Systematising knowledge of Drosophila pathway members to fuel biological discovery

Giulia Antonazzo, Helen Attrill, Joshua L. Goodman, Nicholas H. Brown 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.

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
4. Use GO annotation to build pathway reports

Accurate pathway member lists fuel biological analyses

FlyBase data
- Expression
- Protein interactions
- Genetic interactions
- Gene ages

Machine learning
- Gene
- Pathway membership

Predictions
<table>
<thead>
<tr>
<th>Gene</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENE1</td>
<td>0.91</td>
</tr>
<tr>
<td>GENE2</td>
<td>0.84</td>
</tr>
<tr>
<td>GENE3</td>
<td>0.83</td>
</tr>
<tr>
<td>GENE4</td>
<td>0.50</td>
</tr>
<tr>
<td>GENE13933</td>
<td>0.02</td>
</tr>
</tbody>
</table>

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

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.

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/MEK/ERK kinase signaling module (left). The insulin receptor and PVR pathways show a high degree of divergence from the 'classical' RTK pathway (right).

Contact: ga362@cam.ac.uk

The Flybase Consortium comprises: Nick Brown, Giulia Antonazzo, Helen Attrill, Phani Garapati, Aoife Larkin, Steven Marrygold, Gillian Milburn, Clare Pilgrim, Sinho Pop, Vitor Trovisco, Jose-Maria Urbano (FlyBase-Cambridge), Norbert Perrimon, Susan Russel Gerhart, Julie Agapite, Kris Brott, Lynn Crosby, Gilberto dos Santos, Kathleen Fols, T. Sion Gramata, Victoria Jenkins, Ian Longden, Beverly Matthews, Carol Sutherland, Christopher Tabone, Pingli Zhou, Mark Zitnik (Flybase-Harvard), Thomas Kaufman, Brian Callis, Josh Goodman, Victor Stieve, Jim Thompson (FlyBase-Indiana), Richard Crippes, Maggie Werner-Washburne, Phillip Baker (Flybase-New Mexico).

This work is supported by the British Medical Research Council (R083674). FlyBase is supported by a grant from the National Human Genome Research Institute at the U.S. National Institutes of Health (R01HG00739). Support is also provided by the Indiana Genomics Initiative and FlyBase users all over the world.