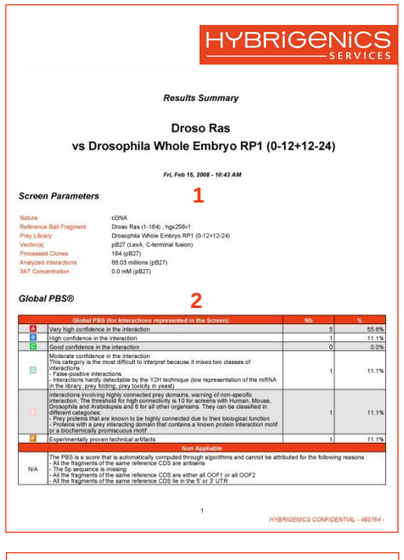
How to read your Results Documents

Your results are spread on 3 different files: a PDF document called **"Results Summary",** an Excel® worksheet which includes **raw data** and a PDF file called **"DomSight"** which includes information on bait and prey structural, functional and interaction domains.
You will find below more details on these documents. Our commercial team is at your disposal should you need further information.



HYBRIGENICS SERVICES

Results Summary

Droso Ras
vs Drosophila Whole Embryo RP1 (0-12+12-24)

Fri, Feb 15, 2008 - 10:43 AM

Screen Parameters **1**

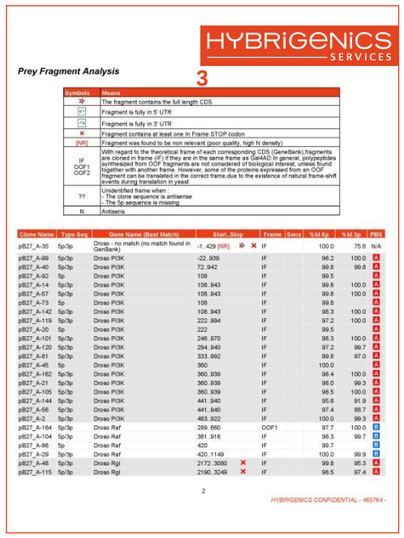
Nature	cDNA
Reference Bait Fragment	Droso Ras (1-164) ; hgx256v1
Prey Library	Drosophila Whole Embryo RP1 (0-12+12-24)
Vector(s)	pB27 (LexA, C-terminal fusion)
Processed Clones	164 (pB27)
Analyzed Interactions	68.03 millions (pB27)
3AT Concentration	0.0 mM (pB27)

Global PBS® **2**

Global PBS (for Interactions represented in the Screen)	Nb	%
A Very high confidence in the interaction	5	55.6%
B High confidence in the interaction	1	11.1%
C Good confidence in the interaction	0	0.0%
D Moderate confidence in the interaction This category is the most difficult to interpret because it mixes two classes of interactions : - False-positive interactions - Interactions hardly detectable by the Y2H technique (low representation of the mRNA in the library, prey folding, prey toxicity in yeast)	1	11.1%
E Interactions involving highly connected prey domains, warning of non-specific interaction. The threshold for high connectivity is 10 for screens with Human, Mouse, Drosophila and Arabidopsis and 6 for all other organisms. They can be classified in different categories: - Prey proteins that are known to be highly connected due to their biological function - Proteins with a prey interacting domain that contains a known protein interaction motif or a biochemically promiscuous motif	1	11.1%
F Experimentally proven technical artifacts	1	11.1%
Non Appliable		
N/A The PBS is a score that is automatically computed through algorithms and cannot be attributed for the following reasons : - The interaction involves a missing reference CDS are antisens - The 5p sequence is missing - All the fragments of the same reference CDS are either all OOF1 or all OOF2 - All the fragments of the same reference CDS lie in the 5' or 3' UTR		

1

HYBRIGENICS CONFIDENTIAL - 465764 -



HYBRIGENICS SERVICES

Prey Fragment Analysis **3**

Symbols	Means
5p	The fragment contains the full length CDS
[5']	Fragment is fully in 5' UTR
[3']	Fragment is fully in 3' UTR
X	Fragment contains at least one In Frame STOP codon
[NR]	Fragment was found to be non relevant (poor quality, high N density)
IF OOF1 OOF2	With regard to the theoretical frame of each corresponding CDS (GeneBank),fragments are cloned in frame (IF),if they are in the same frame as Gal4AD. In general, polypeptides synthesized from OOF fragments are not considered of biological interest, unless found together with another frame. However, some of the proteins expressed from an OOF fragment can be translated in the correct frame,due to the existence of natural frame-shift events during translation in yeast.
??	Unidentified frame when - The clone sequence is antisense - The 5p sequence is missing
N	Antisens

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop		Frame	Sens	%Id 5p	%Id 3p	PBS
pB27_A-35	5p/3p	Droso - no match (no match found in GenBank)	-1..429 [NR]	✲ ✕	IF		100.0	75.6	N/A
pB27_A-99	5p/3p	Droso PI3K	-22..939		IF		96.2	100.0	A
pB27_A-40	5p/3p	Droso PI3K	72..942		IF		99.8	99.8	A
pB27_A-92	5p	Droso PI3K	108		IF		99.5		A
pB27_A-14	5p/3p	Droso PI3K	108..943		IF		99.8	100.0	A
pB27_A-57	5p/3p	Droso PI3K	108..943		IF		99.8	100.0	A
pB27_A-73	5p	Droso PI3K	108		IF		99.8		A
pB27_A-142	5p/3p	Droso PI3K	108..943		IF		98.3	100.0	A
pB27_A-119	5p/3p	Droso PI3K	222..994		IF		97.2	100.0	A
pB27_A-20	5p	Droso PI3K	222		IF		99.5		A
pB27_A-105	5p/3p	Droso PI3K	246..970		IF		98.3	100.0	A
pB27_A-120	5p/3p	Droso PI3K	294..940		IF		97.2	99.7	A
pB27_A-81	5p/3p	Droso PI3K	303..892		IF		99.8	97.0	A
pB27_A-45	5p	Droso PI3K	360		IF		100.0		A
pB27_A-162	5p/3p	Droso PI3K	360..939		IF		98.4	100.0	A
pB27_A-21	5p/3p	Droso PI3K	360..939		IF		98.0	99.3	A
pB27_A-105	5p/3p	Droso PI3K	360..939		IF		98.5	100.0	A
pB27_A-144	5p/3p	Droso PI3K	441..940		IF		95.6	91.9	A
pB27_A-56	5p/3p	Droso PI3K	441..940		IF		97.4	88.7	A
pB27_A-2	5p/3p	Droso PI3K	483..922		IF		100.0	99.3	A
pB27_A-164	5p/3p	Droso Raf	288..660		OOF1		97.7	100.0	B
pB27_A-104	5p/3p	Droso Raf	391..916		IF		98.3	99.7	B
pB27_A-66	5p	Droso Raf	420		IF		99.7		B
pB27_A-29	5p/3p	Droso Raf	420..1149		IF		100.0	99.9	B
pB27_A-48	5p/3p	Droso Rgl	2172..3080	✕	IF		99.8	95.3	A
pB27_A-115	5p/3p	Droso Rgl	2190..3249	✕	IF		98.5	97.4	A

2

HYBRIGENICS CONFIDENTIAL - 465764 -

1- Results Summary

1 ***Screen parameters***

It is a summary of the technical parameters of your screen(s).
3-AT concentration: 3-amino-triazole (3-AT) may be used in order to decrease the background.

2 ***Global PBS®***

For each interaction, a Predicted Biological Score (PBS) is computed to assess the interaction reliability. This score represents the probability of an interaction to be non-specific: it is an e-value, primarily based on the comparison between the number of independent prey fragments found for an interaction and the chance of finding them at random (background noise).

The value varies between 0 and 1. Several thresholds have been arbitrarily defined in order to rank the results in 4 categories from A (the highest confidence rank) to D.

PBS D generally represents interactions identified through one unique prey fragment or multiple identical ones. It can be interactions hardly detectable by the Y2H technique (low representation of the mRNA in the library, prey folding, prey toxicity in yeast…) or it can be false-positive interactions.

The PBS is adjusted by integrating the PBS of other interactions from our database in which interaction domains of the involved proteins have been found. For example, reciprocal interactions found in independent screens are technically very reliable and thus tagged as A, B or C. You will find more details about the calculation in the Formstecher *et al.* paper (see. Selected References, page 4).

Two additional categories have been implemented: PBS E and PBS F. The PBS E represents interactions involving prey domains connected to more than 20 different human bait proteins in our entire database. This threshold is lower, at 10 distinct interactions, for Mouse, Drosophila and Arabidopsis whose libraries have been less screened so far. The limit for all other organisms is 6 connections. This arbitrary threshold allows us to flag highly – or relatively highly – connected protein domains. Our experience with more than 7,000 screens performed so far has allowed us to classify highly connected proteins in different categories: i) 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; ii) Proteins 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).
Experimentally proven artifacts of the Y2H technology are flagged with a PBS F. These can be LexA or Gal4 protein binders, binders of the DNA sequence upstream of the reporter gene…

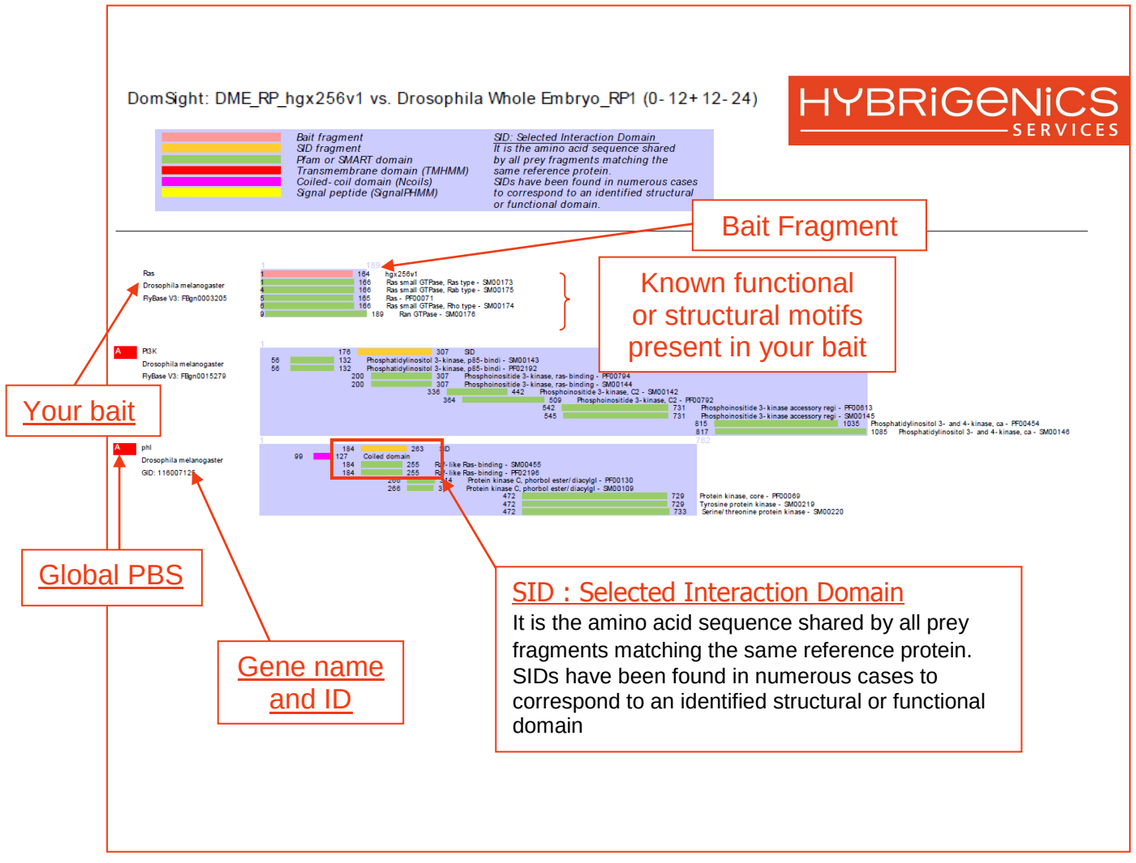
This confidence scoring system was made possible thanks to the reproducibility of the screens (>80%) and the availability of a large data set of interactions.

3 ***Prey Fragment Analysis***

It includes additional information regarding prey fragments.

2- DomSight

DomSight compares the bait fragment and the Selected Interacting Domain (SID) of the prey proteins with the functional and structural domains (PFAM, SMART, TMHMM, SignalP, Coil algorithms) on these proteins.
N.B: N/A annotations are not considered as interactions.



Selected references

- **Rain J.-C.**, *et al.* "The protein-protein interaction map of *Helicobacter pylori*". (2001) *Nature 409, 211-215* : describes the first genome-scale bacterial interactome (genomic Y2H screens) and the computation of our confidence score, the PBS.
- **Formstecher E.**, *et al.* "Protein interaction mapping: A Drosophila case study." (2005) *Genome Research, 15, 376-384* : describes the application of our technology to Drosophila as a model organism to study cancer-related proteins (general characteristics of our cDNA libraries, description of our screening protocol, full description and benchmarking of the confidence score can be found in the Supplemental Material).

3- Excel® worksheet

The worksheet contains raw data, in particular 5p and 3p experimental sequences, and the bioinformatics analysis.

Clone name:

DME_RP: Library nickname
hgx256v1: Bait fragment reference
pB27: Bait vector
A-99: Prey clone number

Name and GID:

Best match and reference sequence from GenBank or dedicated databases
(ex: Flybase for Drosophila)
Proteic Match (PM) = No hit found on nucleic bank but the sequence match on the NR proteic bank.
GenMatch = No hit found on nucleic or NR protein database but the sequence matches on a chromosomic sequence.

Additional Gene Notes:

Annotations from GenBank or dedicated database
and gene identification

Frame:

The reference database is:
IF: In frame with the Gal4 Activation Domain
OOF1, OOF2: Out of frame. Frameshifting is constitutive in yeast.

UTR inclusion:

Warning when the prey fragment is fully included in the 5p or 3p untranslated region (potential false positive)

Alu Warning 5p and 3p:

"Yes" indicates the presence of ALU sequences in the raw sequence data

Frag theoretical Sequence:

Shows the whole theoretical fragment sequence as deduced from the reference gene and the experimental start / stop positions of prey fragments

Microsoft Excel - AP18774484_DME_RP_hgx256v1_pB27.xls

A33 fx DME_RP_hgx256v1_pB27_A-147

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Mon May 15 15:28:05 CEST 2006 HYBRIGENICS CONFIDENTIAL Program = DME_RP Screen = hgx256v1 Vector = pB27 (LexA, C-terminal fusion)																
2	hgx000256-Droso Ras (1-164) vs Drosophila Whole Embryo RP1 (0-12+12-24)																
3	**Clone Name**	**Type Seq**	**Gene Name (Best Match)**	**GID**	**Global PBS**	**Additional Gene Notes**	**Start**	**Stop**	**Frame**	**Orientation**	**UTR Inclusion**	**% Id 5p/3p**	**Alu Warning 5p**	**Alu Warning 3p**	**5p Sequence**	**3p Sequence**	**Frag theoretical Sequence**
4	DME_RP_hgx256v1_pB27_A-45	5p	Droso PI3K	GID: 2	A	Phosphoinosc	360	No Dat	IF	Sense		100.0			CGGCATCCT	CAGCCTCG	GCGAGCGCA
5	DME_RP_hgx256v1_pB27_A-162	5p 3p	Droso PI3K	GID: 2	A	Phosphoinosc	360	939	IF	Sense		98.4 / 100.0			CGGCATCC	GACTGCTC	CGGCATCCT
6	DME_RP_hgx256v1_pB27_A-144	5p 3p	Droso PI3K	GID: 2	A	Phosphoinosc	441	940	IF	Sense		95.6 / 91.9			GATCGNGN	AGCATCTT	GATCGGCAC
7	DME_RP_hgx256v1_pB27_A-56	5p 3p	Droso PI3K	GID: 2	A	Phosphoinosc	441	940	IF	Sense		97.4 / 88.7			GATCGNGA	CNAGANTN	GATCGGCAC
8	DME_RP_hgx256v1_pB27_A-2	5p 3p	Droso PI3K	GID: 2	A	Phosphoinosc	483	922	IF	Sense		100.0 / 99.3			CGAGATTG	ACGACTTC	CGAGATTGA
9	DME_RP_hgx256v1_pB27_A-164	5p 3p	Droso Raf	: CG28	B	Raf kinase ; pc	289	660	OOF1	Sense		97.7 / 100.0			CACTATNA	CAGCAACA	CACTATCAA
10	DME_RP_hgx256v1_pB27_A-104	5p 3p	Droso Raf	: CG28	B	Raf kinase ; pc	381	916	IF	Sense		98.3 / 99.7			GCCGGGCA	CAACTCAT	GCCGGGCAG
11	DME_RP_hgx256v1_pB27_A-86	5p	Droso Raf	: CG28	B	Raf kinase ; pc	420	No Dat	IF	Sense		99.7			GATCCTTTT	GCGAGCCCA	CCTGCCCAA
12	DME_RP_hgx256v1_pB27_A-29	5p 3p	Droso Raf	: CG28	B	Raf kinase ; pc	420	1149	IF	Sense		100.0 / 99.9			GATCCTTTT	AGGGCTGT	GATCCTTTTG

Type sequence:

Indicates if 5p and/or 3p sequences are available for prey identification.
For technical reasons, 5p or 3p sequence could be missing. It does not prevent the prey identification. One single sequence (5p or 3p) can be enough to resolve a prey fragment on its full length.

Global PBS:

The Predicted Biological Score indicates the confidence we have in an interaction (see part 1)

Start/Stop:

Position of the 5p and 3p prey fragment ends, relative to the position of the ATG start codon (A = 0)

Orientation:

Sense or antisense in respect to the reference sequence

% Id 5p/3p:

Indicates the % identity of the prey fragment sequences with the gene reference sequence

5p sequence/3p sequence

Prey fragment experimental sequences