Binding term working group

Camp Co-chairs: Ursula Hinz & Ruth Lovering

Minutes: Jim Hu & Damien Lieberherr

Mike Cherry	SGD
Elisabeth Coudert	SIB
Peter D'Eustachio	Reactome
Emily Dimmer	GOA
Harold Drabkin	MGI
Serenella Ferro Rojas	SIB
Rebecca Foulger	GOC editor
Pascale Gaudet	DictyBase
Michelle Gwinn Giglio	PAMGO
Thomas Hayman	RGD
Ben Hitz	SGD
David Hill	MGI
Jim Hu	Ecoli
Suzi Lewis	BBOP
Jane Lomax	GOC editor
Chris Mungall	BBOP
Li Ni	MGI
Lakshmi Pillai	AgBase
Deborah Siegele	Ecoli
Paul Thomas	PANTHER
Shur-Jen Wang	RGD

Agenda

1. Current GO guidelines on binding (Ursula 20-30 min)
- Background and examples
- Quality control checks now operational

2. Summary of ontology development (Chris 15 min)
- How users will retrieve information on specific chemical substrates
- What impact will this have on GO annotation

3. New developments since the GOC meeting in March (Ruth 30-40 min)
- Use of column 16 for function and process terms
- Use of column 16 relationship ontologies
- Not using IPI in conjunction with a catalytic activity annotation

4. Discussion of annotation examples (Ursula/Ruth 15-30 min)

Binding guidelines

http://wiki.geneontology.org/index.php/Binding_Guidelines

Binding background

- ✓Binding a biological entity
- ✓Binding of macromolecules
- ✓Binding of small ligands (ions and organic molecules)

Binding of macromolecules

GO:0005515 protein binding

with numerous child terms for

protein classes

GO:0003779 actin binding
GO:0031493 nucleosomal histone binding

specific proteins

GO:0002039 p53 binding
GO:0043130 ubiquitin binding

parts of proteins

GO:0030274 LIM domain binding
GO:0070064 proline-rich region binding

GO:0005515 protein binding

always annotate with **IPI**

Quality check control for GO:0005515 and IPI:
must use the **"with" column** to indicate the **interaction partner**

Always do reciprocal annotation

Example
Annotation 1: protein A GO:0005515 IPI protein B
Annotation 2: protein B GO:0005515 IPI protein A

GO:0005515: use of specific child terms

Preferentially use the most specific child terms.

Example:

The ubiquitin-specific protease CYLD (Q9NQC7) interacts with the kinase PLK1 (P53350) (PubMed=17495026)

Annotation 1:
PLK1 GO:0002020 (protease binding) IPI CYLD

Annotation 2:
CYLD GO:0019901 (protein kinase binding) IPI PLK1

Annotation using child terms of GO:0005515

Use IPI for interactions with specific proteins
Use IDA for interactions with classes of proteins (where it is not possible/desirable to provide an identifier)

Examples:

NOS3 GO:0003779 actin binding IPI P60709 (ACTB PMID:17502619)
Myrip GO:0003779 actin binding IDA (PMID:12221080)

GO:0005515 protein binding

should not be propagated with the evidence code **ISS**

The qualifier pertains to the **GO term**, not the AC of the interaction partner

SWIS	Q6PIJ4	GO:0005515 protein binding	ISS	⇐	GO_REF	0000024	SPTR	Q6P4R8
onent								
SWIS	Q6PIJ4	GO:0005634 nucleus	ISS	⇐	GO_REF	0000024	SPTR	Q6P4R8

ok: propagation with ISS for child terms of GO:0005515 that name the interaction partner:

GO:0003779 actin binding
GO:0019900 kinase binding
Use curator judgement and annotate what is physiologically relevant

The qualifier “Not” should not be used with GO:0005515 protein binding

The qualifier pertains to the **GO term**, not the AC of the interaction partner

Using the qualifier **“Not”** with **GO:0005515** implies that the protein in question does not bind anything, ever.

Exception: It is acceptable to make a **NOT** annotation to a child term of protein binding - such as **'GO:0051529 NFAT4 protein binding**’.

Binding background

- ✓Binding a biological entity
- ✓Binding of macromolecules
- ✓Binding of small ligands (ions and organic molecules)

  GO:0005509 calcium ion binding
  GO:0005524 ATP binding

Avoid redundant annotation

Enzymes must bind their substrates and cofactors to effect catalysis. Likewise, transporters must bind the molecules or ions they transport

Therefore, one should NOT add GO terms for interactions with substrates, etc

Adding the term “ATP-binding” for an ATPase is redundant.

It is legitimate to add the term “ATP-binding” when a protein binds an ATP analog, and the publication does not show that the protein can hydrolyze ATP

http://gocwiki.geneontology.org/index.php/Binding_terms_working_group#September_4.2C_2009

Human FTO (Q9C0B1)

an iron-dependent dioxygenase involved in DNA repair (PubMed=20376003)

Proposed annotation:

Process:

GO:0006307 (DNA dealkylation involved in DNA repair)

Mol. Function:

GO:0035516 (oxidative DNA demethylase activity)

GO:0043734 (DNA-N1-methyladenine dioxygenase activity)

Definition: Catalysis of the oxidative demethylation of N1-methyladenine and N3-methylcytosine in DNA and RNA, with concomitant decarboxylation of 2-oxoglutarate and releases oxidized methylgroup on N1-methyladenine and N3-methylcytosine as formaldehyde.

Human FTO (Q9C0B1)

an iron-dependent dioxygenase involved in DNA repair (PubMed=20376003)

Process:
GO:0006307 (DNA dealkylation involved in DNA repair)
Mol. Function:
GO:0035516 (oxidative DNA demethylase activity)
GO:0043734 (DNA-N1-methyladenine dioxygenase activity)
GO:0008198 (ferrous iron binding)

Redundant?
GO:0032131 alkylated DNA binding (child of
GO:0003684 damaged DNA binding)
GO:0019825 oxygen binding

Future ontology developments are required to make sure that gene products can also be identified via their ligands

Chris/Jane to discuss

Thank you

Recent developments

- Use of column 16 for function and process terms
- Use of column 16 relationship ontologies
- Not using IPI in conjunction with a catalytic activity annotation

Column 16

- An ID for the target of the annotation can be included in column 16
- This should be a direct target
- Curator judgement is required to interpret the experiments

With Column (8) target

Organizer restriction through modulation of Bozozok stability by the E3 ubiquitin ligase Lnx-like.

Ro H, Dawid IB.

Nat Cell Biol. 2009 Sep;11(9):1121-7.

PMID: 19668196

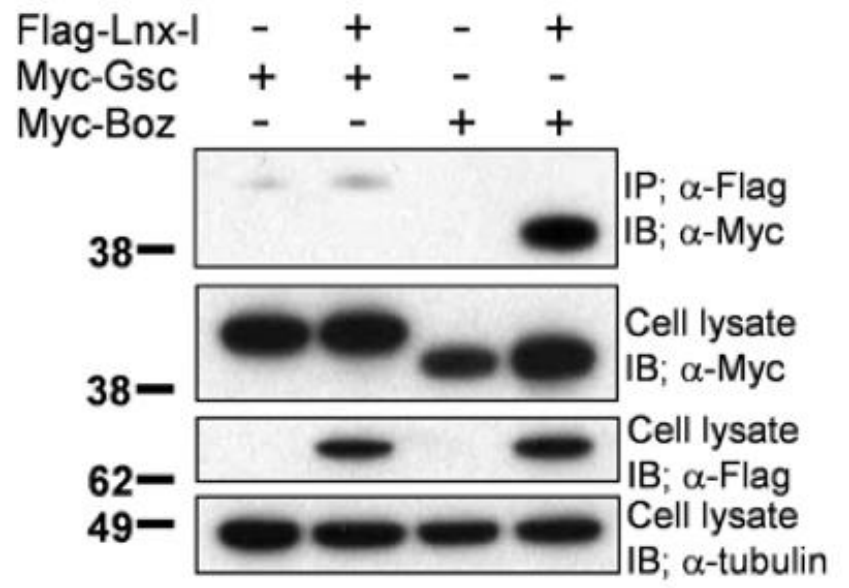
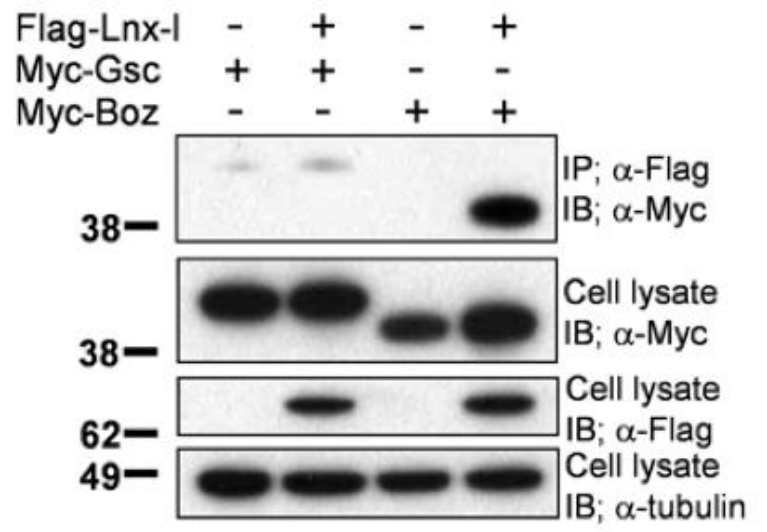


Figure legend: Lnx-l interacts with Boz but not with Gsc. Flag-tagged Lnx-l was co-transfected into 293T cells with 6xMyc-tagged Gsc or Boz. After 48 h, IP and IB were performed with the indicated antibodies.

Paragraph in Results section: To check the specificity of the Boz-Lnx-l interaction we tested whether Gsc, an organizer homeodomain protein and the closest homolog of Boz, could act as a substrate of Lnx-l. Binding between Gsc and Lnx-l was very weak and probably not significant as expression levels of Gsc were not substantially affected by co-transfection with Lnx-l.

GO annotation



Name	ID	GO term	Evidence	With ID	With (name)
Lnx-I	A4VCF7	transcription factor binding	IPI	O93236	Boz
Boz	O93236	ubiquitin protein ligase binding	IPI	A4VCF7	Lnx-I

Column 16 target

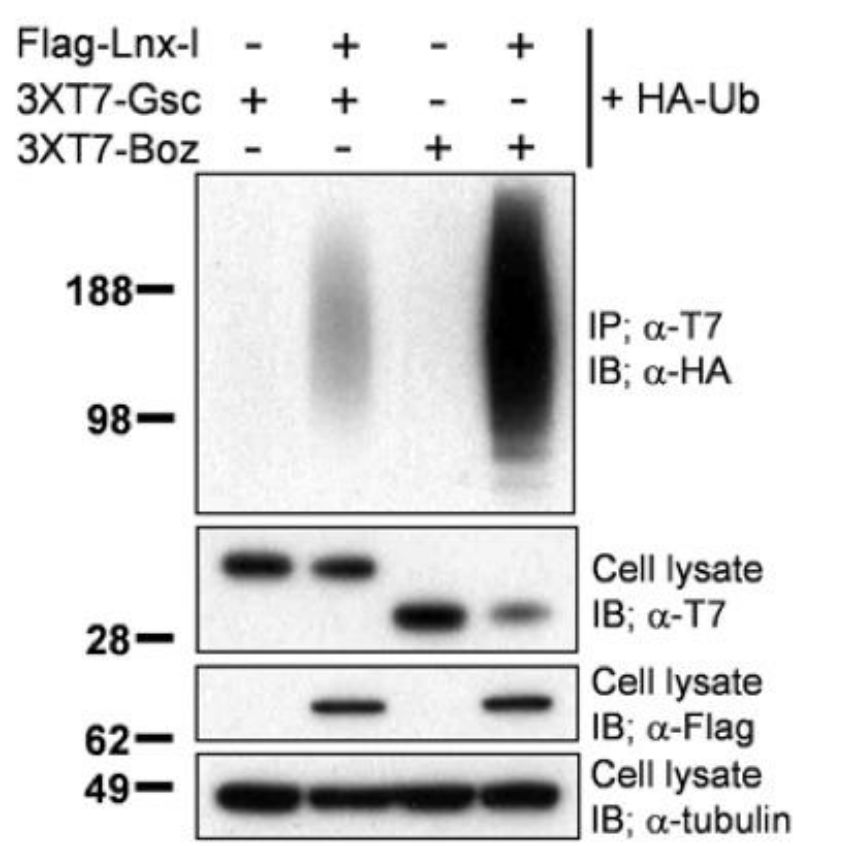
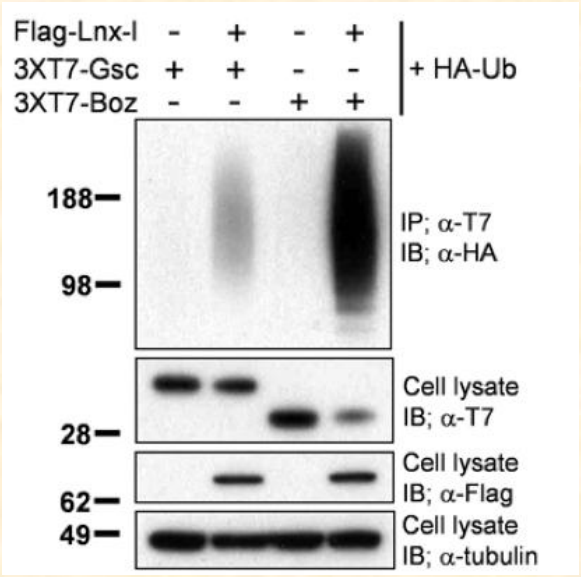


Figure legend: Lnx-l ubiquitinates Boz but not Gsc. Flag-tagged Lnx-l was co-transfected with 3xT7 tagged Gsc or Boz. The yield of polyubiquitinated Gsc is very low compared to that of Boz.

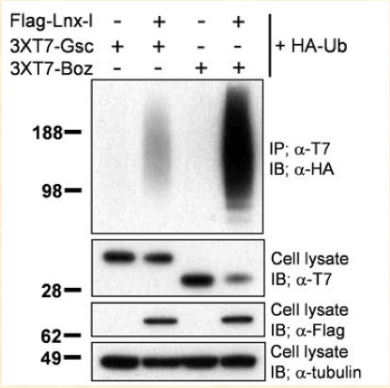
Paragraph in Results section: To check the specificity of the Boz-Lnx-l interaction we tested whether Gsc, an organizer homeodomain protein and the closest homolog of Boz, could act as a substrate of Lnx-l. Polyubiquitination of Gsc by Lnx-l was very weak and probably not significant as expression levels of Gsc were not substantially affected by co-transfection with Lnx-l. These data indicate that Boz is a specific target of Lnx-l.

GO annotation



Name	ID	GO term	Evidence	Column 16 ID	Column 16 (name)
Lnx-I	A4VCF7	ubiquitin-protein ligase activity	IDA	O93236	Boz
Lnx-I	A4VCF7	protein ubiquitination	IDA		

GO annotation

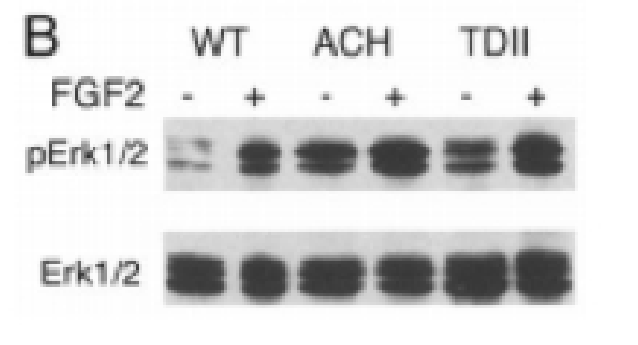


Name	ID	GO term	Evidence	Column 16 ID	Column 16 (name)
Lnx-I	A4VCF7	ubiquitin-protein ligase activity	IDA	O93236	Boz
Lnx-I	A4VCF7	protein ubiquitination	IDA		

Name	ID	GO term	Evidence	With ID	With (name)
Lnx-I	A4VCF7	transcription factor binding	IPI	O93236	Boz
Boz	O93236	ubiquitin protein ligase binding	IPI	A4VCF7	Lnx-I

When not to use column 16

Defective lysosomal targeting of activated fibroblast growth factor receptor 3 in achondroplasia.
Cho JY, Guo C, Torello M, Lunstrum GP, Iwata T, Deng C, Horton WA.
Proc Natl Acad Sci U S A. 2004 Jan 13;101(2):609-14. PMID: 14699054



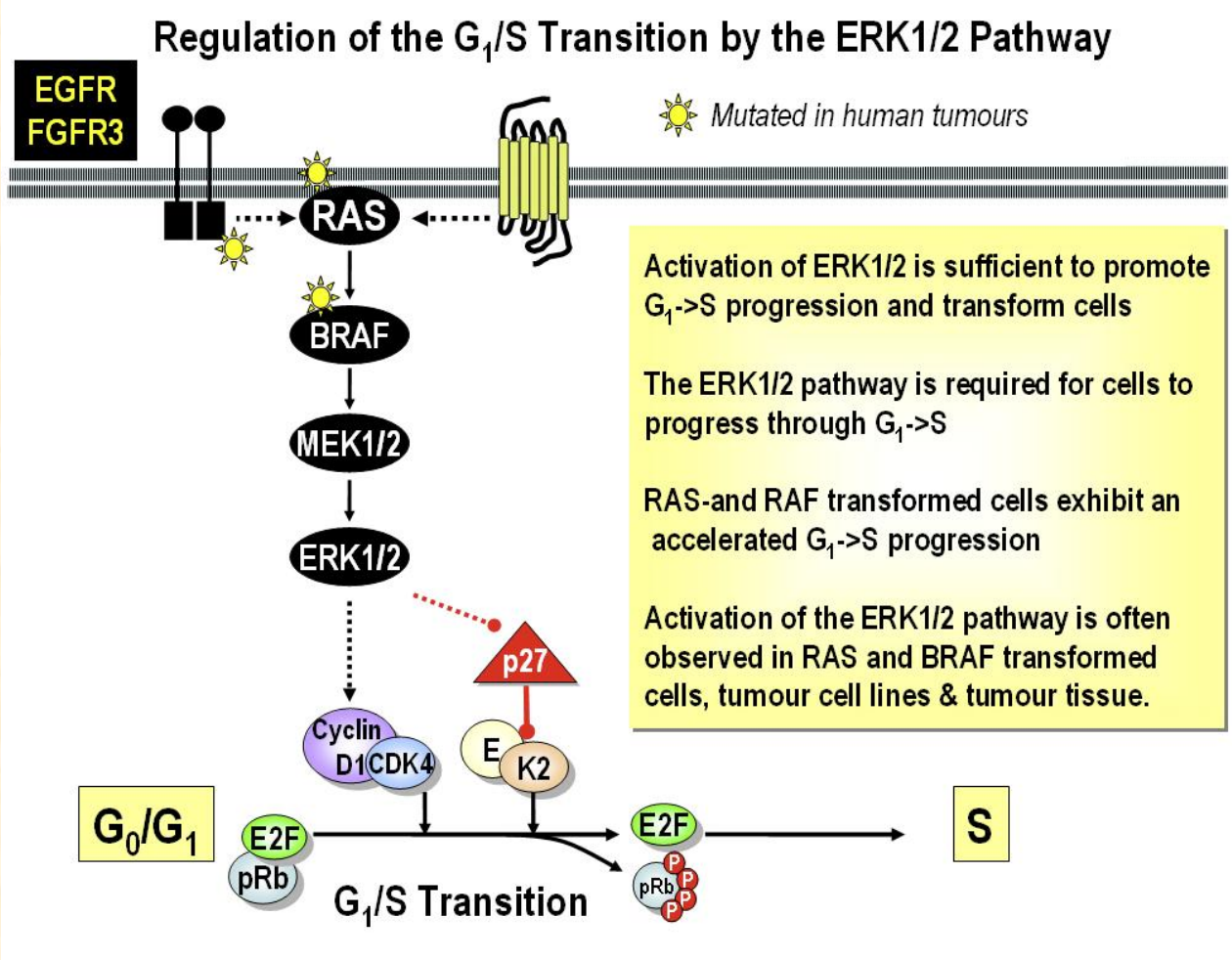
Mouse Fgfr3 constructs transfected into rat chondrocyte cells (RCJ)
WT: Wild type
ACH: G380R
TDII: K650E

Figure legend: Maximal phosphorylation of Erk1/2, a downstream signaling target of FGFR3, requires ligand in the presence of WT FGFR3 but is constitutive in the presence of mutant FGFR3.

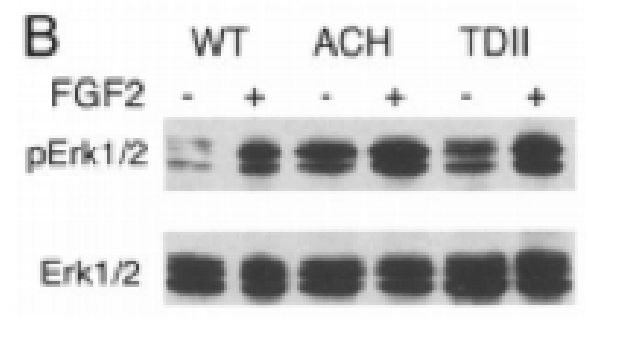
Paragraph in Results section: Erk1/2, which is known to be phosphorylated downstream of FGFR3 activation, is also phosphorylated in the absence of ligand in cells expressing mutant but not WT FGFR3.

Note where is control lane?

ERK1/2 not a direct target of FGFR3



GO annotation



Name	ID	GO term	Evidence	Reference
Fgfr3	Q61851	positive regulation of protein amino acid phosphorylation	IDA	14699054
Fgfr3	Q61851	fibroblast growth factor receptor signaling pathway	IDA	14699054
Fgfr3	Q61851	fibroblast growth factor receptor activity	IDA	14699054

Recent developments

- Use of column 16 for function and process terms
- Use of column 16 relationship ontologies
- Not using IPI in conjunction with a catalytic activity annotation

Column 16 relationship ontology

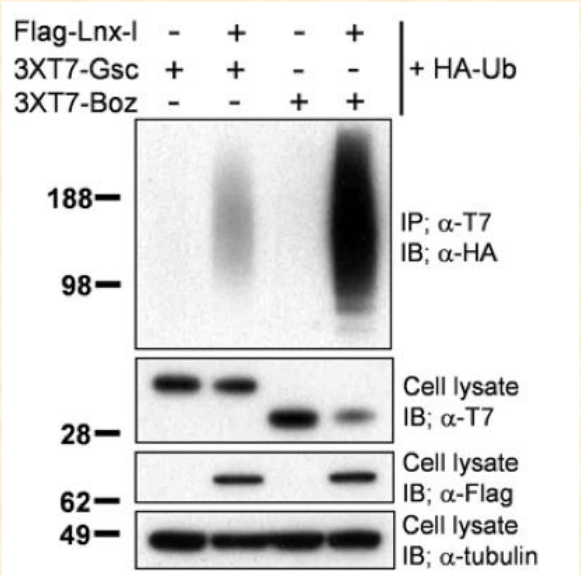
Relation	Column 4 (core term)	Column 16
has_participant	Function	CHEBI or gene product
has_input	Function	CHEBI or gene product
has_output	Function	CHEBI or gene product

has_participant, participates in the reaction

has_input, is the substrate in the reaction

has_output, is the product of the reaction

Column 16 relationships



Organizer restriction through modulation of Bozozok stability by the E3 ubiquitin ligase Lnx-like.
Ro H, Dawid IB.
Nat Cell Biol. 2009 Sep;11(9):1121-7. PMID: 19668196

Name	GO term	Evidence	Relationship	Column 16 ID	Column 16 (name)
Lnx-I	ubiquitin-protein ligase activity	IDA	has_input	O93236	Boz

Column 16 and ChEBI IDs

- Small molecule IDs, such as ChEBI IDs can also be included in column 16

- But only if these IDs add more information than is available from the GO term itself

Annotation of enzymes

Role of steroid 11 beta-hydroxylase and steroid 18-hydroxylase in the biosynthesis of glucocorticoids and mineralocorticoids in humans.
Kawamoto T, Mitsuuchi Y, Toda K, Yokoyama Y, Miyahara K, Miura S, Ohnishi T, Ichikawa Y, Nakao K, Imura H, et al.
Proc Natl Acad Sci U S A. 1992 Feb 15;89(4):1458-62.PMID: 1741400

Table 1. Steroid hydroxylase activity of mitochondria in COS-7 cells transfected with pSV11β(*CYP11B1*) or pSVC18 (*CYP11B2*)

Substrate	Products	Retention time, min	Hydroxylase activity, pmol/mg of protein	
			pSV11β	pSVC18
DOC	Corticosterone	52	482.5 ± 19.0	437.9 ± 13.8
	18-Hydroxycorticosterone	29	2.7 ± 0.6	14.1 ± 4.3
	Aldosterone	24	<0.02	2.0 ± 0.5
11-Deoxycortisol	Cortisol	33	411.4 ± 23.3	392.5 ± 18.5
	18-Hydroxycortisol	21	1.1 ± 0.1	7.0 ± 0.7
	18-Oxocortisol	18	<0.02	1.8 ± 0.1
Corticosterone	18-Hydroxycorticosterone	29	1.5 ± 0.3	10.6 ± 2.8
	Aldosterone	24	<0.02	0.9 ± 0.2
Cortisol	18-Hydroxycortisol	21	0.7 ± 0.2	4.0 ± 0.8
	18-Oxocortisol	18	<0.02	0.4 ± 0.1

ChEBI IDs

Table 1. Steroid hydroxylase activity of mitochondria in COS-7 cells transfected with pSV11β(*CYP11B1*) or pSVC18 (*CYP11B2*)

Substrate	Products	Retention time, min	Hydroxylase activity, pmol/mg of protein	
			pSV11β	pSVC18
Corticosterone	18-Hydroxycorticosterone	29	1.5 ± 0.3	10.6 ± 2.8
	Aldosterone	24	<0.02	0.9 ± 0.2

Name	GO term	Evidence	Relationship	Column 16 ID	Column 16 (name)
CYP11B2	Steroid hydroxylase activity	IDA	has_input	CHEBI: 16827	Corticosterone
CYP11B2	Steroid hydroxylase activity	IDA	has_output	CHEBI: 27584	Aldosterone
CYP11B2	aldosterone biosynthetic process	IDA			

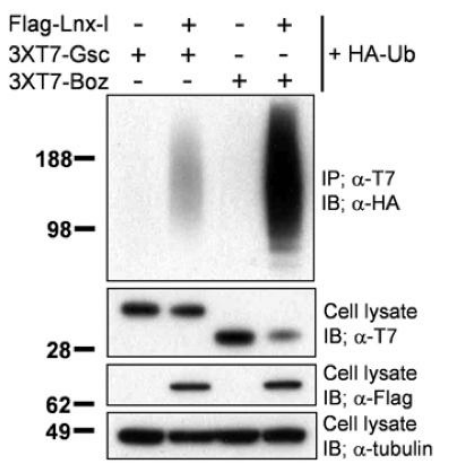
Recent developments

- Use of column 16 for function and process terms
- Use of column 16 relationship ontologies
- Not using IPI in conjunction with a catalytic activity annotation

IPI and catalytic activity

- The IPI evidence code (Inferred from Physical Interaction) - support annotations to 'GO:0005488; binding' and child terms
- Not enough information can be obtained from a binding interaction to support an annotation to 'GO:0003824; catalytic activity' or child terms
- Propose new QC check
  - identification of all IPI annotations to 'GO:0003824; catalytic activity' and child terms
  - advise no further annotations are made in this way
- 144 annotations were initially found in the GOdb that used catalytic activity + IPI - this has gone down to 141

No further IPI annotations like this



Organizer restriction through modulation of Bozozok stability by the E3 ubiquitin ligase Lnx-like.
Ro H, Dawid IB.
Nat Cell Biol. 2009 Sep;11(9):1121-7. PMID: 19668196

Name	ID	GO term	Evidence	With ID	With (name)
Lnx-I	A4VCF7	ubiquitin-protein ligase activity	IPI	O93236	Boz

Name	ID	GO term	Evidence	Column 16 ID	Column 16 (name)
Lnx-I	A4VCF7	ubiquitin-protein ligase activity	IDA	O93236	Boz

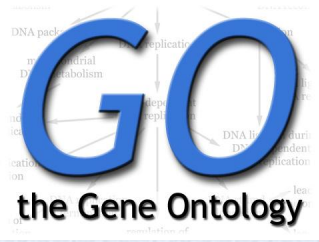
Longer term: Interaction with IMEx Consortium?

- The IMEx Consortium (IntAct, DIP, MINT, BioGrid, MPact, MatrixDB and MPIDB) carry out high-quality manual annotation of protein-protein interactions (http://disber.net/imexdrupal/node/1) from peer-reviewed papers.

- The annotation format used by IMEx captures full details of participants, assay and detection methods and meets the PSI **Minimum Information about a Molecular Interaction eXperiment (MIMIx)** standard.

- IMEx groups are working to produce a complete, non-redundant annotation set, that can be downloaded from an IMEx web service.

- Many GOC groups may already work with an IMEx member, capturing interaction data that is MIMIx or IMEx compliant



Longer term: Interaction with IMEx Consortium?

- Those at the GO Consortium meeting in April agreed it would be sensible to investigate a closer interaction with IMEx curation activities.

- An IMEx central website is being developed that will incorporate an option to contact the consortium to ask for a paper to be curated, making it possible for all GO groups to have their curation requests triaged between the different IMEx databases

- Groups can decide to contribute annotations using IMEx/MIMIx standards or just integrate annotations from IMEx; and decide what types of interaction data they would prefer to display

- Curators in UniProtKB will shortly contribute all protein binding annotations directly to the IntAct group, using a dedicated editor applying MIMIx /IMEx (editor will be integrated into the protein2go tool).

- All binding annotations meeting UniProtKB quality standards will be integrated into the UniProtKB-GOA dataset from IntAct.