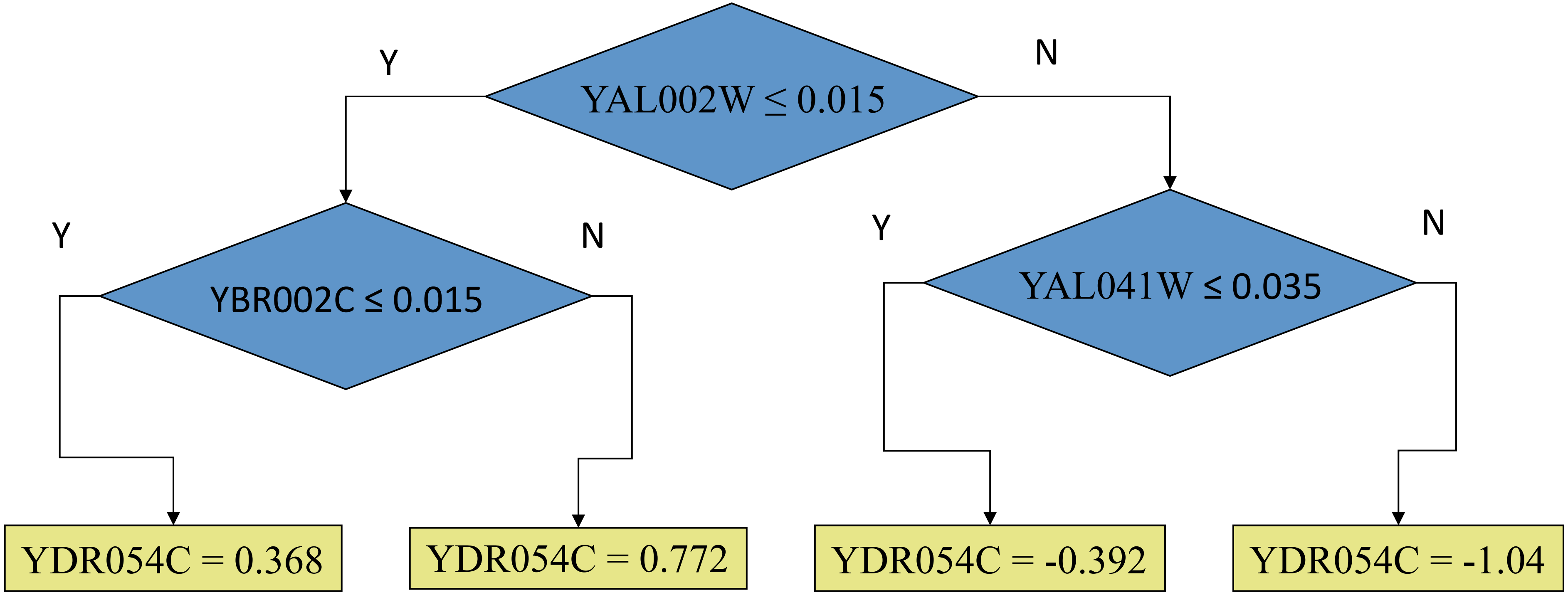
  (after Soinov et al. Genome Biology 2003, 4:R6)

  - Pick a subset of genes of interest and try to predict their expression from the expression levels of the other genes.
  - Let X be a gene-expression matrix, and x_{ij} the expression level of gene i in sample j.
  - Then for gene i construct a regression tree to predict the expression level of gene i given expression levels of all other genes in sample j.

Methods evaluated

- Decision trees: predict one of a finite set of discrete classes
- Regression trees: predict a continuous value by assigning a mean at each leaf
- Model trees: as regression trees but a linear regression at each leaf
- Decision tree forests

```
YAL002W <= 0.015 :
| YBR002C <= 0.015 : YDR054C = 0.368
| YBR002C > 0.015 : YDR054C = 0.772
YAL002W > 0.015 :
| YAL041W <= -0.035 : YDR054C = -0.392
| YAL041W > -0.035 : YDR054C = -1.04
```

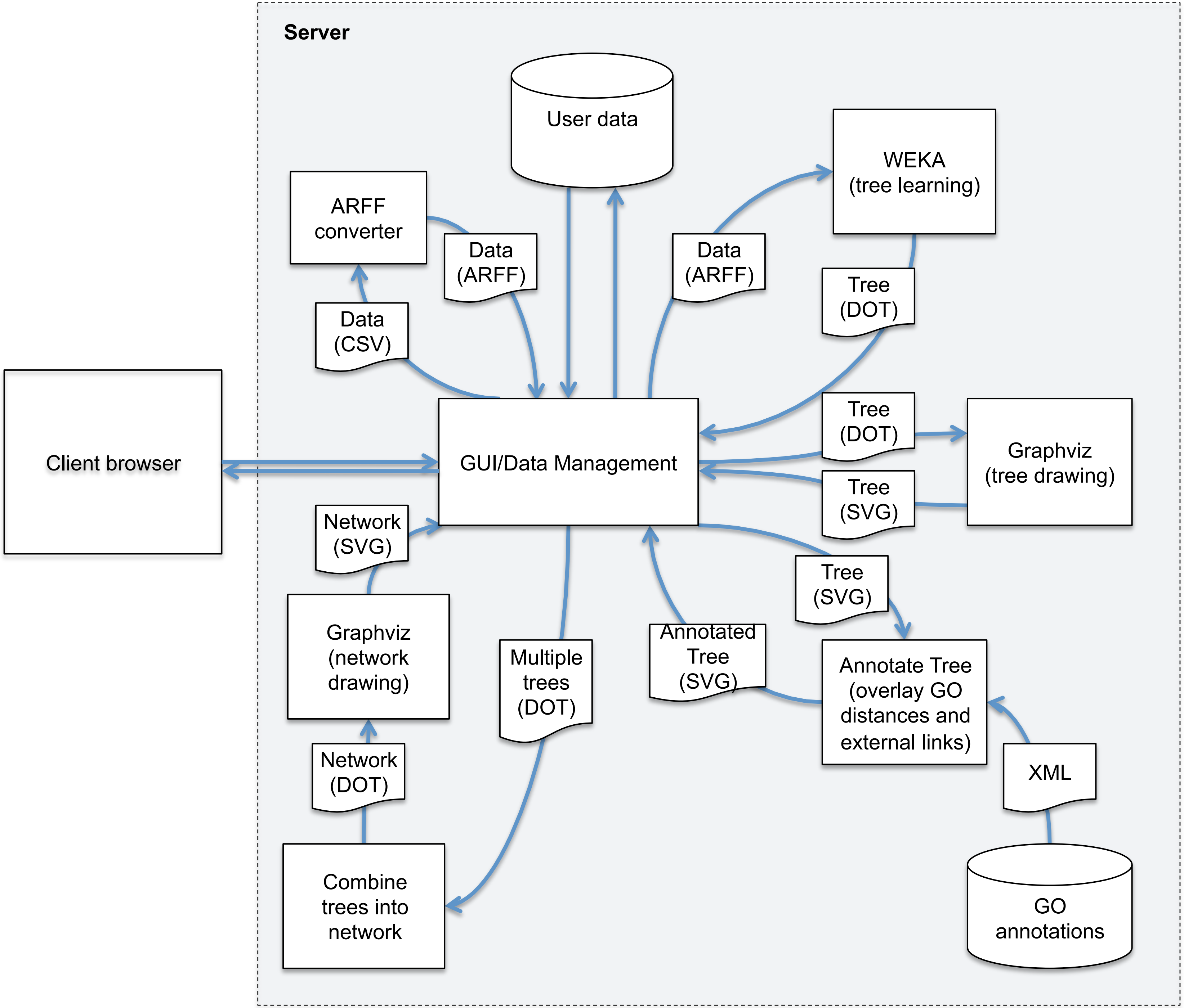


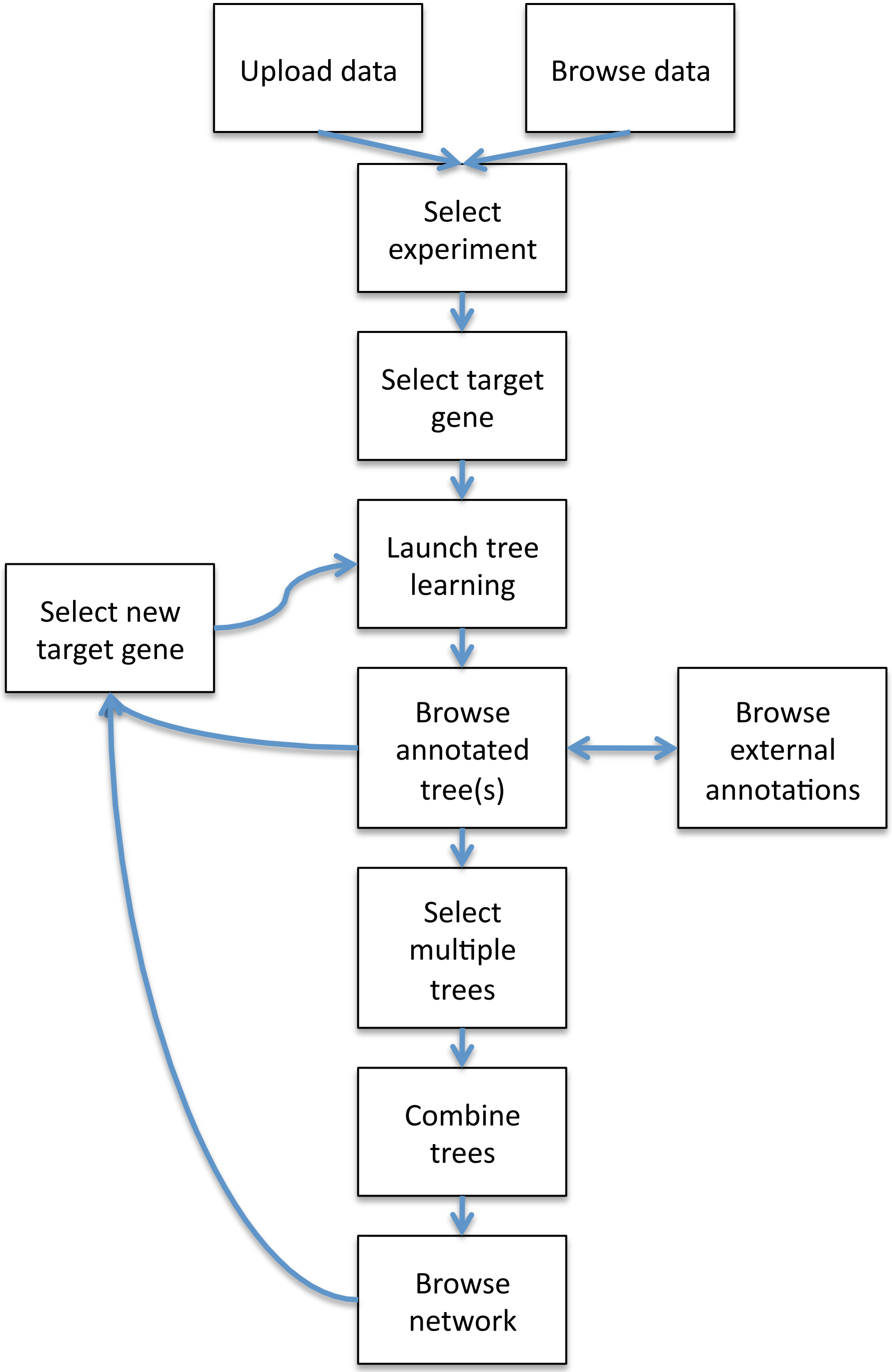
MINER

- A user interface paradigm that allows the use of tree learning to explore potential relationships in gene expression data
- Building the network from the ground up rather than top-down

MINER

- Microarray
- Interactive
- Network
- Exploration
- Representation


Server
User data
WEKA (tree learning)
ARFF converter
Data (ARFF)
Data (ARFF)
Data (CSV)
Tree (DOT)
Client browser
GUI/Data Management
Tree (DOT)
Graphviz (tree drawing)
Tree (SVG)
Network (SVG)
Tree (SVG)
Graphviz (network drawing)
Multiple trees (DOT)
Annotated Tree (SVG)
Annotate Tree (overlay GO distances and external links)
XML
Network (DOT)
Combine trees into network
GO annotations


Upload data
Browse data
Select experiment
Select target gene
Launch tree learning
Select new target gene
Browse annotated tree(s)
Browse external annotations
Select multiple trees
Combine trees
Browse network

You are here: Home > MINER

MINER – Pick an experiment

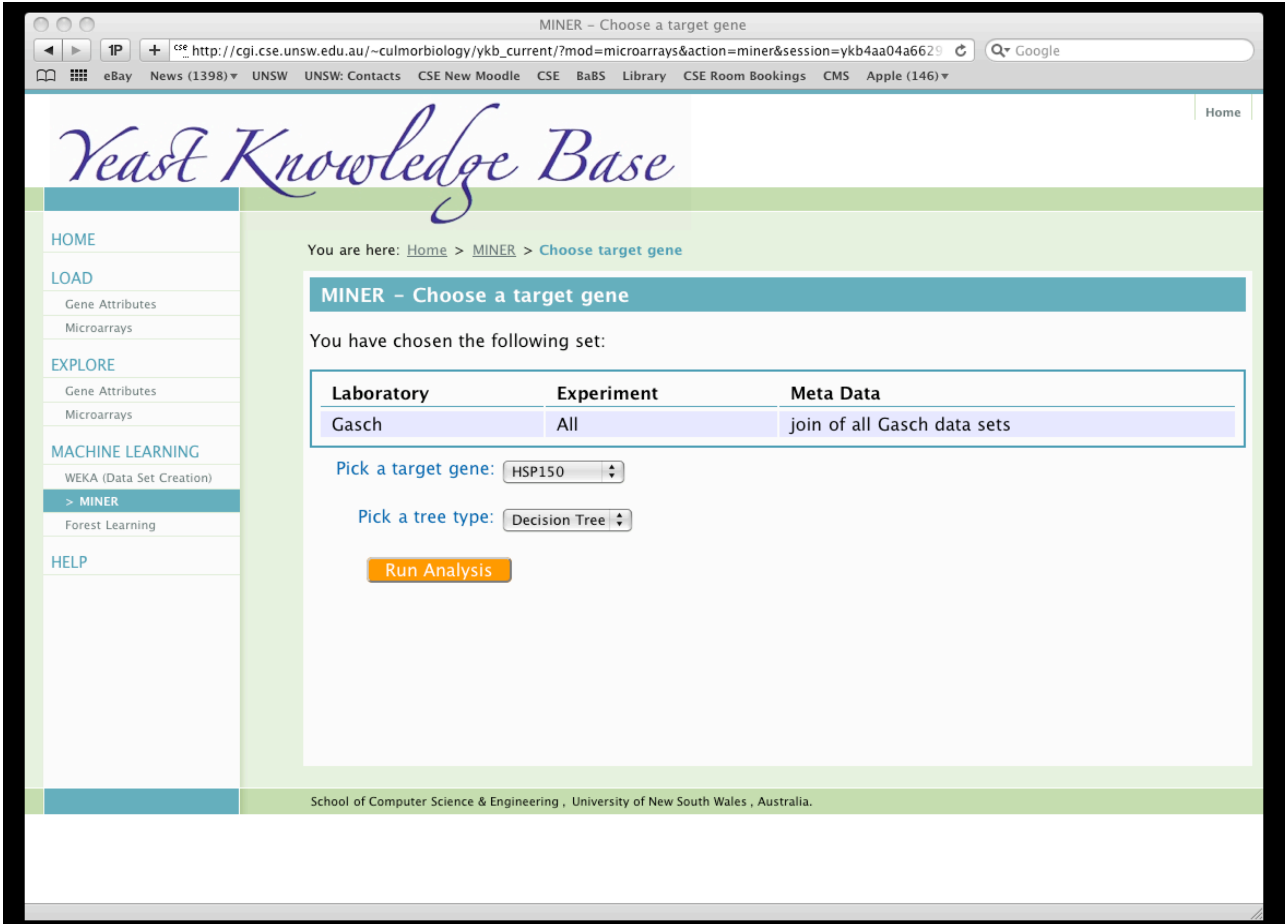
Please choose a microarray experiment set:

Show 10 entries

Search:

Laboratory	Experiment	Meta Data	Select
Causton	Acid		●
Causton	Alkali		○
Causton	H2O2		○
Causton	Heat		○
Causton	Salt		○
Causton	Sorbitol		○
Causton	All	join of all causton data sets	○
Cyert	Expt 1 – Ca2+	"Genome-wide Analysis of Gene Expression Regulated by the Calcineurin/Crz1p Signaling Pathway in Saccharomyces cerevisiae", J. Biol. Chemistry, Yoshimoto et al. (2002)	○
Cyert	Expt 2 – Ca2+ plus FK506	"Genome-wide Analysis of Gene Expression Regulated by the Calcineurin/Crz1p Signaling Pathway in Saccharomyces cerevisiae", J. Biol. Chemistry, Yoshimoto et al. (2002)	○
Gasch	Diamide		○

Showing 1 to 10 of 19 entries



You are here: Home > MINER > Choose target gene

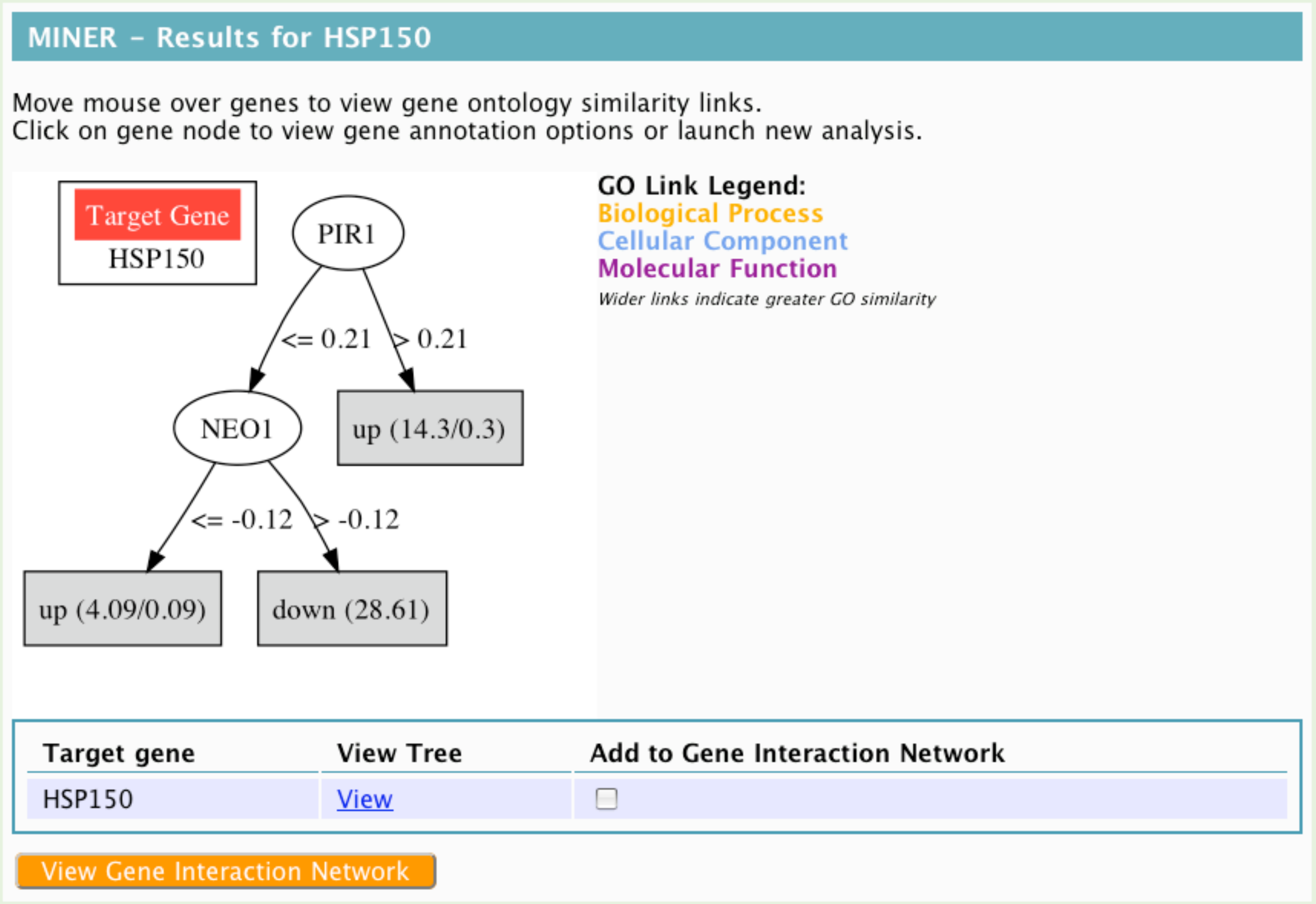
MINER – Choose a target gene

You have chosen the following set:

Laboratory	Experiment	Meta Data
Gasch	All	join of all Gasch data sets

Pick a target gene: HSP150

Pick a tree type: Decision Tree

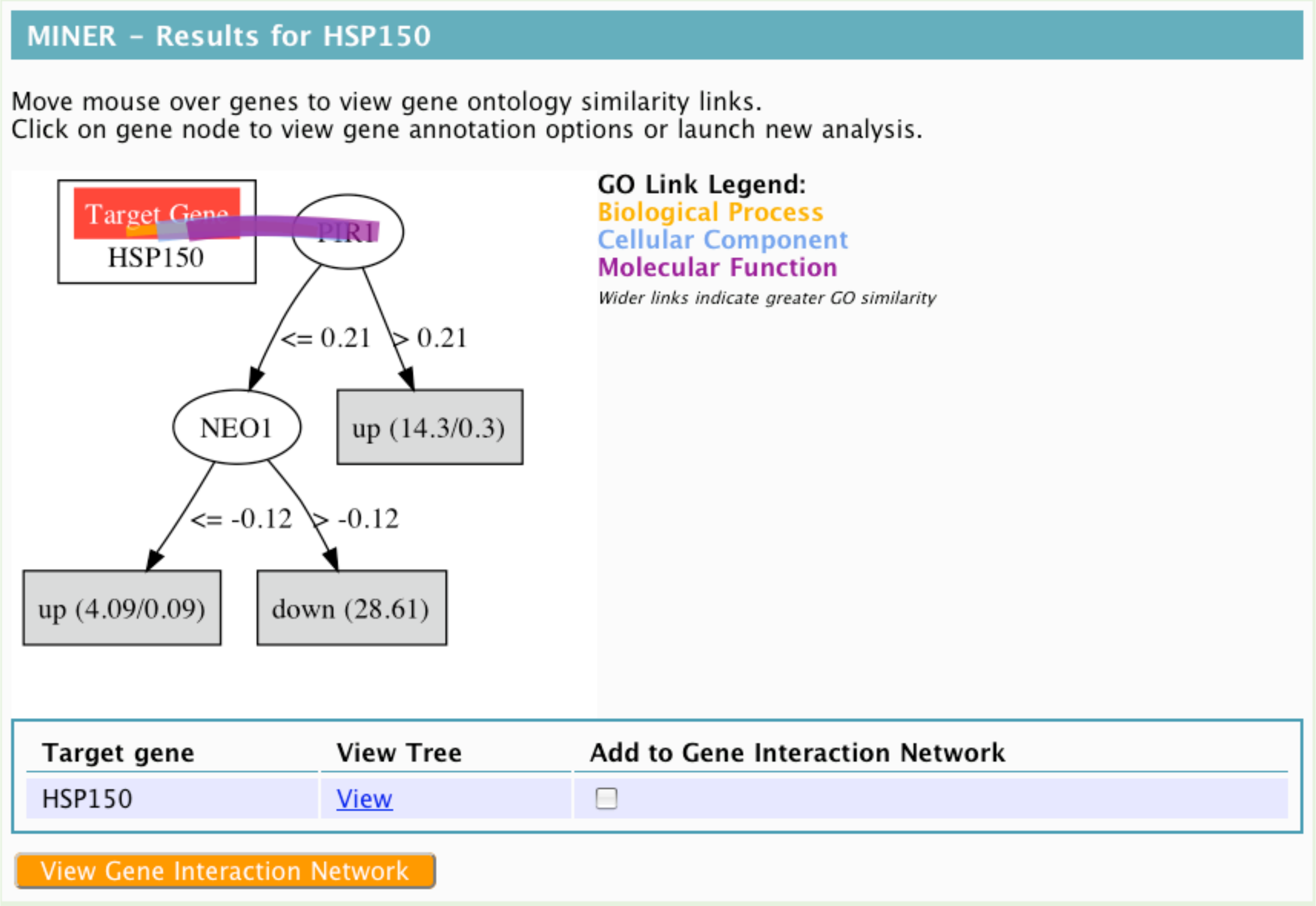


MINER - Results for HSP150

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐

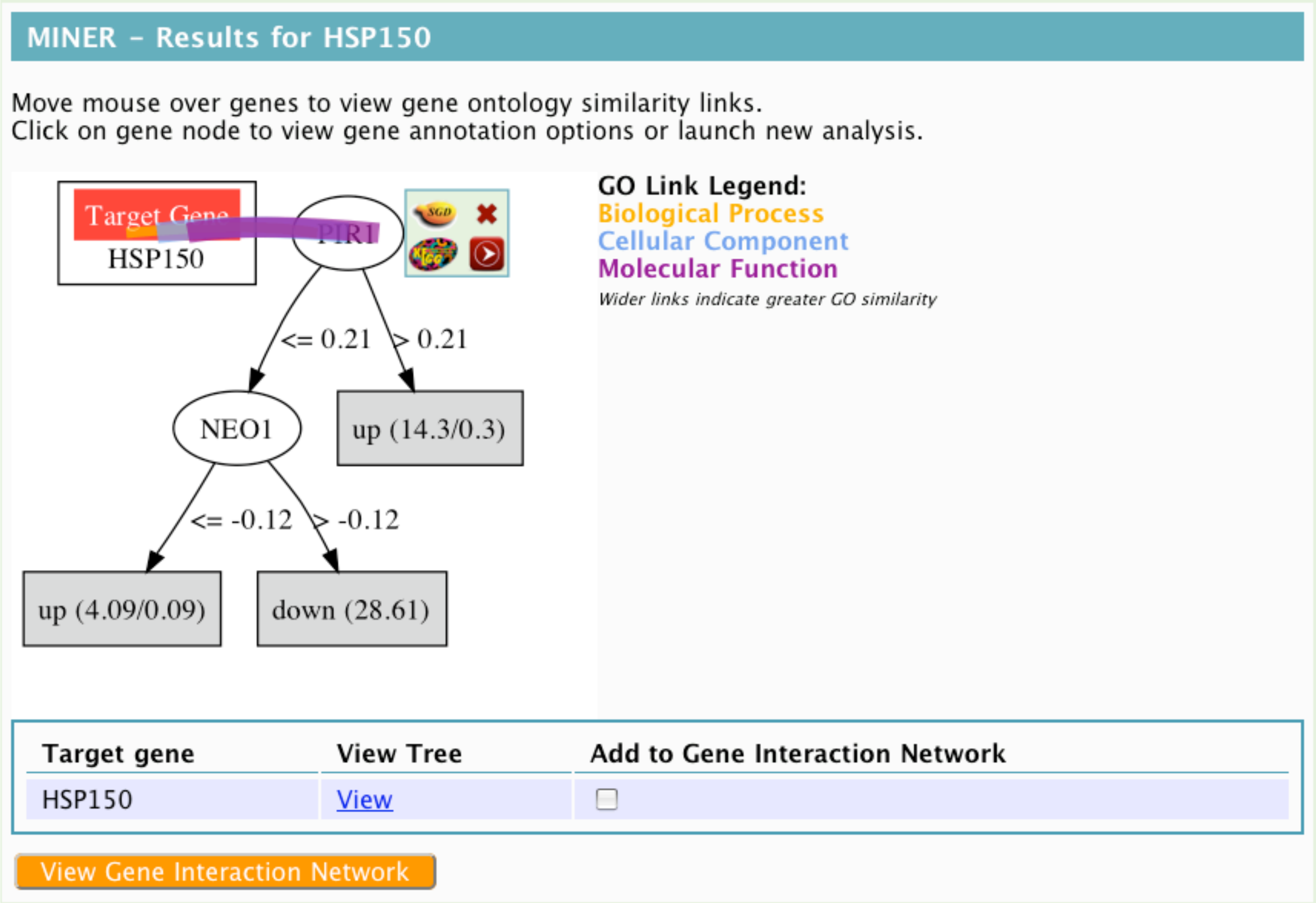


MINER - Results for HSP150

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

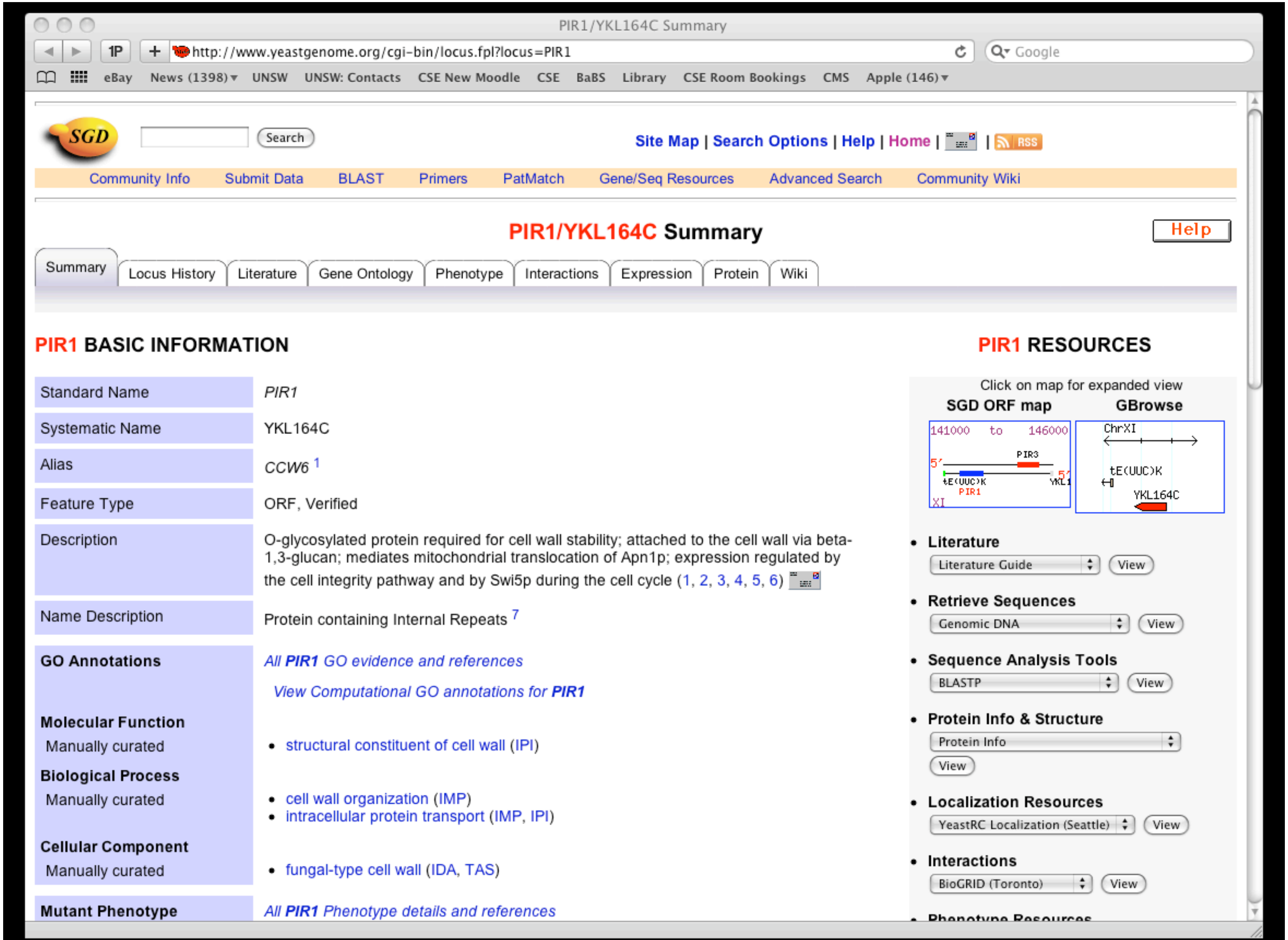
Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐



Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐



PIR1/YKL164C Summary

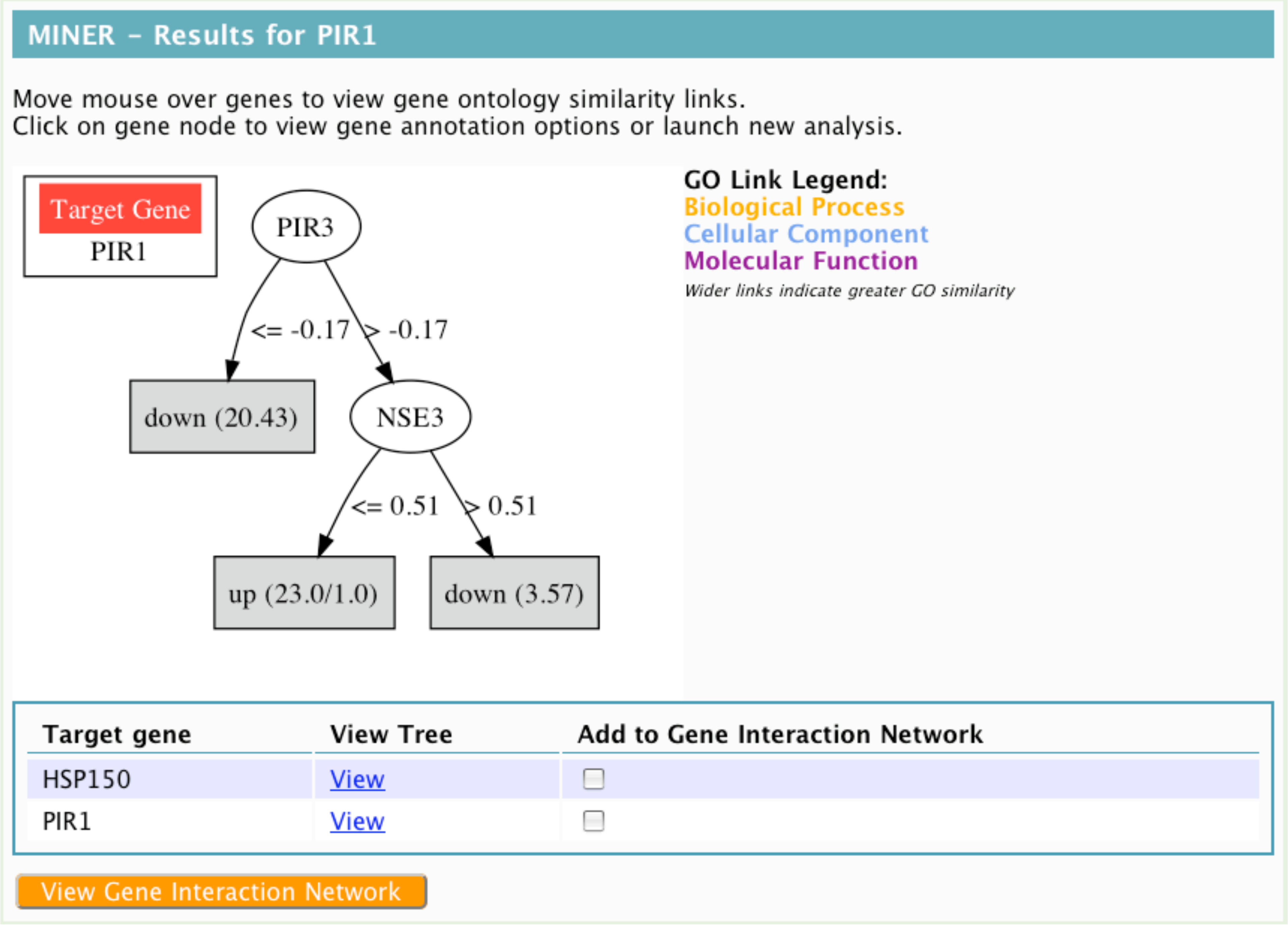
PIR1 BASIC INFORMATION

Standard Name	*PIR1*
Systematic Name	YKL164C
Alias	*CCW6* [1]
Feature Type	ORF, Verified
Description	O-glycosylated protein required for cell wall stability; attached to the cell wall via beta-1,3-glucan; mediates mitochondrial translocation of Apn1p; expression regulated by the cell integrity pathway and by Swi5p during the cell cycle (1, 2, 3, 4, 5, 6)
Name Description	Protein containing Internal Repeats [7]
GO Annotations	*All **PIR1** GO evidence and references* *View Computational GO annotations for **PIR1***
Molecular Function Manually curated	- structural constituent of cell wall (IPI)
Biological Process Manually curated	- cell wall organization (IMP) - intracellular protein transport (IMP, IPI)
Cellular Component Manually curated	- fungal-type cell wall (IDA, TAS)
Mutant Phenotype	*All **PIR1** Phenotype details and references*

PIR1 RESOURCES

Click on map for expanded view

- **Literature**
  Literature Guide — View
- **Retrieve Sequences**
  Genomic DNA — View
- **Sequence Analysis Tools**
  BLASTP — View
- **Protein Info & Structure**
  Protein Info — View
- **Localization Resources**
  YeastRC Localization (Seattle) — View
- **Interactions**
  BioGRID (Toronto) — View
- **Phenotype Resources**

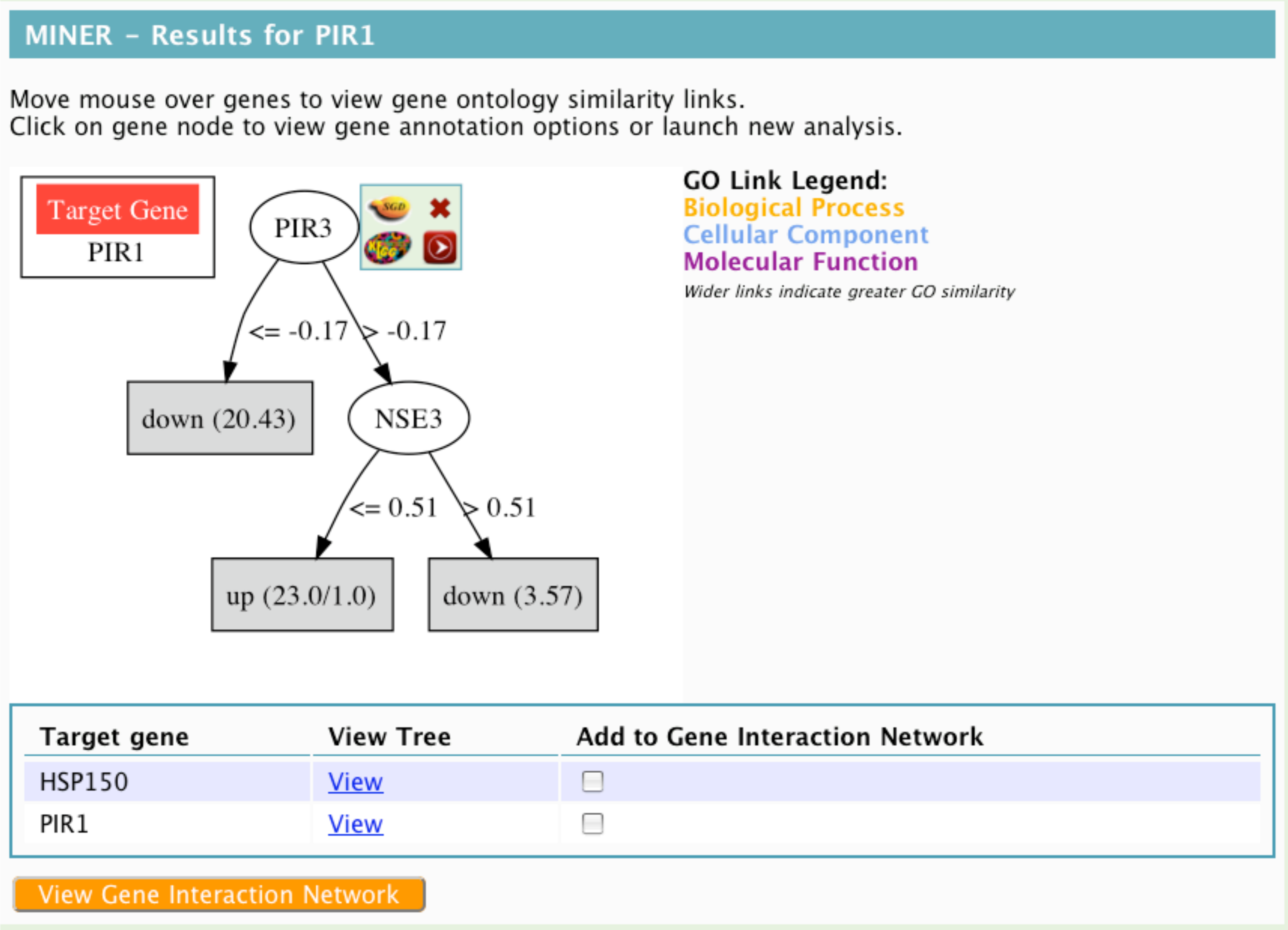


MINER – Results for PIR1

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐
PIR1	View	☐

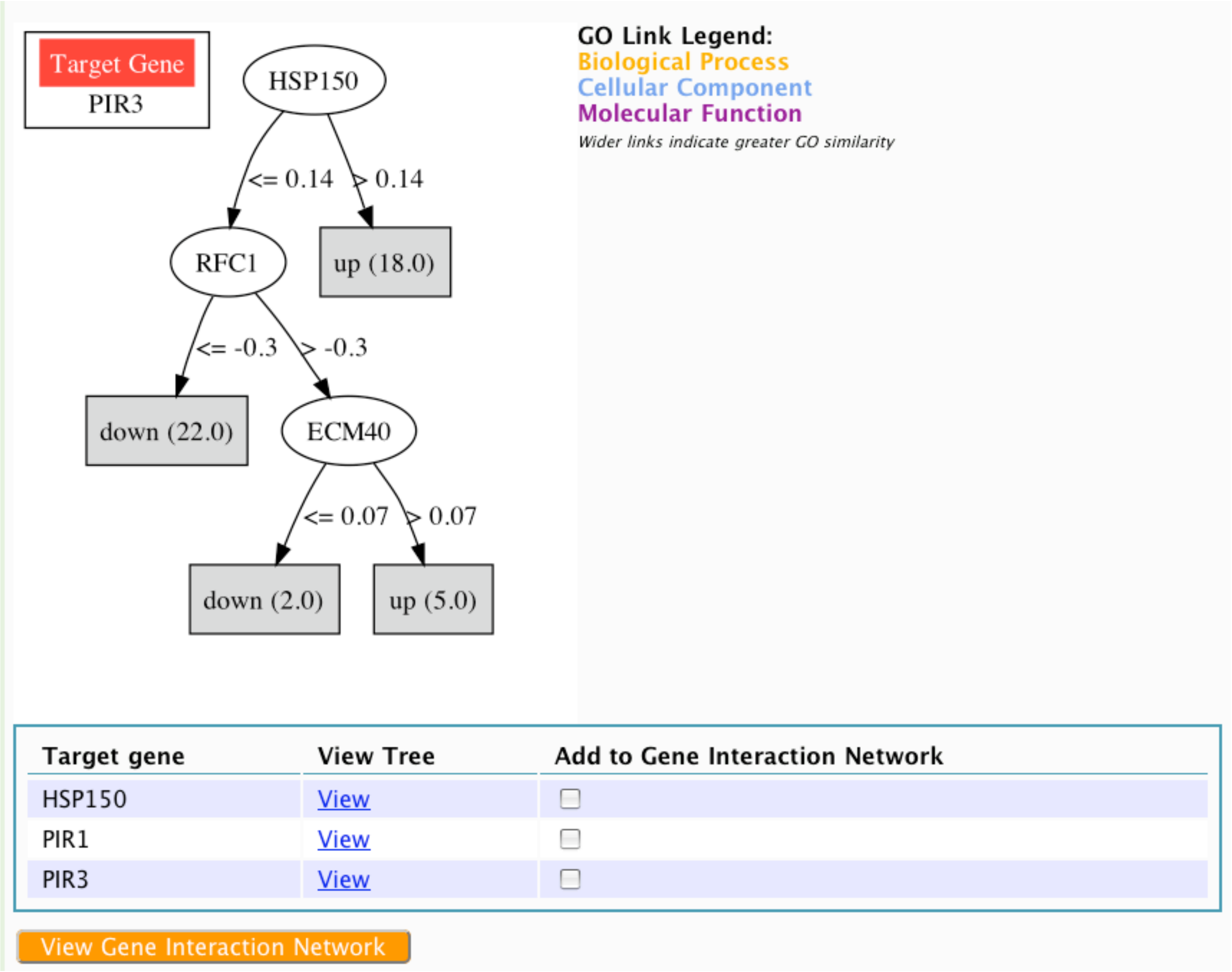


MINER - Results for PIR1

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

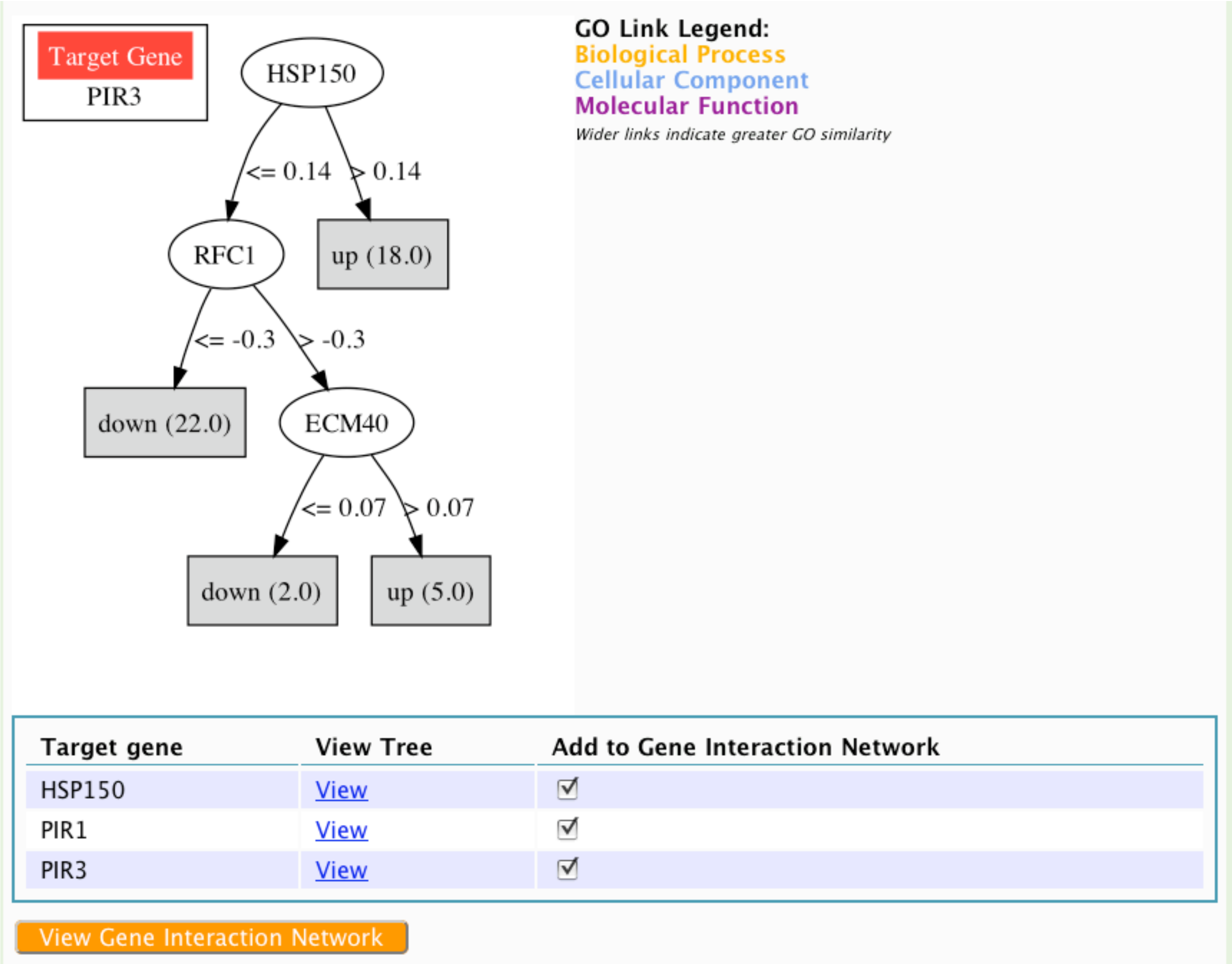
GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐
PIR1	View	☐



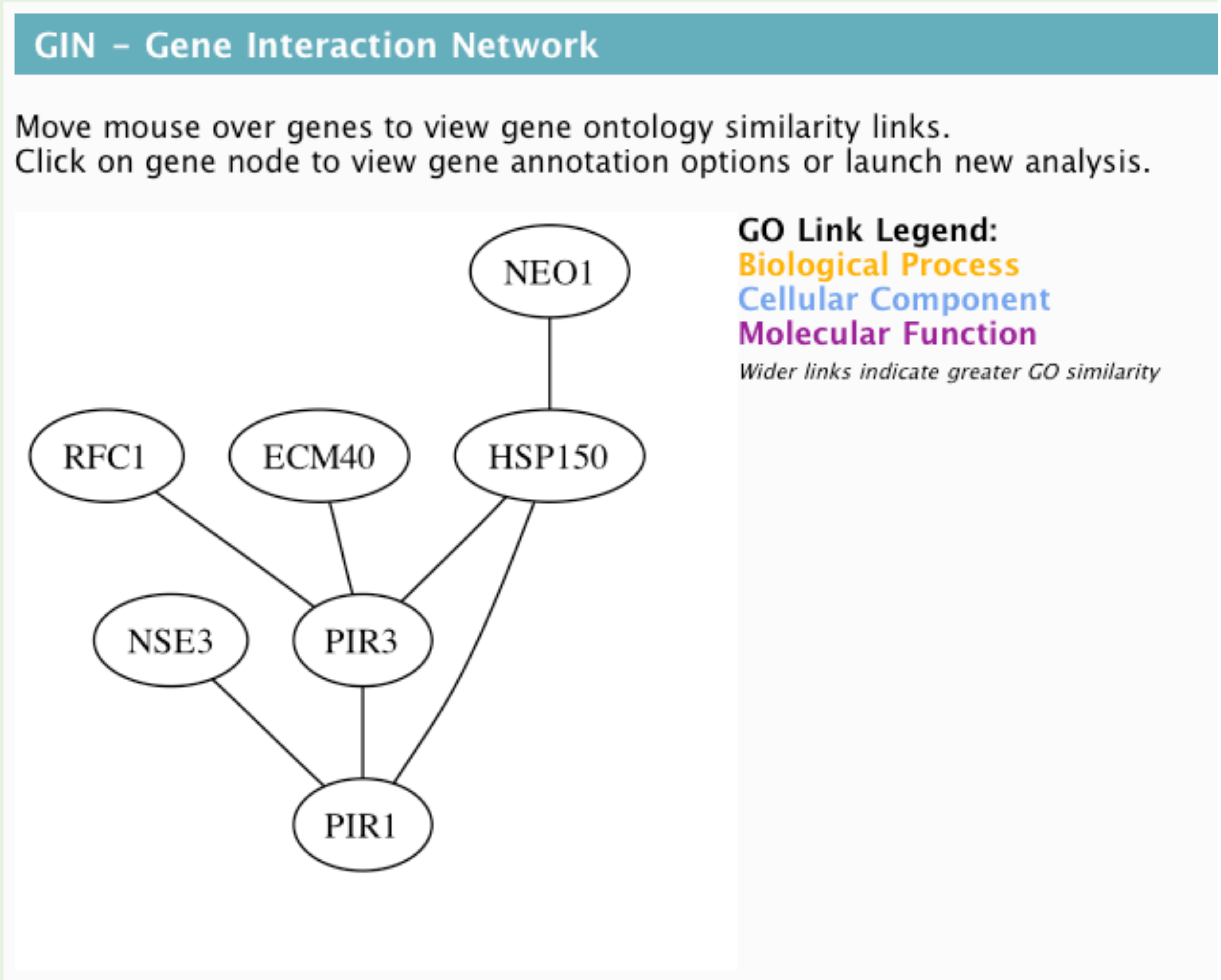
GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☐
PIR1	View	☐
PIR3	View	☐



GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity

Target gene	View Tree	Add to Gene Interaction Network
HSP150	View	☑
PIR1	View	☑
PIR3	View	☑

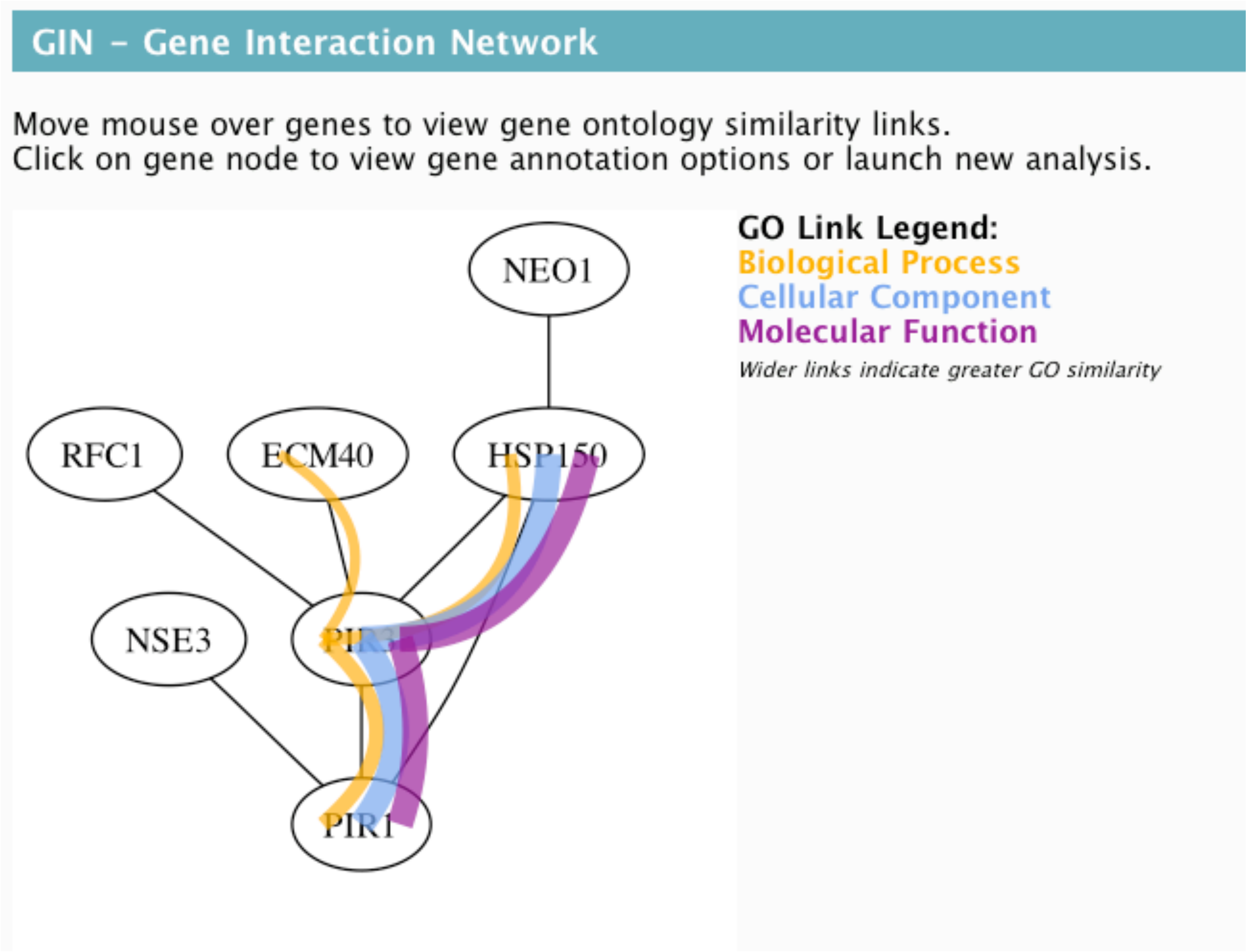


GIN – Gene Interaction Network

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function

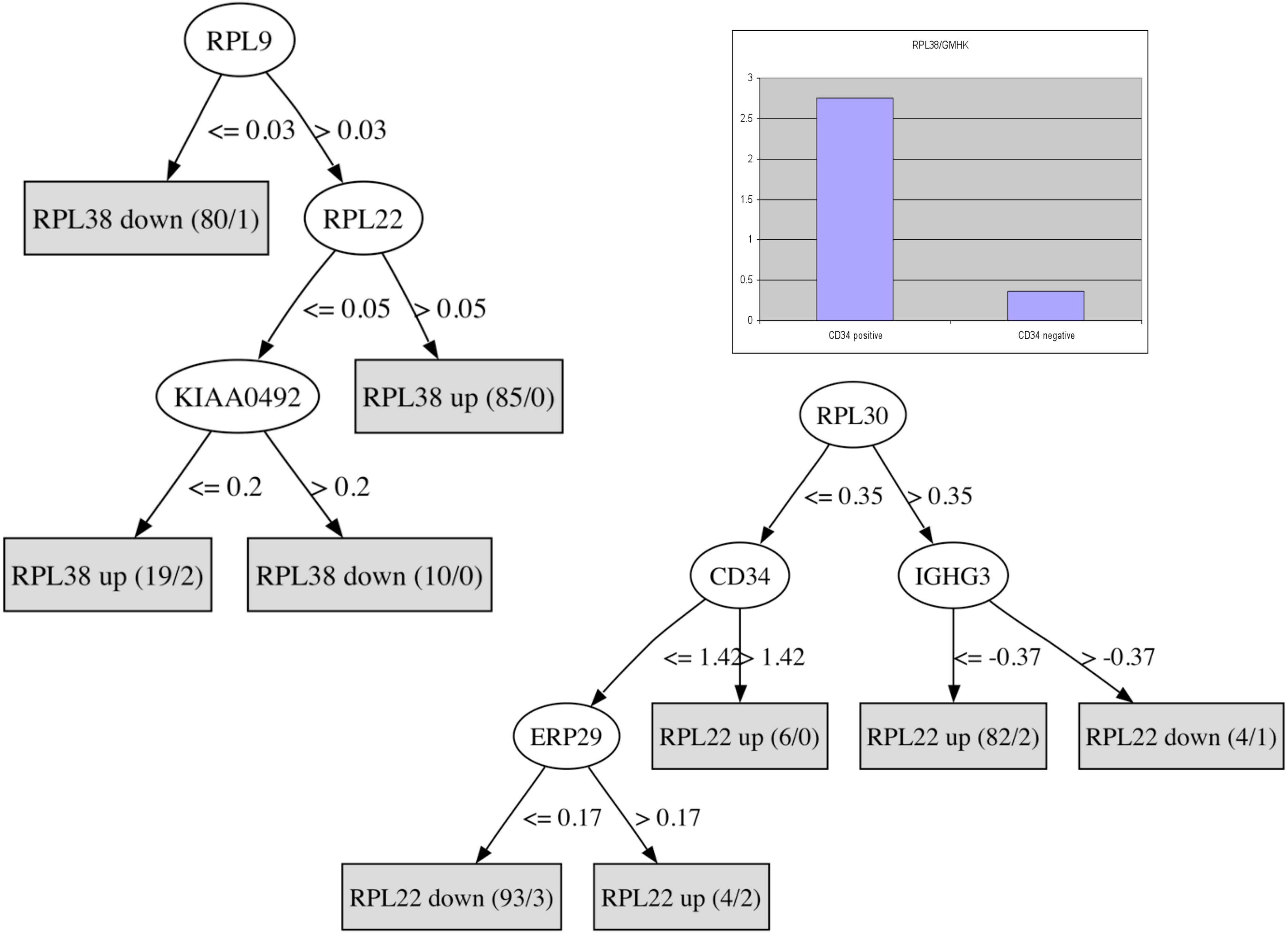
Wider links indicate greater GO similarity



GIN - Gene Interaction Network

Move mouse over genes to view gene ontology similarity links.
Click on gene node to view gene annotation options or launch new analysis.

GO Link Legend:
Biological Process
Cellular Component
Molecular Function
Wider links indicate greater GO similarity


RPL9
<= 0.03
> 0.03
RPL38 down (80/1)
RPL22
<= 0.05
> 0.05
KIAA0492
RPL38 up (85/0)
<= 0.2
> 0.2
RPL38 up (19/2)
RPL38 down (10/0)
RPL38/GMHK
3
2.5
2
1.5
1
0.5
0
CD34 positive
CD34 negative
RPL30
<= 0.35
> 0.35
CD34
IGHG3
<= 1.42
> 1.42
<= -0.37
> -0.37
ERP29
RPL22 up (6/0)
RPL22 up (82/2)
RPL22 down (4/1)
<= 0.17
> 0.17
RPL22 down (93/3)
RPL22 up (4/2)