The Gaggle

Connect ***R*** to Firefox and assorted Java programs

Paul Shannon

Institute for Systems Biology & Seattle Biomedical Research Institute

Overview

- Metaphors: goose, gaggle, boss, broadcast
- Four data types: nameList, matrix, network, hash map
- Semantic flexibility: the low road to data integration
- Firegoose

Motivating Example

- Identify up-regulated genes in R
- Broadcast Entrez GeneIDs to Firefox
- In Firefox, send them to EMBL STRING to explore for protein associations
- Broadcast expanded list and network back to ***R***

Cytokine time course

Locally released cytokines contribute to beta cell dysfunction and apoptosis in Type 1 diabetes. In vitro exposure of insulin producing INS 1E cells to the cytokines interleukin (IL) 1beta + interferon (IFN) gamma leads to a significant increase in apoptosis. To characterize the genetic networks implicated in beta cell dysfunction and apoptosis, we performed a time course analysis using the Affymetrix RG U34A microarry. INS 1E cells were exposed in duplicate to IL 1beta + IFN gamma for six different time points (1, 2, 4, 8, 12, and 24 h).

http://diabetes.diabetesjournals.org/cgi/content/full/52/11/2701

First Step, Always

Start the Gaggle Boss using Java Web Start:

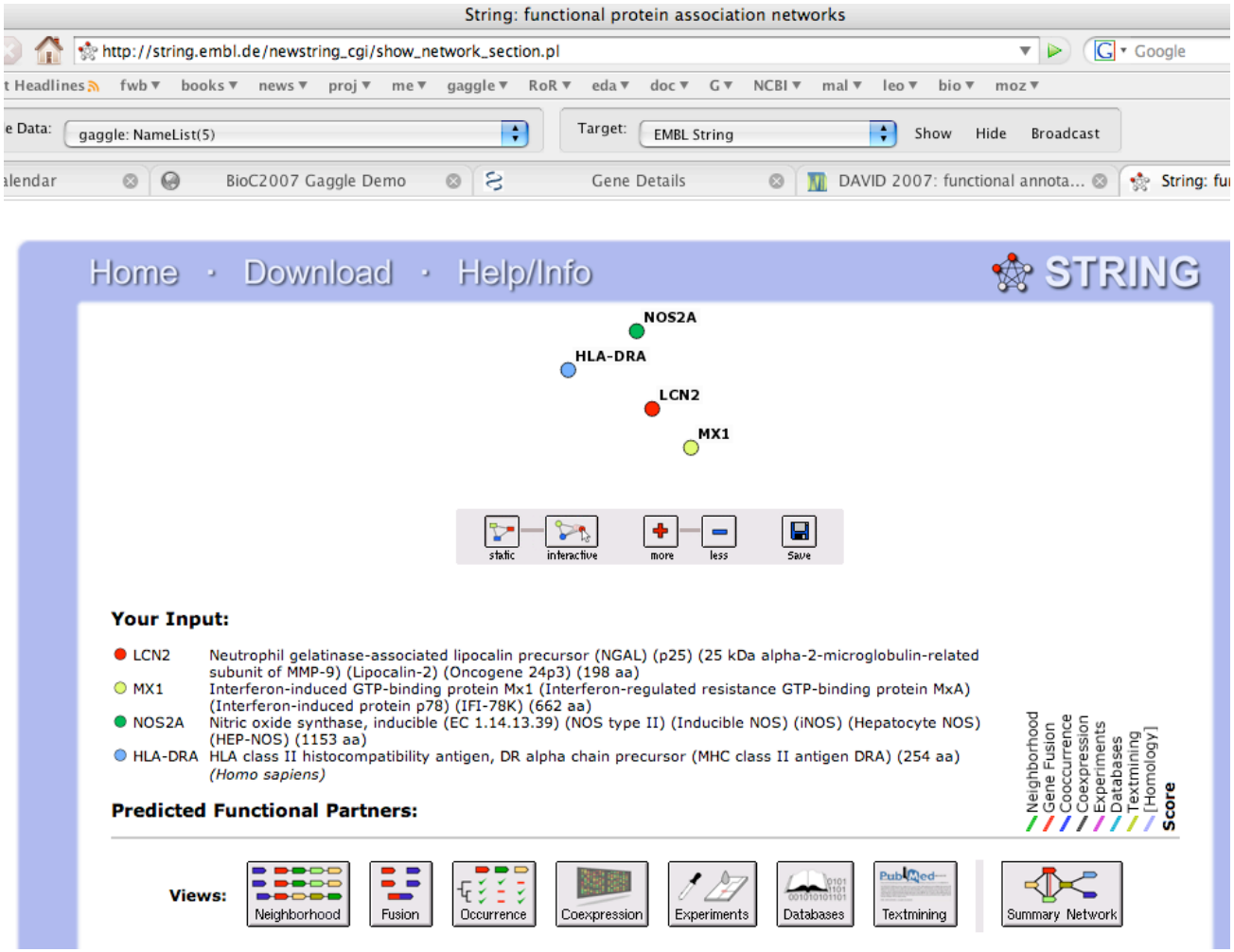
This, and other related web start links, may be found on the workshop web page:

http://gaggle.systemsbiology.org/pshannon/bioc2007/

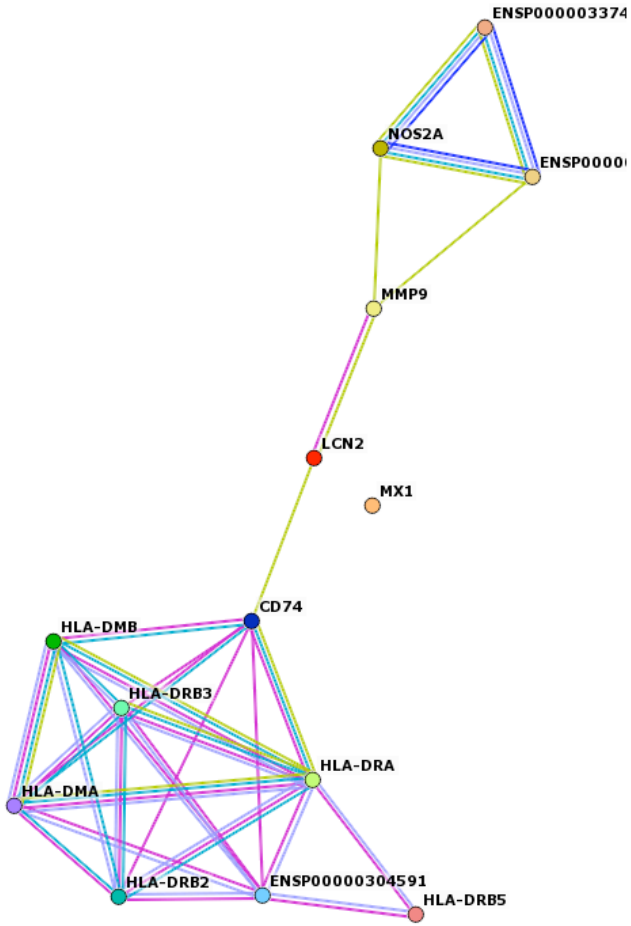
In *R*

```
> library (gaggle)
> gaggleInit ()
> m = read.table ('matrix.tsv', sep='\t')
> rows = which (apply (m, 1, function (row) IQR (row) > 5))
> genes = row.names (m) [rows]
> genes
   [1] "3122" "3934" "4599" "4843" "715”
> geese ()
   [1] "Network"   "Firegoose" "DMV"       "R-05”
> setTargetGoose (“Firegoose”)
> broadcast (genes)
```

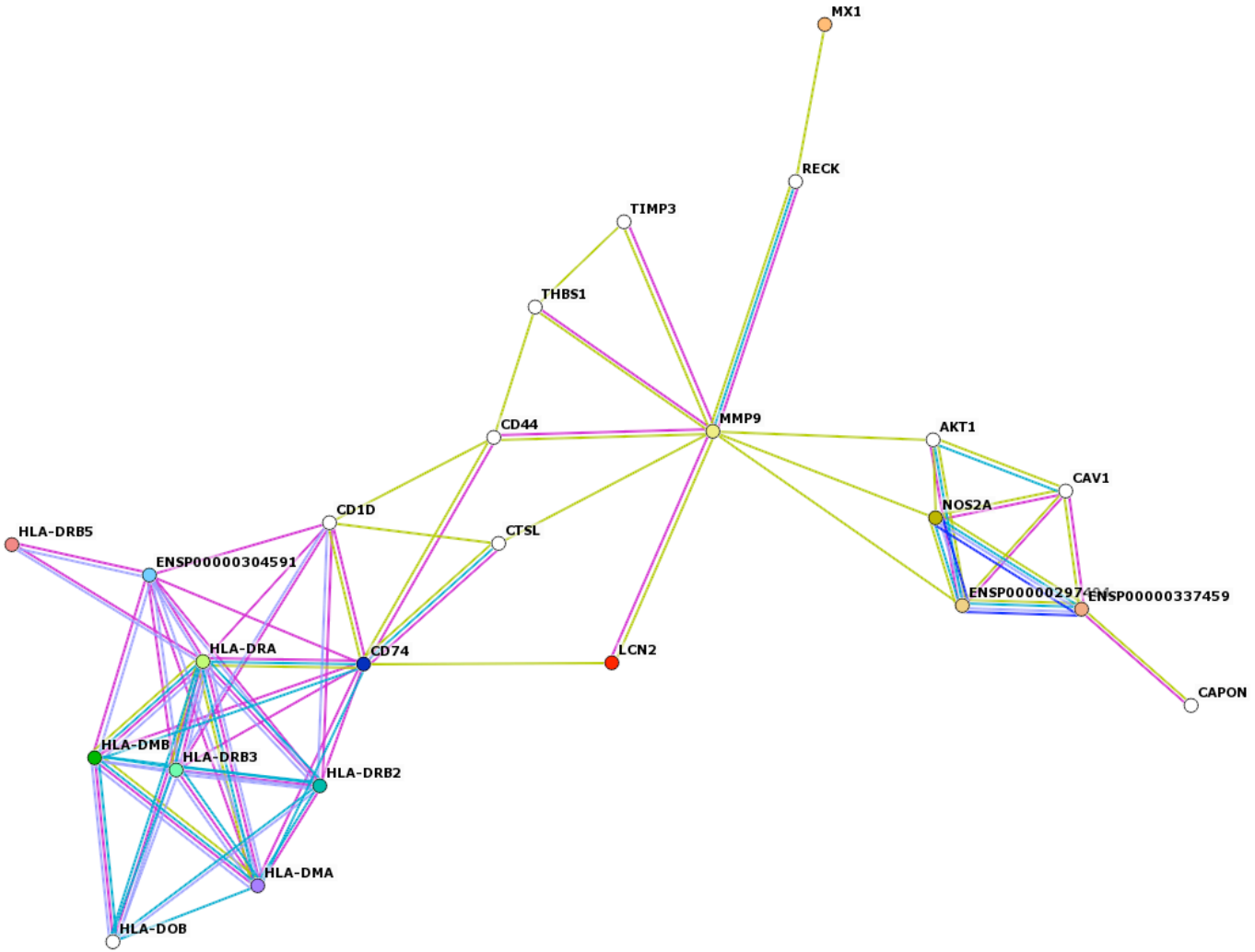
In Firefox: send to STRING



Find associations in STRING



Expand associations in STRING



Evidence in STRING

LCN2 *[ENSP00000277480]*	< - - - - - >	MMP9 *[ENSP00000216953]*
Neutrophil gelatinase-associated lipocalin precursor (NGAL) (p25) (25 kDa alpha-2-microglobulin-related subunit of MMP-9) (Lipocalin-2) (Oncogene 24p3)		Matrix metalloproteinase-9 precursor (MMP-9) (EC 3.4.24.35) (92 kDa type IV collagenase) (92 kDa gelatinase) (Gelatinase B) (GELB) [Contains: 67 kDa matrix metalloproteinase-9; 82 kDa matrix metalloproteinase-9]

Evidence suggesting a functional link:

Neighborhood in the Genome:	none / insignificant.	
Gene Fusions:	none / insignificant.	
Cooccurence Across Genomes:	none / insignificant.	
Co-Expression:	none / insignificant.	
Experimental/Biochemical Data:	yes (score 0.903).	Show
Association in Curated Databases:	none / insignificant.	
Co-Mentioned in PubMed Abstracts:	yes (score 0.900). In addition, putative homologs are mentioned together in 2 other species (score 0.092).	Show
Combined Score:	0.991	

Your Input:

LCN2	Neutrophil gelatinase-associated lipocalin precursor (NGAL) (p25) (25 kDa alpha-2-microglobulin-related subunit of MMP-9) (Lipocalin-2) (Oncogene 24p3) (198 aa)
MX1	Interferon-induced GTP-binding protein Mx1 (Interferon-regulated resistance GTP-binding protein MxA) (Interferon-induced protein p78) (IFI-78K) (662 aa)
NOS2A	Nitric oxide synthase, inducible (EC 1.14.13.39) (NOS type II) (Inducible NOS) (iNOS) (Hepatocyte NOS) (HEP-NOS) (1153 aa)
HLA-DRA	HLA class II histocompatibility antigen, DR alpha chain precursor (MHC class II antigen DRA) (254 aa) *(Homo sapiens)*

Predicted Functional Partners:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
HLA-DMB	HLA class II histocompatibility antigen, DM beta chain precursor (MHC class II antigen DMB) (263 aa)					●	●	●	●	0.999
HLA-DRB3	HLA class II histocompatibility antigen, DRB3-1 beta chain precursor (MHC class I antigen DRB3*1) (266 aa)					●	●	●	●	0.999
HLA-DRB2	HLA class II histocompatibility antigen, DW2.2/DR2.2 beta chain (Fragment) (266 aa)					●	●		●	0.999
ENSP00000304591	HLA class II histocompatibility antigen, DR alpha chain precursor (MHC class II antigen DRA) (254 aa)					●			●	0.999
CD74	HLA class II histocompatibility antigen, gamma chain (HLA-DR antigens associated invariant chain) (Ia antige					●	●	●		0.605

PubMed Abstracts from STRING

Relevant abstracts mentioning your query species (Homo sapiens):

[Expression of matrix metalloproteinase-9 and its complex in the urine of breast cancer patients].
Zhonghua Wai Ke Za Zhi (2003).
○ MMP-9 ● NGAL ...

[Functions of neutrophil gelatinase-associated lipocalin in the esophageal carcinoma cell line SHEEC].
Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) (2003).
○ MMP-9 ● NGAL ...

Microdeformational wound therapy: effects on angiogenesis and matrix metalloproteinases in chronic wounds of 3 debilitated patients.
Ann Plast Surg (2006).
○ MMP-9 ● NGAL ...

Neutrophil granule proteins in bronchoalveolar lavage fluid from subjects with subclinical emphysema.
Am J Respir Crit Care Med (1999).
○ MMP-9, gelatinase B ● HNL ...

The human neutrophil lipocalin supports the allosteric activation of matrix metalloproteinases.
Eur J Biochem (2001).

The human neutrophil lipocalin (HNL (●)), a member of the large family of lipocalins that exhibit various physiological functions, is coexpressed in granulocytes with progelatinase B (MMP-9 (○)). Part of it is covalently bound to the proenzyme and therefore may play a possible role in the activation process of promatrix metalloproteinases. We now report that HNL (●) is able to accelerate the direct activation of promatrix metalloproteinases slightly. A significant enhancement of the activity could be demonstrated for the HgCl2- and the plasma kallikrein-induced activation of all three secretory forms of proMMP-9 and of proMMP-8. The same activating effects were exerted by HNL (●) isolated from granulocytes as well as by the recombinant forms expressed by the yeast Pichia pastoris or by Escherichia coli. This demonstrates that the carbohydrate moiety is not essential for the biological activity of HNL (●). Activation and activity enhancement are obviously mediated by entrapping the remaining N-terminal sequence residues of the partially truncated proenzyme into the hydrophobic binding pocket of the HNL (●). In conclusion these results document that HNL (●) can exert an enzyme-activating effect in the regulation of inflammatory and pathophysiological responses of granulocytes in the physiological activation of MMPs that have been subject to limited proteolytic processing.

Sustained activation of neutrophils in the course of Kawasaki disease: an association with matrix metalloproteinases.
Clin Exp Immunol (2005).
○ MMP9 ● NGAL ...

Gelatinase isoforms in urine from bladder cancer patients.
Clin Chim Acta (2000).
○ MMP-9 ● NGAL ...

Human neutrophil gelatinase and associated lipocalin in adult and localized juvenile periodontitis.
J Dent Res (1996).
○ MMP-9 ● NGAL ...

Up-regulation of the extracellular matrix remodeling genes, biglycan, neutrophil gelatinase-associated lipocalin, and matrix metalloproteinase-9 in familial amyloid polyneuropathy.
FASEB J (2005).

Broadcast network members back to ***R***

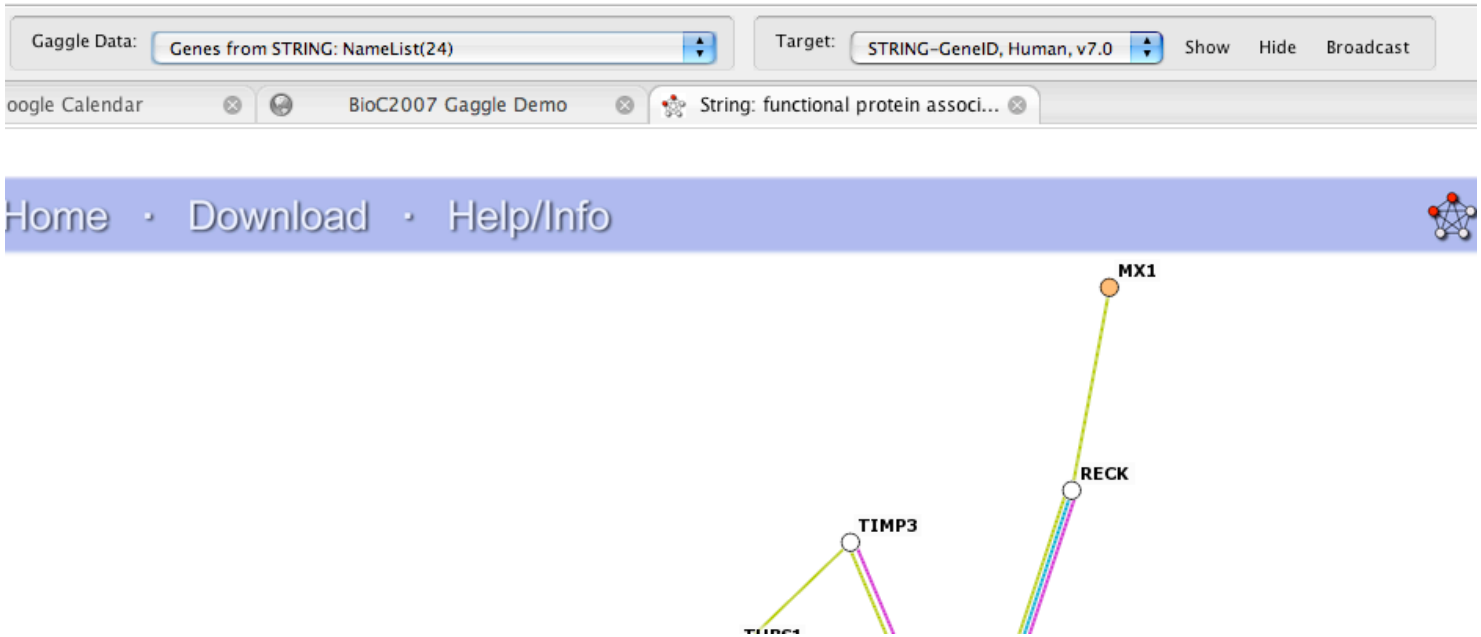
```
 nameList ready, length 24
> from.string = getNameList ()
> from.string
 [1] "ENSP00000346631" "ENSP00000303017" "ENSP00000257498" "ENSP00000299785"
 [5] "ENSP00000293722" "ENSP00000260356" "ENSP00000304591" "ENSP00000289425"
 [9] "ENSP00000339191" "ENSP00000355133" "ENSP00000327251" "ENSP00000242287"
[13] "ENSP00000277480" "ENSP00000266085" "ENSP00000216953" "ENSP00000229825"
[17] "ENSP00000302517" "ENSP00000353099" "ENSP00000297494" "ENSP00000270202"
[21] "ENSP00000288383" "ENSP00000278385" "ENSP00000337459" "ENSP00000009530"
```

A Name Translation Goose

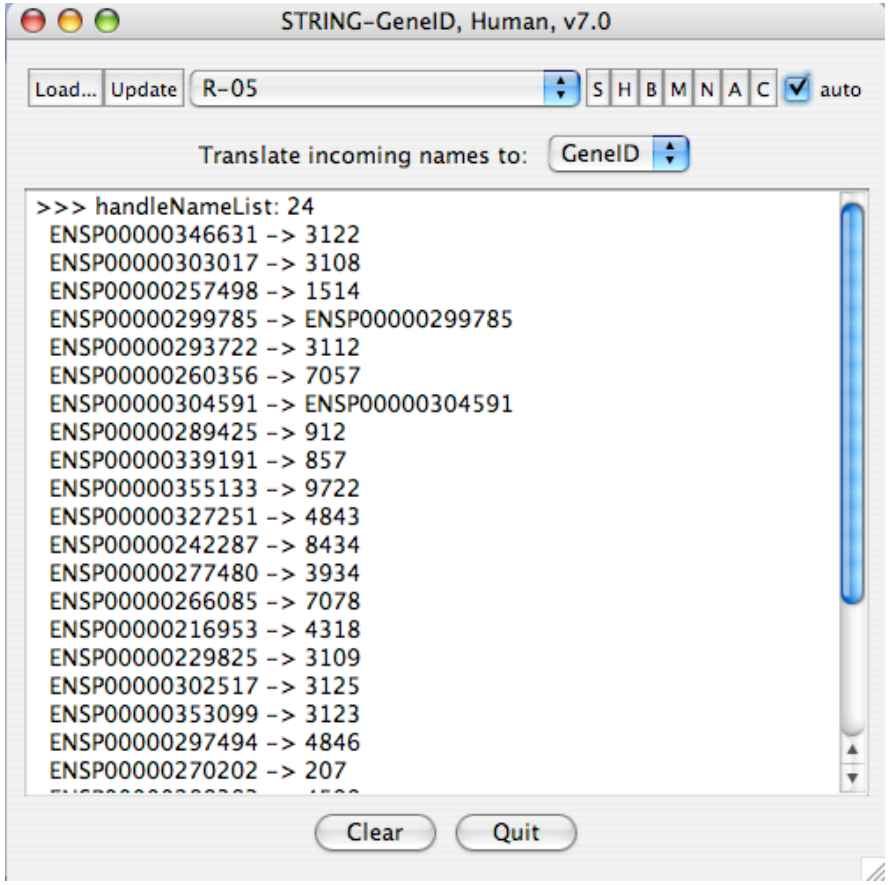
http://gaggle.systemsbiology.org/nameTranslations/humanStringToGeneID.70.jnlp

- Launch with Java Web Start
- Broadcast from Firefox/STRING to this goose
- Set up this goose to automatically broadcast translated names to R
- A good collection of these translators can solve many of the identifier problem bioinformaticians face

Name Translation Goose:
Setup Firefox



Name Translation Goose
In Action



Back in *R*

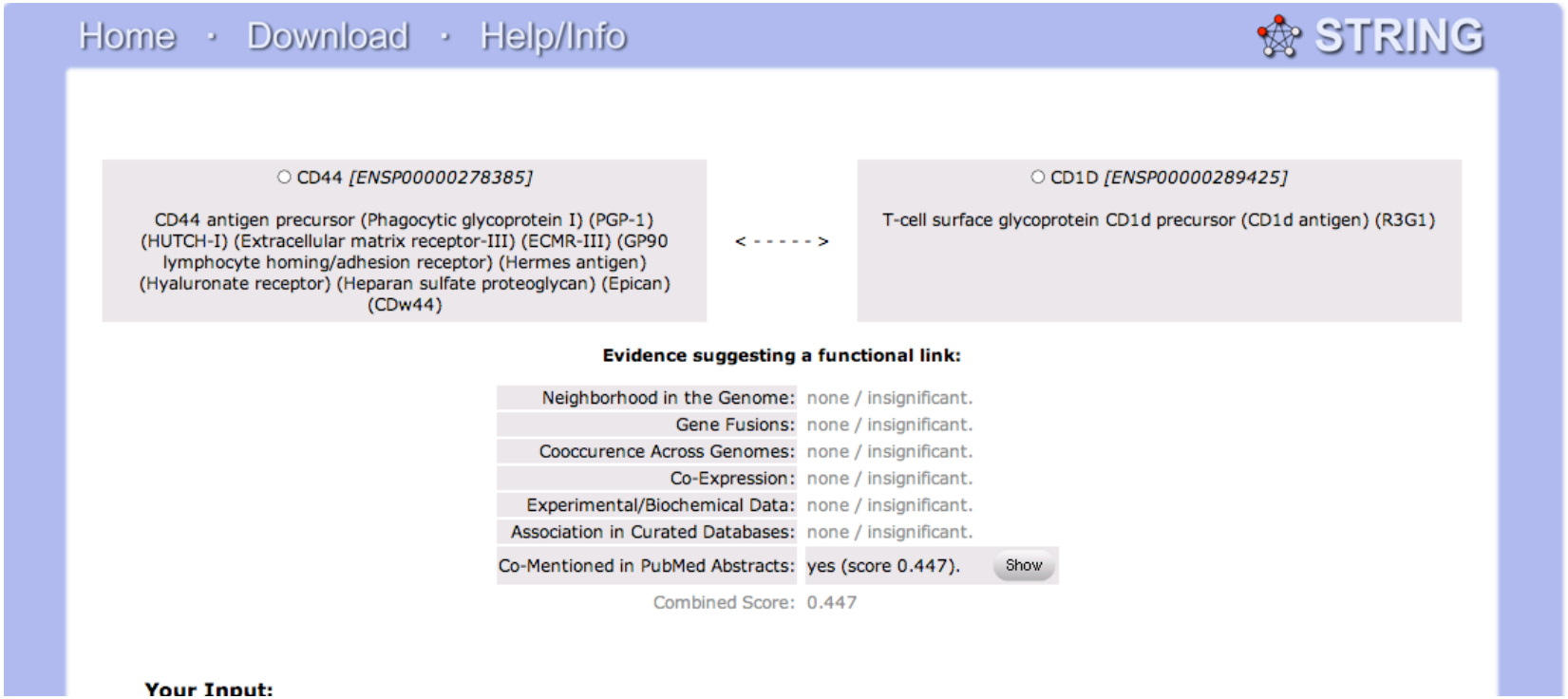
```
> from.string = getNameList ()
> from.string
  [1] "3122"           "3108"           "1514"                  "ENSP00000299785"
  [5] "3112"           "7057"           "ENSP00000304591"       "912"
  [9] "857"            "9722"           "4843"                  "8434"
 [13] "3934"           "7078"           "4318"                  "3109"
 [17] "3125"           "3123"           "4846"                  "207"
 [21] "4599"           "960"            "4842"                  "972"

> length (from.string)
   [1] 24

> length (intersect (from.string, rownames (m)))  # suggested by STRING, in our data
   [1] 13

> new.genes = setdiff (intersect (from.string, rownames (m)), genes)

> sapply (new.genes, function (row) IQR (m [row, ]))
     3108       3112        912       7078       3109       4846        207        960        972
0.6989106  0.2881418  0.6021836  0.1142142  3.8377716  0.9327259  0.3628950  0.0000000  4.1485932
```

Get STRING associations to *R*

```
  network ready, node count 24, edges: 60
> network = getNetwork ()

> edgeData (network) [1]
$`912|960`
$`912|960`$edgeType
[1] "embl string"

$`912|960`$`embl string url`
[1]
"http://string.embl.de/newstring_cgi/show_edge_data.pl?taskId=Ehq4jercyivn&node1=313
732&node2=312768"

$`912|960`$`combined association score`
[1] 0.447

> browseURL (edgeData (network)[[1]][[2]])
```

STRING evidence from *R:*

`browseURL (edgeData (network)[[1]][[2]])`



An Expression Movie

- Overlay (and animate) time-course expression levels on network derived from STRING
- Requires two new geese: DMV and Cytosape
- Cytoscape 2.5 *almost* ready; but using ancient Cytoscape 1.2 for now
- Begin by broadcasting matrix from ***R*** to DMV

An Expression Movie, cont.

Start DMV:

http://gaggle.systemsbiology.net/2005-11/dmv.jnlp

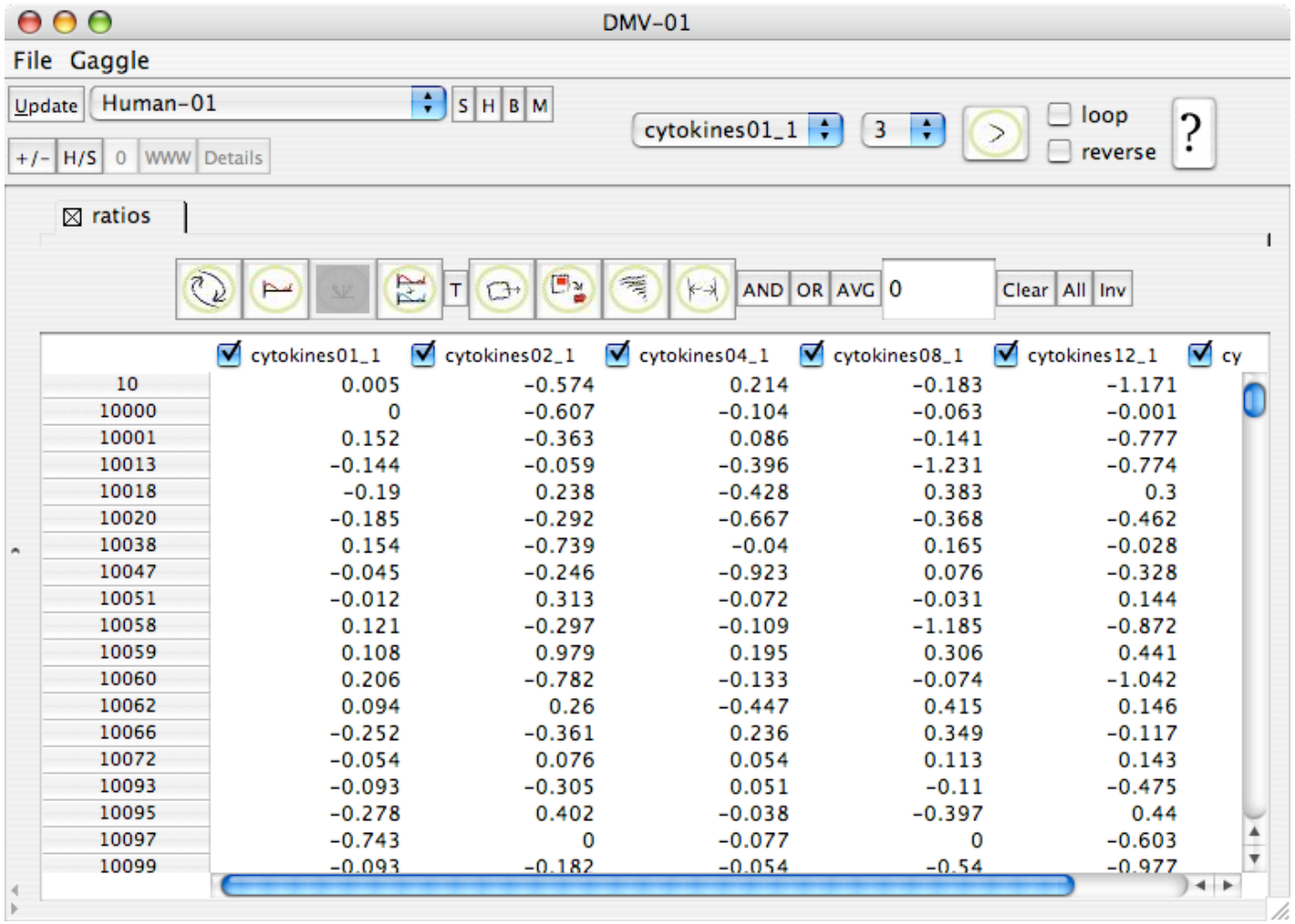
Start Cytoscape:

http://gaggle.systemsbiology.org/pshannon/cy12/blankSlate/human/cy.jnlp

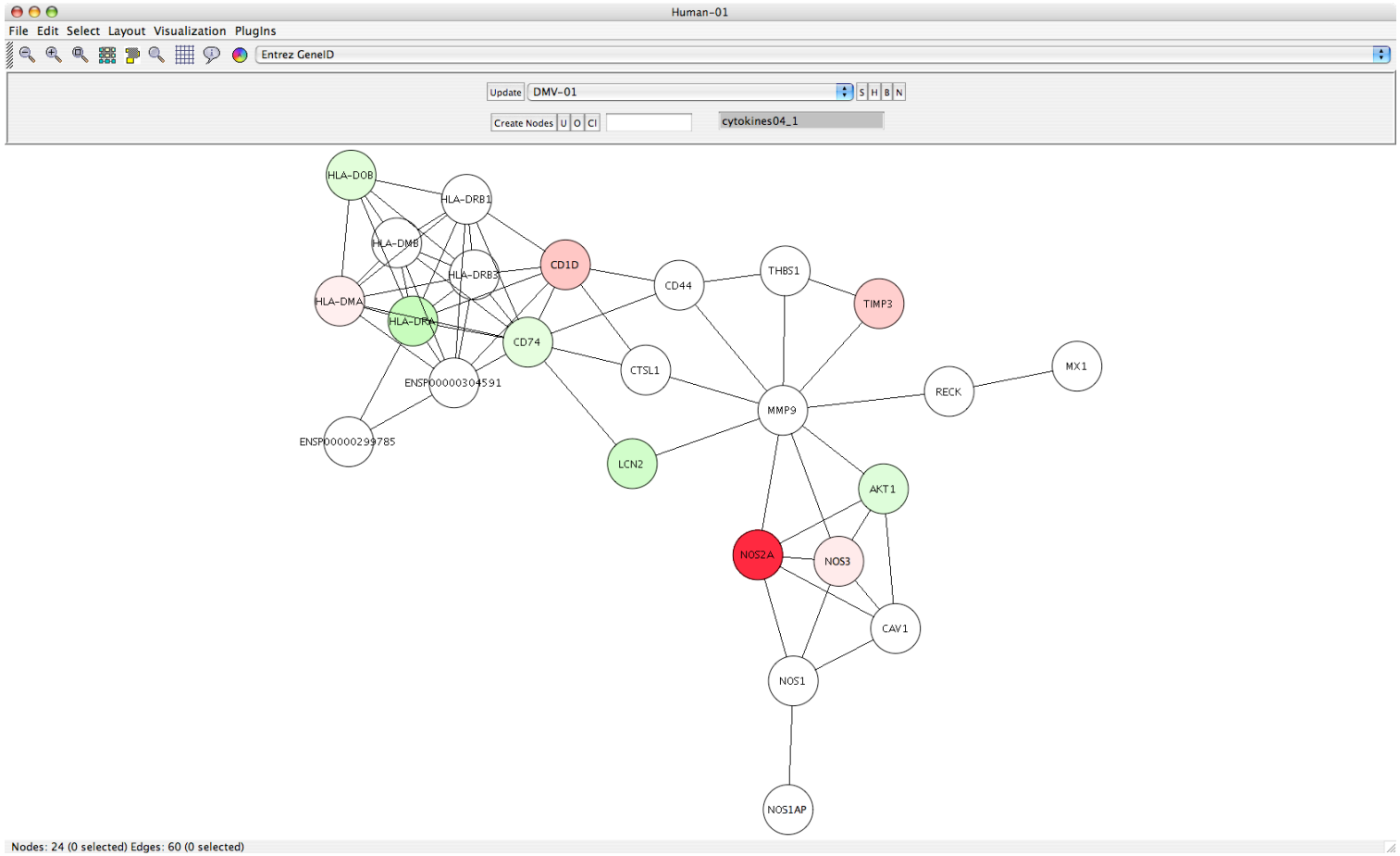
An Expression Movie, cont.

```
> geese ()
[1] "Human-01"                          "DMV-01"
[3] "Network"                           "Firegoose"
[5] "STRING-GeneID, Human, v7.0”    "R-05”

> setTargetGoose (‘DMV-01’)   # or stg (2)

> broadcast (m, ‘ratios’)

> setTargetGoose (‘Human-01’) # or stg (1)

> broadcast (network)

> stg (2); showGoose ()   # bring DMV to the front
```

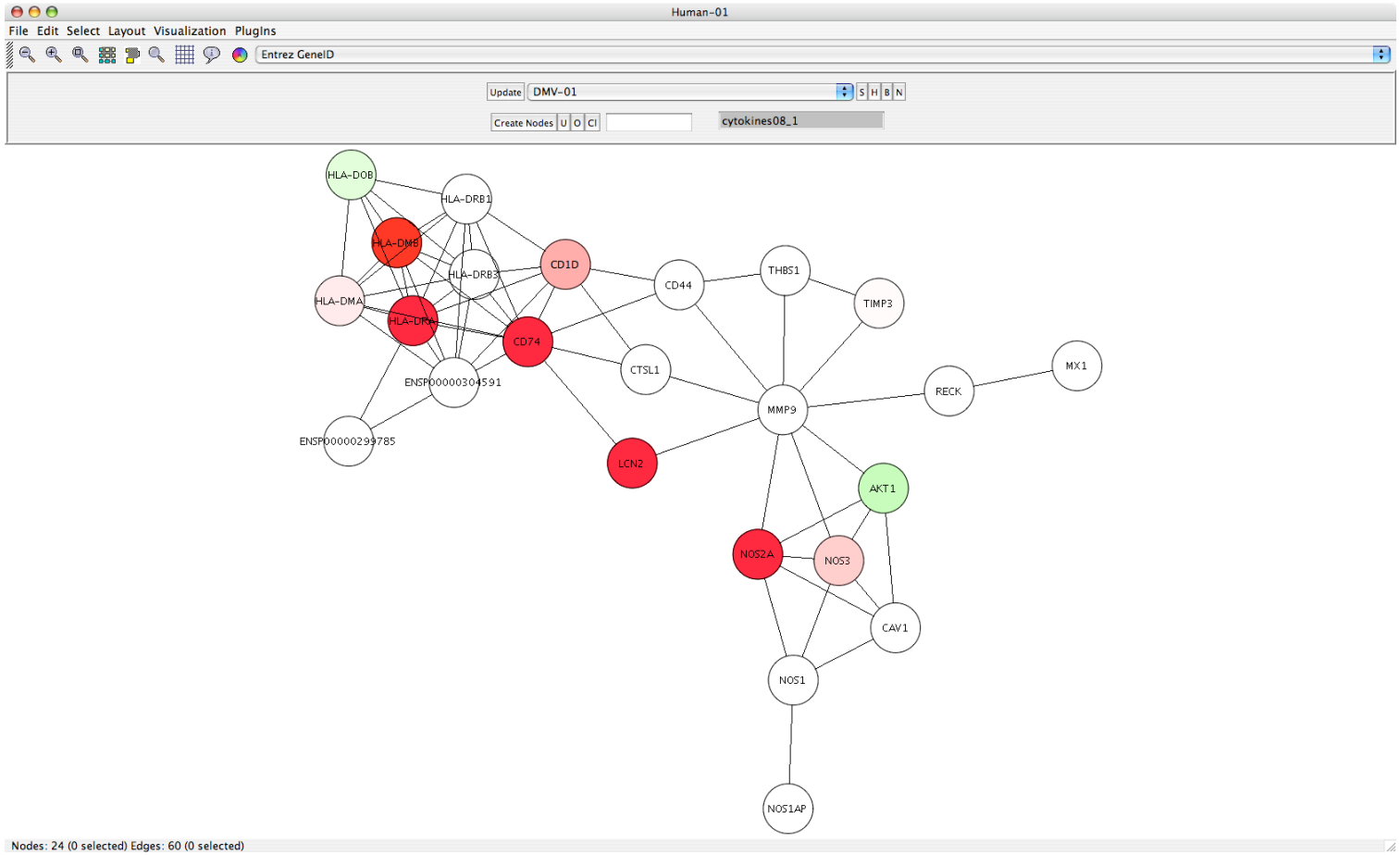
An Expression Movie, cont.



An Expression Movie, cont.
4 Hours



An Expression Movie, cont.
8 Hours



Future Plans

- GWAP: the gaggle web application
- Add a genome browser (Apollo from flybase?)
- Support Cytoscape 2.5
- Embedded gaggle markup at bioinformatics websites
- Systems Biology Markup Language (SBML)
- Better integration of MeV (Multi-experiment viewer)
- Socket communication to complement Java RMI