

# DROSOPHILA PROJECT

## The Drosophila protein Interaction Map

The protein-protein interaction map described here connects 1,727 Drosophila proteins, which represents 12% of Drosophila melanogaster proteome (Formstecher et al. 2005, *Genome Research* 15(3), 76-84 ). Over 2,300 interactions have been identified using a high-throughput proprietary yeast two-hybrid-based technology. This project is the result of a collaboration between Hybrigenics and 29 groups affiliated with the Curie Institute , and focused on understanding key mechanisms involved in cancer. This work was supported in part by GenHomme Network Grant 02490-6088 to Hybrigenics and Institut Curie.

## Collaborators and Sub-projects

The goal of this project was to create a dataset useful to the scientific community investigating oncogenesis. Bait proteins were chosen for the involvement of their human counterparts in cancer or in basic cellular metabolism. 150 baits were selected based on the possible correspondence between human proteins involved in cancer and their Drosophila orthologues. Highly complex cDNA libraries were built from Drosophila embryos and adult Drosophila heads. This program comprises 25 projects contributed by Hybrigenics and a total of 29 different academic groups affiliated with the Curie Institute.

## Baits and cDNA libraries construction

Baits were either directly sub-cloned in the pB27 plasmid derived from the original pBTM116 or were first PCR-amplified and then transferred (Vojtek and Hollenberg, *Methods Enzymol.* 1995, 255:331-342). Inserts were submitted to quality control by full insert sequencing and those who matched the predicted sequence were used further. In some circumstances, specific domains such as transmembrane domain, membrane-anchoring motifs, signal peptide or transactivation domain were deleted from the baits. Mutant alleles with defined biological or biochemical properties were also used as baits. Random-primed cDNA libraries from adult Drosophila heads (provided by Dr. Michael Rosbash, Brandeis University, Waltham, USA) and Drosophila embryos (pool of 0-12 and 12-24 hours embryos, provided by Dr. Peter Maroy, University of Szeged, Hungary) poly(A+) RNA were constructed into the pP6 plasmid derived from the original pGADGH (Bartel et al., 1993) Using the two-hybrid system to detect protein-protein interactions. In *Cellular Interactions in Development: A Practical Approach*, D. A. Hartley, Ed., Oxford University Press, Oxford; pp 153-179). The libraries were then transformed into the Y187 yeast strain and ten millions independent yeast colonies were collected, pooled and stored at &ndash;80°C as equivalent aliquot fractions of the same library. Yeast Two-Hybrid Screening Procedure and Prey Identification The screens were performed using a proprietary mating method (Rain et al. *Nature* 2001, 409:211-215). Each screen was performed to ensure a minimum of 50 millions interactions tested. Up to 384 positive clones per independent screen were picked and the corresponding prey fragments were amplified by PCR and sequenced at their 5' and 3' junctions on a PE3700 Sequencer. 5' and 3' sequences were then filtered by using PHRED (Ewing et al. *Genome Res* 1998, 8:175-185) and masking ALU repeats. Sequence contigs are built using CAP3 (Huang & Madan *Genome Res* 1999, 9:868-877) and compared to the recent release 3.1 of BDGP using BLASTN (Altschul et al. *Nucleic Acids Res* 1997, 25:3389-3402). In the few cases where no matching transcripts were identified, the contigs were compared to the latest release of the GenBank database using the same BLASTN procedure. Interactions Scoring and Results Display For each interaction, the minimal interacting domain or SID® (Selected Interacting Domain) was deduced from the individual fragments obtained in two-hybrid screens. The SID corresponds to the smallest identified protein domain necessary for a particular protein-protein interaction. A PBS® (Predicted Biological Score) was calculated for each interaction thanks to Hybrigenics proprietary algorithms allowing assessment of their reliability. These information as well as protein annotations were integrated in a proprietary application, PIMRider®, designed to explore the Protein Interaction Map.