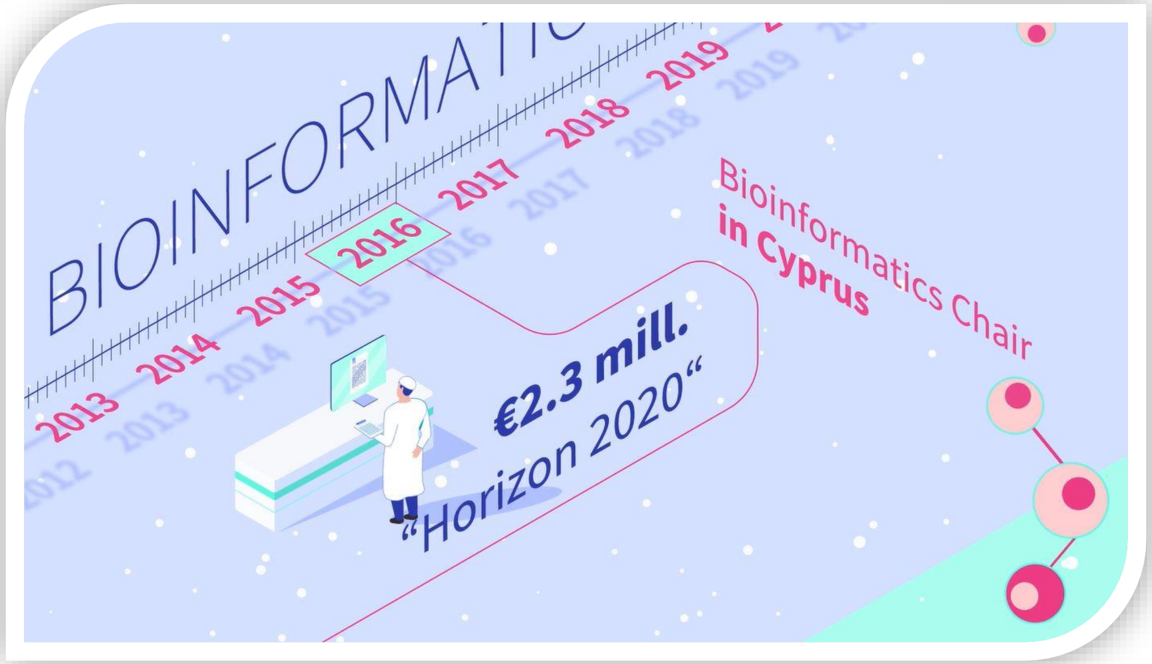
Bioinformatics Applications and Services at NI4OS from the Cyprus Institute of Neurology and Genetics

GEORGE M. SPYROU, PHD

Bioinformatics European Research Area Chair and Department Head,
Bioinformatics Department, The Cyprus Institute of Neurology & Genetics
Professor, The Cyprus School of Molecular Medicine

C-BIG

The **CING Bioinformatics Group (C-BIG)** is a highly inter-disciplinary team founded in ***2016*** to **host the Bioinformatics European Research Area Chair in Cyprus** under the support of the EU H2020 ERA Chair grant, BIORISE (Grant Number 669026) with an EU contribution of EUR 2 273 546 for the period ***2015-2020*** and since then it has been transformed to a **permanent Department within the Cyprus Institute of Neurology and Genetics (CING).**



Mission

C-BIG

- The mission of the Group is to function as a **hub of excellence** in the areas of **applied bioinformatics to early diagnosis, effective prognosis and drug discovery** contributing to the concepts of **Preventive, Personalized and Precise Medicine**.

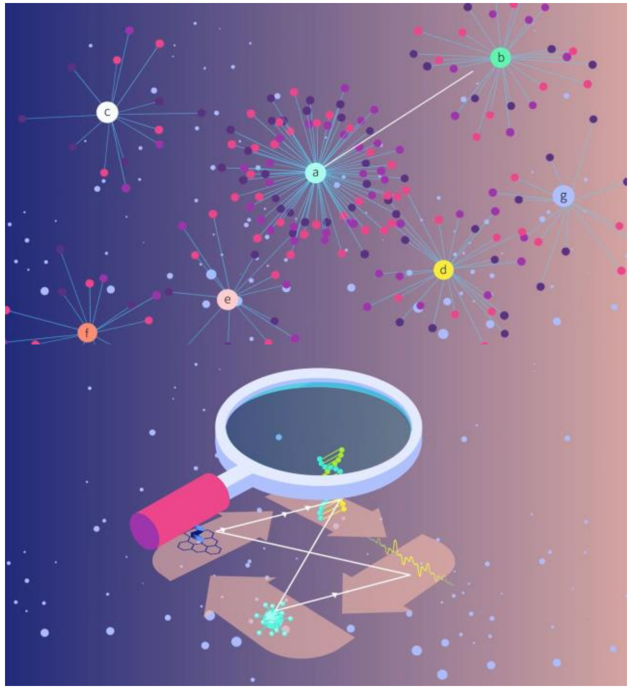
- This is accomplished through **state-of-the-art bioinformatics research**, **advanced education** in postgraduate level, **continuous development of bioinformatics methods and tools** provided either as services or as publicly available applications and participation to **international networks** leading/participating in **highly competitive research grant applications**.

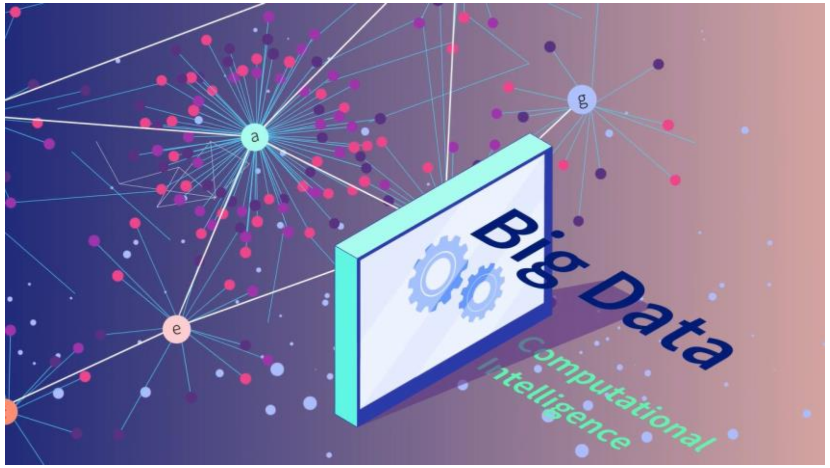
- Further to this, the mission of the Bioinformatics Chair within CING is to **make scientific and societal impact** by educating/informing students, patient groups and physicians about the potentialities of Bioinformatics to Human health and well-being.

Research

The research in C-BIG is focusing on **Network-based Computational Diagnostics and Therapeutics.**

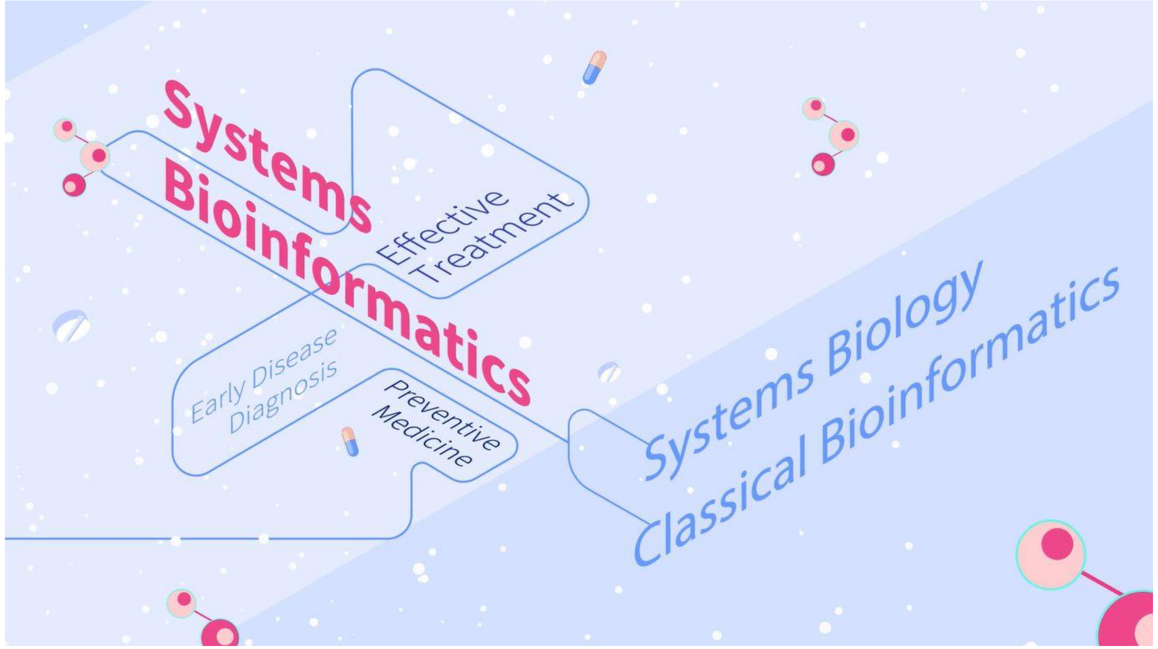
We are **analysing each layer of information separately** (the omics spectrum, signalling and imaging measurements, clinical and other patient's record data) and then we **perform multiomics / multisource integration methods** to **generate descriptive profiles**, to **highlight complex patterns of potential biomarkers**, **provide insights regarding the underlying molecular mechanisms** and **suggest candidate repurposed drugs**.

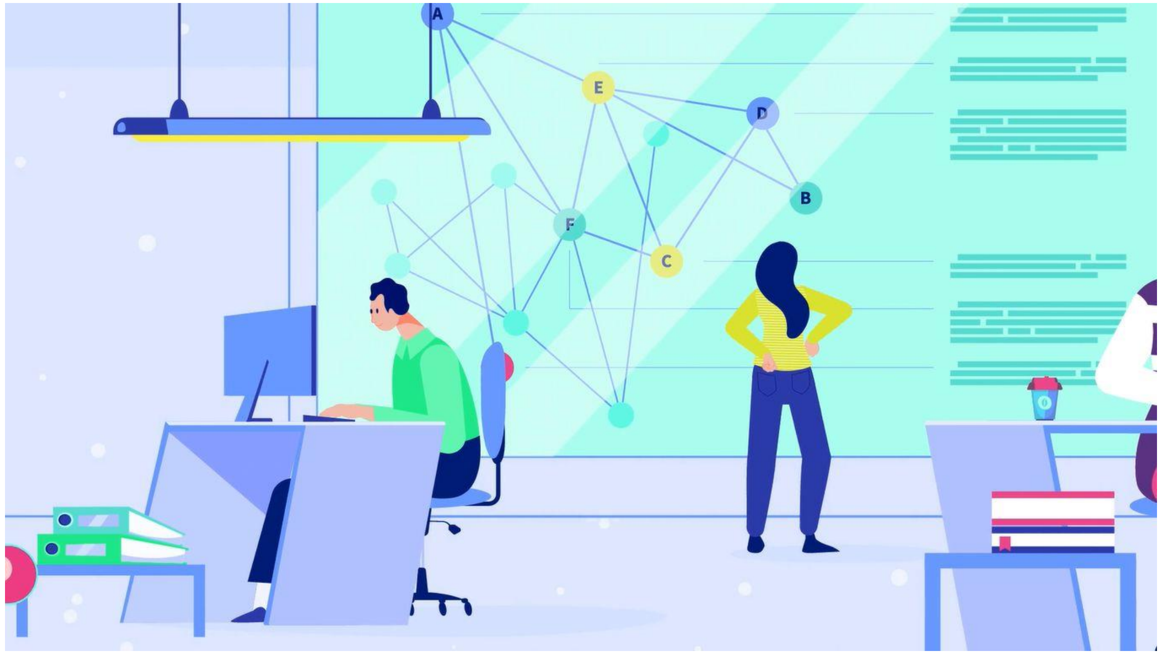




Research

C-BIG





We have **strong interest in network analytics** that **capture the systemic behavior of the biomolecules**, **to integrate multi-omics / multisource data**, to **generate comprehensive disease/patient profiles** that will act as the **stepping stone for AI-fueled computer aided diagnostics**, to **face complexities** and to **recognize criticalities** in biomedical problems, trying to **develop new mathematical/computational methods** to offer a solution if possible. This way we try to **contribute to the vision of personalized and precision medicine.**

Bioinformatics Analytics

C-BIG

- We are providing a number of Bioinformatics Analytics in terms of large-scale genomics (NGS) and proteomics analysis as well as post-omics analysis in the level of functional analysis and systems analysis as well. The already **developed pipelines** within CING are related to:

(1) Whole Genome/Exome Sequence Preprocessing and Analysis,

(2) RNA Seq (Bulk and Single-Cell) Preprocessing and Analysis,

(3) Microarray Data Preprocessing and Analysis,

(4) Proteomics Data Preprocessing and Analysis,

(5) Metabolomics Analysis,

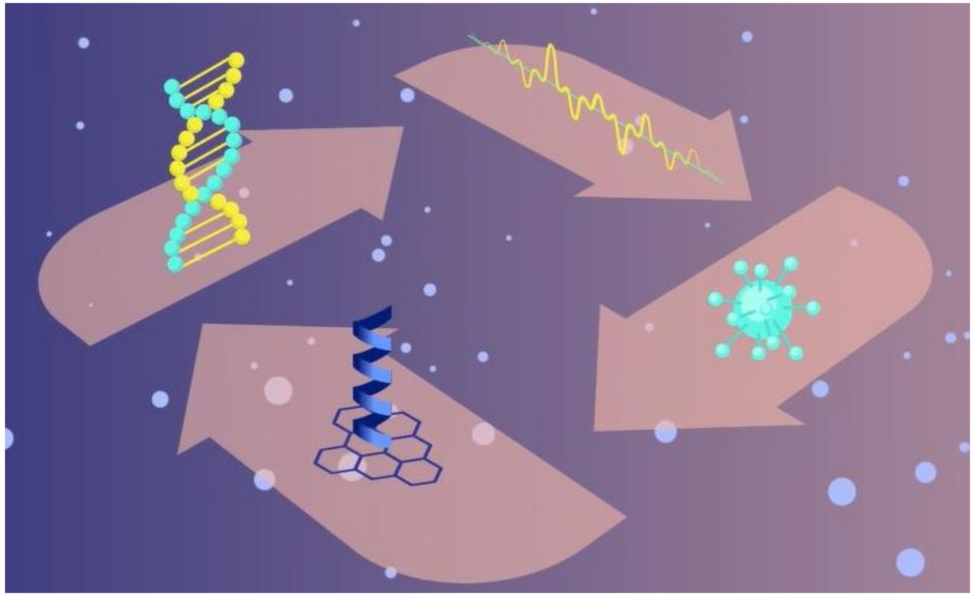
(6) Metagenomics Analysis,

(7) Structural Analysis –Cheminformatics

(8) Functional Analysis – Pathway Analysis,

(9) Network Representation and Analysis,

(10) Multi-omics / Multi-source Data Integration and visualization



Educating/training young scientists

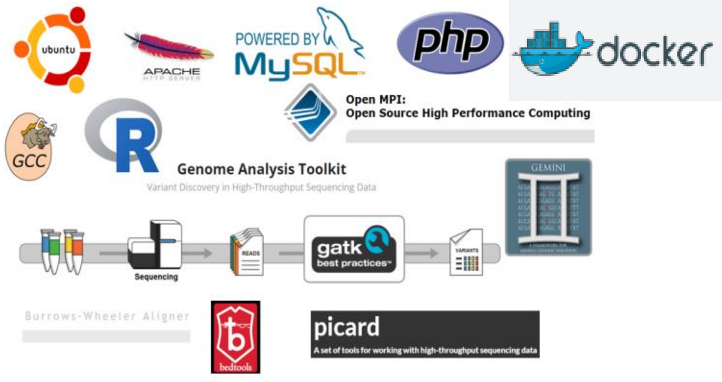
C-BIG

- PhD Supervision/co-Supervision: **3 completed PhDs**, **4 PhDs running**.
- **9 MSc students** from CSMM have already completed their theses within Bioinformatics Group and currently there are **other 3 MSc theses that are running**.
- **4 undergraduate students** have completed their diploma thesis under the co-supervision of the Bioinformatics Group.
- The Bioinformatics Group has attracted up to now **12 ERASMUS+ applications for hosting undergraduate, postgraduate students, and one professor**.
- **53 internship students** have been hosted and trained by the Bioinformatics Group, up to now.

Computational Infrastructure

C-BIG

Local Computing Infrastructure: There are six high performance C-BIG servers (2 Application Servers , 2 Database Servers, 2 High Performance Computing Servers) connected to a Storage System.



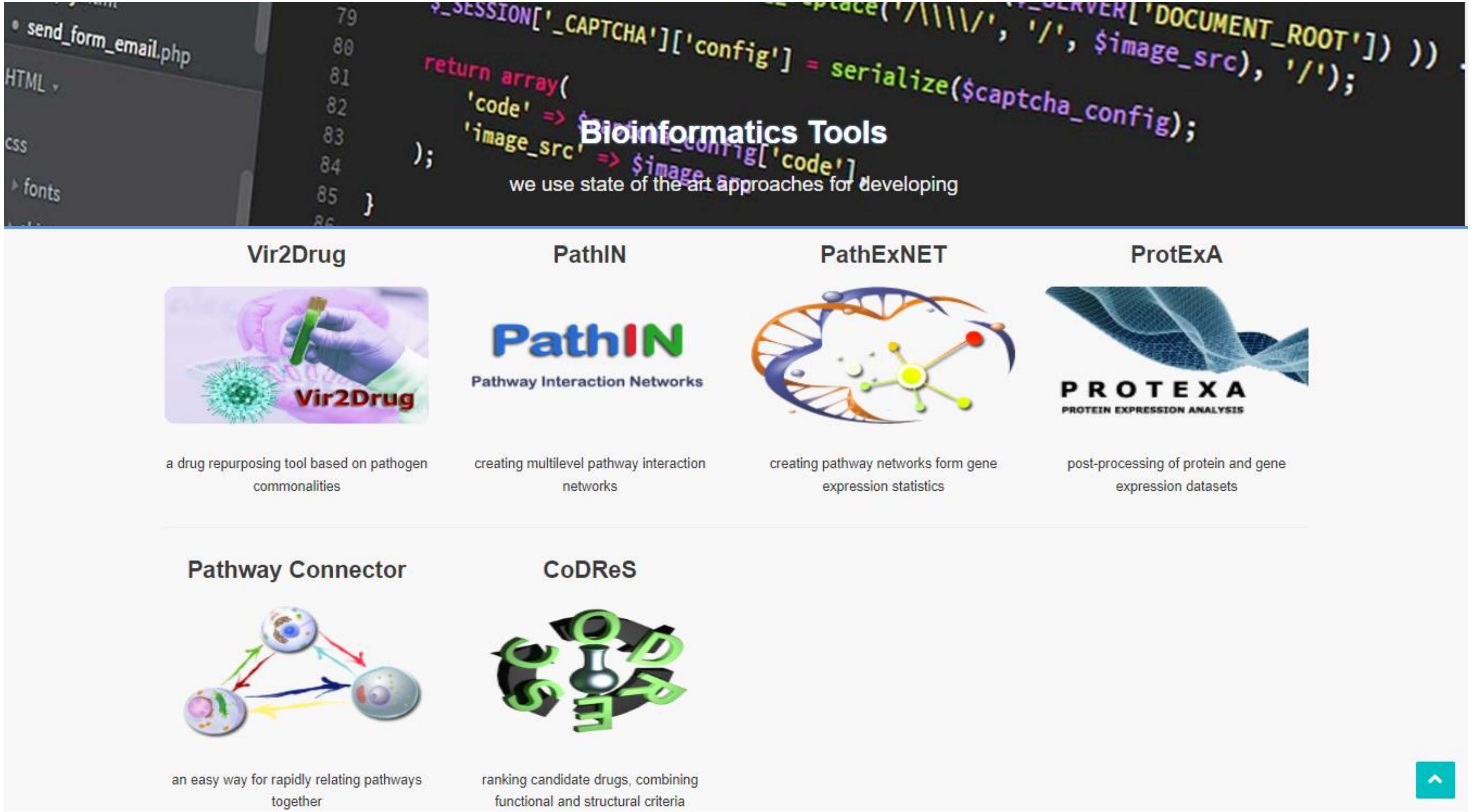
and many more.......

Collaboration with the National Supercomputing Facility for Cyprus:
The HPCF at CyI and the C-BIG are collaborating to provide a set of Bioinformatics tools and resources developed by C-BIG and powered by HPCF-CyI.

Participation in the NI4OS-Europe Catalogue:
Some of the most mature bioinformatics applications developed by C-BIG are indexed at NI4OS-Europe Catalogue

Also, C-BIG participates in the Steering Committee for setting up the **ELIXIR-Cyprus Node.**

Developing Publicly Available Tools



Bioinformatics Tools

we use state of the art approaches for developing

Vir2Drug

a drug repurposing tool based on pathogen commonalities

PathIN

creating multilevel pathway interaction networks

PathExNET

creating pathway networks form gene expression statistics

ProtExA

post-processing of protein and gene expression datasets

Pathway Connector

an easy way for rapidly relating pathways together

CoDReS

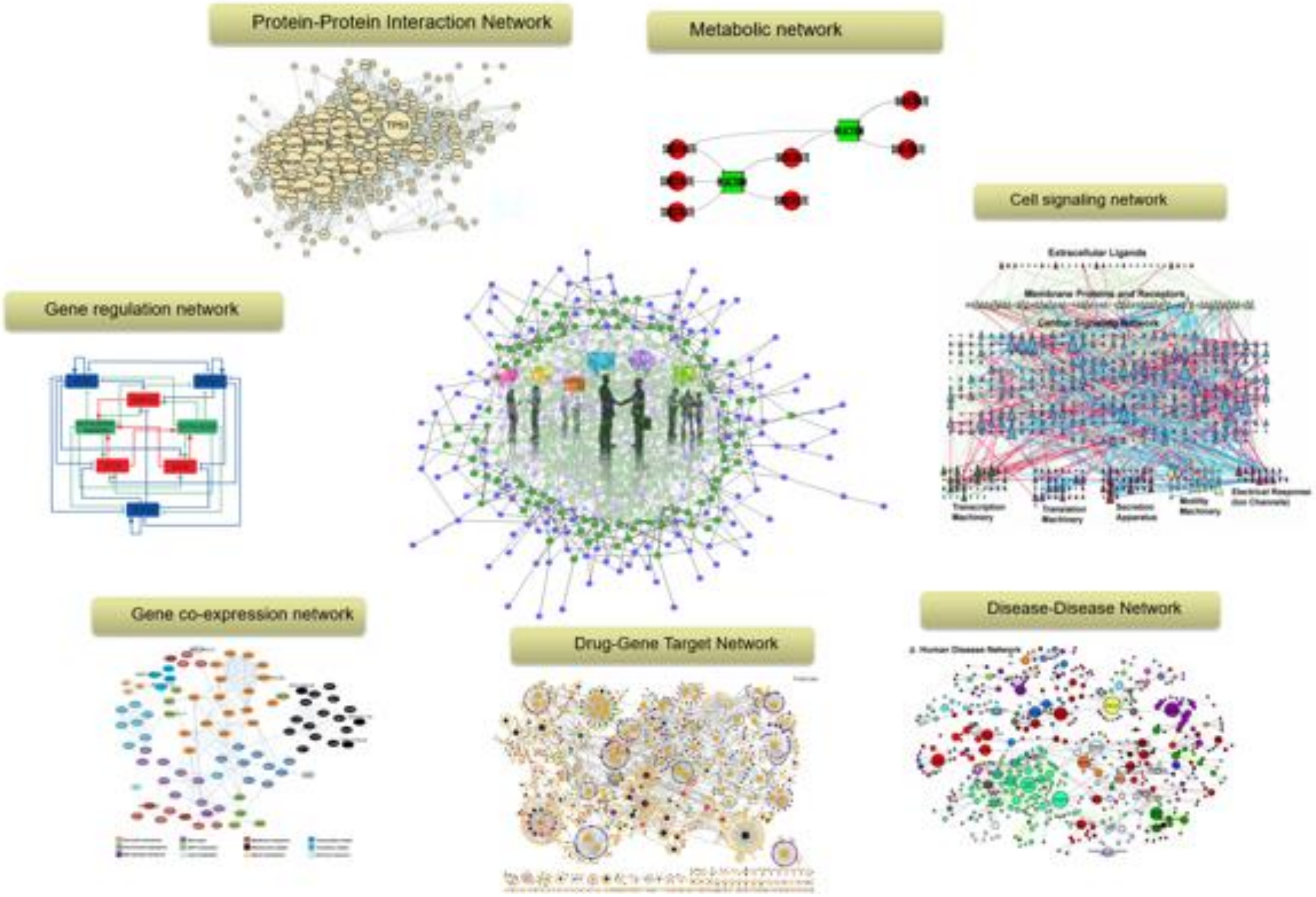
ranking candidate drugs, combining functional and structural criteria

Some basic info in a nutshell

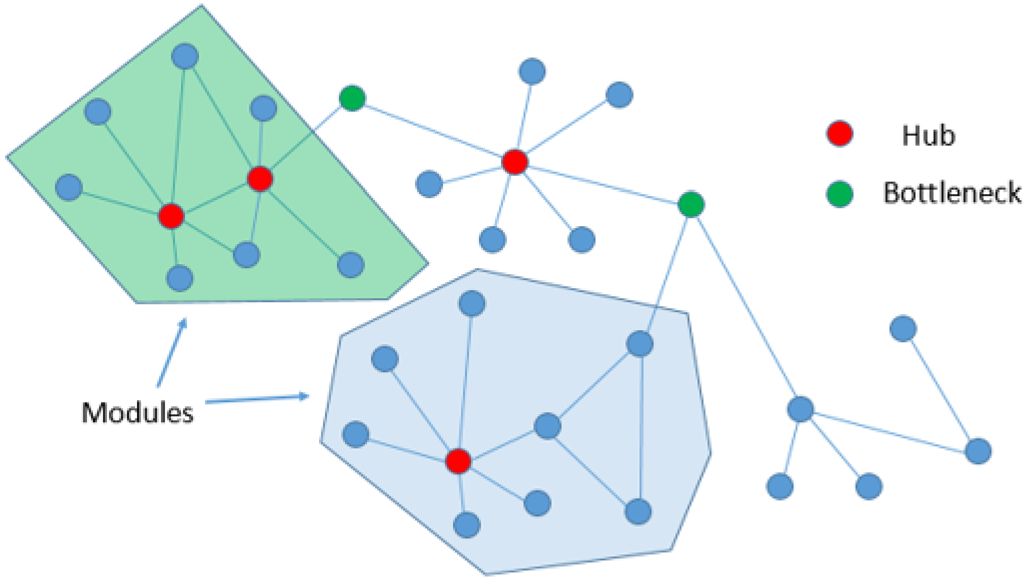
C-BIG

Networks

A visualization of relationships among a set of entities



How to analyze a network ?

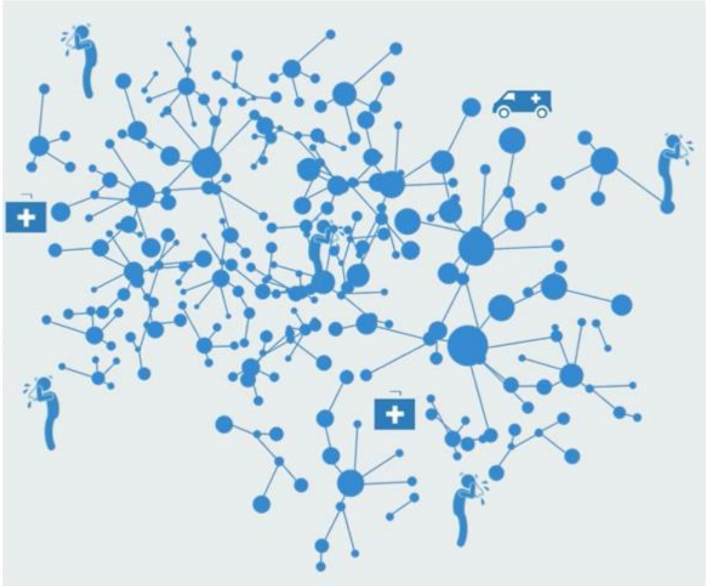


Some basic info in a nutshell

Networks

How to analyze a network ?

How susceptible is network to failure both from random and strategic attack?

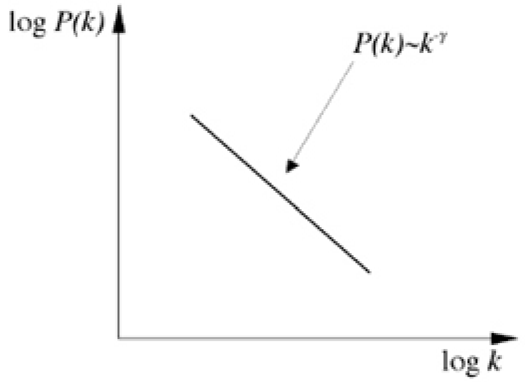


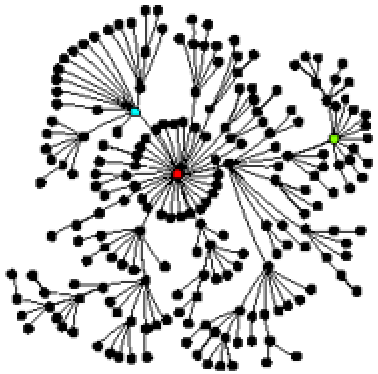
Disease-related perturbations ⇔ Attacks in the connectivity of the network

A **scale-free network** is a network whose degree distribution follows a power law, at least asymptotically.

The highest-degree nodes are often called "**hubs**", and are thought to serve specific purposes in their networks, although this depends greatly on the domain.

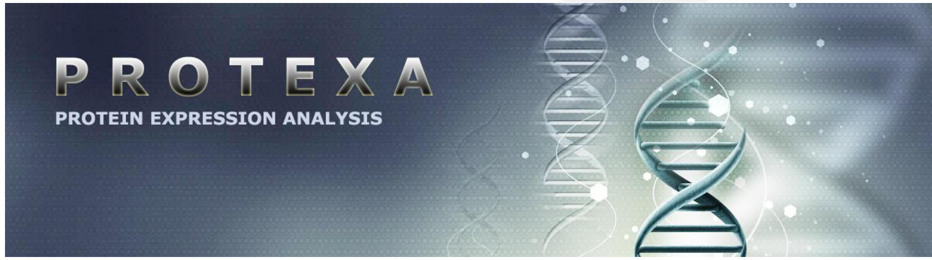
The scale-free property strongly correlates with the network's robustness to failure.





Selected pipelines in a nutshell

Differential Analysis & Network-construction



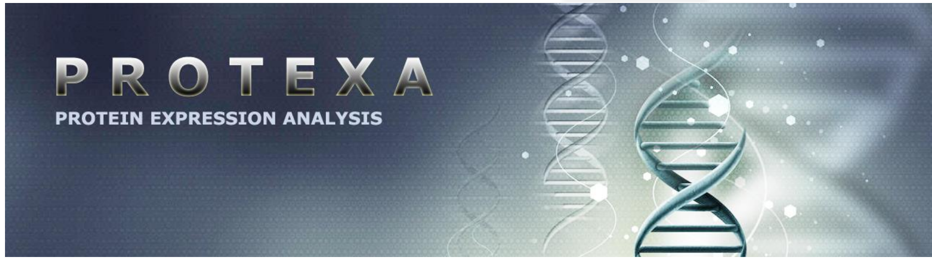
ProtExA is a web-tool for protein post-processing analysis from expression profiling experiments acting as an all-in-one tool for the statistical and functional analysis of protein expression datasets.

It facilitates gene/protein differential expression analysis and protein co-expression network analysis as well as pathway and post-pathway analysis through network-base bioinformatics approaches.

ProtExA combines four main pillars of omics data analysis, commonly used in bioinformatics pipelines, providing significant information about the functional relations between protein and genes.

Selected pipelines in a nutshell

Differential Analysis & Network-construction



- Differential Expression Analysis (DEA) and filtering to identify lists of top-scored proteins-genes
- Enrichment analysis to identify related pathways and their functional connectivity
- Creation of protein-protein co-expression networks
- Network clustering on co-expression networks to identify sub-networks that may related related to the biological condition under study

For a given set of protein expression data across samples, ProtExA: (1) performs statistical analysis and filtering to highlight the differentially expressed proteins, (2) generates protein co-expression networks using a variety of methodologies, (3) applies clustering methodologies to identify sub-networks of co-expressed proteins, (4) performs enrichment analysis to identify top-scored pathways and (5) generates pathway-to-pathway and pathway-to-gene networks.

Selected pipelines in a nutshell

Differential Analysis & Network-construction

PROTEXA

PROTEIN EXPRESSION ANALYSIS

✔ Below you will find the filtered results of **LIMMA** analysis performed on step 2. You can download this list by using the toolbar located on the top of this list.

Protein	Gene.symbol	mouse.protein	mouse.gene	FC	logFC	[abs.logFC]	[P.Value]	adj.P.Val	adj.P.BY	adj.P.FDR	adj.P.bonferroni	adj.P.hommel	adj.P.hochberg	adj.P.holm	t	B	
Q9HBG4	ATP6V0A4	Q920R6	Atp6v0a4	0.1248576563	-3.00164380500	3.00164380500	5.247724941e-03	0.0598156172901	0.551462074117	0.0598156172901	1.000000000000	0.999495787375	0.999495787375	1.000000000000	-4.238252670	-2.218435796152	3.
Q15465	SHH	Q62226	Shh	0.1351990578	-2.88684299700	2.88684299700	4.636178242e-06	0.0021886624784	0.020178080651	0.0021886624784	0.026263949741	0.026097047324	0.026212951780	0.026212951780	-15.082931220	5.006195340069	5.
P02652	APOA2	P09813	Apoa2	0.1626893701	-2.61980810500	2.61980810500	5.037123493e-07	0.0007133826147	0.006576935492	0.0007133826147	0.002853530459	0.002850508185	0.002852019322	0.002852019322	-21.831522660	6.785319491026	6.
Q9H0U9	TSPYL1	O88852	Tspyl1	0.2226845542	-2.16692660100	2.16692660100	5.473590724e-05	0.0077519728632	0.071468275800	0.0077519728632	0.310078914527	0.275321613428	0.307944214145	0.307944214145	-9.916446716	2.645013310137	4.
Q96CP2	FLYWCH2	Q9CQE9	Flywch2	0.2660516005	-1.91022201200	1.91022201200	4.787417659e-02	0.1861408444648	1.000000000000	0.1861408444648	1.000000000000	0.999495787375	0.999495787375	1.000000000000	-2.468767571	-4.577050001814	3.

Showing 1 to 5 of 26 entries

Figure Description: Heatmap showing in color scale the expression values of samples accross proteins. The red color refers to the number of over-expressed proteins while the green color to the under-expressed ones.

Figure Description: Barplot showing the number of over- and under-expressed proteins included in the filtered selected two-class dataset. The red color refers to the number of over-expressed proteins while the green color to the under-expressed ones.

Selected pipelines in a nutshell

Differential Analysis & Network-construction

PROTEXA
PROTEIN EXPRESSION ANALYSIS

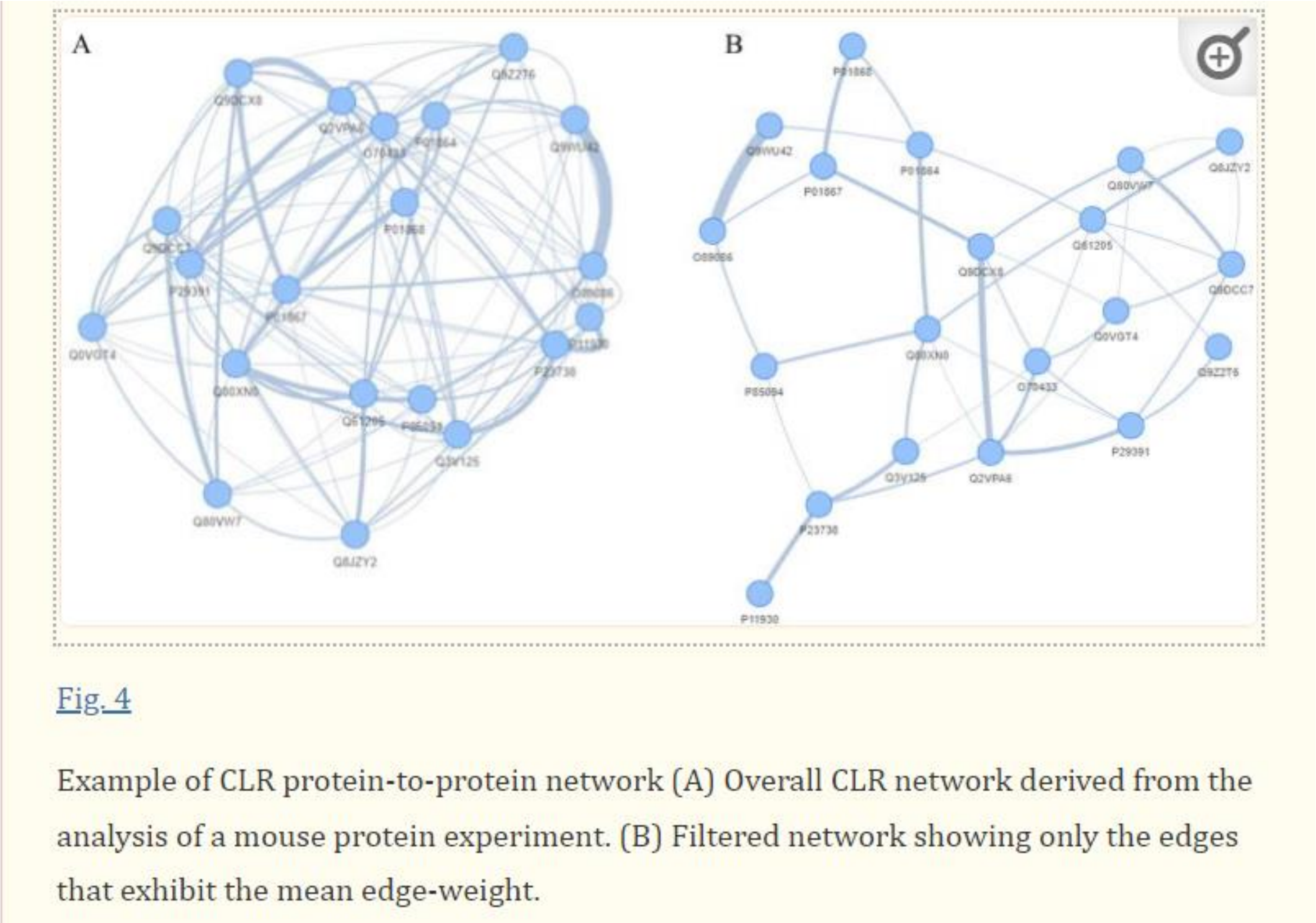


Fig. 4

Example of CLR protein-to-protein network (A) Overall CLR network derived from the analysis of a mouse protein experiment. (B) Filtered network showing only the edges that exhibit the mean edge-weight.

Selected pipelines in a nutshell

Pathway-expression Networks

PathExNET is a web service that allows the creation of pathway expression networks that hold the over- and under-expression information obtained from differential gene expression analyses.

PathExNET holds a large database of reference pathway-to-pathway networks, which have been developed through the freely available information included in the KEGG, Reactome and Wiki Pathways database repositories. Users can upload their differential gene expression statistical analysis, followed with pathways and/or genes of interest, and further chose a score methodology to create and explore the derived pathway-to-pathway expression networks.

In order to provide a concrete set of well-evaluated differential gene expression statistical analyses and to further increase the data-availability and easy data access of **PathExNET**, an additional tool has been rooted in **PathExNET** framework that allows to search and directly import, pre-processed statistic files from the Expression Atlas (EA) data repository of the European Bioinformatics Institute (EMBL-EBI).

Selected pipelines in a nutshell

Pathway-expression Networks

Depending on the input type **PathExNET** provides:

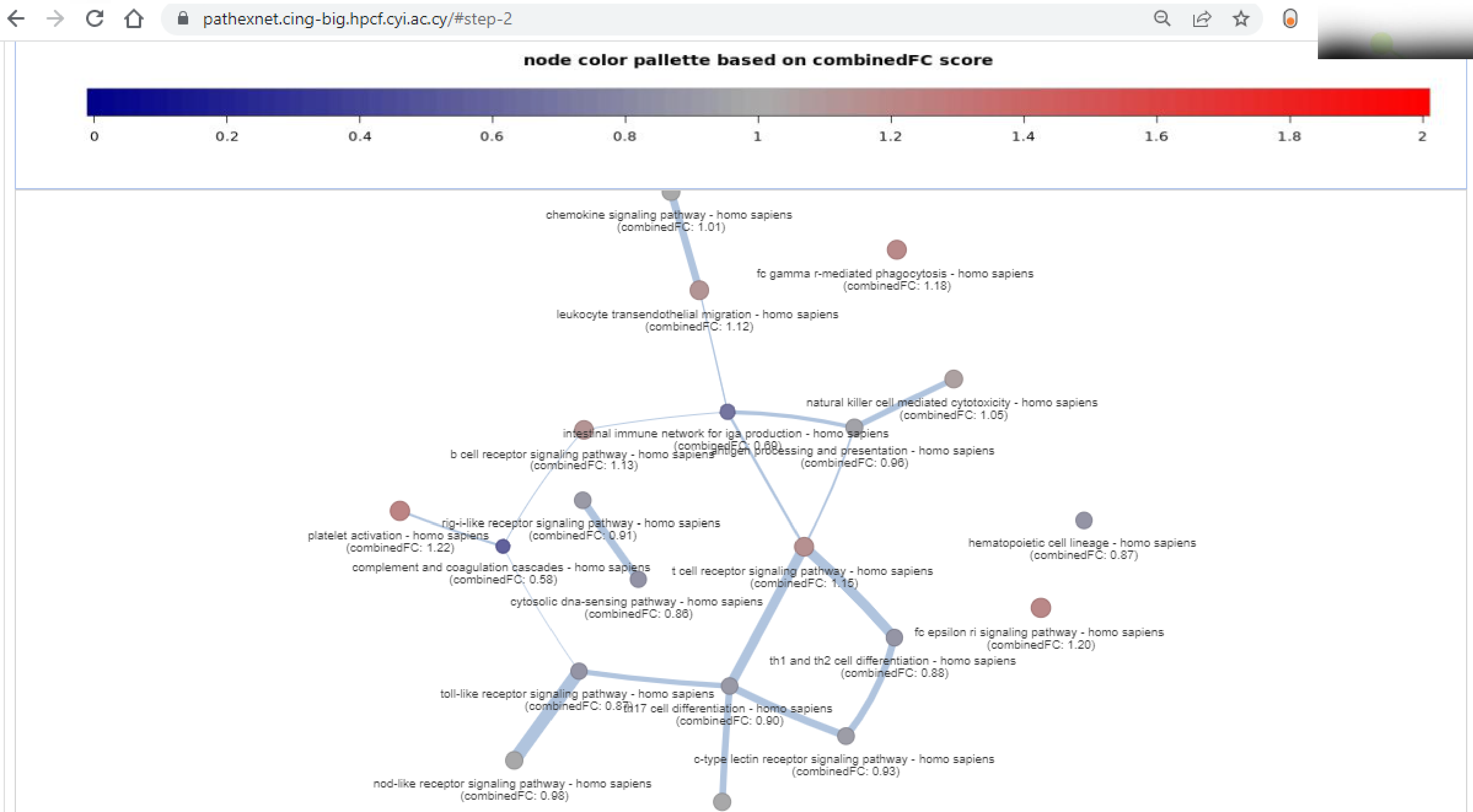
- pathway-to-pathway expression networks where the node size represents a specific score and the edge size represents the number of common genes between two pathways
- pathway-to-gene networks where the edges characterise the relation between pathway and gene.
- a series of pathway expression scores related to the over- and under-expression statistical information of genes
- a network based information on a single pathway and it's first neighbours
- further statistical information on User's input

Selected pipelines in a nutshell

C-BIG

Pathway-expression Networks

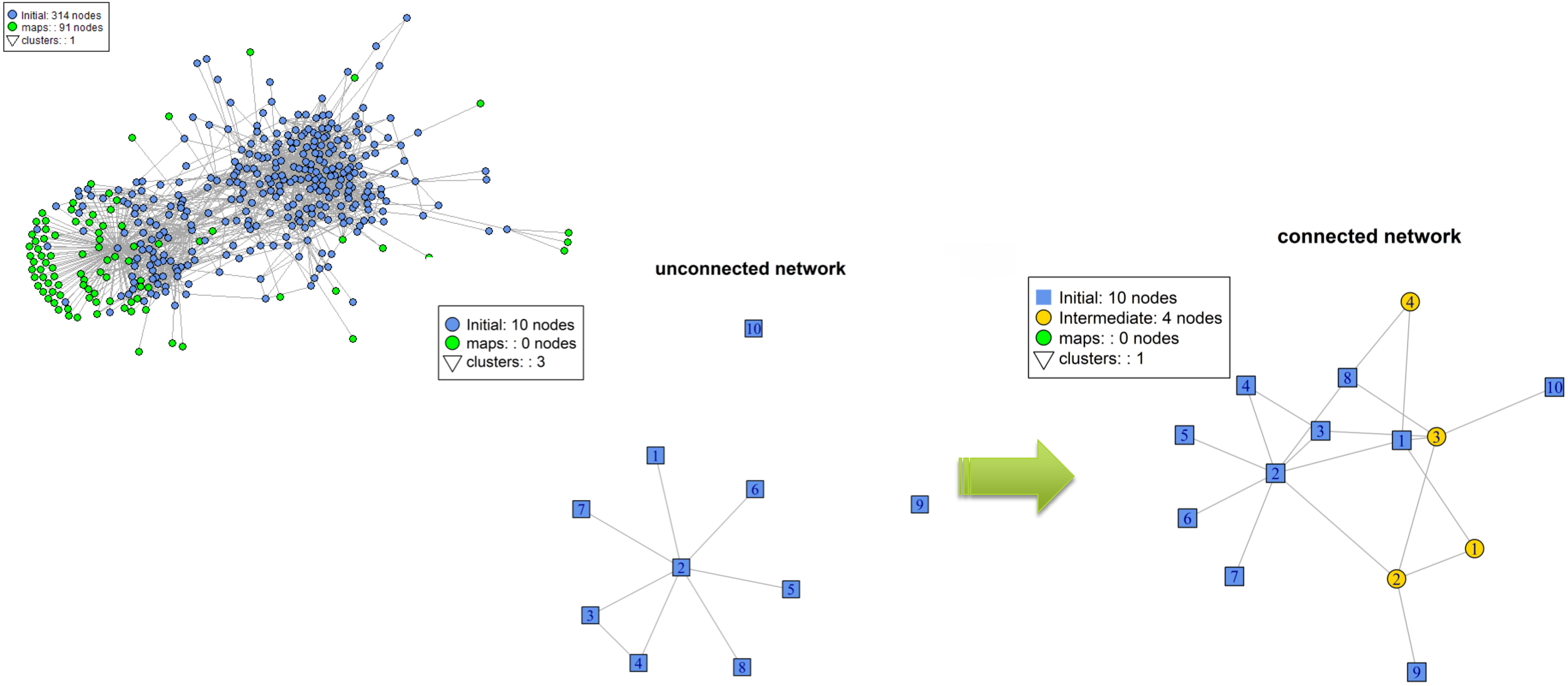
PathExNET
PATHWAY EXPRESSION NETWORK TOOL



Selected pipelines in a nutshell

C-BIG

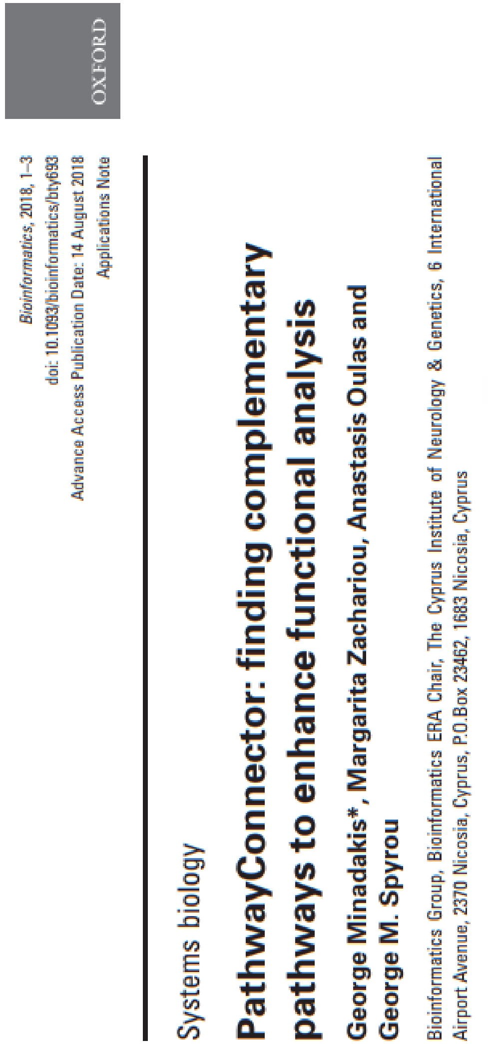
Network-based Inference



Selected pipelines in a nutshell

C-BIG

Network-based Inference



OXFORD

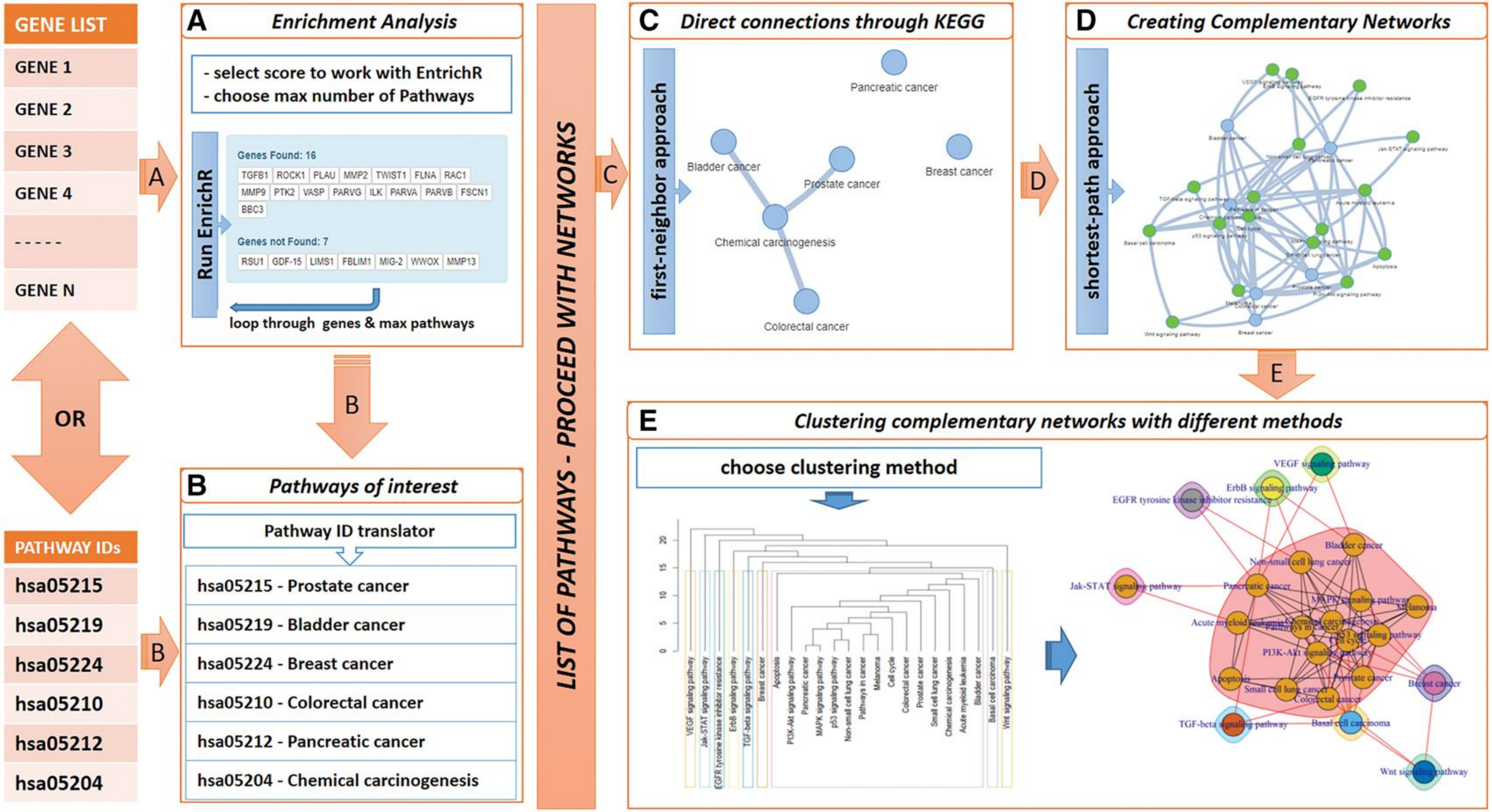
Bioinformatics, 2018, 1–3
doi: 10.1093/bioinformatics/bty693
Advance Access Publication Date: 14 August 2018
Applications Note

Systems biology

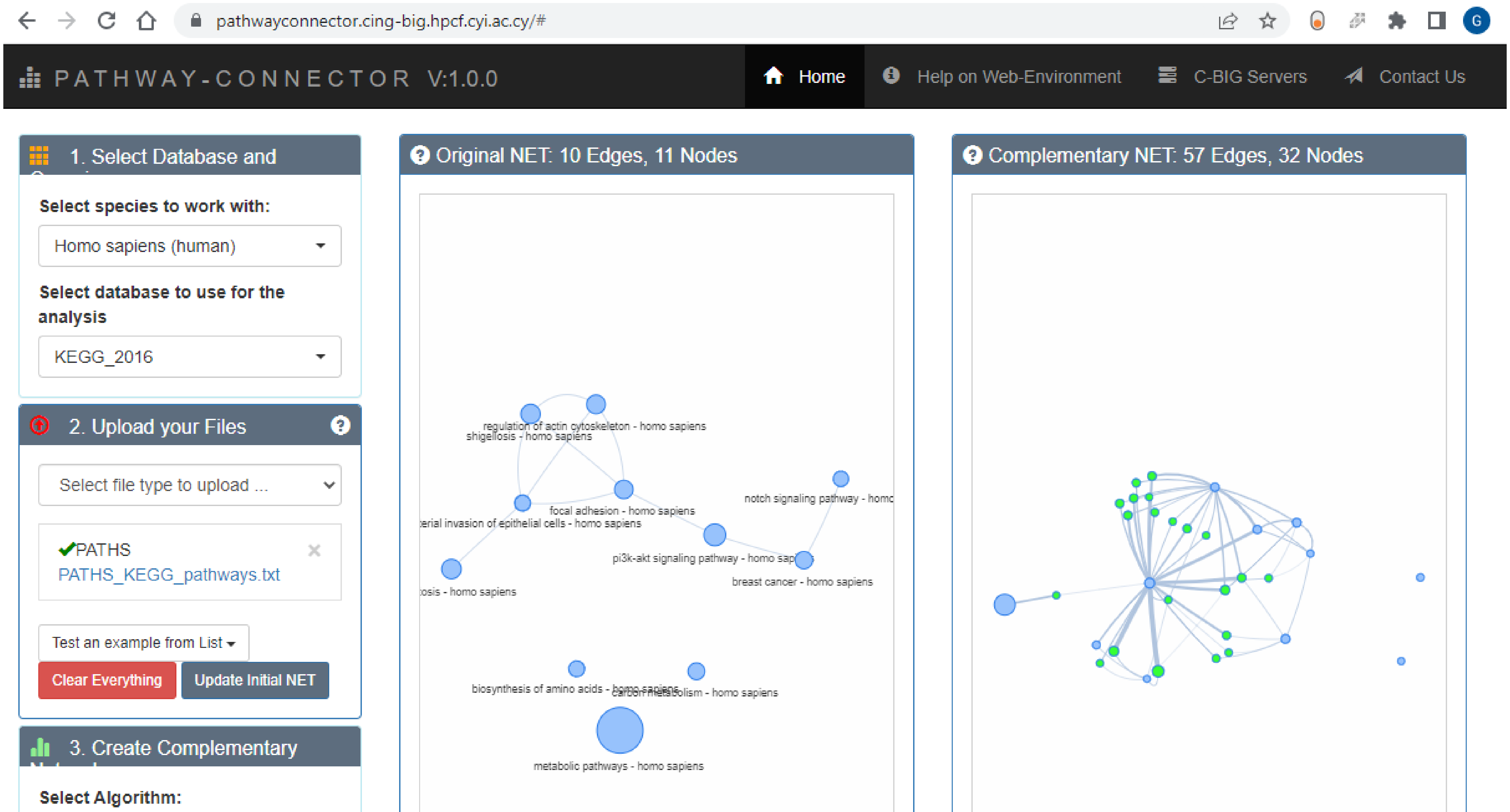
PathwayConnector: finding complementary pathways to enhance functional analysis

George Minadakis*, Margarita Zachariou, Anastasis Oulas and George M. Spyrou

Bioinformatics Group, Bioinformatics ERA Chair, The Cyprus Institute of Neurology & Genetics, 6 International Airport Avenue, 2370 Nicosia, Cyprus, P.O.Box 23462, 1683 Nicosia, Cyprus



Selected pipelines in a nutshell



pathwayconnector.cing-big.hpcf.cyi.ac.cy/#

PATHWAY-CONNECTOR V:1.0.0

1. Select Database and

Select species to work with: Homo sapiens (human)

Select database to use for the analysis: KEGG_2016

2. Upload your Files

✓PATHS
PATHS_KEGG_pathways.txt

3. Create Complementary

Select Algorithm:

Original NET: 10 Edges, 11 Nodes

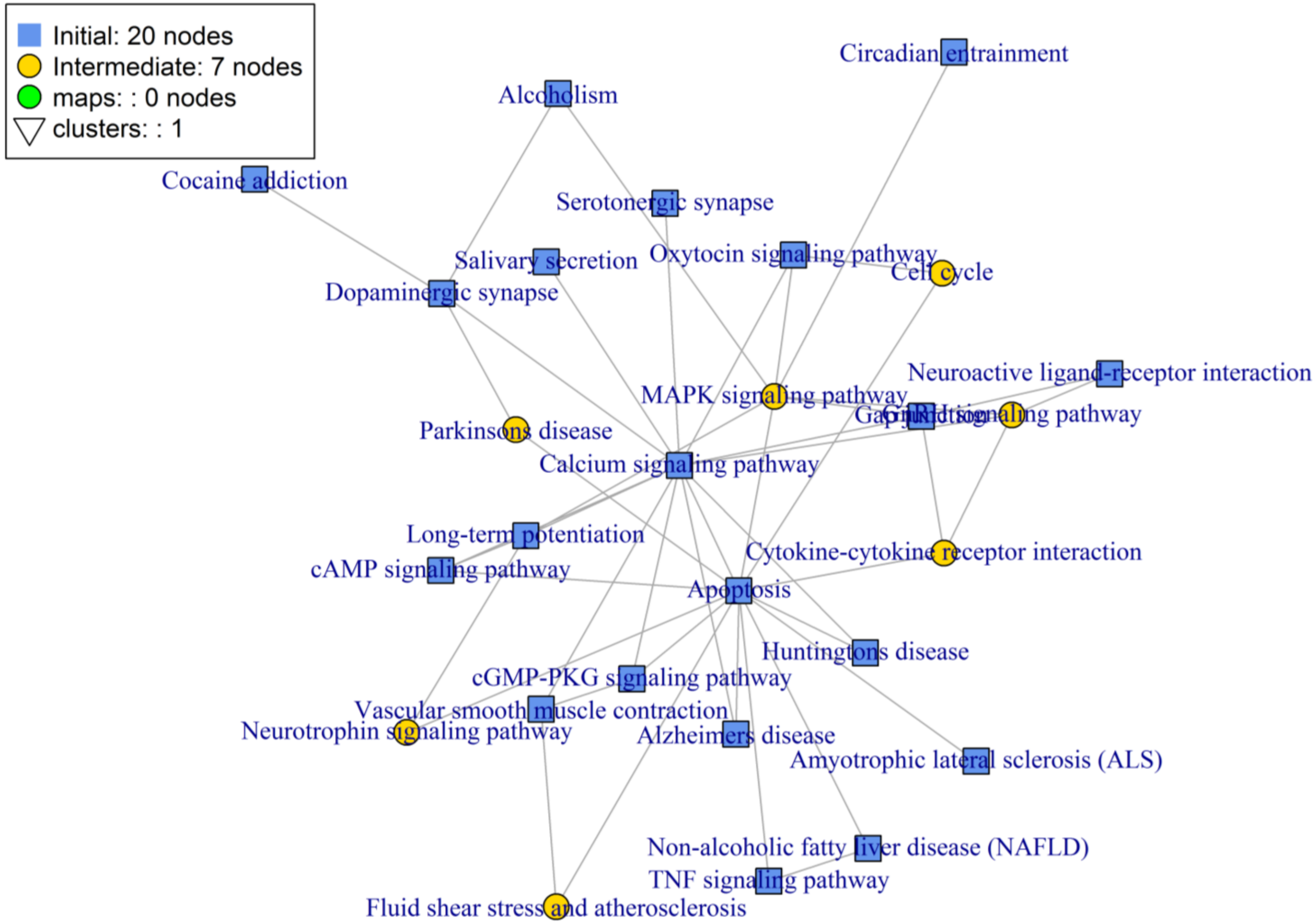
Complementary NET: 57 Edges, 32 Nodes

Selected pipelines in a nutshell

C-BIG

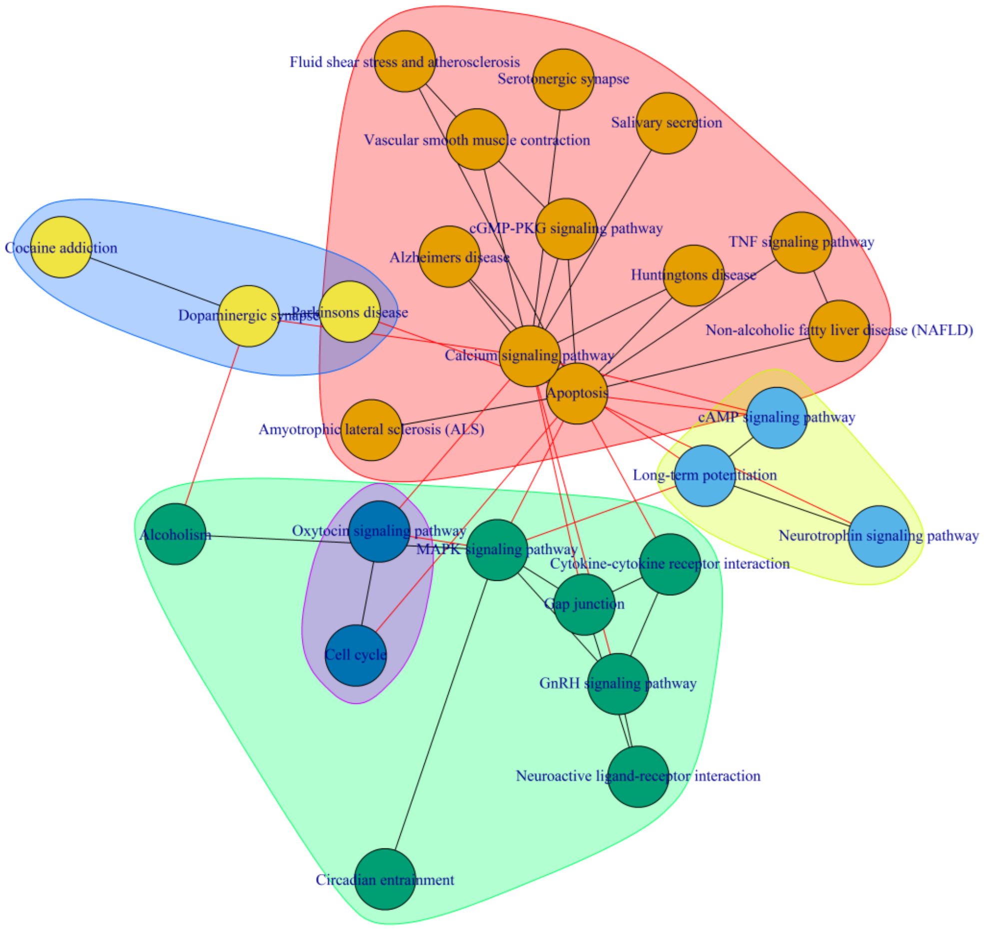
Connected AD-related pathways

connected network



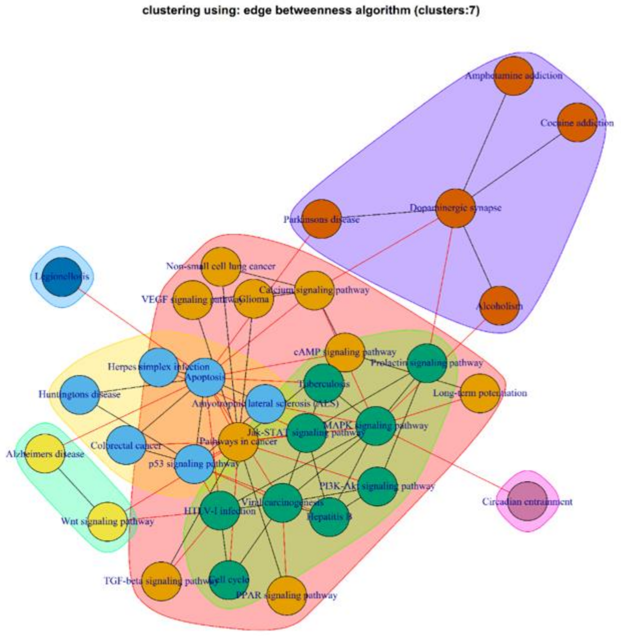
Clusters of pathways in AD

clustering using: edge betweenness algorithm (clusters:5)

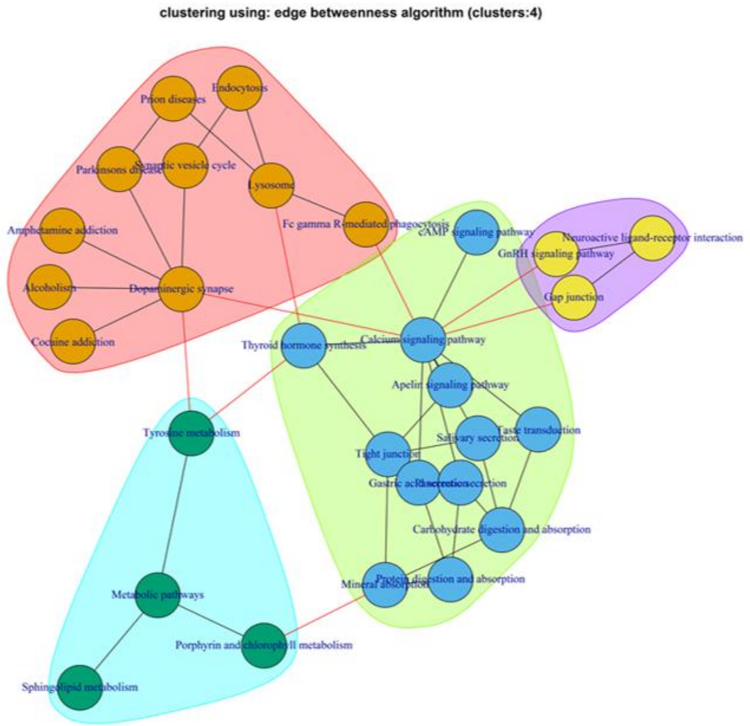


Selected pipelines in a nutshell

Clusters of pathways in HD



Clusters of pathways in GBA2-related diseases



IEEE J Biomed Health Inform. 2018 Aug 30. doi: 10.1109/JBHI.2018.2865569. [Epub ahead of print]

Revealing Clusters of Connected Pathways through Multisource Data Integration in Huntington's disease and Spastic Ataxia.

Kakouri A, Christodoulou CC, Zachariou M, Oulas A, Minadakis G, Demetriou CA, Votsi C, Zamba-Papanicolaou E, Christodoulou K, Spyrou G.

Selected pipelines in a nutshell

C-BIG

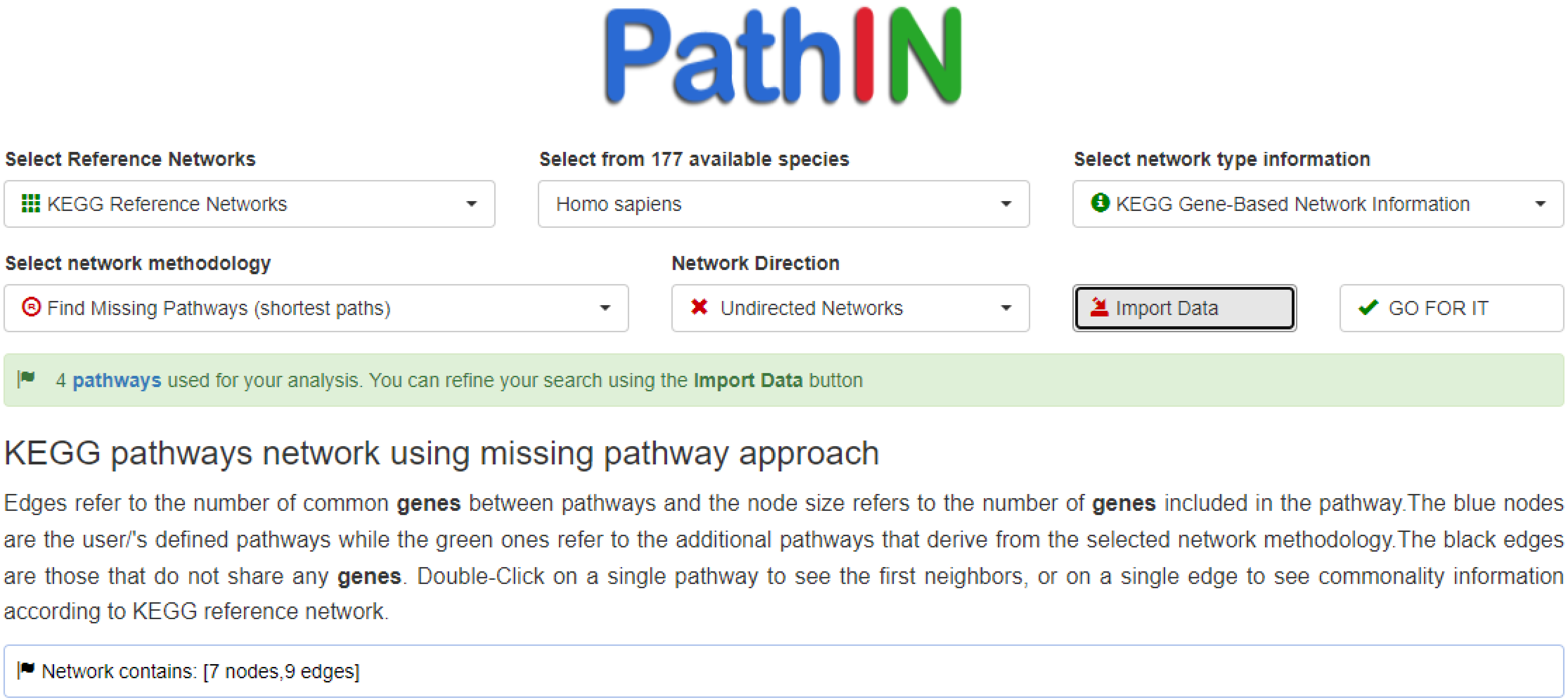
PathIN

PathIN is a web-service that provides an easy and flexible way for rapidly creating pathway based networks, at several functional biological levels: genes, compounds and reactions. The proposed tool holds a large database repository of reference pathway networks, across a large set of species, which have been developed through the freely available information included in the KEGG, Reactome, and Wiki Pathways database repositories.

PathIN provides networks by means of five diverse methodologies: (a) direct connections between pathways of interest, (b) direct connections as well as the first neighbours of the given pathways, (c) direct connections, the first neighbours and the connections in between them, and (d) two additional methodologies for creating complementary pathway-to-pathway networks that involve additional (missing) pathways that interfere in-between pathways of interest.

Selected pipelines in a nutshell

C-BIG



PathIN

Select Reference Networks
KEGG Reference Networks

Select from 177 available species
Homo sapiens

Select network type information
KEGG Gene-Based Network Information

Select network methodology
Find Missing Pathways (shortest paths)

Network Direction
✖ Undirected Networks

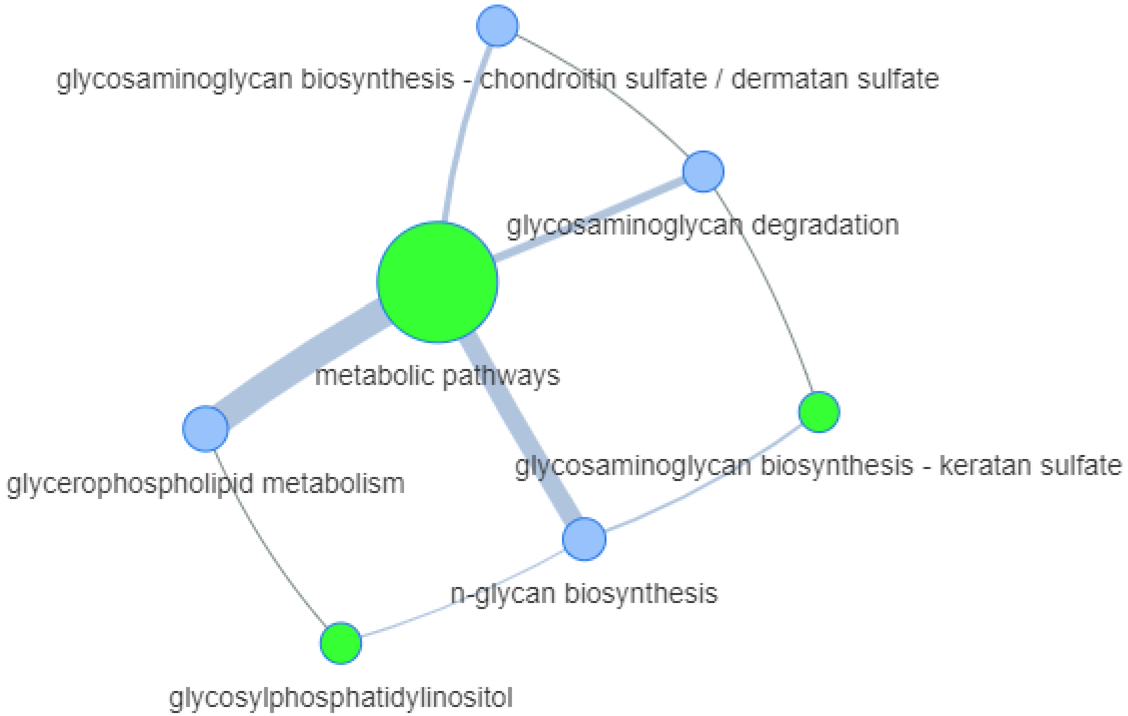
4 **pathways** used for your analysis. You can refine your search using the **Import Data** button

KEGG pathways network using missing pathway approach

Edges refer to the number of common **genes** between pathways and the node size refers to the number of **genes** included in the pathway.The blue nodes are the user/'s defined pathways while the green ones refer to the additional pathways that derive from the selected network methodology.The black edges are those that do not share any **genes**. Double-Click on a single pathway to see the first neighbors, or on a single edge to see commonality information according to KEGG reference network.

Network contains: [7 nodes,9 edges]

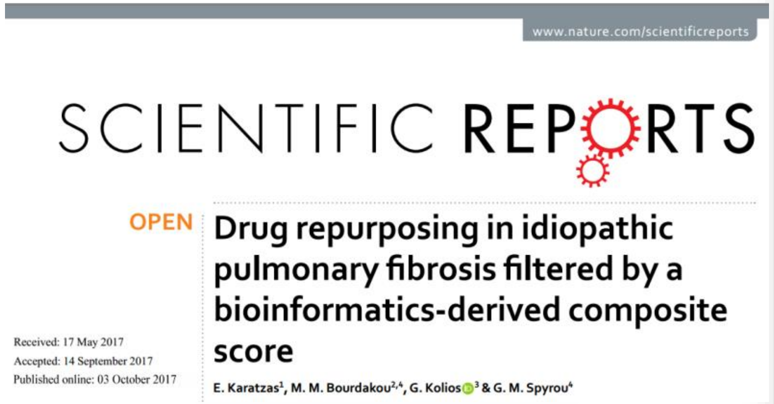
Selected pipelines in a nutshell



Selected pipelines in a nutshell

C-BIG

Drug repurposing and re-ranking


www.nature.com/scientificreports

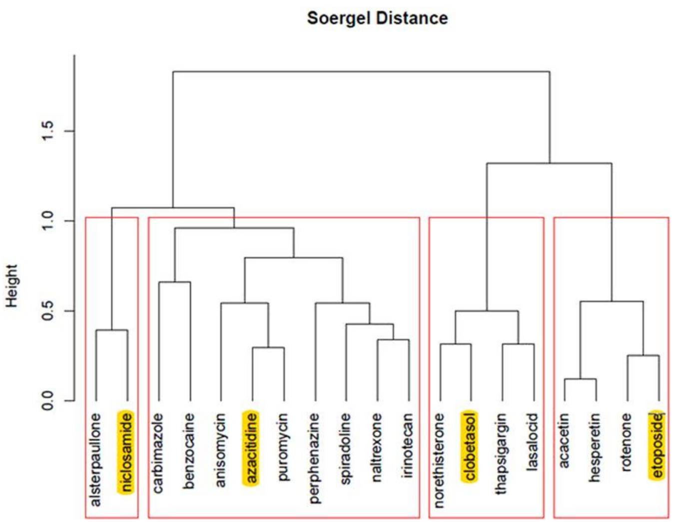
SCIENTIFIC REPORTS

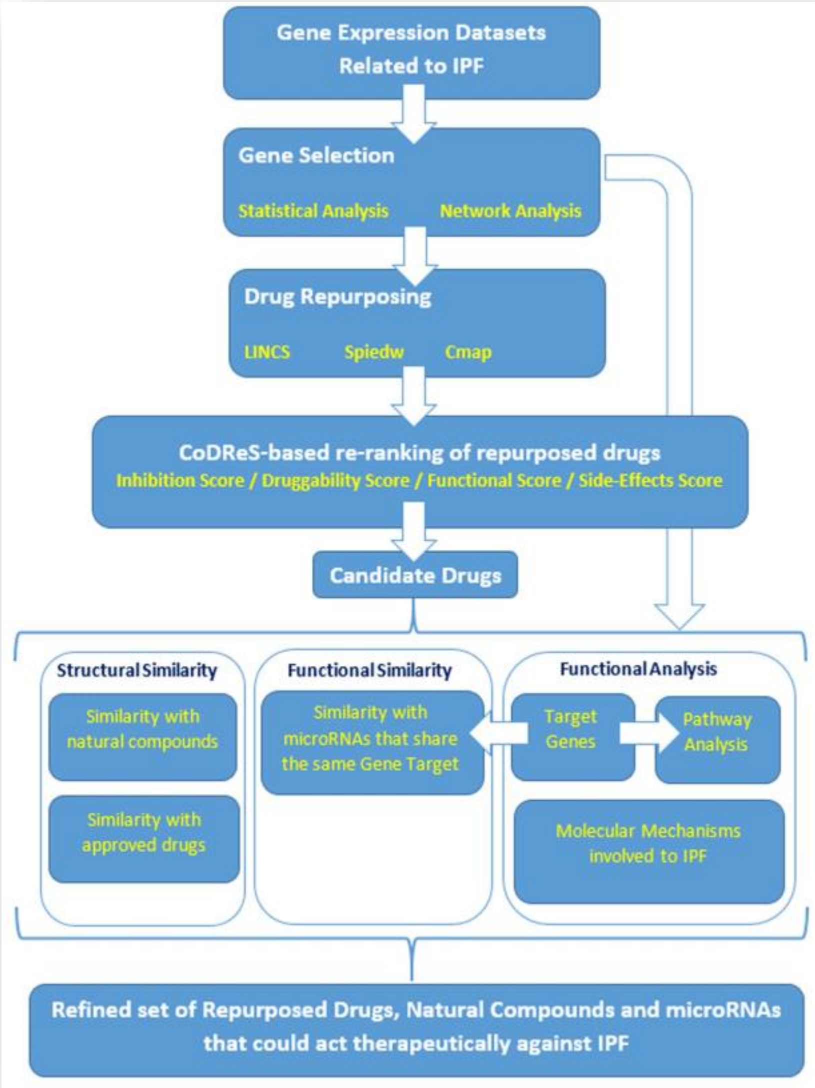
OPEN

Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score

Received: 17 May 2017
Accepted: 14 September 2017
Published online: 03 October 2017

E. Karatzas[1], M. M. Bourdakou[2,4], G. Kolios[3] & G. M. Spyrou[4]




Gene Expression Datasets Related to IPF

Gene Selection
Statistical Analysis
Network Analysis

Drug Repurposing
LINCS
Spiedw
Cmap

CoDReS-based re-ranking of repurposed drugs
Inhibition Score / Druggability Score / Functional Score / Side-Effects Score

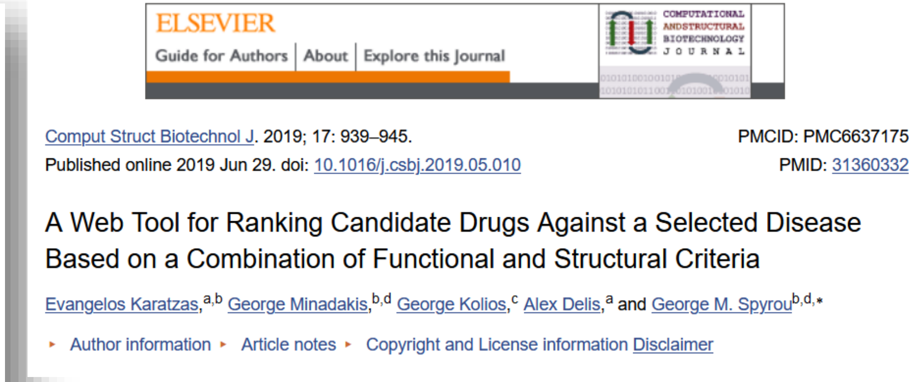
Candidate Drugs

Structural Similarity
- Similarity with natural compounds
- Similarity with approved drugs

Functional Similarity
- Similarity with microRNAs that share the same Gene Target

Functional Analysis
- Target Genes
- Pathway Analysis
- Molecular Mechanisms involved to IPF

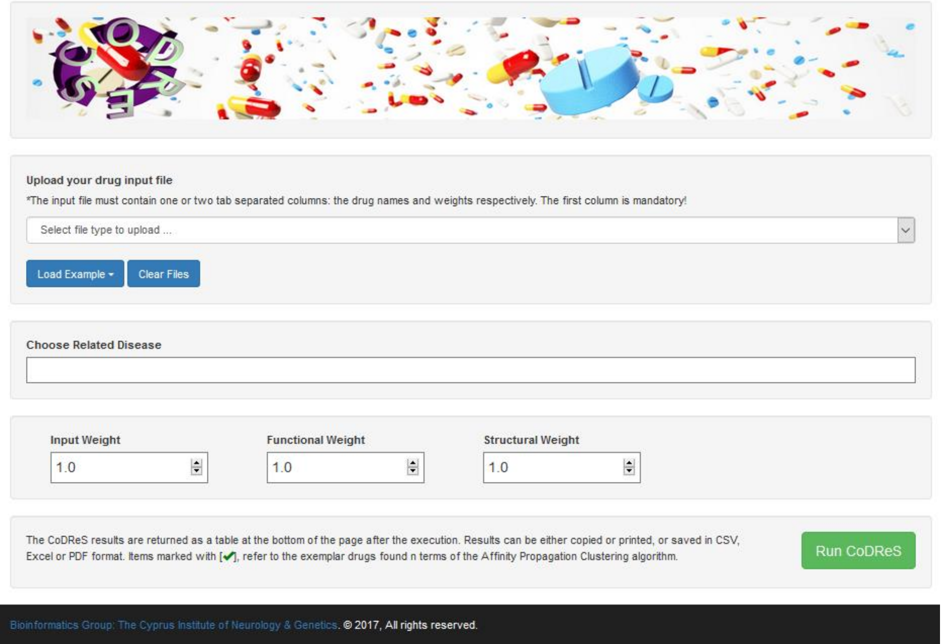
Refined set of Repurposed Drugs, Natural Compounds and microRNAs that could act therapeutically against IPF



Comput Struct Biotechnol J. 2019; 17: 939–945.
Published online 2019 Jun 29. doi: 10.1016/j.csbj.2019.05.010
PMCID: PMC6637175
PMID: 31360332

A Web Tool for Ranking Candidate Drugs Against a Selected Disease Based on a Combination of Functional and Structural Criteria

Evangelos Karatzas,[a,b] George Minadakis,[b,d] George Kolios,[c] Alex Delis,[a] and George M. Spyrou[b,d,*]


Upload your drug input file

*The input file must contain one or two tab separated columns: the drug names and weights respectively. The first column is mandatory!

Choose Related Disease

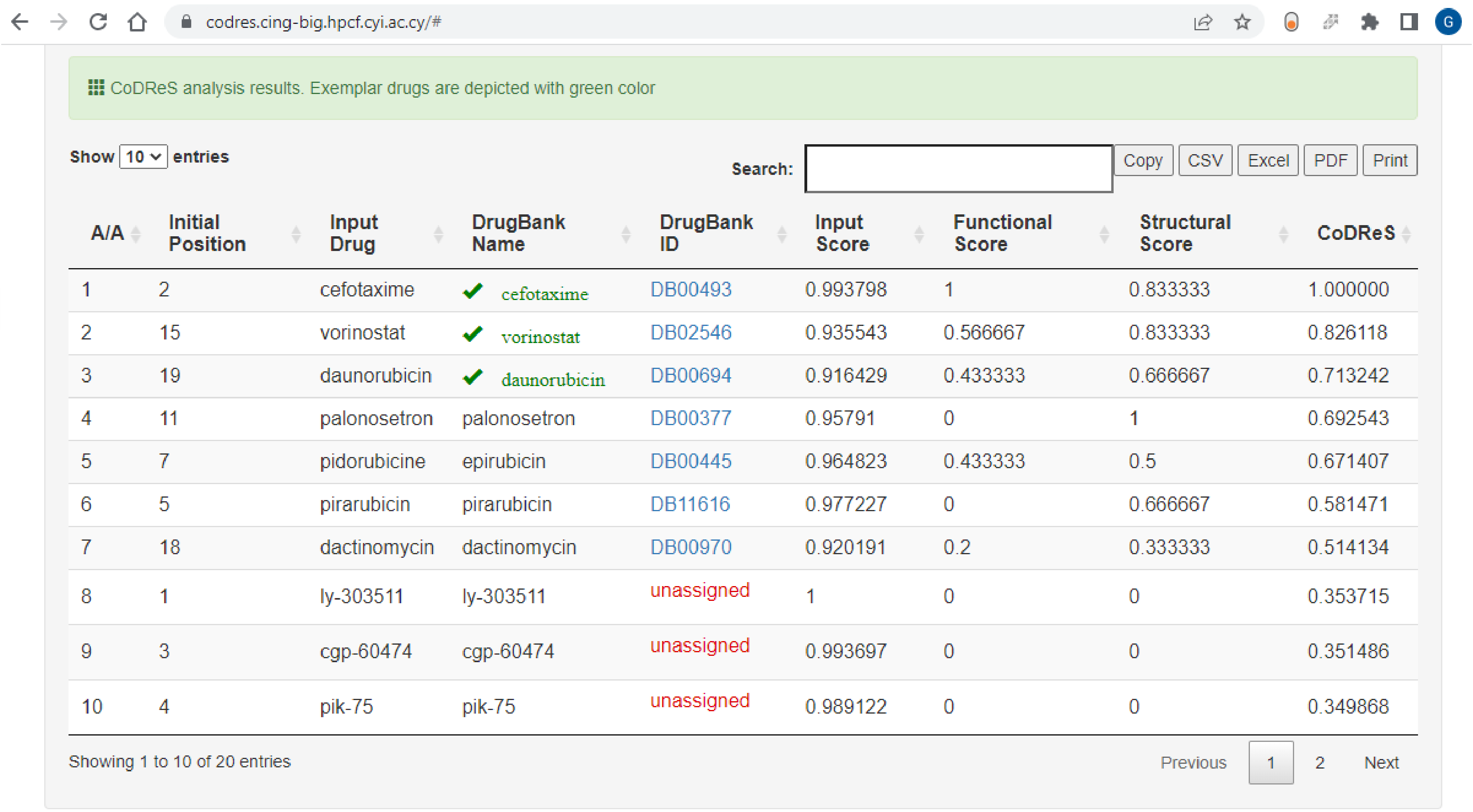
Input Weight: 1.0
Functional Weight: 1.0
Structural Weight: 1.0

The CoDReS results are returned as a table at the bottom of the page after the execution. Results can be either copied or printed, or saved in CSV, Excel or PDF format. Items marked with [✓], refer to the exemplar drugs found n terms of the Affinity Propagation Clustering algorithm.

Bioinformatics Group: The Cyprus Institute of Neurology & Genetics. © 2017, All rights reserved.

Selected pipelines in a nutshell

Drug repurposing and re-ranking



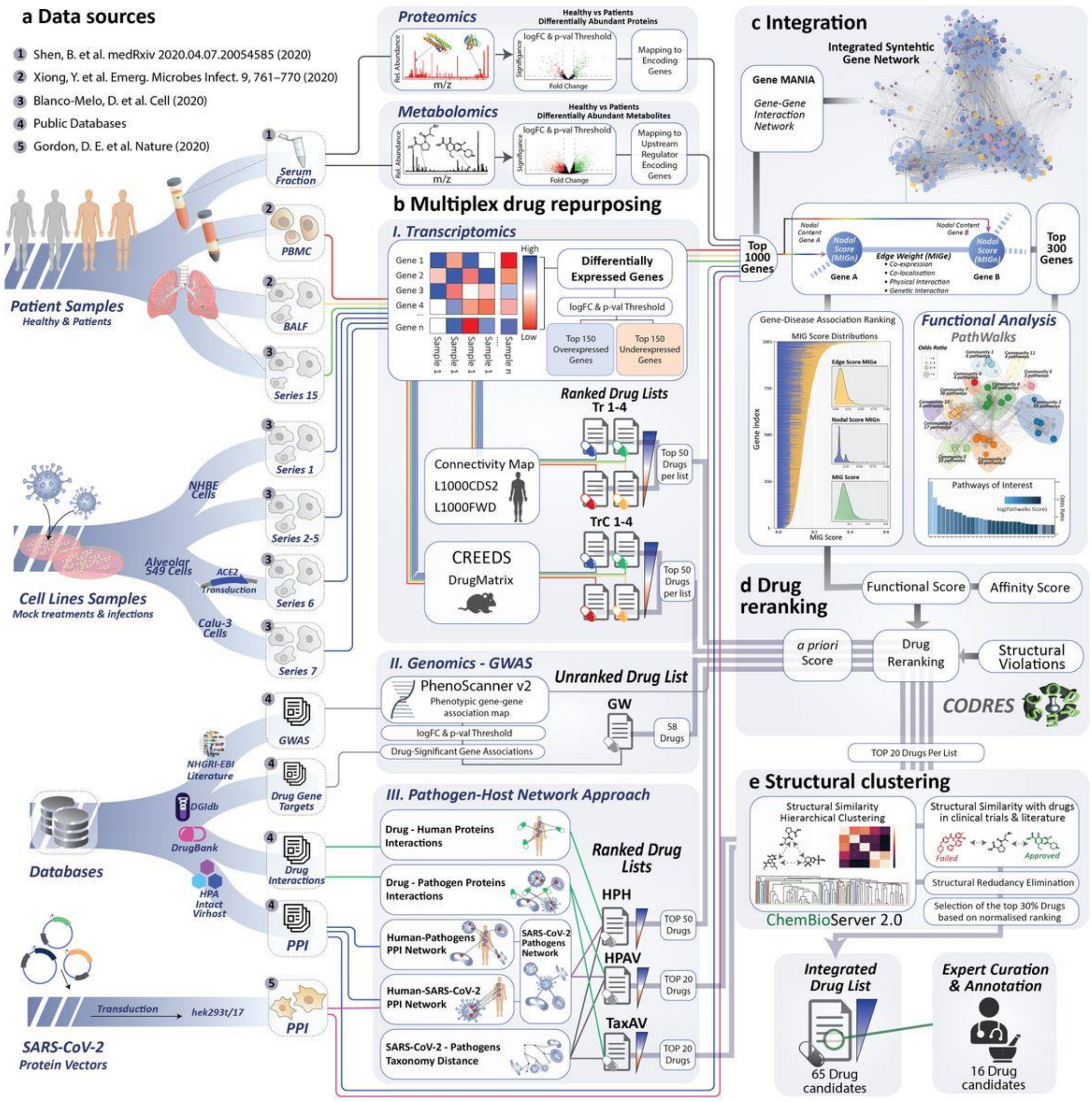
CoDReS analysis results. Exemplar drugs are depicted with green color

A/A	Initial Position	Input Drug	DrugBank Name	DrugBank ID	Input Score	Functional Score	Structural Score	CoDReS
1	2	cefotaxime	✓ cefotaxime	DB00493	0.993798	1	0.833333	1.000000
2	15	vorinostat	✓ vorinostat	DB02546	0.935543	0.566667	0.833333	0.826118
3	19	daunorubicin	✓ daunorubicin	DB00694	0.916429	0.433333	0.666667	0.713242
4	11	palonosetron	palonosetron	DB00377	0.95791	0	1	0.692543
5	7	pidorubicine	epirubicin	DB00445	0.964823	0.433333	0.5	0.671407
6	5	pirarubicin	pirarubicin	DB11616	0.977227	0	0.666667	0.581471
7	18	dactinomycin	dactinomycin	DB00970	0.920191	0.2	0.333333	0.514134
8	1	ly-303511	ly-303511	unassigned	1	0	0	0.353715
9	3	cgp-60474	cgp-60474	unassigned	0.993697	0	0	0.351486
10	4	pik-75	pik-75	unassigned	0.989122	0	0	0.349868

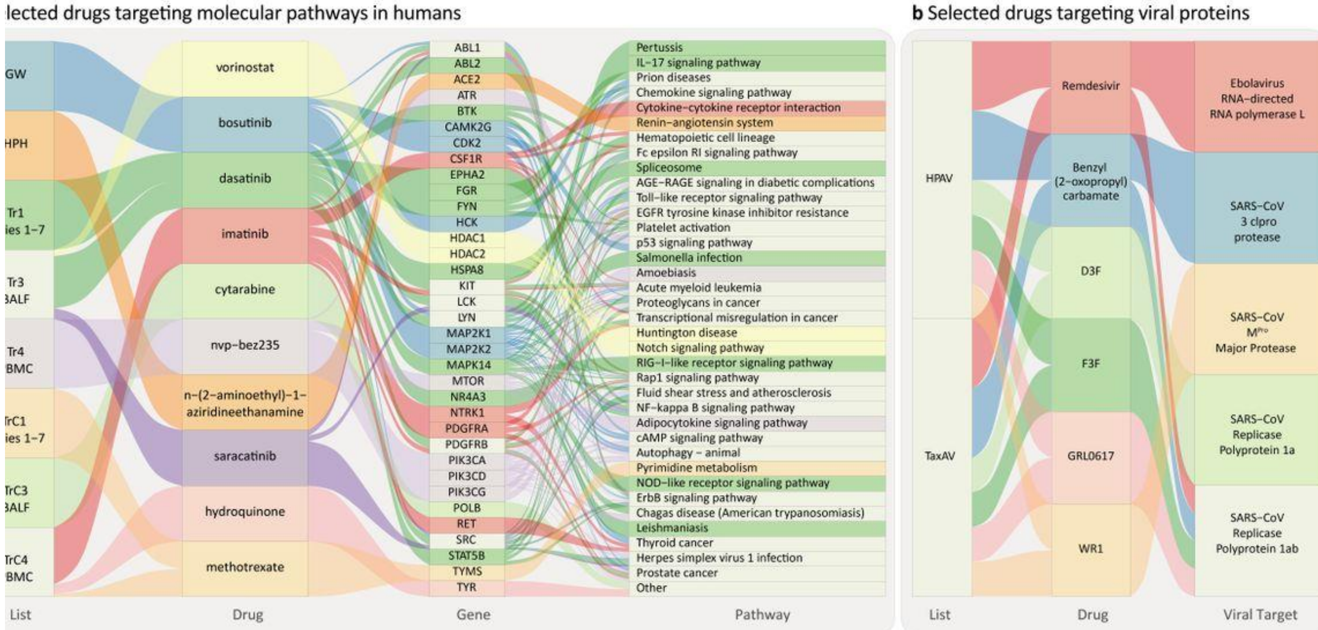
Showing 1 to 10 of 20 entries

Selected pipelines in a nutshell

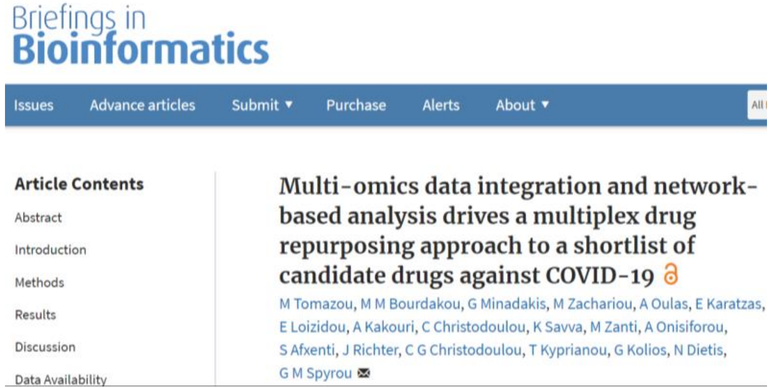
C-BIG



Multiplex drug repurposing against COVID-19



NI4OS
National Initiatives for Open Science in Europe



Article Contents

Abstract

Introduction

Methods

Results

Discussion

Data Availability

Multi-omics data integration and network-based analysis drives a multiplex drug repurposing approach to a shortlist of candidate drugs against COVID-19

M Tomazou, M M Bourdakou, G Minadakis, M Zachariou, A Oulas, E Karatzas, E Loizidou, A Kakouri, C Christodoulou, K Savva, M Zanti, A Onisiforou, S Afxenti, J Richter, C G Christodoulou, T Kyprianou, G Kolios, N Dietis, G M Spyrou

Selected pipelines in a nutshell

We take action in the fight against COVID-19

NI4OS-Europe opens a fast track access channel to its services, tools and software for the Scientific communities that perform extensive research to tackle the COVID-19.

Computational resources have already been allocated to the Bioinformatics European Research Era Chair and the Bioinformatics Group at the Cyprus Institute of Neurology and Genetics. The research team has put their efforts in the multi-omic analysis and network-based integration towards a highly-informed decision regarding a short list of repurposed drugs and related to them natural products against COVID-19.

Thank you!

11001110 10100011
11001110 10110001
11001111 10000010
00100000 11001110
10110101 11001111
10000101 11001111
10000111 11001110
10110001 11001111
10000001 11001110
10111001 11001111
10000011 11001111
10000100 11001111
10001110 00100001
00001101 00001010

TAGCGTACTCTCTCGAGCGACTGTGAT