

Drug repurposing from a network endopharmacology perspective

Emre Güney, PhD

Hospital del Mar Medical Research Institute ([IMIM](#))
& Pompeu Fabra University ([UPF](#))

Sanofi, Chilly Mazarin, Paris

Oct 18th, 2018

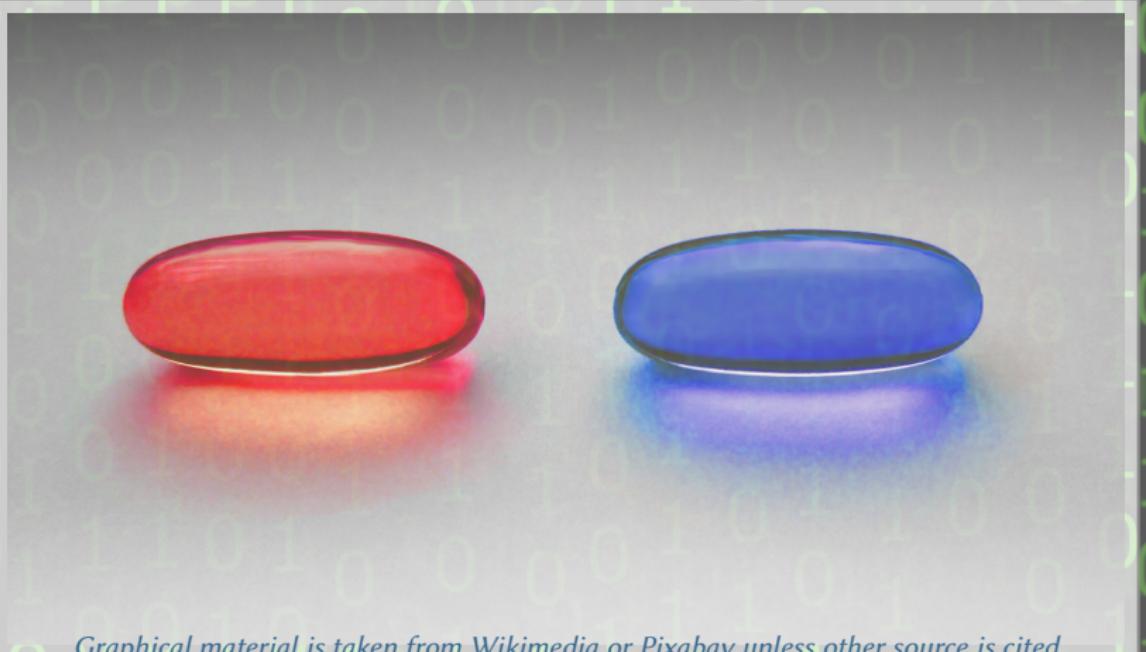


Institut Hospital del Mar
d'Investigacions Mèdiques



RESEARCH
PROGRAMME
ON BIOMEDICAL
INFORMATICS

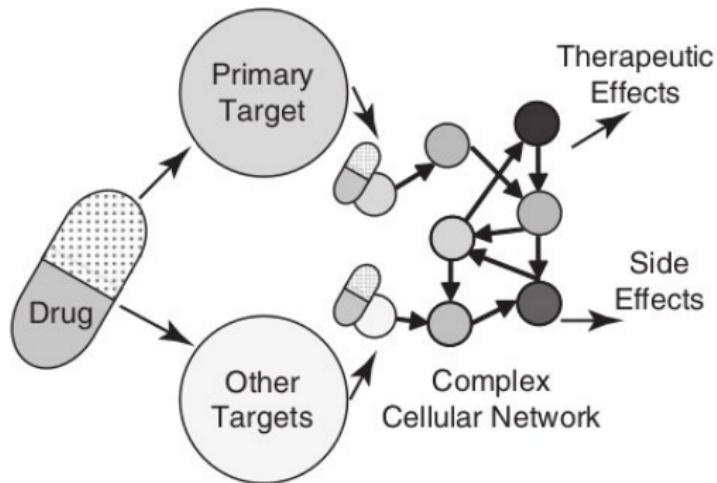




Graphical material is taken from Wikimedia or Pixabay unless other source is cited

Challenges and opportunities in translational medicine

Systems Pharmacology View of Drug Action



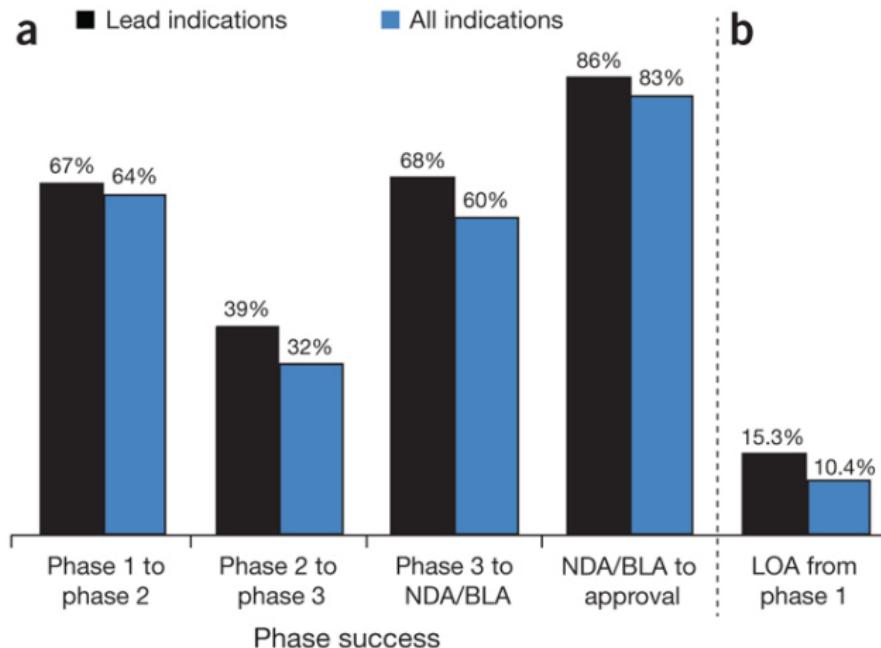
- Difficult to chemically achieve single target specificity
- Acting on multiple targets is likely to be more effective

Berger and Iyengar, 2009, Bioinformatics

Most drugs fail to satisfy desired efficacy and safety requirements

~10%

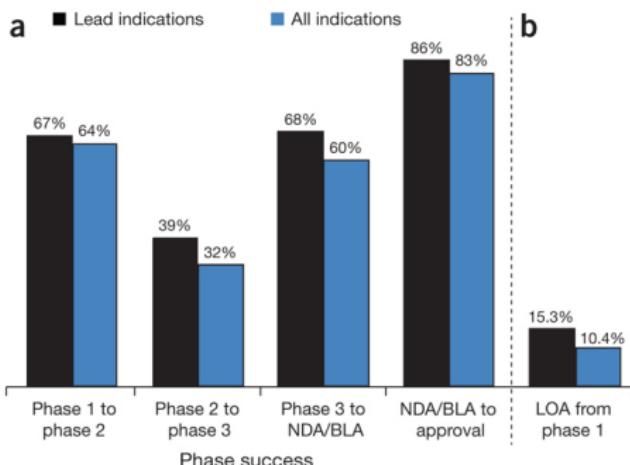
Percentage of drugs that get FDA approval after clinical trials



Few drugs make it to the clinic

~10%

Percentage of drugs that get FDA approval after clinical trials

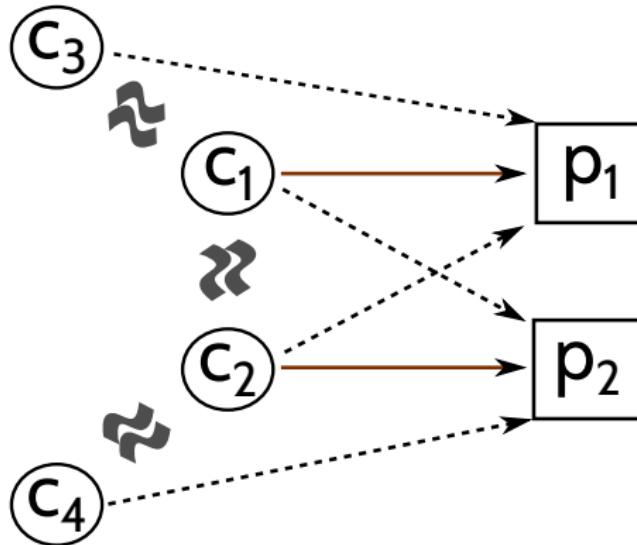


*image from findacure.org.uk

Hay et al., 2014, Nat Biotech

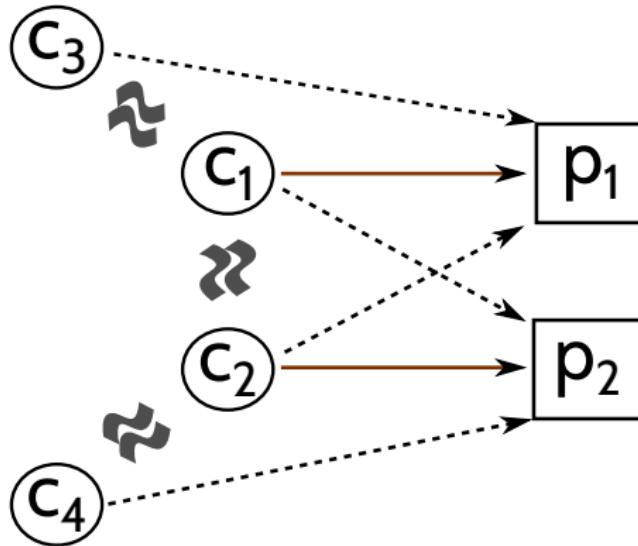
Take 1: Supervised approach (Machine learning)

Similarity-based prediction of drug effect



- chemical formula
- target
- side effect
- gene expression

Similarity-based prediction of drug effect



- chemical formula
- target
- side effect
- gene expression

85-95%
Reported prediction accuracies

Similarity based drug repurposing: Too good to be true?

Vilar and colleagues (2014)

“...*bias* introduced with the information provided in the construction of the *similarity measurement*”

Hodos *et al.* (2016)

“...*reliance on data* existing nearby in pharmacological space”

Reviewer n+1

“...*the paper is not quite complete with respect to the number of papers on the topic*. In fact, *the practical utility* of all these studies is still not well demonstrated in concrete case studies.”

Similarity based drug repurposing: Too good to be true?

Vilar and colleagues (2014)

“...*bias* introduced with the information provided in the construction of the *similarity measurement*”

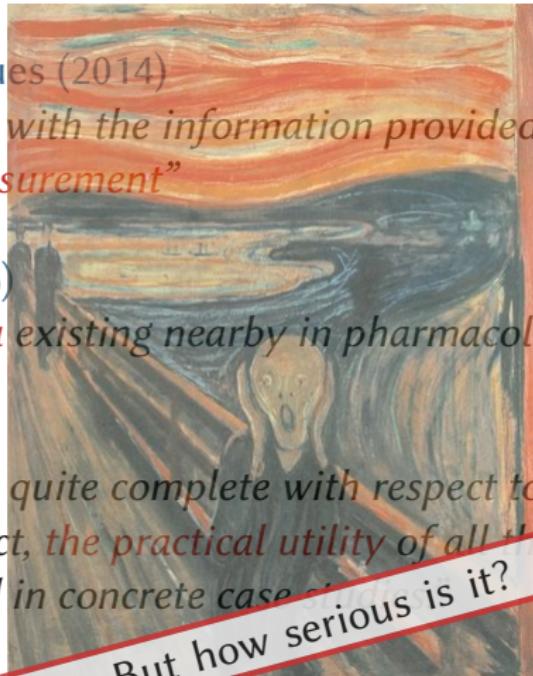
Hodos *et al.* (2016)

“...*reliance on data* existing nearby in pharmacological space”

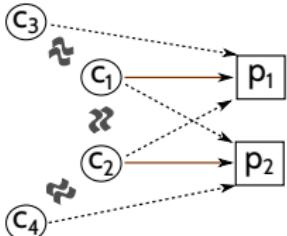
Reviewer n+1

“...*the paper is not quite complete with respect to the number of papers on the topic*. In fact, *the practical utility of all these studies is still not well demonstrated in concrete case*.”

But how serious is it?



Traditional vs disjoint cross validation

| Similarity-based repurposing | Data set | 2-fold cross validation | 2-fold disjoint cross validation |
|--|--|--|--|
| | | Fold 1 | Fold 2 |
|  | $(c_1, p_1, +)$ $(c_3, p_1, -)$ $(c_1, p_2, -)$ $(c_3, p_2, -)$ $(c_2, p_1, -)$ $(c_4, p_1, -)$ $(c_2, p_2, +)$ $(c_4, p_2, -)$ | $(c_1, p_1, +)$ $(c_2, p_2, +)$ $(c_2, p_1, -)$ $(c_1, p_2, -)$ $(c_3, p_1, -)$ $(c_3, p_2, -)$ $(c_4, p_2, -)$ $(c_4, p_1, -)$ | $(c_1, p_1, +)$ $(c_2, p_2, +)$ $(c_1, p_2, -)$ $(c_2, p_1, -)$ $(c_4, p_1, -)$ $(c_3, p_1, -)$ $(c_4, p_2, -)$ $(c_3, p_2, -)$ |

Define non-overlapping (disjoint) drug groups

D : data set containing drug-disease pairs, c : drug, p : disease,
 l : label (1 if c is known to be indicated for p , 0 otherwise), k : number of cross validation folds,
 $fold$: dictionary containing the fold index of each drug-disease pair
 $i := \text{random}([0, 100])$
 $fold := \{\}$
for each $(c, p, l) \in D$ **do**
 $sum := 0$
 for each $x \in \text{characters}(c)$ **do**
 $sum := sum + \text{to_integer}(x)$
 $fold(c, p) := \text{modulo}(sum + i, k)$
return $fold$

Similarity-based classifier

Drug i defined as a binary vector for a given feature

$$X_i^f = [x_1^f, x_2^f, \dots, x_n^f]^T \quad f: \begin{array}{l} \bullet \text{ chemical substructures} \\ \bullet \text{ targets} \\ \bullet \text{ side effects} \end{array}$$

Similarity between two drugs i and j are defined by the Pearson product-moment correlation

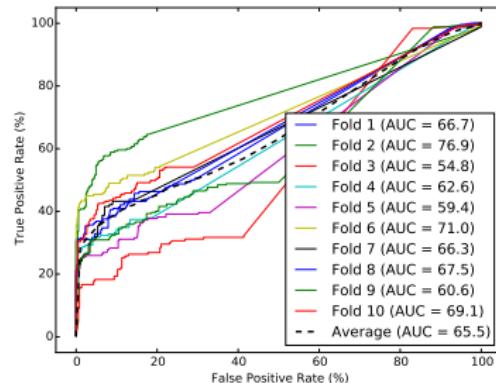
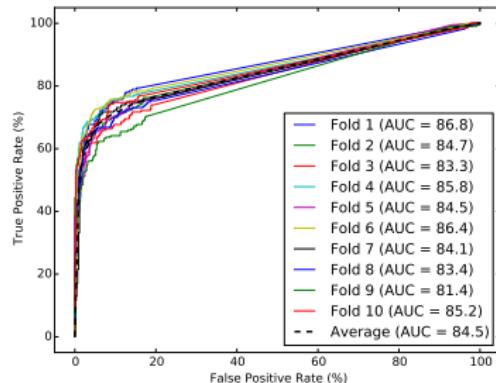
Train logistic regression classifier using similarity-based drug-disease score s_{ip}^f defined as

$$s_{ip}^f = \sum_{j \in \text{NearestNeighbors}(i)} \rho_{ij}^f * I_{jp}$$

where

$$I_{jp} = \begin{cases} 1, & \text{if } j \text{ is indicated for } p \\ 0, & \text{otherwise} \end{cases}$$

Models perform poorly on drugs they have not seen before

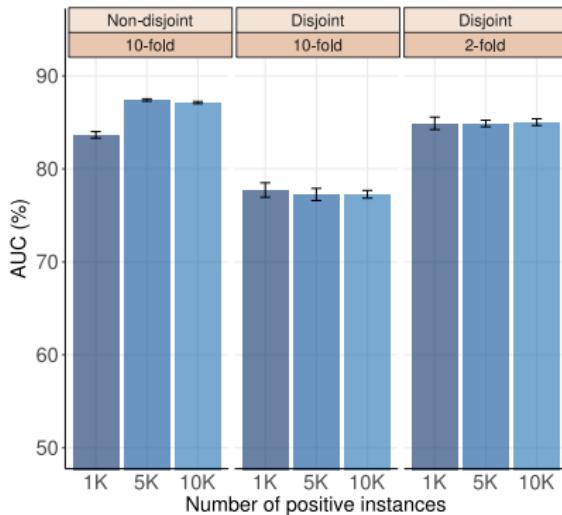


| Disjoint folds | Mean AUC (%) | Mean AUPRC (%) |
|----------------|--------------------|--------------------|
| No | 84.1 (± 0.3) | 83.7 (± 0.3) |
| Yes | 65.6 (± 0.5) | 62.8 (± 0.5) |

Data set: 2,229 drug-disease associations covering 578 diseases and 536 drugs (> 250K possible pairs)

The diversity of the training set has a strong effect on the accuracy

| Number of folds | Mean AUC (%) | Mean AUPRC (%) |
|-----------------|-------------------|-------------------|
| 2 | 80.7 (\pm 0.3) | 79.3 (\pm 0.3) |
| 5 | 73.6 (\pm 0.7) | 71.9 (\pm 0.7) |
| 10 | 65.6 (\pm 0.5) | 62.8 (\pm 0.5) |
| 20 | 59.1 (\pm 0.6) | 57.0 (\pm 0.3) |



Similarity based drug repurposing: Too good to be true?

Vilar and colleagues

“...*bias introduced with the similarity measure*”

Hodos *et al.* (2016)

“...*reliance on data* ex

Reviewer n+1

“...*the paper is not quite on the topic*. In fact, *the practical utility* of all those studies is still not well demonstrated in concrete case”



“...*in the construction of a biological space*”

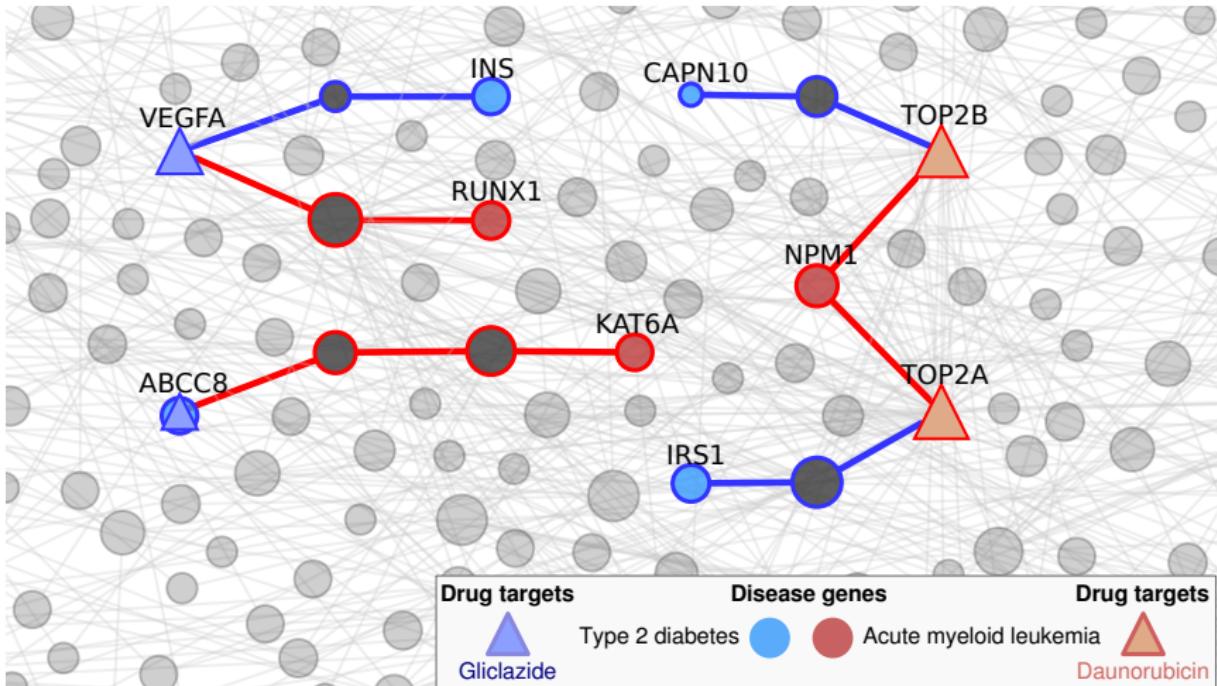
the number of papers

of all those studies is still not well demonstrated in concrete case

How about interpretability?

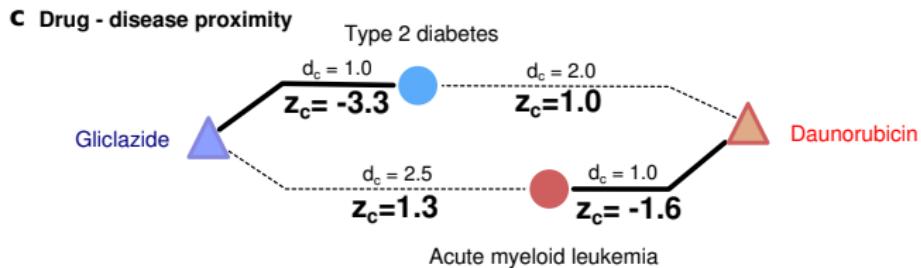
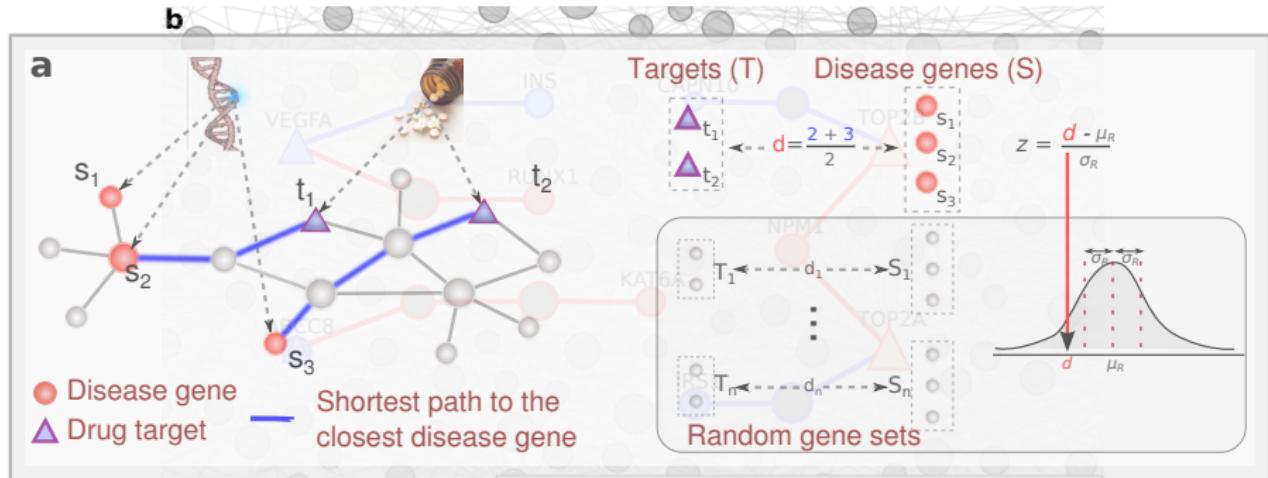
Take 2: Unsupervised approach (Network medicine)

Modeling drug effect via interactome-based proximity



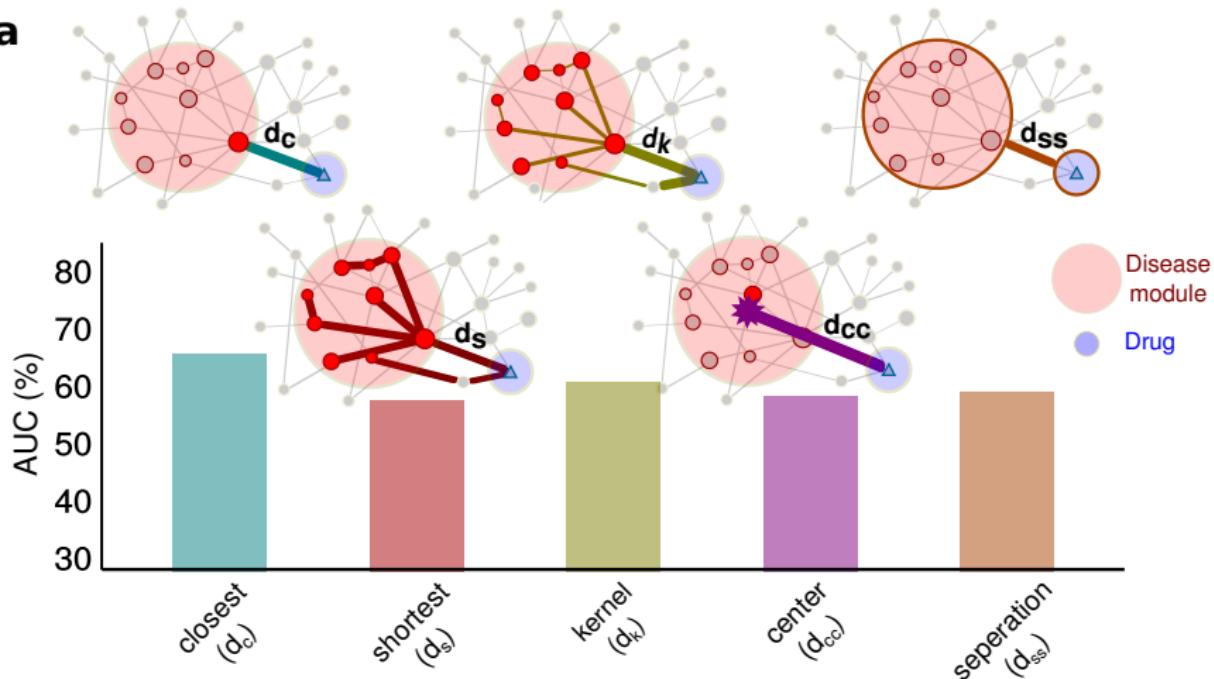
Guney et al., 2016, Nat Comm

Modeling drug effect via interactome-based proximity



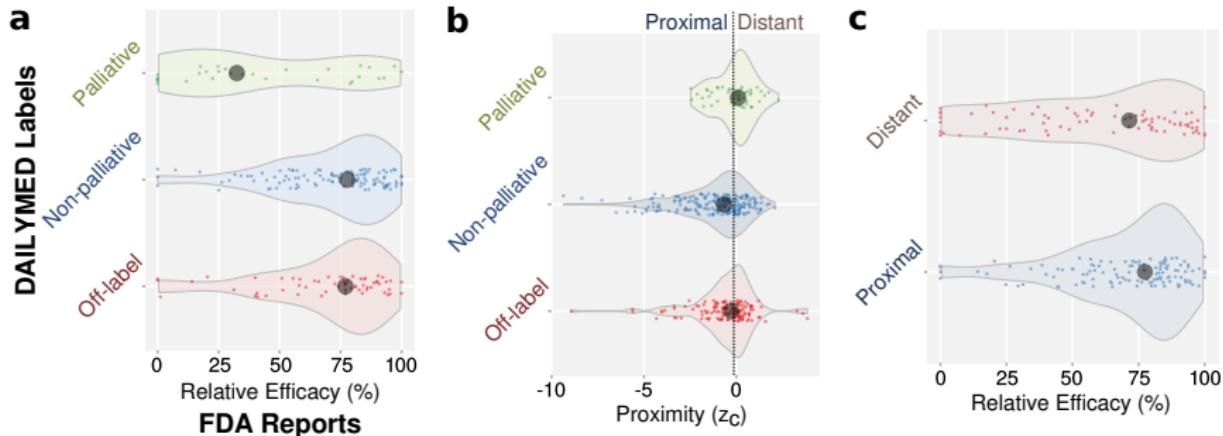
Drugs do not target the disease module as a whole

a



Guney et al., 2016, Nat Comm

Proximity is a good proxy for drug's therapeutic effect



Proximal drug-disease pairs are more likely to correspond to effective treatments

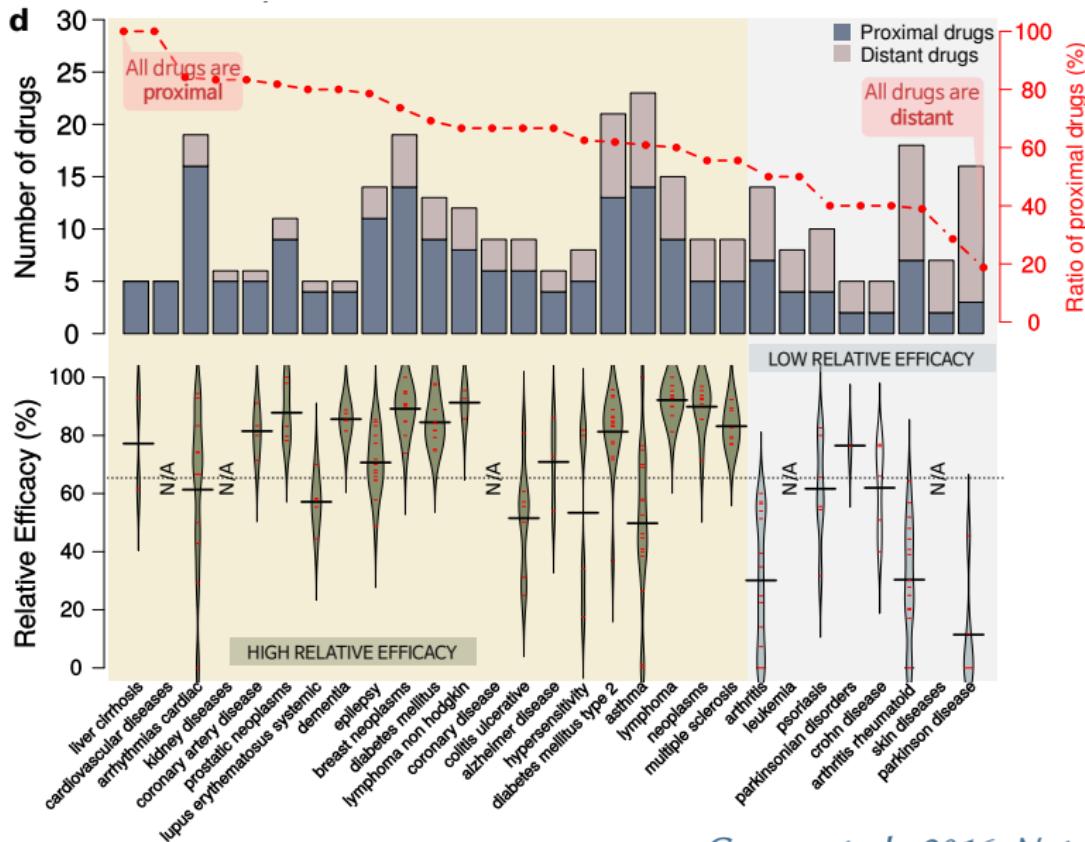
$$P = 7.3 \times 10^{-5}$$
$$P = 7.6 \times 10^{-4}$$
$$(n = 50, 219, 133)$$

$$P = 4.0 \times 10^{-5}$$
$$P = 0.02$$
$$(n = 50, 219, 133)$$

$$P = 0.04$$
$$(n = 237 \text{ vs } 165)$$

Guney et al., 2016, Nat Comm

Proximity highlights treatment bottlenecks



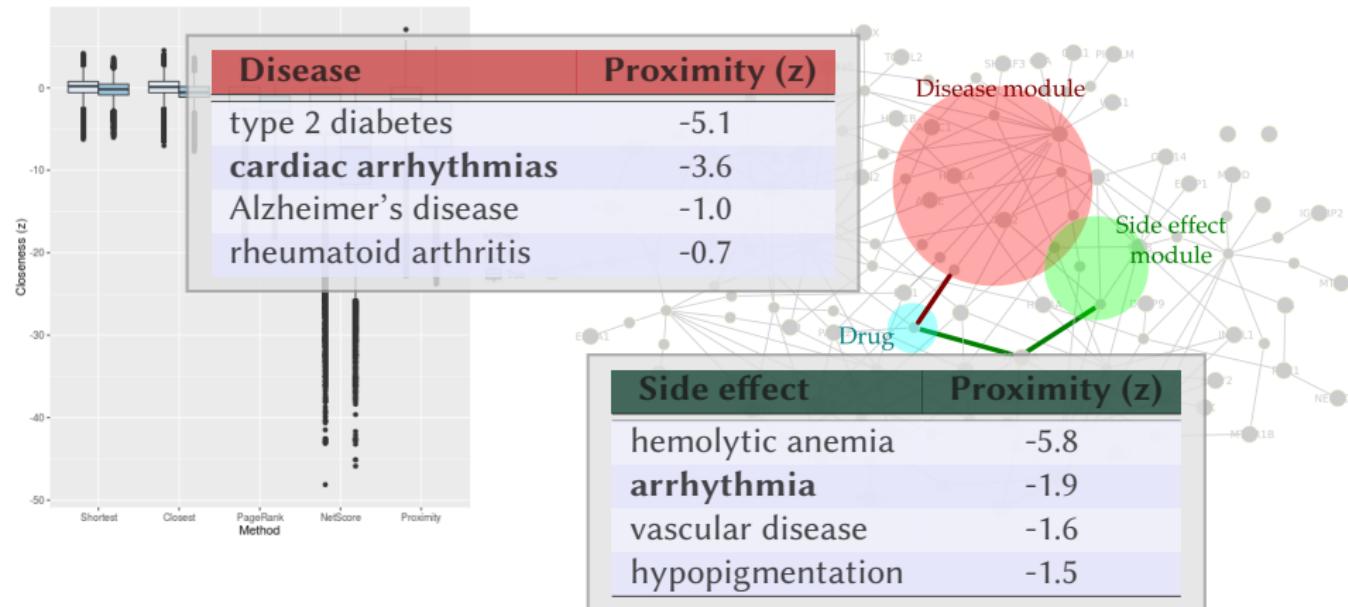
Guney et al., 2016, Nat Comm

Drug repurposing using proximity

| Drug | Description | Phenotype | Proximity (z) |
|---|--|--|---------------|
| Repurposed uses | | | |
| Plerixafor | repurposed to treat non-Hodgkin's lymphoma | non-Hodgkin's lymphoma | -2.4 |
| Ropinirole | repurposed to treat restless legs syndrome | restless legs syndrome | -1.1 |
| Sildenafil | repurposed to treat erectile dysfunction | erectile dysfunction | -1.0 |
| Meta data based observations | | | |
| Dospirenone | confer protection against endometrial cancer | endometrial cancer | -1.1 |
| Levonorgestrel | confer protection against endometrial cancer | endometrial cancer | -1.6 |
| Failures due to lack of efficacy | | | |
| Tabalumab | showed lack of efficacy for systemic lupus erythematosus | systemic lupus erythematosus | 1.8 |
| Preladenant | discontinued trials for parkinson due to lack of improvement compared to placebo | parkinson disease | 0.2 |
| Iniparib | failed to achieve improvement while being tested for squamous non-small-cell lung cancer | squamous cell cancer | 0.0 |
| Failures due to adverse effects | | | |
| Semagacestat | failed trials due to worsening Alzheimer's disease | Alzheimer's disease | -5.6 |
| Terfenadine | withdrawn due to inducing cardiac arrhythmia | cardiac arrhythmia arrhythmia (side effect) | -2.2 -2.6 |

Guney et al., 2016, Nat Comm

Predicting the directionality of drug effect using interactome-based modeling



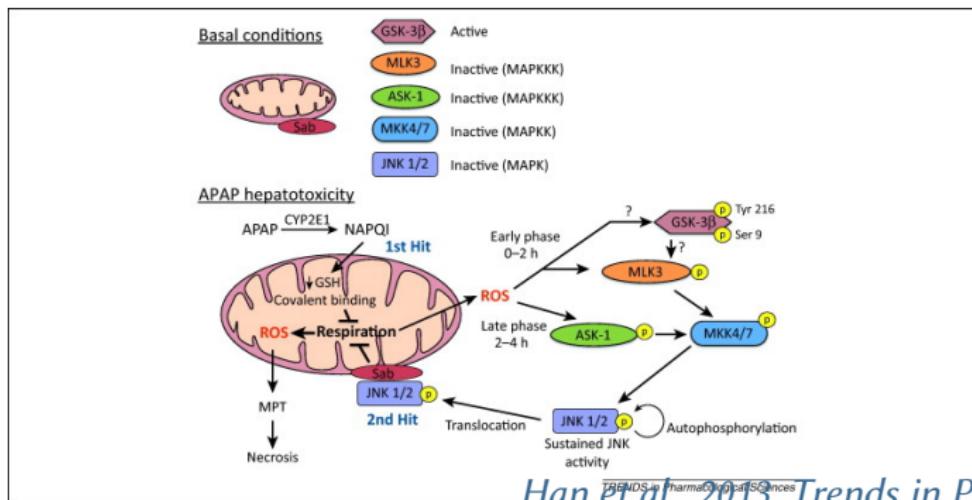
*Guney et al., 2016, Nat Comm
Guney, 2017, Workshop on Complex Networks*

APAP-induced hepatotoxicity

APAP (a.k.a. paracetamol, acetaminophen) overdose alone is estimated to contribute around 40% of all acute liver failure cases in the USA

Ostapowicz et al., 2014, Ann. of Int. Medic.

- The dose that induces hepatotoxicity in a person varies
- The intrinsic response to APAP differs across patients substantially



Adverse outcome pathway discovery through interactome-based analysis

Problem: Find the pathway(s) involved in APAP-induced hepatotoxicity

Given interaction network $G(V, E)$, genes associated to a phenotype of interest (i.e. liver injury) $S \in V$, genes involved in drug response (i.e. targets or differentially expressed genes) $T \in V$:

find $G'(V', E')$ s.t.

$$S \subset V', |V'| \text{ is minimized and } |V' \cap T| \text{ is maximized}$$

Solution: Node-weighted Steiner tree connecting seeds

- A tree that connects given seed nodes through other nodes ($S \in V'$)
- Minimizes the cost associated with including differentially expressed genes ($\operatorname{argmin}_{V'} |V'|$ & $\operatorname{argmax}_{V'} |V' \cap T|$)
- Set the costs such that drug response genes are easier to be included ($w_i = 10^{-|FC_i|}$)

Adverse outcome pathway discovery through interactome-based analysis

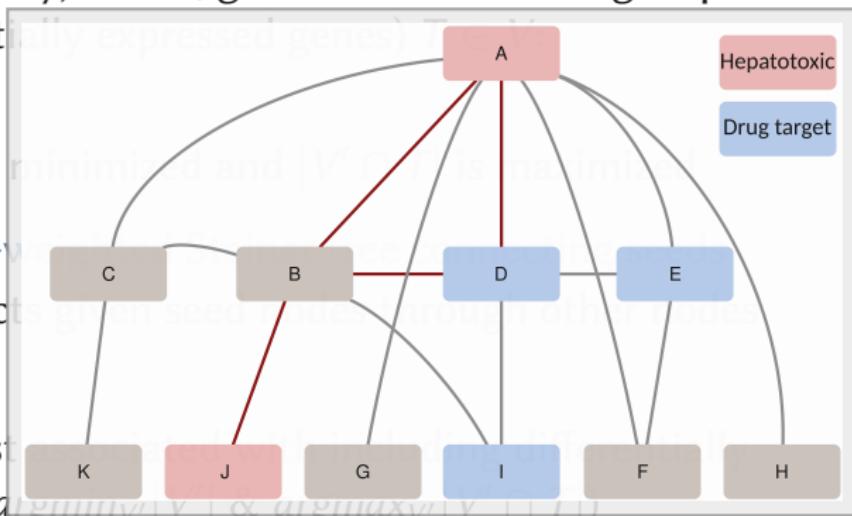
Problem: Find the pathway(s) involved in APAP-induced hepatotoxicity
Given interaction network $G(V, E)$, genes associated to a phenotype of interest (i.e. liver injury) $S \in V$, genes involved in drug response (i.e. targets or different)

find $G'(V', E')$ s.t.

$$S \subset V', |V'| \text{ is}$$

Solution: Node-

- A tree that connects $S \in V'$
- Minimizes the cost expressed genes (a_i)
- Set the costs such that drug response genes are easier to be included ($w_i = 10^{-|FC_i|}$)



Identifying hepatotoxicity pathways in the interactome – Data sets

Liver-specific interaction network

- Protein interaction data from InBioMap (*Score > 0.15*)
Li et al., 2017, Nat Methods
- Filter hub proteins and proteins that are not expressed in liver using GTEx (*TPM > 1*)
GTEx Consortium, 2013, Nat Genetics

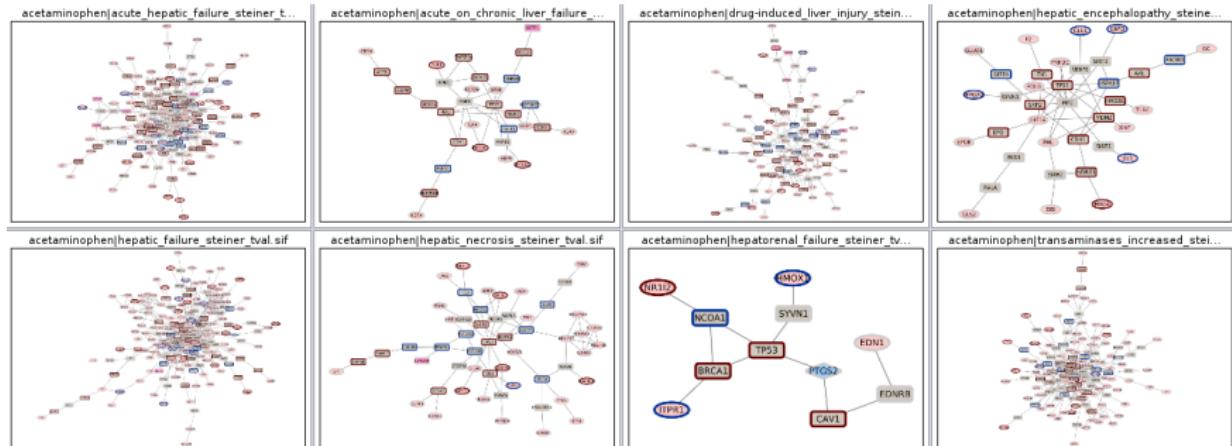
APAP targets

- Drug-target interactions from DrugBank (i.e. pharmacological targets and enzymes)
Wishart et al., 2017, NAR

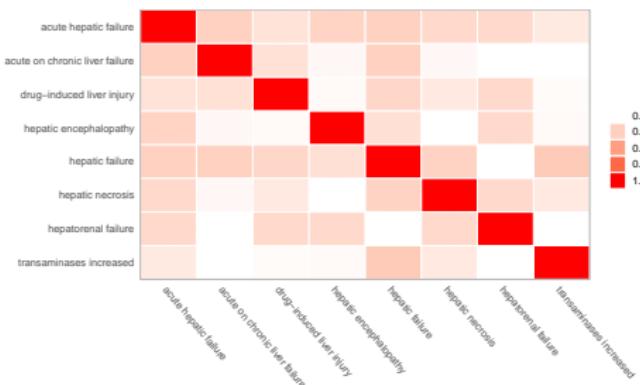
Gene expression

- APAP-induced gene expression (high dose, 24h from TG-GATES)
Igarashi et al., 2015, NAR

Defining adverse outcomes in APAP-induced liver injury



| Phenotype | # of genes |
|--------------------------------|------------|
| Acute hepatic failure | 125 |
| Acute on chronic liver failure | 25 |
| Drug-induced liver injury | 86 |
| Hepatic encephalopathy | 31 |
| Hepatic failure | 173 |
| Hepatic necrosis | 35 |
| Hepatorenal failure | 5 |
| Transaminases increased | 99 |



Patient-level heterogeneity

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
Schizophrenia



2. NEXIUM (esomeprazole)
Heartburn



3. HUMIRA (adalimumab)
Arthritis



4. CRESTOR (rosuvastatin)
High cholesterol



5. CYMBALTA (duloxetine)
Depression



6. ADVAIR DISKUS (fluticasone propionate)
Asthma



7. ENBREL (etanercept)
Psoriasis



8. REMICADE (infliximab)
Crohn's disease



9. COPAYONE (glatiramer acetate, natalizumab, neliastatimab, pegfilgrastim)
Multiple sclerosis, MS, neuastim, aplastic anemia



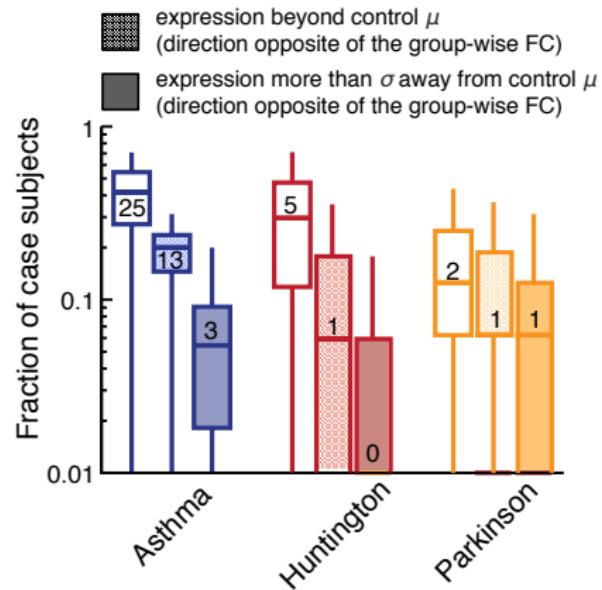
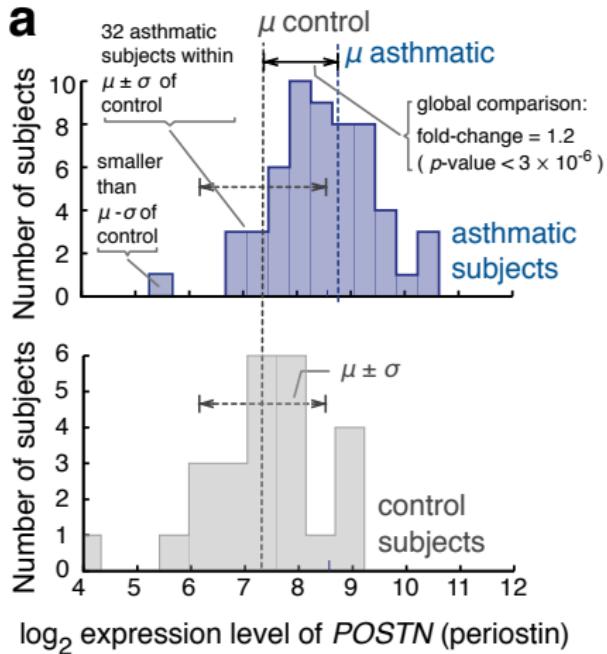
How about heterogeneity?

Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78t.

Schork, 2015, Nature

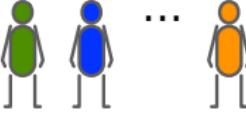
Take 3: Personalized approach (Semi-supervised)

Group-wise differentially expressed genes do not capture transcriptomic heterogeneity

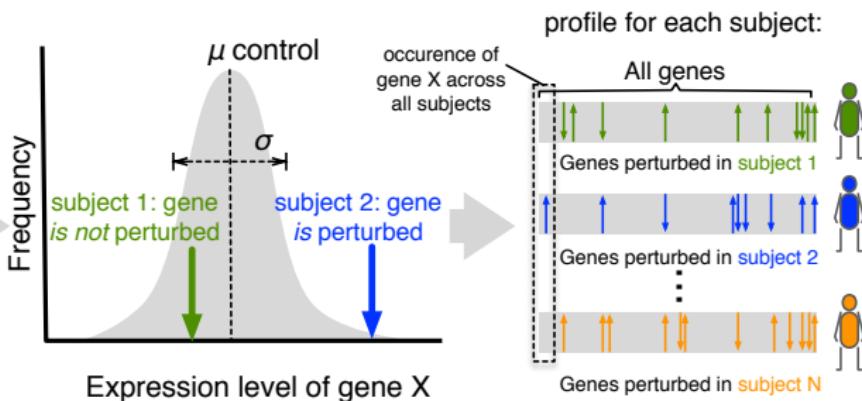


Menche et al., 2017, Npj Sys Bio & App

PeeP: PErsonalized Expression Profile

Individual case subjects

 vs.

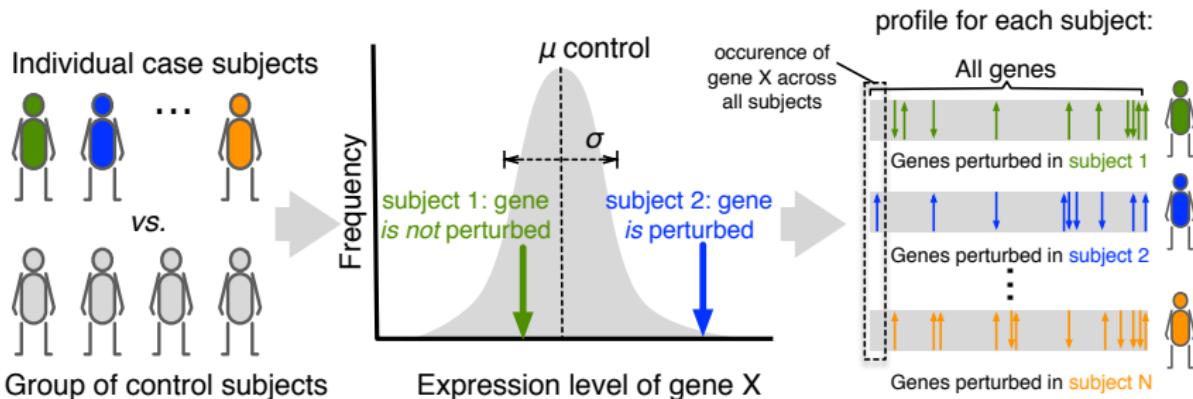
 Group of control subjects



| | Samples | | | |
|---------|---------|---|-----|---|
| Case | x | x | ... | x |
| Control | c | c | c | c |

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

PeeP: PErsonalized Expression Profile



| | Samples | | | |
|---------|---------|---|-----|---|
| Case | x | x | ... | x |
| Control | c | c | c | c |

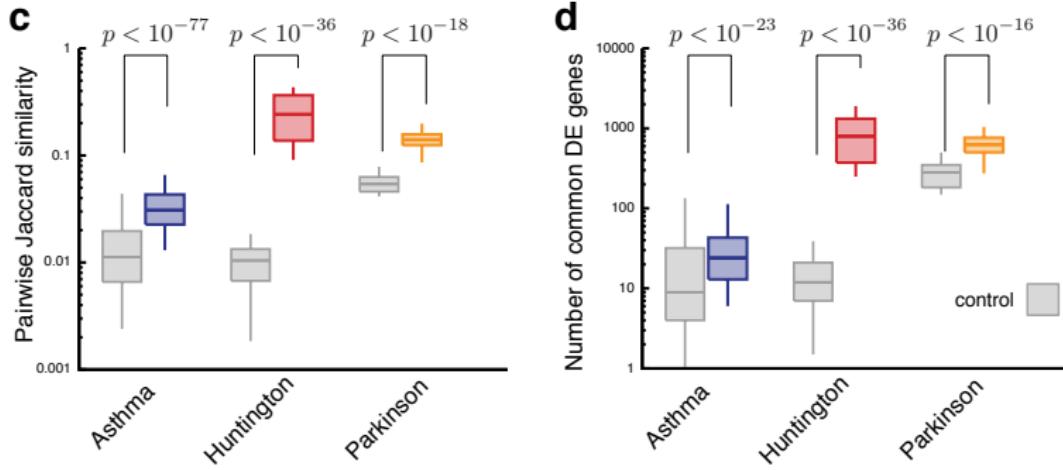
$$z(\text{gene in } \mathbf{x}) = \frac{\text{expression}_{\mathbf{x}}(\text{gene}) - \mu_c(\text{gene})}{\sigma_c(\text{gene})}$$

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

[Genes that are significantly perturbed in each individual]

Menche et al., 2017, *Npj Sys Bio & App*

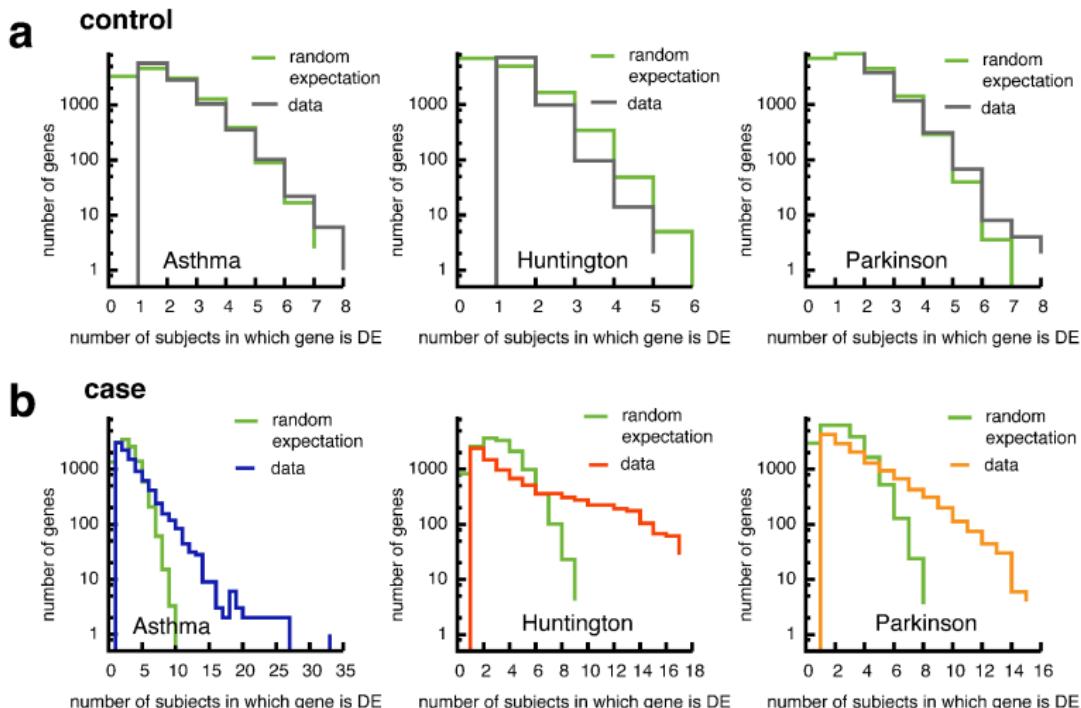
Quantifying the heterogeneity using PeePs



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

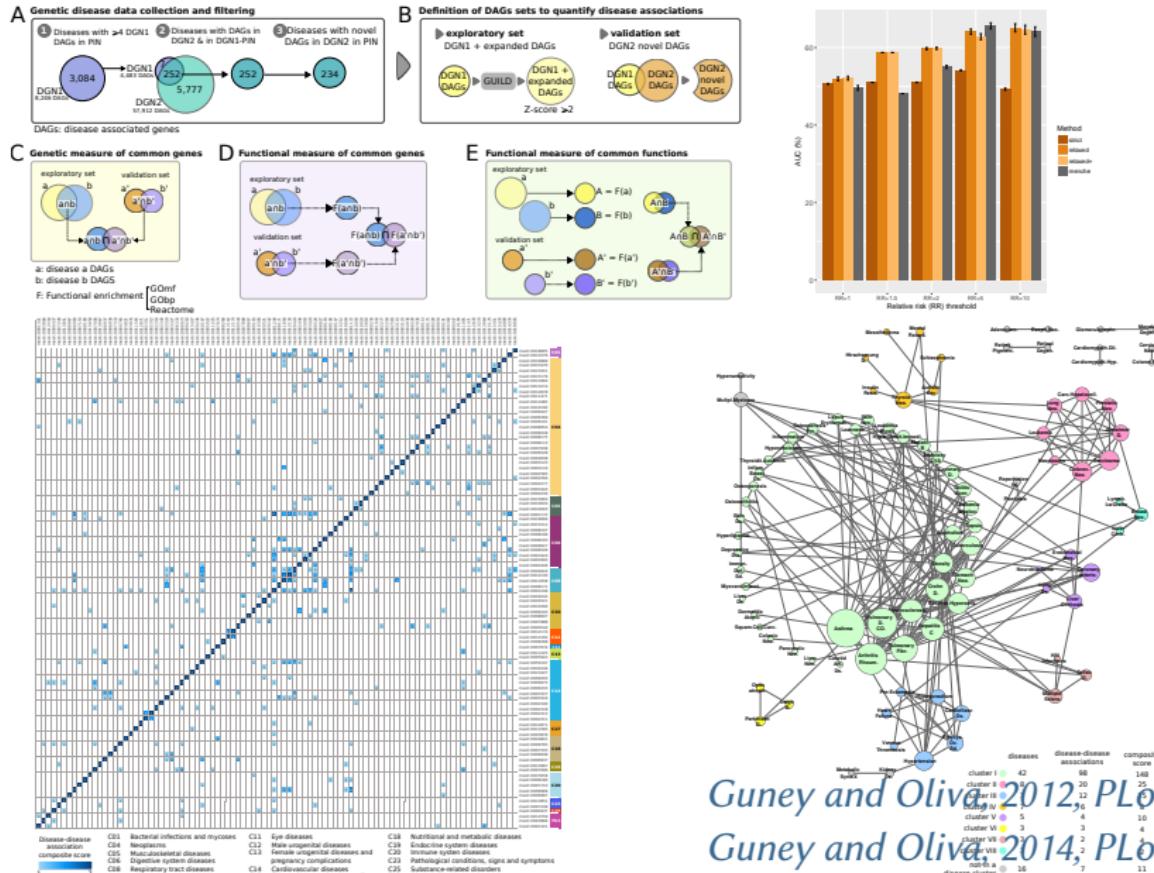
Can PeePs predict disease status?



The number of shared genes between case subjects significantly exceeds the random expectation *Menche et al., 2017, Npj Sys Bio & App*

*Take 4: Leveraging disease-disease
relationships (Endophenotyping)*

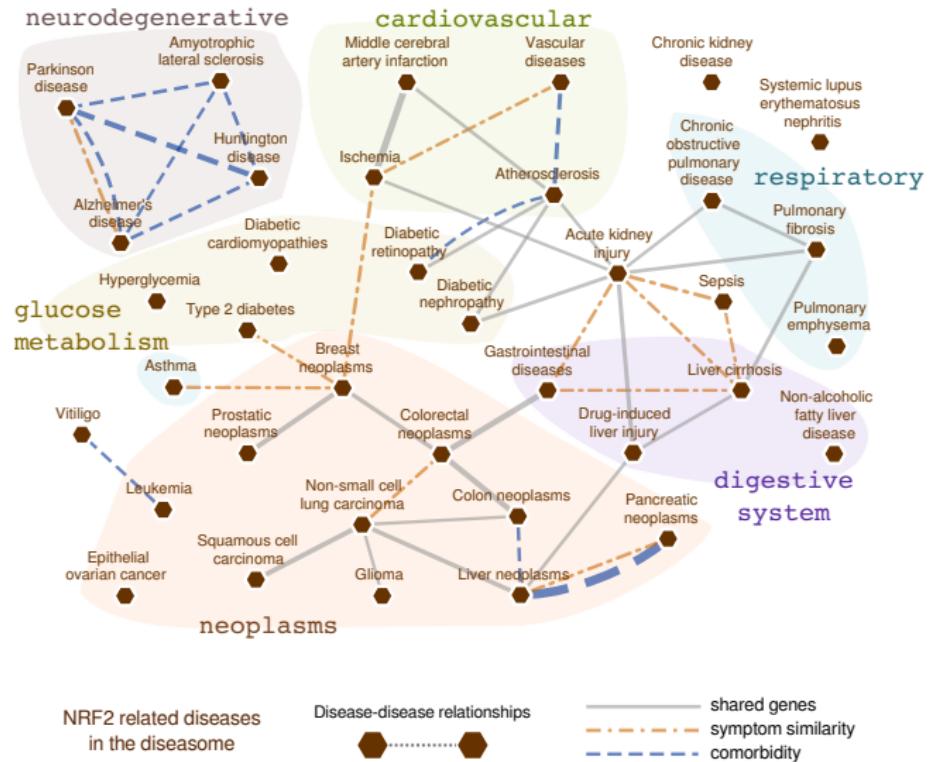
Diseasome+: Shared functions via interactome-based expansion



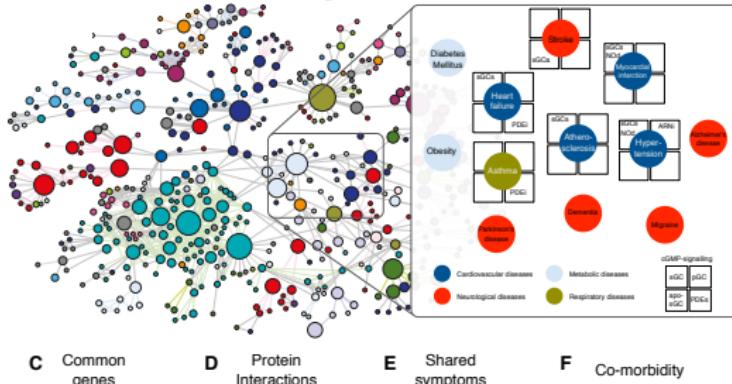
Guney and Oliveira 2012, PLOS ONE
Guney and Oliveira 2014, PLOS ONE

Rubio-Perez et al., 2017, Sci Rep³³

Diseasome++: Shared genetic, functional and clinical signatures across diseases



Leveraging diseasesome for drug repurposing



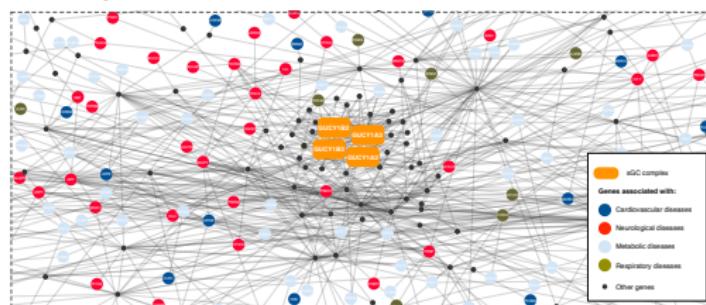
C Common genes

D Protein interactions

E Shared symptoms

F Co-morbidity

G sGC neighborhood in the human interactome



Pathway level connections between T2D and AD

Alzheimer disease

Type 2 diabetes

Pathway

| Pathway | z |
|--|------|
| synthesis of phosphatidylcholine | -3.3 |
| serotonin receptors | -3.3 |
| adenylyl cyclase inhibitory pathway | -2.2 |
| IL-6 signaling | -2.1 |
| the NLRP3 inflammasome | -2.1 |
| regulation of insulin secretion by acetyl-choline | -2.1 |

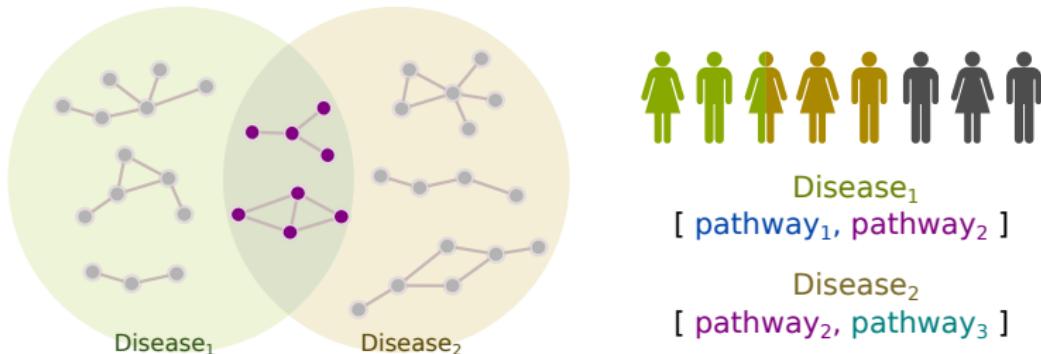
Donepezil

Pathway

| Pathway | z |
|---|------|
| inwardly rectifying K ⁺ channels | -9.0 |
| ABC family proteins mediated transport | -8.5 |
| Inhibition of voltage gated Ca ²⁺ channels via G beta gamma subunits | -4.3 |
| GABA _B receptor activation | -4.1 |
| regulation of insulin secretion by acetyl-choline | -3.3 |
| Na ⁺ /Cl ⁻ dependent neurotransmitter transporters | -3.3 |

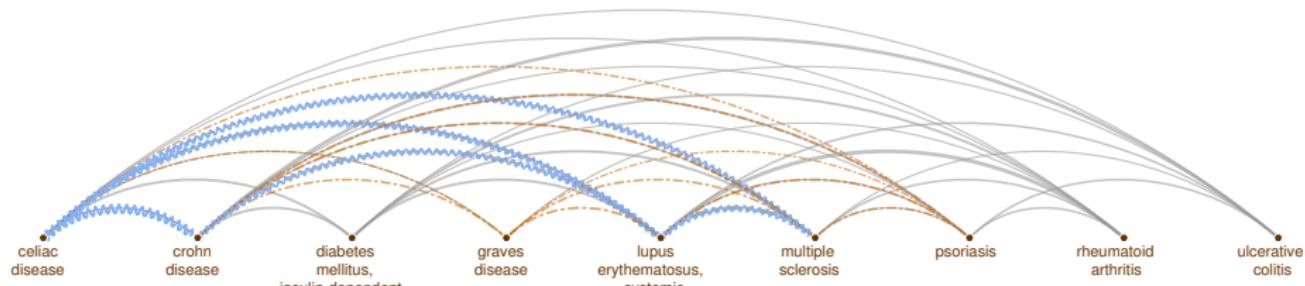
Glyburide

Endophenotyping using common pathways

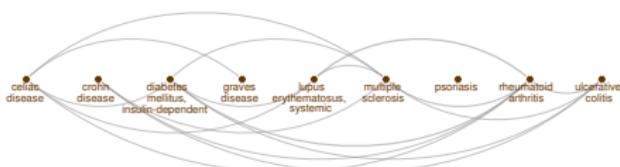


- Perturbations across various biological pathways give rise to diseases
- Existing methods focus on identifying those pathways via enrichment of perturbed genes
- Often only a handful of genes within a pathway that are involved in the disease (*Menche**, *Guney** *et al.*, 2017, *Npj Sys Bio and App*)

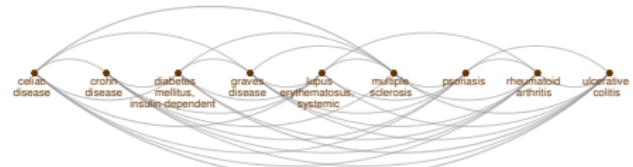
Pathway level commonalities between autoimmune disorders



Genetic (shared genes) and clinical (symptom & comorbidity) similarities



(Overlap between genes)



(Proximity of genes)

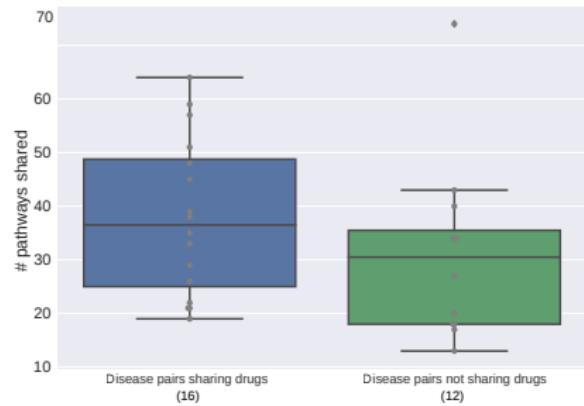
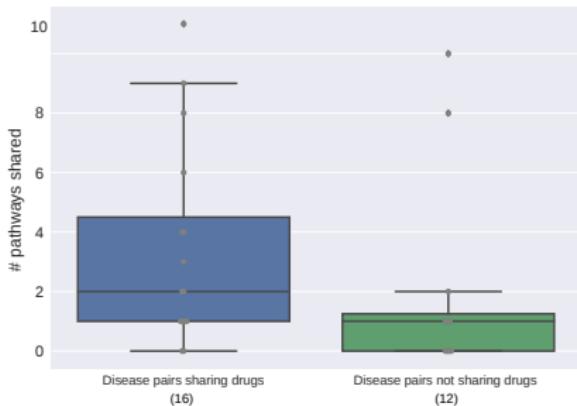
Shared pathways

Aguirre-Plans et al., 2018, Pharmaceuticals

Pathway level commonalities between autoimmune disorders

| Pathway | # of shared diseases | |
|--|----------------------|-----------|
| | overlap | proximity |
| interferon gamma signaling | 5 | 8 |
| costimulation by the CD28 family | 5 | 7 |
| cytokine signaling in immune system | 5 | 7 |
| translocation of ZAP-70 to immunological synapse | 5 | 6 |
| phosphorylation of CD3 and TCR zeta chains | 5 | 6 |
| PD1 signaling | 5 | 4 |
| IL-6 signaling | 4 | 8 |
| generation of second messenger molecules | 4 | 6 |
| TCR signaling | 4 | 6 |
| signaling by ILs | 3 | 9 |
| immune system | 3 | 7 |
| downstream TCR signaling | 3 | 7 |
| interferon signaling | 3 | 7 |
| adaptive immune system | 3 | 3 |

Pathway level commonalities between autoimmune disorders

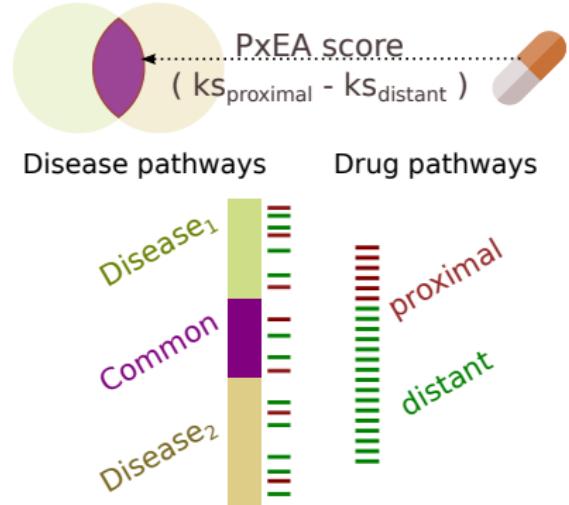
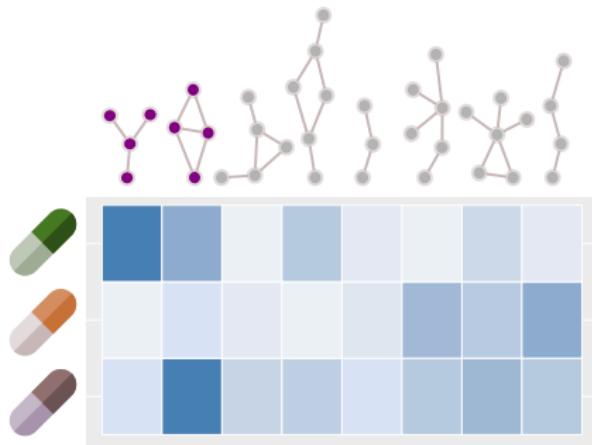


Diseases targeted by the same drugs tend to share more pathways than the rest

Aguirre-Plans et al., 2018, Pharmaceuticals

Take Five: Targeting common pathways (Endopharmacology)

PxEA: ProXimal pathway Enrichment Analysis



A general framework for calculating the enrichment of **group of pathways** ranked with respect to *proximity*

Aguirre-Plans et al., 2018, Pharmaceuticals

PxEA: Enrichment score

D : the pathways ranked with respect their proximity to drug targets

p_i : the pathway in consideration within D

C : the set of pathways of interest

$$ES(D, C) = \sum_{p_i \in P} X_i$$

where,

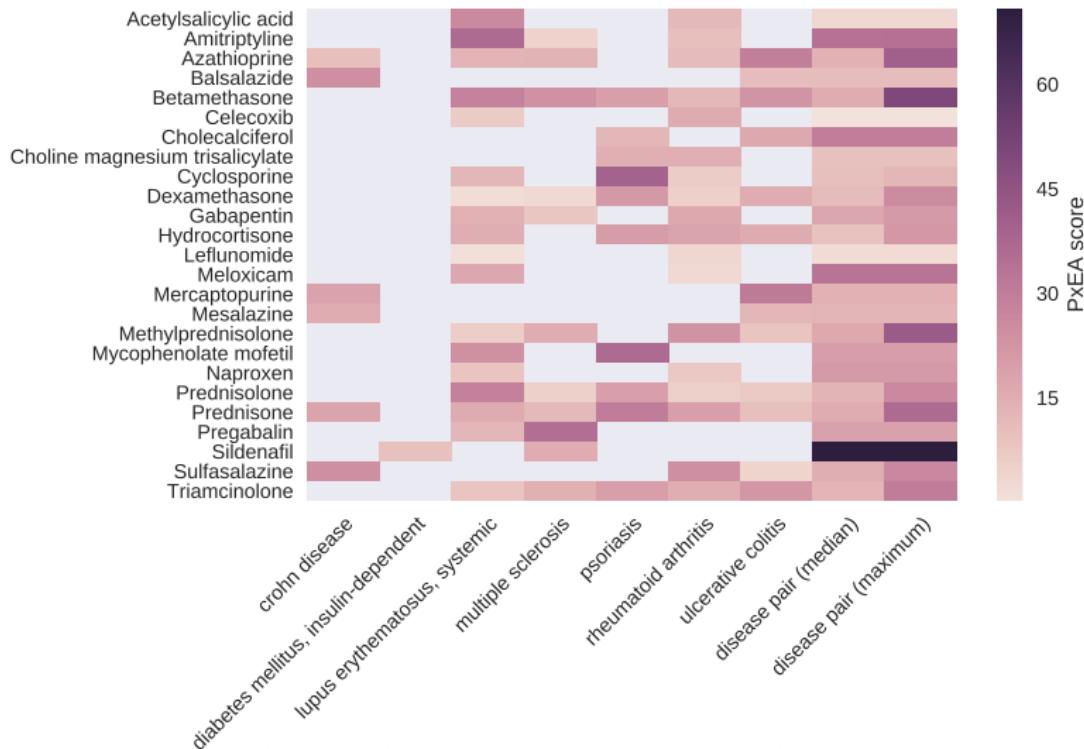
$$X_i = \begin{cases} \sqrt{\frac{|D|-|C|}{|C|}}, & \text{if } p_i \in C \\ -\sqrt{\frac{|C|}{|D|-|C|}}, & \text{otherwise} \end{cases}$$

To calculate significance, repeat the procedure above 10,000 times, shuffling randomly D

$$P = \frac{|ES(D, C) < ES(D^{random}, C)|}{10,000}$$

Aguirre-Plans et al., 2018, Pharmaceuticals

Targeting pathway level commonalities between diseases using PxEA



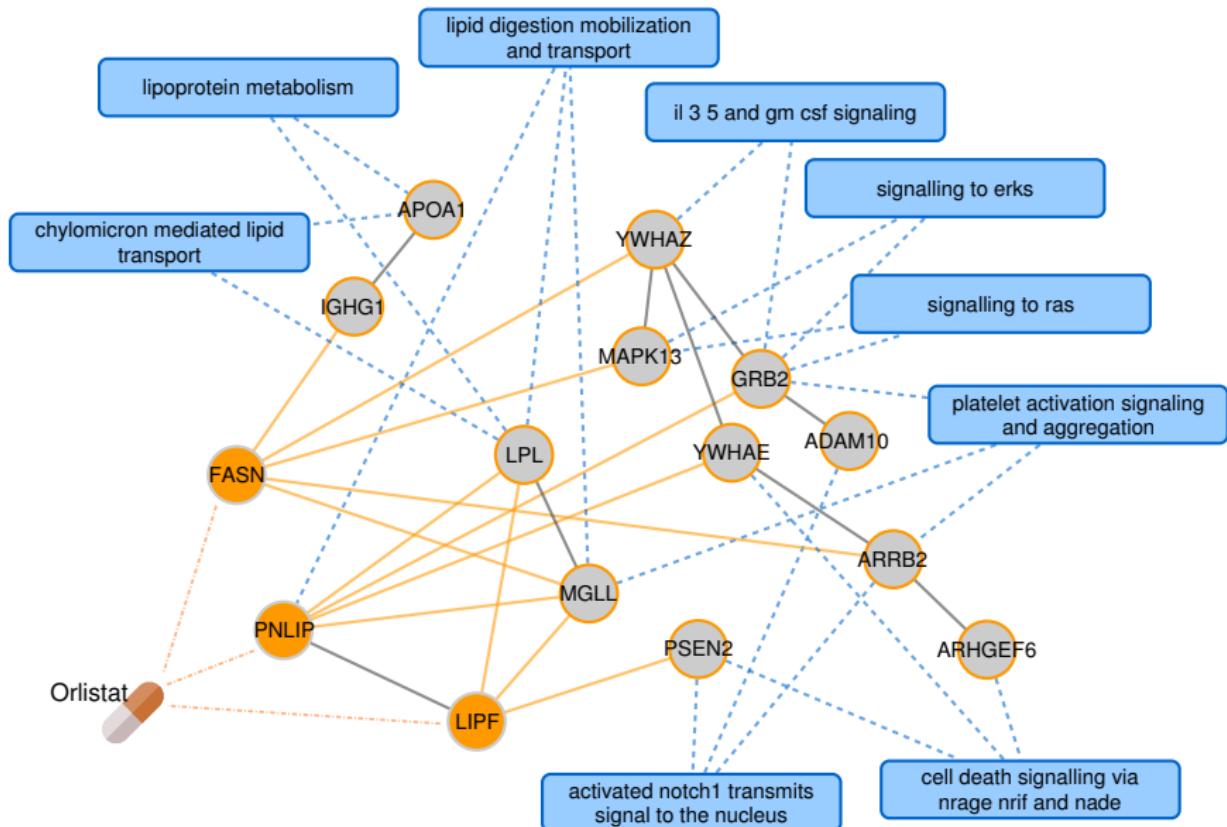
Drugs commonly used across multiple conditions tend to target common pathways between these diseases

PxE predictions using pathways shared between Alzheimer and type 2 diabetes

| Drug | Hetionet indication | PxE score | Adjusted P-value |
|-------------------------|---------------------------|-----------|------------------|
| orlistat | obesity, type 2 diabetes | 94.07 | < 0.001 |
| chenodeoxycholic acid | primary biliary cirrhosis | 74.06 | < 0.001 |
| obeticholic acid | - | 74.06 | < 0.001 |
| practolol | - | 70.55 | < 0.001 |
| esmolol | hypertension | 70.55 | < 0.001 |
| clenbuterol | - | 70.44 | < 0.001 |
| erythrityl tetranitrate | - | 70.32 | < 0.001 |
| bupranolol | - | 68.97 | < 0.001 |
| arbutamine | - | 68.97 | < 0.001 |
| fenoterol | - | 68.97 | < 0.001 |

Aguirre-Plans et al., 2018, Pharmaceuticals

PxEAs: Orlistat subnetwork



Bonus: GUILDify Webserver

GUIDify (v2) prioritizes genes using integrated disease, drug and protein interaction data

[File upload]

Species: Homo sapiens

Tissues: All

Example queries: ["lung carcinoma" (TOGETHER)][+alzheimer +diabetes (AND)][alzheimer diabetes (OR)][TP53; BRCA1; BRCA2 (GENES)]
If not 'quoted', whitespaces act as OR, you can convert them to AND by using a '+'. See quick start guide below for details.

⌚ Quick start guide

+

💻 Pre-calculated examples

+

⌚ Retrieve results

+

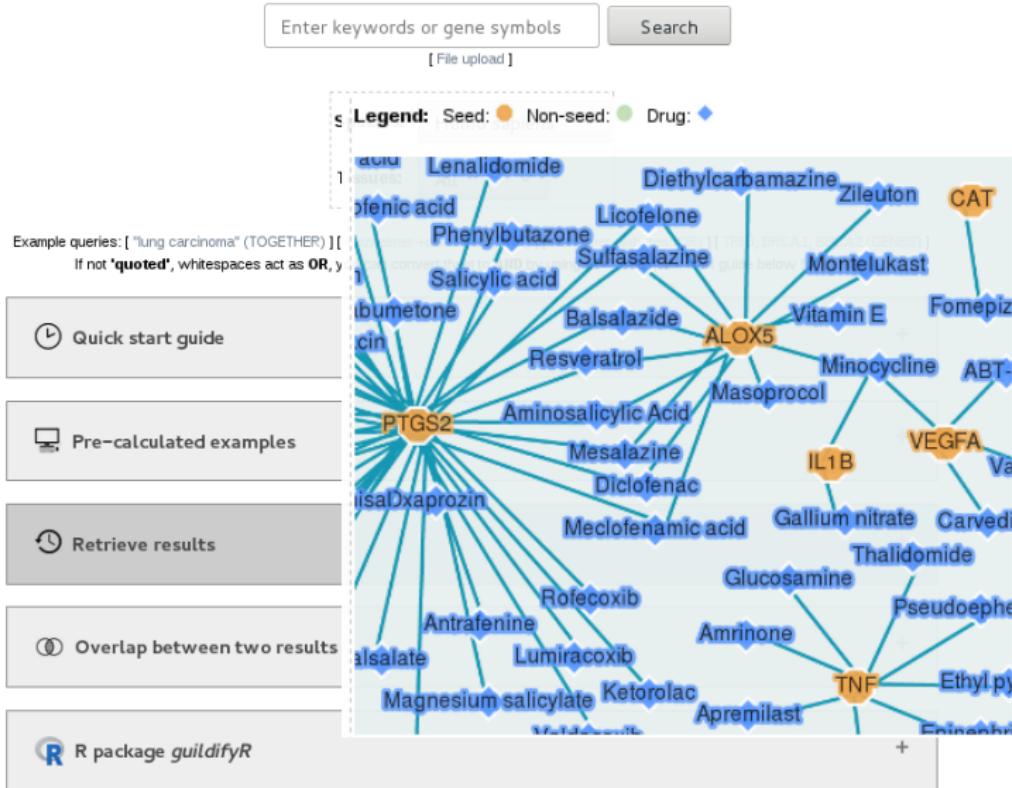
⌚ Overlap between two results

+

⌚ R package *guidifyR*

+

GUIDify (v2) prioritizes genes using integrated disease, drug and protein interaction data



GuILDify (v2) identifies functional connections between diseases

Overlap among functions of top ranking proteins

rheumatoid arthritis (1) common asthma (2)



| N common | N1 | N2 | N total | Odds Ratio | P value |
|----------|-----|-----|---------|------------|----------|
| 179 | 551 | 523 | 49064 | 67.378 | 2.2E-220 |

| GO ID | GO term name | P-value 1 | P-value 2 | Comb. p-value |
|------------|---|-----------|-----------|---------------|
| GO:0042102 | positive regulation of T cell proliferation | 8.2E-10 | 1.6E-05 | 4.5E-13 |
| GO:0031663 | lipopolysaccharide-mediated signaling pathway | 1.8E-07 | 1.8E-07 | 9.9E-13 |
| GO:0005125 | cytokine activity | 4.0E-08 | 1.0E-06 | 1.3E-12 |
| GO:0001934 | positive regulation of protein phosphorylation | 5.9E-09 | 5.7E-04 | 9.3E-11 |
| GO:0048661 | positive regulation of smooth muscle cell proliferation | 8.6E-08 | 7.6E-05 | 1.8E-10 |
| GO:0071407 | cellular response to organic cyclic compound | 1.6E-04 | 4.7E-08 | 2.0E-10 |
| GO:0031622 | positive regulation of fever generation | 3.1E-06 | 3.1E-06 | 2.5E-10 |
| GO:0032729 | positive regulation of interferon-gamma production | 1.1E-04 | 1.4E-07 | 3.8E-10 |
| GO:0060559 | positive regulation of calcidiol 1-monooxygenase activity | 7.8E-07 | 2.5E-04 | 4.6E-09 |
| GO:0043491 | protein kinase B signaling | 1.6E-02 | 2.7E-08 | 9.8E-09 |
| GO:0051384 | response to glucocorticoid | 8.5E-07 | 2.9E-03 | 5.2E-08 |
| GO:0006954 | inflammatory response | 7.1E-05 | 7.1E-05 | 1.0E-07 |

In summary...

- Data-driven models are powerful, yet prone to overfitting (especially on small data sets)
- Interactome-based modeling can offer improved **interpretability**
- Personalized profiles could be used to characterize and account for transcriptomic **heterogeneity**
- Exploiting functional relationships across diseases offers novel opportunities for drug repurposing via **endopharmacology**

Acknowledgements



IMIM

Janet Piñero
Ferran Sanz
Laura Furlong

UPF

Joaquim Aguirre-Plans
Baldo Oliva

CeMM

Joerg Menche

CCNR

Albert-László Barabási

CCSB

Marc Vidal

UM

Terezinha de Souza

Ana I Casas

Harald HHW Schmidt

ISCIII-FEDER (PI13/00082, CP10/00524, CPII16/00026), MINECO (BIO2014-57518-R), IMI-JU (No:116030 - TransQST & No:777365 - eTRANSAFE) with contributions from the EU H2020 and EFPIA companies



Appendix

Data set

Zhang and coworkers (2013)

- 536 drugs and their targets (DrugBank) & chemical structures (PubChem)
- 2,229 drug-disease associations (NDF-RT) covering 578 diseases
- 40,455 drug-side effect associations (SIDER) covering 1,252 side effects

The data set is publicly available online at

astro.temple.edu/~tua87106/drugreposition.html

Prediction accuracy evaluation

Positive instances: 2,229 known drug-disease associations

Potential negative instances: Remaining possible associations
between 536 drugs and 578 diseases
 $(536 \times 578 - 2,229 = 307,579)$ associations

Negative instances: Randomly sample twice as many negative instances as positives

- Calculate the area under ROC curve (AUC) / Precision-Recall curve (AUPRC)
- Use k-fold cross validation scheme ($k = 2, 5, 10, 20$)
- Report the mean AUC over 10 repetitions of the cross validation procedure

Defining similarity between drugs

Drug i defined as a binary vector for a given feature

$$X_i^f = [x_1^f, x_2^f, \dots, x_n^f]^T \quad f: \begin{array}{l} \bullet \text{ chemical substructures} \\ \bullet \text{ targets} \\ \bullet \text{ side effects} \end{array}$$

Similarity between two drugs i and j are defined by

$$\rho_{ij}^f = \frac{C_{ij}^f}{\sqrt{C_{ii}^f * C_{jj}^f}}$$

where C_{ij}^f given by

$$C_{ij}^f = \text{cov}(X_i^f, X_j^f) = E[(X_i^f - E(X_i^f))(X_j^f - E(X_j^f))]$$

Similarity-based logistic regression classifier

Probability of observing an association between the drug i and the disease p

$$P(Y_{ip} = 1 | s_{ip}^{\text{chem.}}, s_{ip}^{\text{target}}, s_{ip}^{\text{s.effect}}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * s_{ip}^{\text{chem.}} + \beta_2 * s_{ip}^{\text{target}} + \beta_3 * s_{ip}^{\text{s.effect}})}}$$

where the similarity-based drug-disease score s_{ip}^f is defined as

$$s_{ip}^f = \sum_{j \in \text{NearestNeighbors}(i)} \rho_{ij}^f * I_{jp}$$

and

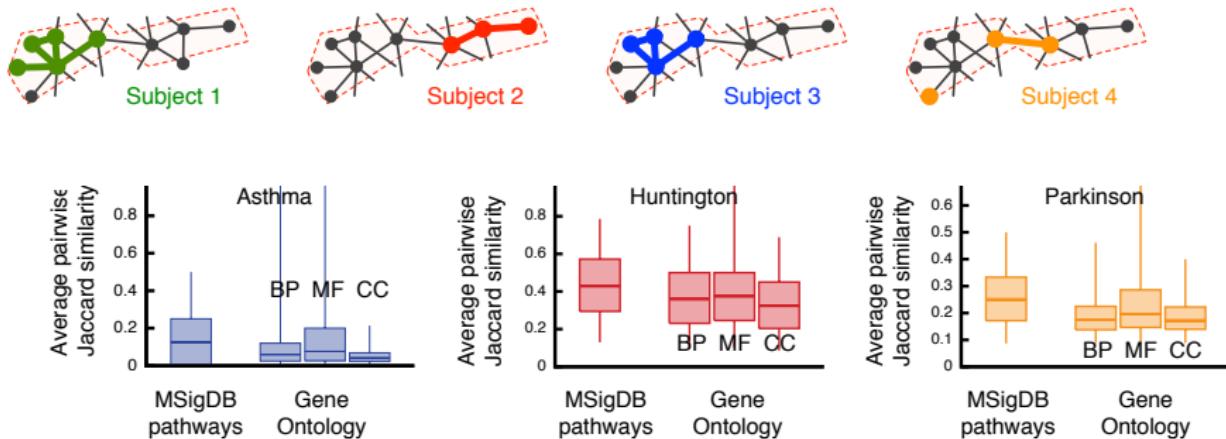
$$I_{jp} = \begin{cases} 1, & \text{if } j \text{ is indicated for } p \\ 0, & \text{otherwise} \end{cases}$$

(Subject to L2 regularization, i.e., find β maximizing $\sum_{k=1}^m \log(y_k | x_k, \beta) - \sum_{l=0}^3 \beta_l^2$)

Repurpose in action

| Drug | Non-disjoint cross validation | | | | Disjoint cross validation | | | |
|--|-------------------------------|--------------|-------------------|-------------|---------------------------|--------------|-------------------|-------------|
| | Chemical score | Target score | Side effect score | Probability | Chemical score | Target score | Side effect score | Probability |
| <i>Hypercholesterolemia drugs</i> | | | | | | | | |
| fenofibrate | 0.76 | 0.71 | 1.10 | 0.82 | 0.57 | 0 | 0.46 | 0.36 |
| lovastatin | 1.93 | 1.97 | 2.92 | 0.99 | 0 | 0 | 0 | 0.14 |
| <i>Juvenile rheumatoid arthritis drugs</i> | | | | | | | | |
| ibuprofen | 0.82 | 3.50 | 1.08 | 1.00 | 0 | 0.50 | 0.43 | 0.43 |
| sulfasalazine | 1.39 | 1.99 | 0.43 | 0.96 | 0 | 0.50 | 0.43 | 0.43 |
| <i>Acute myeloid leukemia drugs</i> | | | | | | | | |
| daunorubicin | 1.77 | 1.50 | 0 | 0.87 | 0 | 0 | 0 | 0.15 |
| idarubicin | 0.78 | 2.00 | 0.81 | 0.97 | 0 | 0 | 0 | 0.14 |

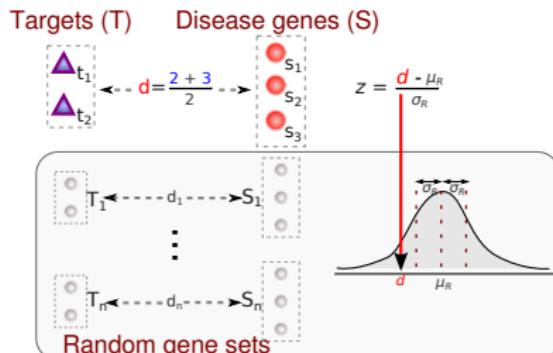
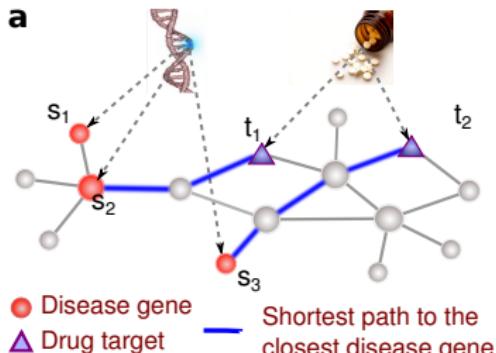
From personalized gene-level signatures to pathway-level signatures



Enrichment of disease-specific pathways in PeePs (assessed by Fisher's test followed by Bonferroni correction) reveals that

- almost all the individuals show significant perturbations in disease-specific pathways
- the specific perturbations differ greatly across subjects

Evaluating proximity

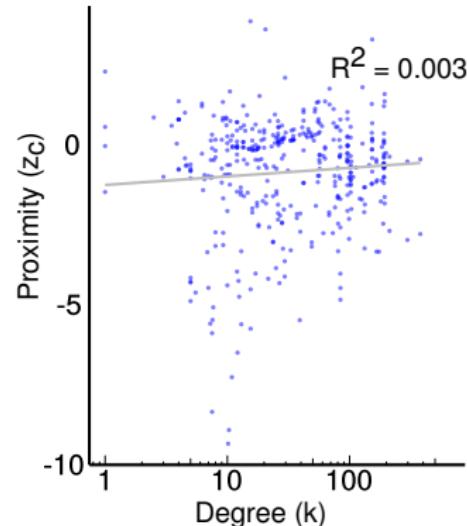
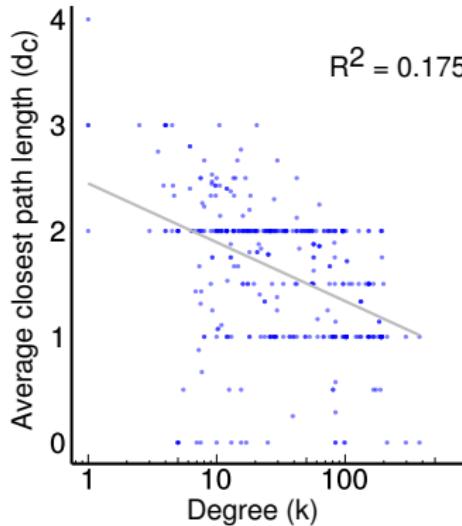


78 diseases x 238 drugs
18,564 possible drug-disease pairs

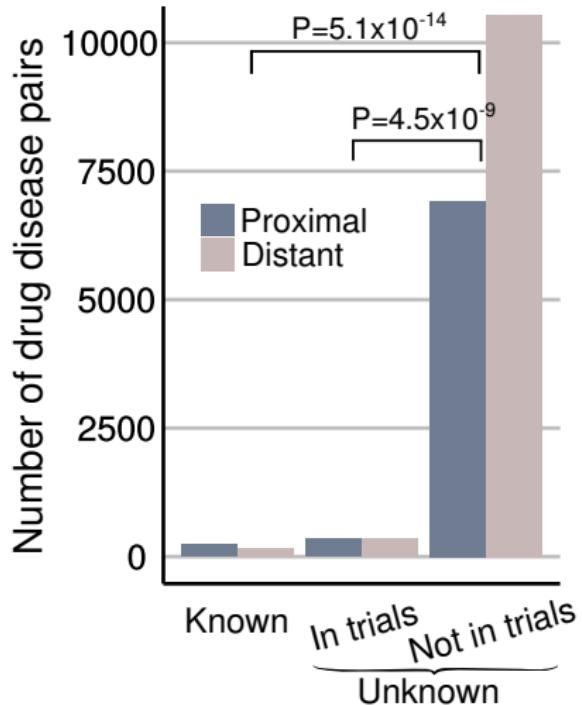
| | Type 2 diabetes | Acute myeloid leukemia |
|--------------|-----------------|------------------------|
| Gliclazide | ✓ | - |
| Daunorubicin | - | ✓ |

402 known vs 18,162 unknown drug-disease pairs based on MEDI

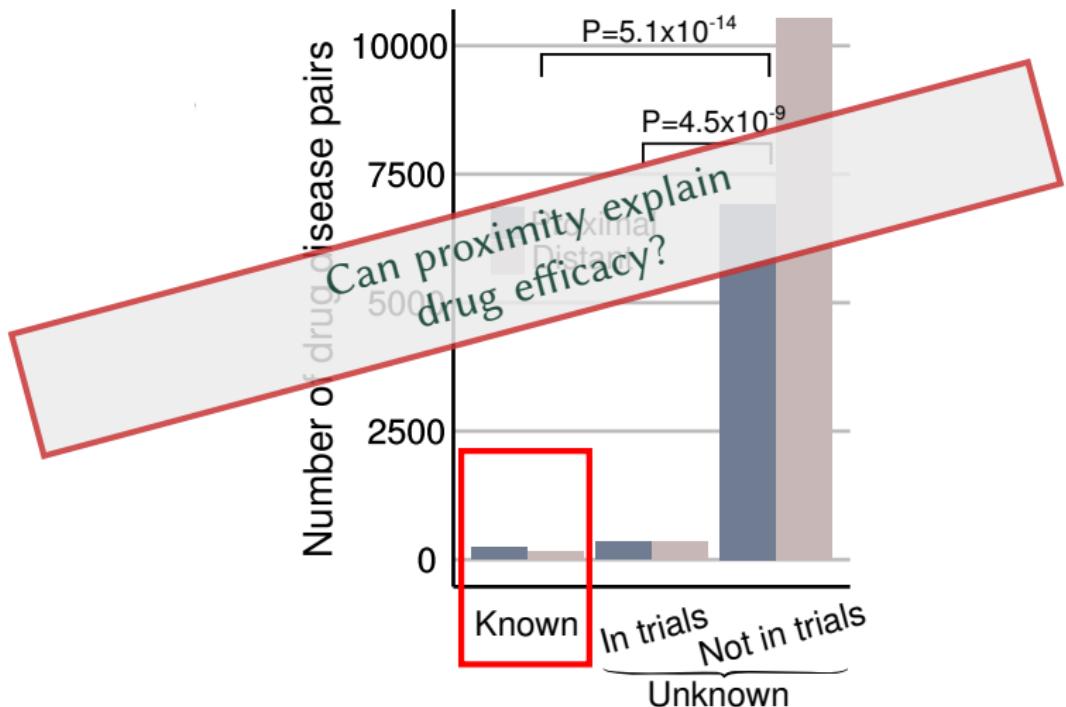
Proximity is not correlated with the degree of the targets



Proximal drug-disease pairs are enriched among known associations



Proximal drug-disease pairs are enriched among known associations



Assessing drug efficacy



About
Updates
@openFDA Twitter

API basics

Drugs

Devices

Foods

Drugs

$$\text{Relative Efficacy (RE)} = 1 - \frac{\text{\# of adverse events reporting inefficacy}}{\text{\# of most common adverse event}}$$

Adverse events

Labeling

Enforcement rep

Ask a question

Report a bug

"meta": {
 "disclaimer": "openFDA is a beta research project and not for clinical use. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated.",
 "license": "http://open.fda.gov/license",
 "last_updated": "2014-08-06"
},

"results": [
 {

"term": "PYREXIA",
 "count": 36

},

{
 "term": "RHEUMATOID ARTHRITIS",
 "count": 32

},

{
 "term": "ARTHRALGIA",
 "count": 28

},

{
 "term": "DRUG INEFFECTIVE",
 "count": 26

},

{
 "term": "PNEUMONIA",
 "count": 23

},

{
 "term": "INTERSTITIAL LUNG DISEASE",
 "count": 23

},

Adverse drug event reports since 2004

This is the openFDA API endpoint for adverse drug events. An adverse event is submitted to the FDA to report any undesirable experience associated with the use of a drug, including serious drug side effects, product use errors, product quality problems, and therapeutic failures.

Reporting of adverse events by healthcare professionals and consumers is voluntary in the United States. Increases in the total number of adverse events are likely caused by improved reporting. News, enforcement actions, and other phenomena can also spur reporting.



Proximity vs therapeutic efficacy

Ketoprofen capsules USP are indicated for **the management of the signs and symptoms of rheumatoid arthritis and osteoarthritis.**

Ketoprofen capsules USP are indicated for **the management of pain.**

<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=c2c99853-1268-4998-a44b-2bf0c0b70fd2>

Proximity to rheumatoid arthritis (z) = 1.5

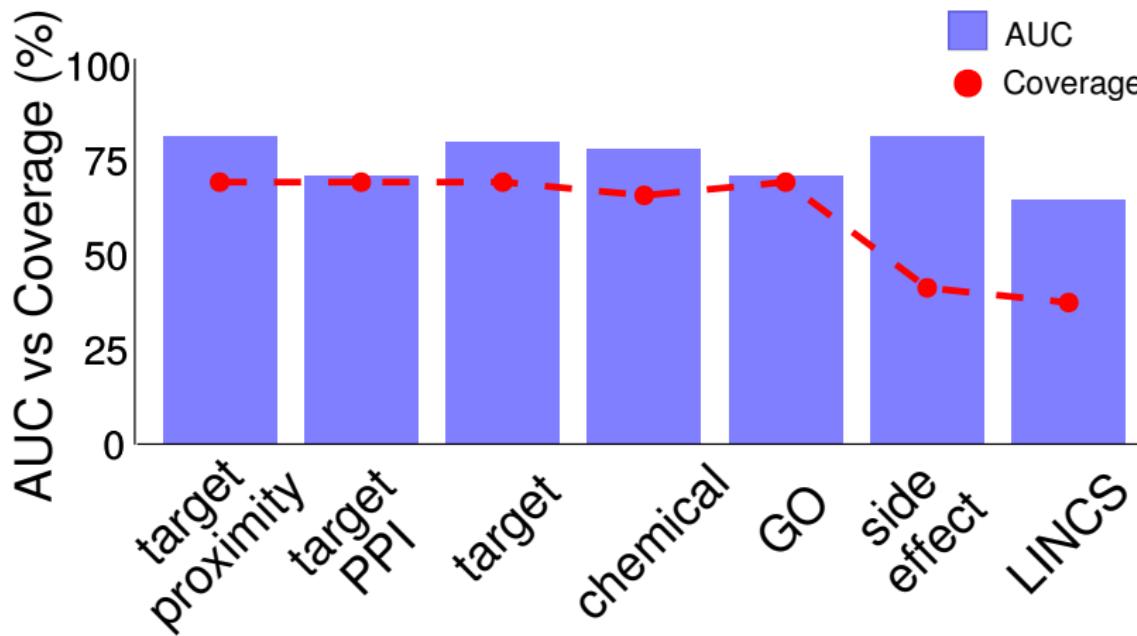
Ropinirole Tablets are indicated for **the treatment of the signs and symptoms of idiopathic Parkinson's disease.**

<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=9a25f575-09ab-4d32-b73e-5426f08c00c4>

Proximity to Parkinson's disease (z) = 0.8

Categorizing drugs based on label info: Off-label | Palliative | Non-palliative

Proximity in comparison to drug similarity-based approaches



Drug-drug similarity based classification

Guney et al., 2016, Nat Comm

Drug repurposing using proximity: Glimepiride

| Disease | Proximity (z) |
|----------------------|---------------|
| type 2 diabetes | -5.1 |
| cardiac arrhythmias | -3.6 |
| Alzheimer's disease | -1.0 |
| rheumatoid arthritis | -0.7 |

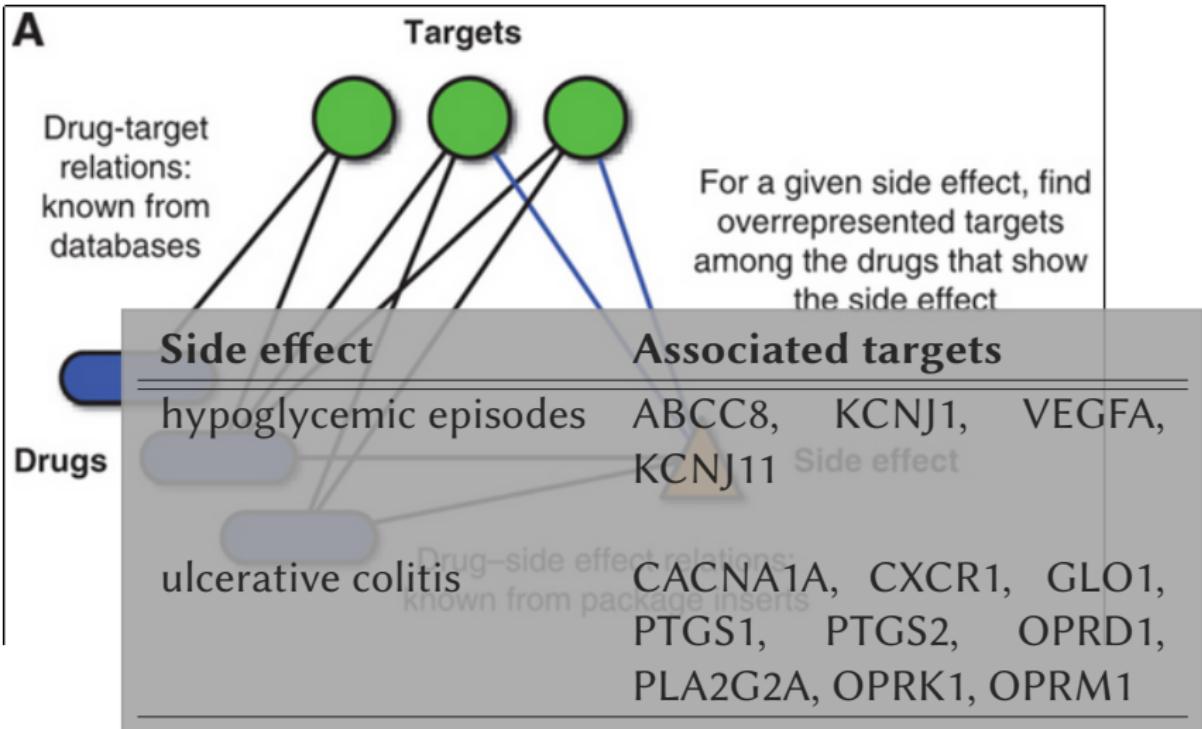
Glimepiride is approved for use in type 2 diabetes.

Mortality and cardiovascular risk associated with different insulin secretagogues compared with metformin in type 2 diabetes, with or without a previous myocardial infarction: a nationwide study. Schramm,

T. K. et al. *Eur. Heart J.*, 32:1900 (2011).

Defining side effect module

Kuhn et al., 2013, Mol. Sys. Bio.



Controlling for data quality

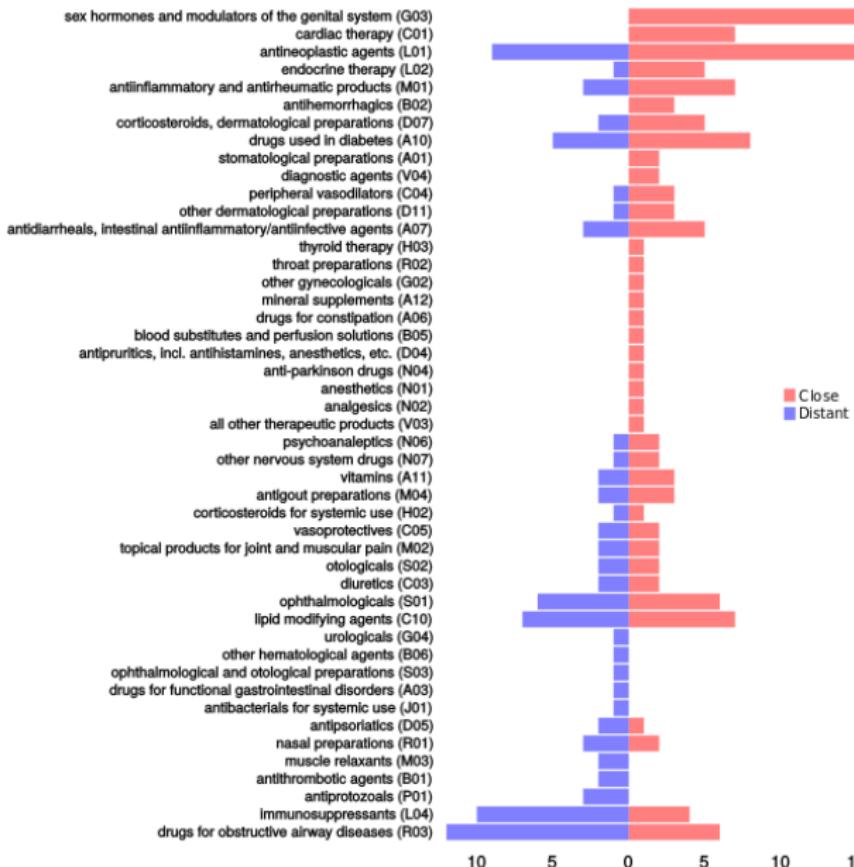
| Data set | Number of diseases | Number of drugs | Number of drug-disease pairs | AUC (%) |
|----------------------------------|--------------------|-----------------|------------------------------|---------|
| Original | 78 | 238 | 402 | 65.7 |
| <i>Protein interactions</i> | | | | |
| Binary interactome | 50 | 129 | 226 | 58.3 |
| STRING | 77 | 233 | 396 | 61.3 |
| <i>Disease-gene associations</i> | | | | |
| OMIM | 35 | 114 | 155 | 71.2 |
| GWAS | 44 | 157 | 260 | 60.2 |
| <i>Drug-target associations</i> | | | | |
| STITCH | 73 | 212 | 359 | 64.8 |
| <i>Disease-drug associations</i> | | | | |
| NDF-RT | 61 | 160 | 233 | 66.2 |
| KEGG | 16 | 74 | 76 | 71.3 |

Controlling for data quality

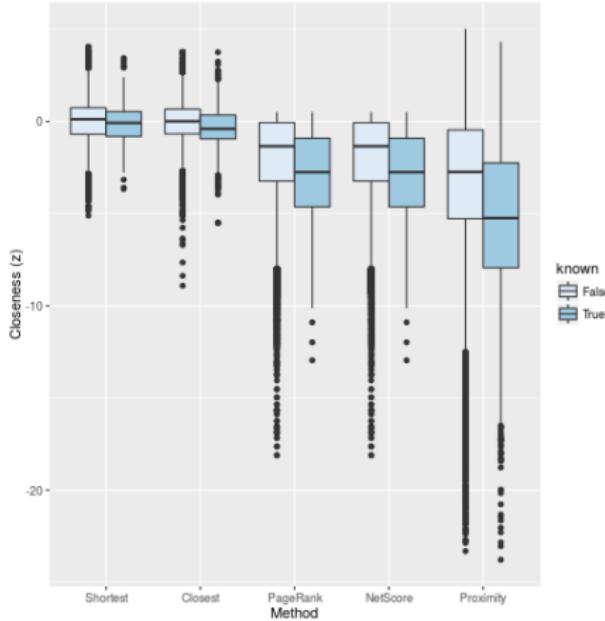
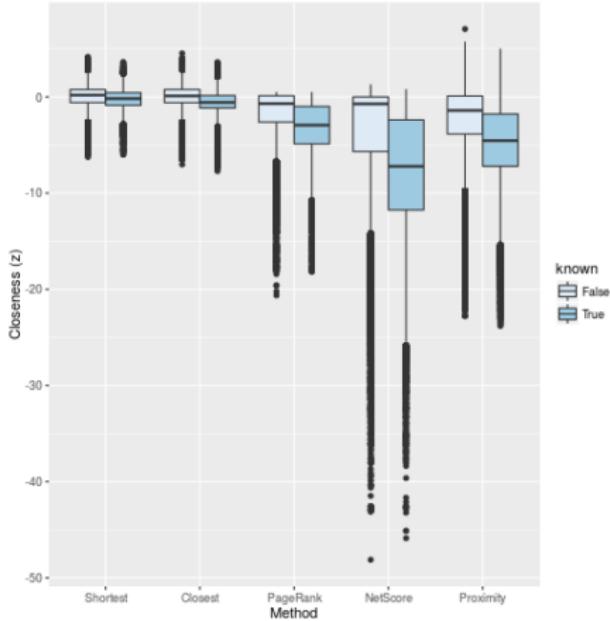
| Data set | # of diseases | # of drugs | # of drug-disease pairs | AUC (%) |
|--|---------------|------------|-------------------------|---------|
| <i>Diseases</i> | | | | |
| $n_{gene} \geq 20$ | 78 | 238 | 402 | 65.7 |
| $n_{gene} \geq 1$ | 304 | 462 | 1192 | 58.6 |
| $n_{gene} \geq 20$, broad terms filtered | 53 | 205 | 282 | 67.2 |
| <i>Drugs</i> | | | | |
| $n_{target} \geq 3$ | 49 | 95 | 144 | 64.6 |
| $n_{target \cap gene} = 0$ | 76 | 227 | 384 | 64.5 |

Guney et al., 2016, Nat Comm

ATC classification of proximal and distant drugs



Network-based closeness of side effects



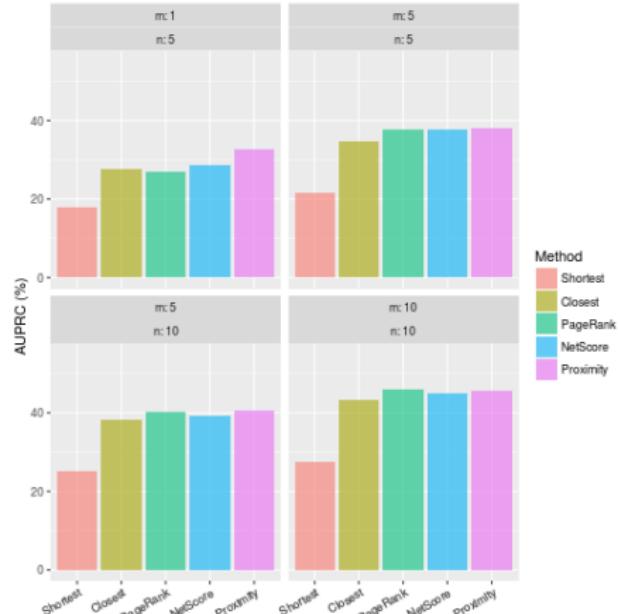
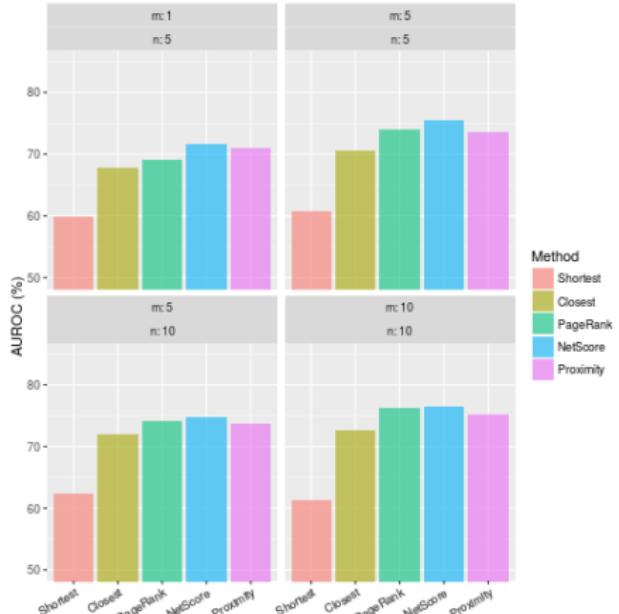
Prediction accuracy of network-based side effect detection

| | AUROC (%) | | AUPRC (%) | | Correct at top (%) | |
|-----------|-----------|----------|-----------|----------|--------------------|----------|
| | SIDER | OFFSIDES | SIDER | OFFSIDES | SIDER | OFFSIDES |
| Shortest | 59.8 | 53.9 | 17.8 | 7.1 | 15.9 | 8.2 |
| Closest | 67.9 | 57.7 | 27.6 | 8.5 | 79.6 | 28.6 |
| PageRank | 69.0 | 59.6 | 27.0 | 8.6 | 55.8 | 13.0 |
| NetScore | 71.7 | 61.9 | 28.8 | 9.6 | 52.1 | 14.5 |
| Proximity | 71.1 | 63.6 | 32.8 | 11.4 | 56.7 | 11.5 |

[†]Area under ROC curve (AUROC)

[‡]Area under Precision-Recall curve (AUPRC) *Guney, 2017, Workshop on Complex Networks*

Robustness of network-based drug side effect detection



Guney, 2017, Workshop on Complex Networks

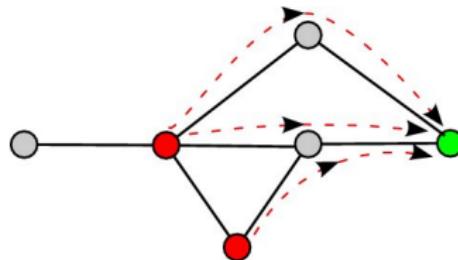
Top 10 side effects predicted for tamoxifen using ProXide

| Rank | Side effect | in SIDER | ProXide score (z) |
|------|----------------------------|----------|-------------------|
| 1 | muscular weakness | 0 | -12.9 |
| 2 | musculoskeletal discomfort | 1 | -12.3 |
| 3 | alopecia | 1 | -12.1 |
| 4 | neuropathy peripheral | 0 | -12.1 |
| 5 | drug interaction | 1 | -11.7 |
| 6 | hepatitis | 1 | -11.7 |
| 7 | diarrhoea | 1 | -11.7 |
| 8 | myalgia | 1 | -11.6 |
| 9 | injury | 1 | -11.5 |
| 10 | discomfort | 1 | -11.3 |

Guney, 2017, Workshop on Complex Networks

Network-based disease-gene prioritization

A message passing based algorithm to consider multiple shortest paths between two nodes



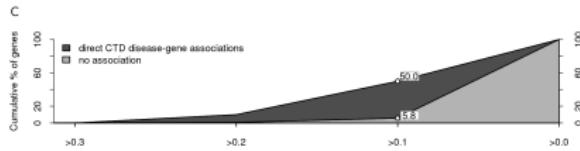
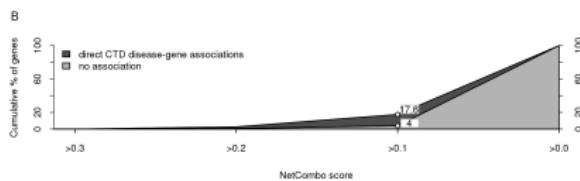
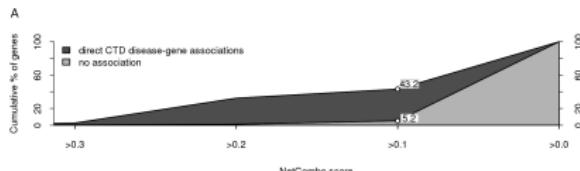
$$message = \left\{ \begin{array}{l} \textit{source} \\ \textit{timestamp} \\ \textit{score} \end{array} \right\}$$

$$\begin{matrix} \mathbf{u} \\ m_u \\ \vdots \\ m_v \\ \vdots \end{matrix}$$

Guney and Oliva, 2012, PLoS ONE

GUILD scores distinguish disease-genes from the rest of the genes

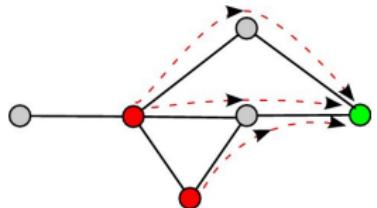
| Data Set | Metric | NetScore | NetZcore | NetShort | NetCombo | Func. Flow | PageRank | Random Walk | Network Prop. |
|----------|--------|--------------|----------|----------|--------------|------------|----------|-------------|---------------|
| OMIM | AUC | 67.49 | 62.99 | 65.63 | 72.09 | 58.55 | 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 |



Guney and Oliva, 2012, PLoS ONE

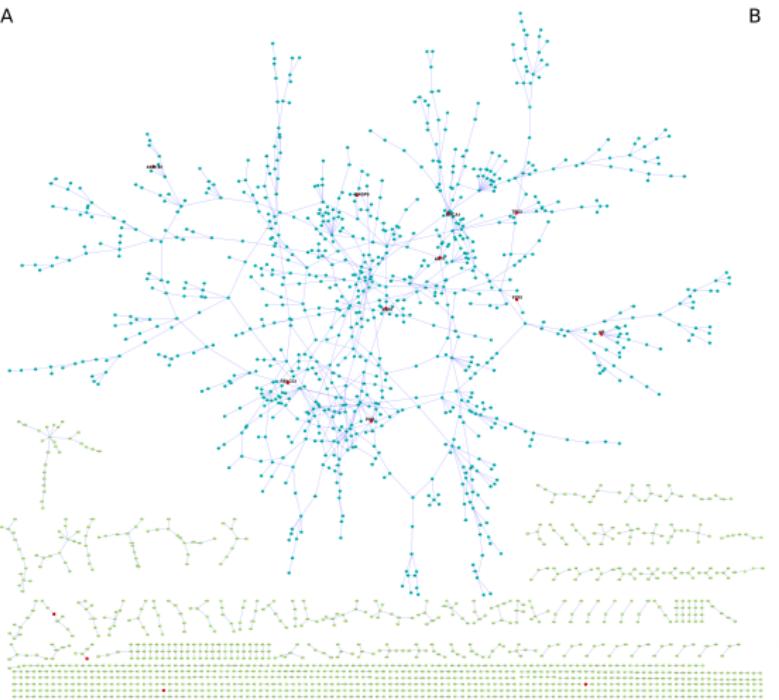
Understanding relationships between diseases via interactome-based modeling

GUILD framework
(free & open source)

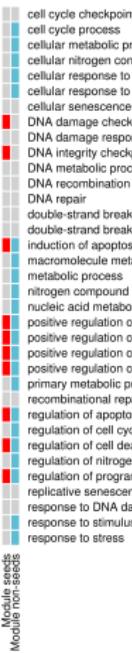


message =
 {
 source
 timestamp
 score
 }
 }

A



B

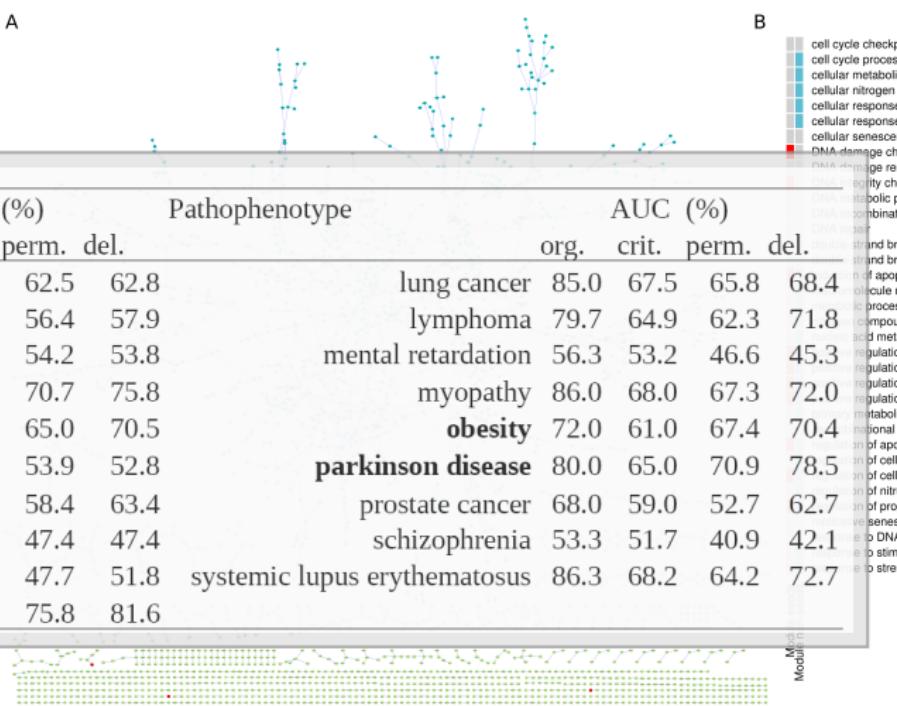


Guney and Oliva, 2012, PLoS ONE

Guney et al., 2014, Bioinformatics

Understanding relationships between diseases via interactome-based modeling

GUIDL framework (free & open source)



Guney and Oliva, 2014, PLoS ONE