Interactome and gene expression based analyzes for translational medicine

Emre Güney, PhD

Hospital del Mar Research Institute (IMIM)
& Pompeu Fabra University(UPF)

MSc on Omics Data Analysis - Bioinformatics Applications

January 17th, 2018

Facebook friends

$\sim 85\%$

Percentage of your Facebook friends from the same country

(Ugander et al., 2011, arXiv:1111.4503)

~85%

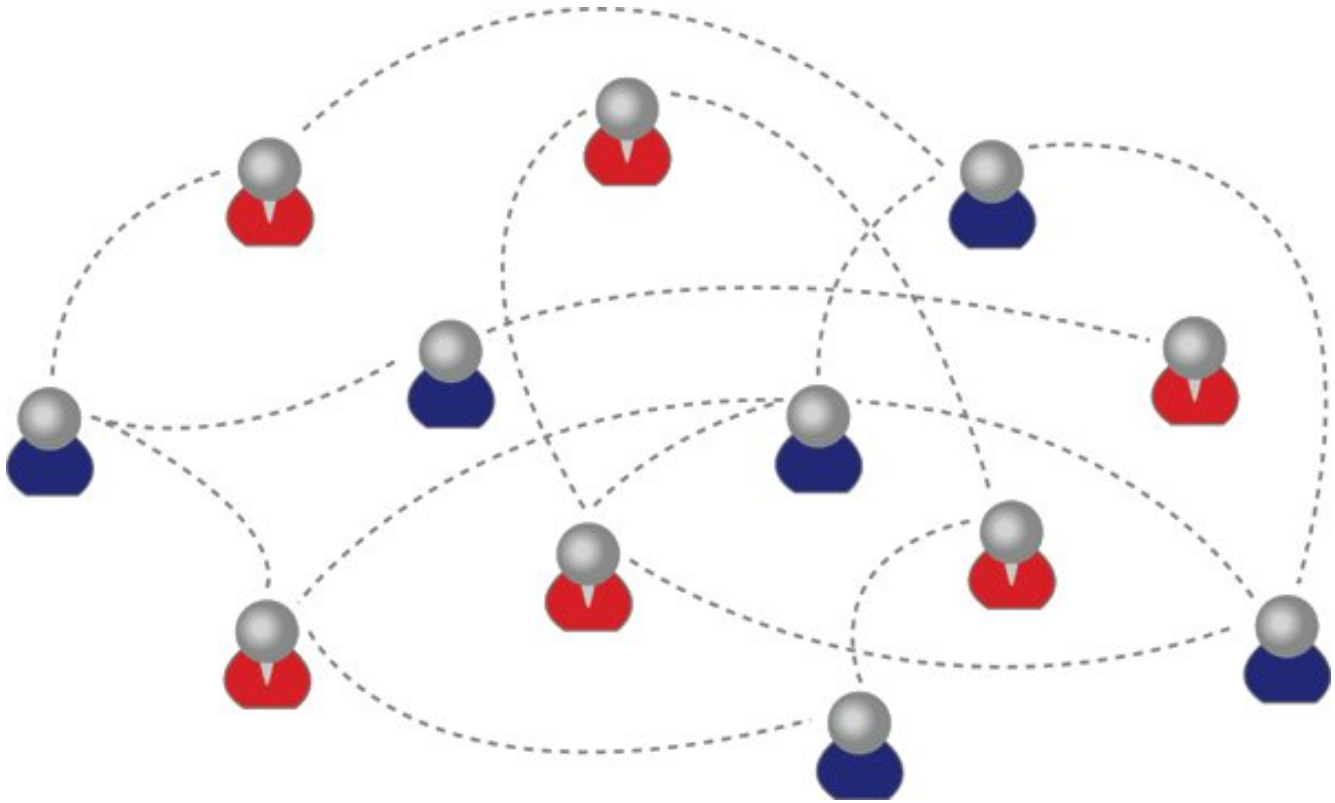
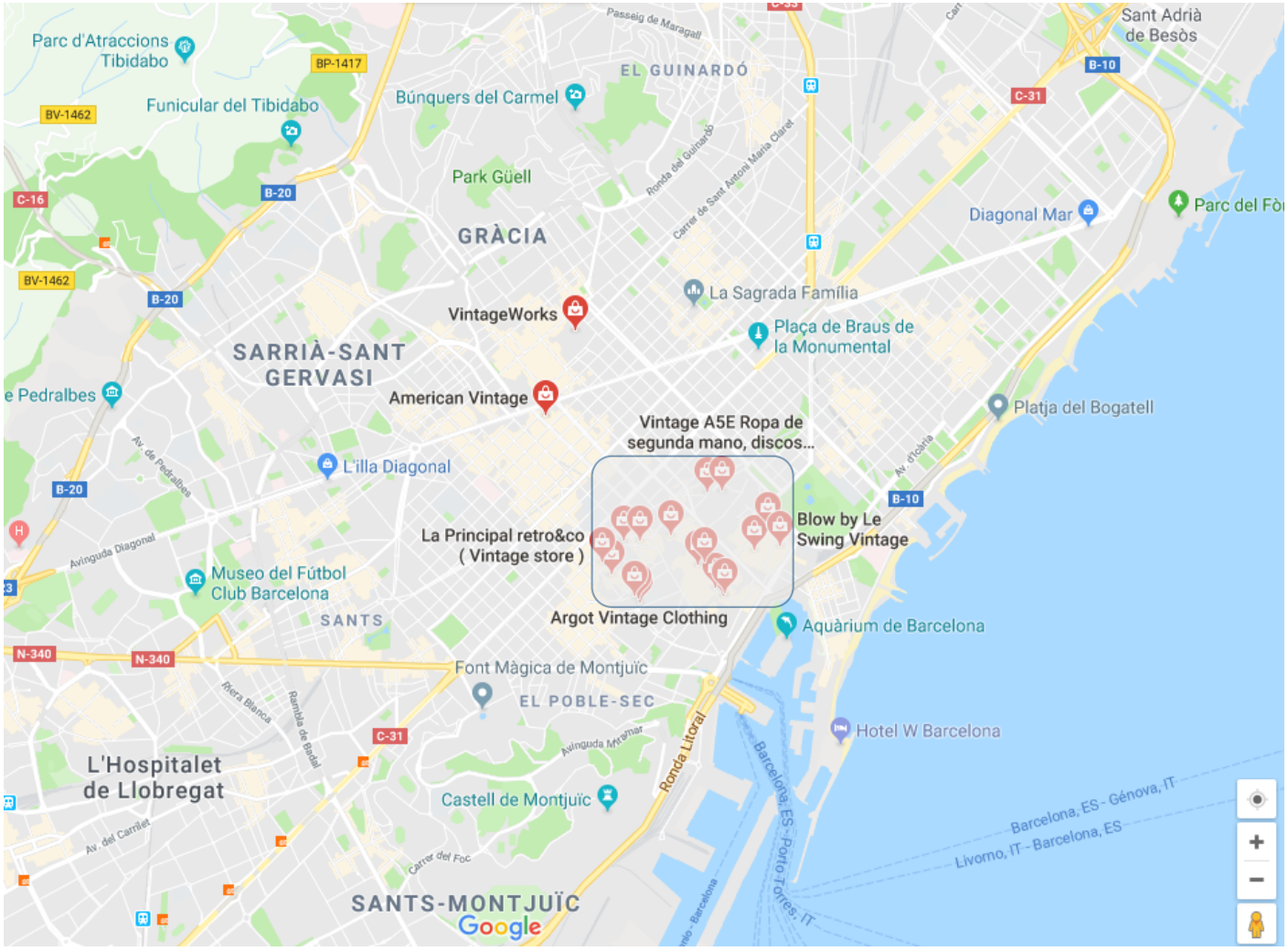


image from relenet.com

Clustering & Guilt-by-association

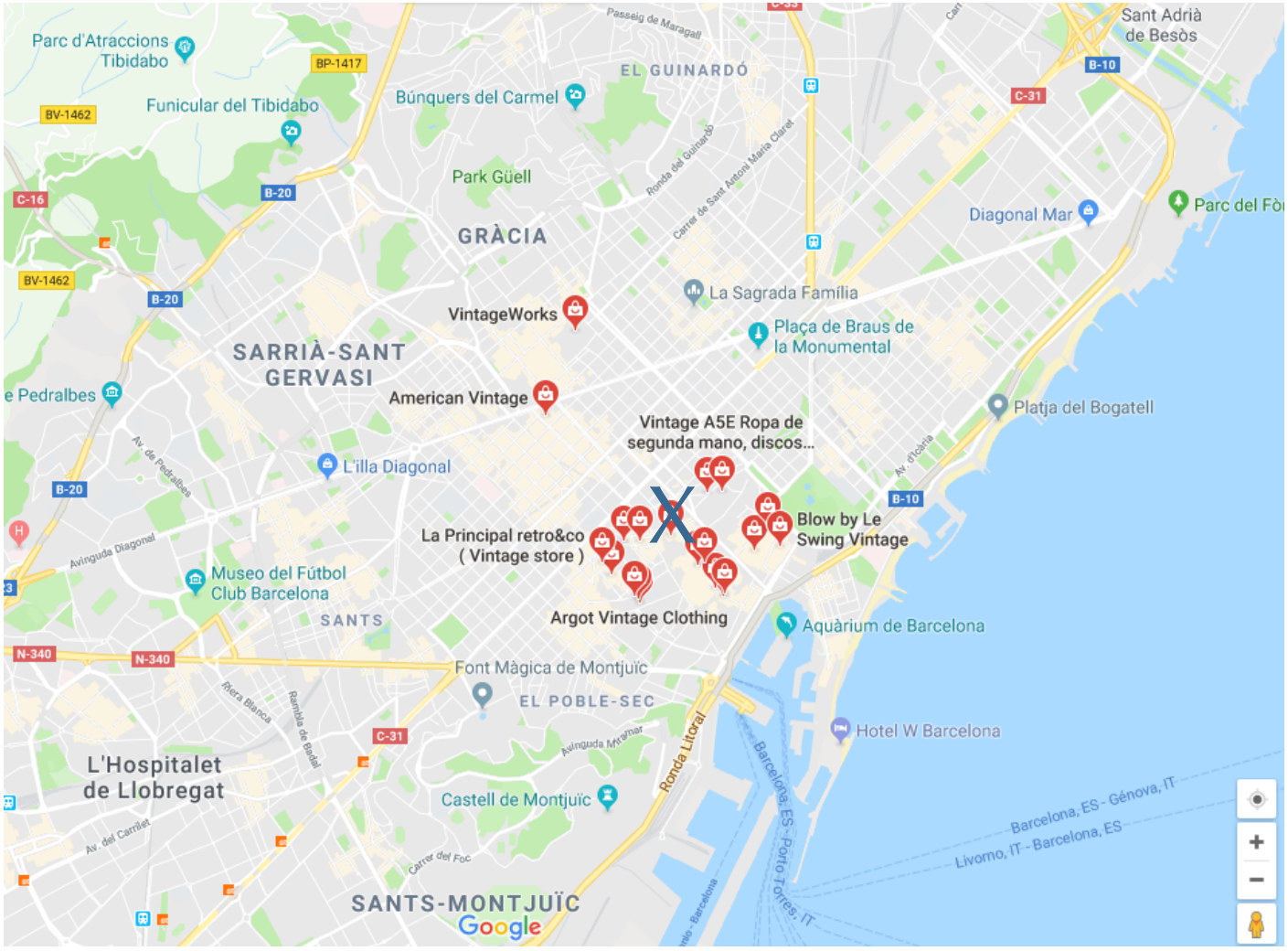
Vintage shops in Barcelona

Clustering & Guilt-by-association



clustering (Vintage shops are located in the center)

Clustering & Guilt-by-association



guilt-by-association (X is a vintage shop)

Interactome: The cellular map

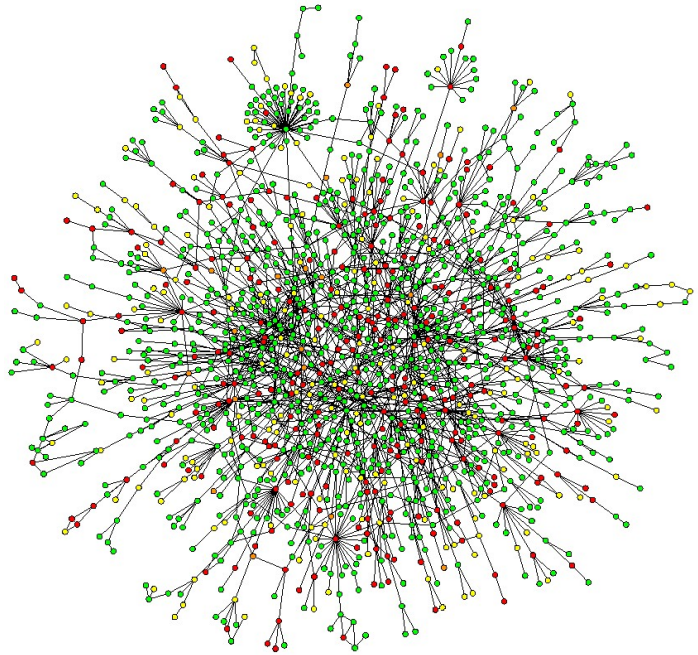
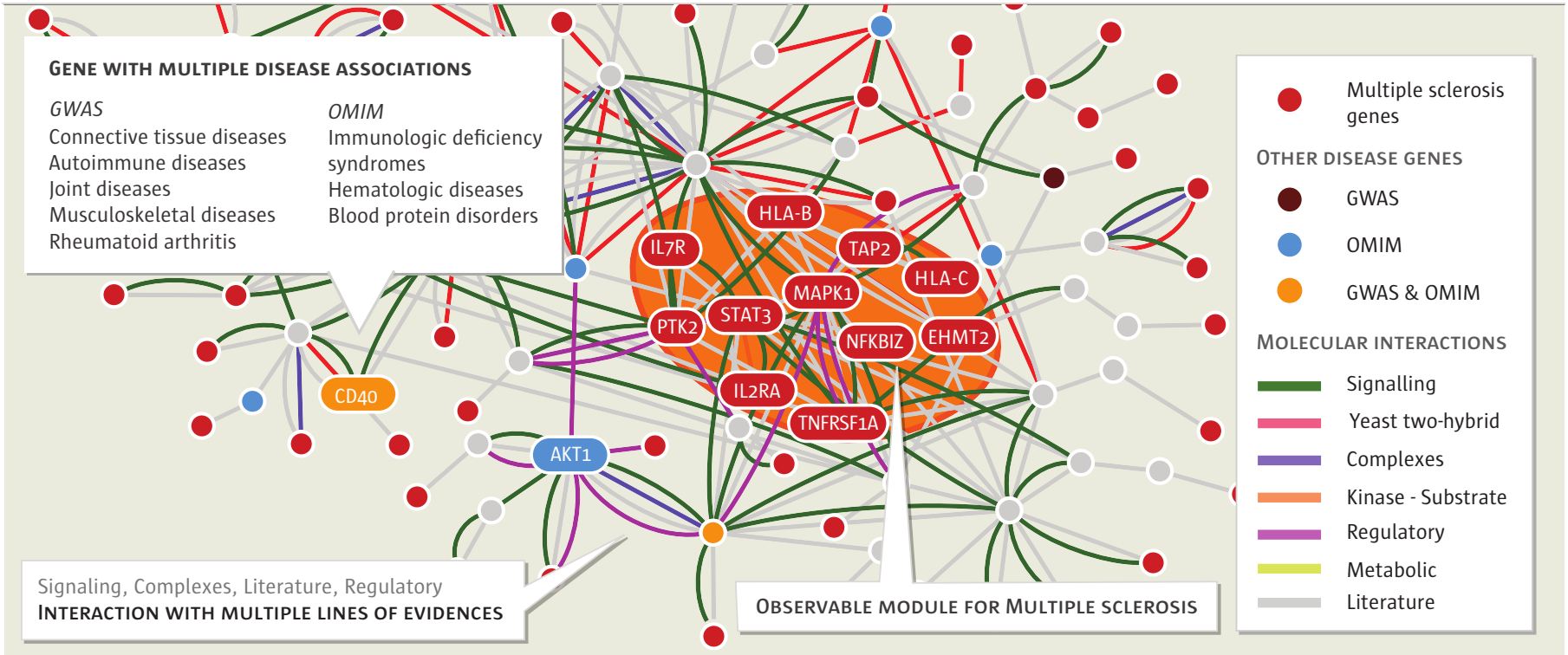


image from bordalierinstitute.com

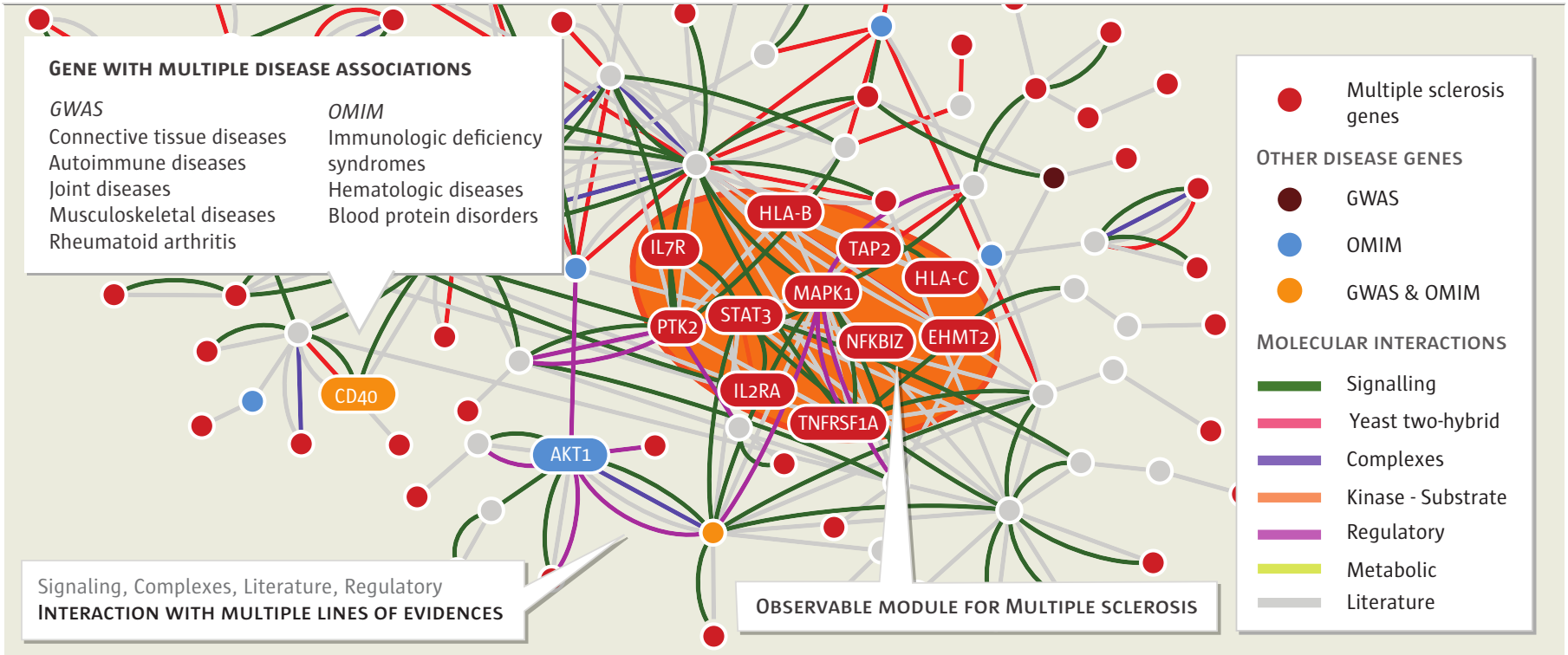
- Proteins “talk to each other” by physically interacting with each other
- These interactions are essential for performing biological processes
- The network of interactions between proteins: Interactome

Guilt-by-association in the interactome



(Menche et al., 2015, Science)

Guilt-by-association in the interactome



(Menche et al., 2015, Science)

How to generate and analyze the interactome?

BIANA: Biological Interaction And Network Analysis

Garcia-Garcia *et al. BMC Bioinformatics* 2010, **11**:56
http://www.biomedcentral.com/1471-2105/11/56

BMC Bioinformatics

SOFTWARE Open Access

Biana: a software framework for compiling biological interactions and analyzing networks

Javier Garcia-Garcia, Emre Guney, Ramon Aragues, Joan Planas-Iglesias, Baldo Oliva*

Abstract

Background: The analysis and usage of biological data is hindered by the spread of information across multiple repositories and the difficulties posed by different nomenclature systems and storage formats. In particular, there is an important need for data unification in the study and use of protein-protein interactions. Without good integration strategies, it is difficult to analyze the whole set of available data and its properties.

Results: We introduce BIANA (Biologic Interactions and Network Analysis), a tool for biological information integration and network management. BIANA is a Python framework designed to achieve two major goals: i) the integration of multiple sources of biological information, including biological entities and their relationships, and ii) the management of biological information as a network where entities are nodes and relationships are edges. Moreover, BIANA uses properties of proteins and genes to infer latent biomolecular relationships by transferring edges to entities sharing similar properties. BIANA is also provided as a plugin for Cytoscape, which allows users to visualize and interactively manage the data. A web interface to BIANA providing basic functionalities is also available. The software can be downloaded under GNU GPL license from http://sbi.imim.es/web/BIANA.php.

Conclusions: BIANA's approach to data unification solves many of the nomenclature issues common to systems dealing with biological data. BIANA can easily be extended to handle new specific data repositories and new specific data types. The unification protocol allows BIANA to be a flexible tool suitable for different user requirements: non-expert users can use a suggested unification protocol while expert users can define their own specific unification rules.

Protein Interaction Data

Protein-protein interaction (PPI) data is spread across various repositories
(Similar to the social interactions data that is spread across Facebook, Twitter, Instagram, etc...)

MIPS

DIP

Human Protein Reference Database

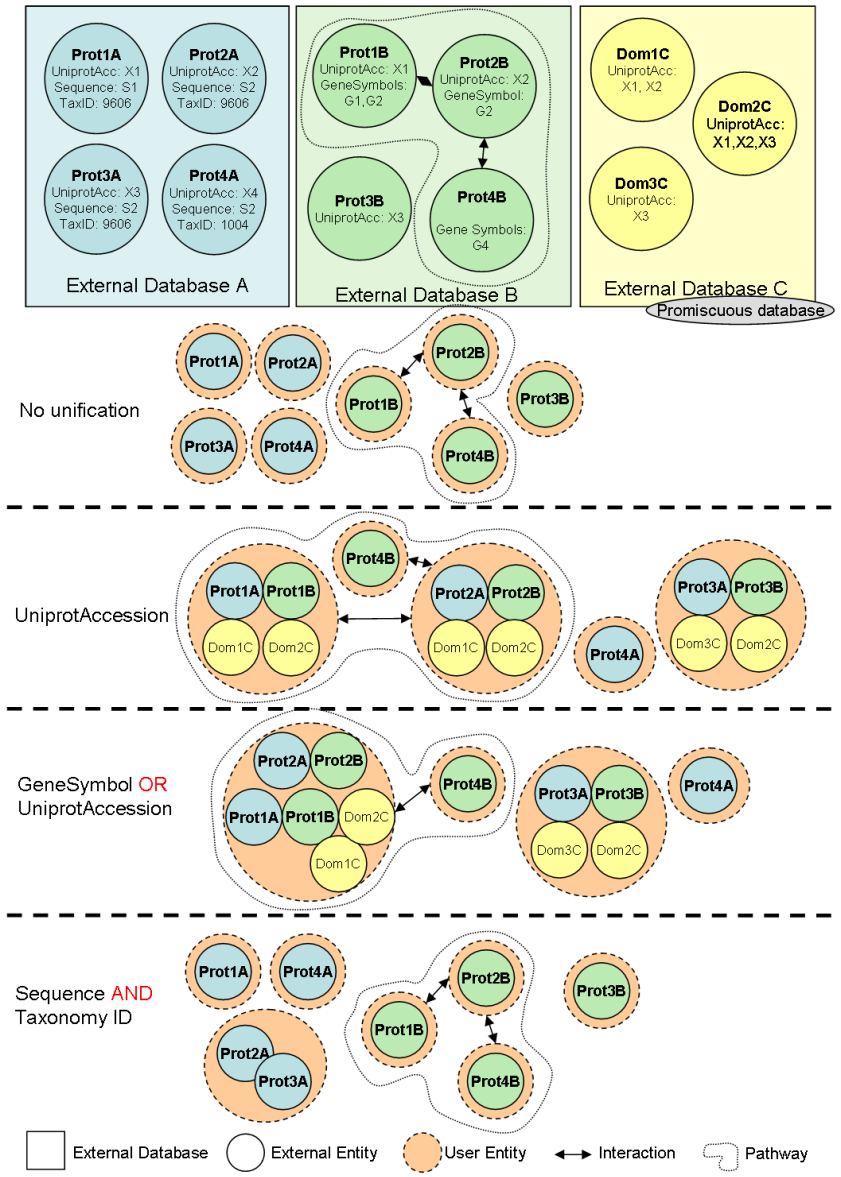
IntAct

STRING

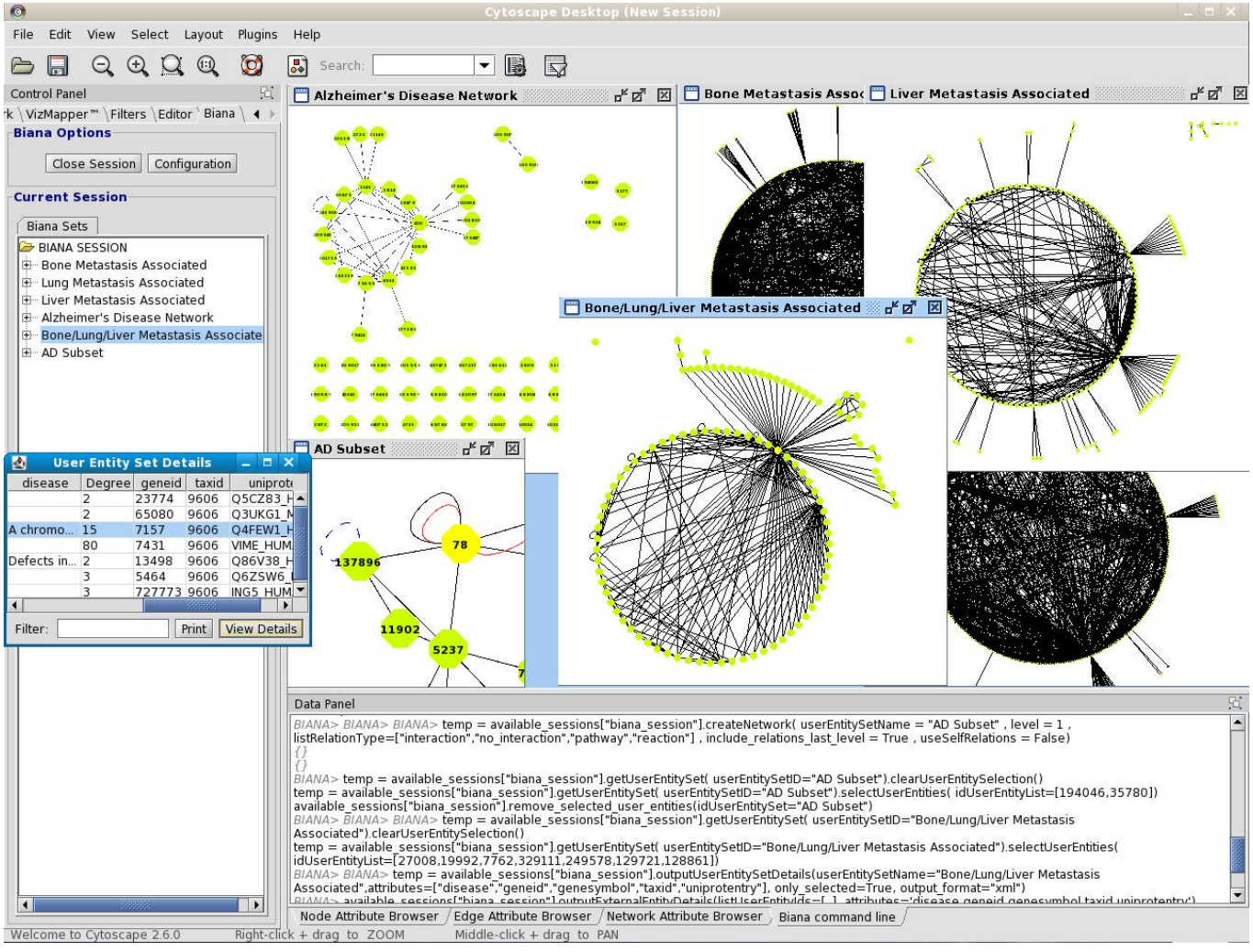
MINT

BioGrid

Integrating Protein Interaction Data



Analyzing Protein Interaction Data



PPICoin

A "new" pseudo-crypto currency for computational biologists

Rules for block generation and transaction verification

- Start a new block with the gene (protein X) associated to *"Nemaline myopathy 3"* in DisGeNET
- Add a protein Y interacting with X to the block such that it interacts with at least one other protein Z
- If there are multiple candidate proteins Y, add the one which has the smallest ENTREZ gene id
- Write that protein Y on a paper (along with your name) and pass it to the person on your left
- When you receive the paper, verify the correctedness of previous transaction and repeat the steps above to generate a new one

Analyzing Protein Interaction Data – Hands On!

Given G(V, E) the interactome represented as a graph (V: proteins, E: interactions) and k is node degree (number of the node's neighbors).

PPICoin:

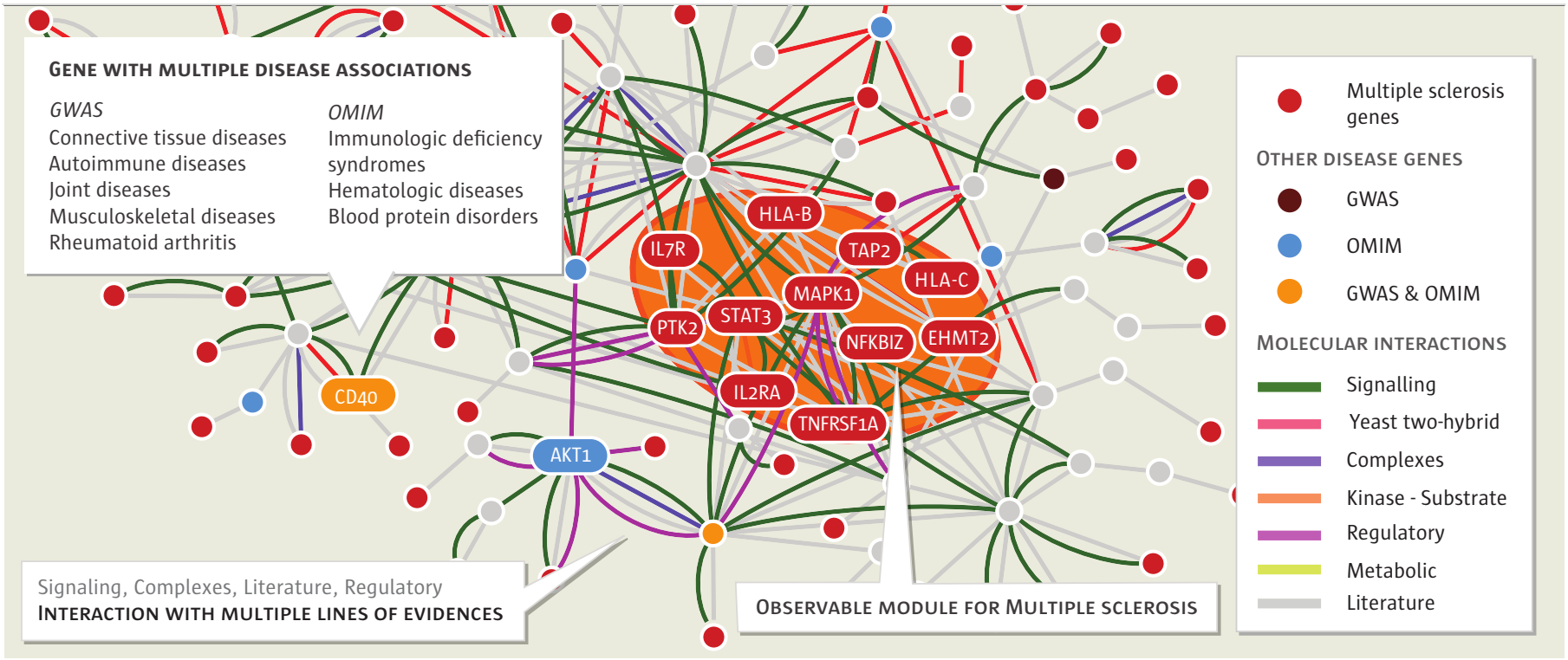
X	Y	Z	...
Emre	Alice	Bob	...

Subject to:

- $X, Y \in V$
- $Candidates(Y) = \{y : (X, y) \in E, k_y > 1\}$
- $Y = argmin_{Geneid}(Candidates(Y))$

Resource	URL
Disease-gene information	
DisGeNET	`www.disgenet.org`
Interactome generation	
BIANA (web server)	`sbi.imim.es/BIANA.php`
BIANA generated files	`emreguney.net/doc/ms_omics.zip`
Interactome visualization	
Cytoscape	`www.cytoscape.org`

Guilt-by-association in the interactome



(Menche et al., 2015, Science)

How to use interactome neighborhood for extracting biologically meaningful information?

A similar problem in a different domain: Image recovery

guilt-by-association

$\Rightarrow$

images from blog.newhumanist.org.uk/2012/08/leonardo-da-vincis-last-supper.html

A similar problem in a different domain: Image recovery

guilt-by-association

⇒

images from blog.newhumanist.org.uk/2012/08/leonardo-da-vincis-last-supper.html

GUILD: Genes Underlying Inheritence Linked Disorders

OPEN ACCESS Freely available online

PLOS ONE

Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization

Emre Guney, Baldo Oliva*

Structural Bioinformatics Group (GRIB), Universitat Pompeu Fabra, Barcelona Research Park of Biomedicine (PRBB), Barcelona, Catalonia, Spain

Abstract

Complex genetic disorders often involve products of multiple genes acting cooperatively. Hence, the pathophenotype is the outcome of the perturbations in the underlying pathways, where gene products cooperate through various mechanisms such as protein-protein interactions. Pinpointing the decisive elements of such disease pathways is still challenging. Over the last years, computational approaches exploiting interaction network topology have been successfully applied to prioritize individual genes involved in diseases. Although linkage intervals provide a list of disease-gene candidates, recent genome-wide studies demonstrate that genes not associated with any known linkage interval may also contribute to the disease phenotype. Network based prioritization methods help highlighting such associations. Still, there is a need for robust methods that capture the interplay among disease-associated genes mediated by the topology of the network. Here, we propose a genome-wide network-based prioritization framework named GUILD. This framework implements four network-based disease-gene prioritization algorithms. We analyze the performance of these algorithms in dozens of disease phenotypes. The algorithms in GUILD are compared to state-of-the-art network topology based algorithms for prioritization of genes. As a proof of principle, we investigate top-ranking genes in Alzheimer's disease (AD), diabetes and AIDS using disease-gene associations from various sources. We show that GUILD is able to significantly highlight disease-gene associations that are not used *a priori*. Our findings suggest that GUILD helps to identify genes implicated in the pathology of human disorders independent of the loci associated with the disorders.

Citation: Guney E, Oliva B (2012) Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization. PLoS ONE 7(9): e43557. doi:10.1371/journal.pone.0043557

Editor: Narcis Fernandez-Fuentes, Aberystwyth University, United Kingdom

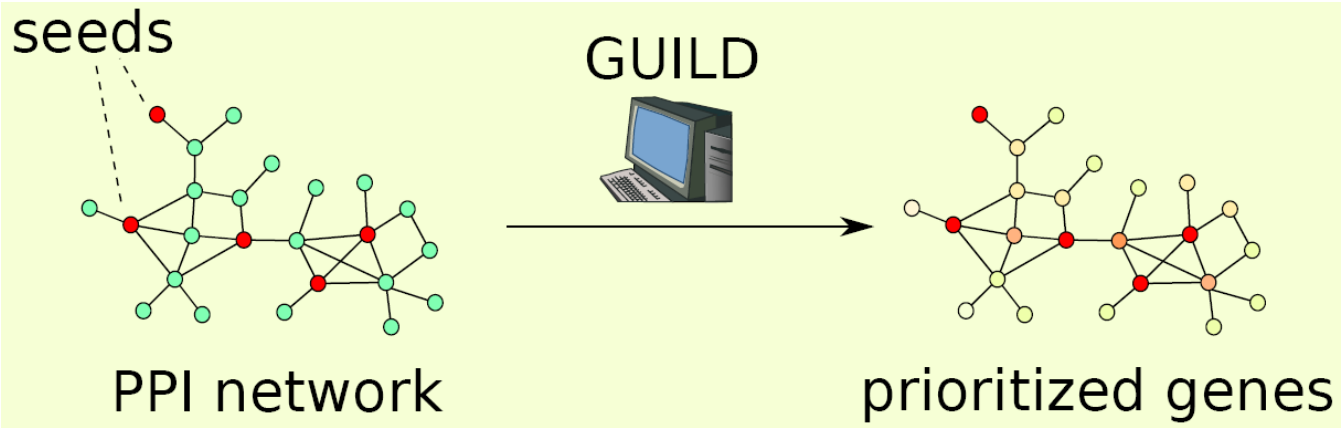
Received March 19, 2012; **Accepted** July 23, 2012; **Published** September 21, 2012

Funding: Departament d'Educació i Universitats de la Generalitat de Catalunya i del Fons Social Europeu (Department of Education and Universities of the Generalitat of Catalonia and the European Social Fons). Spanish Ministry of Science and Innovation (MICINN), FEDER (Fonds Européen de Développement Régional) BIO2008-0205, BIO2011-22568, PSE-0100000-2007, and PSE-0100000-2009; and by EU grant EraSysbio+ (SHIPREC) Euroinvestigación (EUI2009-04018). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: baldo.oliva@upf.edu

GUILD framework

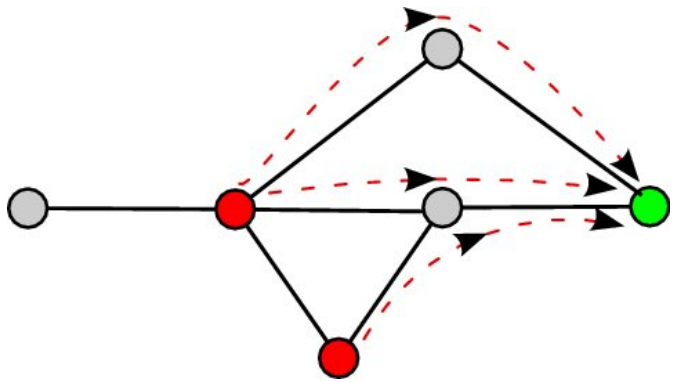


- NetScore
- NetZcore
- NetShort
- NetCombo

- Functional Flow *(Nabieva et al., 2005)*
- Random walk with restart *(Kohler et al., 2008)*
- PageRank with priors *(Chen et al., 2009)*
- Network propagation *(Vanunu et al., 2011)*

NetScore

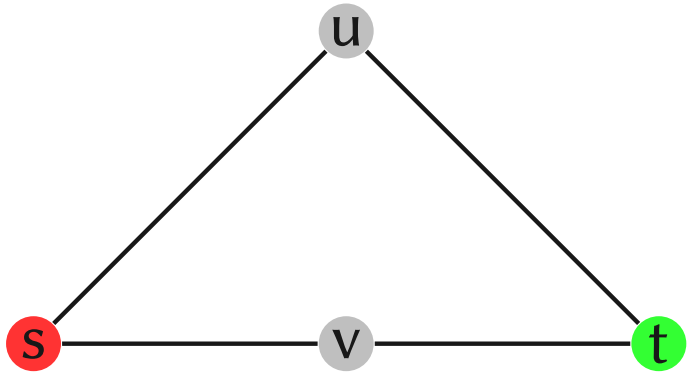
- considers multiple shortest paths in-between the two nodes



$$message = \left\{ \begin{array}{l} source \\ timestamp \\ path\_weight \end{array} \right\}$$

u
m_u
$\vdots$
m_v
$\vdots$

NetScore

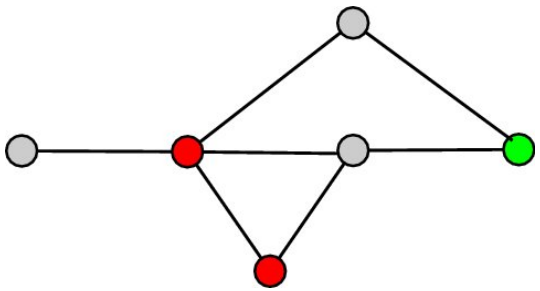


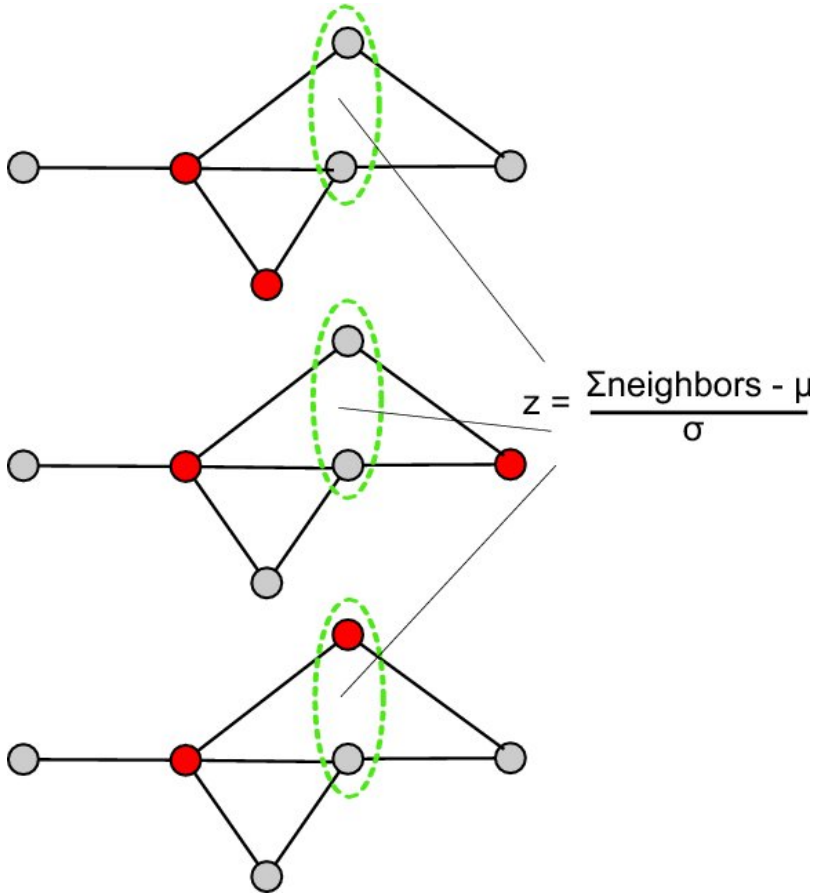

```
Input: G = (V, E) graph with score property score:V → [0,1],
       weight property weight: E → [0,1], nRepetition, nIteration.
Output: G = (V, E) graph with score property score′:V → [0,1].
for i=1 to nRepetition do
   /* Initialize message arrays                                    */
   foreach u ∈ V do
      m.source ← u // source node id
      m.timestamp ← 0 // the iteration when recieved
      m.path_weight ← 1 // weights of the traveled path
      messages(u) ← { m }
   end
   for j=1 to nIteration do
      /* Digest messages                                          */
      foreach u ∈ V do
         foreach v ∈ {(u, v) ∈ E } do
            foreach m ∈ messages(v) do
               /* Do not accept messages from the same node
                  recieved during previous iterations            */
               if AcceptMessage(messages(u), m) then
                  m.timestamp ← j
                  m.path_weight ← m.path_weight ∗ weight(u, v)
                  messages(u) ← messages(u) ∪ m
               end
            end
         end
      end
      /* Update node scores                                       */
      foreach u ∈ V do
         foreach m ∈ messages(u) do
            score(u) ← score(u) + m.path_weight ∗
            score(m.source)
         end
         score(u) ← score(u) / ‖ messages(u) ‖
      end
   end
end
```

	s	u	v	t
initially	m_s^0	m_u^0	m_v^0	m_t^0
iteration 1	m_u^1	m_s^1	m_s^1	m_u^1
iteration 2	m_v^1	m_t^1	m_t^1	m_v^1
	$2m_t^2$	$2m_v^2$	$2m_u^2$	$2m_s^2$

NetZcore

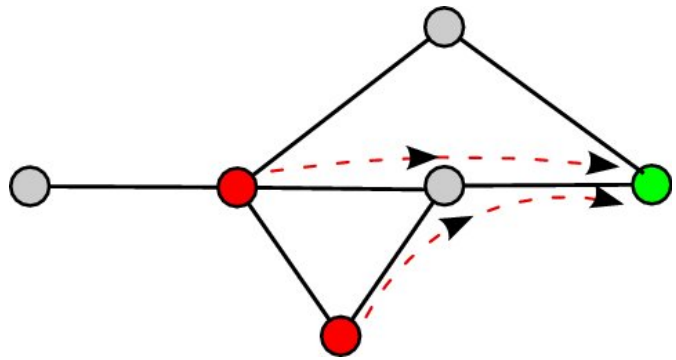
- checks the “significance” of the neighborhood configuration

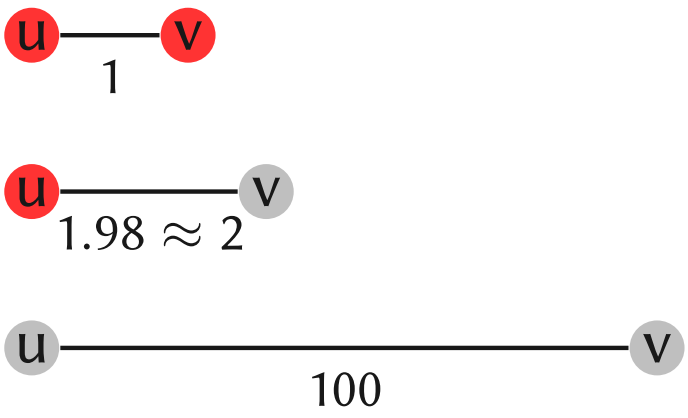




NetShort

- incorporates “disease-relevance” of a path
- the more seeds the path has, the shorter it is

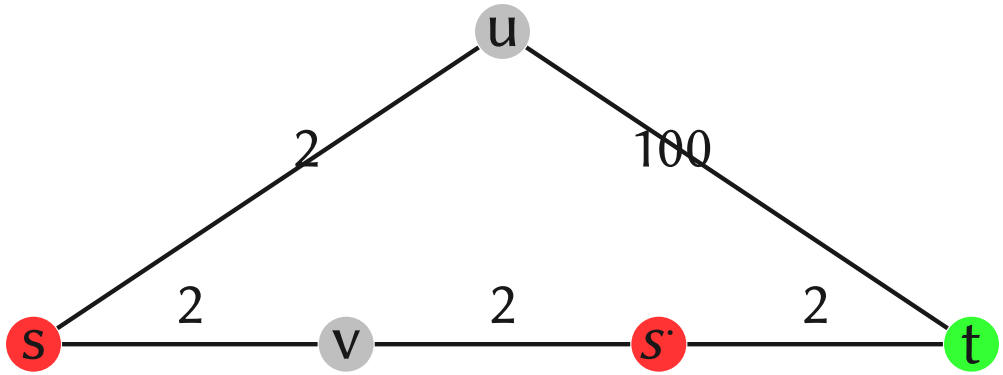


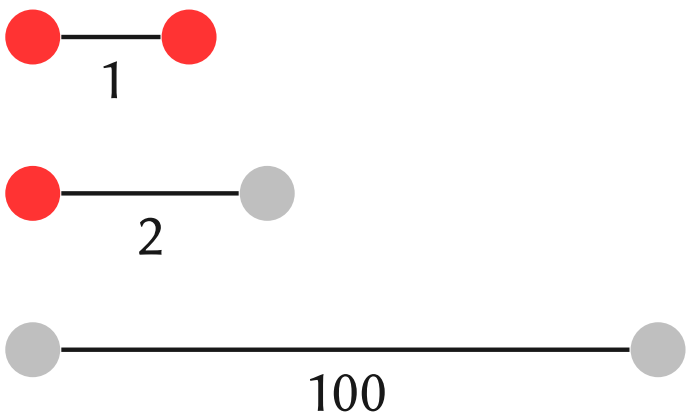


$$score(u) = \begin{cases} 1, & \text{if } u \text{ is seed} \\ 0.01, & \text{otherwise} \end{cases}$$

$$weight(u, v) = \frac{1}{(score(u)+score(v))/2}$$

NetShort





$$path\_length(s,t)_{\{s,u,t\}} = 2 + 100 = 102$$

$$path\_length(s,t)_{\{s,v,s',t\}} = 2 + 2 + 2 = 6$$

NetCombo

- combines the scores of the three methods
- calculates a z-score for each node using the score distribution of each method
- averages the z-score from all three algorithms for each node

$$score_u^{NetCombo} = \frac{1}{3} * \sum_{method \in \{NetScore, NetZcore, NetCombo\}} \frac{score_u^{method} - \mu^{method}}{\sigma^{method}}$$

GUILD accuracy

Data Set	Metric	NetScore	NetZcore	NetShort	NetCombo	Func. Flow	PageRank	Random Walk	Network Prop.
OMIM	AUC	67.49	62.99	65.63	**72.09**	58.55	57.03	55.36	65.97
	Sens.	20.69	19.62	15.41	21.46	22.31	10.76	14.64	**23.24**
Goh	AUC	**67.32**	61.45	55.36	67.08	54.78	52.39	49.35	54.74
	Sens.	**11.61**	11.05	4.88	11.34	6.22	4.00	5.69	8.66
Chen	AUC	75.92	72.80	63.11	**78.41**	63.56	65.30	61.78	69.07
	Sens.	**18.89**	12.84	9.06	17.51	12.43	6.00	9.64	15.30

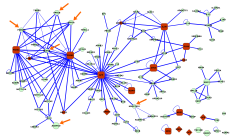
Improvement over existing methods is significant

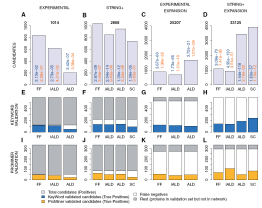
- for NetCombo in all three datasets
- for NetScore in Goh and Chen datasets

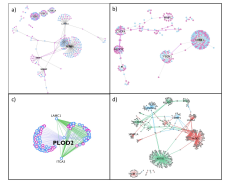
Guney and Oliva, 2012

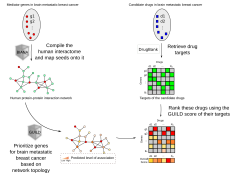
Several applications of GUILD

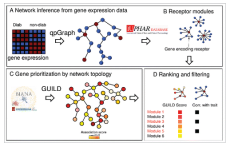
- implicating genes in AD *(Guney and Oliva, 2012, PLoS ONE)*
- extending apoptosis pathway *(Planas et al., 2012, OMICS: A Jour Int Bio)*
- prioritizing genes in bone metastatic breast cancer *(Santana-Codina et al., 2013, Mol & Cel Prot)*
- identifying subnetworks driving brain and lung metastasis in breast cancer *(Engin et al., 2013, PLoS ONE)*
- analysis of functional diversity of disease genes *(Guney and Oliva, 2014, PLoS ONE)*
- drug repurposing in brain metastatic breast cancer *(Martinez-Aranda et al., 2015, Oncotarget)*
- prioritization of regulatory modules in insulin secretion *(Hänzelmann et al., 2015, Islets)*











GUILDify: GUILD web server

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 12 2014, pages 1789–1790
doi:10.1093/bioinformatics/btu092

Systems biology

Advance Access publication February 14, 2014

GUILDify: a web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms

Emre Guney†, Javier Garcia-Garcia and Baldo Oliva*

Departament de Ciencies Experimentals i de la Salut, Structural Bioinformatics Group (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona, 08003 Catalonia, Spain

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: Determining genetic factors underlying various phenotypes is hindered by the involvement of multiple genes acting cooperatively. Over the past years, disease–gene prioritization has been central to identify genes implicated in human disorders. Special attention has been paid on using physical interactions between the proteins encoded by the genes to link them with diseases. Such methods exploit the guilt-by-association principle in the protein interaction network to uncover novel disease–gene associations. These methods rely on the proximity of a gene in the network to the genes associated with a phenotype and require a set of initial associations. Here, we present GUILDify, an easy-to-use web server for the phenotypic characterization of genes. GUILDify offers a prioritization approach based on the protein–protein interaction network where the initial phenotype–gene associations are retrieved via free text search on biological databases. GUILDify web server does not restrict the prioritization to any predefined phenotype, supports multiple species and accepts user-specified genes. It also prioritizes drugs based on the ranking of their targets, unleashing opportunities for repurposing drugs for novel therapies.

Availability and implementation: Available online at http://sbi.imim.es/GUILDify.php

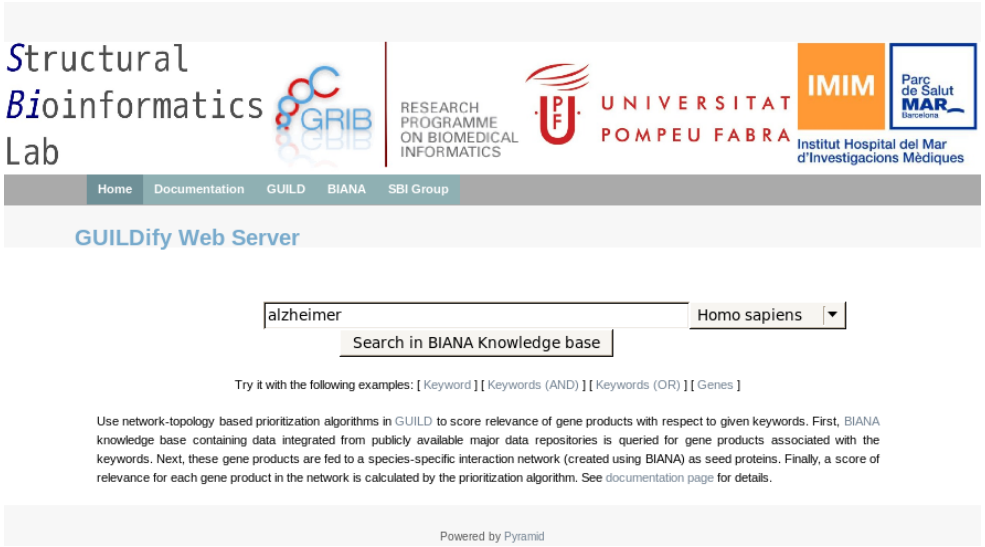
Contact: baldo.oliva@upf.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Methods using protein–protein interactions (PPIs) exploit the 'guilt-by-association' principle over the network topology to uncover new disease–gene associations. The guilt-by-association principle suggests that the genes whose products (proteins) interact with the products of known disease genes are more likely to be disease genes (Aerts *et al.*, 2006; Lage *et al.*, 2007). Recently, we proposed three novel algorithms for genome-wide prioritization of disease genes using PPI networks and showed that a consensus method combining these algorithms improved the prioritization (Guney and Oliva, 2012) when using the disease–gene associations in Online Mendalian Inheritance in Man (OMIM) database (Hamosh *et al.*, 2005). Combined with genomics and proteomics data, the method has been successfully used to identify a gene driving metastasis to bone in breast cancer (Santana-Codina *et al.*, 2013).

Available network-based prioritization tools use either disease–gene annotations from OMIM database (Gottlieb *et al.*, 2011; Kohler *et al.*, 2008) or a set of genes provided by the user (Chen *et al.*, 2009; Kacprowski *et al.*, 2013; Warde-Farley *et al.*, 2010) as initial associations (seed genes). These tools typically output the prioritization for a set of candidate genes; genes lying under a given genomic interval, a set of user-provided genes or top ranking genes (several hundred at most). Furthermore, some of these tools are accessible only through Cytoscape (Saito *et al.*, 2012) as a plugin (Gottlieb *et al.*, 2011; Kacprowski *et al.*, 2013).

GUILDify



- free text search on UniProt, OMIM databases and GO for a given species
- matching genes are retrieved and used as seeds
- the PPI network is compiled from DIP, HPRD, IntAct, MINT, MPact, BioGRID, BIND

Guney and Oliva, 2014

GUILDify

GUILDify Web Server - Matching Proteins

35 BIANA entries are found for the query ***alzheimer*** in *Homo sapiens*. 2 of these entries have no interactions (not in the network).

GUILDify! using selected entries below [Options]

Options

☑ NetScore (Repetition: 3 ▾ Iteration: 2 ▾)

☐ NetZcore (Iteration: 5 ▾)

☐ NetShort

[Hide]

BIANA entries in the network

[Select All / None]

Keep	Gene ID	UniProt ID	Gene Symbol	Description
☑	2	P01023	A2M	{alzheimer disease, susceptibility to}, 104300 (3) omim
☑	26574	Q9NY61	AATF	May function as a general inhibitor of the histone deacetylase HDAC1. Binding to the pocket region of RB1 may displace HDAC1 from RB1/E2F complexes, leading to activation of E2F target genes and cell cycle progression. Conversely, displacement of HDAC1 from SP1 bound to the CDKN1A promoter leads to increased expression of this CDK inhibitor and blocks cell cycle progression. Also antagonizes PAWR mediated induction of aberrant amyloid peptide production in alzheimer disease (presenile and senile dementia), although the molecular basis for this phenomenon has not been described to date. swissprot
☑	1636	P12821	ACE	{alzheimer disease, susceptibility to}, 104300 (3) omim
☑	9507	O75173	ADAMTS4	Cleaves aggrecan, a cartilage proteoglycan, and may be involved in its turnover. May play an important role in the destruction of aggrecan in arthritic diseases. Could also be a critical factor in the exacerbation of neurodegeneration in alzheimer disease. Cleaves aggrecan at the '392-Glu-|-Ala-393' site. swissprot
☑	323	Q92870	APBB2	{alzheimer disease, late-onset}, 104300 (3) omim
☑	348	P02649	APOE	alzheimer disease swissprot Defects in APOE are a cause of hyperlipoproteinemia type 3 (HLPP3) [MIM:107741]; also known as familial dysbetalipoproteinemia. Individuals with HLPP3 are clinically characterized by xanthomas, yellowish lipid deposits in the palmar crease, or less specific on tendons and on elbows. The disorder rarely manifests before the third decade in men. In women, it is usually expressed only after the menopause. The vast majority of the patients are homozygous for APOE*2 alleles. More severe cases of HLPP3 have also been observed in individuals heterozygous for rare APOE variants. The influence of APOE on lipid levels is often suggested to have major implications for the risk of coronary artery disease (CAD). Individuals carrying the common APOE*4 variant are at higher risk of CAD. Genetic variations in APOE are associated with alzheimer disease type 2 (AD2) [MIM:104310]. It is a late-onset neurodegenerative disorder characterized by progressive dementia, loss of cognitive abilities, and deposition of fibrillar amyloid proteins as intraneuronal neurofibrillary tangles, extracellular amyloid plaques and vascular amyloid deposits. The major constituent of these plaques is the neurotoxic amyloid-beta-APP 40-42 peptide (s), derived proteolytically from the transmembrane precursor protein APP by sequential secretase processing. The cytotoxic C-terminal fragments (CTFs) and the caspase-cleaved products such as C31 derived from APP, are also implicated in neuronal death. Note=The APOE*4 allele is genetically associated with the common late onset familial and sporadic forms of alzheimer disease. Risk for AD increased from 20% to 90% and mean age at onset decreased from 84 to 68 years with increasing number of APOE*4 alleles in 42 families with late onset AD. Thus APOE*4 gene dose is a major risk factor for late onset AD and, in these families, homozygosity for APOE*4 was virtually sufficient to cause AD by age 80. The mechanism by which APOE*4 participates in pathogenesis is not known. Defects in

GUILDify

GUILDify Web Server - Results

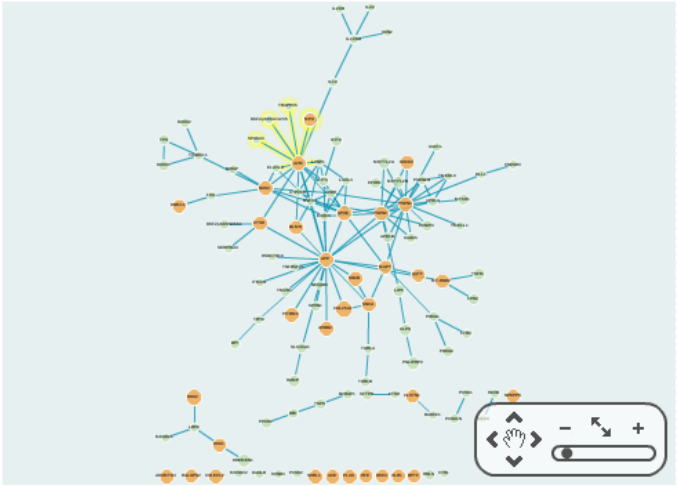
Calculated scores for proteins in the BIANA ***Homo sapiens*** interactome:

[Download all scores] [Download seed proteins] [Download interactome]

Proteins **1** - **20** of 11238 ||<< << >> >>||

Rank	Gene ID	UniProt ID	Gene Symbol	GUILD Score
1	84570	Q9BXS0	seed COL25A1	1.0000
2	27123	Q9UBU2	seed DKK2	0.9398
3	348	P02649	seed APOE	0.8481
4	6620	Q16143	seed SNCB	0.8435
5	10531	Q5JRX3	seed PITRM1	0.8373
6	5664	P49810	seed PSEN2	0.8123
7	5663	P49768	seed PSEN1	0.7905
8	6653	Q92673	seed SORL1	0.7861
9	1795	Q8IZD9	seed DOCK3	0.7858
10	1636	P12821	seed ACE	0.7848
11	5328	P00749	seed PLAU	0.7824
12	351	P05067	seed APP	0.7774
13	323	Q92870	seed APBB2	0.7756
14	9507	O75173	seed ADAMTS4	0.7754
15	4846	P29474	seed NOS3	0.7717
16	-	Q8IVG9	seed MT-RNR2	0.7705
17	2	P01023	seed A2M	0.7599
18	642	Q13867	seed BLMH	0.7556
19	81035	Q5KU26	seed COLEC12	0.7523
20	6622	P37840	seed SNCA	0.7518

Visualization options: Top 1% Top 5% Top 10%



Selected 4 nodes:

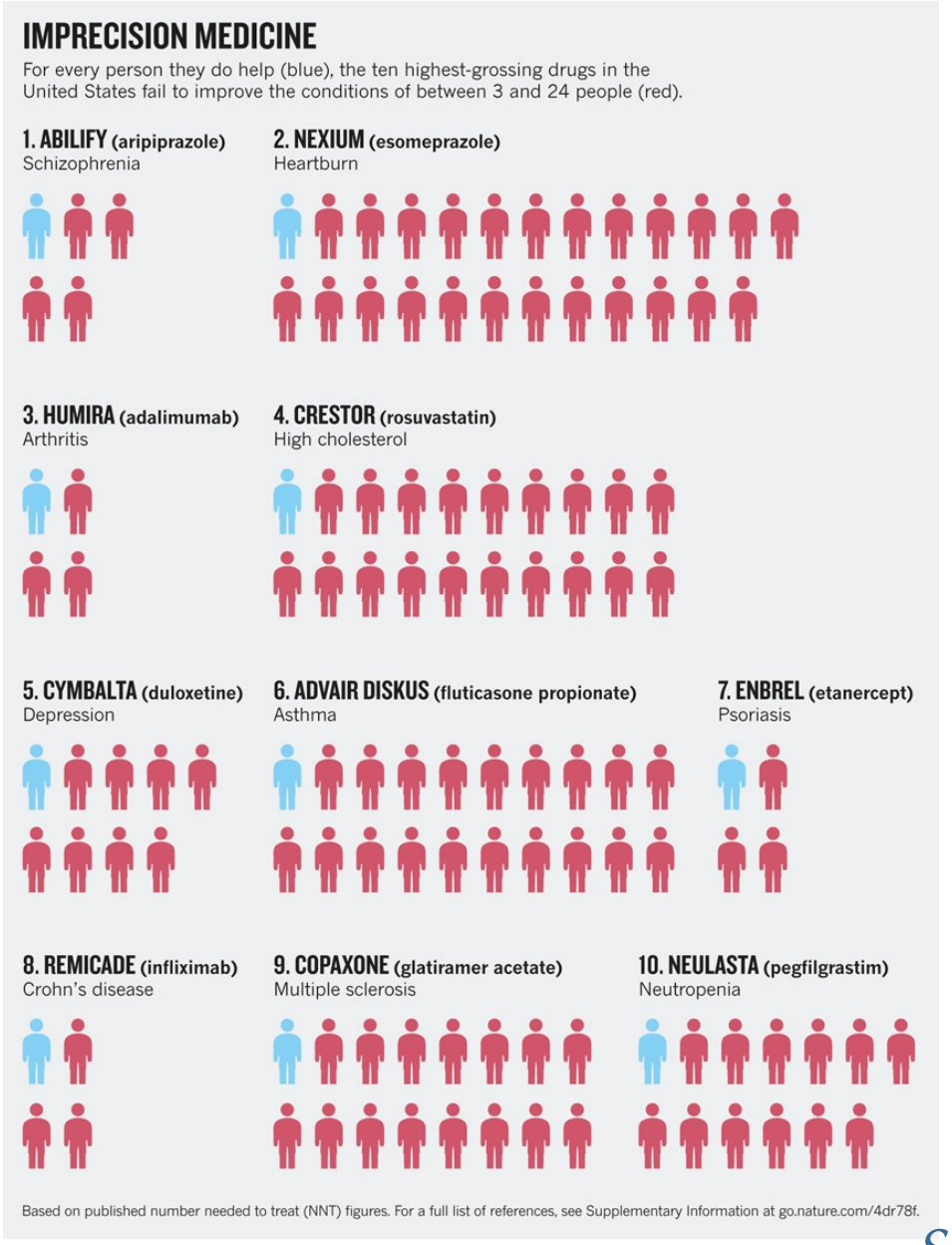
Rank	Gene ID	UniProt ID	Gene Symbol	GUILD Score	Type
96	124912	Q8IXA5	SPACA3	0.227186	non_seed
97	126003	Q8IUR0	TRAPPC5	0.227186	non_seed
78	1562	P33260	DKFZp686I24235	0.274228	non_seed
24	4353	P05164	MPO	0.721588	seed

Relationship between Alzheimer's disease and type 2 diabetes

Investigate the common genetic factors involved in both diseases

Resource	URL
Disease-gene information	
DisGeNET	`www.disgenet.org`
Interactome-based analysis	
GUILDify web server	`sbi.imim.es/GUILDify2.php`
Scripting	
Your favorite tool	bash / python / R / perl / ...

Patient-level heterogeneity



Schork, 2015, Nature

PEPPER: PErsonalized exPression ProfilER

npj | Systems Biology and Applications

www.nature.com/npjsba

ARTICLE OPEN

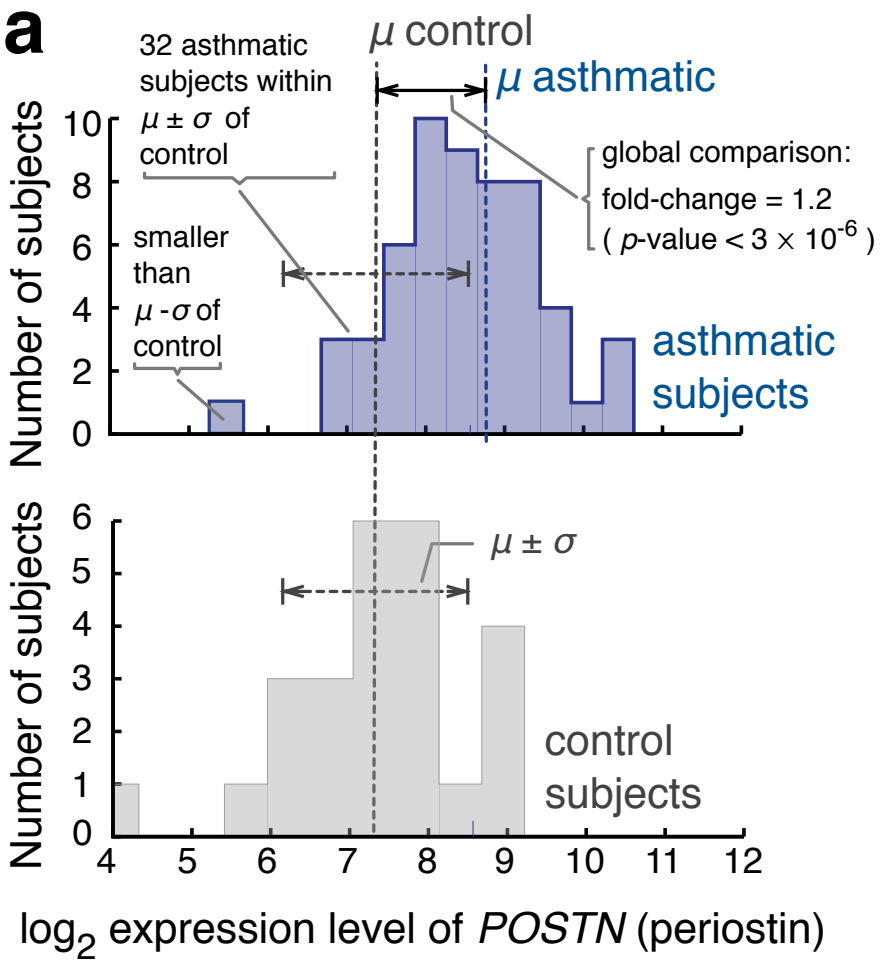
Integrating personalized gene expression profiles into predictive disease-associated gene pools

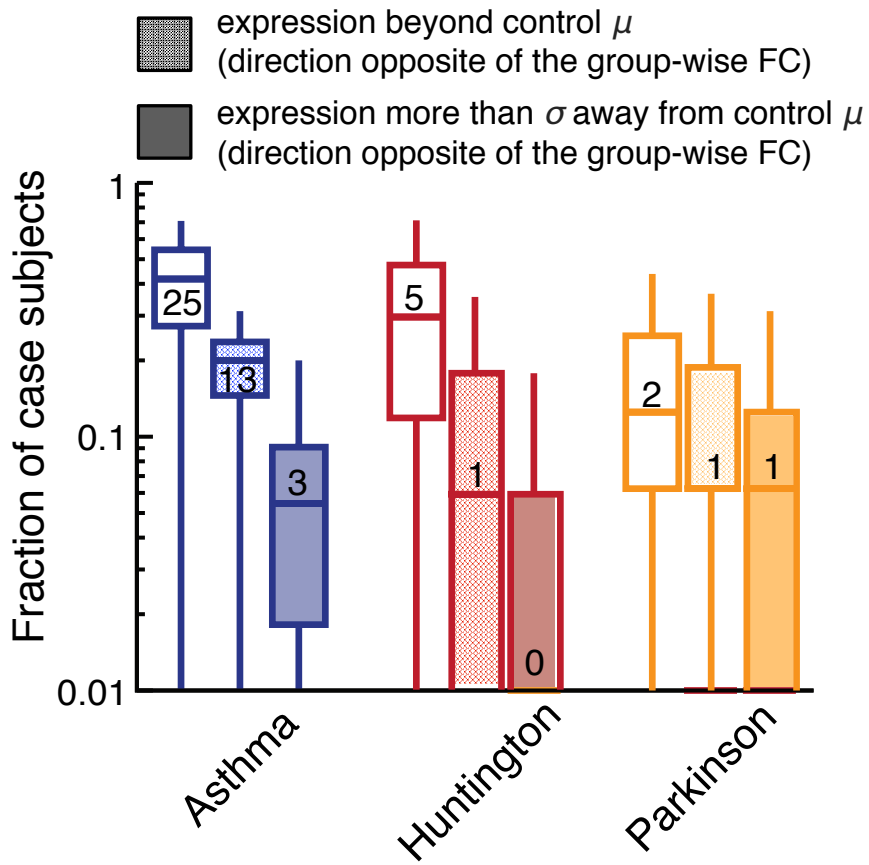
Jörg Menche[1,2,3], Emre Guney[1,4], Amitabh Sharma[1,4,5], Patrick J. Branigan[6], Matthew J. Loza[6], Frédéric Baribaud[6], Radu Dobrin[6] and Albert-László Barabási[1,2,4,5]

Gene expression data are routinely used to identify genes that *on average* exhibit different expression levels between a case and a control group. Yet, very few of such differentially expressed genes are detectably perturbed in individual patients. Here, we develop a framework to construct *personalized* perturbation profiles for individual subjects, identifying the set of genes that are significantly perturbed in each individual. This allows us to characterize the heterogeneity of the molecular manifestations of complex diseases by quantifying the expression-level similarities and differences among patients with the same phenotype. We show that despite the high heterogeneity of the individual perturbation profiles, patients with asthma, Parkinson and Huntington's disease share a broadpool of sporadically disease-associated genes, and that individuals with statistically significant overlap with this pool have a 80–100% chance of being diagnosed with the disease. The developed framework opens up the possibility to apply gene expression data in the context of precision medicine, with important implications for biomarker identification, drug development, diagnosis and treatment.

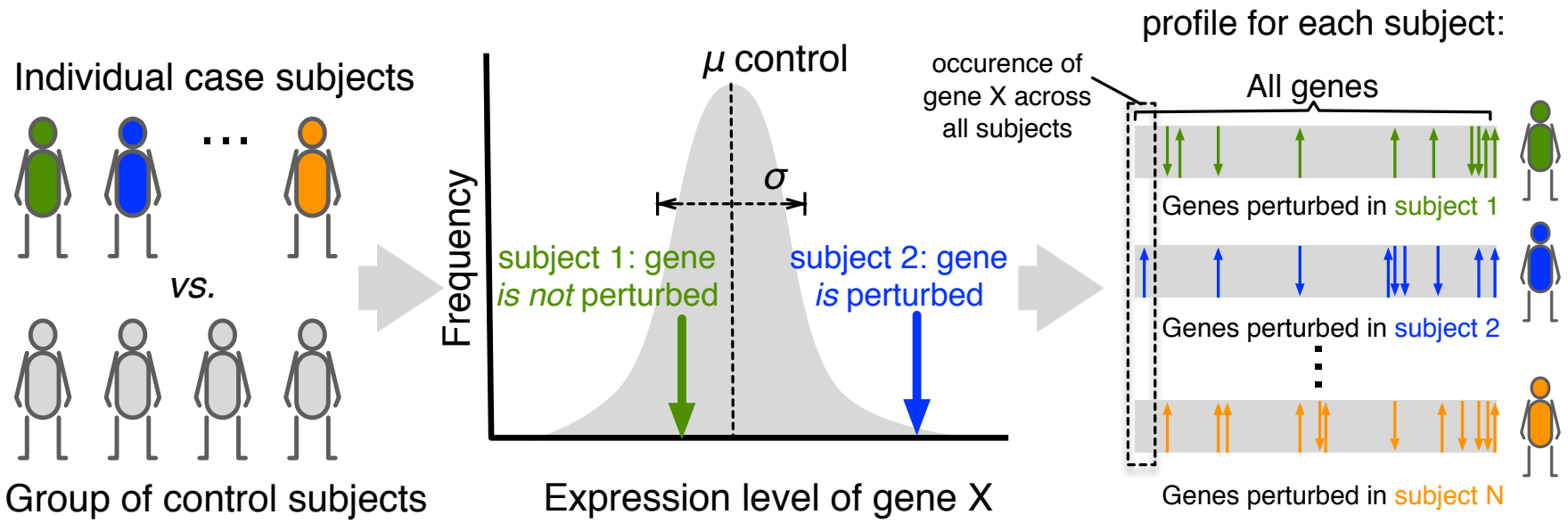
npj Systems Biology and Applications (2017)3:10; doi:10.1038/s41540-017-0009-0

Group-wise differentially expressed genes do not capture transcriptomic heterogeneity





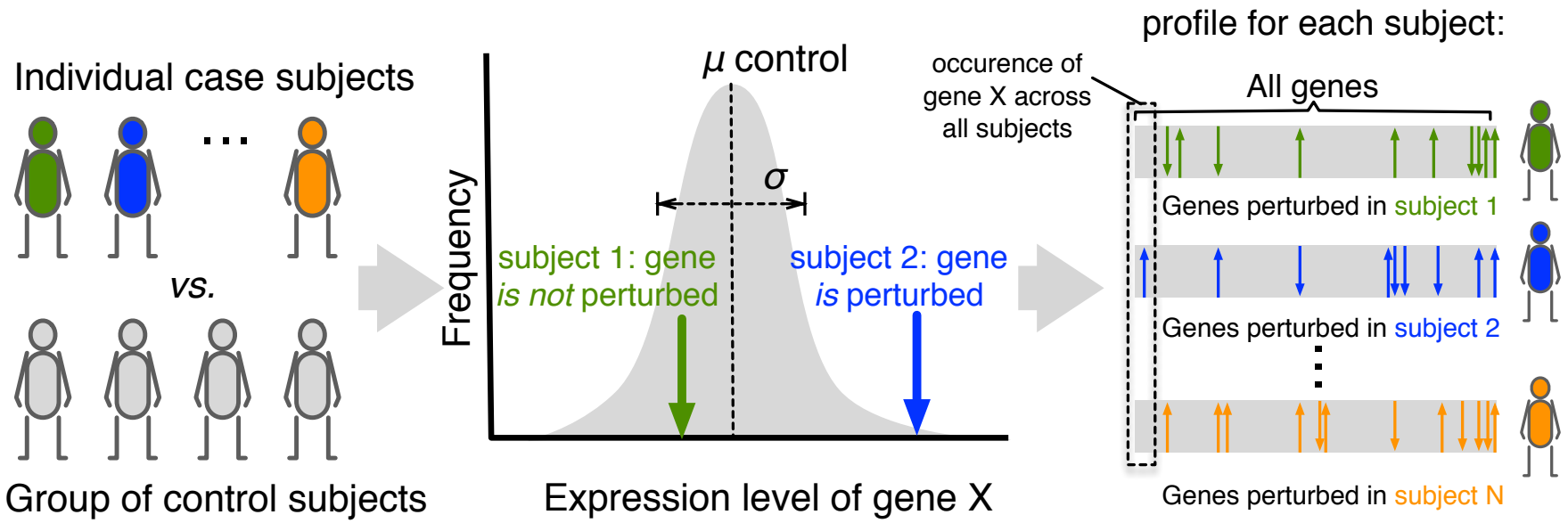
PeeP: PErsonalized Expression Profile



	Samples			
Case	x	x	...	x
Control	c	c	c	c

$$z(gene \text{ in } x) = \frac{expression_x(gene) - \mu_c(gene)}{\sigma_c(gene)}$$

PeeP: PErsonalized Expression Profile



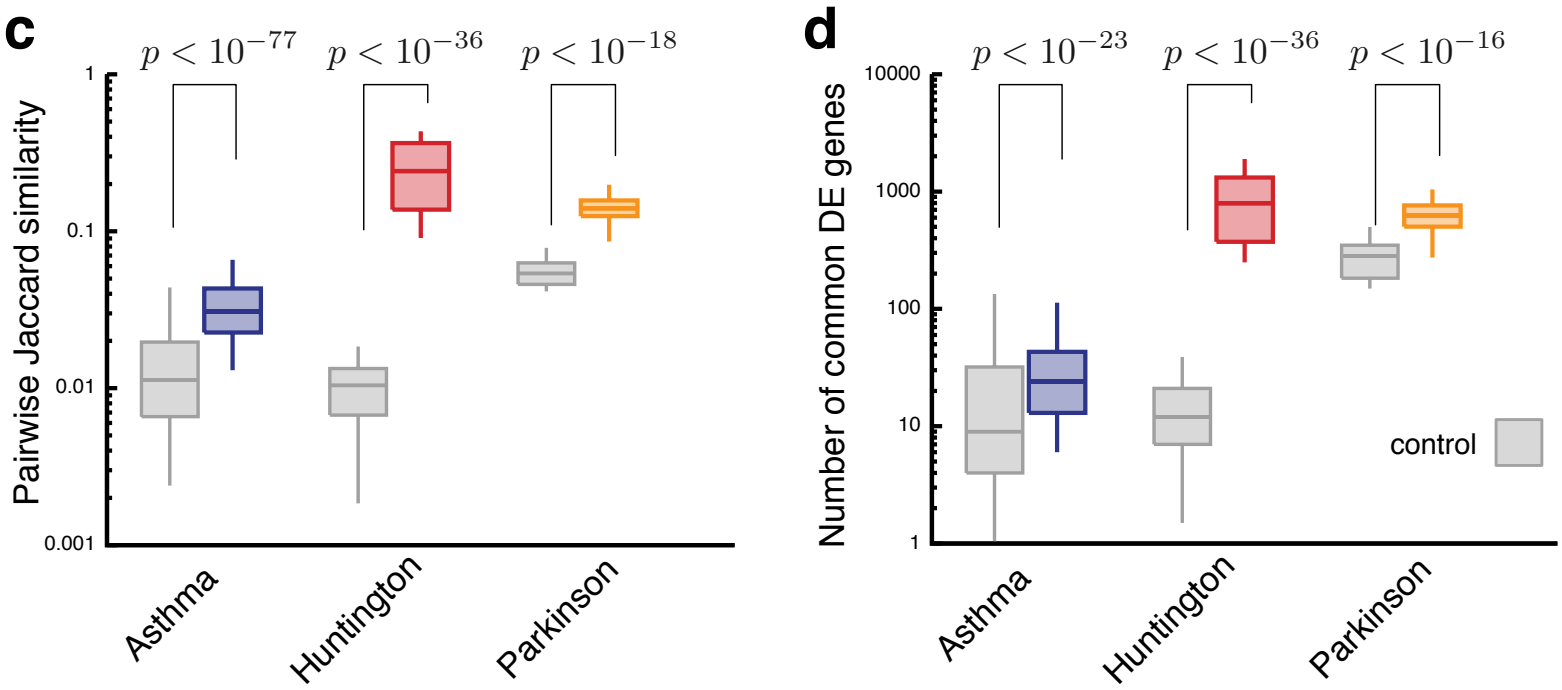
	Samples			
Case	x	x	...	x
Control	c	c	c	c

$$z(gene \text{ in } \mathrm{x}) = \frac{expression_{x}(gene) - \mu_{c}(gene)}{\sigma_{c}(gene)}$$

$$PeeP(x) : \forall gene \; |z(gene \text{ in } \mathrm{x})| > z_{threshold}$$

[Genes that are significantly perturbed in each individual]

Quantifying the heterogeneity using PeePs



The overlap between PeePs of two individuals with the same disease

- is low ($< 30\%$), suggesting high heterogeneity at the transcription level
- is higher than the overlap between the PeePs of healthy subjects

Patient-level gene signatures in Parkinson disease

Identify genes significantly perturbed across patients

Resource	URL
Gene expression data	
NCBI GEO - GSE7621	`ncbi.nlm.nih.gov/geo`
Interactome-based analysis	
PEPPER R package	`emreguney.net/doc/PEPPER.tgz`

Concluding remarks

- Biological processes involve coordinated activity of proteins that physically interact with each other
- Interactome, the network of interactions between proteins, provides a framework for understanding and characterizing biological processes
- Network medicine is an emerging field that aims to use interactome-based analyzes to identify genes associated to diseases and develop novel therapeutics
- Several tools available for the generation, visualization and analysis of the interactome as well as transcriptomic heterogeneity across patients