Ivnstigatng* network-centric plasticity of human diseases using disease-gene prioritization

Emre Guney, Baldo Oliva

Pompeu Fabra University, Dr. Aiguader, 88, Barcelona, 08003, Catalunya, Spain
Emails: {emre.guney|baldo.oliva}@upf.edu

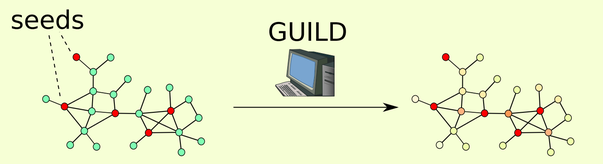
Motivation

A fundamental characteristic of biological systems is tolerance to noise (*e.g., auto-correction of the title of this poster). The ability to counteract both internal mechanistic failures and changes in environmental conditions plays a central role in the survival of the organism. Following the emergence of high-throughput experimental techniques that produce large amount of biological data, several studies have investigated robustness of a complex system in relation to the underlying network topology. Here, we analyze the robustness of network-based disease-gene prioritization methods that use protein-protein interaction (PPI) network to rank genes in diseases. Due to the fact that these methods use the connectivity between genes associated with the disease, we hypothesize that they may serve to distinguish diseases with respect to the predictability of causative genes via investigating the behavior of network-based prioritization under noisy network models.

Methods

We tested the five-fold cross validation Area Under ROC Curve (AUC) performance of five network-based disease gene prioritization algorithms implemented in GUILD framework [1] upon perturbations on the input data sets. Given a network and known disease associations (seeds), GUILD ranks genes in the network by transfering the known annotations through the links of the network. We used the human interactome was created via BIANA tool [2] that integrates data from various PPI repositories such as IntAct, HPRD, DIP, BioGRID, MIPS. We retrieved disease-gene associations for 23 disease phenotypes in OMIM for each of which OMIM entries containing the disease keyword were merged. See [S1,S2,S3] for other tools used in this study.

Network-based disease-gene prioritization

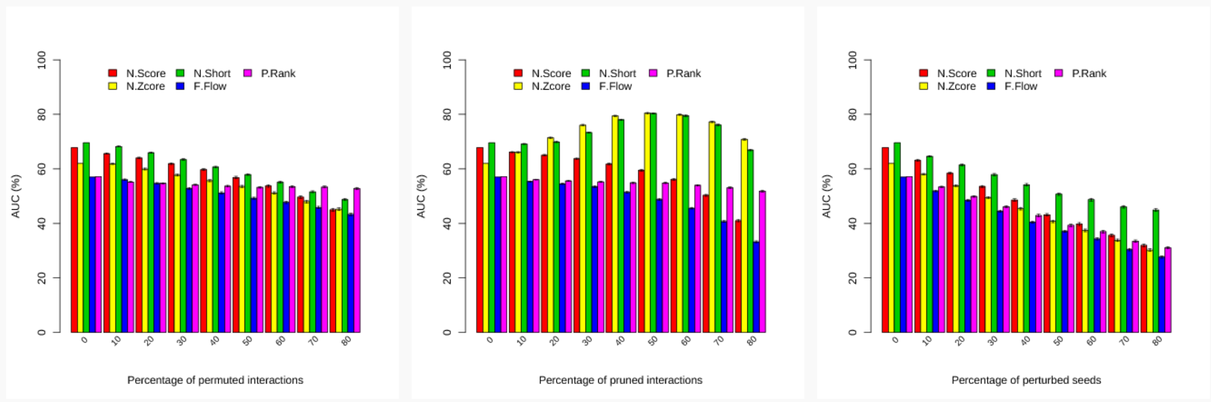


Three types of perturbation
(at various levels; 10% to 80%)
- interaction rewiring
- interaction deletion
- swapping seeds

Five-fold AUC performance (capacity to accurately recover hidden associations) upon perturbations was evaluated

Results

In silico analysis points out to alternative routes connecting the genes involved in diseases



NetZcore and NetShort were consistently effective in ranking genes when interactions were removed, while PageRank was instead more robust against the introduction of false interactions. The interactome might tolerate an interaction removal rate as high as 50%, the point at which methods using alternative paths start suffering from low prediction performance.

Categorizing network-centric plasticity of diseases

Pathophenotypic plasticity is the incapacity to affect the pathophenotype with the use of chemical intervention [4] and we model such possible interventions as perturbations introduced in the interaction network. Based on the average prediction performance the prioritization method had on the networks in which half of the interactions were perturbed, we grouped the diseases into two. A disease was considered robust (i.e. bearing high network-centric plasticty) if the average AUC on perturbed networks was higher than a "critical" AUC and non-robust otherwise. The critical AUC was defined as follows:

$$AUC(p)_{crit} = AUC(p)_{org} - (AUC(p)_{org} - 0.5) / 2$$

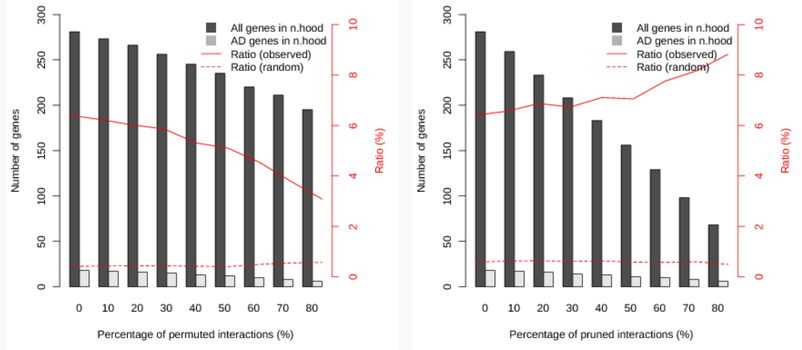
Pathophenotype	AUC (%)				Pathophenotype	AUC (%)			
	org.	crit.	perm.	del.		org.	crit.	perm.	del.
alzheimer	78.3	64.2	62.5	62.8	lung cancer	85.0	67.5	65.8	68.4
anemia	70.3	60.2	56.4	57.9	lymphoma	79.7	64.9	62.3	71.8
asthma	31.7	n/a	34.8	28.4	mental retardation	56.3	53.2	46.6	45.3
ataxia	62.6	56.3	54.2	53.8	myopathy	86.0	68.0	67.3	72.0
breast cancer	76.7	63.4	70.7	75.8	neuropathy	42.5	n/a	38.1	33.2
cardiomyopathy	69.5	59.8	65.0	70.5	**obesity**	72.0	61.0	67.4	70.4
cataract	72.0	61.0	53.9	52.8	**parkinson disease**	80.0	65.0	70.9	78.5
diabetes	61.4	55.7	58.4	63.4	prostate cancer	68.0	59.0	52.7	62.7
epilepsy	62.1	56.1	47.4	47.4	schizophrenia	53.3	51.7	40.9	42.1
hypertension	70.0	60.0	47.7	51.8	spastic paraplegia	31.7	n/a	31.7	31.1
insulin	80.0	65.0	71.6	73.6	systemic lupus erythematosus	86.3	68.2	64.2	72.7
leukemia	84.6	67.3	75.8	81.6					

GO terms enriched in diseases bearing network-centric plasticity

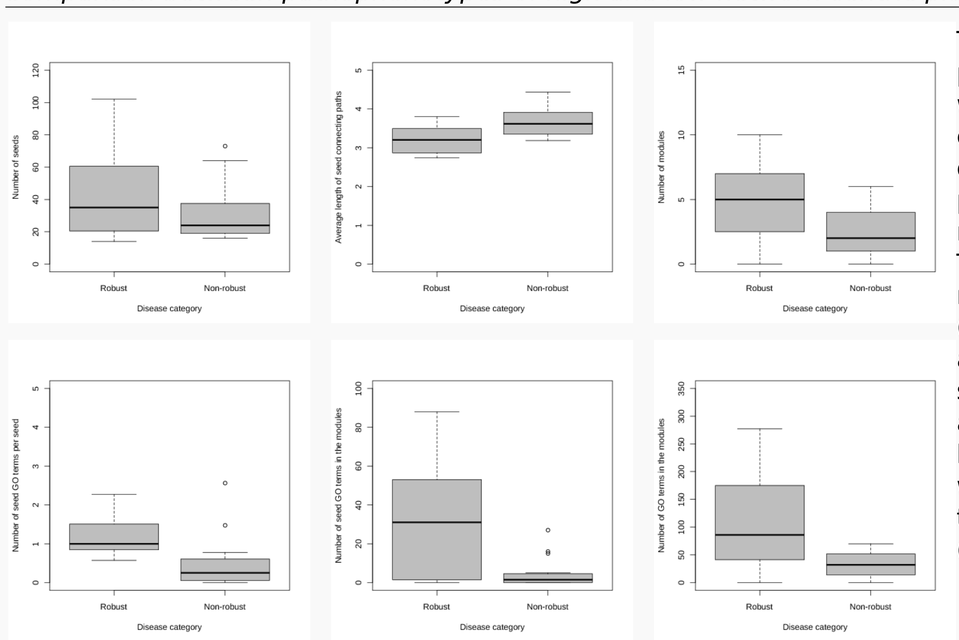
Highly plastic diseases implicated biological processes associated with regulation of cellular metabolism which supported the role of positive and negative feedback mechanisms in defining robustness.

The effect of perturbing interactions on the pathophenotype: the case of Alzheimer's Disease

As the percentage of rewired interactions increased, the ratio of AD-related genes [3] to all genes in the neighborhood of AD-seeds decreased. However, in the case of interaction deletion, this ratio increased suggesting that AD-related genes tended to remain connected with at least one AD-seed in the interaction network.



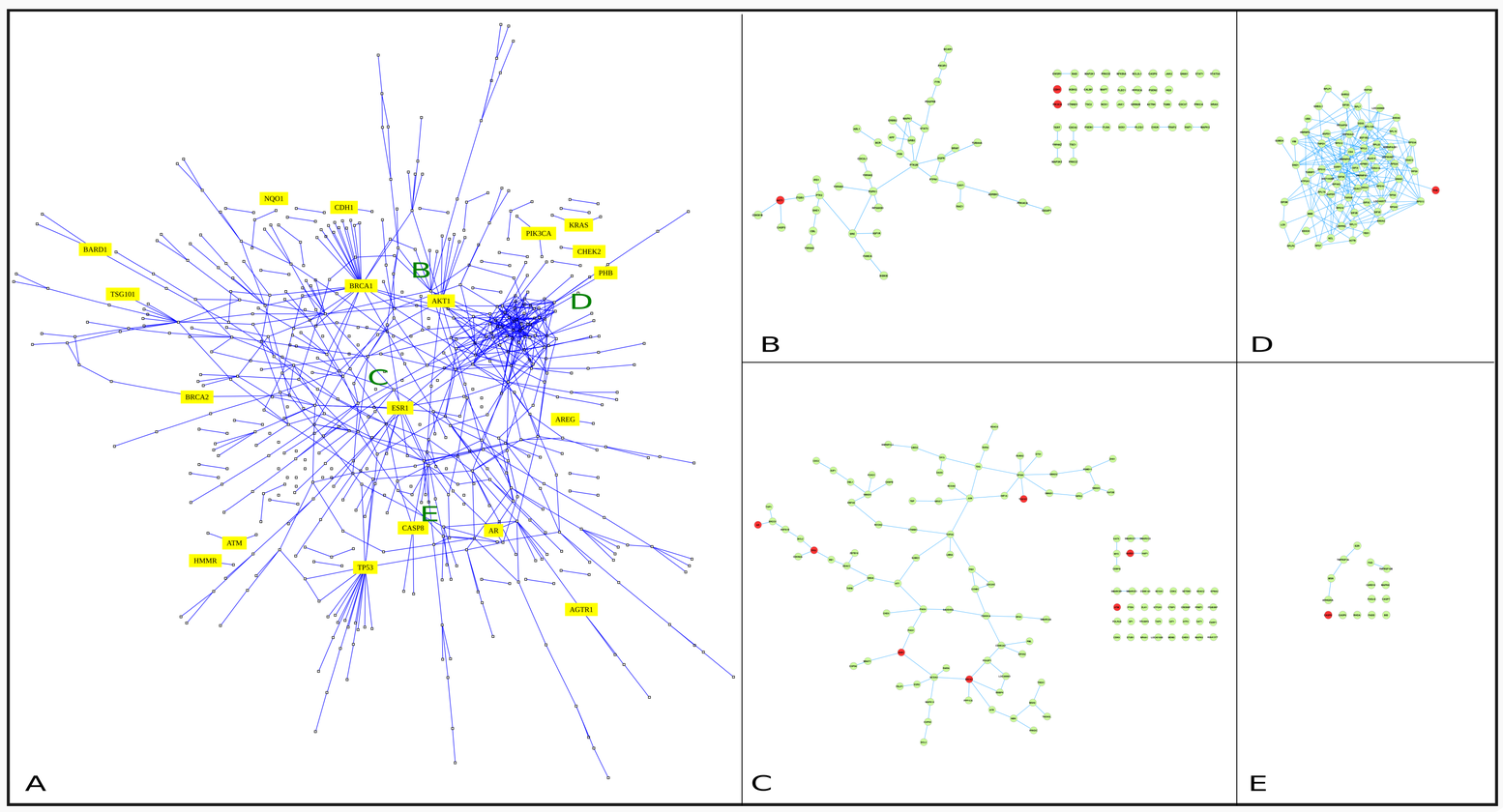
Comparison between pathophenotypes of high and low network-centric plasticity



The seeds of high plasticity diseases were significantly connected to each other with shorter paths compared to rest (*p=0.03*).
They also contained more seed GO terms (functions enriched among seeds) per seed, indicating that a larger number of biological functions were involved in these diseases (*p=0.01*).

Role of redundancy and modularity in network-centric plasticity: the case of breast cancer

Among the perturbed networks in which 80% of the interactions were randomly removed, the one that yielded highest prediction accuracy is visualized below (A). We observed that key proteins involved in breast cancer such as the proteins encoded by BRCA1, TP53, ESR1, AR, AKT1 remained connected in the network. Furthermore, we found evidence for modular segregation of certain biological functions relevant to cancer pathology such as cell signaling (B), cell cycle regulation (C), translational regulation (D) and apoptosis (E).



Bibliography

[1] Guney E. and Oliva B., 2012, *Plos ONE*
[2] Garcia-Garica J. *et al*., 2010, *BMC Bioinformatics*
[3] Krauthammer M. *et al*., 2004, *PNAS*
[4] Kitano H., 2007, *Nature Rev. Drug Disc.*

[S1] ROCR (ROC analysis): Sing T. *et al*., 2005, *Bioinformatics*
[S2] MCL (Modularity analysis): Enright A.J. *et al*., 2002, *NAR*
[S3] FuncAssoc (GO analysis): Berriz GT., 2009, *Bioinformatics*

Special thanks to *Swiss Institute of Bioinformatics* (***SIB***) for awarding a travel fellowship for the presentation of this poster in *ECCB'12*.

Acknowledgements

This work was supported by the following grants: AGAUR-FI, MICINN-FEDER BIO2011-22568; EraSysbio+ (SHIPREC) Euroinvestigación (EUI2009-04018).