Institute for Digital Biology

>IDB

0101…ATGC

Welcome to GO: a User's Perspective

Fiona McCarthy*, Nan Wang*, Susan Bridges** and Shane Burgess**

FAQs

1. “How can I use the GO to model my functional genomics dataset?”

2. “Not all of my genes/proteins have GO annotation, how do I get more?”

3. “Why isn’t my species represented in the GO?”

FAQs

1. “How can I use the GO to model my functional genomics dataset?”

   Case study: chicken proteomics.

2. “Not all of my genes/proteins have GO annotation, how do I get more?”

   Case study: chicken microarrays.

3. “Why isn’t my species represented in the GO?”

   Your species and the GO annotation effort.

Case Study: Chicken Proteomics

FAQ: “How can I use the GO to model my functional genomics dataset?”

1. Why I started using the GO.
2. Why and how I became a GO annotator.

We identified 3,602 chicken B cell and stroma proteins.

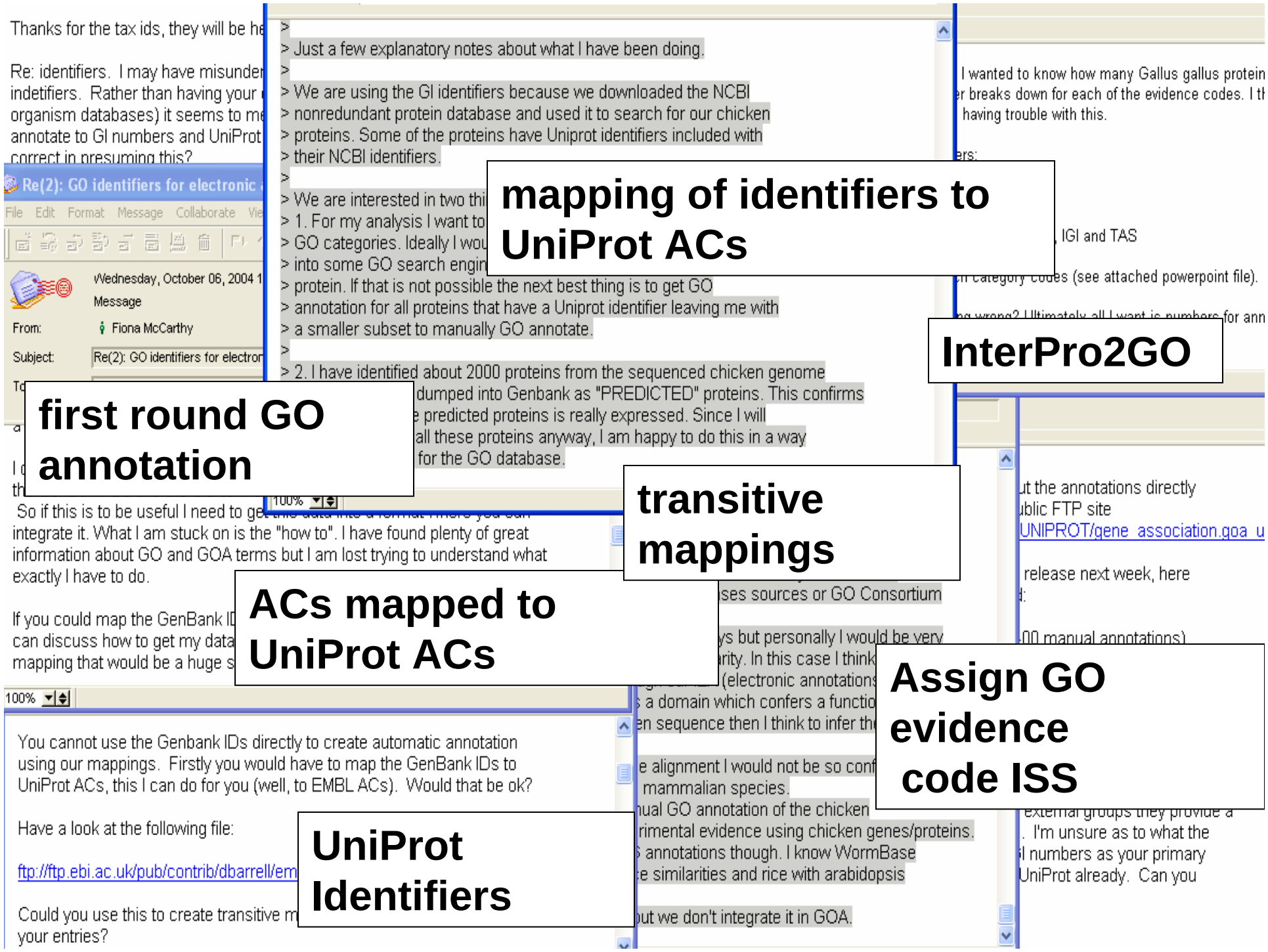
	Reference	Accession	Peptides (Hi	Score
#1	;al d 5)	113575	255 (244 6 1	2508.355
#2		113990	94 (84 4 4 2	904.4369
#3	Fibrinop	1706798	40 (37 1 1 0	386.3528
#4	oumin; H	1127086	34 (30 3 0 1	328.2493
#5		9954387	28 (0 22 3 2	204.7303
#6	C Chain C, Crystal Structure Of Native Chicken Fibrinogen	8569623	15 (15 0 0 0	150.3351
#7	I50711 complement C3 precursor - chicken	2118406	11 (9 2 0 0 0	106.3393
#8	TTHY_CHICK Transthyretin precursor (Prealbumin) (TBPA)	136463	10 (10 0 0 0	100.3767
#9	TIM2_CHICK Metalloproteinase inhibitor 2 precursor (TIMP-2) (Tissue inhib	3122960	13 (0 11 1 0	96.44643
#10	AAA6469	3645997	10 (9 0 1 0 0	96.43947
#11	MYH9_CHICK Myosin heavy chain, nonmuscle (Cellular myosin heavy cha	127759	14 (5 0 5 3 1	94.4
#12	S19188 myosin-V - chicken	104779	13 (4 3 3 2 1	92.33327
#13	FIBB_CHICK Fibrinogen beta chain precursor [Contains: Fibrinopeptide B]	399491	9 (8 1 0 0 0)	88.22259
#14	A Chain A, Crystal Structure Of Wild Type Turkey Delta 1 Crystallin (Eye l	14278427	12 (0 2 10 0	76.49026
#15	type I polyketide synthase AVES 2 [Streptomyces avermitilis MA-4680]	29827480	8 (6 1 0 1 0)	72.40443
#16	Hyperion protein, 419 kD isoform [Gallus gallus] 0	4582571	11 (0 5 3 3 0	70.23222
#17	vitronectin [Gallus gallus] ovirus 3]	1922282	7 (7 0 0 0 0)	70.21583
#18	CA36_CHICK Collagen alpha 3(VI) chain precursor	1345652	10 (3 2 3 1 1	70.21163
#19	paired-type homeobox Atx [Gallus gallus] I beta su	18252581	11 (0 4 5 1 1	68.6834
#20	I51298 transforming protein sno-N - chicken	2147397	7 (6 1 0 0 0)	68.24525
#21	TP2A_CHICK DNA topoisomerase II, alpha isozyme	13959708	7 (5 2 0 0 0)	66.27445
#22	ITA6_CHICK Integrin alpha-6 precursor (VLA-6)	124948	12 (0 5 1 4 2	66.25118
#23	glucose regulated thiol oxidoreductase protein precursor [Gallus gallus]	22651801	11 (1 3 4 0 3	64.65224
#24	spectrin alpha chain [Gallus gallus] rsor	1334744	9 (3 2 1 2 1)	62.33337
#25	ATP-binding cassette transporter 1 [Gallus gallus]	18028983	9 (2 3 2 1 1)	62.31025
#26	cone-type transducin alpha subunit [Gallus gallus]	11066401	8 (1 6 0 1 0)	62.15586
#27	condensin complex subunit [Gallus gallus] s] hick	26801168	12 (0 2 4 4 2	60.69801
#28	BA2B_CHICK Bromodomain adjacent to zinc finger domain 2B (Extracellul	22653663	8 (3 1 3 1 0)	60.14009
#29	ryanodine receptor type 3 [Gallus gallus]	1212912	9 (0 5 2 1 1)	58.38206
#30	type I polyketide synthase AVES 4 [Streptomyces avermitilis MA-4680]	29827484	7 (2 4 1 0 0)	58.30331
#31	structural muscle protein titin [Gallus gallus] n k	7024535	9 (0 5 2 0 2)	56.29744
#32	breast cancer susceptibility protein [Gallus gallus]	19568157	7 (4 1 1 0 1)	56.23917
#33	FAS_CHICK Fatty acid synthase [Includes: EC 2.3.1.38; EC 2.3.1.39; EC	1345958	8 (1 4 1 1 1)	54.35799

	Reference	Accession	Peptides (Hi	Score
#1	ALBU_CHICK Serum albumin precursor (Alpha-livetin) (Allergen Gal d 5)	113575	255 (244 6 1	2508.355
#2	APA1_CHICK Apolipoprotein A-I precursor (Apo-AI)	113990	94 (84 4 4 2	904.4369
#3	FIBA_CHICK Fibrinogen alpha/alpha-E chain precursor [Contains: Fibrinop	1706798	40 (37 1 1 0	386.3528
#4	Mol_id: 1; Molecule: Ovotransferrin; Chain: Null; Synonym: Conalbumin; H	1127086	34 (30 3 0 1	328.2493
#5	PB2 protein [Influenza A virus (A/chicken/Taiwan/7-5/99(H6N1))]	9954387	28 (0 22 3 2	204.7303
#6	C Chain C, Crystal Structure Of Native Chicken Fibrinogen	8569623	15 (15 0 0 0	150.3351
#7	I50711 complement C3 precursor - chicken	2118406	11 (9 2 0 0 0	106.3393
#8	TTHY_CHICK Transthyretin precursor (Prealbumin) (TBPA)	136463	10 (10 0 0 0	100.3767
#9	TIM2_CHICK Metalloproteinase inhibitor 2 precursor (TIMP-2) (Tissue inhib	3122960	13 (0 11 1 0	96.44643
#10	AAA6469	3645997	10 (9 0 1 0 0	96.43947
#11	MYH9_CHICK Myosin heavy chain, nonmuscle (Cellular myosin heavy cha	127759	14 (5 0 5 3 1	94.4
#12	S19188 myosin-V - chicken	104779	13 (4 3 3 2 1	92.33327
#13	FIBB_CHICK Fibrinogen beta chain precursor [Contains: Fibrinopeptide B]	399491	9 (8 1 0 0 0)	88.22259
#14	A Chain A, Crystal Structure Of Wild Type Turkey Delta 1 Crystallin (Eye L	14278427	12 (0 2 10 0	76.49026
#15	type I polyket		8 (6 1 0 1 0)	72.40443
#16	Hyperion prot		11 (0 5 3 3 0	70.23222
#17	vitronectin [G		7 (7 0 0 0 0)	70.21583
#18	CA36_CHICK		10 (3 2 3 1 1	70.21163
#19	paired-type h		11 (0 4 5 1 1	68.6834
#20	I51298 transf		7 (6 1 0 0 0)	68.24525
#21	TP2A_CHICK		7 (5 2 0 0 0)	66.27445
#22	ITA6_CHICK Integrin alpha-6 precursor (VLA-6)	124948	12 (0 5 1 4 2	66.25118
#23	glucose regulated thiol oxidoreductase protein precursor [Gallus gallus]	22651801	11 (1 3 4 0 3	64.65224
#24	spectrin alpha chain [Gallus gallus] rsor	1334744	9 (3 2 1 2 1)	62.33337
#25	ATP-binding cassette transporter 1 [Gallus gallus]	18028983	9 (2 3 2 1 1)	62.31025
#26	cone-type transducin alpha subunit [Gallus gallus]	11066401	8 (1 6 0 1 0)	62.15586
#27	condensin complex subunit [Gallus gallus] s] hick	26801168	12 (0 2 4 4 2	60.69801
#28	BA2B_CHICK Bromodomain adjacent to zinc finger domain 2B (Extracellul	22653663	8 (3 1 3 1 0)	60.14009
#29	ryanodine receptor type 3 [Gallus gallus]	1212912	9 (0 5 2 1 1)	58.38206
#30	type I polyketide synthase AVES 4 [Streptomyces avermitilis MA-4680]	29827484	7 (2 4 1 0 0)	58.30331
#31	structural muscle protein titin [Gallus gallus] n k	7024535	9 (0 5 2 0 2)	56.29744
#32	breast cancer susceptibility protein [Gallus gallus]	19568157	7 (4 1 1 0 1)	56.23917
#33	FAS_CHICK Fatty acid synthase [Includes: EC 2.3.1.38; EC 2.3.1.39; EC	1345958	8 (1 4 1 1 1)	54.35799

How do I turn this laundry list into a useful biological model?

Why I started using the GO.

- Need to know the function of all 3,602 chicken proteins.
- Google searches for the proteins name and “function” continually found the GO pages.
- Only 8% of my chicken proteins had any functional GO annotation (even in our well-studied model system).
- Contacted GO to get more GO annotations…


mapping of identifiers to UniProt ACs
InterPro2GO
first round GO annotation
transitive mappings
ACs mapped to UniProt ACs
Assign GO evidence code ISS
UniProt Identifiers
Re(2): GO identifiers for electronic
Wednesday, October 06, 2004
Message
From: Fiona McCarthy
Subject: Re(2): GO identifiers for electron
> Just a few explanatory notes about what I have been doing.
> We are using the GI identifiers because we downloaded the NCBI
> nonredundant protein database and used it to search for our chicken
> proteins. Some of the proteins have Uniprot identifiers included with
> their NCBI identifiers.
> We are interested in two thi
> 1. For my analysis I want to
> GO categories. Ideally I wou
> into some GO search engin
> protein. If that is not possible the next best thing is to get GO
> annotation for all proteins that have a Uniprot identifier leaving me with
> a smaller subset to manually GO annotate.
> 2. I have identified about 2000 proteins from the sequenced chicken genome
dumped into Genbank as "PREDICTED" proteins. This confirms
predicted proteins is really expressed. Since I will
all these proteins anyway, I am happy to do this in a way
for the GO database.
So if this is to be useful I need to ge
integrate it. What I am stuck on is the "how to". I have found plenty of great
information about GO and GOA terms but I am lost trying to understand what
exactly I have to do.
If you could map the GenBank ID
can discuss how to get my data
mapping that would be a huge s
100%
You cannot use the Genbank IDs directly to create automatic annotation
using our mappings. Firstly you would have to map the GenBank IDs to
UniProt ACs, this I can do for you (well, to EMBL ACs). Would that be ok?
Have a look at the following file:
ftp://ftp.ebi.ac.uk/pub/contrib/dbarrell/em
Could you use this to create transitive m
your entries?
Thanks for the tax ids, they will be he
Re: identifiers. I may have misunder
indetifiers. Rather than having your
organism databases) it seems to m
annotate to GI numbers and UniProt
correct in presuming this?
I wanted to know how many Gallus gallus protein
breaks down for each of the evidence codes. I th
having trouble with this.
IGI and TAS
(see attached powerpoint file).
ut the annotations directly
ublic FTP site
UNIPROT/gene_association.goa_u
release next week, here
ses sources or GO Consortium
ys but personally I would be very
arity. In this case I think
(electronic annotations
a domain which confers a functio
en sequence then I think to infer th
e alignment I would not be so conf
mammalian species.
ual GO annotation of the chicken
rimental evidence using chicken genes/proteins.
annotations though. I know WormBase
e similarities and rice with arabidopsis
ut we don't integrate it in GOA.
. I'm unsure as to what the
l numbers as your primary
UniProt already. Can you

GO Chicken: From MudPIT to Swamp

GO Users' Meeting, Chicago 2004

Fiona M. McCarthy, Amanda M. Cooksey, G. Todd Pharr and Shane C. Burgess

College of Veterinary Medicine, Mississippi State University

Why I became a GO annotator…

- Chicken GO annotations provided via EBI-GOA Project
- No dedicated effort to annotate chicken
- The only way to get more GO annotation is for me to add it!

GO Annotation training to become a GO curator.

GO Training at MGI, Maine

[Contact us]
[Help]

AgBase
[Version: 1.02]

Animals
Plants
Microbes
Parasites

Search
Text
BLAST
Taxonomy
Gene Ontology
Proteogenomics

Tools

Downloads &
Statistics

Welcome to AgBase

AgBase is a curated, open-source, web-accessible resource for functional analysis of agricultural plant and animal gene products.
Read more...

AgBase also provides tools for proteomic analysis and functional analysis using the GO.

Search AgBase

Select AgBase Database: AgBase

Browse By:
>Select name type

Enter Multiple Queries (new line separated)

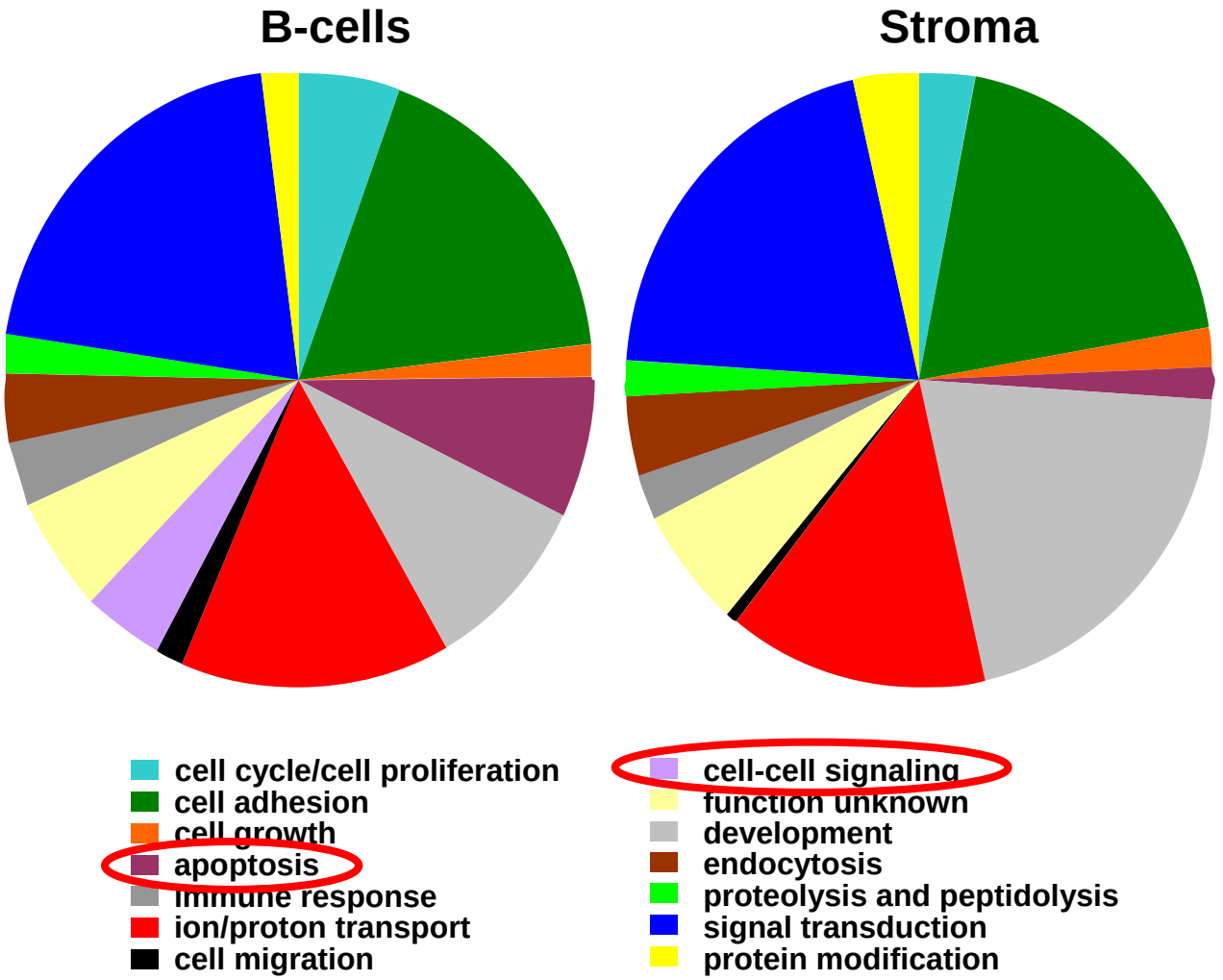
Modeling a whole organ using proteomics: The avian bursa of Fabricius

Fiona M. McCarthy, Dr. [1]*, Amanda M. Cooksey [1], Nan Wang [2], Susan M. Bridges [2], G. Todd Pharr [1], Shane C. Burgess [1]

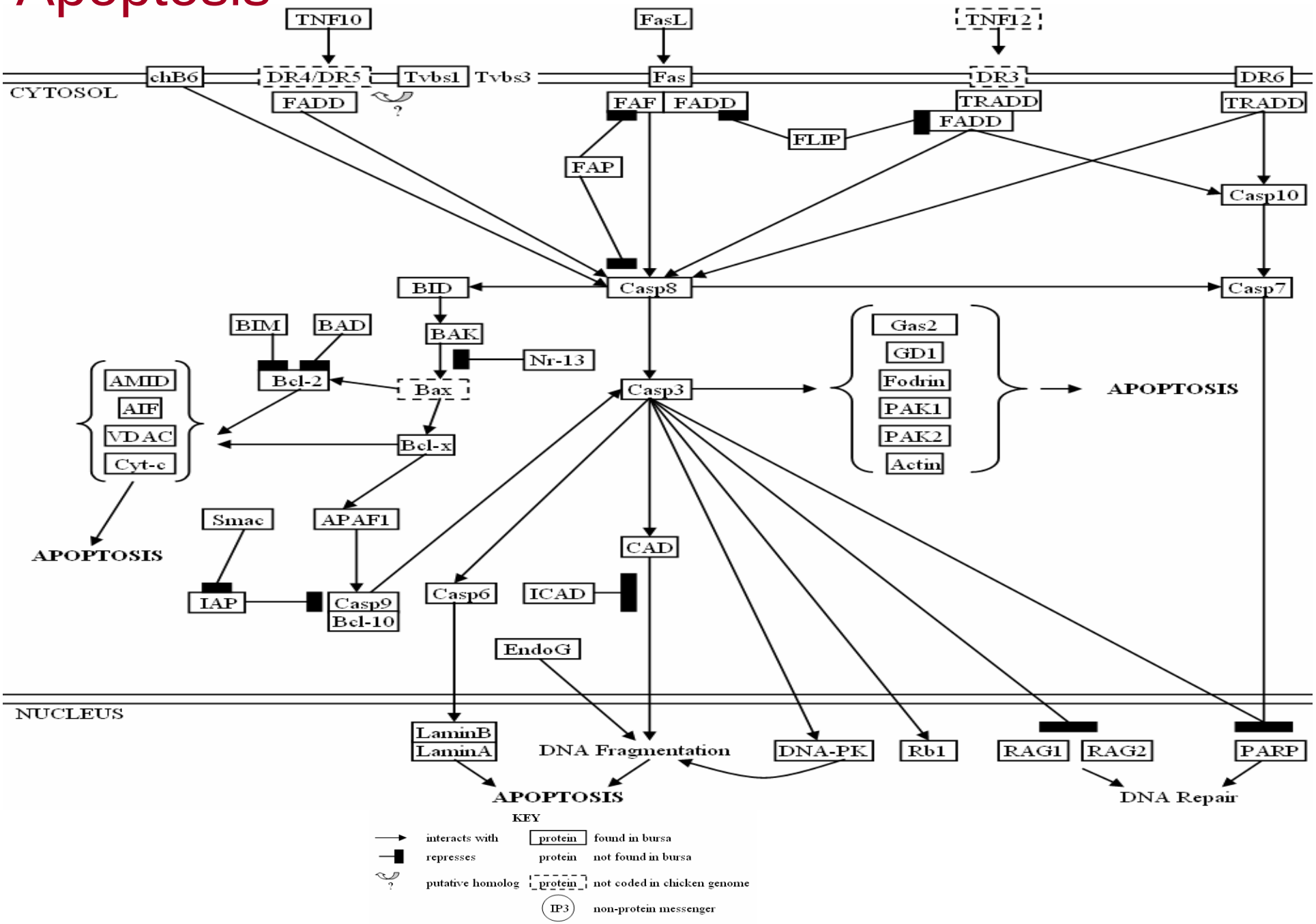
[1]Department of Basic Science, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA

[2]Department of Computer Science and Engineering, College of Engineering, Mississippi State, MS, USA

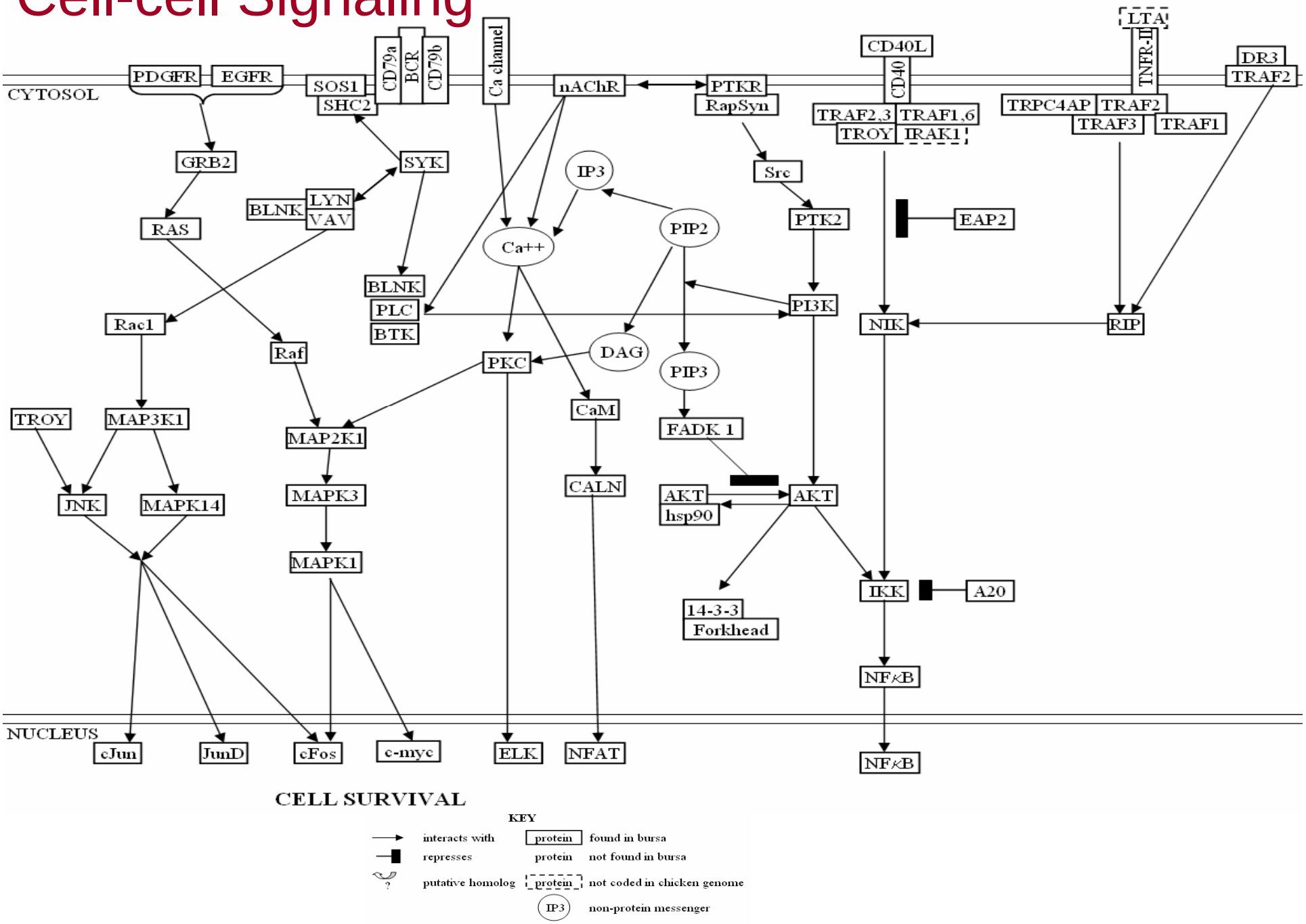
Comparison of Biological Process: Membrane Proteins



Apoptosis


TNF10
FasL
TNF12
chB6
DR4/DR5
Tvbs1
Tvbs3
Fas
DR3
DR6
CYTOSOL
FADD
?
FAF
FADD
TRADD
FADD
TRADD
FLIP
FAP
Casp10
BID
Casp8
Casp7
BIM
BAD
BAK
Nr-13
Gas2
GD1
Fodrin
PAK1
PAK2
Actin
AMID
AIF
VDAC
Cyt-c
Bcl-2
Bax
Bcl-x
Casp3
APOPTOSIS
Smac
APAF1
CAD
APOPTOSIS
IAP
Casp9
Bcl-10
Casp6
ICAD
EndoG
NUCLEUS
LaminB
LaminA
DNA Fragmentation
DNA-PK
Rb1
RAG1
RAG2
PARP
APOPTOSIS
DNA Repair
KEY
interacts with
protein
found in bursa
represses
protein
not found in bursa
putative homolog
protein
not coded in chicken genome
IP3
non-protein messenger

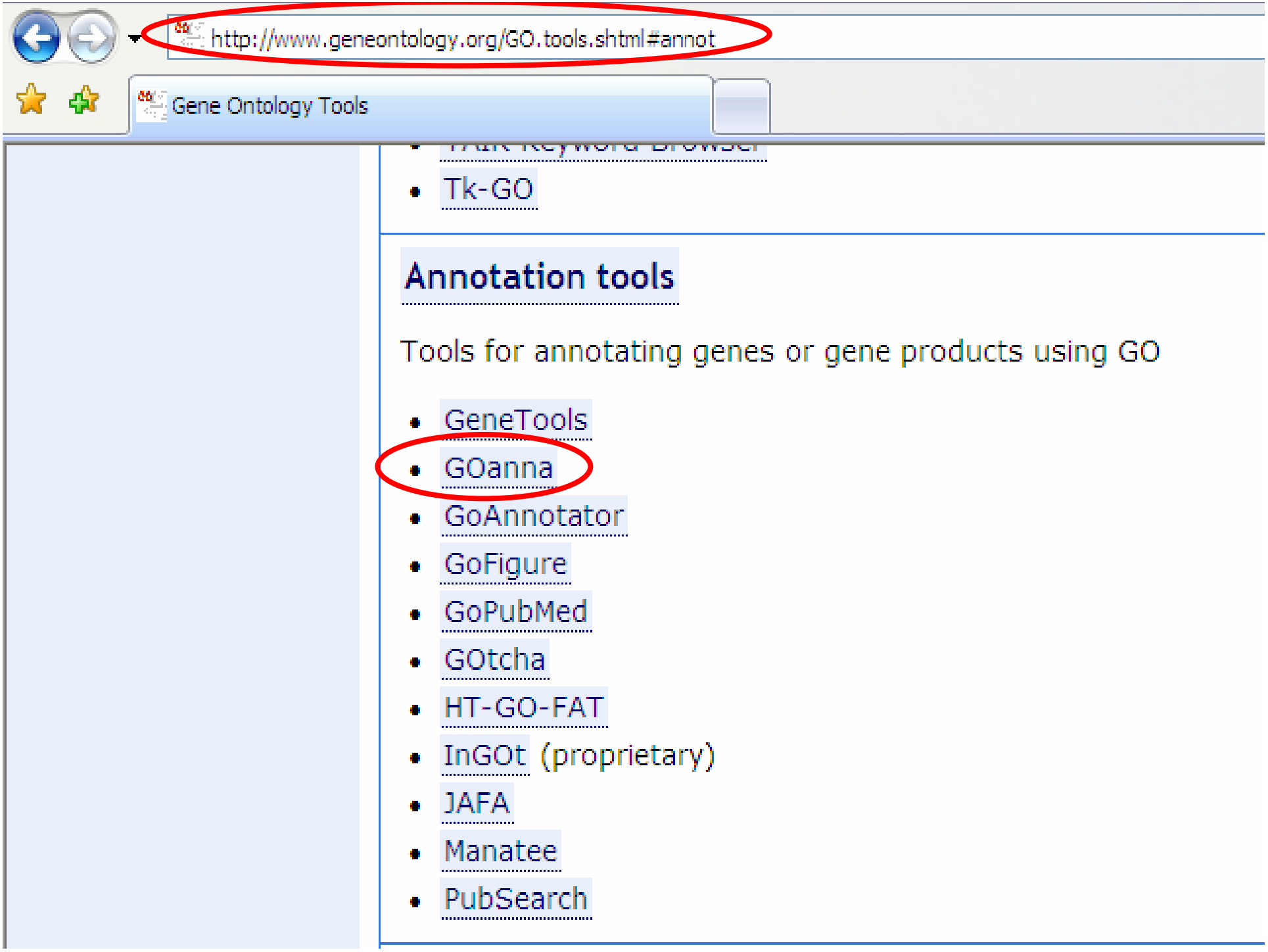
Cell-cell Signaling



Case Study: Chicken Microarrays

FAQ: “Not all of my genes/proteins have GO annotation, how do I get more?”

1. Tools for adding GO based on homology.
2. Getting involved: community annotation.


http://www.geneontology.org/GO.tools.shtml#annot
Gene Ontology Tools
Tk-GO
Annotation tools
Tools for annotating genes or gene products using GO
GeneTools
GOanna
GoAnnotator
GoFigure
GoPubMed
GOtcha
HT-GO-FAT
InGOt (proprietary)
JAFA
Manatee
PubSearch

Tools for adding GO

[Contact us]
[Help]

AgBase
[Version: 1.02]

Animals
Plants
Microbes
Parasites

Search
Text
BLAST
Taxonomy
Gene Ontology
Proteogenomics

Tools

Downloads & Statistics

Journal Database

Educational Resources

AgBase Tools

AgBase provides tools designed to assist with the analysis of proteomics data and tools to evaluate experimental datasets using the GO.

Proteomics Analysis

ProtIDer (ESTPlus, PIE Tool) [Available on request]
Enhances proteomics based on EST or EST assemblies by generating databases of matching highly homologous proteins.

The Proteogenomic Pipeline [Available on request]
The Proteogenomic Pipeline provides experimentally-based structural annotations at a complete genome level.

Functional Analysis Using GO

GOProfiler
Summarizes existing GO annotations in AgBase.

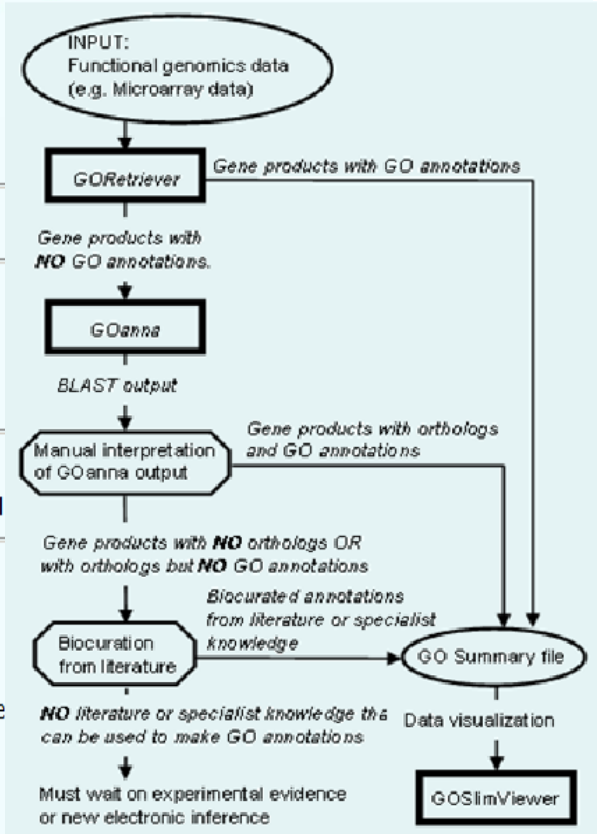
GO Annotation Pipeline

These tools may be used independently, or as a pipeline (shown below) to provide GO annotations for experimental

GORetriever
Returns AgBase GO annotations for a list of proteins IDs.

GOanna
Allows users to submit a list of protein or nucleotide IDs or a FASTA file containing sequences and returns AgBase

GoSlimViewer
Summarizes GO annotations for datasets using a GO Slim.



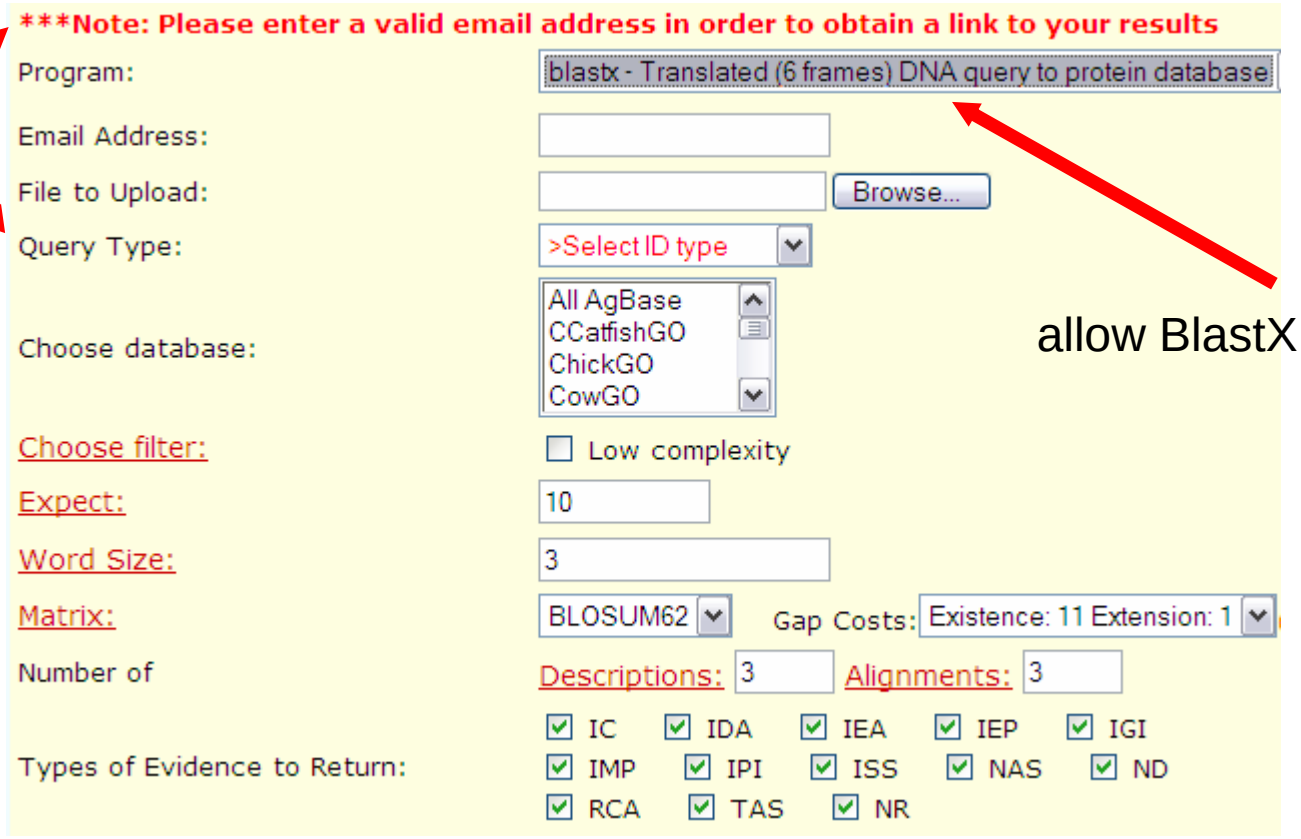
Case Study: Chicken Microarrays

modify GOanna tool to assist with GO annotation of microarrays:

allow larger searches by returning results via email

modify to accept Fasta sequence files

allow BlastX searches



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Contig_11378.1	Chick																
2		EMP2_HU	Epithelial r	246	1.00E-64													
3			UniProt	P54851	EMP2_HUMAN		GO:00082	PMID:899	TAS		P	cell proliferation		IPI000088	protein	9606	9/4/2003	PINC
4			UniProt	P54851	EMP2_HUMAN		GO:00160	PMID:888	TAS		C	integral to membrane		IPI000088	protein	9606	9/4/2003	PINC
5		Q66HH2_F	Epithelial r	232	3.00E-60													
6		Q548I4_M	Epithelial r	231	6.00E-60													
7	Contig_12369.1	Chick																
8		Q8N828_H	CDNA FL	98	6.00E-20													
9		Q6JQN1_H	ACAD 10	98	6.00E-20													
10		Q8NAP2_H	CDNA FL	96	2.00E-19													
11	Contig_3035.2	Chick																
12		COMD9_H	COMM dor	275	3.00E-73													
13		Q3MIE7_F	COMM dor	274	4.00E-73													
14		Q53FR9_H	COMM dor	272	2.00E-72													
15	Contig_145.1	Chick																
16		ERP29_CH	Endoplasm	447	e-125													

reformat output to a GEO style

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	**EST Contigs**	**Matching Gene Products**		**BLAST Statistics**					**O Cellular Componer**		**O Molecular Functio**		**O Biological Preoces**	
2	**Sequence ID**	**UniProtKB Accession**	**UniProt Name**	**query length**	**match length**	**alignment**	**Score**	**Eval**	**GO ID**	**GO Term**	**GO ID**	**GO Term**	**GO ID**	**GO Term Nar**
3	Contig_11378.1	P54851	Epithelial membrane protein 2	2055	167	118/168 (70%)	246	1.00E-64	GO:00160	integral to membrane			GO:00082	cell proliferatic
4	Contig_145.1	P81628	Endoplasmic reticulum protein E	2026	228	226/228 (99%)	447	e-125	GO:00057	endoplasmic reticulum; endoplasmic reticulum				
5	Contig_145.1	P30040	Endoplasmic reticulum protein E	2026	261	181/259 (69%)	363	1.00E-99	GO:00057	endoplasm	GO:00037	protein dis	GO:00064	protein folding
6														

Summary of GO Annotations of Del-Mar 14K Array		
Total Contigs on Del-Mar 14K Array	**14,049**	**% of Contigs with Protein ID/GO Terms**
Number of Contigs with Protein ID	9,587	68.23973
Number of Contigs with CC GO Terms	1982	14.10777
Number of Contigs with MF GO Terms	2576	18.33582
Number of Contigs with BP GO Terms	2331	16.59193
Number of Singlets on Del-Mar 14K Array	**3034**	**% of singlets with Protein ID/GO Terms**
Number of Singlets with Protein ID	1400	46.1437
Number of Singlets with CC GO Terms	354	11.66777
Number of Singlets with MF GO Terms	263	8.668425
Number of Singlets with BP GO Terms	297	9.789057

prioritise chicken protein matches with no annotation for future GO annotation

Adding GO via Homology

- Rapidly gives “breadth” of coverage.
- GO annotation based on homology adds general or “higher order” GO functions.
- GO annotation based on homology does not add direct functional annotation for the species of interest.
- Manual annotation of the literature provides detailed, species specific GO annotations.

[Contact u
[Help]

AgBase
[Version: 1.02]

Animals
Plants
Microbes
Parasites

Search
Text
BLAST
Taxonomy
Gene Ontology
Proteogenomics

Tools

Downloads &
Statistics

Community Request and Submissions

Members of the agricultural research communities can submit their own GO annotations to AgBase or request that a particular protein be annotated by the AgBase curators. To request that a protein be annotated, fill out the **User Information**, the **Gene Product Information**, and the **Literature Evidence** sections below. If you would like to submit an annotation to the AgBase database, all three sections above and the **Optional Information** must be completed.

User Information:

Name:

Email address/Login:

Institution:

Research Area:

Community Annotation:
bridging the gap between functional genomics and trained GO annotators.

Additional information regarding these fields can be found here.

* - Denotes a required field.

Gene Product Information:

*Database: >Please select

You are AgBase visitor no. **8725**

Your Species and the GO

FAQ: “Why isn’t my species represented in the GO?”

1. Finding out more about your species in the GO.
2. GO educational resources.
3. GO annotation training.

Your species & the GO

Species directed GO annotation efforts:

[Contact us]

[Help]

AgBase
[Version: 1.02]

Animals
Plants
Microbes
Parasites

Search
Text
BLAST
Taxonomy
Gene Ontology
Proteogenomics

Tools

Downloads &
Statistics

Welcome to AgBase

AgBase is a curated, open-source, web-accessible resource for functional analysis of agricultural plant and animal gene products.
Read more...

AgBase also provides tools for proteomic analysis and functional analysis using the GO.

Search AgBase

Select AgBase Database: AgBase

Browse By:

>Select name type

Enter Multiple Queries (new line separated)

AgBase

[Contact u
[Help]

AgBase
[Version: 1.02]

Animals
Plants
Microbes
Parasites

Search
Text
BLAST
Taxonomy
Gene Ontology
Proteogenomics

GOProfiler

GOProfiler provides a summary of the GO annotations available in AgBase. The user provides a species (taxon id) and GOProfiler displays the number of GO associations and the number of annotated proteins for that species. The results are listed by evidence code and a separate list of unannotated proteins is also provided.

Enter Taxon ID: 9913 (e.g. 9031)
use Taxonomy Browser to find the taxon ID for a species
Submit Reset

Our AgBase GoProfiler Summary Page for taxonomy: 9913

Number Of Annotated proteins	**7445**
Number Of assigned annotions	**40422**
Number Of unannoted proteins	2509

Evience Code	Number of Annotations	Number of Proteins
IC	10	9
IDA	148	58
IEA	36020	7400
IEP	3	3
IMP	11	7
IPI	56	35

GOProfiler: summarizes the number of GO annotations available based on GO evidence codes.

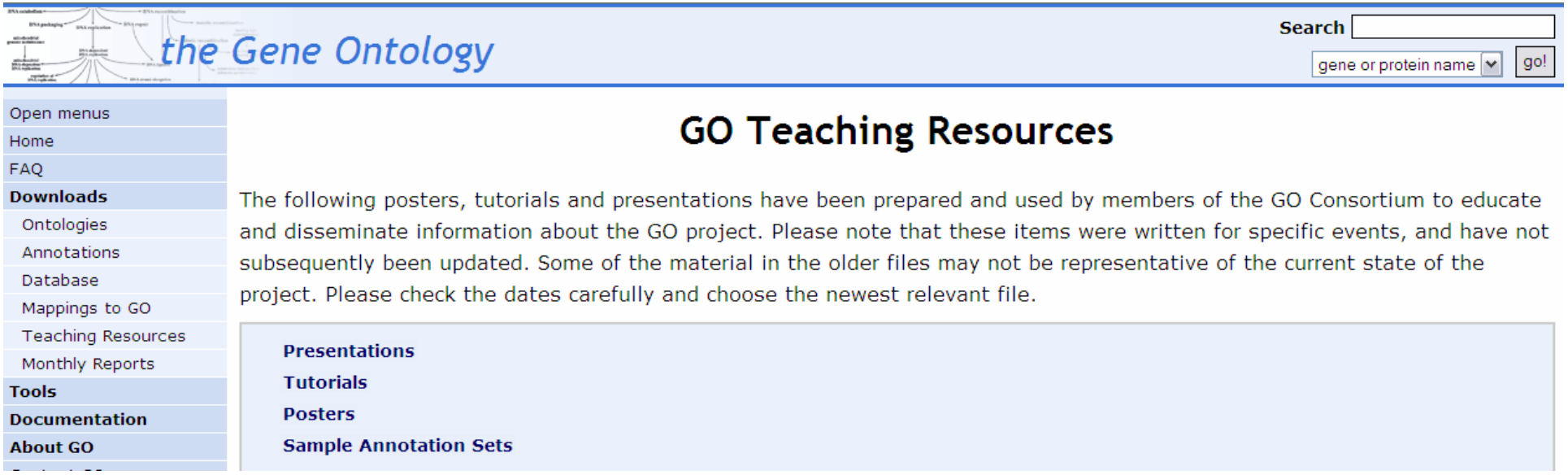
GO Educational Resources

AgBase Educational Resources

The following material has been prepared by members of AgBase and includes seminars and tutorials about GO annotation. They are presented here to encourage agricultural researchers to learn more about how to use the GO and its associated computational tools to add value to their functional genomics datasets.

We also suggest that users interested in GO visit the GO Consortium Homepage Teaching Resources.

- **"MSU GO Annotation class notes"**(Fall 2005)
- **"AgBase and the Ruminant Genomes: GOing towards greener pastures"** (PAGXIV 2006)
- **"ChickGO Takes Flight"** (2005 GO User Meeting)
- **"The AgBase GO Annotation Tools"** (2005 GO User Meeting)
- **"AgBase GO Toolkit: A Gene Ontology Annotation Toolkit for Large Scale Protein Datasets"**(2005 Biocurator Meeting)



GO Annotation Training

- AgBase: 2 week intensive GO annotation training with MGI mentor group

- GO now runs regular GO annotation training courses
  - focus on **manual** GO annotations
  - 2007: Annotation Training Camp at TIGR

With Thanks To:

GO Consortium Member mentor

MGI: Judith Blake, David Hill, Mary Dolan, Harold Drabkin

GOA: Evelyn Camon, Dan Barrell

GO Editorial Office: Jen Clark, Midori Harris

dictyBase: Rex Chisholm, Eric Just

Plant Associated Microbe Gene Ontology (PAMGO) / Institute of Digital Biology - AgBase.

Booth 206