

# Building and Using Ontologies to do biology

Robert Stevens  
BioHealth Informatics Group  
School of Computer Science  
University of Manchester  
[Robert.Stevens@Manchester.ac.uk](mailto:Robert.Stevens@Manchester.ac.uk)

# Ernest Rutherford



“All science is either physics or stamp  
collecting”

# The Basics

- Unless you know what you've got, it is not much use in analysis
- Unless you share that knowledge about what you've got, then other people's data isn't much use either
- If we can make our information more intelligent, then we can make more sophisticated analyses

# The Annotation Baseline

- Most uses of ontology is to supply a vocabulary for annotation
- “This is what my thing is”, “This is what my thing does”
- Get me all gene products labelled as being involved in “cell motility”
- Also gets those gene products labelled with kinds of “cell motility”

# What happens when we've annotated everything?

- Reliable querying across resources
- Wherever one needs to know what something is...
- Knowing how to recognise an entity is the heart of ontology
- Computational recognition of entities
- Capturing *some* of human knowledge in computational form
- Ontological definitions are not just written in English (though they should be this as well)
- Write the definition in logical form for the computer to use...

# AmiGO on “Cell motility”

Go to cell motility involved in somitogenic axis elongation

Ancestors and Children   Inferred Tree View   Graph View   Other Views   Downloads   Mappings

- GO:0008150 biological process [405310 gene products]
  - GO:0009987 cellular process [241906 gene products]
  - GO:0051179 localization [58312 gene products]
    - GO:0006928 cellular component movement [7621 gene products]
    - GO:0051674 localization of cell [5206 gene products]
    - GO:0040011 locomotion [9689 gene products]
    - GO:0048870 cell motility [5197 gene products]
      - GO:0070358 actin polymerization-dependent cell motility [17 gene products]
      - GO:0071976 cell gliding [0 gene products]
      - GO:0016477 cell migration [4292 gene products]
      - GO:0003411 cell motility involved in camera-type eye morphogenesis [0 gene products]
      - GO:0021814 cell motility involved in cerebral cortex radial glia guided migration [0 gene products]
      - GO:0090247 cell motility involved in somitogenic axis elongation [0 gene products]
      - GO:0021805 cell movement involved in somal translocation [0 gene products]
      - GO:0061334 cell rearrangement involved in Malpighian tubule morphogenesis [0 gene products]
      - GO:0071975 cell swimming [0 gene products]
      - GO:0001539 ciliary or flagellar motility [670 gene products]
      - GO:0060327 cytoplasmic actin-based contraction involved in cell motility [18 gene products]
      - GO:2000146 negative regulation of cell motility [497 gene products]
      - GO:2000147 positive regulation of cell motility [1050 gene products]
      - GO:2000145 regulation of cell motility [1829 gene products]
      - GO:0030317 sperm motility [165 gene products]
      - GO:0043107 type IV pilus-dependent motility [14 gene products]

# Associations on “Cell motility”

## cell motility

Term associations Term information Term lineage External references

### Gene Product Associations to cell motility ; GO:0048870 and children

Download all association information in:  gene association format  RDF/XML

#### Current filters

Species:

#### Filter associations displayed

Filter by Gene Product

Gene Product Type	Data source	Species	Filter by Association
All	All	Gallus gallus	Evidence Code
complex	ASAP	Geobacillus stear...	All
gene	AspGD	Geobacter sulfurr...	IBA
gene product	CGD	Homo sapiens	IBD
			IKR

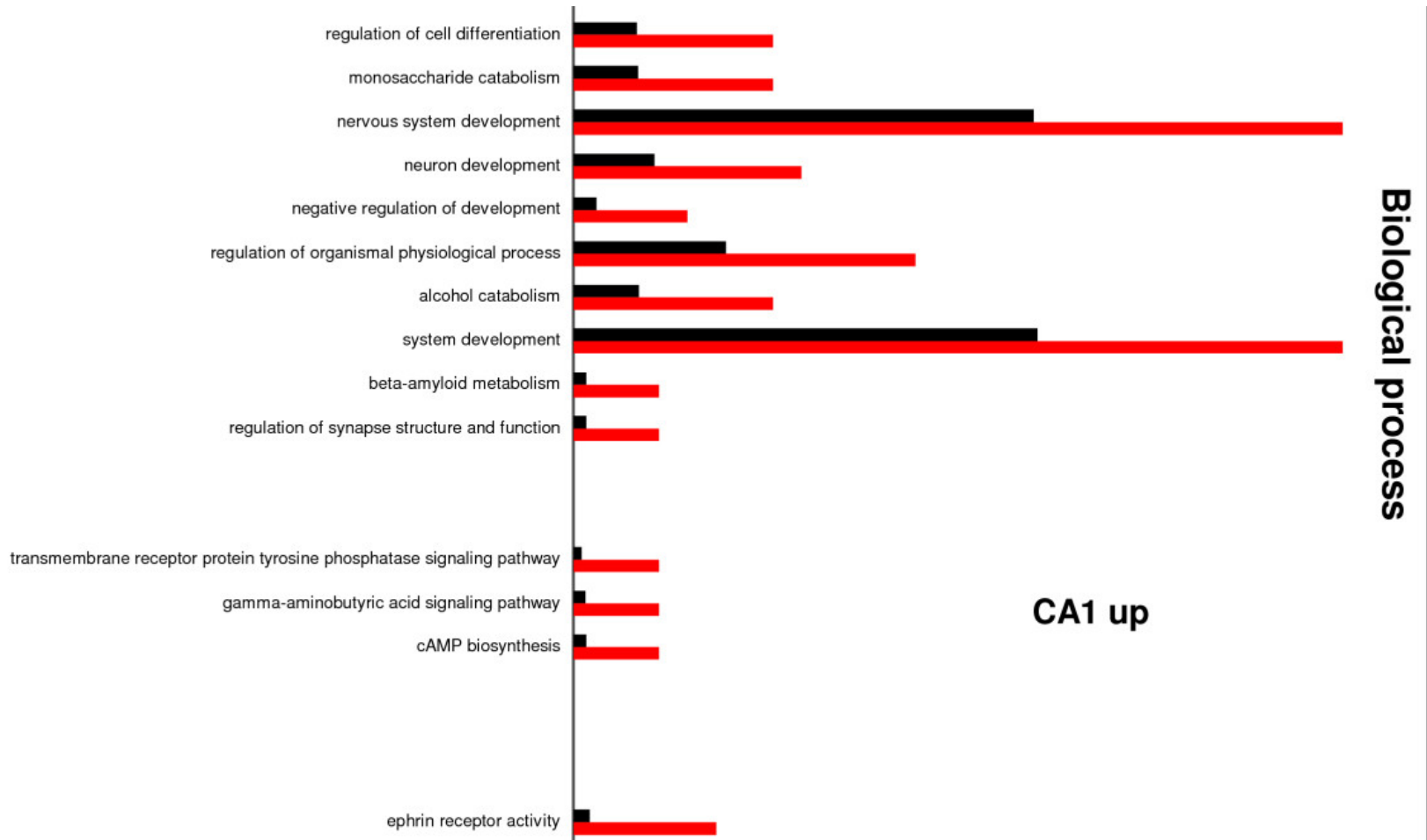
View associations:  All  Direct associations

1 2 3 4 5 6 7 8 9 ... 29 View all results

### cell motility ; GO:0048870 [\[show def\]](#) [\[view in tree\]](#)

Symbol, full name	Information	Qualifier	Evidence	Reference	Assigned by
<a href="#">ADAM17</a> Disintegrin and metalloproteinase domain-containing protein 17	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	ISS With UniProtKB:Q9Z0F8	PMID:10433800	BHF-UCL (via UniProtKB)
<a href="#">DST</a> Dystonin	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IMP	PMID:19403692	UniProtKB
<a href="#">ENG</a> Endoglin	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IMP	PMID:18223685	BHF-UCL (via UniProtKB)
<a href="#">ETS1</a> Protein C-ets-1	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IMP	PMID:15247905	BHF-UCL (via UniProtKB)
<a href="#">FSCN1</a> Fascin	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IDA	PMID:9571235	UniProtKB
<a href="#">ITGB4</a> Integrin beta-4	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IMP	PMID:19403692	UniProtKB
<a href="#">MAP2K1</a> Dual specificity mitogen-activated protein kinase kinase 1	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IEA With Ensembl:ENSMUSP00000005066	GO REF:0000019	Ensembl (via UniProtKB)
<a href="#">RAC1</a> Ras-related C3 botulinum toxin substrate 1	<a href="#">view associations</a> <a href="#">BLAST</a>	protein from <i>Homo sapiens</i>	IDA	PMID:19403692	UniProtKB

# Over expression analysis



Newrzella *et al.* *BMC Genomics* 2007 8:370 doi:  
10.1186/1471-2164-8-370



# Annotating more than gene products

- Ontologies spreading from genotype to phenotype
- Ontologies for describing the experiments that produce data and the information they produce
- All comes down to knowing what we've got and being able to compare
- Semantic description of bioinformatics services

# RACE Data and Methods Ontology ([http:// edamontology.sf.net](http://edamontology.sf.net))

## Bioinformatics data, formats, identifiers, operations and topics

Terms ▾

Jump To:

Details Visualization Notes (0) Term Mappings (0) Term Resources

- ⊕ Data
  - ⊖ 1 Alignment data
  - ⊖ 1 Article data
  - ⊖ 1 Biological model
  - ⊖ 1 Classification data
  - ⊖ 1 Codon usage data
  - ⊖ 1 Data index data
  - ⊖ 1 Data reference
  - ⊖ 1 Experimental data
  - ⊖ 1 Gene expression data
  - ⊖ 1 Genotype and phenotype data
  - ⊖ 1 Identifier data
  - ⊖ 1 Image or plot
  - ⊖ 1 Map data
  - ⊖ 1 Matrix
  - ⊖ 1 Metadata and annotation
  - ⊖ 1 Molecular data
  - ⊖ 1 Molecular interaction data
  - ⊖ 1 Ontological data
  - ⊖ 1 Parameter or primitive
  - ⊖ 1 Pathway or network data
  - ⊖ 1 Phylogenetic data
  - ⊖ 1 Provenance data
  - ⊖ 1 Reaction data
  - ⊖ 1 Schema
  - ⊖ 1 Sequence data
  - ⊖ 1 Structural data
  - ⊖ 1 Workflow data
- ⊕ Format
- ⊕ Identifier
- ⊕ Operation
- ⊕ Topic

Preferred Name	Data
Definitions	A type of data in common use in bioinformatics.
ID	EDAM:0000006
Full Id	<a href="http://purl.bioontology.org/ontology/EDAM/EDAM_0000006">http://purl.bioontology.org/ontology/EDAM/EDAM_0000006</a>
Subset	data
Xref Definition	EDAM:EBI "EMBRACE definition"

# Making Deductions

- “All french people like wine”
- “Pierre is a Frenchman”
- Thus we know “Pierre likes wine”
- We’ve already seen the application of this in AmiGO
- ..., but there’s a whole lot more “reasoning”

# The GOAL User Interface

- <http://owl.cs.manchester.ac.uk/goal/>;
- GWT application for building simple queries against Gene Ontology Annotations (GOA);
- Uses OWL API and uses HermiT reasoner;
- Simply browse and “add” classes to a conjunctive query;
- Write your own DL queries (can be dangerous...)

# The Gene Product Class

- Class: 'Gene product'
- SubClassOf:
  - has\_function some 'molecular function',
  - participates\_in some 'biological process',
  - located\_in some 'cellular component'

# Two Mouse Gene Products as Primitive OWL classes

- Class: Frem1

- SubClassOf:
- has\_function some 'sugar binding',
- participates\_in some 'cell adhesion',
- located\_in some 'proteinaceous extracellular matrix',
- located\_in some 'extracellular region'

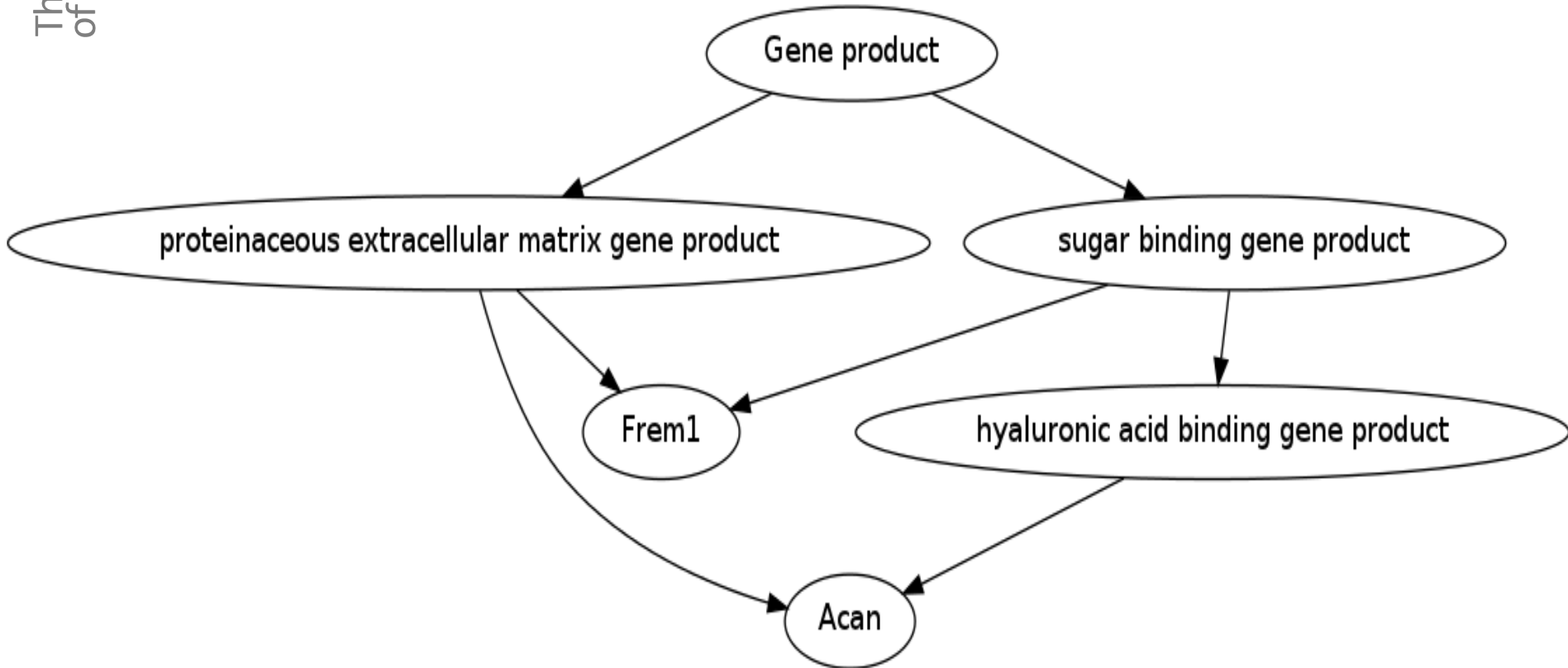
- Class: Acan

- SubClassOf:
- has\_function some 'hyaluronic acid binding',
- participates\_in some 'collagen fibril organization',
- located\_in some 'proteinaceous extracellular matrix'

# Two Defined Classes

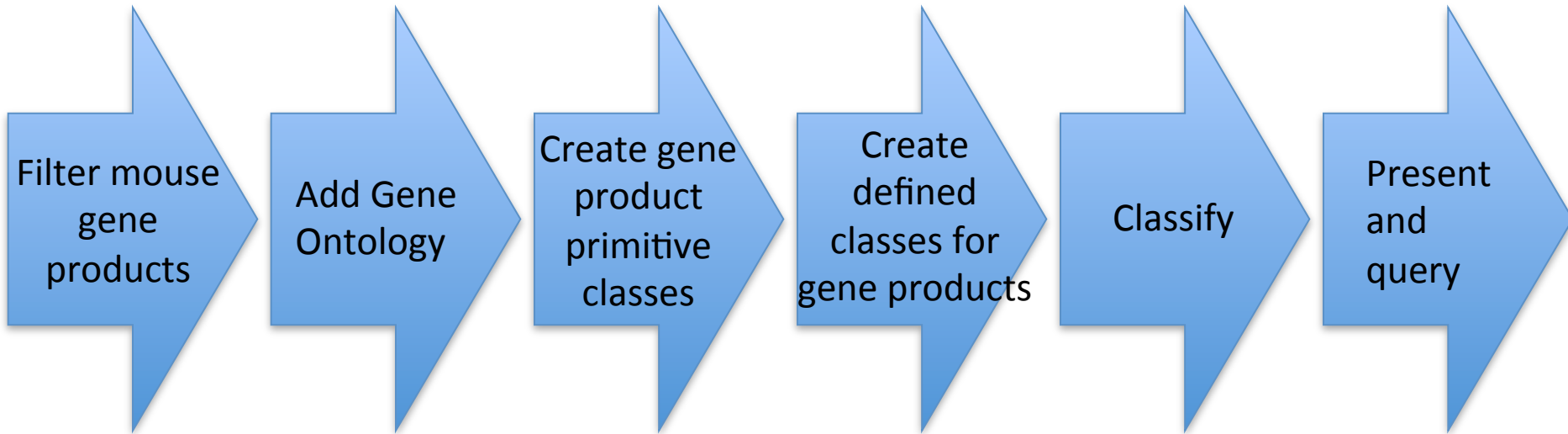
- Class: 'proteinaceous extracellular matrix gene product'
- EquivalentTo:
  - 'Gene product' and located\_in some 'proteinaceous extracellular matrix'
- Class: 'sugar binding gene product'
- EquivalentTo:
  - 'Gene product' and has\_function some 'sugar binding'

# The Resulting Hierarchy





# Workflow



# Basic GOAL interface

## GOAL Browser

Logical Gene Ontology Annotations

This simple application demonstrates how we can exploit OWL semantics to query over Gene Ontology annotations. This application currently covers only a subset of the mouse uniprot proteins.

You can download all the ontologies from [here](#), they are best viewed in Protege 4.1 and can be classified on a desktop machine with [HermiT](#) in around 5 minutes.

### Gene Product Ontology Browser

Use ctrl/cmd key to select multiple terms

---

▼ **GeneProduct**

- ▶ cellular\_component gene product
- ▶ biological\_process gene product
- ▶ molecular\_function gene product

### Class Description

---

### Enter DL Query

---

This query is evaluated and returns all descending classes that are uniprot proteins. Any inferred gene product classes are removed from the results table

# GOAL query

## Gene Product Ontology Browser

Use ctrl/cmd key to select multiple terms

### ▼ GeneProduct

- ▼ cellular\_component gene product
  - ▶ cell part gene product
  - ▶ synapse part gene product
  - ▶ extracellular region gene product
  - ▶ virion part gene product
- ▼ organelle part gene product
  - ▶ intracellular organelle part gene product
  - ▶ plasma membrane-derived photosystem I gene product
  - ▶ organellar ribosome gene product
  - ▶ contractile fiber part gene product
  - ▶ flagellum part gene product
  - ▶ plasma membrane-derived thylakoid photosystem II gene product
  - ▶ axoneme part gene product

## Class Description

**Class:** flagellum part gene product

**Super class description:** 'cell projection part gene product'

'intracellular part gene product'

'organelle part gene product'

Add to query

**Equivalent class description:** 'GeneProduct and (located\_in some flagellum part)'

## Enter DL Query

This query is evaluated and returns all descending classes that are uniprot proteins. Any inferred gene product classes are removed from the results table

'flagellum part gene product'

DL Query

## Gene Results Table

Protein ID

Super Classes

# GOAL results

plasma membrane-derived photosystem II gene product

- ▶ organellar ribosome gene product
- ▶ contractile fiber part gene product

▶ **flagellum part gene product**

- ▶ plasma membrane-derived thylakoid photosystem II gene product
- ▶ axoneme part gene product
- ▶ organelle membrane gene product
- ▶ plasma membrane-derived thylakoid membrane gene product
- ▶ plasma membrane-derived thylakoid ribulose biphosphate carboxylase complex gene product
- ▶ stereocilium membrane gene product
- ▶ symbiont-containing vacuolar space gene product
- ▶ organelle lumen gene product
- ▶ organelle membrane contact site gene product
- ▶ kinocilium gene product
- ▶ stereocilium gene product
- ▶ axoneme gene product
- ▶ extracellular region part gene product

'flagellum part gene product'

DL Query

## Gene Results Table

Protein ID	Super Classes
Calcr	'regulation of mRNA stability gene product' 'GeneProduct and (located_in some acrosomal vesicle)' 'acrosomal vesicle gene product' 'microtubule-based flagellum part gene product' 'MouseGeneProduct' 'elevation of cytosolic calcium ion concentration gene product' 'activation of adenylate cyclase activity by G-protein signaling pathway gene product' 'calcitonin receptor activity gene product' 'GeneProduct and (participates_in some elevation of cytosolic calcium ion concentration)' 'osteoclast differentiation gene product' 'GeneProduct and (participates_in some regulation of mRNA stability)' 'GeneProduct and (located_in some microtubule-based flagellum part)' 'GeneProduct and (molecular_function some calcitonin receptor activity)' 'GeneProduct and (participates_in some activation of adenylate cyclase activity by G-protein signaling pathway)' 'GeneProduct and (participates_in some osteoclast differentiation)'
Fscb	'GeneProduct and (located_in some microtubule-based flagellum part)' 'peptidyl-serine phosphorylation gene product' 'GeneProduct and (participates_in some peptidyl-serine phosphorylation)' 'GeneProduct and (molecular_function some calcium ion binding)' 'microtubule-based flagellum part gene product'

# More complex GOAL query

- ▶ response to stimulus gene product
- ▶ establishment of localization gene product
- ▶ death gene product
- ▶ pigmentation gene product
- ▶ reproductive process gene product
- ▶ phosphorus utilization gene product
- ▶ nitrogen utilization gene product
- ▶ locomotion gene product
- ▶ sugar utilization gene product
- ▶ developmental process gene product
- ▶ multicellular organismal process gene product
- ▶ reproduction gene product
- ▶ cell killing gene product
- ▶ cellular component organization or biogenesis gene product
- ▶ viral reproduction gene product
- ▶ sulfur utilization gene product
- ▶ rhythmic process gene product
- ▶ biological regulation gene product
- ▶ growth gene product
- ▶ cell proliferation gene product

This query is evaluated and returns all descending classes that are uniprot proteins. Any inferred gene product classes are removed from the results table

```
'flagellum part gene product' and 'biological regulation gene product'
```

DL Query

## Gene Results Table

Protein ID	Super Classes
Calcr	'regulation of mRNA stability gene product' 'GeneProduct and (located_in some acrosomal vesicle)' 'acrosomal vesicle gene product' 'microtubule-based flagellum part gene product' 'MouseGeneProduct' 'elevation of cytosolic calcium ion concentration gene product' 'activation of adenylate cyclase activity by G-protein signaling pathway gene product' 'calcitonin receptor activity gene product' 'GeneProduct and (participates_in some elevation of cytosolic calcium ion concentration)' 'osteoclast differentiation gene product' 'GeneProduct and (participates_in some regulation of mRNA stability)' 'GeneProduct and (located_in some microtubule-based flagellum part)' 'GeneProduct and (molecular_function some calcitonin receptor activity)' 'GeneProduct and (participates_in some activation of adenylate cyclase activity by G-protein signaling pathway)' 'GeneProduct and (participates_in some osteoclast differentiation)'

# What does an OWL ontology look like?

Class: 'Ranunculus Repens'

Annotations: [in flower]  
 label "Ranunculus Repens",  
 comment "Common buttercup"^^string

SubClassOf: [in flower]  
 \* ActinomorphicFlower,  
 \* 'Apopetalous Flower',  
 \* 'Aposepalous Flower',  
 \* 'Heterosporangiate Flower',  
 Flower  
 and (hasFlowerSymmetry some 'Radial Flower Symmetry')  
 and (hasPart some  
 (Androecium  
 and (hasAndroecialFusion some Apostemonous)  
 and (hasPart some  
 (Stamen  
 and (hasPart some Filament)  
 and (hasPart some  
 (Anther  
 and (hasAntherAttachment some 'Adnate Anther Attachment')  
 and (hasDehiscenceType some 'Longitudinal Dehiscence'))))))))  
 and (hasPart some  
 (Gynoecium  
 and (hasGynoecialFusion some Apocarpous)  
 and (hasPart some  
 (Pistil  
 and (hasPart some Carpel)  
 and (hasPart some Style)  
 and (hasPart some  
 (Stigma  
 and (hasStickiness some Stickiness)  
 and (hasStigmaShape some 'Hooked Stigma Shape'))  
 and (hasPart only  
 (Carpel  
 or Stigma  
 or Style))))  
 and (hasSexualPartArrangement some 'Spiral Arrangement'))))

# Actinomorphic Flower

Class: 'Actinomorphic flower'

Annotations:

label "Actinomorphic flower"

EquivalentTo: [in flower]

\* 'Actinomorphic flower',

Flower

and (hasFlowerSymmetry some 'Radial Flower Symmetry')

# Who knows about biology?

- Biologists know about biology;
- Easier to turn a biologist into an ontologist, rather than a CS person to an ontologists then a biologist
- Do we need to craft axioms to deliver domain knowledge to an ontology?



# Can biologists write OWL?

- Yes, but why should they?
- Once we have the ontology design patterns, domain experts can fill them in
- Spreadsheets often used – tabulation of relationships between entities;
- Map contents of spreadsheets to the underlying ontology's axioms

# How can we enable biologists to semantically describe data?

- Scientists like using spreadsheets (and for good reasons)
- Put the data into spreadsheets and suck it out into ontologies
- Constrain the descriptions in the spreadsheet to conform to a particular description style
- *Semantic Spreadsheets* – RightField (<http://www.rightfield.org>) and Populous (<http://www.populous.org>)

# RightField Application

The screenshot displays the RightField application interface. On the left is a spreadsheet with columns labeled A through Q and rows numbered 1 through 48. The spreadsheet is currently empty. On the right is a configuration panel with the following sections:

- HIERARCHY**: A large empty area with the text "No ontologies loaded" centered in the middle.
- TYPE OF ALLOWED VALUES**: A section containing five radio button options:
  - "Free text"
  - "Direct subclasses"
  - "Subclasses"
  - "Instances"
  - "Direct instances"
- ALLOWED VALUES**: A section that is currently empty.

# The User View

	A	B	C	D	E	F
1	# Use this template for 3' or whole Gene expression studies when summarization probe set data will be provided in Matrix table format					
2						
3						
4	<b>Series</b>					
5	# This section describes the overall experiment					
6	title					
7	Summary					
8	Overall Design description					
9	Experiment Class (assay Type)	methylation_profiling				
10	Experiment Design type	ExperimentDesignType				
11	Technology type	microarray				
12	quality control type	QualityControlDescriptionType				
13	contributer (person or SEEK ID)	biological_replicate				
14	project	dye_swap_quality_control				
15		peer_review_quality_control				
16	<b>Samples</b>	real_time_PCR_quality_control				
17	# The Sample names in the first column must match the column headers of the Matrix table (see next worksheet).					
18	#To find the Affymetrix platform accession numbers (GPLxxxx), go to <a href="http://www.ncbi.nlm.nih.gov/projects/geo/query/browse.cgi?mode=findplatform">http://www.ncbi.nlm.nih.gov/projects/geo/query/browse.cgi?mode=findplatform</a>					
19	Sample name	title	CEL file	source name	organism	characteristics: Strain
20	SAMPLE 1				organism	
21	SAMPLE 2				organism	
22	SAMPLE 3				organism	
23	SAMPLE 4				organism	
24	SAMPLE 5				organism	
25	SAMPLE 6				organism	
26	SAMPLE 7				organism	
27	SAMPLE 8				organism	
28	SAMPLE 9				organism	
29	SAMPLE X				organism	
30						

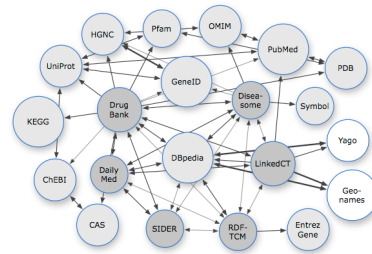
Ontology terms for marked-up cells in drop-down boxes

- QualityControlDescriptionType
- biological\_replicate
- dye\_swap\_quality\_control
- peer\_review\_quality\_control
- real\_time\_PCR\_quality\_control
- reverse\_transcription\_PCR\_quality\_control
- spike\_quality\_control
- technical\_replicate



# What can reasoning do for me?

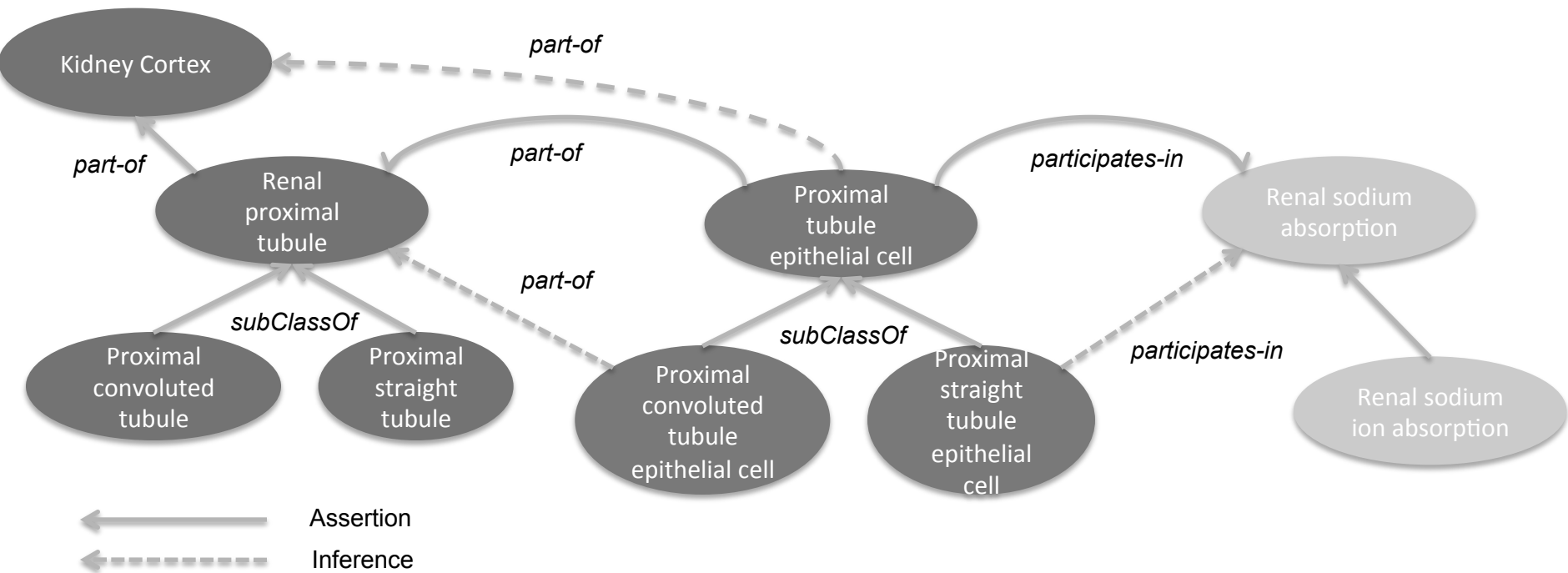
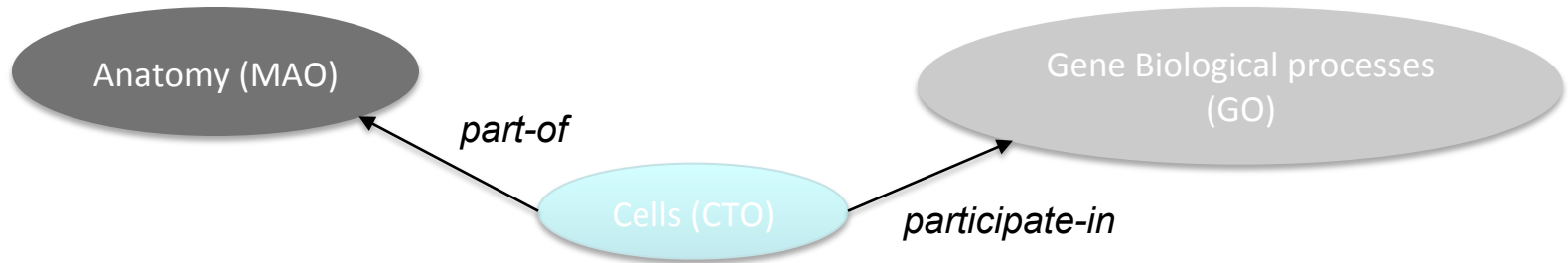
# Kidney and Urinary Knowledge Base and Ontology



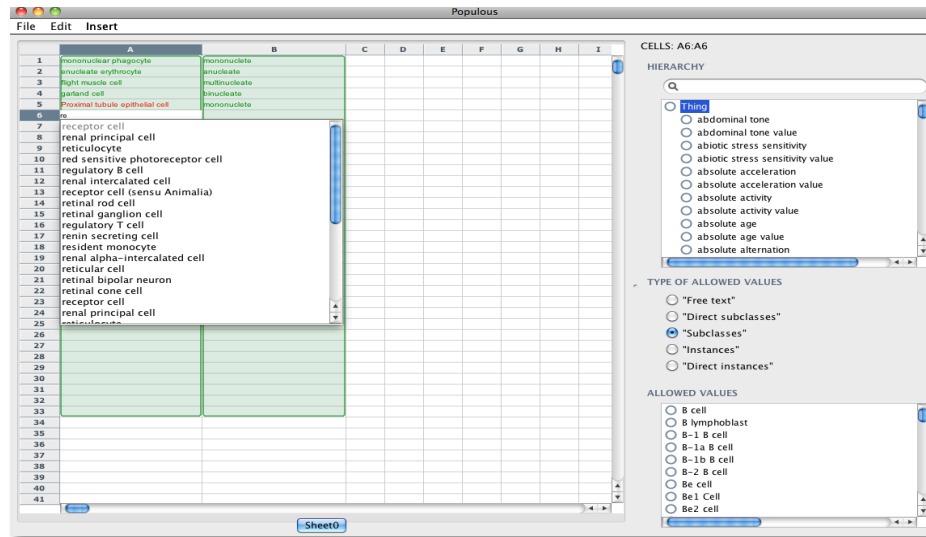
- Specialised repository of KUP related data
- KUP ontology for integration, query and inference
- Background knowledge for data mining experiments
- Collaborative update by the community

# KUPKB: KUP ontology

Each kidney cell is currently described by its localisation and function



# Populous



<http://www.e-lico.eu/populous>

- Generic tool for populating ontology templates
- Spreadsheet style interface
- Supports validation at the point of data entry
- Expressive pattern language for OWL Ontology generation



# KUPO cells in Populous

Cell type Ontology

Mouse anatomy Ontology

Biological Process (GO)

Populous - /Users/simon/Documents/e-lico/kupo\_cells\_sept10\_Populous.xls

	A	B	C	D
1	Cell Term	cell label	part_of	participates_in
2	kidney cell	renal cell	kidney	
3	kidney epithelial cell		kidney epithelium	
4	renal tubule epithelial cell		renal tubule epithelium	
5	kidney glomerular epithelial cell		kidney glomerular epithelium	
6	renal tubule cell		renal tubule	
7	kidney cortex cell	renal cortex cell	kidney cortex	
8	renal cortex tubule cell		renal cortex tubule	
9	kidney medulla cell	renal medullary cell	kidney medulla	
10	kidney outer medulla cell	renal outer	outer renal medulla	
11	kidney inner medulla cell	renal inner medullary	inner renal medulla	
12	inner renal medulla loop of henle cell		inner renal medulla loop of henle	
13	juxtaglomerular complex cell	juxtaglomerular	juxtaglomerular complex	regulation of glomerular filtration, regulation of blood circulation
14	kidney blood vessel cell	renal blood vessel	kidney blood vessel	
15	kidney arterial blood vessel cell	renal arterial blood	kidney arterial blood vessel	
16	kidney capillary endothelial cell	renal capillary cell	kidney capillary, capillary endothelium	
17	kidney venous blood vessel cell	renal blood vessel cell	kidney venous blood vessel	
18	renal corpuscule cell	cell of the renal	renal corpuscule	
19	mesangial cell		mesangium	phagocytosis, extracellular matrix constituent
20	glomerular mesangial cell		glomerular mesangium	
21	extraglomerular mesangial cell		extraglomerular mesangium	
22	podocyte	visceral epithelial	glomerular visceral epithelium	glomerular filtration, regulation of glomerular anatomical structure arrangement
23	bowmans capsule epithelial cell	epithelial cell of the glomerular parietal	bowmans capsule	
24	parietal epithelial cell		glomerular parietal epithelium	
25	glomerular cell		glomerulus	
26	glomerular capillary endothelial cell	glomerular capillary	glomerular capillary endothelium	glomerular filtration, regulation of glomerular regulation of glomerular filtration
27	renal afferent arteriole cell	afferent arteriole cell	afferent arteriole	
28	renal afferent arteriole endothelial cell	afferent arteriole	afferent arteriole, arteriole endothelium	
29	juxtaglomerular cell		part of afferent arteriole forming	renin secretion into blood stream, detection
30	renal afferent arteriole smooth muscle	afferent arteriole	afferent arteriole, arteriole smooth	
31	renal efferent arteriole cell	efferent arteriole cell	efferent arteriole	regulation of glomerular filtration
32	renal efferent arteriole endothelial cell	efferent arteriole	efferent arteriole, arteriole endothelium	
33	renal efferent arteriole smooth muscle	efferent arteriole	efferent arteriole, arteriole smooth	
34	proximal tubule epithelial cell	PTEC, proximal	renal proximal tubule	renal sodium ion absorption, potassium ion
35	proximal convoluted tubule epithelial		proximal convoluted tubule	
36	proximal straight tubule cell		proximal straight tubule	
37	loop of henle epithelial cell	henle's loop	loop of henle	extracellular matrix constituent secretion
38	loop of henle ascending limb	henle's loop	loop of henle ascending limb	renal sodium ion absorption, potassium ion
39	loop of henle thick ascending limb	henle's loop thick	loop of henle ascending limb thick	
40	loop of henle thin ascending limb	henle's loop thin	loop of henle ascending limb thin	
41	loop of henle medullary thick	henle's loop	distal straight tubule premacula	

CELLS: B17:B17

HIERARCHY

Thing

- abdominal tone
- abdominal tone value
- abiotic stress sensitivity
- abiotic stress sensitivity value
- absolute acceleration
- absolute acceleration value
- absolute activity
- absolute activity value
- absolute age
- absolute age value
- absolute alternation

TYPE OF ALLOWED VALUES

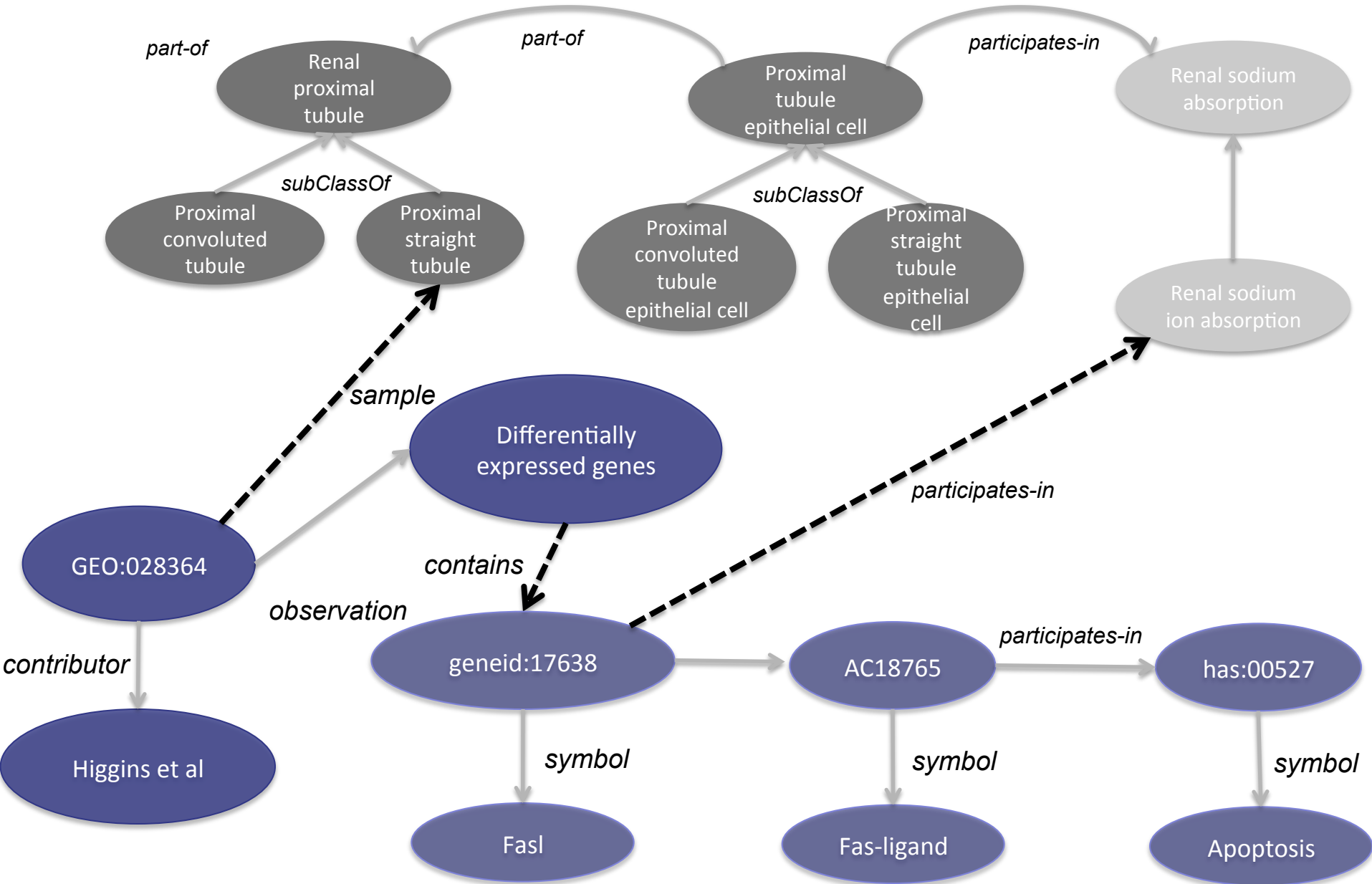
- "Free text"
- "Direct subclasses"
- "Subclasses"
- "Instances"
- "Direct instances"

ALLOWED VALUES

Any

Sheet1

# Connecting the graphs



# www.kupkb.org



## The Kidney & Urinary Pathway Knowledge Base

[iKUP Browser](#) [About](#) [Submit Data](#) [Contact](#) [Acknowledgements](#) [FAQ](#)

Search

The KUPKB is a collection of omics datasets that have been extracted from scientific publications and other related renal databases. The iKUP browser provides a single point of entry for you to query and browse these datasets.

Simply enter your gene, protein or miRNA <sup>?</sup> of interest into the query box and press search. You can search for multiple entities per line and we support a range of identifiers including entrez gene ids, gene names, uniprot ids and miRNA ids from MirBase DB. e.g. Search for TGFB1 or transforming growth factor or 3172. We have currently collected over 160 experiments, a summary of all the experiments collected is available [here](#). If you would like to submit your own datasets please choose the submit data tab above

### Results View

The results table shows the KUPKB experiments that reference your search terms. You can sort the results table by clicking on the column headers. The navigation tree below gives you a summary of your results and can be used to filter the results table.

Entity id	Species	Anatomy	Disease/Model	▲ Expression	Experiment	Type
Cell (0)						
Disease/Model (0)						
Anatomy (0)						

# iKUP search results

## Cell (1)













- ▶ Somatic cell (1)
- ▶ Epithelial cell (1)
- ▶ Barrier cell (1)
- ▶ Kidney cell (1)
- ▶ Secretory cell (1)
- ▶ Contractile cell (1)
- ▶ Motile cell (1)
- ▶ Stem cell (1)
- ▶ Defensive cell (1)
- ▶ Leucocyte (1)
- CL\_0000461 (1)
- ▶ Lower urinary tract cell (1)
- ▶ Electrically active cell (1)

## Anatomy (3)

- ▶ Kidney (3)
- ▶ Pelvis (1)
- ▶ Lower urinary tract (1)

## Disease/Model (2)

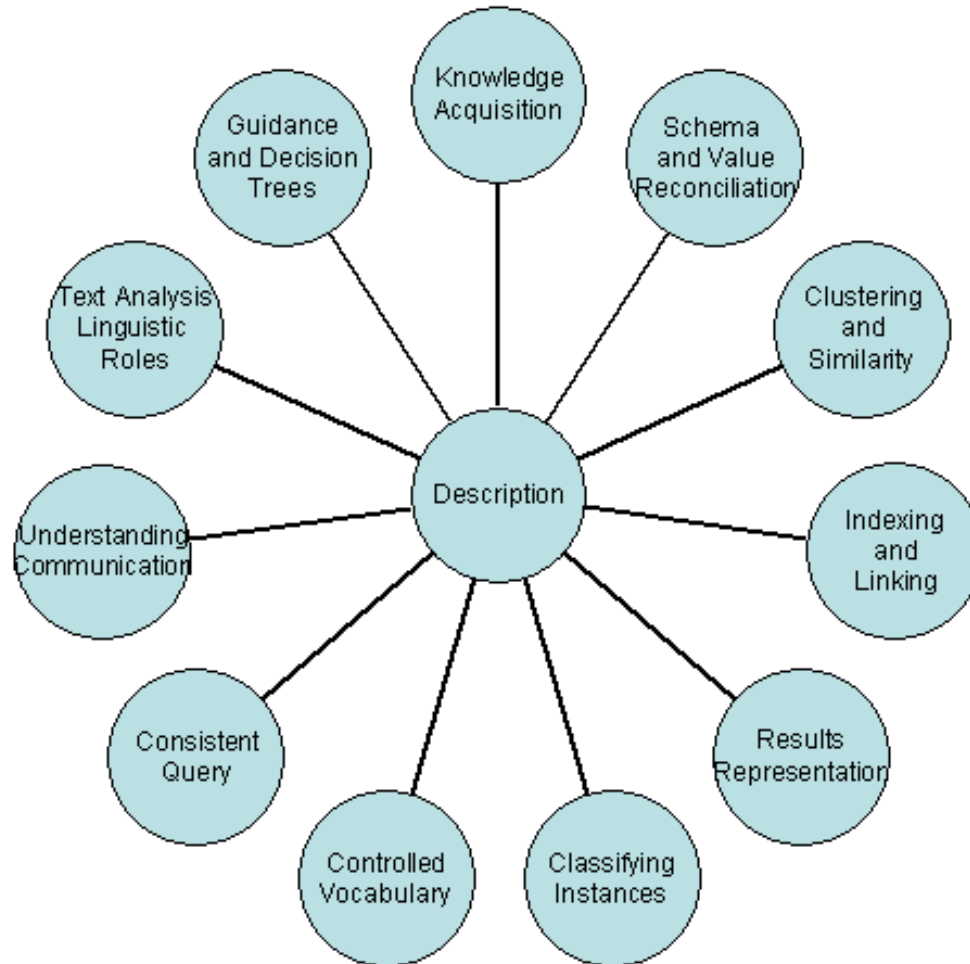
- ▶ Animal model (1)
- ▶ Urinary system disease (1)

Entity id	Species	Anatomy	Disease/Model	Expression	Experiment	Type
<a href="#">Tgfb1</a>	Rat (mature)	Distal straight tubule premacula segment	Healthy	 Absent	<a href="#">Yu, GSE13672</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (mature)	Renal proximal tubule	Healthy	 Absent	<a href="#">Yu, GSE13672</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (neonatal)	Kidney	Polycystic kidney disease model	 Unmodified	<a href="#">Koupepidou, BMC Nephrol, 2010</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (juvenile)	Kidney	Polycystic kidney disease model	 Up	<a href="#">Koupepidou, BMC Nephrol, 2010</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (neonatal)	Kidney	Polycystic kidney disease model	 Unmodified	<a href="#">Koupepidou, BMC Nephrol, 2010</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (mature)	Kidney cortex	Thy-1 nephritis rat model	 Up	<a href="#">Sadler, Kidney Int, 2005</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (mature)	Kidney	Interstitial fibrosis and tubular atrophy model	 Up	<a href="#">von Toerne, Am J Transplant, 2009</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (mature)	Kidney	Interstitial fibrosis and tubular atrophy model	 Up	<a href="#">von Toerne, Am J Transplant, 2009</a> <sup>?</sup>	mRNA
<a href="#">Tgfb1</a>	Rat (mature)	Kidney	Interstitial fibrosis and tubular atrophy model	 Up	<a href="#">von Toerne, Am J Transplant, 2009</a> <sup>?</sup>	mRNA
<a href="#">TGFB1</a>	Human	Kidney	Acute renal allograft rejection	 Down	<a href="#">Head, GSE1563</a> <sup>?</sup>	mRNA
<a href="#">TGFB1</a>	Human	Kidney	diabetic nephropathy (severe)	 Up	<a href="#">Rooney, FEBS Lett, 2011</a> <sup>?</sup>	mRNA
						

# Doing some Biology

- With the iKUP interface biologists find things to test
- These facts already existed, but difficult to find
- Easy to find if you know to look
- Integration and presentation allows one to see things you weren't looking for....

# Uses of Ontology in Bioinformatics



# Is all this doing biology?

- Ontologies enable biology;
- Ontological machinery can be used to “find new things”;
- Most often through a level of indirection to a *real* biologist
- Enabling things to be found more easily, enables hypothesis generation to be easier
- Unforeseen connections enabled through common language and common definition
- “I see these genes differentially regulated across a range of diseases”
- Can only be asked or found through common language and common description
- Can automate the application of knowledge through deduction, but need a whole lot more...

# Acknowledgements

- Simon Jupp, Julie Klein and Joost Schanstra for KUPKB
- Carole Goble and Simon Jupp for slides
- Katy Wolstencroft, Stuart Owen, Matt Horridge for RightField
- Simon Jupp for Populous
- ... and the rest of the *legion*