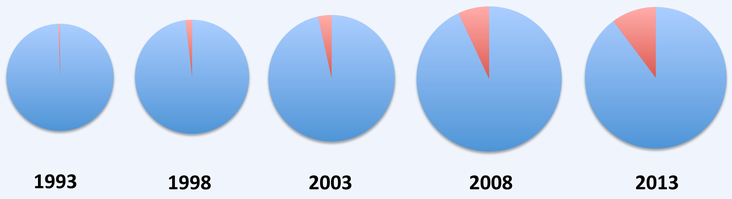
Representing *Drosophila* models of Human Disease in FlyBase

The FlyBase Consortium (Poster presenter: Gillian Millburn)

Introduction

The number and proportion of *Drosophila* articles that mention disease is increasing year-on-year. Many of these papers describe *Drosophila* models of human diseases or investigate functional conservation between human-fly orthologs. To help researchers find genes and alleles that model human diseases in *Drosophila*, FlyBase has started to collect and display this information from primary research articles.



Drosophila papers containing "disease" in the abstract or title

Drosophila models of Human Disease in FlyBase

What disease models do we capture ?

- Data on **genetic models** which result in a **phenotype** that **recapitulates** some aspect of **human disease**.
- The genetic models in this case are **alleles**.

How do we identify relevant papers ?

- Newly published papers: authors tag the paper with "Human Disease" using our **Fast Track Your Paper** tool.
- Archive of previously published papers:
  - We searched for papers with "disease" in the title or abstract.
  - We examined papers known to use human transgenes.

What alleles are included ?

- *Drosophila* genes: both at-locus mutations and transgenic constructs.
- Transgenic constructs expressing human genes.
- Alleles which cause a phenotype that models a human disease.
- Alleles which modify a disease-model phenotype.

How do we link alleles to disease ?

- We associate the allele with standardized human disease terms from the Disease Ontology (http://disease-ontology.org).

DISEASE ONTOLOGY

```
disease
|__disease of anatomical entity
   |__nervous system disease
      |__central nervous system disease
         |__neurodegenerative disease
            |__Parkinson's disease
```

The Disease Ontology (DO) provides standardized terms to describe human diseases and orders them in a hierarchy.

- We record whether the allele is:
  - a **model of** the human disease.

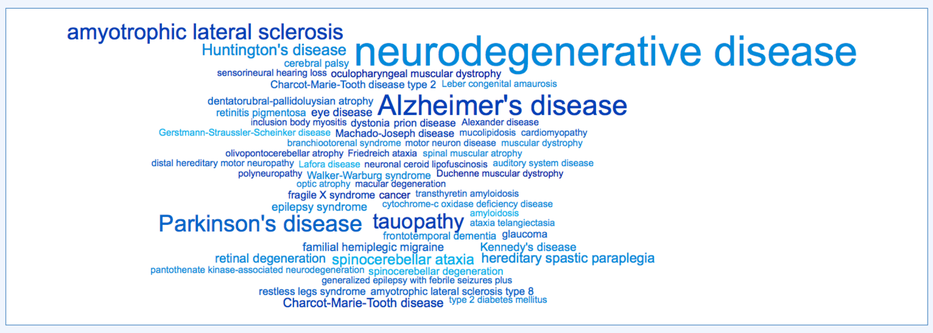
  OR
  - a modifier, which **ameliorates** or **exacerbates** a disease **modeled by** another allele.
- When >1 allele models the disease, we record that they act "**in combination with**" each other.
- We use a **DOES NOT** prefix to record unexpected negative results *eg.* when a fly ortholog of a disease-causing human gene does not recapitulate the disease phenotype.
- Using these simple expressions there are multiple ways in which the relationship between an allele and a disease can be described.

The different types of disease annotation

	allele X			
model	model of	Disease	inferred from mutant phenotype	
	model of	Disease	in combination with	allele Y
modifier	ameliorates	Disease	modeled by	allele Y
	exacerbates	Disease	modeled by	allele Y
negative statements	DOES NOT model	Disease	inferred from mutant phenotype	
	DOES NOT model	Disease	in combination with	allele Y
	DOES NOT ameliorate	Disease	modeled by	allele Y
	DOES NOT exacerbate	Disease	modeled by	allele Y

The data so far

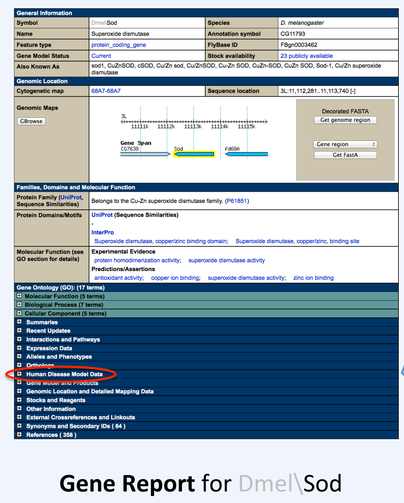
- FlyBase started capturing disease annotations at the beginning of 2014.
- The FB_2014_05 (September 9th 2014) release contains:
  - ~**2,500** disease annotations from ~**500** references.
  - **>1,500** alleles from **>750** genes are annotated as either a human-disease model or a modifier.
- Models of **126** different human diseases have been annotated.
- Approximately **two-thirds** of the models are of neurological diseases.



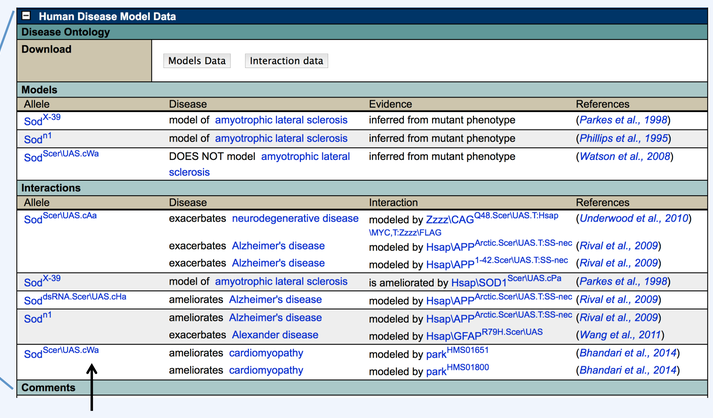
How to find human disease models in FlyBase

1. Is a particular gene/allele associated with a disease model ?

There is now a **Human Disease Model Data** section in each **Gene** and **Allele Report**.



Gene Report for Dmel\Sod

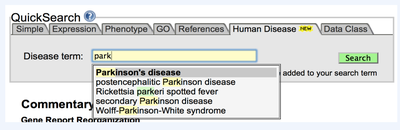


Clicking on a specific allele will take you to the associated **Allele Report** page.

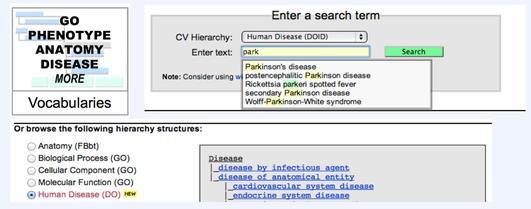
2. What genes/alleles are associated with a particular disease model ?

There are two ways to directly search for a disease: **QuickSearch** and the dedicated Controlled Vocabulary search tool **Vocabularies.** Both have autocomplete functions, but Vocabularies will also allow you to explore the Disease Ontology hierarchy.

QuickSearch "Human Disease" tab

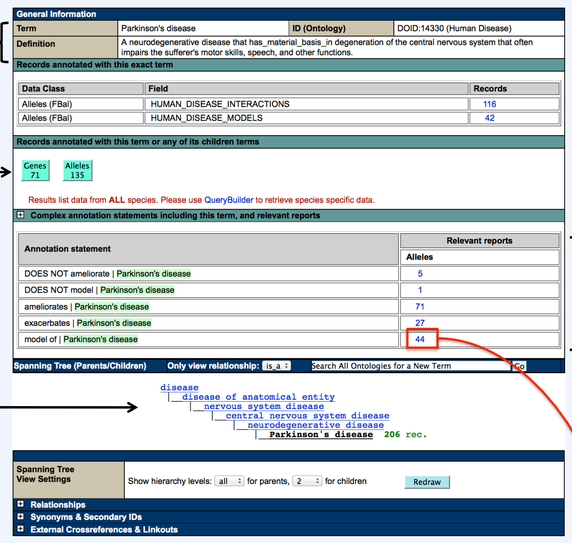


Vocabularies



The Disease Term Report page

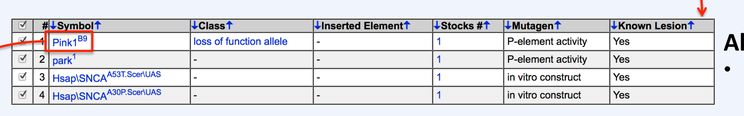
Disease details



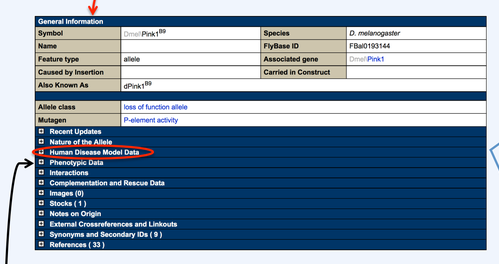
Click for a list of all genes or alleles associated with this disease term.

Disease Ontology tree view
- Links take you to the associated **Disease Term Report**.
- Explore the hierarchy to find related disease terms.
- Numbers indicate the number of disease annotations for this term and its child terms.

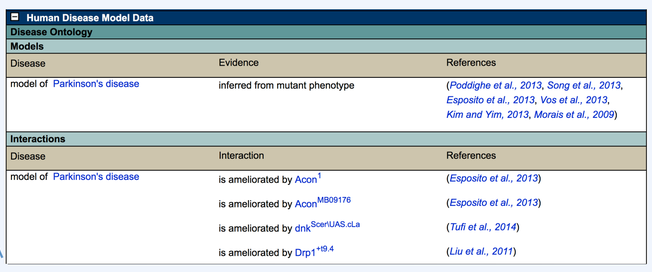
Annotation Statements
- Clicking on the links will give a hit list of matching alleles.
- Statements are grouped according to whether the alleles are models or modifiers of the disease.
- Negative statements are grouped separately.



Allele Hit list
- A section of the allele hit list for model of | Parkinson's disease.

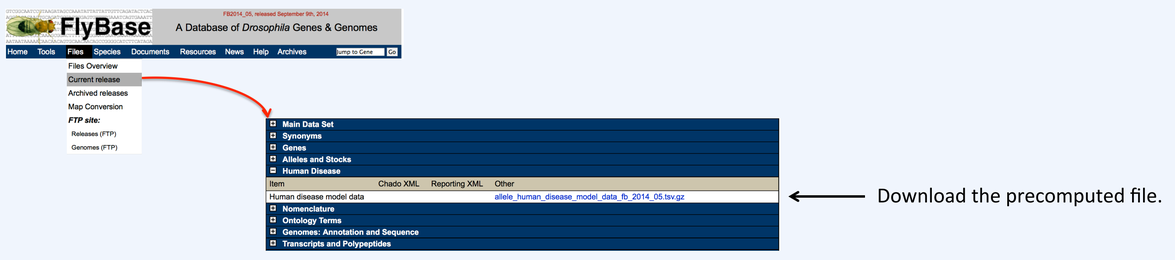


Allele Report for Pink[B9]



Phenotypic data
- Look here for a detailed description of the phenotype.

3. How do I get the complete set of data ?



Download the precomputed file.

Future Improvements .....

- We are still looking at past papers to identify potential disease models and text-mining will be employed to help trawl the archives.
- **Please help us** make this resource as comprehensive and useful as possible by providing feedback, suggestions and alerting us to disease models we have missed. Use the **Contact FlyBase** link (found at the bottom of each page in FlyBase).

Acknowledgements

- Thank you to all authors who flagged "Human Disease" using the **Fast Track Your Paper** tool!

The FlyBase Consortium comprises: *Nick Brown, Boris Adryan, Helen Attrill, Marta Costa, Steven Marygold, Peter McQuilton, Gillian Millburn, Laura Ponting, Ray Stefancsik, Susan Tweedie (FlyBase-Cambridge), William Gelbart, Kris Broll, Lynn Crosby, Gil dos Santos, David Emmert, L. Sian Gramates, Kathleen Falls, Beverley Matthews, Susan Russo, Andy Schroeder, Susan St. Pierre, Pinglei Zhou, Mark Zytkovicz (FlyBase-Harvard), Thomas Kaufman, Josh Goodman, Gary Grumbling, Victor Strelets, Jim Thurmond (FlyBase-Indiana), Richard Cripps, Maggie Werner-Washburne (FlyBase-NewMexico).*

FlyBase is supported by a grant from the National Human Genome Research Institute at the U.S. National Institutes of Health (U41HG000739).
Support is also provided by the British Medical Research Council (G1000968) , the Indiana Genomics Initiative, and the National Science Foundation through XSEDE resources provided by Indiana University.