

CONTENTS**About us**

- > Edito: Happy New Year!
- > Computation of a confidence score: the PBS
- > Highly connected proteins

News**Events****Interview****NEWS**

We have added a human H7 Embryonic Stem Cells library to our collection of highly complex cDNA libraries. The H7 cell line meets the criteria for the use of human embryonic stem cells by US federally funded researchers. As always, this library contains 10 million independent and random-primed cDNA fragments ranging from 600 to 800 bp and is ready to be screened with ULTimate Y2H. We have recently presented the first results using this library at the 1st International Symposium on Human Embryonic Stem Cell Research in Evry, France, from January 31st to February 2nd. Later in the year we will be presenting new data at the International Society for Stem Cell Research 6th Annual Meeting to be held in Philadelphia, USA, June 11-14. Meet us there!

**EXPERTISE #2**

hybrigenics services | Newsletter February 2008

HAPPY NEW YEAR!

The whole team of Hybrigenics Services joins me to wish you a prosperous New Year 2008! We hope your research projects move on a big step forward this year.

Luc Selig, Ph.D.

Director, Sales and Marketing

This second edition of our Expertise Newsletter focuses on one of the two major matters that come with any technology: the false positives. False positives are experimental data that are observed with the settings of the experiment but which are not true in the context of a cell.

The yeast two-hybrid (Y2H) system, invented in 1989 by Stanley Fields and co-workers, has since then suffered from a reputation of yielding numerous so-called "sticky proteins". These are proteins often found as preys in screens dealing with non-related bait proteins and are considered as the major source of false positives.

As the protein interactions expert, having performed more than 3,500 ULTimate Y2H™ screens to date, we would like to share with you our knowledge of the various types of false positives that co-exist in Y2H screens, detail different types of "highly connected preys", and share above all the way we rank protein interactions according to a technical confidence score we compute: the PBS® or Predicted Biological Score.

COMPUTATION OF A CONFIDENCE SCORE: THE PBS**EXHAUSTIVITY IS MANDATORY**

ULTimate Y2H allows the testing of 97 million interactions on average per screen by a proprietary cell-to-cell mating protocol and therefore achieves library screening to saturation (10x coverage of 10 million clones cDNA libraries). Screening to saturation is necessary to obtain reproducibility and exhaustivity, the main characteristics necessary for the bioinformatics computation of the Predicted Biological Score (PBS).

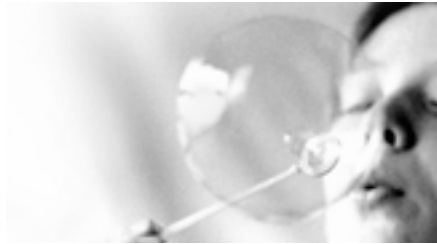
PBS COMPUTATION

The PBS is computed to assess interaction reliability. It represents an estimate of the probability for an interaction between a bait and the prey interacting domain to be non-specific: it is an e-value (e.g. 3.4E-14), primarily based on the comparison between the number of independent prey fragments

EVENTS

Hybrigenics celebrates its 10 years anniversary in 2008!

Hybrigenics is the first spin-off of the Institut Pasteur de Paris, and was founded in 1998 on its core patented ULTimate Y2H technology. This screening platform was automated and industrialized to take into account internal needs for large Protein Interaction Maps (PIMs). Sophisticated bioinformatics tools were further developed to analyze the tremendous amount of protein interaction data that was produced. These tools include the PIMRider® software, a tool dedicated to the visualization and analysis of large PIMs.



INTERVIEW

Barbara Ruggiero, Sales Manager, Hybrigenics Services

"I talk to my customers every day and I feel very happy when they are satisfied with my answers. This proximity allows me, each time, to meet new individual needs. My challenge is to find the answers to these needs to make their research project move forward."



Contact us

hybrigenics services
services@hybrigenics.com
+33 (0)1 58 10 38 29
www.hybrigenics-services.com

defining an interaction in a given screen and the chance of drawing them at random (background noise in statistical terms). In a second step of PBS computation, the PBS scores from all the screens that were performed with the same library are pooled automatically to take into account additional information such as reciprocity (bait 1 -> prey 2 and bait 2 -> prey 1) and cycles (e.g. bait 1 -> prey 2, bait 2 -> prey 3, bait 3 -> prey 1). The final value of the PBS varies between 0 and 1. Several thresholds have been predefined to rank the results in 4 categories from A (the highest confidence score) to D. PBS A, B and C are high confidence interactions identified through two or more overlapping fragments. We have published in Nature and Genome Research a **very good correlation between the PBS score and the validation of the interaction *in vivo***. For more details on PBS computation, please refer to these two papers.

PBS D: SO TRUE, SO FALSE

PBS D corresponds to protein interactions with moderate confidence identified through a unique prey fragment (singleton) or multiple identical ones. Interactions in this category are more difficult to interpret as they represent either **true interactions hardly detectable** by Y2H different than ULTimate Y2H (low representation of the mRNA in the library, non optimal folding of the bait or the prey fragment, prey toxicity...) or **background noise** due to stochastic variations of the system or rare random events.

HIGHLY CONNECTED PROTEINS

In a final step of PBS computation, protein interactions identified within the same organism are analyzed. Prey interacting domains which have been found in 6 or more independent screens are considered highly connected (HC) and the corresponding interactions are flagged with a PBS E. Our experience with 3,500 screens performed has allowed us to classify highly connected proteins in 3 categories:

1. Proteins that are known to be highly connected due to their biological function (modifying enzyme, chaperones, protein degradation enzymes...) and which therefore easily exceed the threshold of 6 when many screens are performed within the same organism.

Naturally highly connected

Prey protein	HC domain length [aa]	Number of interactions	InterPro domain contained in HC domain
HSPCB Heat shock protein HSP 90-beta	25	22	IPR009079 Four-helical cytokine
UBE2I ubiquitin-conjugating enzyme E2	112	37	IPR000608 Ubiquitin-conjugating enzyme E2

2. Preys with a prey interacting domain that contains a known protein interaction motif (e.g. PDZ) or a biochemically promiscuous motif (e.g. highly charged residues). Their higher connectivity may be specific or the consequence of the biochemical promiscuity (interaction biochemically relevant but not necessarily occurring in a physiological setting).

Known or promiscuous interaction motif

Prey protein	HC domain length [aa]	Number of interactions	InterPro domain contained in HC domain
PTPN3 Tyrosine-protein phosphatase non-receptor type 3	74	19	IPR001478 PDZ/DHR/GLGF
YWHAB Protein kinase C inhibitor protein 1	40	11	IPR000308 14-3-3 protein

3. Artifacts of the Y2H technology. These can be LexA or Gal4 protein binders, binders of the DNA sequence upstream of the reporter gene...

Technical Y2H artifacts

Prey protein	HC domain length [aa]	Number of interactions	InterPro domain contained in HC domain
COPS5 COP9 signalosome complex subunit 5	189	26	IPR000555 Mov34/MPN/PAD-1
SNRP70 var1/2 small nuclear ribonucleoprotein 70kDa polypeptide	53	55	IPR000504 RNA-binding region RNP-1
RANBP9 Ran-binding protein 9	79	146	IPR006594 Lissencephaly type-1-like homology motif
HEYL Hairy/enhancer-of-split related with YRPW motif-like protein	26	228	IPR003650 Orange

All these categories of proteins represent highly connected preys that we flag with a PBS E to warn researchers on their non-specificity. We consider the wording "sticky proteins" is misleading since it is associated with non-naturally occurring interactions while some of them actually do occur in cells.

**PROTEIN INTERACTIONS,
WHO ELSE CAN DO IT?**