

LEGO: modular GO annotation

Paul Thomas

Motivation and Goal

- Ontology is very rich and quite mature
- However, annotations take a very simple form:
 - Gene_product relation GO_class
- Enable curators to use the GO to express rich biological statements from the literature
 - Maximize knowledge captured by curator

Building blocks

- Molecular function: break into blocks
- Relations
 - Function-function
 - Function-process (-component)
 - Process-process
- Specificity

Molecular function

- Biochemical function, e.g. binding, catalysis
- Specificity (“target”) of biochemical function
 - Binding partner: fructose binding
 - Substrate: 1-aminocyclopropane-1-carboxylate deaminase activity
 - Product: 6-carboxy-5,6,7,8-tetrahydropterin synthase activity
- Molecular role, e.g. protein kinase activator activity
 - These are often annotated to noncovalent activators rather than covalent activators (enzymes), e.g. MAPKK

- The epithelial Na(+) channel (ENaC) regulates Na(+) absorption in epithelial tissues including the lung, colon and sweat gland, and in the distal nephrons of the kidney. When Na(+)-channel function is disrupted, salt and water homeostasis is affected.... Previously we showed that a proline-rich region of the alpha subunit of the Na(+) channel bound to a protein of 116 kDa from human lung cells. Here we report the identification of this protein as human Nedd4, a ubiquitin-protein ligase that binds to the Na(+)-channel subunits via its WW domains. Further, we show that WW domains 2, 3 and 4 of human Nedd4 bind to the alpha, beta and gamma Na(+)-channel subunits but not to a mutated beta subunit. In addition, when co-expressed in *Xenopus* oocytes, human Nedd4 down-regulates Na(+)-channel activity.

Relations between processes

BP

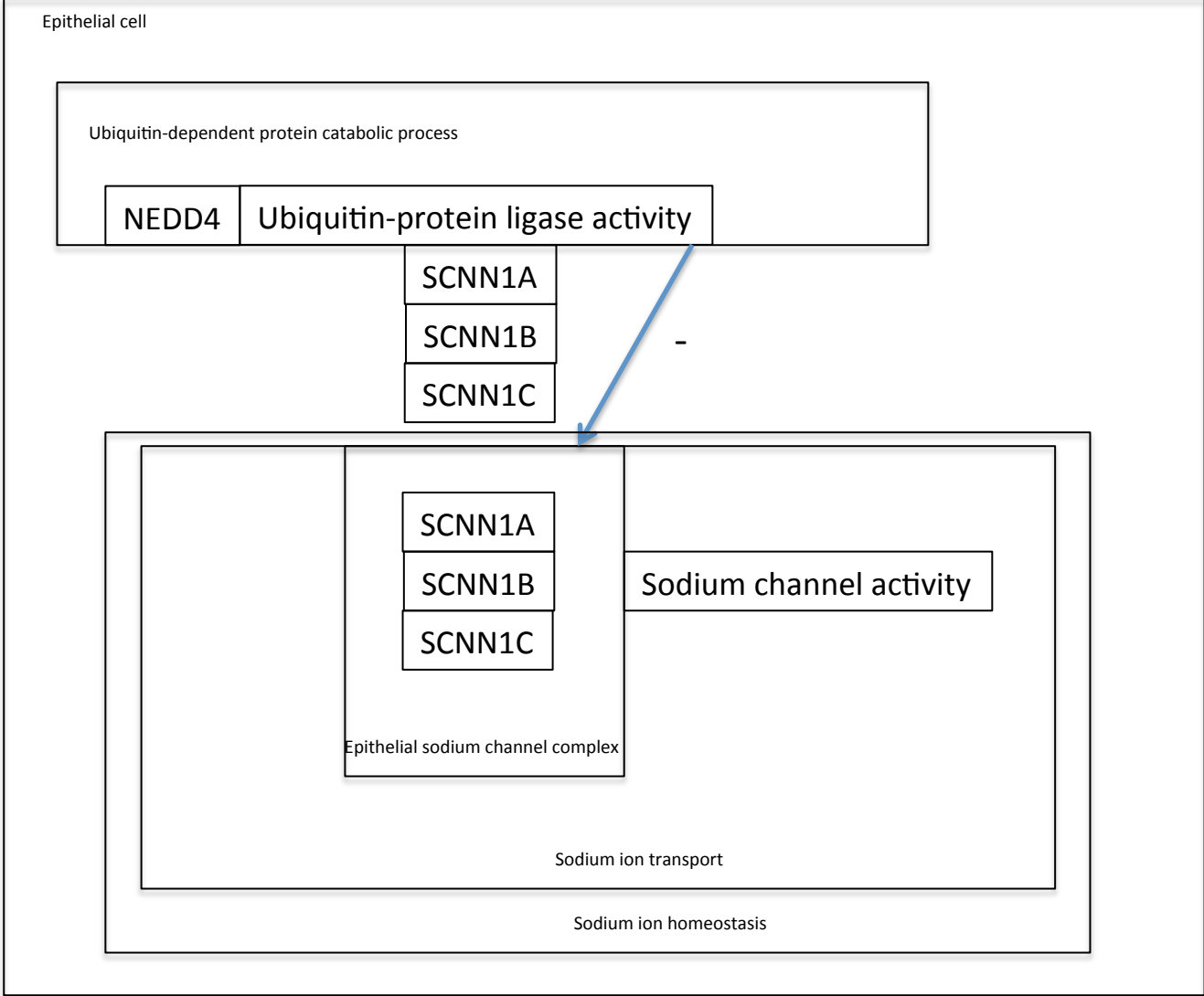
NEDD4

Term	Reference	ECO
cellular response to UV	PMID:17996703	IMP
glucocorticoid receptor signaling pathway	PMID:8649367	IDA
negative regulation of sodium ion transport	PMID:10642508	IDA
negative regulation of transcription from RNA polymerase II promoter	PMID:17996703	IMP
neuron projection development	PMID:9990509	IEP
pathogenesis	PMID:15126635	IMP
positive regulation of nucleocytoplasmic transport	PMID:17218261	IDA
positive regulation of phosphoinositide 3-kinase cascade	PMID:17218260	IMP
positive regulation of protein catabolic process	PMID:14973438	IDA
progesterone receptor signaling pathway	PMID:8649367	IDA
protein targeting to lysosome	PMID:17116753	IDA
protein ubiquitination during ubiquitin-dependent proteolysis	PMID:17218260	IMP
receptor catabolic process	PMID:18544533	IDA
receptor internalization	PMID:18544533	IDA
response to calcium ion	PMID:9405440	TAS
transmission of virus	PMID:15126635	IMP

MF

NEDD4

Term	Reference	ECO	With
RNA polymerase binding	PMID:17996703	IPI	UniProtKB:P24928
beta-2 adrenergic receptor binding	PMID:18544533	IDA	
proline-rich region binding	PMID:11342538	IPI	UniProtKB:Q15038
protein domain specific binding	PMID:12907594	IPI	UniProtKB:Q969W9
ubiquitin binding	PMID:9990509	IDA	



Function-Process
and
Process-Process
relation (negatively
regulates)

- β -arrestin2 binds at least two E3 ubiquitin ligases, Mdm2 and Nedd4, serving different purposes in β 2AR regulation: Mdm2, which mediates β -arrestin ubiquitination (12) and regulates the initial step of receptor endocytosis, and Nedd4, which mediates receptor ubiquitination that targets receptors to lysosomal compartments... We also demonstrate that β -arrestin2 functions as an E3 ubiquitin ligase adaptor to recruit Nedd4 to the activated β 2AR.

Part of process, or preceding process?

BP

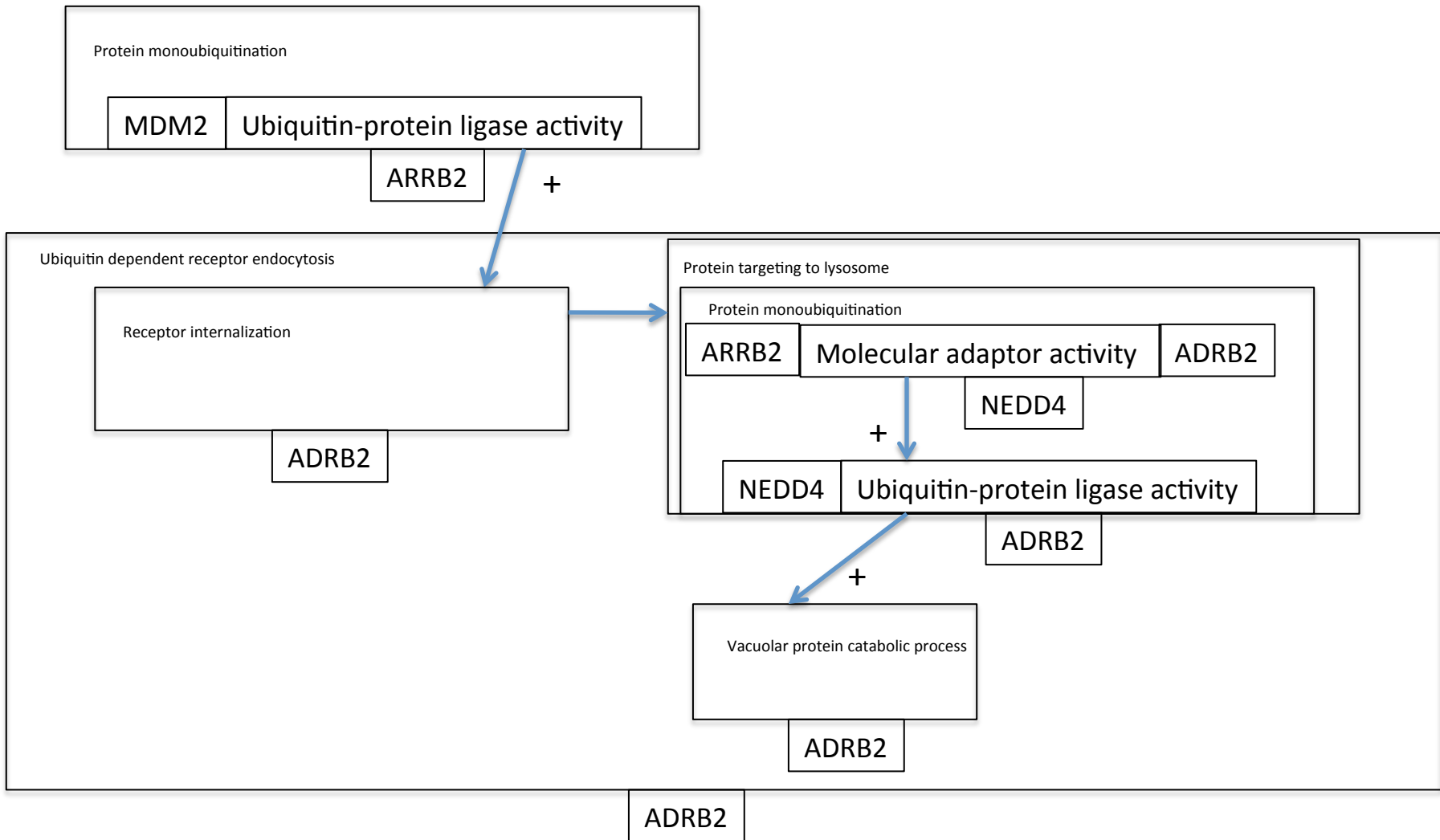
NEDD4

Term	Reference	ECO
cellular response to UV	PMID:17996703	IMP
glucocorticoid receptor signaling pathway	PMID:8649367	IDA
negative regulation of sodium ion transport	PMID:10642508	IDA
negative regulation of transcription from RNA polymerase II promoter	PMID:17996703	IMP
neuron projection development	PMID:9990509	IEP
pathogenesis	PMID:15126635	IMP
positive regulation of nucleocytoplasmic transport	PMID:17218261	IDA
positive regulation of phosphoinositide 3-kinase cascade	PMID:17218260	IMP
positive regulation of protein catabolic process	PMID:14973438	IDA
progesterone receptor signaling pathway	PMID:8649367	IDA
protein targeting to lysosome	PMID:17116753	IDA
protein ubiquitination during ubiquitin-dependent proteolysis	PMID:17218260	IMP
receptor catabolic process	PMID:18544533	IDA
receptor internalization	PMID:18544533	IDA
response to calcium ion	PMID:9405440	TAS
transmission of virus	PMID:15126635	IMP

MF

NEDD4

Term	Reference	ECO	With
RNA polymerase binding	PMID:17996703	IPI	UniProtKB:P24928
beta-2 adrenergic receptor binding	PMID:18544533	IDA	
proline-rich region binding	PMID:11342538	IPI	UniProtKB:Q15038
protein domain specific binding	PMID:12907594	IPI	UniProtKB:Q969W9
ubiquitin binding	PMID:9990509	IDA	



- UV-induced RNA polymerase II (RNAPII) ubiquitylation and degradation are important DNA damage responses, conserved from yeast to man....The defects in RNAPII ubiquitylation observed in CS cells are caused by an indirect mechanism: these cells shut down transcription in response to DNA damage, effectively depleting the substrate for ubiquitylation, namely elongating RNAPII. Instead, we identified Nedd4 as an E3 that associates with and ubiquitylates RNAPII in response to UV-induced DNA damage in human cells.

Disjoint annotations

BP

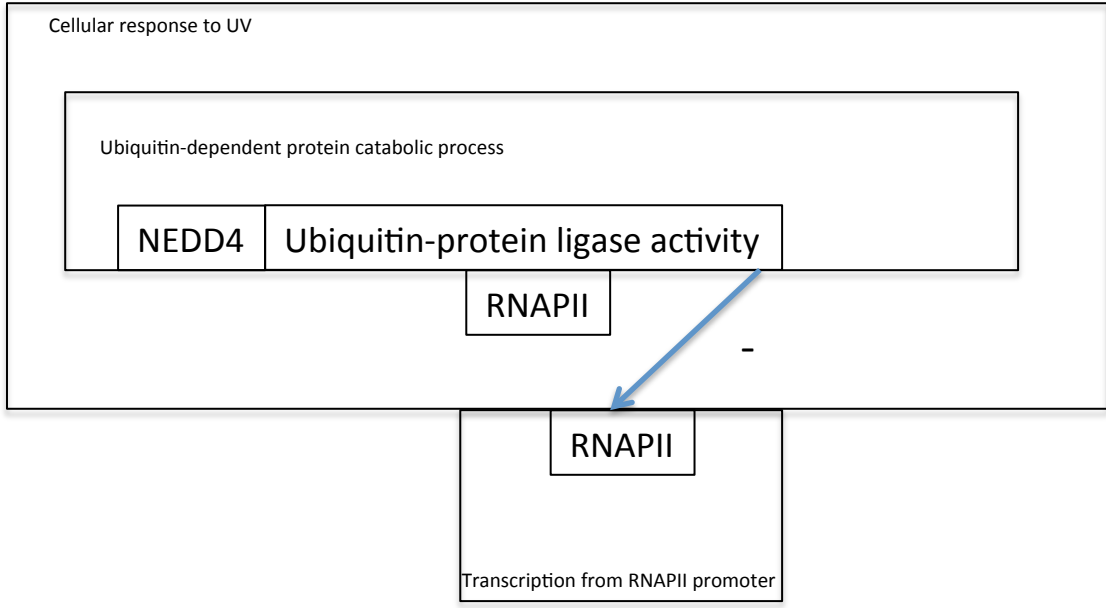
NEDD4

Term	Reference	ECO
cellular response to UV	PMID:17996703	IMP
glucocorticoid receptor signaling pathway	PMID:8649367	IDA
negative regulation of sodium ion transport	PMID:10642508	IDA
negative regulation of transcription from RNA polymerase II promoter	PMID:17996703	IMP
neuron projection development	PMID:9990509	IEP
pathogenesis	PMID:15126635	IMP
positive regulation of nucleocytoplasmic transport	PMID:17218261	IDA
positive regulation of phosphoinositide 3-kinase cascade	PMID:17218260	IMP
positive regulation of protein catabolic process	PMID:14973438	IDA
progesterone receptor signaling pathway	PMID:8649367	IDA
protein targeting to lysosome	PMID:17116753	IDA
protein ubiquitination during ubiquitin-dependent proteolysis	PMID:17218260	IMP
receptor catabolic process	PMID:18544533	IDA
receptor internalization	PMID:18544533	IDA
response to calcium ion	PMID:9405440	TAS
transmission of virus	PMID:15126635	IMP

MF

NEDD4

Term	Reference	ECO	With
RNA polymerase binding	PMID:17996703	IPI	UniProtKB:P24928
beta-2 adrenergic receptor binding	PMID:18544533	IDA	
proline-rich region binding	PMID:11342538	IPI	UniProtKB:Q15038
protein domain specific binding	PMID:12907594	IPI	UniProtKB:Q969W9
ubiquitin binding	PMID:9990509	IDA	



- LAPT_M5 contains three PY motifs (L/PPxY), which bind Nedd4-WW domains, and a ubiquitin-interacting motif (UIM) motif. The Nedd4-LAPT_M5 complex recruits ubiquitinated GGA3, which binds the LAPT_M5-UIM; this interaction does not require the GGA3-GAT domain. ... ubiquitination-impaired LAPT_M5 can still traffic to the lysosome, suggesting that Nedd4 binding to LAPT_M5, not LAPT_M5 ubiquitination, is required for targeting. ... These results demonstrate a novel mechanism by which the ubiquitin-ligase Nedd4, via interactions with GGA3 and cargo (LAPT_M5), regulates cargo trafficking to the lysosome without requiring cargo ubiquitination.

Part of process, or preceding process?

BP

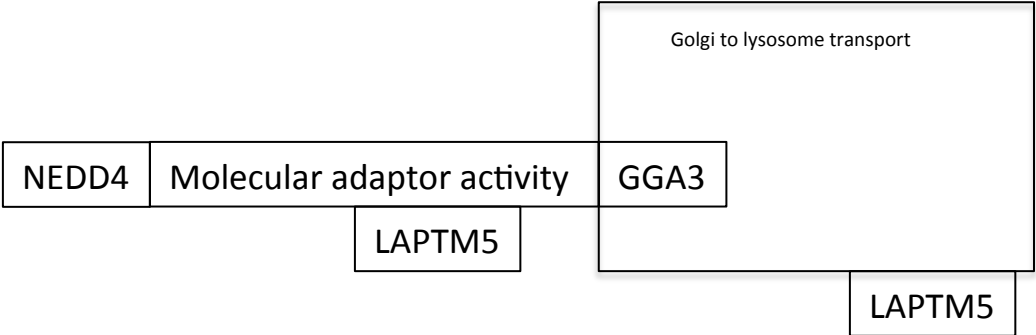
NEDD4

Term	Reference	ECO
cellular response to UV	PMID:17996703	IMP
glucocorticoid receptor signaling pathway	PMID:8649367	IDA
negative regulation of sodium ion transport	PMID:10642508	IDA
negative regulation of transcription from RNA polymerase II promoter	PMID:17996703	IMP
neuron projection development	PMID:9990509	IEP
pathogenesis	PMID:15126635	IMP
positive regulation of nucleocytoplasmic transport	PMID:17218261	IDA
positive regulation of phosphoinositide 3-kinase cascade	PMID:17218260	IMP
positive regulation of protein catabolic process	PMID:14973438	IDA
progesterone receptor signaling pathway	PMID:8649367	IDA
protein targeting to lysosome	PMID:17116753	IDA
protein ubiquitination during ubiquitin-dependent proteolysis	PMID:17218260	IMP
receptor catabolic process	PMID:18544533	IDA
receptor internalization	PMID:18544533	IDA
response to calcium ion	PMID:9405440	TAS
transmission of virus	PMID:15126635	IMP

MF

NEDD4

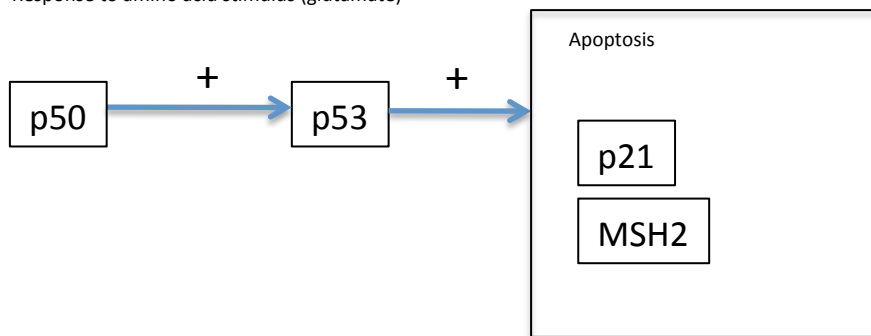
Term	Reference	ECO	With
RNA polymerase binding	PMID:17996703	IPI	UniProtKB:P24928
beta-2 adrenergic receptor binding	PMID:18544533	IDA	
proline-rich region binding	PMID:11342538	IPI	UniProtKB:Q15038
protein domain specific binding	PMID:12907594	IPI	UniProtKB:Q969W9
ubiquitin binding	PMID:9990509	IDA	



- Fifteen minute exposure of primary cultures of cerebellar granule cells to micromolar concentrations of glutamate results in apoptotic cell death. Among the intracellular events triggered by glutamate, we identified two transcriptional factors, i.e. the p50 member of the NF-kappaB family and the tumor suppressor phosphoprotein p53, that are apparently linked by a sequential transcriptional program....We also found that two other proteins, the cyclin dependent kinase inhibitor p21 and DNA mismatches repair MSH2, whose encoding genes are well known target of p53, were upregulated by glutamate. On these bases, we propose NF-kappaB, p53, p21 and MSH2 as relevant contributors of the glutamate-induced pro-apoptotic pathway.

Cerebellar granule cell

Response to amino acid stimulus (glutamate)



What do our end-users want from GO annotation?

- Succinct description of gene function at-a-glance
- Support for analysis of genomic data sets in terms of common processes
- Most effective when annotation is
 - Comprehensive
 - As many genes as possible
 - As much information as possible about each gene
 - Accurate

Possibilities for the future?

- “Systems” approach: expand what we consider to be important for GO annotation
 - Complete gene sets across many genomes
 - Evolutionary history and related genes
 - Active site / critical sites
 - Gene structure refinement
 - Biological system: capture more of the curators’ knowledge!
 - Representation of *context* for a particular class of molecular function, and how it links to biological process
 - Extend the ontology

Ubiquitin-dependent protein catabolic process

Protein polyubiquitination

Protein X Ubiquitin activating enzyme activity

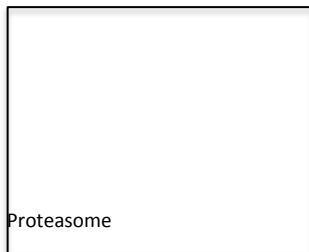
Protein Y Ubiquitin conjugating enzyme activity

Protein X

Protein Z Ubiquitin-protein ligase activity Protein A

Protein Y

Protein A



ATP-dependent protease activity

Protein A

Proteolysis

Protein A

Function-Process
and
Process-Process
relation (negatively
regulates)