Systematising knowledge of Drosophila pathway members to fuel biological discovery

Giulia Antonazzo[1], Helen Attrill[1], Joshua L. Goodman[2], Nicholas H. Brown[1] and the FlyBase Consortium

[1] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3DY, UK.
[2] Department of Biology, Indiana University, 1001 East 3rd Street, Bloomington, Indiana 47405-7005, USA.

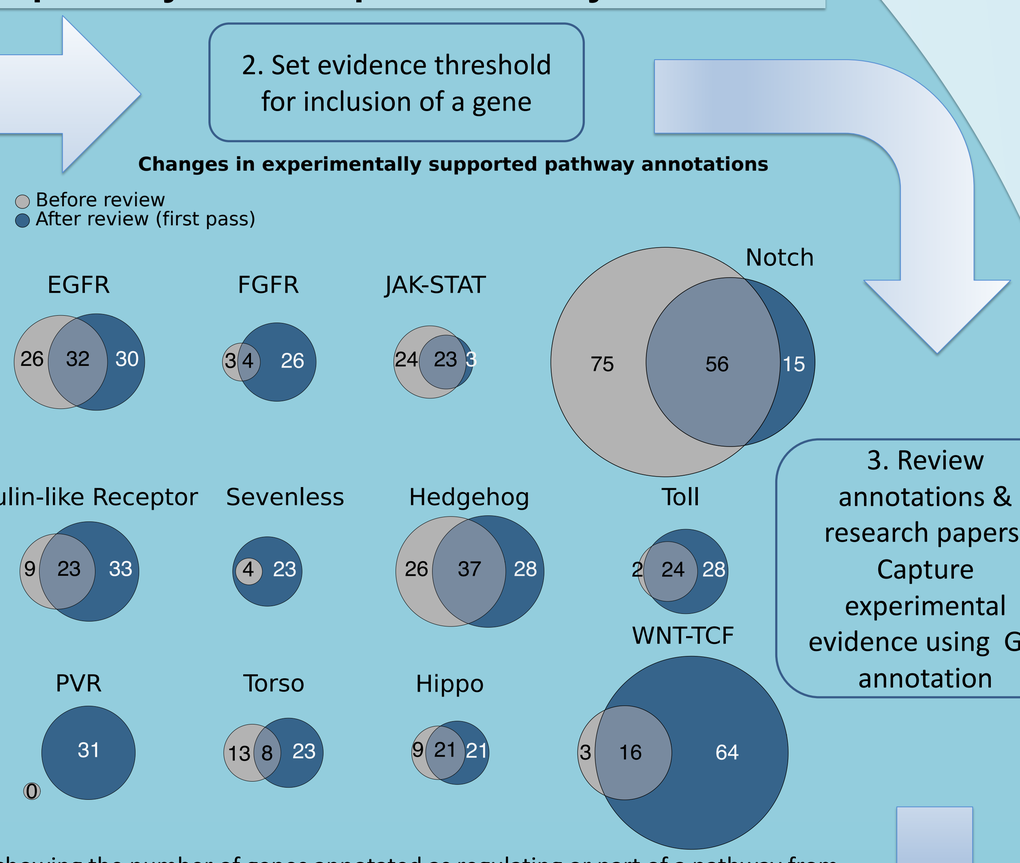
MRC Medical Research Council

The new pathway curation process in FlyBase

1. Define characteristics of a core member vs regulator

2. Set evidence threshold for inclusion of a gene

3. Review annotations & research papers. Capture experimental evidence using GO annotation



Venn diagrams showing the number of genes annotated as regulating or part of a pathway from an experimental observation, showing the overlap between the sets before and after the first pass review. In the review, we found that many genes annotated as being a member of a pathway were actually far downstream or upstream of a pathway and these were removed.

4. Use GO annotation to build **pathway reports**

More pathways, references and regulators will be added. Pathway pages will be kept up-to-date with paper curation at FlyBase.

GO annotation of pathway components are used to populate pathway reports.

General Information			
Name	Notch Signaling Pathway Core Components	Species	
Symbol	NTCH-C	FlyBase ID	
Date last reviewed	2019-01-23	Number of members	
Description			
Description	The Notch receptor signaling pathway is activated by the binding of the transmembrane receptor Notch (N) to transmembrane ligands, Dl or Ser, presented on adjacent cells. This results … N, releasing the intracellular domain (NICD). NICD translocates into the nucleus, interacting with Su(H) and mam to form a transcription complex, which up-…tch-responsive genes. (Adapted from FBrf0225731 and FBrf0192604).		
	…re required for signaling from the sending cell and response in the receiving cell.		
Biological Process Gene Ontology (GO) term(s)	Notch signaling pathway		
Related Gene Groups			
Parent group(s)	Notch Signaling Pathway		
Protein Complex group(s)	CSL-NOTCH-MASTERMIND TRANSCRIPTION FACTOR COMPLEX GAMMA SECRETASE COMPLEX		
Other related group(s)	NOTCH LIGANDS		
Members (12)			
For all members:	View Orthologs	Export to HitList	Export to Batch Download

Links to analysis tools

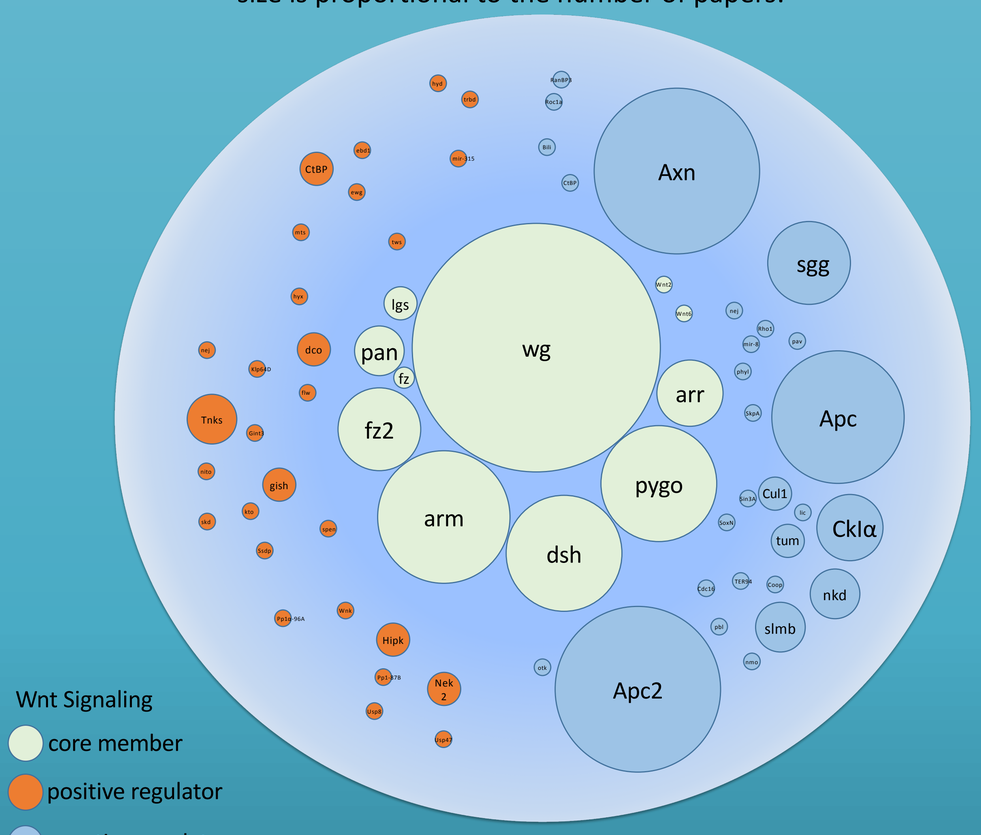
Gene Symbol	Gene Name	Gene Group Membership	GO Molecular Function (Experimental)	# Refs
aph-1	anterior pharynx defective 1	GAMMA SECRETASE COMPLEX	endopeptidase activity	2
Dl	Delta	NOTCH LIGANDS	Notch binding receptor ligand activity	9
kuz	kuzbanian	ADAM METALLOPROTEASES	metalloendopeptidase activity Notch binding	5
mam	mastermind	CSL-NOTCH-MASTERMIND TRANSCRIPTION FACTOR COMPLEX		5
N	Notch	CSL-NOTCH-MASTERMIND TRANSCRIPTION FACTOR COMPLEX	transmembrane signaling receptor activity chromatin binding	11
Nct	Nicastrin	GAMMA SECRETASE COMPLEX		7
pen-2	presenilin enhancer	GAMMA SECRETASE COMPLEX		2
Psn	Presenilin	GAMMA SE…SE COMPLEX	endop…activity protein …dimerization activity	

Functional pointers: membership of FlyBase gene groups and display of experimentally characterized molecular functions

Research references used to link gene to pathway from GO annotation

The weight of experimental evidence

By counting the number of annotated papers, we can show the relative weight of experimental evidence for each gene's involvement in a pathway. Here, node size is proportional to the number of papers:



Accurate pathway member lists fuel biological analyses

FlyBase data — Expression, Protein interactions, Genetic interactions, Gene ages

Machine learning

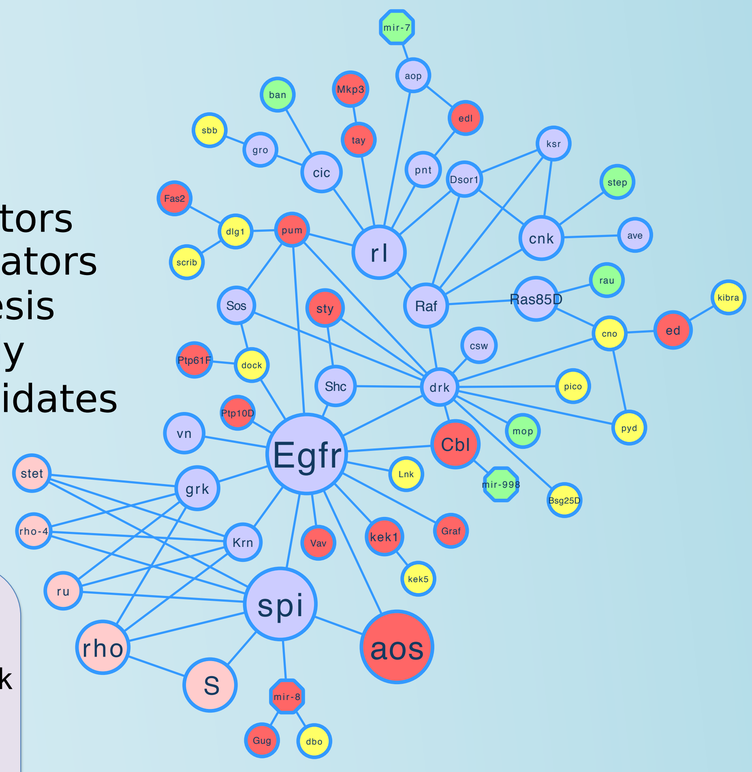
Predictions

Gene	*Pathway membership probability*
GENE1	0.91
GENE2	0.84
GENE3	0.83
GENE4	0.50
…	
GENE13933	0.02

By using the curated pathway member lists as training sets, we can train machine learning models aiming to predict novel pathway members. We use various forms of functional genomics data stored in FlyBase as features for training.

Pathway network models and biological properties

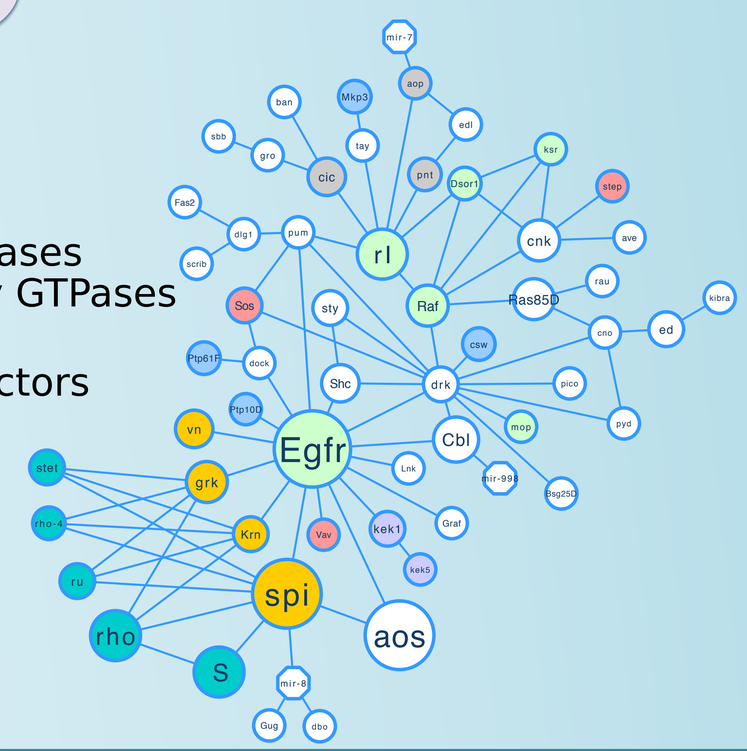
- Core Members
- Positive Regulators
- Negative Regulators
- Ligand Biogenesis
- Computationally Predicted Candidates



EGFR signaling network
Accurate pathway membership assignments allows us to build network models using interaction data. In this representation, the size of each gene node is based on the weight of curated experimental evidence.

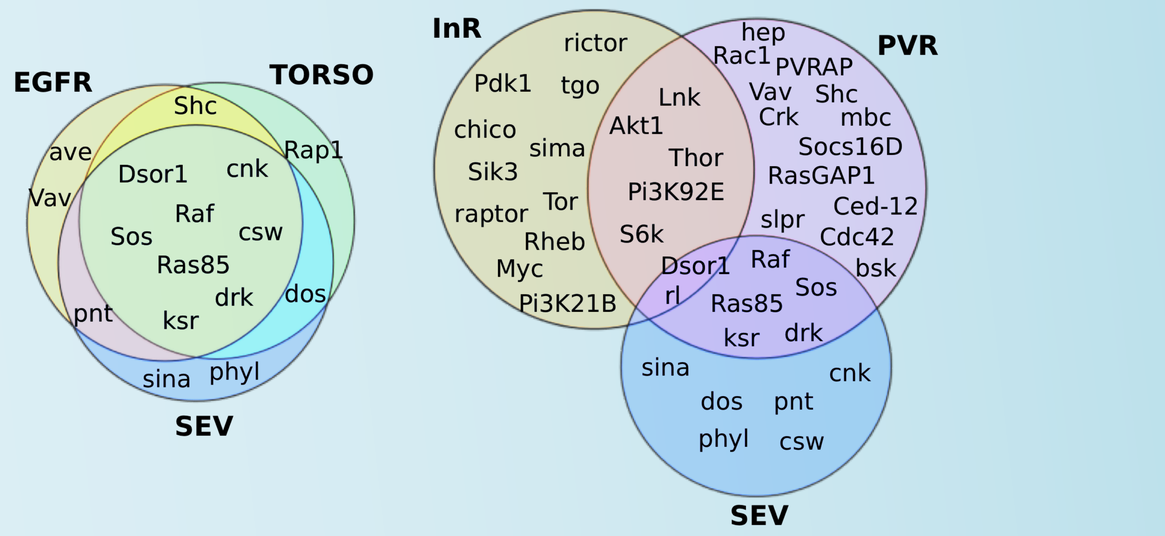
- Kinases
- Phosphatases
- EGFR-Agonists
- Rhomboid Proteases
- RAS superfamily GTPases
- Kekkons
- Transcription Factors

Other data can be overlaid on the network, here genes are coloured according to FlyBase gene group memberships.



Overlap between receptor tyrosine kinase pathways

A Venn diagram of EGFR, Torso and Sevenless receptor core intracellular pathway members reveals a high degree of overlap in components, corresponding to the Ras/Raf/MAP (Erk) kinase signaling module (left). The Insulin receptor and PVR pathways show a high degree of divergence from the 'classical' RTK pathway (right).



The FlyBase Consortium comprises: *Nick Brown, Giulia Antonazzo, Helen Attrill, Phani Garapati, Aoife Larkin, Steven Marygold, Gillian Millburn, Clare Pilgrim, Sinzi Pop, Vitor Trovisco, Jose-Maria Urbano (FlyBase-Cambridge), Norbert Perrimon, Susan Russo Gelbart, Julie Agapite, Kris Broll, Lynn Crosby, Gilberto dos Santos, Kathleen Falls, L. Sian Gramates, Victoria Jenkins, Ian Longden, Beverley Matthews, Carol Sutherland, Christopher Tabone, Pinglei Zhou, Mark Zytkovicz (FlyBase-Harvard), Thomas Kaufman, Brian Calvi, Josh Goodman, Victor Strelets, Jim Thurmond (FlyBase-Indiana), Richard Cripps, Maggie Werner-Washburne, Phillip Baker (FlyBase-New Mexico).*

This work is supported by the British Medical Research Council (#RG83674). FlyBase is supported by a grant from the National Human Genome Research Institute at the U.S. National Institutes of Health #U41 HG000739.
Support is also provided by the Indiana Genomics Initiative and FlyBase users all over the world.

Contact: ***ga362@cam.ac.uk***