



# Aligning the FMA and the GO; Connecting DBs and KBs

**John H. Gennari and Adam Silberfein**

**Biomedical & Health Informatics**




**The Information School**

**University of Washington**





# Motivation





- 
- 
- 
- Genomic knowledge sources do not use anatomic knowledge
  - Knowledge sources should be aligned / connected
  - Given a GO / FMA alignment, what new functionality can we provide?



View the GO databases from within the FMA knowledge base







# The FMA

- 
- 
- 
- 
- The Foundational Model of Anatomy:  
A large KB about *human* anatomy
    - Over 70,000 terms
    - Based on physical *structure*
    - Includes a rich set of relations: is-a, part-of, adjacency, contained-in, etc.
    - From gross anatomy to “cellular anatomy”







# The gene ontology (GO)

- 
- 
- 
- 
- Goal: A consensus vocabulary for annotating gene products
  - A vocabulary organized into 3 trees:
    - Biological process
    - Molecular function
    - Cellular component
  - Has (only) 2 relations:
    - Part-of
    - Is-a







# Gene annotation DBs

- 
- 
- 
- 
- Many genomic “DBs” use GO:
    - Flybase (drosophilia)
    - MGI (mouse)
    - Wormbase (C. Elegans)
    - TIGR (many species)
    - Etc...
  - Some DBs are not in relational database format







# Mappings: FMA $\leftrightarrow$ GO

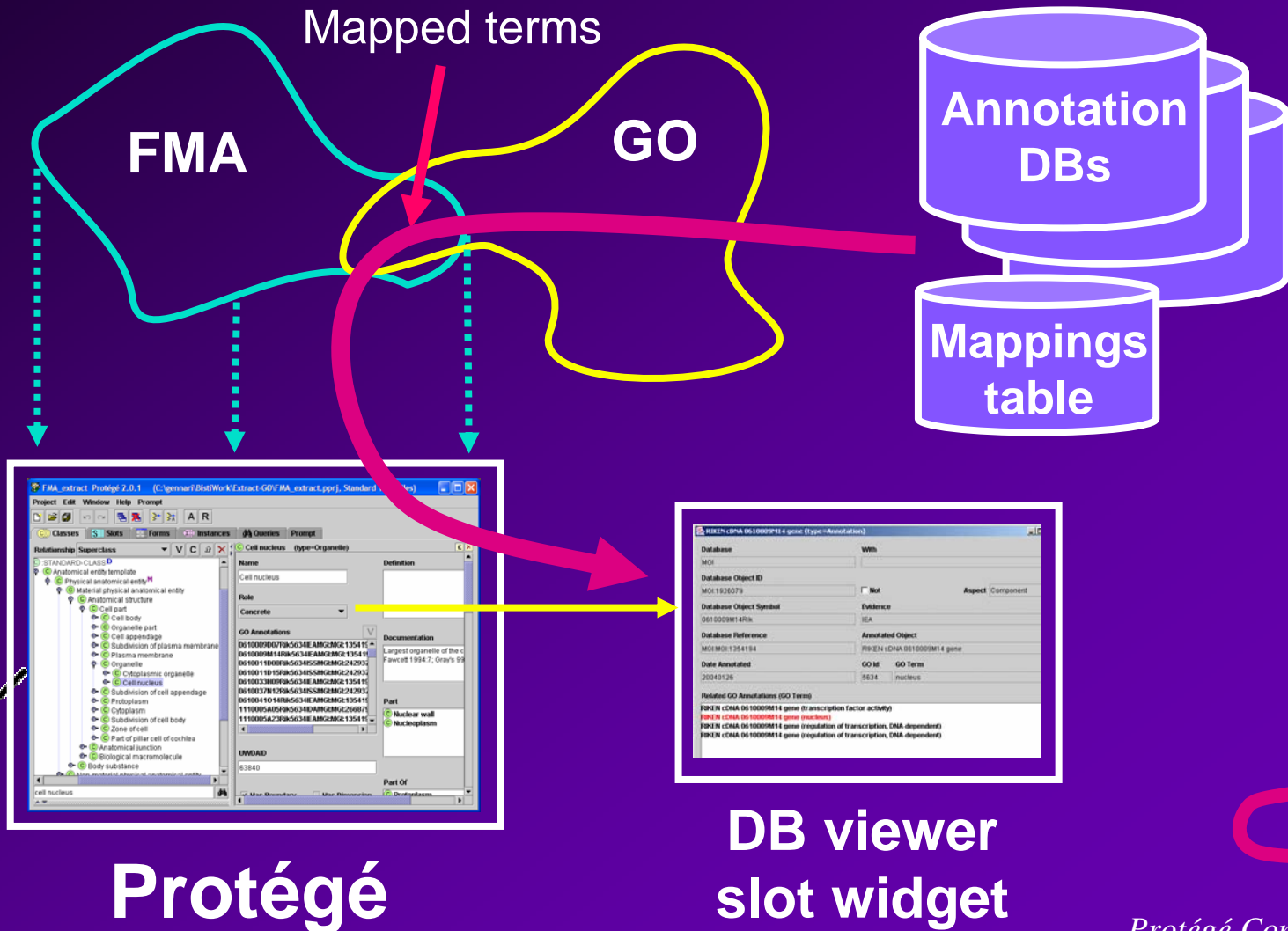
- 
- 
- 
- 
- Align cell anatomy in FMA with cellular component in GO—How?
    - Prompt
    - Model management (Bernstein)
  - Assume others will generate maps
  - Hand-connected about 150 GO terms to the FMA
    - Mostly 1-to-1 mappings
    - Complex mappings ignored



# Viewing the GO DBs

- 
- 
- 
- 
- Given
    - FMA → GO mappings
    - An annotation DB (e.g. MGI)
    - A mapped FMA term (e.g. cell nucleus)
  - Show all relevant genes
  - How? -- a DB viewer slot widget and a mappings table





# Diagrammatic overview







# The slot widget

- 
- 
- 
- 
- Added own slot (“GO annotation”) to FMA classes:  
Cell, Cell part, Cell Junction, Cell cavity, Cell substance, Macromolecule
  - Slot has no value – just an associated widget
  - Widget uses mapped GO term to retrieve annotations from DBs (via JDBC)

# Go annotations slot widget

The screenshot shows the Protégé 2.0.1 interface. On the left, a class hierarchy is displayed under 'Relationship Superclass'. The 'Cell nucleus' class is selected. On the right, the 'Cell nucleus (type=Organelle)' slot widget is shown. The 'GO Annotations' section is highlighted with a pink circle and contains the following text:

GO Annotations

- 0610009D07Rik5634IEAMGI:MGI:135419
- 0610009M14Rik5634IEAMGI:MGI:135419
- 0610011D08Rik5634ISSMGI:MGI:242937
- 0610011D15Rik5634ISSMGI:MGI:242937
- 0610033H09Rik5634IEAMGI:MGI:135419
- 0610037N12Rik5634ISSMGI:MGI:242937
- 0610041014Rik5634IEAMGI:MGI:135419
- 1110005A05Rik5634IDAMGI:MGI:266875
- 1110005A23Rik5634IEAMGI:MGI:135419

A blue callout box on the right contains the text: "DB viewer slot widget w/gene list". A pink arrow points from this box to the GO Annotations list. Another pink arrow points from the bottom of the GO Annotations list to the 'Part Of' section, which includes 'Nuclear wall' and 'Nucleoplasm'.

# Annotated object (from MGI)

RIKEN cDNA 0610009M14 gene (type=Annotation)

<b>Database</b>	With	
MGI		
<b>Database Object ID</b>	<input type="checkbox"/> Not	<b>Aspect</b> Component
MGI:1926079		
<b>Database Object Symbol</b>	<b>Evidence</b>	
0610009M14Rik	IEA	
<b>Database Reference</b>	<b>Annotated Object</b>	
MGI:MGI:1354194	RIKEN cDNA 0610009M14 gene	
<b>Date Annotated</b>	<b>GO Id</b>	<b>GO Term</b>
20040126	5634	nucleus
<b>Related GO Annotations (GO Term)</b>		
RIKEN cDNA 0610009M14 gene (transcription factor activity)		
RIKEN cDNA 0610009M14 gene (nucleus)		
RIKEN cDNA 0610009M14 gene (regulation of transcription, DNA-dependent)		
RIKEN cDNA 0610009M14 gene (regulation of transcription, DNA-dependent)		



# Limitations



## ■ User interface:

- What's the primary key?
- How to display tabular data?
- How not to overwhelm user?

## ■ KB – DB connections

- Can mappings be generated?
- Are there other candidate knowledge sources to link?

# Future work

- Make generic “DB viewer” plug-in with configuration panel:
  - Connect to any DB table
  - Options for tabular display
- Connect the FMA with other genomic knowledge sources

# Databases vs. ontologies

- Blurring the distinction
  - End-users don't care
  - KBs and DBs aren't really that different
- Some concepts best expressed in an ontology
- Some data best stored in a simple tabular DB
- Sometimes we want to connect these!



# Thank You!

The work was partially funded by NLM  
via a Bisti planning grant

We thank Jesse Wiley and Onard Mejino  
for expert help with cellular biology and anatomy