Annotating to Downstream Processes

Downstream Process Working Group

GO Annotation Camp
Geneva
16-18 June 2010

Downstream Process Working Group members

Co-chairs; Rachael Huntley and Varsha Khodiyar

Members;

Yasmin Alam-Faruque	GOA	Suzi Lewis	BBOP
Rama Balakrishnan	SGD	Ruth Lovering	BHF-UCL
Alex Diehl	MGI	Tim Lowry	RGD
Serenella Ferro-Rojas	SIB	Fiona McCarthy	AgBase
Pascale Gaudet	DictyBase	Li Ni	MGI
Ursula Hinz	SIB	Sylvain Poux	SIB
Ranjana Kinshore	WormBase	Alan Ruttenberg	MIT
Val Wood	PomBase	Jennifer Smith	RGD

Minutes; Yasmin Alam-Faruque and Ursula Hinz

Outline of session

Definition of downstream processes

Types of downstream annotations (examples)

Results of initial survey

Proposed guidelines

Results from annotation camp survey

Discussion

Definition of downstream process

"For any gene product a downstream process is any biological process which that gene product is not an essential component of or does not directly regulate, but whose action initiates a series of steps leading to a subsequent biological process"

Downstream Process annotations are made from experiments describing:

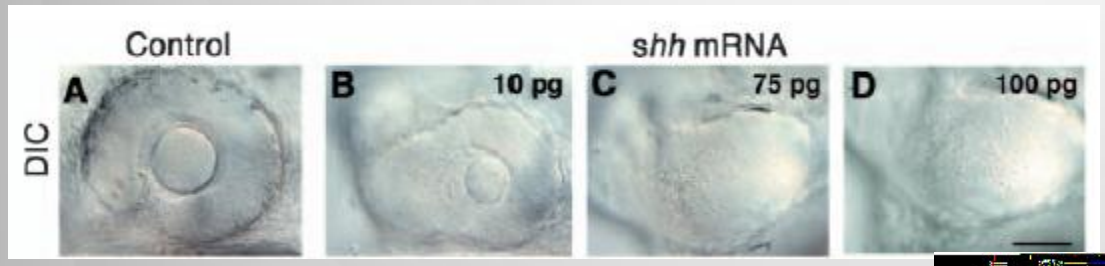
1. Development

2. Mutant phenotypes (HTP/LTP)

3. Signaling cascades

4. Insufficient evidence for a direct role of the gene product

1. Example of a downstream process annotation in development

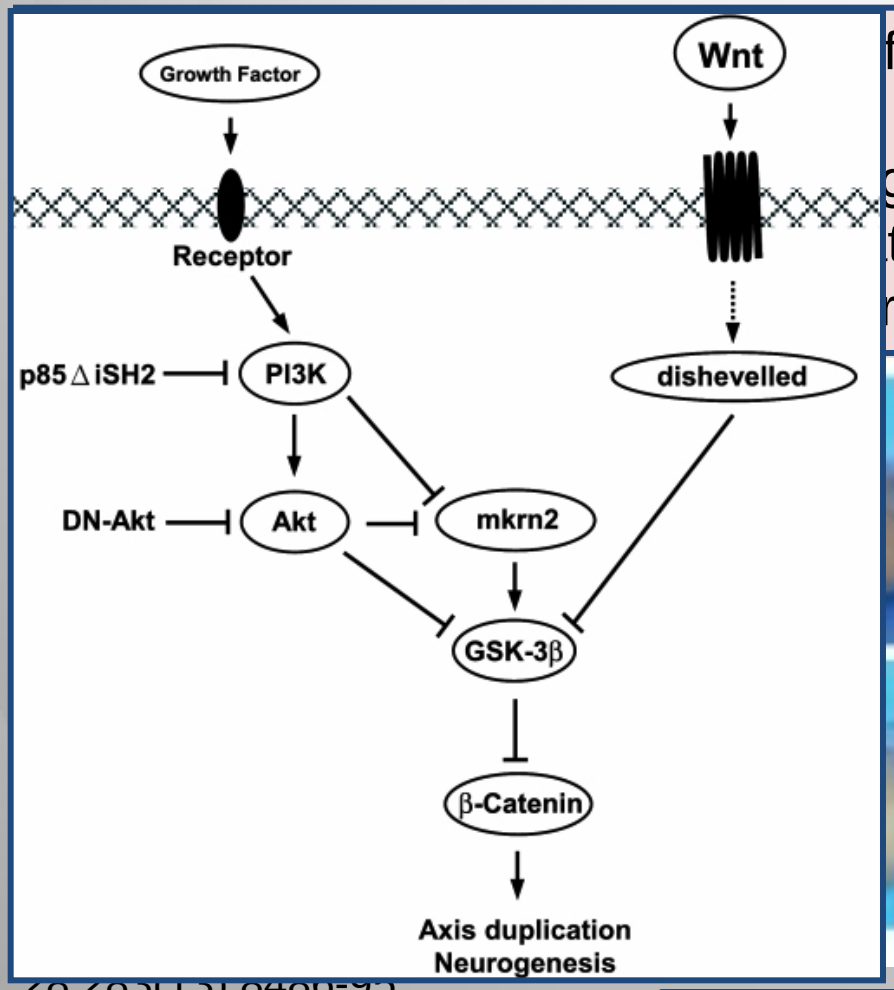


***Example*:** Zebrafish Sonic hedgehog protein A (intercellular signaling)
PMID: 15728669 Shha inhibits formation of eye lens.

"We find that *shh* mRNA injected embryos lack lens"

***Annotation of Shha*:**
GO:0043010 camera-type eye development IDA

2. Example of a downstream process annotation from a mutant phenotype

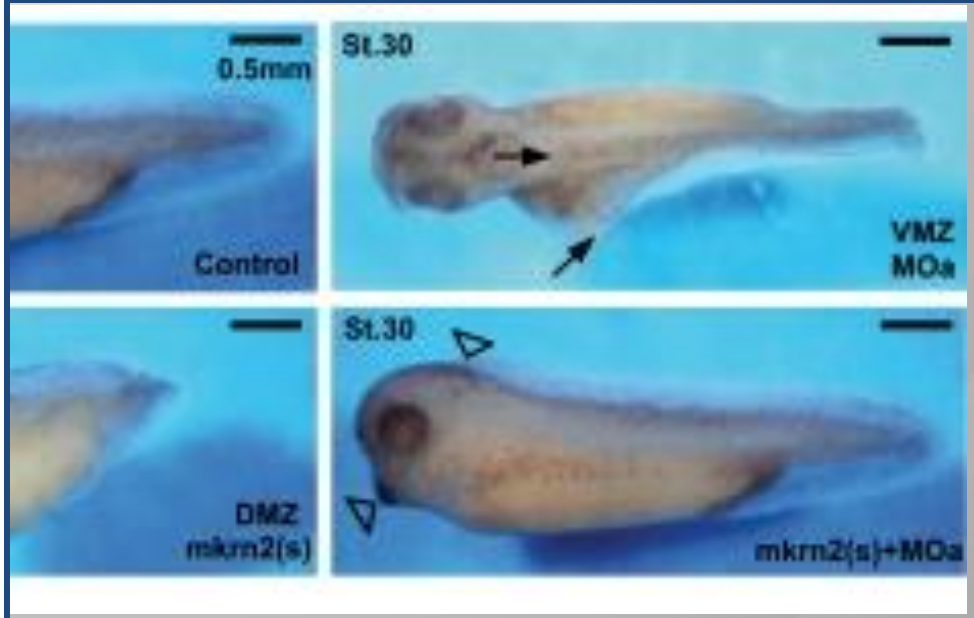


28;283(13):8486-95

fu[illegible]nction causes embryonic death

Fig 8 J Biol Chem. 2008; 28;283(13):8486-95

g tadpoles displayed axis duplication….
t mkrn2 plays a negative role in
ring *Xenopus* embryonic development."

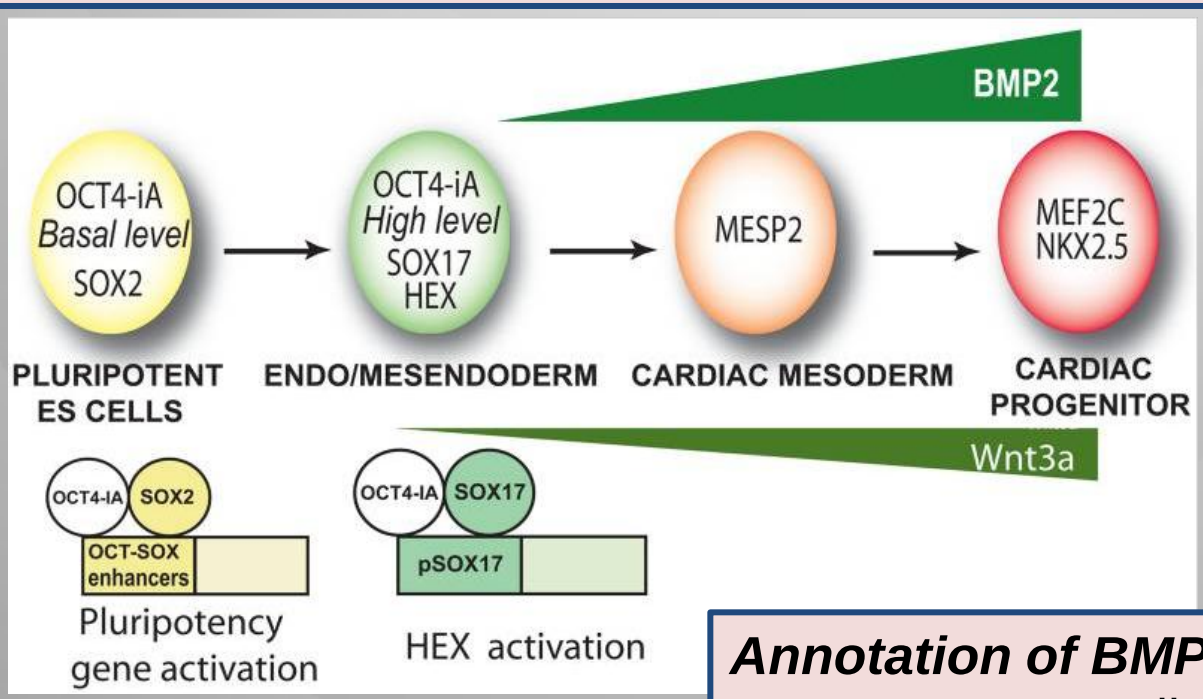


***Annotation of Mkrn2*:**
GO:0009798 axis specification IMP
GO:0050768 negative regulation of neurogenesis IMP

3. Example of a downstream process annotation in a signaling cascade

***Example*:** human growth factor BMP2 signals via Oct4 and Sox17 causing stem cells to become cardiac progenitors
PMID:19736317

"Oct4-iA was up-regulated threefold in response to BMP2." "Endodermal genes *Hex* and *Sox17* were also switched on by Oct4-iA"



***Annotation of BMP2*:**
GO:0035051 cardiac cell differentiation IMP

4. Example of downstream process annotations made when there is insufficient evidence for a direct role of the gene product

***Example*:** C. elegans rpb-2 RNA polymerase II subunit.
Only HTP studies available

***Annotation of rpb-2*:**
GO:0000003 reproduction IMP
GO:0002119 nematode larval development IMP
GO:0008340 determination of adult lifespan IMP
GO:0040010 positive regulation of growth rate IMP
GO:0009792 embryonic development ending in birth or egg hatching IMP

Initial survey

- *Do you annotate downstream processes? (Yes/No/Sometimes)*

  - All answered **Yes** or **Yes, in the absence of other information**

- *If yes/sometimes when do you annotate, when do you not annotate and why?*

  - Several groups directly annotate the evidence in the paper

  - All agreed that if that is the only evidence for a gene product, they would annotate to downstream processes

  - Most groups would annotate a gene product involved in a known core process, e.g. transcription, with that process alone

Initial survey (cont'd)

• *Do you revise your annotations sets when more is known about a gene product?*

• Most groups do not revise annotation sets, unless the annotations are found to be incorrect. Some groups, e.g. SGD, PomBase, do remove annotations later found to be to indirect terms

Dependent on;

- curator time
- availability of alternative way to display data, e.g. Phenotype database
- the need to capture all papers from community
- the desire to provide a 'complete story' of a gene product

Proposed Guidelines

Guideline 1

Quite often it is the case that the most relevant GO term will not exist.
It is desirable to request terms which describe the involvement of a process in another process, if that will give more specificity to the annotation.

For example, to describe a gene product's 'intent' to change the 'state' of the cell;

- Growth factor BMP2 is instrumental in cardiac cell differentiation

- Following stimulation with BMP2, large numbers of genes are up/down regulated

Requesting the new GO term 'BMP signaling involved in cardiac cell differentiation' may be preferable to annotating to the separate terms 'BMP signaling' and 'cardiac cell differentiation' as it will be clear how the gene product is involved in cardiac cell differentiation.

i.e. qualify how the gene product is involved in the downstream process in preference to annotating to the downstream process term

Guideline 2

For small scale experiments**, curators should annotate to the experimental evidence in the paper.**

However, curator judgement should be used, taking into account what the curator knows about:

a) the gene product; does it have a central role causing it to affect multiple processes, or does it have few specific targets?

b) the quality of the experimental assays performed in the paper; are they fully explained and the evidence supplied convincing?

(See separate guidelines for annotation of high-throughput experiments.)

Guideline 2 working examples

Example 1. Gene product involved in core process.

a) Yeast RNA polymerase II subunit RPB2

- has core function of RNA polymerase activity
- likely to affect large number of processes unrelated to its function
- most curators agree should annotate only to ‘transcription’

b) Yeast spliceosome

- in S. cerevisiae several genes are components of spliceosome
- when mutated the strains have defects in translation
- later evidence confirmed the genes’ involvement in mRNA splicing, NOT translation
- since most splicing in yeast is to ribosome genes the effect on translation was seen
- so annotations to ‘translation’ were removed from the spliceosome components

Guideline 2 working examples (cont'd)

Example 2. Gene product involved in core and specific process(es).

S. pombe gene Sre1

• direct transcriptional regulator of genes which have a role in heme and lipid biosynthesis (PMID:16537923)

• the curator judged this to be important information for this gene product

• annotations were made to:
  - specific RNA polymerase II transcription factor activity
  - regulation of transcription
  - positive regulation of heme biosynthesis
  - positive regulation of lipid biosynthesis

We would recommend that new terms are requested for;
- Regulation of transcription involved in heme biosynthesis
- Regulation of transcription involved in lipid biosynthesis

Guideline 3

If a gene product has limited experimental literature, such as a newly characterised protein, it is acceptable to annotate to more general 'downstream' process terms that may represent a phenotype

As more functional information is published about a gene product, these annotations to potential downstream processes may be removed if they are deemed by the annotating group as indirect, or they may be kept depending on each MOD's strategy

Always remove annotations that are incorrect or are from substandard evidence (NAS/TAS/IC) when replaced with better evidence to the same or more-granular term

Guideline 3 working example

***Example*:** C. elegans rpb-2 RNA polymerase II subunit.
Only high-throughput studies available

***Annotation of rpb-2*:**
GO:0000003 reproduction IMP
GO:0002119 nematode larval development IMP
GO:0008340 determination of adult lifespan IMP
GO:0040010 positive regulation of growth rate IMP
GO:0009792 embryonic development ending in birth or egg hatching IMP

We would consider these annotations OK

Guideline 4

Annotate ligand-receptor signaling pathways as shown in following diagrams

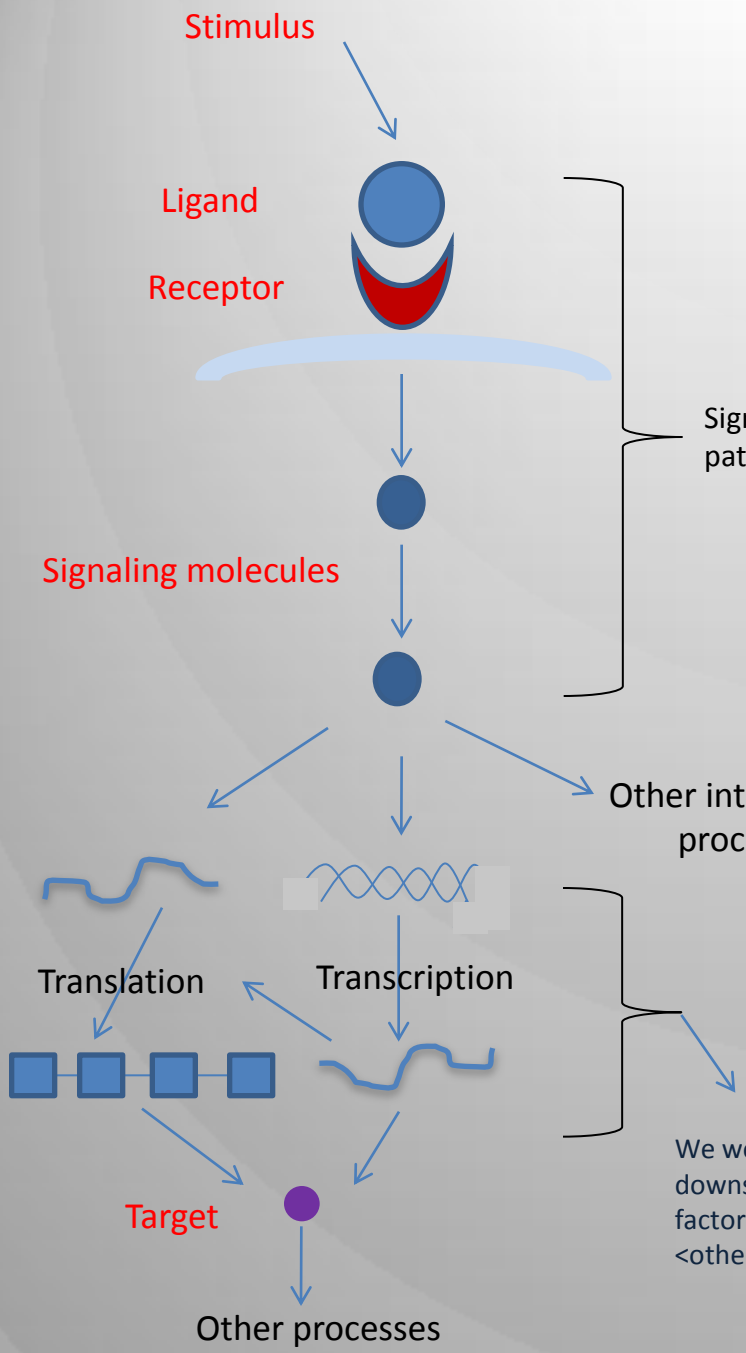
General considerations;

- For a signaling pathway that goes between cells in a multicellular organism (intercellular), the ligand is part of the pathway, e.g. the insulin signaling pathway. In this case, a factor which limits/increases the availability of a ligand to a receptor should be annotated as regulating the ligand/receptor pathway.

- For a signal transduction pathway within a single cell (intracellular), the ligand regulates the pathway. An example here is the regulation of the T cell receptor signaling pathway by CD4 (or CD8).

N.B. Clarification of the start/end of a signaling pathway by the signaling group will allow us to refine these guidelines

General ligand-receptor pathway



Suggested biological process annotations

Stimulus (intercellular pathway):
Regulation of signaling pathway

Ligand (intercellular pathway):
Signaling pathway
Regulation of <other> process(es)

Ligand (intracellular pathway):
Regulation of signaling pathway
Regulation of <other> process(es)

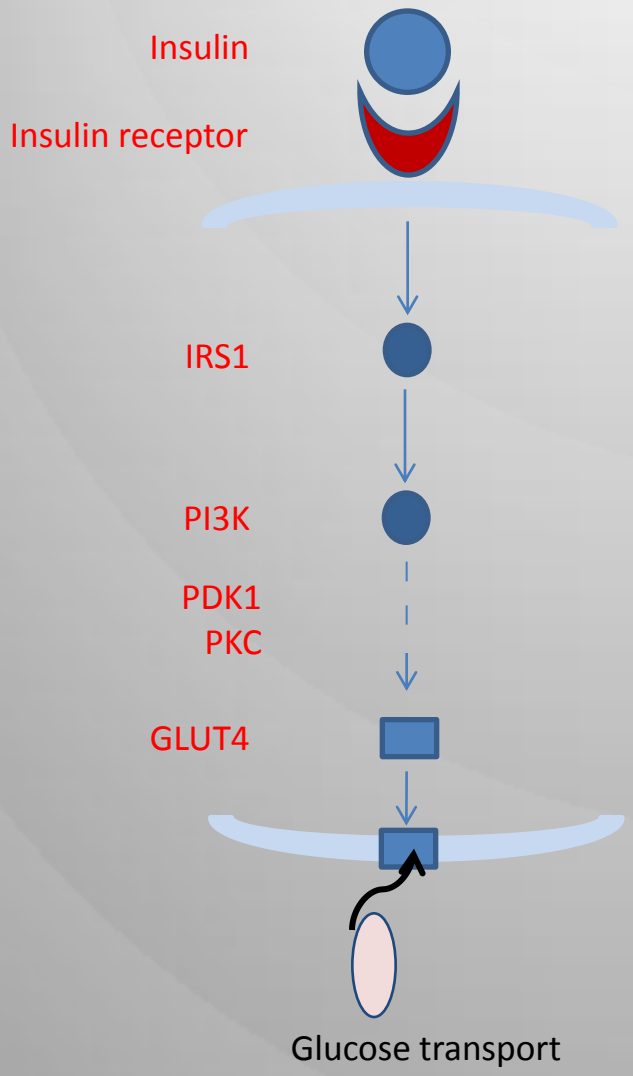
Receptor:
Signaling pathway
Regulation of <other> process(es)
Cellular response to stimulus/ligand

Signaling molecules:
Signaling pathway
Regulation of gene-specific transcription
Regulation of translation
(Regulation of) transcription in response to <stimulus/ligand>
(Regulation of) transcription involved in <other> process(es)
(Regulation of) <other intracellular> process(es)
Cellular response to stimulus/ligand

Target:
Cellular response to stimulus
<Other> process(es)
Regulation of <other> process(es)

We would not consider annotating the core transcription machinery to the downstream (other) processes that the target is involved in unless the transcription factor is gene-specific, in which case we would annotate to "regulation of transcription involved in <other> process(es)".

Regulation of glucose transport



Suggested biological process annotations

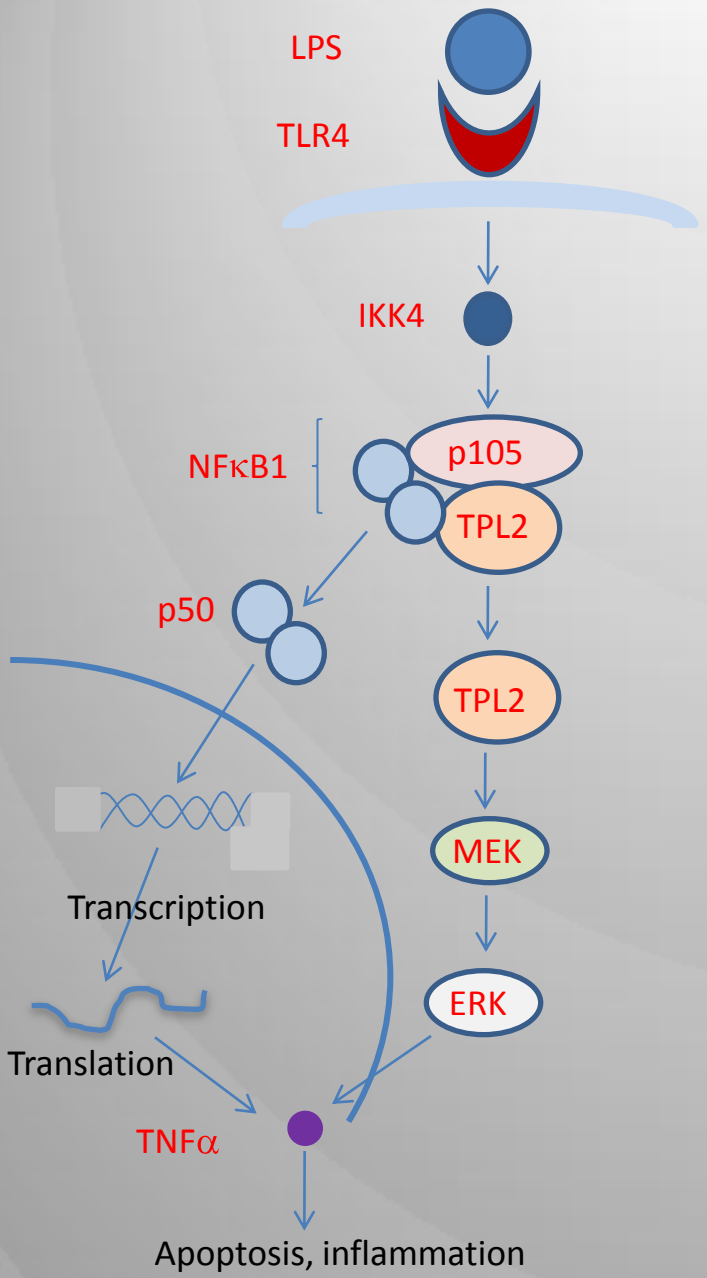
Insulin (ligand):
Insulin receptor signaling pathway
Regulation of glucose transport/homeostasis

Insulin receptor (receptor):
Insulin receptor signaling pathway
Regulation of glucose transport/homeostasis
Cellular response to insulin

IRS1, PI3K, PDK1, PKC (signaling molecules):
Insulin receptor signaling pathway
Regulation of glucose transport/homeostasis
Cellular response to insulin
Protein localization at cell surface (NTR: involved in response to insulin)

GLUT4 (target):
Cellular response to insulin
Glucose transport/homeostasis

Simplified NFκB pathway



Suggested biological process annotations

LPS (Ligand):
No annotation

TLR4 (Receptor):
NFκB cascade
Regulation of apoptosis
Regulation of inflammatory response

IKK4, TPL2, MEK, ERK (signaling molecules):
NFκB cascade
Cellular response to lipopolysaccharide
Regulation of apoptosis
Regulation of inflammatory response

NFκB (p50/p105):
NTR: Regulation of transcription involved in apoptosis
NTR: Regulation of transcription involved in inflammation

TNFα (Target):
Induction of apoptosis
Inflammatory response
Cellular response to lipopolysaccharide

Diagram adapted from Beinke and Ley 2004. PMID: 15214841

Documentation for Revision of Annotation Sets

(not a guideline)

Relevant to gene products with little annotatable evidence,
e.g. the C. elegans rpb-2 RNA polymerase II subunit.

When further information about a gene product is obtained, there are two options for the annotation set:

1.- Remove annotations to indirect/downstream processes (or update them to 'regulation' terms). This 'deleted' information is usually stored in the annotating group's phenotype database.

2.- Do not remove annotations to indirect/downstream processes because;

a) downstream annotations are supported by good evidence / want to keep as history of annotation / want to give a complete overview of knowledge about the gene product.
b) do not have resources to revise annotation sets / do not have alternative place to store data

NOTE: MODs that keep annotations will be a source of these downstream terms to MODS which do not keep these terms, via ISS from orthologs.

Suggested Quality Control Check

Check for co-annotation of a less-granular term with a more-granular term in the same path

Any action from this check is optional for each group as it may still be appropriate to keep both annotations, for example, it is acceptable to retain the less-granular annotation if;

- It has a ‘better’ evidence code
- The curator feels it adds weight to the more-granular annotation
- Both annotations add value, e.g. ‘histone methylation’ and ‘protein amino acid methylation’

Results from annotation camp survey Q1

Abstract: CYLD regulates angiogenesis by mediating the spreading and migration of vascular endothelial cells. Silencing of CYLD dramatically decreases microtubule dynamics in endothelial cells and inhibits endothelial cell migration by blocking the polarization process

What GO terms should one annotate?

A) The underlying mechanism

GO:0070507 regulation of microtubule cytoskeleton organization

B) Downstream effects

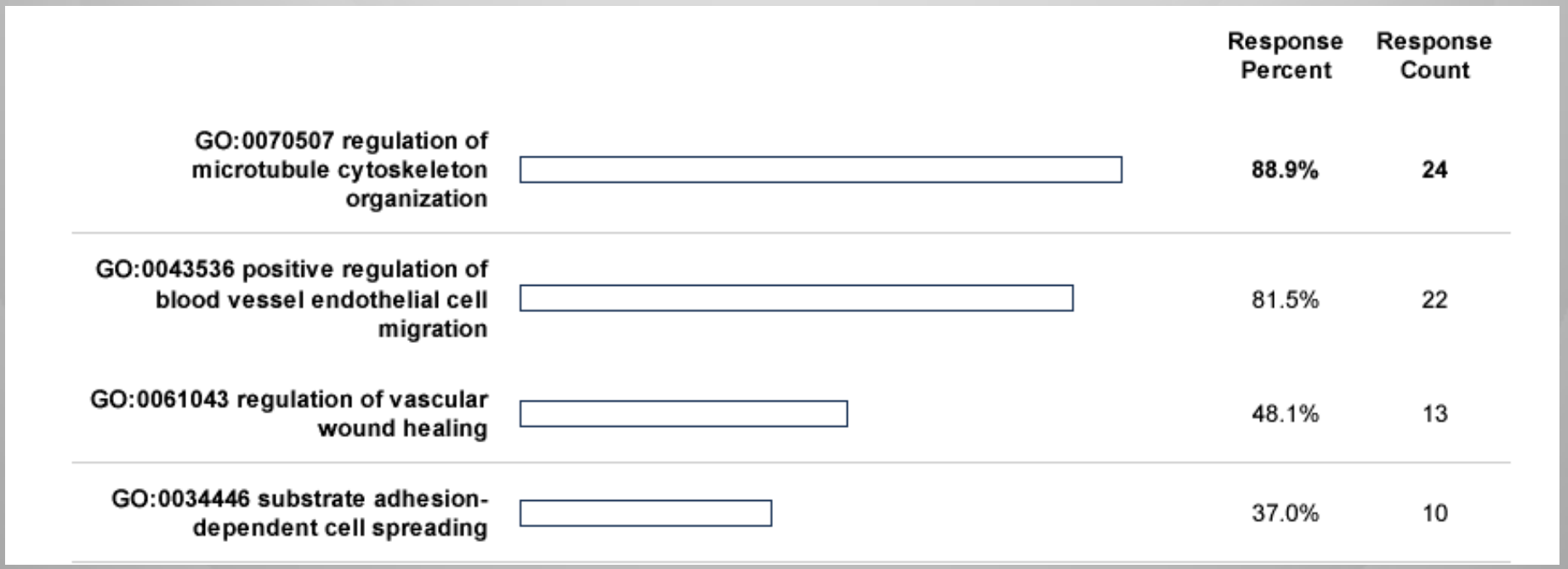
GO:0043536 positive regulation of blood vessel endothelial cell migration

GO:0061043 regulation of vascular wound healing

GO:0034446 substrate adhesion-dependent cell spreading

Results from annotation camp survey Q1

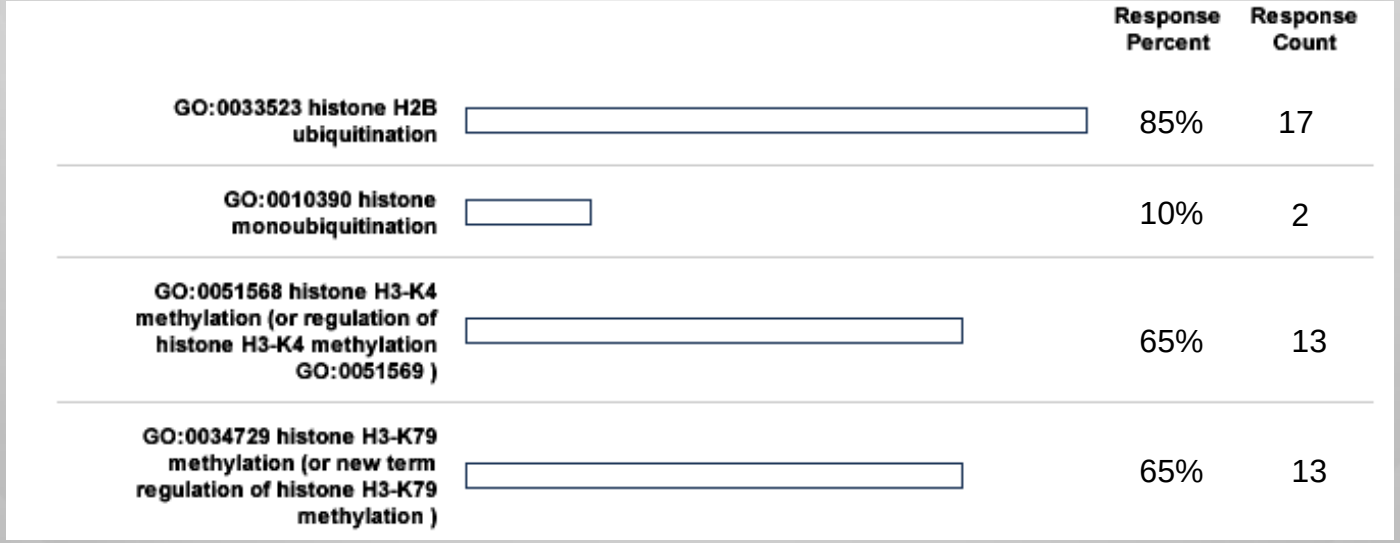
CYLD is a **ubiquitin-specific protease** that regulates a number of pathways, including **NF-kappa-B activation**, **Wnt signaling** and **microtubule dynamics**.



	Response Percent	Response Count
GO:0070507 regulation of microtubule cytoskeleton organization	88.9%	24
GO:0043536 positive regulation of blood vessel endothelial cell migration	81.5%	22
GO:0061043 regulation of vascular wound healing	48.1%	13
GO:0034446 substrate adhesion-dependent cell spreading	37.0%	10

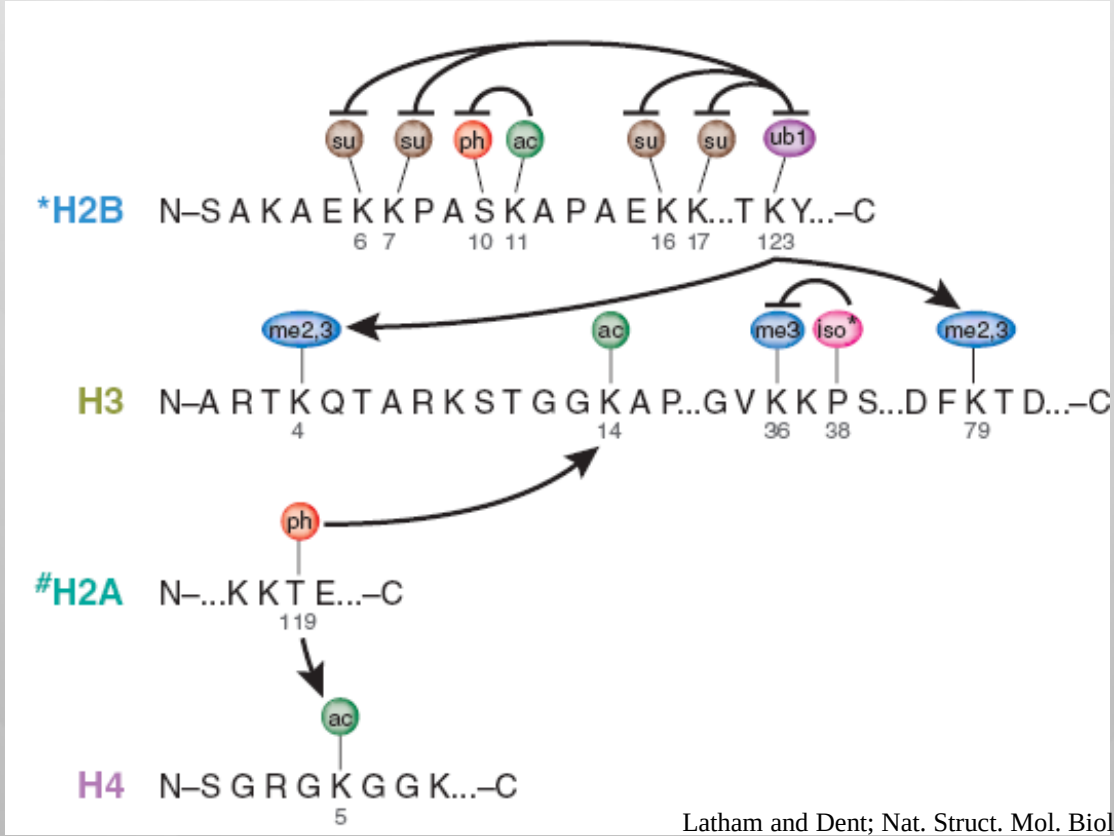
Results from annotation camp survey Q2

Human Bre1: ubiquitination vs. methylation



- The people who chose the methylation terms would have used ‘regulation of methylation’
- Most people felt there was insufficient evidence for the H2B monoubiquitination annotation
- One person suggested NTR: regulation of histone H2B ubiquitination involved in histone methylation

Histone code



Latham and Dent; Nat. Struct. Mol. Biol. 14:1017-1024(2007).

Histone tails are post-translationally modified by a complex of modifications that cross-react together. For instance, histone H2B monoubiquitination of 'Lys-120' of (H2BK120ub) is a prerequisite for histone H3 'Lys-4' and 'Lys-79' methylation (H3K4me and H3K79me, respectively).

Results from annotation camp survey Q2

When annotating enzymes that mediate modifications of histone tails, we would therefore recommend to only annotate the primary function of the enzyme:

GO:0004842; F:ubiquitin-protein ligase activity
GO:0033523; P:histone H2B ubiquitination
GO:0010390; P:histone monoubiquitination

And not all the downstream effects on histone modifications, otherwise most chromatin-modification enzymes will be over-annotated with the same terms that are not directly related to our proteins.

GO:0051568; P:histone H3-K4 methylation
GO:0034729; P:histone H3-K79 methylation
□

Summary of proposed guidelines

1. Request new terms as needed to qualify how the gene product is involved in the downstream process in preference to annotating to the downstream process term

2. For small scale experiments, curators should annotate to the experimental evidence in the paper

3. If a gene product has limited experimental literature, such as a newly characterised protein, it is acceptable to annotate to more general 'downstream' process terms that may represent a phenotype

4. We would like to provide annotators with a diagram summarising the downstream annotations which can be made to components of signaling pathways

Outstanding issues

1. What is the process term for a specific transcription factor – ACTION: transcription ontology revision

2. Define start and end of signaling processes – ACTION: signaling working group

3. Some MODs keep legacy annotations, some prefer to remove them – is this a problem?

4. Form a working group to look into phenotype/development/IMP issues