Systems Pharmacology

DAVID RAMÍREZ

dramirezs@udec.cl

Web lab: ***ramirezlab.github.io***

Pharmacoinformatics & Drug Design Lab
Departamento de Farmacología
Facultad de Ciencias Biológicas
Universidad de Concepción

Systems pharmacology

Image Credit: LeoWolfert/Shutterstock.com

Any ideas???

Systems pharmacology

Image Credit: LeoWolfert/Shutterstock.com

Systems pharmacology aims to model the effect of drugs on complex biological systems (both experimentally and computationally), instead of focusing on specific interactions between molecules as classic reductionism models.

Systems pharmacology

1. What are the characteristics of diseases for which drugs at a single target may not be therapeutically efficacious?

2. How does intracellular and intercellular networking give rise to adverse events?

3. How do we relate the efficacy of (poly)pharmacology to the genomic status of the individual, and how does genomic status interact with environment and behavior to control (poly)pharmacology efficacy?

4. How do we determine what combinations of targets are most likely to be effective for polypharmacology of complex diseases?

5. Can we use the human interactome and the genomic status of the individual to predict therapeutic efficacy and adverse event probability prior to commencement of therapy?

Effect of drugs on complex biological systems

Effect of drugs on complex biological systems

https://doi.org/10.33548/SCIENTIA686

Effect of drugs on complex biological systems

Without treatment

With treatment

Effect of drugs on complex biological systems

Effect of drugs on complex biological systems

The integration of high-throughput data technologies with computational data analysis and modeling offers new opportunities to overcome the one disease, one target, one drug approach.

What really happens when you take a drug?

What really happens when you take a drug?

Can we predict drug efficacy and toxicity?
Can we reuse old drugs?
Can we design personalized medicines and treatments?

Schematic representation of the multiscale networks needed to understand and predict drug action

Annu Rev Pharmacol Toxicol. 2012 ; 52: 505–521

Relationships between proteins/genes involved with disease, drugs, and drug targets

Relationships between proteins/genes involved with disease, drugs, and drug targets

https://doi.org/10.1002/msj.20191 (2010)

Relationships between proteins/genes involved with disease, drugs, and drug targets

https://doi.org/10.1002/msj.20191 (2010)

Diseaseome - drugome (~3,200): Are those targets important?

How can we work in the lab?

Development of whole cell models to predict drug responses

Healthy cells

Healthy cells
+ Insult

Healthy cells
+ Insult
+ Know drug

Healthy cells
+ Insult
+ Tested drug

Development of whole cell models to predict drug responses

Healthy cells

Healthy cells
+ Insult

Healthy cells
+ Insult
+ Know drug

Healthy cells
+ Insult
+ Tested drug

Development of whole cell models to predict drug responses

Genomics

GO Category	P-value
cell cycle	1.2E-34
cell cycle process	8.1E-33
mitotic cell cycle	9.7E-33
mitotic cell cycle process	3.5E-29
chromosome segregation	4.8E-26
cell division	6.4E-24
ribosome biogenesis	1.1E-23
nuclear chromosome segregation	1.1E-22
rRNA processing	1.3E-22
rRNA metabolic process	4.3E-22
chromosome organization	6.9E-20
ribonucleoprotein complex biogenesis	3.0E-18
DNA replication	3.9E-18
sister chromatid segregation	6.4E-18
nuclear division	1.9E-17
ncRNA processing	2.5E-17
cell cycle phase transition	5.4E-16
organelle fission	7.7E-16
mitotic cell cycle phase transition	1.2E-15
DNA metabolic process	3.9E-15

GO Category	P-value
nervous system development	6.6E-53
trans-synaptic signaling	3.3E-34
synaptic signaling	3.3E-34
chemical synaptic transmission	2.2E-33
anterograde trans-synaptic signaling	2.2E-33
generation of neurons	2.5E-33
neurogenesis	1.0E-32
cell-cell signaling	4.2E-30
neuron differentiation	1.9E-26
cell-cell adhesion via plasma-membrane adhesion molecules	4.0E-25
homophilic cell adhesion via plasma membrane adhesion molecules	1.7E-24
regulation of membrane potential	1.5E-22
neuron development	3.2E-22
behavior	1.7E-20
cell-cell adhesion	2.7E-20
modulation of synaptic transmission	3.1E-20
cell adhesion	8.3E-20
biological adhesion	1.9E-19
regulation of nervous system development	2.2E-19
neuron projection dev	3.2E-18

Development of whole cell models to predict drug responses

Transcriptomics

Development of whole cell models to predict drug responses

Proteomics

Development of whole cell models to predict drug responses

Metabolomics

Wu RQ, J Dent Res. 2011

Development of whole cell models to predict drug responses

from Yugi et al., Trends Biotechnol. 2016 Apr;34(4):276–290

Development of whole cell models to predict drug responses

Healthy cells

Healthy cells
+ Insult

Healthy cells
+ Insult
+ Know drug

Healthy cells
+ Insult
+ Tested drug

Some examples

Annu Rev Pharmacol Toxicol. 2012 ; 52: 505–521.

An intracellular network to explain how drug-induced adverse events can propagate across scales of organization. Drug interaction with the target leads through the network to an alteration of channel activity, which leads to a change in duration of the myocyte action potential, which leads to prolongation of the QT interval as seen in the electrocardiogram. This can result in fatal arrhythmias such as torsades de pointes. This network explains how drugs used to treat very different pathophysiologies such as diarrhea (loperamide) and cancer (dasatanib) can lead to long QT syndrome as an adverse event. However, this network does not explain why only some people show drug-induced long QT syndrome and why only some patients with long QT syndrome develop fatal arrhythmias. Additional information on genomic status and tissue- and multiorgan-level networks may be needed to explain individual susceptibility.

The drugs are shown in purple boxes. Red boxes are drug targets, green boxes are intermediate nodes, and blue boxes are channels responsible for the various phases of the myocyte action potential. The light blue box represents a node that is an intermediate and also a channel involved in myocyte action potential, and each brown box represents a node that is both an intermediate and also a drug target. Black arrows indicate edges that are either undirected or directed with an unknown effect type (inhibition or activation), red arrows indicate edges that are activating, and blue arrows indicate edges that are inhibitory.

Annu Rev Pharmacol Toxicol. 2012 ; 52: 505–521.

Polypharmacology of the cancer drug sorafenib

Phylogenetic tree of the human kinome with sorafenib targets highlighted in red circles.

Sorafenib targets (*b*, *c* and *d*) with positive (red) binding site. ERK (*e*) is not a target of sorafenib and its modeled binding site exhibits lower negative electrostatic potential (blue).

Birtwistle, Marc R., et al. "Systems pharmacology: an overview." *Systems Pharmacology and Pharmacodynamics* (2016): 53-80.

What do we want?

Example:

Molecular pathogenesis and mechanisms of thyroid cancer

Example:

Molecular pathogenesis and mechanisms of thyroid cancer

Nature Reviews Cancer . 13, pages 184–199 (2013)

Example:

Molecular pathogenesis and mechanisms of thyroid cancer

Maybe we are able to open the door now

What are we doing

Pharmacoinformatics & drug design

Systems Pharmacology

Classic
one drug - one target

Systems Biology

Systems Pharmacology

Pritchard, et al. DDW-online 2018 - Dec 28

Systems Pharmacology

Vogt, et al. M*edicine, Health Care and Philosophy* 19.2 (2016): 307-323.

Systems Pharmacology - graph theory

Functional nodes

Healthy person

Systems Pharmacology - graph theory

Functional nodes

Person with symptoms

Systems Pharmacology - graph theory

Functional nodes

Person sick

Protein-Protein interaction (PPI) networks

Functional nodes

PPI networks

https://www.metro.cl/tu-viaje/plano-de-red

PPI networks

Systems pharmacology

- Relational network-based analysis of biological systems
- Analysis of the patterns and prediction of the missing links

- Analysis of molecular transcriptional signatures
- Comparison of drugs mechanism of action

- Data extrapolation
- Data analysis and visualization

- Target similarity assessment
- Prediction of ligand cross-reactivity

- Statistical pattern recognition
- Model development for biological system analysis

- Detection of putative binding pockets
- Ligandability assessment

- Ligand similarity assessment
- Prediction of ligand cross-reactivity

- Prediction of ligand-protein molecular interactions
- Drug-based repurposing
- Reverse docking for target-based repurposing

March-Villa. Et al. (2017) *Front. Pharmacol.*, https://doi.org/10.3389/fphar.2017.00298

Network pharmacology

Network pharmacology allows the representation and analysis of drug-target systems using tools derived from **graph theory**.

Graph theory and the idea of **topology** was first described by the Swiss mathematician Leonard Euler as applied to the problem of the seven bridges of Königsberg.

From : https://www.ebi.ac.uk/training/online/

Network pharmacology

Network pharmacology allows the representation and analysis of drug-target systems using tools derived from **graph theory**.

Graph theory and the idea of **topology** was first described by the Swiss mathematician Leonard Euler as applied to the problem of the seven bridges of Königsberg.

From : https://www.ebi.ac.uk/training/online/

Network pharmacology

Network pharmacology allows the representation and analysis of drug-target systems using tools derived from **graph theory**.

Graph theory and the idea of **topology** was first described by the Swiss mathematician Leonard Euler as applied to the problem of the seven bridges of Königsberg.

From : https://www.ebi.ac.uk/training/online/

Graph theory: graph types and edge properties

Networks can represent many different types of data. The **nodes** represent different entities (e.g. proteins or genes in biological networks), and **edges** convey information about the links between the nodes.

Network edges:

Undirected edges

This type of edge is found in protein-protein interaction networks (PPIs). The relationship between the nodes is a simple connection, without a given 'flow' implied, since the evidence behind the relationship only tells us that A binds B.

From : https://www.ebi.ac.uk/training/online/

Graph theory: graph types and edge properties

Networks can represent many different types of data. The **nodes** represent different entities (e.g. proteins or genes in biological networks), and **edges** convey information about the links between the nodes.

Network edges:

Directed edges

This is the kind of connection found, for example, in metabolic or gene regulation networks. There is a clear flow of signal implied and the network can be organised hierarchically.

From : https://www.ebi.ac.uk/training/online/

Graph theory: graph types and edge properties

Networks can represent many different types of data. The **nodes** represent different entities (e.g. proteins or genes in biological networks), and **edges** convey information about the links between the nodes.

Network edges:

Weighted edges

Directed or undirected edges can also have weight or a quantitative value associated with them. This is used to depict concepts such a reliability of an interaction, the quantitative expression change that a gene induces over another or even how closely related two genes are in terms of sequence similarity. Edges can also be weighted by their centrality values or several other topological parameters.

From : https://www.ebi.ac.uk/training/online/

Graph theory: adjacency matrices

Every network can be expressed mathematically in the form of an adjacency matrix. In these matrices the rows and columns are assigned to the nodes in the network and the presence of an edge is symbolised by a numerical value. By using the matrix representation of the network we can calculate network properties such as degree, and other centralities by applying basic concepts from linear algebra.

Undirected

	A	B	C	D	E	F	G	Degree
A	0	1	1	1	1	1	0	5
B	1	0	0	0	0	1	0	2
C	1	0	0	0	0	0	0	1
D	1	0	0	0	0	0	0	1
E	1	0	0	0	0	0	0	1
F	1	1	0	0	0	0	1	3
G	0	0	0	0	0	1	0	1

Directed

	A	B	C	D	E	F	G	Out-degree
A	0	1	0	0	0	0	0	1
B	0	0	1	1	0	1	1	4
C	0	0	0	1	0	0	0	1
D	0	0	0	0	1	0	0	1
E	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0

Weighted

	A	B	C	D	E	F	G	Degree
A	0	8	12	12	12	16	12	72
B	8	0	0	0	0	4	0	12
C	12	0	0	0	0	0	0	12
D	12	0	0	0	0	0	0	12
E	12	0	0	0	0	0	0	12
F	16	4	0	0	0	0	12	32
G	12	0	0	0	0	12	0	24

Graph theory: adjacency matrices

Try your self, complete the adjacency matrix by adding a 0 or 1 into each space in the matrix:

	A	B	C	D	E	F	G
A							
B							
C							
D							
E							
F							
G							

From : https://www.ebi.ac.uk/training/online/

Graph theory: adjacency matrices

Try your self, complete the adjacency matrix by adding a 0 or 1 into each space in the matrix:

	A	B	C	D	E	F	G
A	0	1	1	1	1	0	0
B	1	0	0	0	0	0	0
C	1	0	0	0	0	0	0
D	1	0	0	0	0	1	1
E	1	0	0	0	0	1	0
F	0	0	0	1	1	0	0
G	0	0	0	1	0	0	0

From : https://www.ebi.ac.uk/training/online/

Graph theory: network topology

The purpose of any type of network analysis is to work with the complexity of the network to extract meaningful information that you would not have if the individual components were examined separately -> *classic pharmacology.*

Topological properties can help us identify relevant sub-structures within a network.

Topology is the **way in which the nodes and edges are arranged within a network**. Topological properties can apply to the network as a whole or to individual nodes and edges.

The degree of a network – The degree is the number of edges that connect to a node. It is a fundamental parameter that influences other characteristics, such as the centrality of a node. The degree distribution of all nodes in the network helps define whether a network is scale-free or not. In the figure, the degree of each node is indicated and reflected in its size and colour. Directed network nodes have two values for degree: out-degree for those edges coming out of the node and in-degree for those edges coming into the node.

From : https://www.ebi.ac.uk/training/online/

Graph theory: network topology

Shortest paths – Shortest paths, or the shortest distance between any two nodes, is used to model how information flows. This is especially relevant in many biological networks. In the figure, the shortest path between nodes *A* and *B* is highlighted and takes five steps.

Scale-free networks – In scale-free networks most of the nodes are connected to a low number of neighbours and there are a small number of high-degree nodes (hubs) that provide high connectivity to the network. In the figure, hubs are highlighted in orange

From : https://www.ebi.ac.uk/training/online/

Graph theory: network topology

Transitivity – Transitivity relates to the presence of tightly interconnected nodes in the network called clusters or communities. These are groups of nodes that are more internally connected than they are with the rest of the network. These groups are also called topological clusters and are highlighted in the figure.

Centralities – There are different flavours of centrality, each representing different concepts. Centrality can be measured for nodes and for edges and gives an estimation on how important that node/edge is for the connectivity or the information flow of the network. The degree of a node has a direct influence on many centrality measures, most prominently on the ‘degree centrality’. Its significance is reduced in more sophisticated types of centrality measures, for example, betweenness centrality. In the figure, the most central nodes (according to their betweenness centrality) are highlighted in warm colours and the node size reflects its degree

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of information can be represented in the shape of networks in order to model the cell. Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Protein-protein interaction networks

- Represent the physical relationships between proteins. They are central to practically every process that takes place in the cell
- Proteins are represented as nodes that are linked by undirected edges

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Metabolic networks

- Represent the biochemical reactions that allow an organism to grow, reproduce, respond to the environment and maintain its structure
- Metabolites and enzymes take the role of nodes and the reactions describing their transformations are represented as directed edges
- Edges can represent the direction of the metabolic flow or regulatory effects of a specific reaction

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Genetic interaction networks

- Genetic interaction is the synergistic phenomenon where the phenotype resulting from simultaneous mutations in two or more genes is significantly different from the phenotype that would result from adding the effects of the individual mutations
- Represent a functional relationship between different genes, rather than a physical one
- Genes are represented as nodes and their relationships as edges. Depending on the type of evidence behind the interaction, directionality can be inferred for the edges

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Gene / transcriptional regulatory networks

- Represent how gene expression is controlled
- Genes and transcription factors are represented as nodes, while the relationship between them is depicted by different types of directional edges
- Regulatory RNAs and other mechanisms can also form part of this type of network
- Usually generated via databases representing consensus knowledge of gene regulation (e.g. Reactome or KEGG), although large-scale experimental datasets are increasingly available

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Cell signalling networks

- Cell signalling is the communication system that controls cellular activities
- Signalling pathways represent the ordered sequences of events and model the information flow within the cell
- Gene regulation networks can be considered as a sub-type of cell signalling networks, focusing on a specific signalling event which is often the final stage of a signalling cascade
- Elements in the pathway (e.g. proteins, nucleic acids, metabolites) are represented as nodes and the flow of information is represented by directed edges

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of data will also produce different general network characteristics in terms of connectivity, complexity and structure, where edges and nodes potentially convey multiple layers of information.

Cell signalling networks

- Are systematically represented by two types of resources:
  - **Pathway databases** (also known as 'process description' resources) such as Reactome, KEGG or Wikipathways. These aim to provide a formal representation of the current scientific consensus on cell signalling pathways. They are generated by manual curation and organise the information in the form of reactions, with substrates and products being affected by the action of catalysers. This information must be converted according to specific rules in order to be represented as a network. Some information loss can occur during this process
  - **Reaction network databases** (also known as 'activity flow' resources) such as Signor, SignaLink or SPIKE. These aim to capture known binary relationships in cell signalling, such as activation, phosphorylation, etc. They are generally manually curated, but not always. In contrast with the pathway databases, they are already graphs in the mathematical sense and require no transformation in order to be represented as a network

From : https://www.ebi.ac.uk/training/online/

Types of biological networks

Different types of information can be represented in the shape of networks in order to model the cell. The meaning of the nodes and edges used in a network representation depends on the type of data used to build the network and this should be taken into account when analysing it.

From : https://www.ebi.ac.uk/training/online/

PPI networks and their importance

Molecular interaction data can be generated using many different techniques, all of which have their strengths and weaknesses. However, it is important to stress that **all molecular interaction data is to some degree artifactual**. No single method can accurately reproduce a true binary interaction observed under physiological conditions.

Molecular interactions are important to molecular biologists because:

1. They help us to understand a protein's function and behaviour.
2. They can help us to predict the biological processes that a protein of unknown function is involved in:
   - We may assume "Guilt by association" if a protein of unknown function associates with one of known function
   - Proteins involved in the same process should cluster together in network maps
3. They can help us to characterise protein complexes and pathways; interaction networks can be used as a draft 'map' to add detail to biological processes and pathways and can help discover new pathways, complexes and functional modules within the cell.

From : https://www.ebi.ac.uk/training/online/

PPI - and their importance

From : https://www.ebi.ac.uk/training/online/

Properties of PPIs: small world effect

Protein-protein interaction networks show a small world effect meaning that there is great connectivity between proteins. In other words, it can be said that the network's diameter (the maximum number of steps separating any two nodes) is small, no matter how big the network is. This usually means that any two nodes are separated by less than six steps, more or less, reflecting the now widely popularised "six degrees of separation" theory used in social sciences.

This level of connectivity has important biological consequences, since it allows for an efficient and quick flow of signals within the network. However, it also poses an interesting question: if the network is so tightly connected, why don't perturbations in a single gene or protein have dramatic consequences for the network?

Biological systems are extremely robust and can cope with a relatively high amount of perturbations in single genes/proteins. In order to explain how can this happen, we need to have a look at another fundamental property of PPINs: they are scale-free networks.

From : https://www.ebi.ac.uk/training/online/

Properties of PPIs: scale-free networks

Protein-protein interaction networks are scale-free networks. The majority of nodes (proteins - blue) in scale-free networks have only a few connections to other nodes, whereas some nodes (hubs - orange) are connected to many other nodes in the network.

Scale-free network

- **Stability**
  - If failures occur at random, and the vast majority of proteins are those with a small degree of connectivity, the likelihood that a hub would be affected is small
  - If a hub-failure occurs, the network will generally not lose its connectedness, due to the remaining hubs
- **Invariant to changes of scale**
  - No matter how many nodes or edges the network has, its properties remain stable
  - The presence of hubs is what allows for the small-world effect to be present regardless of the size of the network
- **Vulnerable to targeted attack**
  - If we lose a few major hubs from the network, the network is turned into a set of rather isolated graphs
  - Hubs are enriched with essential/lethal genes. For example, many cancer-linked proteins are hub proteins (e.g. the tumour suppressor protein p53)

From : https://www.ebi.ac.uk/training/online/

Properties of PPIs: transitivity

Another crucial characteristic of PPINs is their modularity.
The **transitivity** or **clustering coefficient** of a network is a measure of the tendency of the nodes to cluster together. High transitivity means that the network contains communities or groups of nodes that are densely connected internally. Following an analogy from the social sciences, “the friends of my friends are my friends”. In biological networks, finding these communities is very important, because they can reflect functional modules and protein complexes.

From : https://www.ebi.ac.uk/training/online/

Building and analysing PPIs

From : https://www.ebi.ac.uk/training/online/

PPIs representation and analysis tools

Cytoscape

Cytoscape is one of the most popular network analysis tools. It is an open-source, Java-based, multi-platform desktop application that is widely used for network representation, integration and analysis. It was originally designed for the analysis of biological networks, which remains as its main application, but can also be used for general purpose network analysis.

Gephi

A non-programmatic option for handling large networks is **Gephi**. Gephi is capable of dealing with hundreds of thousands of nodes, and millions of edges, albeit processing and especially drawing of such nets requires massive computer power.

The benefits of Gephi are that it is open source, multi-platform, and has a wide range of advanced network-related algorithms (often not found anywhere else) in the form of plugins. The one disadvantage is the lack of any capability for processing specifically biological information. It is a general network tool, and should be treated as such and used for enumerating, statistics, and visualisation.

From : https://www.ebi.ac.uk/training/online/

PPIs representation and analysis tools

Programmatic solutions

Programmatic solutions for large scale network analysis include packages such as **igraph** (for R, Python and C), **NetworkX** (for Python) or **graph-tool**. These are scripting packages that have much lower demand on your computer resources and are more amenable for automated tasks. This means they can be easily implemented as part of larger bioinformatics analysis pipelines. For example, the R implementation of igraph is often used hand-in-hand with other biostatistics packages available through this language.

igraph NetworkX graph-tool

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Computational

Experimental

STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:

Genomic Context Predictions

High-throughput Lab Experiments

(Conserved) Co-Expression

Automated Textmining

Previous Knowledge in Databases

Coverage

The STRING database currently covers 67'592'464 proteins from 14'094 organisms.

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Computational

Experimental

STRING

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases.

Data Sources

Interactions in STRING are derived from five main sources:

- Genomic Context Predictions
- High-throughput Lab Experiments
- (Conserved) Co-Expression
- Automated Textmining
- Previous Knowledge in Databases

Coverage

The STRING database currently covers 67'592'464 proteins from 14'094 organisms.

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

High-throughput:

Yeast two-hybrid

[1] The Binding (BD) domain fused to the bait protein (X) and the Active (AD) domain fused to prey protein (Y) are expressed in yeast cells. [2] If proteins X and Y interact, BD binds DNA and AD activates RNA polymerase. An example readout [3] of a Y2H assay with two bait proteins (Bait 1 and Bait 2) and five prey proteins (1 to 5)

PPI - Where the data come from?

High-throughput: Yeast two-hybrid

Advantages:

- Fast
- Inexpensive
- Scalable
- An in vivo system in which binding sites can be accurately mapped
- Provides binary relationships (despite limitations, see below)
- Clone libraries readily available are common

Disadvantages:

- Interactions occur between proteins that would not normally be present in the same cellular compartment, in the same cell type, or at the same time
- False positives occur when a yeast protein acts as a bridge for the interaction
- Limitations inherent to heterologous expression: Both bait and prey proteins can fail to be expressed or might be toxic to the cell, big proteins might need to be cut down into manageable fragments, etc
- Needs a clone library!

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

High-throughput:

Affinity purification mass spectrometry

The bait protein (yellow) is immobilised on a matrix [1]. A protein mixture is passed through and only the interacting partners (prey) are retained [2]. In the following step the prey proteins are removed, digested with a protease and the resulting peptides are analysed by MS [3].

PPI - Where the data come from?

High-throughput: Affinity purification mass spectrometry

Advantages:

- The technique detects complexes of several proteins interacting together
- Potentially, depending on the sensitivity of your MS-approach and the affinity of the interacting partners, this method has the ability to examine interactions among multiple proteins at subpicomole concentrations
- The prey proteins are present in their native state (so long as they are not affected by the sample lysis process) and concentration

Disadvantages:

- The technique detects complexes of several proteins interacting together, but not binary relationships
- Prey proteins without a peptide signature recognisable by MS (owing to obscure post-translational modifications) or present in very low amounts will not be identified
- Relevant transient and/or weak interactions may be missed
- Mixing of compartments during cell lysis/purification is a potential source of false positives. For example, interactions between proteins that would not normally be in the same cellular compartment may confound your results

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Co-immunoprecipitation

Protein complex immunoprecipitation (Co-IP) method. [1] Addition of antibody to protein extract. [2] Target proteins are immunoprecipitated with the antibody. [3] Coupling of antibody to beads. [4] Isolation of protein complexes.

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

X-ray crystallography

X-ray crystallography is used to to obtain detailed structural and chemical insights for selected interactions. The figure shows a model of the cullin complex

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Interaction databases

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Interaction databases

Primary databases are those that collect experimental molecular interaction data exclusively from peer-reviewed scientific publications. IntAct, MINT and MatrixDB are examples of this type of database. They can be further classified by the level of detail that they use to represent the information and the **depth** of their curation policies.

Secondary databases, also known as **meta-databases**, seek to integrate the data curated by several primary databases in one, integrated repository. APID and PINA are examples of this type.

Predictive databases combine the experimentally inferred data taken from primary databases with computational predictions of molecular interactions. Examples include STRING and UniHI.

From : https://www.ebi.ac.uk/training/online/

PPI - Where the data come from?

Interaction databases

Different levels of detail in their curation procedures -> **IMEx guidelines** (The International Molecular Exchange Consortium)

Assessing reliability and measuring confidence

An important concern in network analysis is whether the interaction network can be trusted to represent a “real” biological interaction. Given the noise inherent in current interactome information, it is important to be stringent when evaluating the protein-protein interaction data we use in our analysis. It is important to take into account that interactome coverage is also incomplete and patchy, so we will not always have the luxury of filtering out less reliable evidence.

There are many different methods for ascertaining reliability and giving a measure of confidence. Some strategies make use of:

- **Contextual biological information** regarding the proteins or molecules involved in the interaction. For example, overlapping co-expression patterns.
- **Count how many times a given interaction has been reported in the literature**, as a measure of experimental orthogonal validation. This is a popular and straightforward approach and there are more elaborate variations of this strategy, such as MIscore
- **Aggregated methods** that use a number of different strategies and integrate them in a single score, such as INTscore.

From : https://www.ebi.ac.uk/training/online/

Assessing reliability and measuring confidence

The MIscore method

MIscore is a method for assessing the reliability of protein-protein interaction data based on the use of standards. MIscore gives an estimation of confidence weighting on all available evidence for an interacting pair of proteins. The method allows weighting of evidence provided by different sources, provided the data is represented following the standards created by the IMEx consortium.

The method weights the:

- number of publications
- detection method
- interaction evidence type

Different interaction detection methods and interaction types have different weights, assigned by a group of expert curators. These parameters are aggregated for each interacting pair and then normalised, giving a quantitative measure of how much experimental evidence there is behind a given interaction.

From : https://www.ebi.ac.uk/training/online/

Topological PPIs analysis

From : https://www.ebi.ac.uk/training/online/

Centrality analysis

Centrality gives an estimation on how important a node or edge is for the connectivity or the information flow of the network. It is a useful parameter in signalling networks and it **is often used when trying to find drug targets.**

Centrality analysis in PPIs (node-based centralities) usually aims to answer the following question:

Which protein is the most important and why?

Node centrality represented in a network. Bigger, redder nodes have higher centrality values in this representation.

From : https://www.ebi.ac.uk/training/online/

Centrality analysis

The definition of ‘central’ varies with the context or purpose of our analysis. Centrality can be measured using different metrics and criteria:

- **Degree of the nodes**

The degree of the nodes can be used as a rough estimate of centrality. Nodes with a high degree (hubs) are key in maintaining some characteristics of scale-free networks such as their robustness and the small-world effect. However, this is a **local** measure since it does not take into account the rest of the network and the importance we give to its value depends strongly on the network’s size.

- **Global centrality measures**

Global centrality measures take into account the whole of the network. They are relative measurements that provide a normalised value that is independent of network size. There are many different types of global centrality measures, each addressing a slightly different definition of centrality. Two of the most widely used global centrality measures are **closeness** and **betweenness** centralities.

- **Other measures of centrality**

More complex measures of centrality can be defined depending on the specific method used to calculate it. For example, centralities are often calculated using ‘random walks’ where random nodes are chosen as a starting point and the ‘time’ or ‘speed’ needed to reach other nodes in the network is calculated. This can be combined with the weights assigned to nodes or edges in the graph to influence the centrality calculation derived from other features. This is the method used by the Google PageRank algorithm to assign weight to each webpage.

Note that centrality parameters can also be calculated taking the directionality of edges into account which can cause slight changes their definitions. Since we are focusing on PPIs, we will consider edges to be undirected in this course.

From : https://www.ebi.ac.uk/training/online/

Betweenness centrality

Betweenness centrality is based on communication flow. Nodes with a high betweenness centrality are interesting because they lie on communication paths and can control information flow. These nodes can represent important proteins in signalling pathways and **can form targets for drug discovery**.

The calculation of betweenness centrality is not standardised and there are many ways solve it. It is basically defined as the number of shortest paths in the graph that pass through the node divided by the total number of shortest paths.

Betweenness centrality measures how often a node occurs on all shortest paths between two nodes. Hence, the betweenness of a node *N* is calculated considering couples of nodes (*v1, v2*) and counting the number of shortest paths linking those two nodes, which pass through node *N.* Next the value is related to the total number of shortest paths linking *v1* and *v2*.

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i) / g_{jk}$$

Where g_{jk} = the number of geodesics (shortest paths) connecting *jk*, and $g_{jk}(n_i)$ = the number that node *i* is on.

To give a more intuitive example, in the graph, either node *B* or node *C* can be removed and there will still be paths leading to node *E*. Node *D*, however, is quite central, since it is required for any path leading to node *E*.
Imagine this graph as a city map and our analysis is telling us that *D* is the cross roads where traffic jams are more likely to occur. Betweenness centrality can in fact be used in city planning and there are studies aiming to optimise city transport based on this and related metrics.

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

Looking for communities in a network is a nice strategy for reducing network complexity and extracting functional modules (e.g. protein complexes) that reflect the biology of the network. There are several terms that are commonly used when talking about clustering analysis:

Community / Cluster

A general, catch-all term that can be defined as a group of nodes that are more connected within themselves than with the rest of the network. The precise definition for a community will depend on the method or algorithm used to define it. When talking about PPINs, communities fall into two categories: functional modules and protein complexes.

Module

In biology, modules are exchangeable functional units in which the nodes (proteins) do not have to be interacting in the same time or space. The most important characteristic of a module is that its intrinsic functional properties do not change when it is placed in a different context.

Complex

A complex is a group of proteins that interact with each other at the same time and in the same space, forming relatively stable multi-protein machinery.

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

Looking for communities in a network is a nice strategy for reducing network complexity and extracting functional modules (e.g. protein complexes) that reflect the biology of the network. There are several terms that are commonly used when talking about clustering analysis:

Motif

Motifs are statistically over-represented sub-graphs in a network. They correspond with a pattern of connections that generates a characteristic dynamical response (e.g. a negative feedback loop). They are less important for PPIs, but are quite useful in directed networks.

Clique

A subset of nodes in which every node is connected with every other member of the clique. A maximal clique is a clique that cannot be extended by adding an additional node not previously included in the clique. There are several different types of cliques and they can be used as the basis of algorithms that use topological criteria to look for communities.

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

Here we will focus on methods that exclusively use the topology of the network to find closely-connected components. This is generally known in graph theory as 'community detection methods'. No assumptions are made about the internal structure of these communities, **we are just looking at high-density regions**.

It is important to note that finding the best community structure is algorithmically extremely complex and is only possible for very small networks. For this reason, many approximation methods, often addressing different scenarios, have been developed. Some examples include:

- Clique-percolation method
- Markov Clustering Algorithm (MCL)
- Fuzzy C-Means
- Affinity Propagation
- Chinese Whispers Clustering
- Label Propagation Clustering

Another way to address the search for communities within a network is to use a combination of the topology of the network and some external property, such as protein expression values, as an additional layer defining communities. A good example of this popular method is the jActiveModules app for Cytoscape (17). This app "[…] searches a molecular interaction network to find expression activated sub-networks. Such sub-networks are connected regions of a network that show significant changes in expression over particular subsets of conditions" (18). In essence, connected regions within a network with differential expression can be identified using this tool.

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

Popular methods used to analyse PPIs: **Newman-Girvan fast greedy algorithm** and the **MCODE algorithm**.

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

Newman-Girvan fast greedy algorithm

- Developed for the study of networks in general, with a special focus on social and biological networks
- Identifies communities by using the edge betweenness centrality measure. Edges that connect different communities have higher centrality values, since a larger proportion of shortest paths will pass through them
- To define communities it uses the edge betweenness centrality scores to rank the edges of the network, then removes the most central edges and then re-calculates the betweenness scores until no edges are left. Edges affected by the removal are deemed to be part of the same community
- Can be considered a ‘naïve’ approach that will define communities even when they are only marginally more connected than the rest of the network

From : https://www.ebi.ac.uk/training/online/

Clustering analysis

MCODE algorithm

- Developed to find protein complexes in PPI networks
- Can be considered to be more stringent than the Newman-Girvan algorithm, since it aims to find only those sub networks that are very highly interconnected, representing relatively stable, multi-protein complexes that function as a single entity in time and space
- The parameters of the algorithm can be adjusted to make it less stringent, so that a looser definition of a community is used
- The algorithm uses a three-stage process:
    1. Weighting: a higher score is given to those nodes whose neighbours are more interconnected
    2. Molecular complex prediction: starting with the highest-weighted node (seed), recursively move out, adding nodes to the complex that are above a given threshold
    3. Post-processing: applies filters to improve the cluster quality

It is important to note that when we speak about ‘stringency’ we are talking about how interconnected the nodes within a sub-network must be in order to be considered a separate community (Figure 31). This changes depending on the biological question underlying the analysis. It is not the same to look for stable protein complexes, such as the proteasome, as it is to look for functional sub-modules representing a specific step in a signalling pathway.

From : https://www.ebi.ac.uk/training/online/

Annotation enrichment analysis

There are many different approaches that can be used to understand the biological context of protein-protein interaction networks. **Annotation enrichment analysis** is one of the most popular methods. Although it is not strictly speaking a network analysis tool, it is often used in combination with topological network analysis.

Annotation enrichment analysis uses gene/protein annotations provided by knowledge-bases such as Gene Ontology (GO) or Reactome to infer which annotations are over-represented in a list of genes/proteins that can be taken from a network.

Essentially, annotation tools perform some type of statistical test that tries to answer the following question:

"When sampling X *proteins (test set) out of* N *proteins (reference set; graph or annotation), what is the probability that* x, *or more, of these proteins belong to a functional category* C *shared by* n *of the* N *proteins in the reference set."*

The result of this test provides us with a list of terms that describe the list/network, or rather a part of it, as a whole.

This type of analysis is most frequently performed using GO annotation as a reference, but tools such as the Cytoscape apps BiNGO and ClueGO can also manage other annotation databases such as Reactome and KEGG. This is a widely used technique that helps characterise the network as a whole or sub-sets of it, such as inter-connected communities found through topological clustering analysis

From : https://www.ebi.ac.uk/training/online/

Annotation enrichment analysis

BiNGO

ClueGO

From : https://www.ebi.ac.uk/training/online/

Multi-target drug design against neurodegenerative diseases

Alzheimer's disease

Multi-target drug design against neurodegenerative diseases

Alzheimer's disease

Alzheimer's disease

Towards systems pharmacology to design multi-target directed ligands as therapeutic alternatives for Alzheimer’s disease.

Towards systems pharmacology to design multi-target directed ligands as therapeutic alternatives for Alzheimer's disease.

What caused the accident?

Towards systems pharmacology to design multi-target directed ligands as therapeutic alternatives for Alzheimer's disease.

what caused the AD?

Towards systems pharmacology to design multi-target directed ligands as therapeutic alternatives for Alzheimer's disease.

Hunting Alzheimer's disease targets

Literature

Hunting Alzheimer's disease targets

Databases

National Center for Biotechnology Information
NCBI
https://www.ncbi.nlm.nih.gov/

EMBL-EBI
https://www.ebi.ac.uk/

Open Targets
https://www.targetvalidation.org/

ChEMBL
https://www.ebi.ac.uk/chembl/

RCSB PDB
PROTEIN DATA BANK
https://www.rcsb.org/

DRUGBANK
https://go.drugbank.com/

PubChem
https://pubchem.ncbi.nlm.nih.gov/

TTD
http://db.idrblab.net/ttd/

UniProt
https://www.uniprot.org/

Hunting Alzheimer's disease targets

Open Targets

Alzheimer's disease

EFO: EFO_0000249 | ICD10: G30.9, G30 | OMIM: 608907, 502500, 615590 | UMLS: C0002395 | MeSH: D000544 | NCIt: C2866, C34524, C38778 | MedDRA: 10001896 | MONDO: 0004975

Associated targets | Profile

7103 targets associated with **Alzheimer's disease**

Filter by

Evidence-specific filters

- Data Types

Target-specific filters

- Pathway Types
- Target Classes
- Tractability Antibody
- Tractability Other Modalities
- Tractability PROTAC
- Tractability Small Molecule

Search

Download table as JSON TSV

Hunting Alzheimer's disease targets

BIG DATA ANALYTICS **DECISIONS**

Hunting Alzheimer's disease targets

T1: Associated with FDA approved (Phase IV) and marketed drugs	*-> (score 1.0)*	**565 targets**
T2: Associated with drugs in Phase I-III	*-> (score 0.7)*	
T3: Associated with drugs in preclinical phase	*-> (score 0.4)*	
T4: Targets that interact with T1, T2 and/or T3	*-> (score 0.1)*	**3048 targets**

Total = 3613 targets

Alzheimer's disease – PPI

Which targets should be modulated simultaneously to have a therapeutic impact on AD?

PPI network analysis via **topological** and **functional module** identification

- Cytoscape network analyzer
- R studio
- Python

MTGO

PPI: Protein-Protein interaction n

PPI network topological analysis

Vent diagram

Proteins in the Top-50% for each topological parameter

↓

Key proteins = **108 targets**

PPI network topological analysis

The degree of a network – The degree is the number of edges that connect to a node. It is a fundamental parameter that influences other characteristics, such as the centrality of a node.

Betweenness – is used to model how information flows. This is especially relevant in many biological networks. In the figure, the shortest path between nodes *A* and *B* is highlighted and takes five steps.

Centralities – Centrality can be measured for nodes and for edges and gives an estimation on how important that node/edge is for the connectivity or the information flow of the network. The degree of a node has a direct influence on many centrality measures, most prominently on the ‘degree centrality’. Its significance is reduced in more sophisticated types of centrality measures, for example, betweenness centrality. In the figure, the most central nodes (according to their betweenness centrality) are highlighted in warm colours and the node size reflects its degree.

PPI network topological analysis

Vent diagram

Proteins in the Top-50% for each topological parameter

↓

Key proteins = **108 targets**

PPI network topological analysis

3613 targets

Key proteins = **108 targets**

PPI network functional module analysis

MTGO (Module detection via Topological information and Gene Ontology knowledge)

Most representative GO term = 20

PPI network functional module analysis

PPI network functional module analysis

15 key targets

Uniprot ID	Group	Gene	Protein
Q14416	T2T4	GRM2	Metabotropic glutamate receptor 2 (mGluR2)
P21554	T2T4	CNR1	Cannabinoid receptor 1 (CB-R) (CB1) (CANN6)
Q14832	T2T4	GRM3	Metabotropic glutamate receptor 3 (mGluR3)
P34972	T4	CNR2	Cannabinoid receptor 2 (CB-2)
P41145	T4	OPRK1	Kappa-type opioid receptor (K-OR-1) (KOR-1)
P01189	T4	POMC	Pro-opiomelanocortin (POMC) (Corticotropin-lipotropin)
P01213	T4	PDYN	Proenkephalin-B (Beta-neoendorphin-dynorphin) (Preprodynorphin)
P41146	T4	OPRL1	Nociceptin receptor (Kappa-type 3 opioid receptor) (KOR-3)
P41143	T4	OPRD1	Delta-type opioid receptor (D-OR-1) (DOR-1)
P01303	T4	NPY	Pro-neuropeptide Y
P21917	T2T4	DRD4	D(4) dopamine receptor
O75899	T1T4	GABBR2	GABA-B receptor 2
Q9UBS5	T1T4	GABBR1	GABA-B receptor 1
P08908	T2T4	HTR1A	5-hydroxytryptamine receptor 1A (5-HT-1A)
P30872	T4	SST	Somatostatin receptor type 1 (SS-1-R) (SS1-R) (SS1R) (SRIF-2)

PPI network functional module analysis

15 key targets

Uniprot ID	Group	Gene	Protein
Q14416	T2T4	GRM2	Metabotropic glutamate receptor 2 (mGluR2)
P21554	T2T4	CNR1	Cannabinoid receptor 1 (CB-R) (CB1) (CANN6)
Q14832	T2T4	GRM3	Metabotropic glutamate receptor 3 (mGluR3)
P34972	T4	CNR2	Cannabinoid receptor 2 (CB-2)
P41145	T4	OPRK1	Kappa-type opioid receptor (K-OR-1) (KOR-1)
P01189	T4	POMC	Pro-opiomelanocortin (POMC) (Corticotropin-lipotropin)
P01213	T4	PDYN	Proenkephalin-B (Beta-neoendorphin-dynorphin) (Preprodynorphin)
P41146	T4	OPRL1	Nociceptin receptor (Kappa-type 3 opioid receptor) (KOR-3)
P41143	T4	OPRD1	Delta-type opioid receptor (D-OR-1) (DOR-1)
P01303	T4	NPY	Pro-neuropeptide Y
P21917	T2T4	DRD4	D(4) dopamine receptor
O75899	T1T4	GABBR2	GABA-B receptor 2
Q9UBS5	T1T4	GABBR1	GABA-B receptor 1
P08908	T2T4	HTR1A	5-hydroxytryptamine receptor 1A (5-HT-1A)
P30872	T4	SST	Somatostatin receptor type 1 (SS-1-R) (SS1-R) (SS1R) (SRIF-2)

How to modulate multiple targets with the same compound?

How to modulate multiple targets with the same compound?

How to modulate multiple targets with the same compound?

Strategies:

1. *Ligand-Based Virtual Screening using Machine learning*
2. *Pharmacophore-Based Virtual Screening*

How to modulate multiple targets with the same compound?

Strategies:

1. *Ligand-Based Virtual Screening using Machine learning*
2. *Ensemble Pharmacophore-Based Virtual Screening*

Results: Multi-target drug design against neurodegenerative diseases

UniProt-ID > ChEMBL-KNIME-WF > bioactive ligands (active, inactive, intermediate)

Illuminatins the druggable genome	$IC_{50}/Ki/EC_{50}$ (M)		
	Active (≤)	> Intermediate <	inactive (>)
Kinases: <= 30nM	3.00E-08		3.00E-07
GPCRs: <= 100nM	1.00E-07		1.00E-06
Nuclear Receptors: <= 100nM	1.00E-07		1.00E-06
Ion Channels: <= 10μM	1.00E-05		1.00E-04
Non-IDG Family Targets: <= 1μM	1.00E-06		1.00E-05

Bosc, N., Atkinson, F., Felix, E., Gaulton, A., Hersey, A., & Leach, A. R. (2019). Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. Journal of cheminformatics, 11(1), 1-16.

Results: Multi-target drug design against neurodegenerative diseases

UniProt-ID > ChEMBL-KNIME-WF > bioactive ligands (active, inactive, intermediate)

Illuminatins the druggable genome

	$IC_{50}/Ki/EC_{50}$ (M)		
	Active (≤)	**> Intermediate <**	**inactive (>)**
Kinases: <= 30nM	3.00E-08		3.00E-07
GPCRs: <= 100nM	1.00E-07		1.00E-06
Nuclear Receptors: <= 100nM	1.00E-07		1.00E-06
Ion Channels: <= 10μM	1.00E-05		1.00E-04
Non-IDG Family Targets: <= 1μM	1.00E-06		1.00E-05

Input: List with key UniProt-IDs + ChEMBL database

Output 1: Excel file with all multi-target ligands reported in ChEMBL (related with the input proteins)

Output 2: Excel files (one per protein) with all ligands + clean SMILEs

Output 3: Excel file with # ligands (active, inactive, intermediate) per protein.

Results: Multi-target drug design against neurodegenerative diseases

Output 1: Excel file with all multi-target ligands reported in ChEMBL (related with the input proteins)

Results: Multi-target drug design against neurodegenerative diseases

CHEMBL114484

P41145: Kappa-type opioid receptor
P41143: Delta-type opioid receptor
P41146: Nociceptin receptor
P21917: D(4) dopamine receptor

Common binding sites?
Common Pharmacophores?
Common receptophores?

How to modulate multiple targets with the same compound?

Strategies:

1. *Ligand-Based Virtual Screening using Machine learning*

Output 2: Excel files (one per protein) with all ligands + clean SMILEs

Filtering ligands (ADME/Tox)

Fingerprints + ML algorithm

ML-Model (one per protein)

How to modulate multiple targets with the same compound?

Strategies:

1. *Ligand-Based Virtual Screening using Machine learning*

How to modulate multiple targets with the same compound?

Strategies:

2. *Ensemble Pharmacophore-Based Virtual Screening*

Open for Innovation KNIME

Active ligands clustering based on Tanimoto similarity

Output 2: Excel files (one per protein) with all ligands + clean SMILEs

Significant clusters all different between themself
(Active compounds with structural diversity)
One per protein

Cluster Pharmacophore modeling (per cluster/per protein)

Results: Multi-target drug design against neurodegenerative diseases

UniProt-ID > ChEMBL-KNIME-WF > bioactive ligands (active, inactive, intermediate)

Illuminatins the druggable genome

	$IC_{50}/Ki/EC_{50}$ (M)		
	Active (≤)	**> Intermediate <**	**inactive (>)**
Kinases: <= 30nM	3.00E-08		3.00E-07
GPCRs: <= 100nM	1.00E-07		1.00E-06
Nuclear Receptors: <= 100nM	1.00E-07		1.00E-06
Ion Channels: <= 10μM	1.00E-05		1.00E-04
Non-IDG Family Targets: <= 1μM	1.00E-06		1.00E-05

Input: List with key UniProt-IDs + ChEMBL database

Output 1: Excel file with all multi-target ligands reported in ChEMBL (related with the input proteins)

Output 2: Excel files (one per protein) with all ligands + clean SMILEs

Output 3: Excel file with # ligands (active, inactive, intermediate) per protein.

How to modulate multiple targets with the same compound?

Output 1: Excel file with all multi-target ligands reported in ChEMBL (related with the input proteins)

Polypharmacology Interaction matrix

Selecting multiple Targets
(e.g. T1, T3, T4 & T8)

T1-ML model
T3-ML model
T4-ML model
T8-ML model

1. ADME/Tox filtering

2. Massive LBVS-ML of multiple commercial databases

MolPort

Enamine

ZINC

10^6 -> 10^3

3. Ensemble PBVS
10^3 -> 10^2

4. Docking + MMGBSA
10^2 -> 10^1

Best ~10 multi-target candidates

-> WetLab

Future perspectives

- What physicochemical and electronic characteristics do 2 or more targets have in common to be modulated by the same molecule?

↓

Common pharmacophores among key targets

15 key targets

PPI - Parkinson's disease

1167 targets

Key proteins = **65 targets**

Merge PPI-AD + PPI-PD

854 commun nodes (proteins) - 5619 edges

Ramirez Lab

Ramírez Lab

Ramírez Lab

About us

Research

Publications

Members

Resources

Join us

Events & News

Contact

Welcome to the Ramírez Lab

We are a multidisciplinary group focused on the study of biomolecular systems by using theoretical and experimental approaches. We aim to use computational polypharmacology with wet lab analyses to strengthen the drug design and development processes. Our work involves collaboration with medicinal chemists, biochemists, and biologists, and we are part of the Pharmacology Department - Faculty of Biological Sciences at University of Concepción.

Address
Universidad de Concepción
Facultad de Ciencias Biológicas
Departamento de Farmacología
Pharmacoinformatics & Drug Design Lab
Barrio Universitario s/n; Casilla 160 C – Correo 3
Concepción, Chile

Cel: (+56) 934211932
e-mail: dramirezs@udec.cl

Web lab: ***ramirezlab.github.io***

Research Assistant:

Carlos Peña
Ignacio Martinez
Paulina Valenzuela

PhD Students:

Estanislao Márquez
Kevin Cobos
Jessica Valero

Master students:

Alexis Sepulveda
Javiera Armijo
Alejandro Amaya (Col)

Undergraduate students:

Jordan Alegría
Victoria Oviedo
Alexandra Cigna
Tabata Barbosa (Col)
Laura Rodriguez (Col)
Claudia Martinez (Col)

Dr. Miguel Reyes-Parada
Centro de Investigación Biomédica y Aplicada
Universidad de Santiago de Chile

Dra. Ana Martínez
Química Médica y Biológica traslacional
CSIC-CIB - España

Dra. Janneth González
Computational and Structural Biology
Pontificia Universidad Javeriana - Colombia

Dr. Germán Combariza
Universidad Externado - Colombia

Dr. Camilo Ramirez
Politécnico Grancolombiano - Colombia

Agencia Nacional de Investigación y Desarrollo
Ministerio de Ciencia, Tecnología, Conocimiento e Innovación
Gobierno de Chile

Fondecyt 1220656
CONICYT-PCI REDES190074