

# Gene Groups in FlyBase

Helen L. Attrill, Steven J. Marygold, Susan Tweedie and the FlyBase consortium

## Abstract

FlyBase is undertaking a review of gene level data and creating a Gene Group resource for *Drosophila melanogaster*. Gene families/groups (e.g. paralogs, multi-protein complex components) are defined based on recent literature, and the associated data in FlyBase reviewed. This strategy has allowed us to make significant improvements to the consistency of Gene Ontology (GO) annotation and nomenclature, which was not being addressed by our usual paper-by-paper curation approach. It is envisioned that the collection of gene groups will be used as a basis to form a Gene Group portal on the FlyBase website.

## 1. Gene Groups

### How are Gene Groups selected for review?

- Well-defined groups, with a clear common connection (e.g. gene families and complexes).
- Source of genes must be derived from peer-reviewed publications or expert-led databases.
- Groups which may have a more fluid membership or be hard to delimit, such as pathways and processes, are excluded at present.

Gene Group	No. of genes	Gene Group	No. of genes
Accessory gland proteins	121	Lamins	2
Acetylcholine (nicotinic) receptor subunits	10	Metallothioneins	5
Actin-related protein 2/3 complex	8	Methuselah genes	16
Actin-related proteins	9	Mon1-Ccz1 complex	2
Actins	6	Myosin heavy chains	13
ADAM Metalloproteases	5	Myosin light chains & calmodulin	6
Adaptins	13	N-ethylmaleimide-sensitive fusion (NSF) proteins	2
Adaptor & coat protein complexes	91	Neuropeptides/peptide & protein hormones	48
Anaphase-promoting complex	12	Neurof genes	10
Annexins	3	Nucleoporins	29
ATAC Complex	13	Odorant binding proteins	52
BAR Domain Proteins	21	Odorant receptors	61
Biogenesis of lysosome-related organelles complexes	14	p24 transporters	9
Cadherins	31	Pickpocket genes	31
Cullins	5	Potassium channels	26
Cyclic nucleotide-gated ion channels	5	Proteasome subunits (26S)	50
Deubiquitinating enzymes	45	Ribosomal proteins	168
Dorsocross	3	RZZ complex	3
Drososmins	7	Sepsins	5
E(Esp)C genes	8	Serpins	29
Fliotins	2	SM proteins	5
Glutamate (Ionotropic) receptor subunits	80	SNAP	2
Glutathione transferases (cytoplasmic)	36	SNARE	26
G-Protein Coupled Receptors	92	Soring Neuin	12
GTPase Activating Proteins	66	Synaptotagmin	1
GTPases (small)	79	Tethering Complexes	42
Guanine Nucleotide Exchange Proteins	54	Trafs	3
Gustatory receptors	60	Transcription Factor I1H	10
Halloween genes	6	Transient receptor potential channels	13
Heterochromatin Protein 1 genes	5	Tubulins	12
Heterotrimeric G protein: Galpha subunits	8	Ubiquitin	5
Iminease	8	Vacuolar H <sup>+</sup> ATPase subunits (Vha)	7
Integrins	7	Wnt genes	7
KASH domain proteins	2	yellow genes	14
Kinesin-like proteins	26		

List of Gene Groups. To date 70 groups (shown above) have been reviewed, representing 134 individually reviewed sub-groups and 1588 genes.

## 2a. Reviewing Gene Group Data: Nomenclature

### Reviewing gene nomenclature based on gene groups

Gene names and symbols in FlyBase have traditionally been assigned in isolation, according to the first peer-reviewed paper to characterize the gene. By reviewing the nomenclature of genes in the context of the group to which they belong, we have been able to improve the consistency and clarity of nomenclature of related genes, whilst respecting precedence and/or the preferred usage in the literature. Experts in the relevant field are consulted as necessary to ensure any proposed changes are appropriate.

### Example: Review of Adaptor protein complex nomenclature

During the review of the adaptor protein (AP) complex AP-1 and AP-2 subunits, symbols/names were standardized with the 'AP-' / Adaptor Protein complex' prefix and given a unique suffix based on the well-established complex and subunit nomenclature. The AP-3 subunits have retained their phenotype-based nomenclature owing to popular use and precedence.

Complex	Subunit	Old Symbol	Old Name	New Symbol	New Name
AP-1	γ	AP-1γ	AP-1γ	AP-1γ	Adaptor Protein complex 1, gamma subunit
AP-1	μ1	AP-47	AP-47	AP-1μ	Adaptor Protein complex 1, mu subunit
AP-1	σ1	AP-1σ	AP-1σ	AP-1σ	Adaptor Protein complex 1, sigma subunit
AP-1 & 2	β1/2	Bap	β Adaptin	AP-1 β	Adaptor Protein complex 1/2, beta subunit
AP-2	α	α-Adaptin	α-Adaptin	AP-2α	Adaptor Protein complex 2, alpha subunit
AP-2	μ2	AP-50	AP-50	AP-2μ	Adaptor Protein complex 2, mu subunit
AP-2	σ2	AP-2σ	AP-2σ	AP-2σ	Adaptor Protein complex 2, sigma subunit
AP-3	Δ	g	garnet	g	garnet
AP-3	β3	rb	ruby	rb	ruby
AP-3	α3	cm	cornicle	cm	cornicle
AP-3	σ3	or	orange	or	orange

Symbols and names are greyed-out to indicate that no change was made.

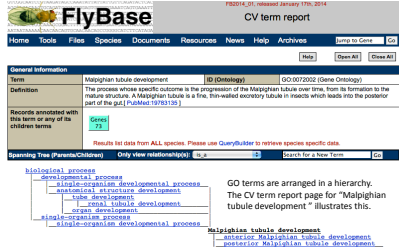
## 2b. Reviewing Gene Group Data: Gene Ontology

What is Gene Ontology? Gene Ontology (GO) uses standardized common terms (controlled vocabulary, CV) to describe the nature and attributes of a gene product.

GO terms are divided into three areas:

1. **molecular function** (such as enzymatic activity, binding)
2. **biological process** (pathways or processes influenced)
3. **cellular component** (sub-cellular localisation or complex)

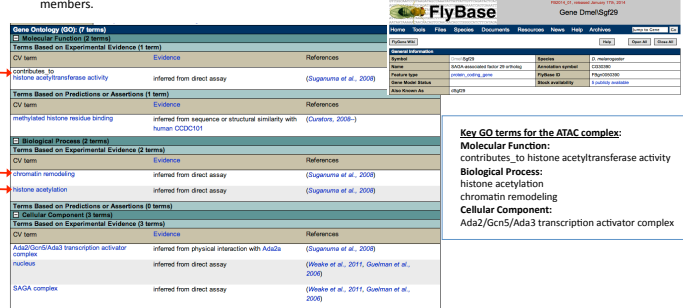
GO terms can be supported by experimental evidence or inferred from sequence evidence.



GO terms are arranged in a hierarchy. The CV term report page for "Magpsh1an tubulin development" illustrates this.

### Example: SAGA-associated factor 29 ortholog (Sgf29) gene

Sgf29 is a component of the Ada2a-containing (ATAC) histone acetyltransferase complex. In FlyBase, GO terms are displayed in **Gene Report** pages. In the example shown here, the GO terms associated with the Sgf29 gene are shown. The red arrows indicate the GO terms that were associated with Sgf29 as a result of reviewing ATAC complex members.



**Key GO terms for the ATAC complex:**  
**Molecular Function:** contributes to histone acetyltransferase activity  
**Biological Process:** histone acetylation, chromatin remodeling  
**Cellular Component:** Ada2/Gcn5/Ada3 transcription activator complex

### Reviewing GO data based on Gene Groups

Groups of genes, by definition, will share certain biological features. This allows us to make the assertion that most/all members of a group should share some defining or "key" GO terms.

#### Aims

1. Add GO terms that reflect a gene's central biological role by defining a set of "key" terms.
2. Review GO terms that have become stale due to up-dated gene models or orthology.
3. Add more descriptive GO terms in-line with changes in terms available and orthologal data.
4. Add GO terms based on experimental evidence where possible to help researchers find key data.

### GO term review of Gene Groups to date:

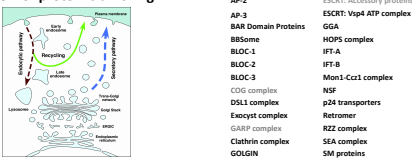
134 groups  
 1588 genes reviewed  
 3248 GO terms added  
 354 GO terms removed

## 3. Future perspectives

- On-going review of gene data using the gene grouping strategy
- The gene groups covered as part of this review will form the basis of **Gene Group** report pages in FlyBase (see right).

- Reviewing groups as part of larger pathway and process groups e.g. protein trafficking (panel below), protein degradation, oxidative phosphorylation

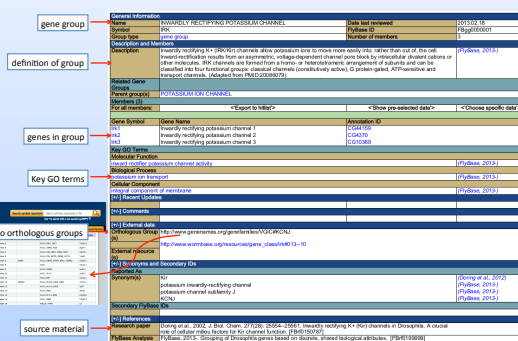
### Review of protein trafficking



Thirty eight gene groups (representing 142 genes) were reviewed as part of examining complexes and gene families involved in protein trafficking. Approximately 790 GO terms were added. Greyed-out groups could not be reviewed as their members were not sufficiently well-defined in literature.

- AP-1
- AP-2
- BAR Domain Proteins
- BBSome
- BLOC-1
- BLOC-2
- BLOC-3
- CDG complex
- DSL1 complex
- Exocyst complex
- GAU3 complex
- RZZ complex
- Clathrin complex
- GOLGIN
- SM proteins
- COP1 complex
- COP2 complex
- COR1E complex
- STONED
- ESCR1-1
- ESCR1-2
- ESCR1-3
- ESCR1-4
- ESCR1-5
- ESCR1-6
- ESCR1-7
- ESCR1-8

### Mock-up of a Gene Group page



Mock-up of a Gene Group page showing definition of group, genes in group, key GO terms, and source material.

Nominate a gene group for review

