Towards a GO annotation tool

CURATION ACCELERATOR SOFTWARE
FlyBase Literature Curators

- Read papers
- Extract information for FlyBase
- Per paper, genes mentioned: per gene, any alleles: new constructs, phenotypes, controls
Annotation with Gene ontologies

- How can we make this easy? or farm it out to The Community (wikipedia model)
About Gene Ontology - GO - annotation

• Each gene should have up to 3 GO terms associated with it, describing
  • molecular function
  • biological process
  • cellular component

• Common to GMOD databases: each GO term has a standard definition designed to have the same biological meaning across the entire living world, ie for all disciplines, any organism.

• GO ontology terms designed to standardise the biology
### Search: thickening

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0052386</td>
<td>cell wall thickening</td>
</tr>
<tr>
<td>GO:0021991</td>
<td>neural plate thickening</td>
</tr>
<tr>
<td>GO:0052482</td>
<td>defense response by cell wall thickening</td>
</tr>
<tr>
<td>GO:0052105</td>
<td>induction by symbiont of defense-related host cell wall thickening</td>
</tr>
<tr>
<td>GO:0052161</td>
<td>modulation by symbiont of defense-related host cell wall thickening</td>
</tr>
<tr>
<td>GO:0052394</td>
<td>induction by organism of defense-related symbiont cell wall thickening</td>
</tr>
<tr>
<td>GO:0052446</td>
<td>modulation by organism of defense-related symbiont cell wall thickening</td>
</tr>
<tr>
<td>GO:0052539</td>
<td>positive regulation by symbiont of defense-related host cell wall thickening</td>
</tr>
<tr>
<td>GO:0052540</td>
<td>positive regulation by organism of defense-related symbiont cell wall thickening</td>
</tr>
<tr>
<td>GO:0052290</td>
<td>induction by organism of defense-related cell wall thickening in other organism involved in symbiotic interaction</td>
</tr>
<tr>
<td>GO:0052300</td>
<td>modulation by organism of defense-related cell wall thickening in other organism involved in symbiotic interaction</td>
</tr>
<tr>
<td>GO:0052538</td>
<td>positive regulation by organism of defense-related cell wall thickening in other organism involved in symbiotic interaction</td>
</tr>
<tr>
<td>GO:0075048</td>
<td>cell wall strengthening in symbiont involved in entry into host</td>
</tr>
<tr>
<td>GO:0043679</td>
<td>axon terminus</td>
</tr>
<tr>
<td>GO:0043195</td>
<td>terminal button</td>
</tr>
<tr>
<td>GO:0043049</td>
<td>otic placode formation</td>
</tr>
<tr>
<td>GO:0061174</td>
<td>type I terminal button</td>
</tr>
<tr>
<td>GO:0060790</td>
<td>tooth placode formation</td>
</tr>
<tr>
<td>GO:0061175</td>
<td>type II terminal button</td>
</tr>
<tr>
<td>GO:0021501</td>
<td>prechordial plate formation</td>
</tr>
<tr>
<td>GO:0030910</td>
<td>olfactory placode formation</td>
</tr>
</tbody>
</table>
ID: GO:0043195
Name: terminal button
Ontology: Cellular Component
Definition: Terminal inflated portion of the axon, containing the specialized apparatus necessary to release neurotransmitters. The axon terminus is considered to be the whole region of thickening and the terminal button is a specialized region of it.

All GO terms displayed in this chart can be right-clicked to display term definitions and child (descendent) terms.

Relationship key:
- part of
- is a
- has part
- regulates
+ve regulates
-ve regulates
<table>
<thead>
<tr>
<th>DB</th>
<th>ID</th>
<th>Att</th>
<th>Symbol</th>
<th>Taxon</th>
<th>Qualifier</th>
<th>GO ID</th>
<th>GO Term name</th>
<th>Reference</th>
<th>Ev</th>
<th>With</th>
<th>A</th>
<th>Date</th>
<th>From</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniProtKB</td>
<td>A0A761</td>
<td>ct-2</td>
<td></td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A2AFLG5</td>
<td>L1cam</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20060512</td>
<td>MGI</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A2AFLG6</td>
<td>L1cam</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20060512</td>
<td>MGI</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A2AFLG7</td>
<td>L1cam</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20060512</td>
<td>MGI</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A2AFLG8</td>
<td>L1cam</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20060512</td>
<td>MGI</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A2VCP3</td>
<td>P2rx7</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A3KMF0</td>
<td>Grik2</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A3KMF1</td>
<td>Grik2</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A3KN68</td>
<td>Ptpdn2</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A4LB49</td>
<td>DRD4</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A4LB56</td>
<td>DRD4</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A5JHE0</td>
<td>Kcnma1</td>
<td>10090</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A6HKB8</td>
<td>GRIK4</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A6PSC6</td>
<td>KCNMA1</td>
<td>9796</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8JNM4</td>
<td>Ank2</td>
<td>7227</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8JNM5</td>
<td>Ank2</td>
<td>7227</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8JNM6</td>
<td>Ank2</td>
<td>7227</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8JNM7</td>
<td>Ank2</td>
<td>7227</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8JNR2</td>
<td>guk</td>
<td>7227</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8KOH7</td>
<td>A8KOH7</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8K139</td>
<td>A8K139</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8K220</td>
<td>A8K220</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8K7F0</td>
<td>A8K7F0</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>A8KAF4</td>
<td>A8KAF4</td>
<td>9606</td>
<td></td>
<td></td>
<td>GO:0043195</td>
<td>terminal button</td>
<td>Compara</td>
<td>IEA</td>
<td>PMID:16537644</td>
<td>C</td>
<td>20100904</td>
<td>ENSEMBL</td>
</tr>
</tbody>
</table>
GO annotation

• Common problem: S.pombe, Zfin, Dictybase, WormBase, Protein2GO, GO consortium

• Changing format for GO annotation lines (16, 17)

• Production of GO lines in standard EDIT format eg for FlyBase

• or in standard GO format required for data input to the GO consortium database (FB submits periodically)
The GO curation tools

- Usually perl (java) front end. All directly input to database
- Web-based
- Combined ontology exploration tools - trees, graphical views, definitions - previously created ontology lines.
- (FB does not have one) FB curators edit text files.
- FB: curation process is independent of database parsing. Highly Recommended
- Database can be developed without reference to the curation interface
- Exploit this to design beautifully engineered GUI based tools
Strategy

• User designed GUI (speed + accuracy)
• Modular design of GO editor with xml config
• OOP with API for different GMODs
• Proof of concept: desktop GO tool to speed up curation > XML config > API for general use
• Translate to web service
Programming

- Java - OOP - GUI - JSP/Tomcat
- Our tool focus on reducing mouse-work, tool switches, ‘manual’ editing and textual errors
- One window, many tools
GO TOOL

Software houses GO lines
Clever editor .. makes GO lines
GO TOOL
Config XML describes which components to use.

For each component, source file database ids:

API
Type of component with display text

source file parsing database queries output formats
GO Tool +

Hello hif22. Ready.
Other tools
**GO Editor**

![Image of the GO Editor interface](image.png)

**Table:**

<table>
<thead>
<tr>
<th>Curation record</th>
<th>Provenance</th>
<th>Qualifier</th>
<th>GO search</th>
</tr>
</thead>
<tbody>
<tr>
<td>??</td>
<td>FlyBaseFB</td>
<td>??</td>
<td>??</td>
</tr>
</tbody>
</table>

**Buttons:**

- Save Copy
- Save
- Clear

**Additional GO Reference Records**

**Clear additional records**

**GO line editor**
GO EDITOR

FlyBaseFB: keratan sulfate biosynthetic process; GO:0018146 | ?? ??

Curation record | Provenance | Qualifier | GO search
??? | FlyBaseFB | keratan sulfate biosynthetic process; GO:0018146

Additional GO Reference Records | Clear additional records
### GO Editor

**FlyBaseFB**: cell wall thickening : GO:0052386 | IMP ??

<table>
<thead>
<tr>
<th>Curation record</th>
<th>Provenance</th>
<th>Qualifier</th>
<th>GO search</th>
<th>kening</th>
<th>Evidence Code</th>
<th>With/From</th>
</tr>
</thead>
<tbody>
<tr>
<td>??</td>
<td>lyBaseFB</td>
<td></td>
<td>cell wall thickening : GO:0052386</td>
<td></td>
<td>IMP</td>
<td>??</td>
</tr>
</tbody>
</table>

**Curation Record**

**Additional GO Reference Records**

**Clear additional records**

**GO line editor**
<table>
<thead>
<tr>
<th>Curation record</th>
<th>Provenance</th>
<th>Qualifier</th>
<th>GO search</th>
</tr>
</thead>
<tbody>
<tr>
<td>??</td>
<td>FlyBaseFB</td>
<td></td>
<td>cell wall thickening : GO:0052386</td>
</tr>
</tbody>
</table>

Curation Record

Additional GO Reference Records

Clear additional records
GO editor

FlyBaseFB: cell wall thickening; GO:0052386 | IMP with

Curation record | Provenance | Qualifier | GO search | kening | Evidence Code | With/From |
---|---|---|---|---|---|---|
?? | FlyBaseFB | cell wall thickening; GO:0052386 | | | IMP | with |

Curation Record

Additional GO Reference Records | Clear additional records
Add evidence gene
GO editor
GO editor
GO editor
**Information**

**FlyBaseFB: cell wall thickening ; GO:0052386 | IMP with FLYBASE:barb; FB:FBgn0066372**

**Curation record**

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>MacOS</th>
<th>GO-line</th>
</tr>
</thead>
</table>

**Evidence gene 0 bar**

**cell wall thickening**

id: GO:0052386
name: cell wall thickening
namespace: biological_process
def: "A type of cell wall modification in which the cell wall is reinforced and made thicker." [GOC:mtg_p is_a: GO:0042545 ! cell wall modification

**barb FBgn0066372**

<table>
<thead>
<tr>
<th>Evidence Code</th>
<th>With/From</th>
</tr>
</thead>
</table>

**Additional GO Reference Records**

**Clear additional records**

**GO line editor**
Save work
Open curation record
# Open Curation Record

<table>
<thead>
<tr>
<th>TimeStamp</th>
<th>MacOS</th>
<th>GO-line</th>
</tr>
</thead>
</table>

**wg FlyBaseFB: cell wall thickening** | GO:0052386 | IMP with **FLYBASE:barb**; FB:FBgn0066372

**cell wall thickening**
- **id:** GO:0052386
- **name:** cell wall thickening
- **namespace:** biological_process
- **def:** "A type of cell wall modification in which the cell wall is reinforced and made thicker."
- **is_a:** GO:0042545 | cell wall modification

**barb FBgn0066372**
- **a barb**
- **z FBgn0066372**
- **e barbos**
- **T supplementary material**
- **T learning or memory** | GO:0007611 | IMP --- FlyBase @ 20060803
- **T effectors:binding** | GO:0008255 | IMP --- FlyBase @ 20060803

**Additional GO Reference Records**
- **null:** FlyBaseFB: cell wall thickening | GO:0052386 | IMP with FLYBASE:barb; FB:FBgn0066372
- **st1290.edit FBrf0190314:wg** frizzled-2 binding | GO:0005110 | IPI with FLYBASE:zf2; FB:FBgn0016797
- **st1290.edit FBrf0190314:zf2** Wnt–protein binding | GO:0017147 | IPI with FLYBASE:wg; FB:FBgn0004009

Curation record(s): /Users/hif22/astrecords/st1290.edit
Open curation record
EDIT EXISTING
EDIT EXISTSING

FlyBaseFB: Wnt–protein binding; GO:0017147 | IPI with

Curation record

wg

Provenance

FlyBaseFl

Quali

Wnt–protein binding; GO:0017147

Evidence Code

kening

With/From

IPI

null

FlyBaseFB: cell wall thickening; GO:0052386 | IMP with FLYBASE:barb; FB:FBgn0066372

st1290.edit FBr0190314: wg frizzled–2 binding; GO:0005110 | IPI with FLYBASE: fz2; FB:FBgn0016797

st1290.edit FBr0190314: arr Wnt–protein binding; GO:0017147 | IPI with FLYBASE: wg; FB:FBgn0004009

Curation record(s): /Users/hif22/astrecords/st1290.edit
The end.
almost
DETAILS

- Mac OS X (testing) 10.5 and 10.6 preAlpha_v002
- 3.8GB RAM but may use less
- Flat file input but also talks to postgres
- Configurable (text file)
- Designed by the CURATORS
- Thanks to FlyBase Curators: Steve Marygold, Gillian Millburn, Steph Bunt, Simon Reeve, Pete McQuilton, Dave Osumi Sutherland, Ray Stefancsik, Nick Brown and Advisory Board members
Future

- chado queries
- XML config/s for clever editor
- API in collaboration
- JSP or J2EE/SOAP web interface
  - Helen Imogen Field
  - hif22 @ gen.cam.ac.uk