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

	String: functional protein association networks							
🔾 🏠 🔄 http://string.embl.de/newstring_cgi/show_network_section.pl 🔹 🕨 💽 🕻 Google							• Google	
t Headlines 🔊 🛛 fv	b▼ books▼ news▼	proj 🔻 me 🔻	gaggle▼ RoR	v eda v doc v G v	NCBI▼ mal▼ leo▼ k	oio▼ moz▼		
e Data: gaggle:	NameList(5)		•	Target: EMBL String	🛟 Show	w Hide Broadcast		
alendar ©	BioC2007 G	aggle Demo	8	Gene Details	Ø M DAVID 2007:	functional annota ⊗	🜸 String: fui	



Find associations in STRING



Expand associations in STRING



Evidence in STRING

LCN2 [ENSP00000277480]		MMP9 [ENSP00000216953]	
Neutrophil gelatinase-associated lipocalin precursor (NGAL alpha-2-microglobulin-related subunit of MMP-9) (Lipocali 24p3)	.) (p25) (25 kDa <> n-2) (Oncogene	Matrix metalloproteinase-9 precursor (MMP-9) (EC 3.4.24 IV collagenase) (92 kDa gelatinase) (Gelatinase B) (GELB) matrix metalloproteinase-9; 82 kDa matrix metallop	.35) (92 kDa type [Contains: 67 kDa proteinase-9]
	Evidence suggesting a func	tional link:	
Neighborhood in the Genome: none / insig	nificant.		
Gene Fusions: none / insig	nificant.		
Cooccurence Across Genomes: none / insig	nificant.		
Co-Expression: none / insig	nificant.		
Experimental/Biochemical Data: yes (score 0	.903).		Show
Association in Curated Databases: none / insig	nificant.		
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) (Opcogene 24p3) (198 aa)	
MX1	Interferon-induced GTP-binding protein Mx1 (Interferon-regulated resistance GTP-binding protein MxA) (Interferon-induced protein p78) (IEI-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)	8-850
HLA-DRA	HLA class II histocompatibility antigen, DR alpha chain precursor (MHC class II antigen DRA) (254 aa) (Homo sapiens)	borhoc Fusion urrenc ressio ressio ases ases ases alogy] ology]
Predicted Funct	ional Partners:	Neigh Gene Coocc Coexp Experi Textrr Textrr Flom Scor
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
O	The class I instructing anagen, sites I beta chain precarsor (nine class I anagen bites I) (200 day	
HLA-DRB2	HLA class II histocompatibility antigen, DW2.2/DR2.2 beta chain (Fragment) (266 aa)	 0.999
HLA-DRB2 ENSP0000030459:	HLA class II histocompatibility antigen, DW2.2/DR2.2 beta chain (Fragment) (266 aa) HLA class II histocompatibility antigen, DR alpha chain precursor (MHC class II antigen DRA) (254 aa)	• • 0.999 • 0.999

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).	Pub Aed
	⊙ MMP-9 ● NGAL …	
Þ	[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).	Pub Aed
\triangleright	Microdeformational wound therapy: effects on angiogenesis and matrix metalloproteinases in chronic wounds of 3 debilitated patients. Ann Plast Surg (2006).	Pub Aed
	○ MMP-9 ● NGAL …	
Þ	Neutrophil granule proteins in bronchoalveolar lavage fluid from subjects with subclinical emphysema. Am J Respir Crit Care Med (1999).	Pub Aed
	○ MMP-9, gelatinase B ● HNL …	
\bigtriangledown	The human neutrophil lipocalin supports the allosteric activation of matrix metalloproteinases. Eur J Biochem (2001).	Pub Med
	The human neutrophil lipocalin (HNL (\bullet)), a member of the large family of lipocalins that exhibit various physiological functions, is coexpressed in gran progelatinase B (MMP-9 (\bullet)). Part of it is covalently bound to the proenzyme and therefore may play a possible role in the activation process of promatrix meta We now report that HNL (\bullet) is able to accelerate the direct activation of promatrix metalloproteinases slightly. A significant enhancement of the activation for the HgCl2- and the plasma kallikrein-induced activation of all three secretory forms of proMMP-9 and of proMMP-8. The same activating exerted by HNL (\bullet) isolated from granulocytes as well as by the recombinant forms expressed by the yeast Pichia pastoris or by Escherichia coli. This demonst carbohydrate moiety is not essential for the biological activity of HNL (\bullet). Activation and activity enhancement are obviously mediated by entrapping to N-terminal sequence residues of the partially truncated proenzyme into the hydrophobic binding pocket of the HNL (\bullet). In conclusion these results document can exert an enzyme-activating effect in the regulation of inflammatory and pathophysiological responses of granulocytes in the physiological activation of MI been subject to limited proteolytic processing.	ulocytes with lloproteinases. ivity could be g effects were rates that the the remaining : that <u>HNL</u> (\bullet) MPs that have
Þ	Sustained activation of neutrophils in the course of Kawasaki disease: an association with matrix metalloproteinases. Clin Exp Immunol (2005).	Pub Aed
	◎ MMP9 ● NGAL	
\triangleright	Gelatinase isoforms in urine from bladder cancer patients. Clin Chim Acta (2000).	Pub Aed
	○ MMP-9 ● NGAL	
	Human neutrophil gelatinase and associated lipocalin in adult and localized juvenile periodontitis. J Dent Res (1996).	Pub Aed
	○ MMP-9 ● NGAL …	
Þ	Up-regulation of the extracellular matrix remodeling genes, biglycan, neutrophil gelatinase-associated lipocalin, and matrix metalloproteinase-9 in familial amyloid polyneuropathy. FASEB J (2005).	Pub

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" "ENSP0000009530"

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

Gaggle Data:	Genes fron	n STRING: Na	meList(2	4)	+	Target:	STRING-GenelD), Human, v7.0 🛟	Show	Hide	Broadcast	
oogle Calendar	0	Θ ві	ioC2007	Gaggle Demo	🛛 🌸 String	g: functiona	I protein associ	. 🛛				
Home ·	Dow	vnload	•	Help/Info)							*
							ТІМРЗ	RECK				

Name Translation Goose In Action

● ⊖ ⊖	STRING-GeneID, Huma	in, v7.0
Load Update	(R-05	S H B M N A C 🗹 auto
	Translate incoming names to:	GenelD
>>> handleN ENSP000003 ENSP000002 ENSP000002 ENSP000002 ENSP000003 ENSP000003 ENSP000003 ENSP000003 ENSP000003 ENSP000003 ENSP000002 ENSP000002 ENSP000002 ENSP000002 ENSP000002	lameList: 24 46631 -> 3122 303017 -> 3108 257498 -> 1514 299785 -> ENSP00000299785 293722 -> 3112 260356 -> 7057 304591 -> ENSP00000304591 289425 -> 912 39191 -> 857 355133 -> 9722 327251 -> 4843 242287 -> 8434 277480 -> 3934 266085 -> 7078 216953 -> 4318	
ENSP000002 ENSP000003 ENSP000003 ENSP000002 ENSP000002	229825 -> 3109 302517 -> 3125 353099 -> 3123 297494 -> 4846 270202 -> 207	▲ ▼
	Clear Qui	t

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 (r	ow) 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	0.000000	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]])

Home · Download · Help/Info		😒 STRING
O CD44 [ENSP00000278385]		O CD1D [ENSP00000289425]
CD44 antigen precursor (Phagocytic glycoprotein I) (PGP-1) (HUTCH-I) (Extracellular matrix receptor-III) (ECMR-III) (GP90 lymphocyte homing/adhesion receptor) (Hermes antigen) (Hyaluronate receptor) (Heparan sulfate proteoglycan) (Epican) (CDw44)	<>	T-cell surface glycoprotein CD1d precursor (CD1d antigen) (R3G1)
Evidence su	iggesting a fund	nctional link:
Neighborhood in the	Genome: none	e / insignificant.
Gen	e Fusions: none,	e / insignificant.
Cooccurence Across	Genomes: none,	e / insignificant.
C0-E: Everimental/Biochem	xpression: none,	e / insignificant.
Association in Curated D	Databases: none	e / insignificant.
Co-Mentioned in PubMed	Abstracts: yes (s	(score 0.447). Show
Combin	ned Score: 0.447	47

Your Input:

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.

0	0			DMV-01			
File	e Gaggle						
<u>U</u> pe	date Human-0	1 Details	etails				?
🖂 ratios							
	C		j T 🕞 💽		or avg 0	Clear All Inv	
		🗹 cytokines01_1	<pre> cytokines02_1 </pre>	🗹 cytokines04_1	🗹 cytokines08_1	🗹 cytokines 12_1	🗹 су
	10	0.005	-0.574	0.214	-0.183	-1.171	-
	10000	0	-0.607	-0.104	-0.063	-0.001	
	10001	0.152	-0.363	0.086	-0.141	-0.777	
	10013	-0.144	-0.059	-0.396	-1.231	-0.774	
	10018	-0.19	0.238	-0.428	0.383	0.3	
	10020	-0.185	-0.292	-0.667	-0.368	-0.462	
~	10038	0.154	-0.739	-0.04	0.165	-0.028	
	10047	-0.045	-0.246	-0.923	0.076	-0.328	
	10051	-0.012	0.313	-0.072	-0.031	0.144	
	10058	0.121	-0.297	-0.109	-1.185	-0.872	
	10059	0.108	0.979	0.195	0.306	0.441	
	10060	0.206	-0.782	-0.133	-0.074	-1.042	
	10062	0.094	0.26	-0.447	0.415	0.146	
	10066	-0.252	-0.361	0.236	0.349	-0.117	
	10072	-0.054	0.076	0.054	0.113	0.143	
	10093	-0.093	-0.305	0.051	-0.11	-0.475	
	10095	-0.278	0.402	-0.038	-0.397	0.44	
	10097	-0.743	0	-0.077	0	-0.603	
	10099	-0.093	-0.182	-0.054	-0.54	-0.977	Ψ
) • •
Þ							//

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