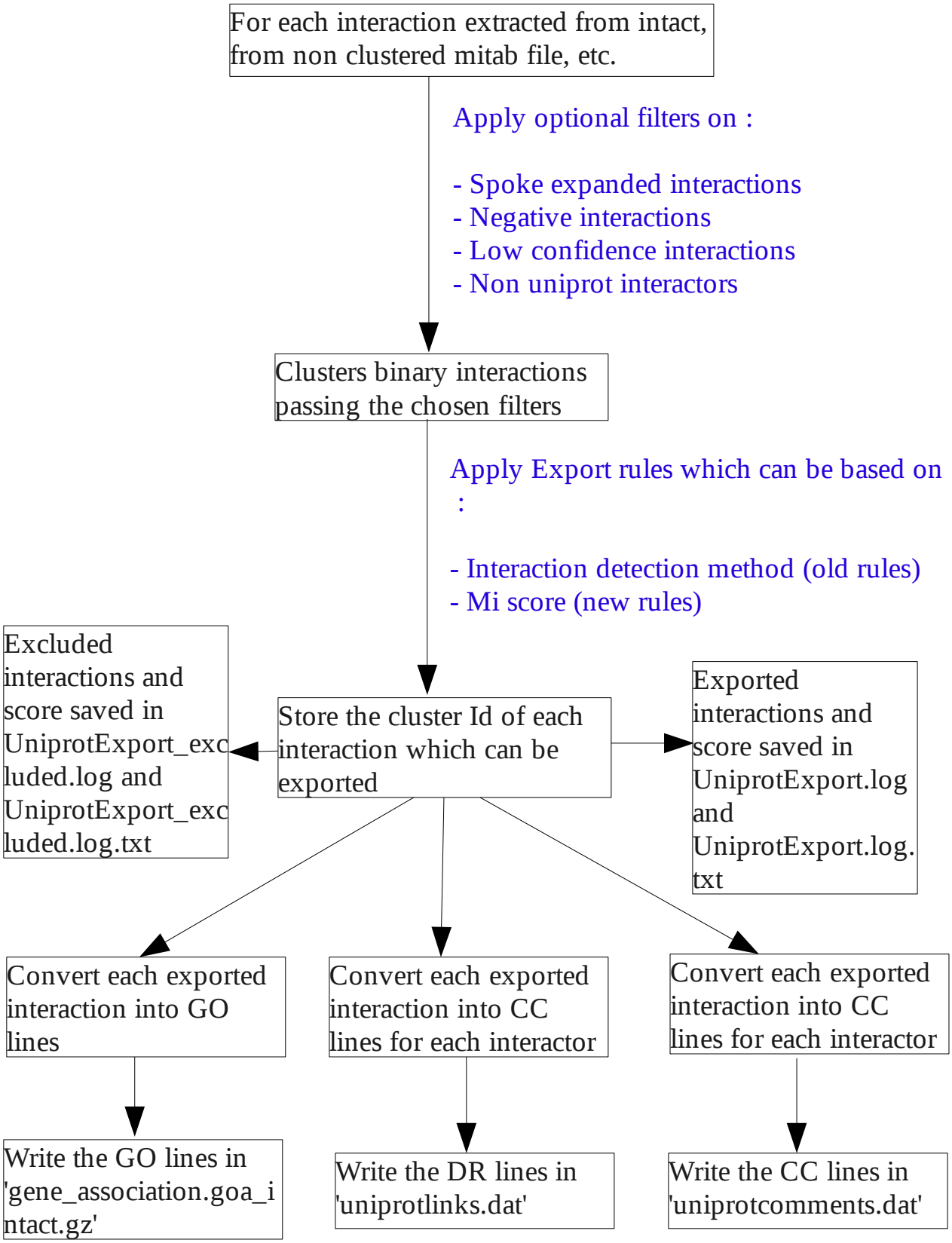
General Uniprot Export Data flow



DR line export specifications

Each DR line must have the following format :

*DR IntAct; **UNIPROT**; **INTERACTIONS**.*

Where :
UNIPROT : is the uniprot AC of a protein in Intact. Isoforms and feature chain ids are not accepted and must be remapped to the parent uniprot AC. It is a mandatory field.

INTERACTIONS : is the number of distinct PPI interactions in IntAct involving this protein. This number must be strictly superior to 0.

- Old Uniprot Export specifications : Only true positive binary interactions involving uniprot proteins must be taken into account. The low confidence interactions must be ignored (interactions with author confidence not matching the 'uniprot-dr-export' annotation of the experiment or interactions with 'uniprot-dr-export' = 'no').
- New Uniprot Export specifications : Spoke expanded binary interactions involving uniprot proteins must be taken into account as well as true binary interactions. Negative interactions are accepted. The low confidence interactions must be ignored (interactions with author confidence not matching the 'uniprot-dr-export' annotation of the experiment or interactions with 'uniprot-dr-export' = 'no').

CC line export specifications

1) Old Uniprot export specifications

Each CC line must have the following format :

For each uniprot protein :

AC ***UNIPROT_MASTER***
CC -!- INTERACTION

For each binary interaction involving the protein :

- If self interactions (two participants having the same uniprot ac and stoichiometry = 0 or 1 for each participant OR one participant with stoichiometry = 2) and same organisms :

*CC Self; NbExp=**INTERACTIONS**; IntAct=**FIRST_INTACT**, **SECOND_INTACT**;*

- If self interactions and different organisms :

*CC Self (xeno); NbExp=**INTERACTIONS**; IntAct=**FIRST_INTACT**, **SECOND_INTACT**;*

- If two different interactors and same organism :

*CC **SECOND_UNIPROT**:**GENE**; NbExp=**INTERACTIONS**; IntAct=**FIRST_INTACT**, **SECOND_INTACT**;*

- If two different interactors and different organisms :

*CC **SECOND_UNIPROT**:**GENE** (xeno); NbExp=**INTERACTIONS**; IntAct=**FIRST_INTACT**, **SECOND_INTACT**;*

…

//

Where :
UNIPROT_MASTER is the uniprot AC of a protein in Intact. Isoforms and feature chain ids are not accepted and must be remapped to the parent uniprot AC. It is a mandatory field and must be unique in a same file.

SECOND_UNIPROT is the uniprot AC of a protein in Intact interacting with the UNIPROT_MASTER. Isoforms and feature chain ids are accepted. It is a mandatory field and must be unique in a same CC line.

GENE is the gene name associated with the SECOND_UNIPROT. It is a mandatory field. If no gene names are available, first look at the locus name, then at the ORF name.

INTERACTIONS is the number of positive interaction evidences in IntAct involving this proteins. This number must be strictly superior to 0. Only evidences of true binary interactions involving uniprot proteins must be taken into account. The low confidence interactions must be ignored (interactions with author confidence not matching the 'uniprot-dr-export' annotation of the experiment or interactions with 'uniprot-dr-export' = 'no'). All evidences must have interaction detection methods passing export rules (described later).

FIRST_INTACT is the intact AC of the protein in Intact which is associated with the UNIPROT_MASTER. It is a mandatory field.

SECOND_INTACT is the intact AC of the protein in Intact which is associated with the SECOND_UNIPROT. It is a mandatory field.

2) New Uniprot export specifications (copy of the document from Uniprot)

The CC line topic INTERACTION conveys information about binary protein-protein interactions which is derived from the IntAct database.

A description of its current format is available in the UniProtKB User Manual.

This format will be modified to include further supplementary experimental information which will allow the user to more easily gauge the level of support or confidence for any given protein interaction. The new format will also allow the representation of isoform- and chain-specific interactions (such as those involving specific portions of viral polyproteins) as well as negative interactions.

Each binary interaction will be represented by a block of several lines:

* The first line describes whether the experimental evidence indicates that the two proteins interact or do not interact and provides cross-reference(s) to other databases storing the experimental data.

* The second line is an optional comment about the interaction.

* The third and fourth lines identify the interacting proteins.

* The fifth and following line(s) (where present) contain supplementary experimental information:

- the type of interaction that was inferred from the experiment(s)
- the experimental method used to demonstrate the interaction
- the source of the data (literature reference)

During data import from IntAct, experiments which share identical interaction type and experimental method are condensed into one line (see examples).

The new format of INTERACTION in the flat file is:

CC -!- INTERACTION:
(CC Interact=(Yes|No); Xref=db:db_id(, db:db_id);*
(CC Comment=free_text;)?
CC Protein1=label [id(:ft_id)?];
CC Protein2=label [id(:ft_id)?];(Organism=tax_name [NCBI_TaxID:tax_id];)?
(CC InteractionType=type; Method=method; Source=db:db_id(, db:db_id);)*
+)+
//

Where:

* Interact= indicates if there is experimental evidence that the two proteins interact (Yes) or do not interact (No) under the experimental conditions described in the cited publication(s).

* db:db_id : Database cross-reference, consisting of database name and unique database identifier.

- Xref= contains currently only one cross-reference to the IntAct database. May contain cross-references to other protein-protein interactions databases in the future. It can be composed of Intact Ac1, Intact Ac2 which is the identifier of the binary interaction.
- Source= contains currently only cross-references to the PubMed database. May contain cross-references using DOIs in the future.

* Comment= contains additional information concerning the interaction (like PTMs, subcellular location, etc.).

* Protein1= and Protein2= describe the two interacting proteins:
  - label: Biologically meaningful label or name for the protein, or chain thereof. This field will generally take the value of the gene name (as shown in the GN line Name= field of the UniProtKB entry). When no gene name is available then the value of the corresponding OrderedLocusNames= or ORFNames= field of the GN line of the UniProtKB entry may be used. In the absence of any available gene designation, this field will be filled with a dash "-".
    If the protein contains multiple chains (e.g. a viral polyprotein), and the interaction involves only one of these chains, then the chain name (from the FT line) is displayed and the identifier which follows in square brackets will contain the FTId (unique feature identifier) for that chain.
  - id: UniProtKB accession number or isoform identifier (IsoId= field).
  - ft_id: UniProtKB feature identifier (FTId= field).
  - ft_id: UniProtKB feature identifier (FTId= field). If Protein2 is derived from a distinct species, that is indicated in the Organism= field:
    - tax_name: Scientific name of the organism.
    - tax_id: NCBI taxonomy database identifier of the organism.

* InteractionType= describes the interaction type:

  - type: PSI-MI controlled vocabulary for interaction type.

* Method= describes the interaction detection method:

  - method: PSI-MI controlled vocabulary for interaction detection method.

Note: Perl-style multipliers indicate whether a pattern (as delimited by parentheses) is optional (?), may occur 0 or more times (*), or 1 or more times (+). Alternative values are separated by a pipe symbol (|).

Examples:

Experimental information deriving from different publications is condensed:

```
CC   -!- INTERACTION:
CC       Interact=Yes; Xref=IntAct:EBI-359343,EBI-79792;
CC         Comment=FANCD2 is ubiquitinated (PubMed:15199141).
CC         Protein1=FANCD2 [Q9BXW9];
CC         Protein2=BRCA2 [P51587];
CC         InteractionType=association; Method=anti bait coimmunoprecipitation;
Source=PubMed:18212739;
```

CC InteractionType=physical association; Method=chromatography technology; Source=PubMed:15115758;
CC InteractionType=physical association; Method=coimmunoprecipitation; Source=PubMed:15115758, PubMed:15199141;

Positive and negative isoform-specific interactions:

CC -!- INTERACTION:
CC Interact=Yes; Xref=IntAct:EBI-1018629,EBI-1569435;
CC Protein1=PIM1 [P11309-1];
CC Protein2=ABCG2 [Q9UNQ0];
CC InteractionType=physical association; Method=anti bait coimmunoprecipitation; Source=PubMed:18056989;
CC InteractionType=physical association; Method=anti tag coimmunoprecipitation; Source=PubMed:18056989;
CC InteractionType=physical association; Method=pull down; Source=PubMed:18056989;
CC InteractionType=physical association; Method=two hybrid; Source=PubMed:18056989;
CC Interact=No; Xref=IntAct:EBI-1018633,EBI-1569435;
CC Protein1=PIM1 [P11309-2];
CC Protein2=ABCG2 [Q9UNQ0];
CC InteractionType=physical association; Method=pull down; Source=PubMed:18056989;

Non-physiological heterologous interaction between rat and human proteins:

CC -!- INTERACTION:
CC Interact=Yes; Xref=IntAct:EBI-296087,EBI-1636616;
CC Protein1=AKT1 [P31749];
CC Protein2=Arrb2 [P29067]; Organism=Rattus norvegicus [NCBI_TaxID:10116];
CC InteractionType=physical association; Method=anti bait coimmunoprecipitation; Source=PubMed:18191226;
CC InteractionType=physical association; Method=pull down; Source=PubMed:18191226;

Physiological heterologous chain-specific virus-host interaction: virus entry

CC -!- INTERACTION:
CC Interact=Yes; Xref=IntAct:EBI-710918,EBI-515315;
CC Protein1=Non-structural protein 5A [Q9WMX2:PRO_0000037551];
CC Protein2=FYN [P06241]; Organism=Homo sapiens [NCBI_TaxID=9606];
CC InteractionType=physical association; Method=anti bait coimmunoprecipitation; Source=PubMed:14993658;

Physiological heterologous chain-specific host-virus interaction: host entry

CC -!- INTERACTION:
CC Interact=Yes; Xref=IntAct:EBI-515315,EBI-710918;
CC Protein1=FYN [P06241];
CC Protein2=Non-structural protein 5A [Q9WMX2:PRO_0000037551]; Organism=Hepatitis C virus genotype 1b (isolate Con1) [NCBI_TaxID=333284];
CC InteractionType=physical association; Method=anti bait coimmunoprecipitation; Source=PubMed:14993658;

GO line export specifications

Each binary interaction exported in the CC lines will be exported in the GO lines.

Each GO line must have the following format :

- If two different interactors :

UniProt ***FIRST_UNIPROT*** *GO:0005515* *PMID:**Id1**|**Id2**|**Id3**|...* *IPI* *UniProt:**SECOND_UNIPROT***
IntAct

- If self interaction :

UniProt ***FIRST_UNIPROT*** *GO:0042802* *PMID:**Id1**|**Id2**|**Id3**|..* *IPI* *UniProt: **SECOND_UNIPROT***
IntAct

Where :

FIRST_UNIPROT is the uniprot AC of a protein in Intact. Feature chain ids are not accepted and must be remapped to the parent uniprot AC. It is a mandatory field.

SECOND_UNIPROT is the uniprot AC of a protein in Intact interacting with the FIRST_UNIPROT. Feature chain ids are not accepted and must be remapped to the parent uniprot AC. It is a mandatory field.

SECOND_UNIPROT is the uniprot AC of a protein in Intact interacting with the FIRST_UNIPROT. Feature chain ids are not accepted and must be remapped to the parent uniprot AC. It is a mandatory field.

Id1|Id2|Id3|... are the pubmed Ids which are referring to the interaction. It is a mandatory field and must not be empty.

Export rules based on the detection method

If the binary interaction has at least one detection method with an 'uniprot-dr-export' annotation = 'yes', the binary interaction is exported.

If the binary interaction has only detection method with an 'uniprot-dr-export' annotation = 'no', the binary interaction is NOT exported.

If the binary interaction has only detection method with an 'uniprot-dr-export' annotation = 'CONDITION', the binary interaction is exported only if the condition is respected for at least one detection method (Ex 2 hybrid method has a condition of 2 to be exported. If two '2-hybrid' methods are found for this interactions, the condition is respected and the interaction can be exported).

List of Detection methods

term: sandwich immunoassay

yes

term: potassium permanganate footprinting

yes

term: luminescence based mammalian interactome mapping

yes

term: lambda repressor two hybrid

yes

term: tandem affinity purification

export: yes

term: 2h fragment pooling: two hybrid fragment pooling approach
id: MI:0399
export: 2

term: adenylate cyclase: adenylate cyclase complementation
id: MI:0014
export: 2

term: affinity chrom: affinity chromatography technologies
id: MI:0004
export: no

term: affinity techniques: affinity technologies
id: MI:0400
export: no

term: anti bait coip: anti bait coimmunoprecipitation
id: MI:0006
export: yes

term: anti tag coip: anti tag coimmunoprecipitation

id: MI:0007
export: yes

term: array technologies
id: MI:0008
export: no

term: bacterial display
id: MI:0009
export: yes

term: beta galactosidase: beta galactosidase complementation
id: MI:0010
export: 2

term: beta lactamase: beta lactamase complementation
id: MI:0011
export: 2

term: biochemical
id: MI:0401
export: no

term: biophysical
id: MI:0013
export: no

term: bret: bioluminescence resonance energy transfer
id: MI:0012
export: yes

term: cd: circular dichroism
id: MI:0016
export: no

term: chromatography: chromatography technologies
id: MI:0091
export: no

term: coip: coimmunoprecipitation
id: MI:0019
export: yes

term: collagen film assay
id: MI:0513
export: yes

term: coloc fluoresc probe: colocalization by fluorescent probes cloning
id: MI:0021
export: no

term: comigration in gel: comigration in non denaturing gel electrophoresis
id: MI:0404
export: no

term: competition binding
id: MI:0405
export: yes

term: complementation: protein complementation assay
id: MI:0090

export: 2

term: cosedimentation
id: MI:0027
export: yes

term: crosslink: cross-linking studies
id: MI:0030
export: yes

term: cytoplasmic compl: cytoplasmic complementation assay
id: MI:0228
export: 2

term: deacetylase assay
id: MI:0406
export: yes

term: density sedimentatio: cosedimentation through density gradients
id: MI:0029
export: yes

term: dhfr reconstruction: dihydrofolate reductase reconstruction
id: MI:0111
export: 2

term: display technologies
id: MI:0034
export: no

term: dls: dynamic light scattering
id: MI:0038
export: no

term: electron microscopy
id: MI:0040
export: yes

term: electron resonance
id: MI:0043
export: yes

term: elisa: enzyme linked immunosorbent assay
id: MI:0411
export: yes

term: endor: electron nuclear double resonance
id: MI:0041
export: yes

term: enzymatic studies
id: MI:0415
export: no

term: epr: electron paramagnetic resonance
id: MI:0042
export: yes

term: experimental
id: MI:0045
export: no

term: facs: fluorescence-activated cell sorting
id: MI:0054
export: no

term: far western blotting
id: MI:0047
export: yes

term: fcs: fluorescence correlation spectroscopy
id: MI:0052
export: yes

term: filamentous phage: filamentous phage display
id: MI:0048
export: yes

term: filter binding
id: MI:0049
export: yes

term: fluorescence: fluorescence technologies
id: MI:0051
export: no

term: fluorescence imaging: fluorescence microscopy
id: MI:0416
export: no

term: fluorescence spectr: classical fluorescence spectroscopy
id: MI:0017
export: yes

term: footprinting
id: MI:0417
export: yes

term: fps: fluorescence polarization spectroscopy
id: MI:0053
export: yes

term: fret: fluorescent resonance energy transfer
id: MI:0055
export: yes

term: gallex: lex-a dimerization assay
id: MI:0369
export: 2

term: gfp complementation: green fluorescence protein complementation assay
id: MI:0229
export: 2

term: gtpase assay
id: MI:0419
export: yes

term: htrf: homogeneous time resolved fluorescence
id: MI:0510
export: yes

term: imaging techniques

id: MI:0428
export: no

term: in gel kinase assay: in-gel kinase assay
id: MI:0423
export: yes

term: in gel phosphatase: in gel phosphatase assay
id: MI:0514
export: yes

term: inference
id: MI:0362
export: no

term: inferred by author
id: MI:0363
export: no

term: inferred by curator
id: MI:0364
export: no

term: ion exchange chrom: ion exchange chromatography
id: MI:0226
export: no

term: itc: isothermal titration calorimetry
id: MI:0065
export: yes

term: kinase htrf: kinase homogeneous time resolved fluorescence
id: MI:0420
export: yes

term: kinase spa: kinase scintillation proximity assay
id: MI:0425
export: yes

term: lambda phage: lambda phage display
id: MI:0066
export: yes

term: light microscopy
id: MI:0426
export: no

term: light scattering
id: MI:0067
export: no

term: mappit: mammalian protein protein interaction trap
id: MI:0231
export: 2

term: membrane compl: membrane bound complementation assay
id: MI:0230
export: 2

term: methyltransferase as: methyltransferase assay
id: MI:0515

export: yes

term: molecular sieving
id: MI:0071
export: no

term: mrna display
id: MI:0073
export: yes

term: ms of complexes: mass spectrometry studies of complexes
id: MI:0069
export: yes

term: nmr: nuclear magnetic resonance
id: MI:0077
export: yes

term: peptide array
id: MI:0081
export: yes

term: phage display
id: MI:0084
export: yes

term: phosphatase assay
id: MI:0434
export: yes

term: phosphatase htrf: phosphatase homogeneous time resolved fluorescence
id: MI:0509
export: yes

term: pisa: protein in situ array
id: MI:0092
export: yes

term: protease assay
id: MI:0435
export: yes

term: protease htrf: protease homogeneous time resolved fluorescence
id: MI:0511
export: yes

term: protein array
id: MI:0089
export: yes

term: protein crosslink: protein cross-linking with a bifunctional reagent
id: MI:0031
export: yes

term: protein kinase assay
id: MI:0424
export: yes

term: pull down
id: MI:0096
export: yes

term: radiolabeled acetate: deacetylase radiometric assay
id: MI:0508
export: yes

term: radiolabeled methyl: methyltransferase radiometric assay
id: MI:0516
export: yes

term: reverse phase chrom: reverse phase chromatography
id: MI:0227
export: no

term: reverse rrs: reverse ras recruitment system
id: MI:0097
export: 2

term: ribosome display
id: MI:0098
export: yes

term: saturation binding
id: MI:0440
export: yes

term: seldi chip: proteinchip(r) on a surface-enhanced laser desorption/ionization
id: MI:0095
export: no

term: sem: electron tomography
id: MI:0410
export: no

term: sls: static light scattering
id: MI:0104
export: no

term: solution sedimentati: cosedimentation in solution
id: MI:0028
export: yes

term: spa: scintillation proximity assay
id: MI:0099
export: yes

term: spr: surface plasmon resonance
id: MI:0107
export: yes

term: t7 phage: t7 phage display
id: MI:0108
export: yes

term: tem: transmission electron microscopy
id: MI:0020
export: yes

term: toxcat: tox-r dimerization assay
id: MI:0370
export: 2

term: transcription compl: transcriptional complementation assay

id: MI:0232
export: 2

term: two hybrid
id: MI:0018
export: 2

term: two hybrid array
id: MI:0397
export: 2

term: two hybrid pooling: two hybrid pooling approach
id: MI:0398
export: 2

term: ub reconstruction: ubiquitin reconstruction
id: MI:0112
export: 2

term: x-ray: x-ray crystallography
id: MI:0114
export: yes

term: yeast display
id: MI:0115
export: yes

term: zymography
id: MI:0512
export: yes

Export rules based on the mi score

The binary interaction must have a score superior to a Threshold value (0.43 or 0.40 for the moment).

The binary interaction must have at least one evidence of true binary interaction which is not COLOCALIZATION.