Protein-Protein Interaction

Elements of Biophysics

Emidio Capriotti
http://biofold.org/

Biomolecules Folding and Disease

Department of Pharmacy and
Biotechnology (FaBiT)
University of Bologna

Classification of the Methods

Wide variety of techniques and methods have been developed to generate PPI data and can be subdivided in:

- High throughput techniques
- Low throughput techniques

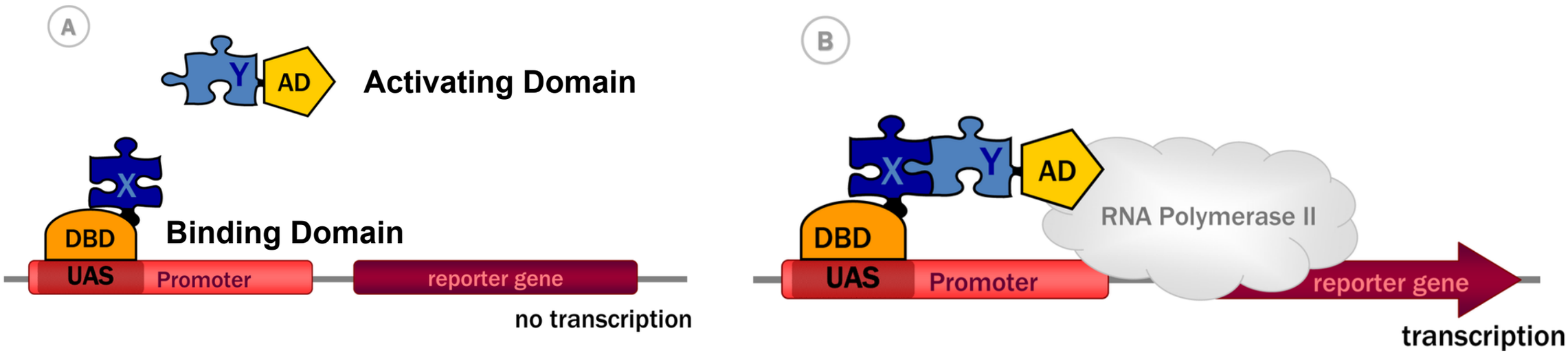
These techniques can be further divided in:

- techniques that detect direct physical interactions between two proteins, called binary methods
- techniques that detect interactions among groups of proteins that may not form physical contacts — co-complex methods.

High Throughput Techniques

The main binary methods for measuring of direct physical interactions between protein pairs is Yeast two-hybrid (Y2H).

The strategy interrogates two proteins, called bait (X) and prey (Y), coupled to two halves of a transcription factor and expressed in yeast. If the proteins make contact, they reconstitute a transcription factor that activates a reporter gene.

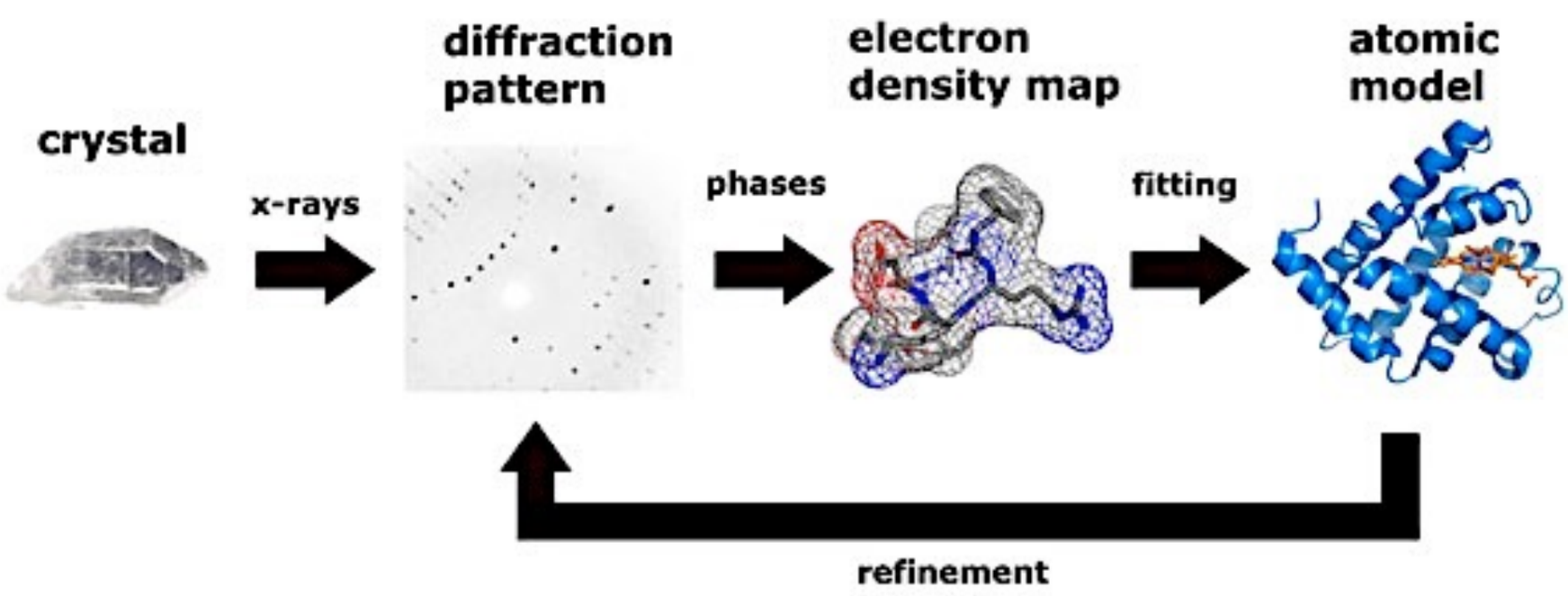


Anna Brückner, et al., Int. J. Mol. Sci. 2009

Low Throughput Techniques

Some low throughput techniques provide deeper insight certain characteristic of an interaction, such as FRET, NMR and X-ray crystallography.

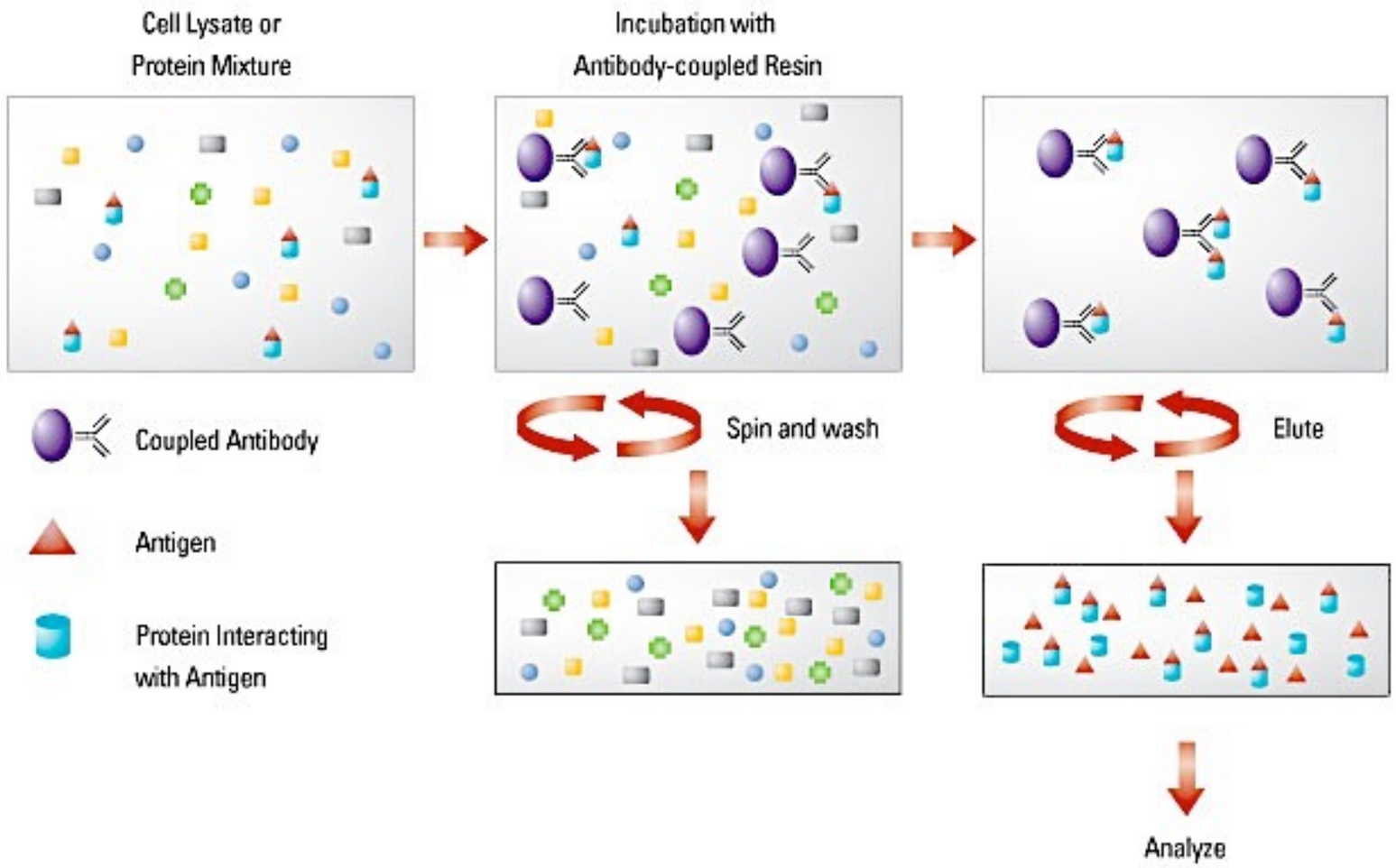
X-ray crystallography is considered the gold standard for PPI, since provide high quality data of binding surfaces to the level of individual atoms and binding sites.



From Thomas Splettstoesser (www.scistyle.com)

Co-complex Method

The most common co-complex method is co-immunoprecipitation (co-IP) coupled with mass spectrometry (MS)**.** In this approach, the bait protein, usually expressed in the cell at *in vivo* conditions, is affinity purified and the interacting partners are detected by mass spectrometry.



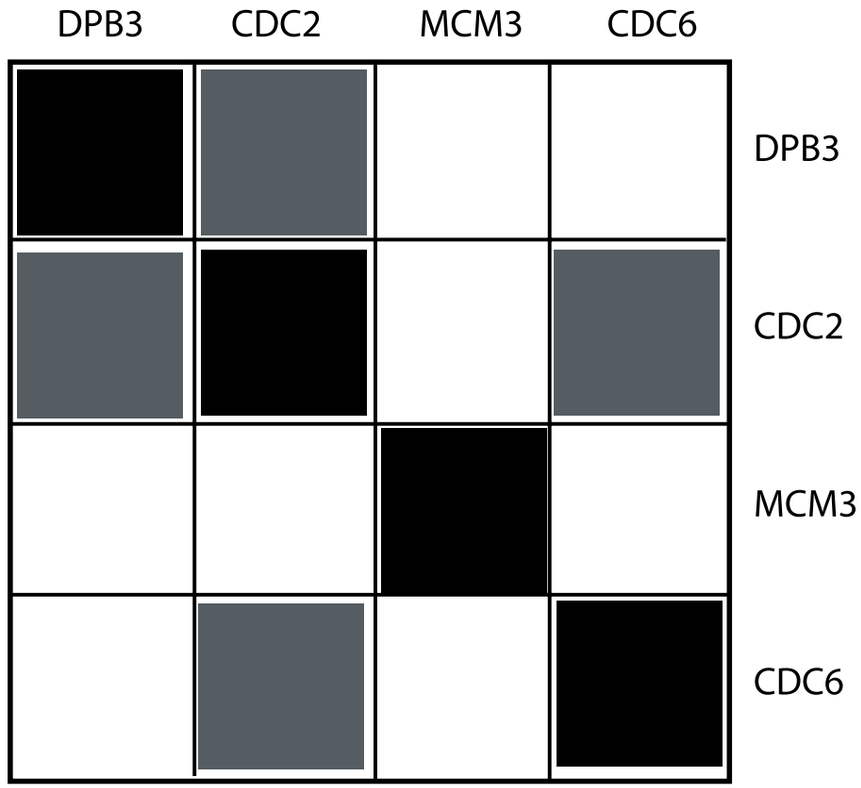
Luana Licata, Systems Biology Course 2015

Gene Co-expression

Function of a protein complex depends on the functionality of all subunits that should be present in the correct stoichiometric concentration. Thus, the gene expression levels of subunits in a complex should be related.

Several studies have tackled the problem of gene co-expression and demonstrated that interacting proteins in yeast are more likely to have their genes coexpressed compared with noninteracting proteins.

The expression levels of physically interacting proteins coevolve, and coevolution of gene expression can be a better predictor of protein interactions than coevolution of amino acid sequences

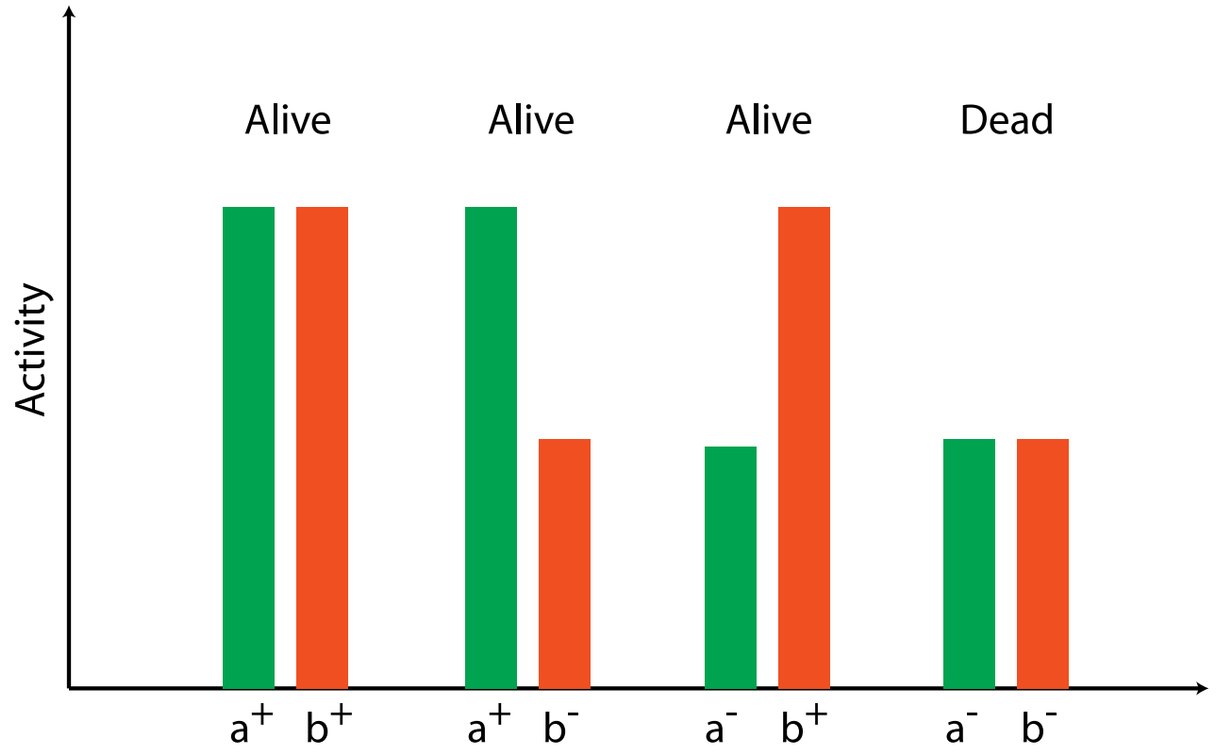


Schoemaker and Panchenko PLOS Comp Biol 2007

Synthetic Lethality

The synthetic lethality method produces mutations or deletions of two separate genes which are viable alone but cause lethality when combined together in a cell under certain conditions.

Synthetic interaction can point to the possible physical interaction between two gene products, their participation in a single pathway, or a similar function.



Schoemaker and Panchenko PLOS Comp Biol 2007

Experimental Techniques

Different experimental techniques for the detection of protein-protein interaction

Methods	HT	Assay	Interaction Type	Characterization
Y2H	+	In vivo	Physical interactions (binary)	Identification
Affinity purification-MS	+	In vitro	Physical interactions (complex)	Identification
DNA microarrays/Gene coexpression	+	In vitro	Functional association	Identification
Protein microarrays	+	In vitro	Physical interaction (complex)	Identification
Synthetic lethality	+	In vivo	Functional association	Identification
Phage display	+	In vitro	Physical interaction (complex)	Identification
X-ray crystallography, NMR spectroscopy	-	In vitro	Physical interactions (complex)	Structural and biological characterization
Fluorescence resonance energy transfer	-	In vivo	Physical interaction (binary)	Biological characterization
Surface plasmon resonance	-	In vitro	Physical interaction (complex)	Kinetic, dynamic characterization
Atomic force microscopy	-	In vitro	Physical interaction (binary)	Mechanical, dynamic characterization
Electron microscopy	-	In vitro	Physical interaction (complex)	Structural and biological characterization

Schoemaker and Panchenko PLOS Comp Biol 2007

Curation and Databases

The results of experiments are published on scientific journal. The curators extract information from the literature and to develop curated databases.

Wet Lab Scientists

Scientific Curators

Molecular Interaction Databases

Europe PMC

PubMed

UniProt

National Center for Biotechnology Information
NCBI

Verification of Interactions

There is no comprehensive gold standard interaction set. Several verification methods have been proposed:

- Expression profile reliability method: based on the observation that interacting proteins are coexpressed.

- Paralogous verification method: if two proteins interact, their paralogs most likely interact. This method identified 40% true interactions at a 1% error rate.

- Protein localization method: defines true positives as interacting proteins that are localized in the same cellular compartment and/or common cellular role. Y2H and co-IP respectively 50% and 100% true positive.

Schoemaker and Panchenko PLOS Comp Biol 2007

Interaction Databases

Molecular interaction databases have been established to archive and subsequently disseminate molecular interaction data in a structured format available to perform searches and bioinformatics analyses.

Molecular interaction databases can be divided in:

- Primary databases: experimentally proven protein interactions coming from either small-scale or large-scale published studies that have been manually curated

- Meta databases: experimentally proven PPIs obtained by consistent integration of several primary databases

- Prediction databases: mainly predicted PPIs derived using different approaches, combined with experimentally proven PPIs.

Database Classification

Type of data captured:

- Only PPIs information as MINT and DIP.
- Interactions between proteins and other molecular types (DNA, RNA, small molecules) as IntAct and MatrixDB.
- PPIs and genetic interactions as BIOGRID.
- Only PPIs related to a specific scientific topic such as : InnateDB (PPIs in the immune system), MPIDB (PPIs in microbes) and MatrixDB (extracellular PPIs).

Type of curation Policy:

- Databases describing PPIs with low level of curation details and quality control procedures
- Databases describing PPIs with high level of curation details and high accuracy of quality control procedures such as IMEx databases.

Important Databases

A complete list of molecular interaction databases is available at: http://www.pathguide.org.

Database name	Data types	Main Taxonomies	Archival/thematic	Curation depth	IMEx Member	PSICQUIC service	Ref.
IntAct	All	Full	Archival	IMEx/MIMIx	Full	Yes	[6]
MINT	PPIs	Full	Archival	IMEx/MIMIx	Full	Yes	[7]
InnateDB	PPIs	Human and mouse	Proteins involved in innate immunity	IMEx/MIMIx	Full	Yes	[10]
MPIDB	PPIs	Bacteria and archaea	Microbial proteins	IMEx/MIMIx	Full	Yes	[9]
I2D	PPIs	Model organisms	Cancer related proteins	IMEx/MIMIx	Full	Yes	
DIP	PPIs	Full	Archival	IMEx	Full	Yes	[1]
MatrixDB	PPIs; PSMIs	Human and mouse	Extracellular matrix	IMEx	Full	Yes	[8]
BioGRID	PPIs	Model organisms	Archival	Limited	Observer	Yes	[13]
HPRD	PPIs	Human	Human	Limited	No	No	[38]
ChEMBL	Drug-target PSMIs	Targets mainly human or pathogens	Drug-target	MIABE [39]/MIMIx	No	Yes	[16]
BindingDB	Drug-target PSMIs	All	Drug-target	MIABE/MIMIx	No	Yes	[40]
PubChem BioAssay	Drug-target PSMIs	Targets mainly human or pathogens	Drug-target	MIABE/MIMIx	No	No	[19]
PrimesDB	PPIs	Human and mouse	EGFR network	Limited	Observer	No	
HPIDB	PPIs	Model organisms and pathogens	Host-pathogen systems	IMEx	Full	Application pending	[34]

IMEX/MIMIx – the database contains both IMEx and MIMIx standards data.

PPIs – Protein-Protein Interactions; PSMIs –Protein-Small Molecule Interactions.

IMEx Consortium

- The International Molecular Exchange Consortium established a collaboration between a group of major public interaction data providers who have agreed to share curation effort (www.imexconsortium.org)

- 13 active molecular interaction databases dedicated to producing high quality, annotated data, curated to the same standards and following the same curation rules

- Data is curated once at a single centre then exchanged between partners

- Users can query a single website to obtain all data

Imex Central

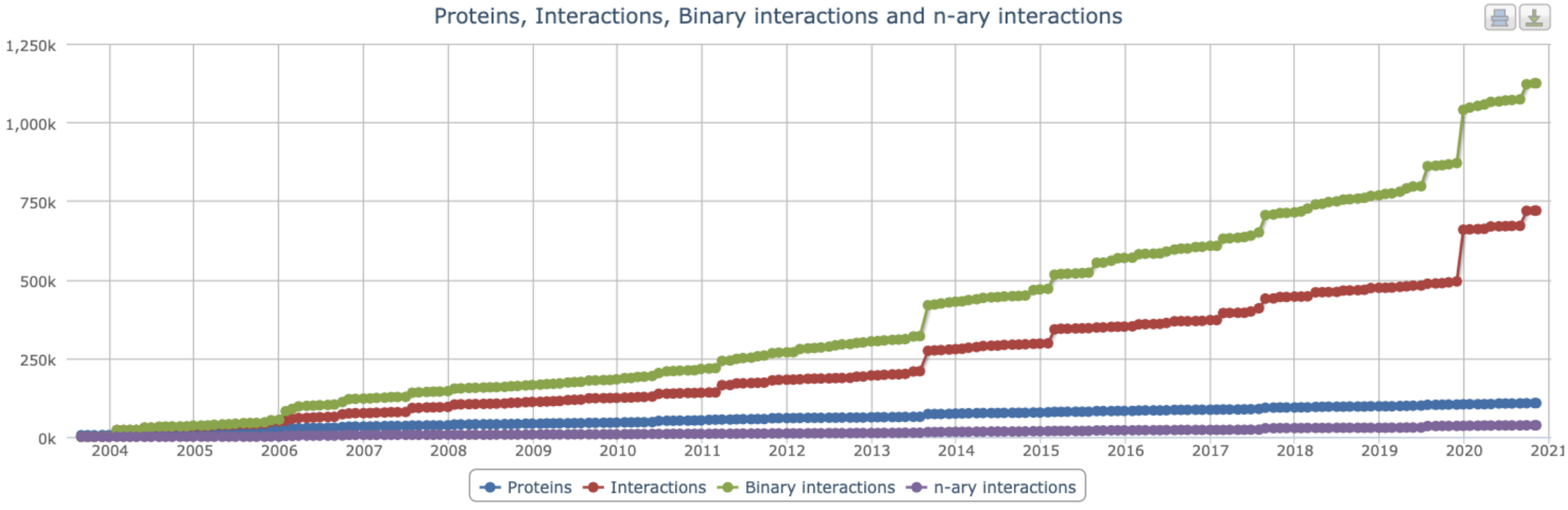
The web service IMEx Central (https://imexcentral.org/icentralbeta/) is a central resource to assign IMEx IDs to the publications curated by IMEx members (version BETA-0.93 has been recently released).

Curators can check by using the NCBI PubMed identifier (PMID) if other IMEx members have curated an already published paper and therefore it allows avoiding work duplication.

MIntAct Project

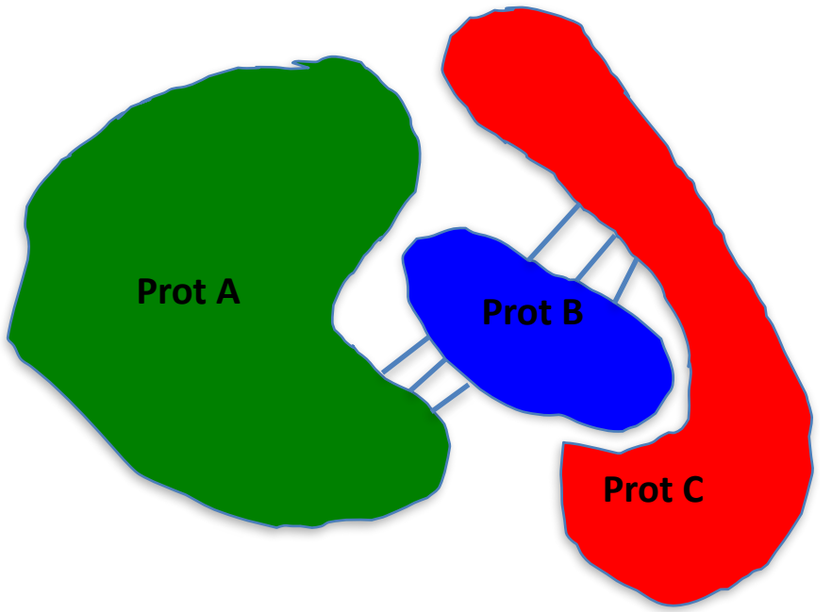
- MINT and IntAct databases were two of the largest databases (number of manuscripts curated and the number of non-redundant interactions).
- Both adopted the highest possible data quality standards.
- Both were founder members of the IMEx Consortium.

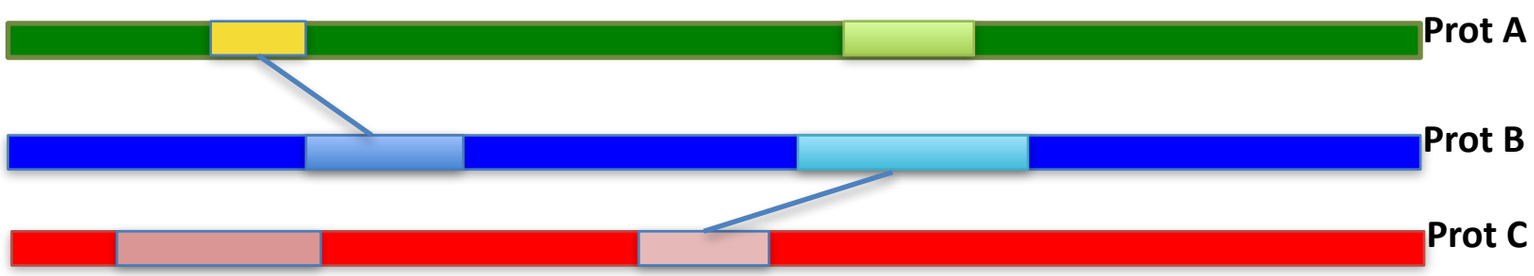
IntAct and MINT joined forces to create a single resource to improve curation and software development efforts.



PPI Representation

Representation of binding domain of interacting proteins in IMEx databases

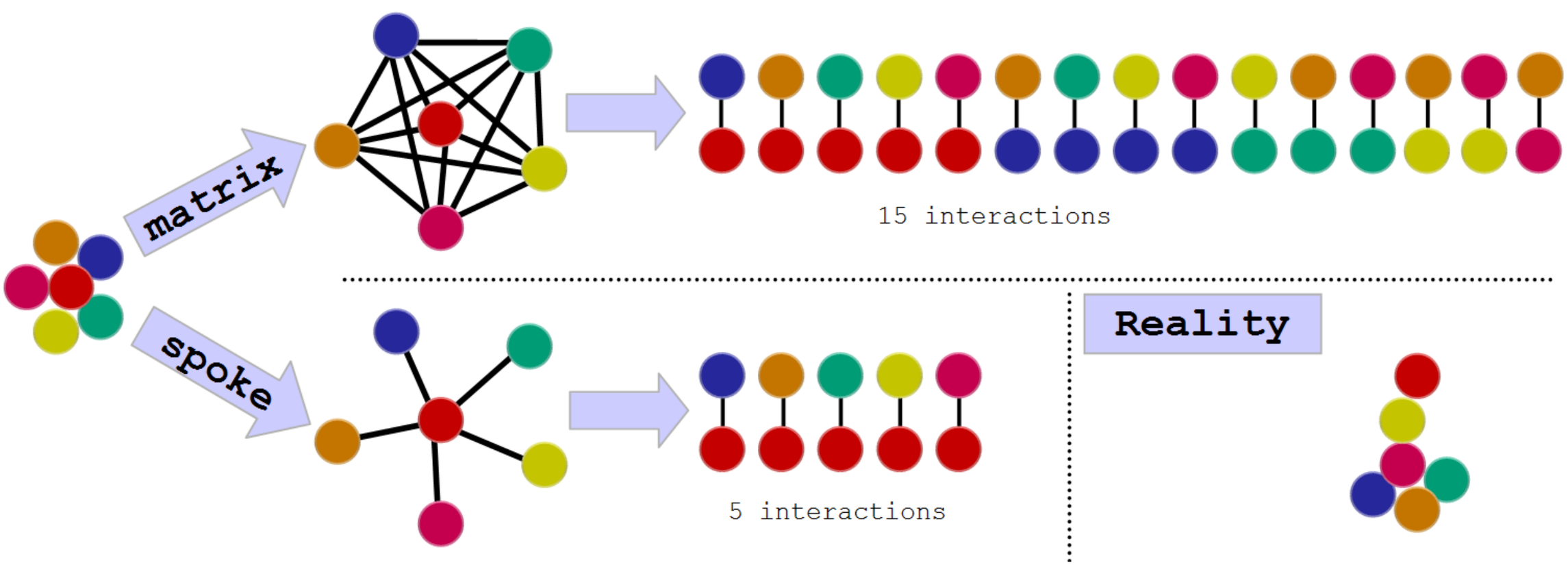




Luana Licata, Systems Biology Course 2015

Complex Representation

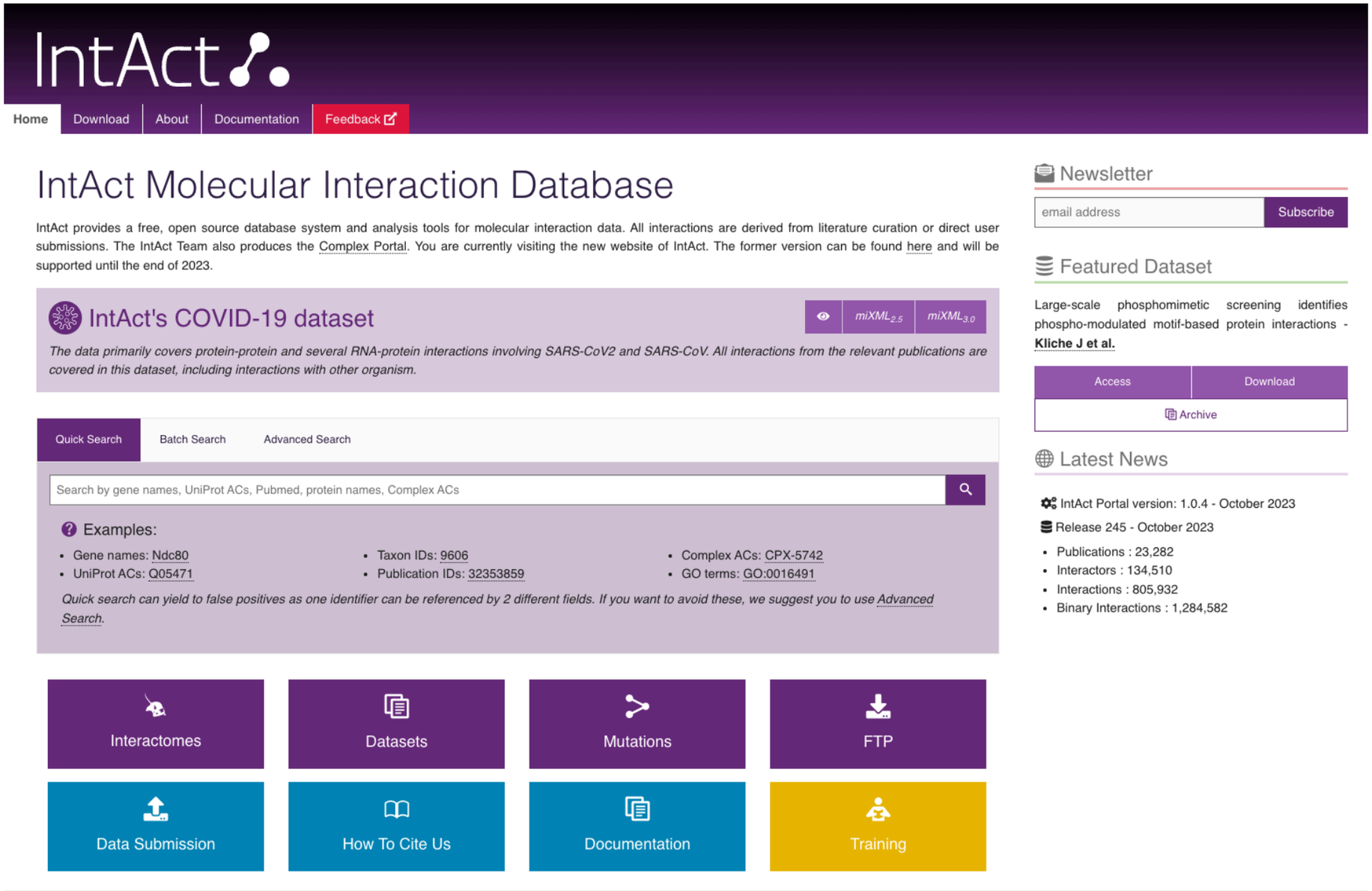
- Several experimental techniques produce complex data: Eg. co-IP coupled with MS
- There are two algorithms available to convert complexes into binary interactions



http://www.ebi.ac.uk

IntAct Interface

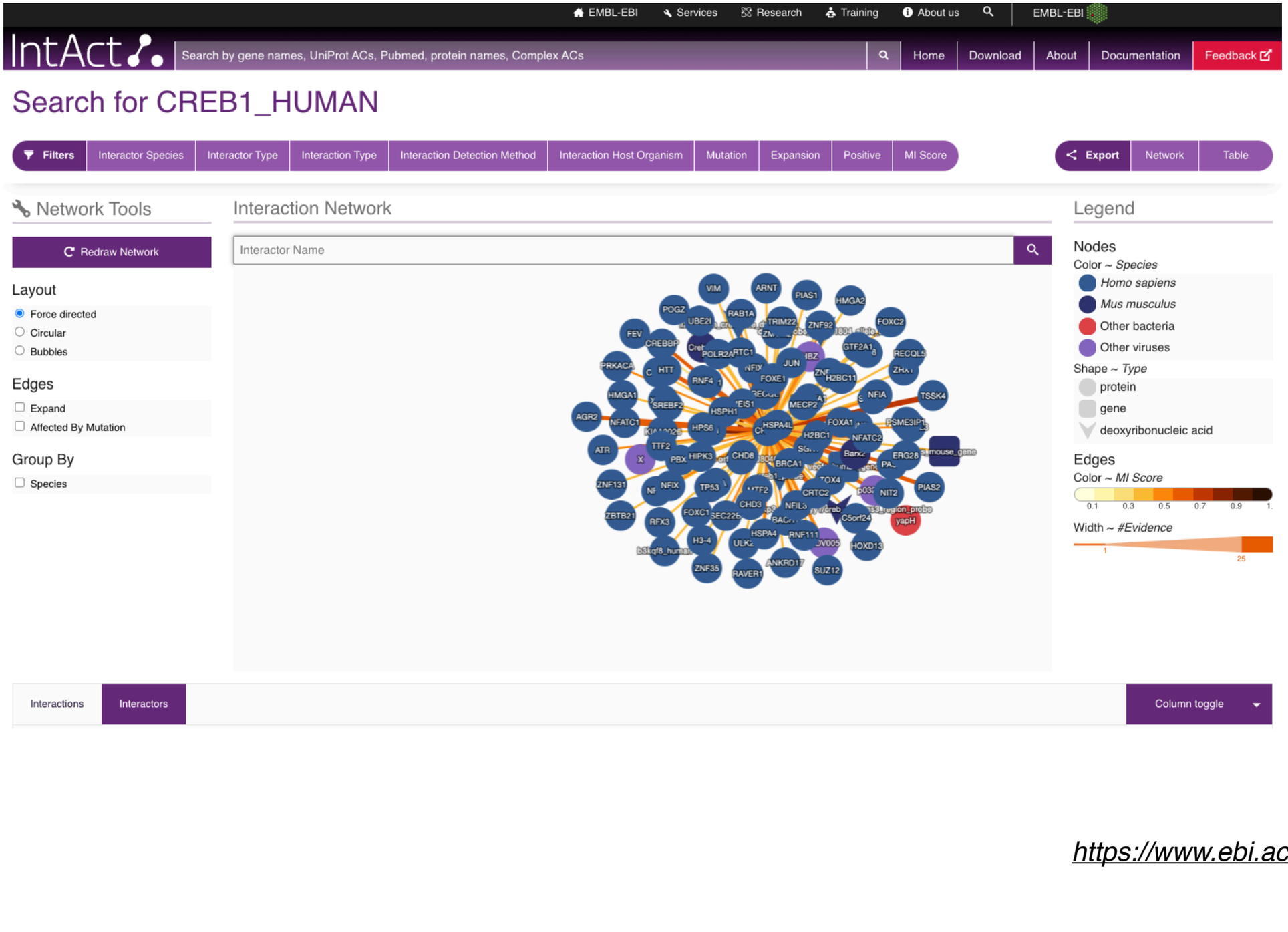
Use the input window to search for the interactions of the the Human Phosphorylation-dependent transcription factor CREB1.



https://www.ebi.ac.uk/intact/

IntAct Output

The CREB1 has 157 interactions, 145 of which are with human proteins



https://www.ebi.ac.uk/intact/

PPI Data Format

The first molecular interaction databases independently established their own dataset formats and curation strategies:

In 2002, The HUPO-Proteomics Standards Initiative (HUPO-PSI) defined community standards for data representation of proteomics data to facilitate data comparison, exchange and verification.

The development of PSI-MI XML schema has facilitated the description of protein-protein interactions.

An Excel-compatible, tab-delimited format, MITAB, has been developed for users who require only minimal information but in a more accessible configuration.

PSI-MITAB File

PSI-MITAB 2.7 Standard Columns (42)

- ID(S) INTERACTORS
- ALT. ID(S) INTERACTORS
- ALIAS(ES) INTERACTORS
- INTERACTION DETECTION METHOD(S) (Col 7)
- PUBLICATION FIRST AUTHOR(S)
- PUBLICATION IDENTIFIER(S)
- TAXID INTERACTORS (Cols 10 -11)
- INTERACTION TYPE (Col 12)
- SOURCE DATABASE(S)
- INTERACTION IDENTIFIER(S)
- CONFIDENCE VALUE(S) EXPANSION METHOD(S)
- BIOLOGICAL ROLE(S)
- EXPERIMENTAL ROLE(S)
- TYPE OF INTERACTORS (Cols 21 - 22)
- PROPERTIES (CROSS REFERENCES) OF INTERACTORS/INTERACTION
- ANNOTATION OF INTERACTORS/INTERACTION
- HOST ORGANISM(S)
- PARAMETER OF INTERACTION
- FEATURE(S) INTERACTORS
- STOICHIOMETRY(S) INTERACTORS
- PARTECIPANT IDENTIFICATION METHODS

Exercise

Download the IntAct.zip file from the the ftp server.

- Search for the interactions of the MEKK1 protein.
  How many interaction you can find? Are all this referring to the same protein?

- Refine your search using the UniProtID Q13233.
  How many interaction you have now?

- Search for the interaction of BRAF. Is BRAF interacting with MEKK1? How many experimental data are supporting the existence of this interaction?

- From the PSI-MITAB file extract all the interactions between human proteins from UniProt. How many unique interactions are present?