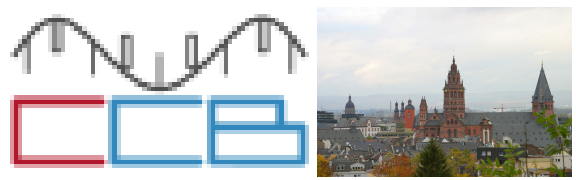
Data-driven analysis and network-based modeling for translational medicine

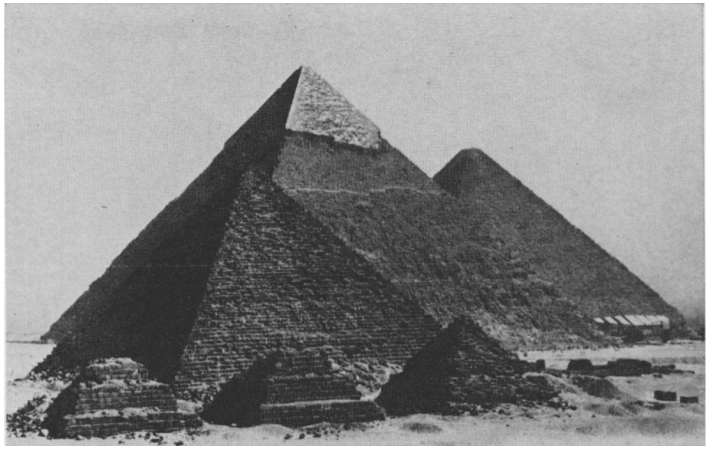
Emre Güney, PhD

Institute for Research in Biomedicine (IRB) Barcelona
& Pompeu Fabra University(UPF)

Nov 20^{th}, 2017

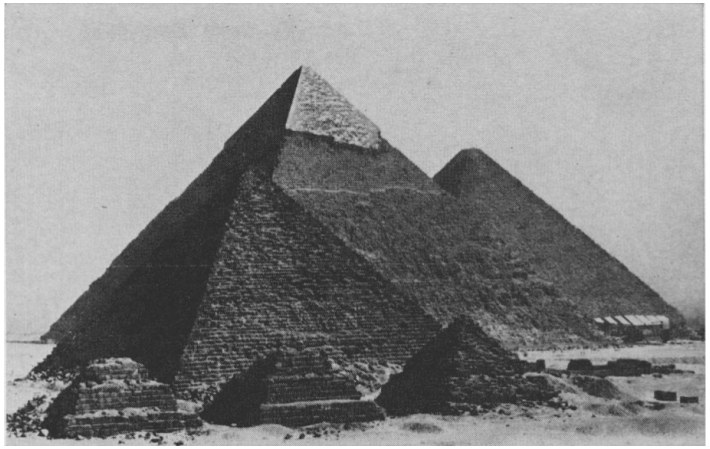


The quest for discovery



- More / better data
- Improved analysis methods

The quest for discovery



- More / better data
- Improved analysis methods

Challenge = Opportunity

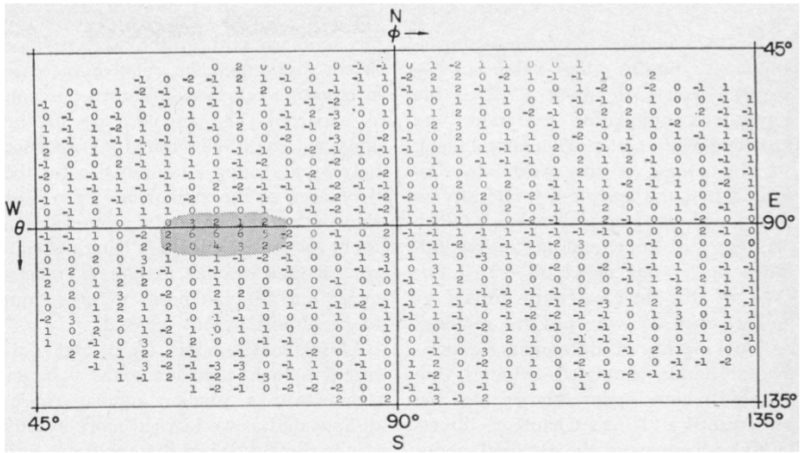
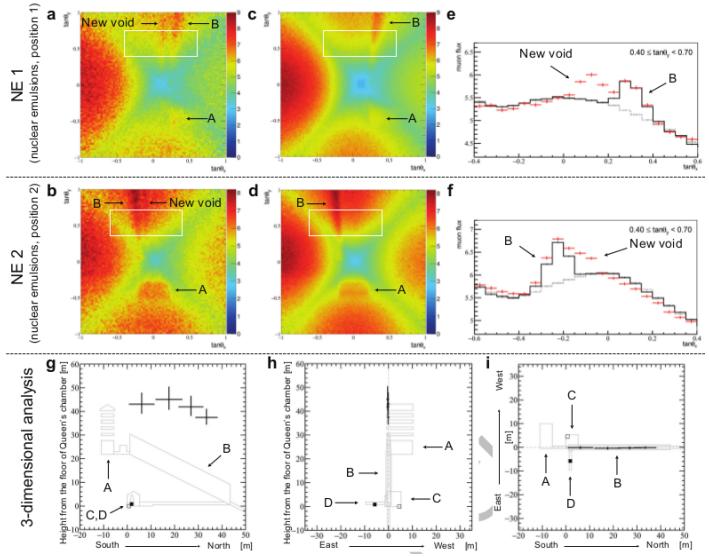


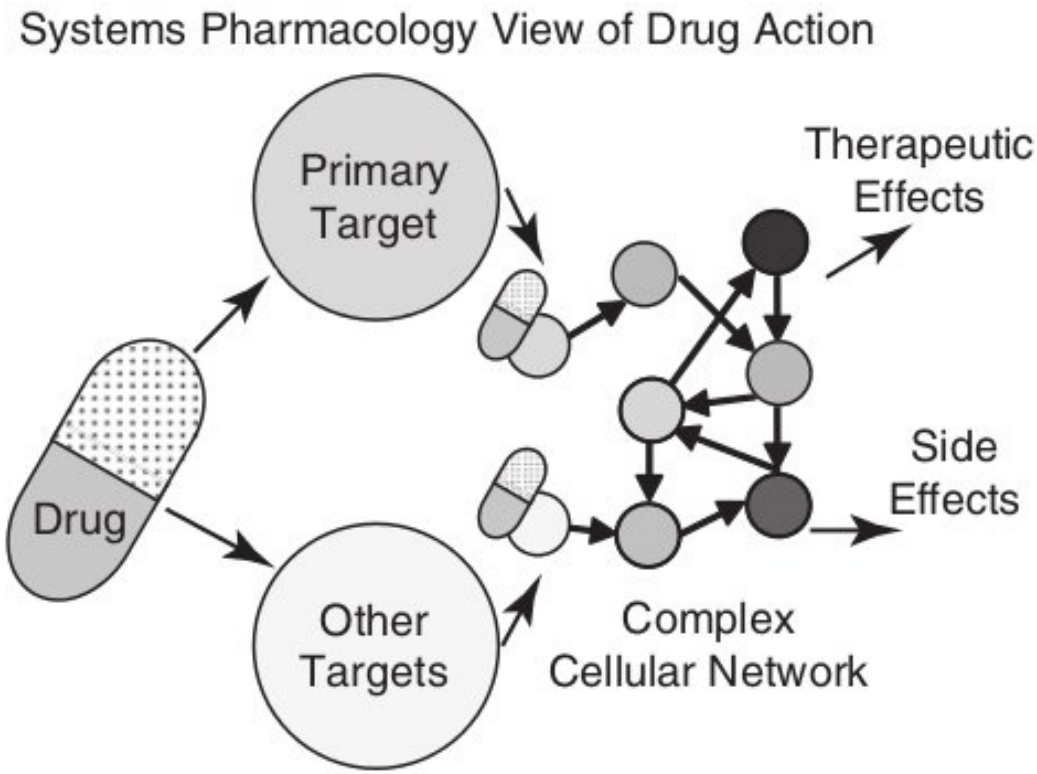
Fig. 11. The display of Fig. 10 as it would have appeared had there been a "King's Chamber" in the pyramid 40 meters above the apparatus. The group of numbers larger than 3 at the center-left (shaded area) indicates the chamber's position.

Search for hidden chambers in the pyramids, Science, 1970



Discovery of a big void in Khufu's Pyramid by observation of cosmic-ray muons, Nature, 2017

Challenges and opportunities in translational medicine



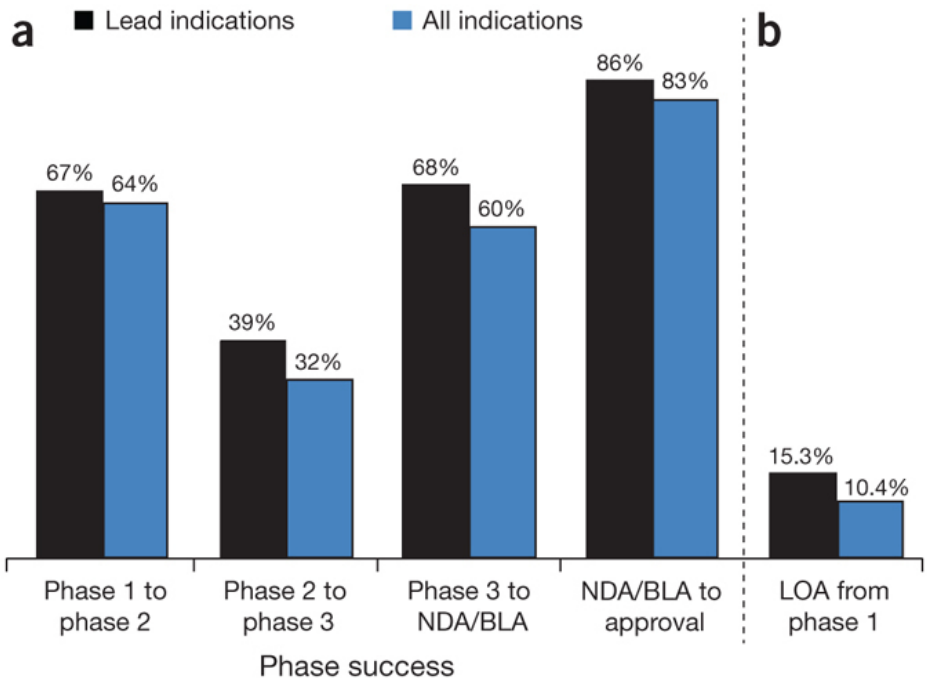
- Difficult to chemically achieve single target specificity
- Acting on multiple targets is likely to be more effective

Berger and Iyengar, 2009, Bioinformatics

Few drugs make it to the clinic

~10%

Percentage of drugs that get FDA approval after clinical trials

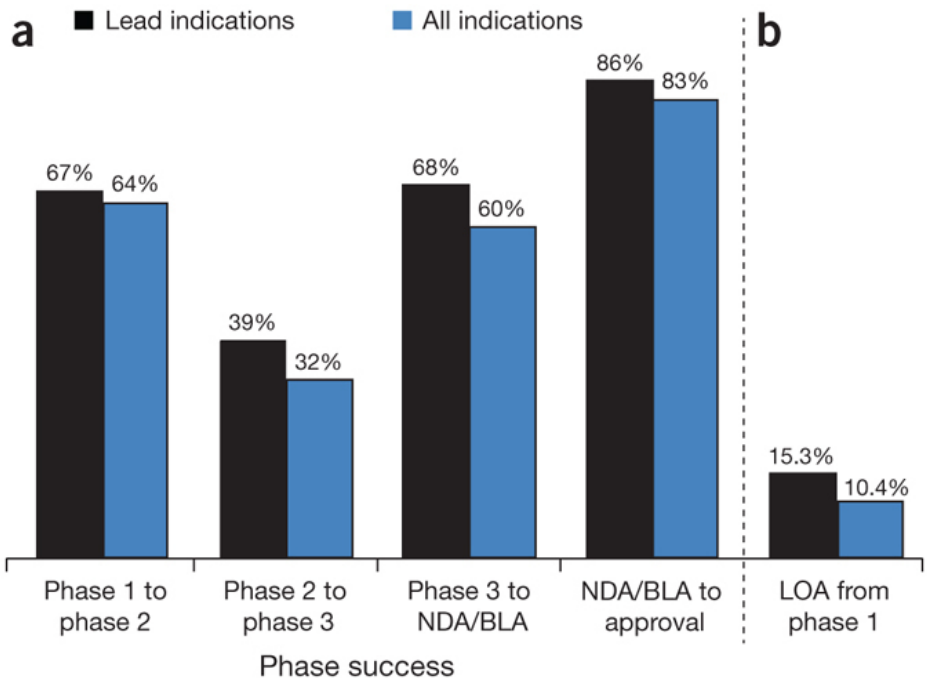


Hay et al., 2014, Nat Biotech

Few drugs make it to the clinic

~10%

Percentage of drugs that get FDA approval after clinical trials



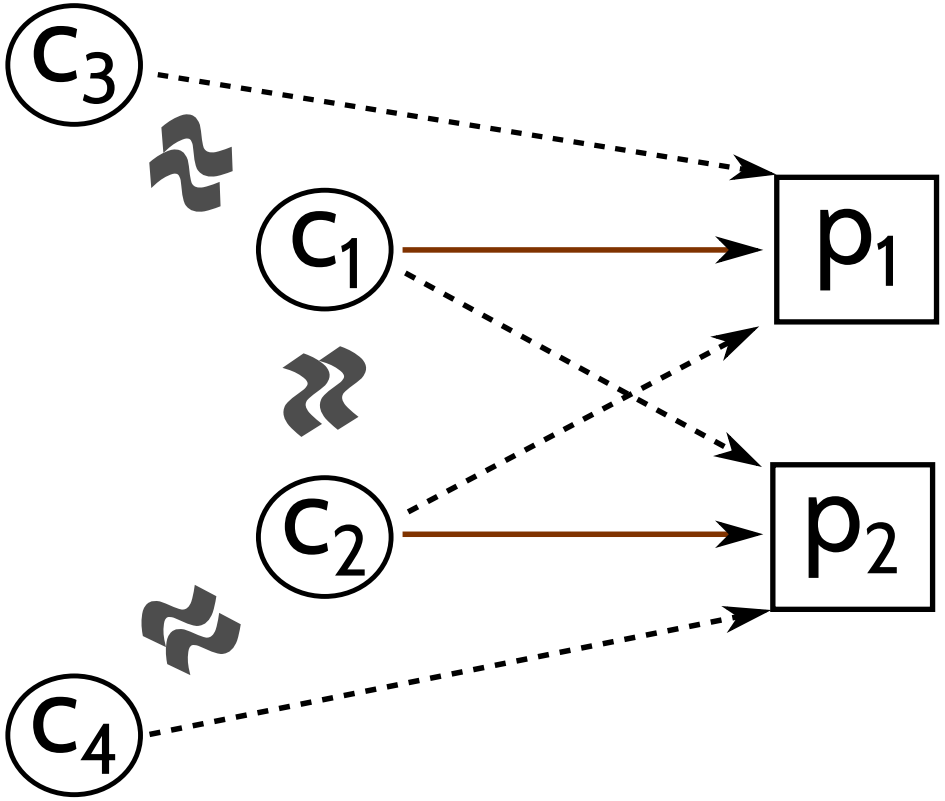
Hay et al., 2014, Nat Biotech

RECYCLE
REPURPOSE

**image from findacure.org.uk*

Reuse existing drugs

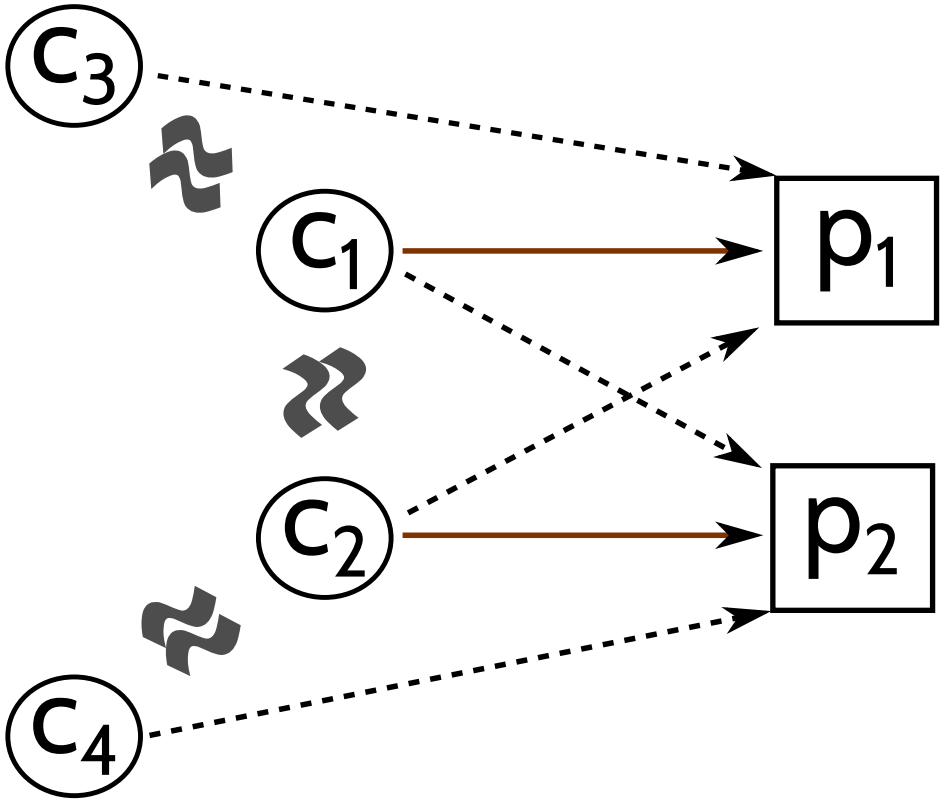
Similarity-based | Guilt-by-association | Knowledge-based



- chemical formula
- target
- side effect
- gene expression

Reuse existing drugs

Similarity-based | Guilt-by-association | Knowledge-based



- 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."

Traditional vs disjoint cross validation

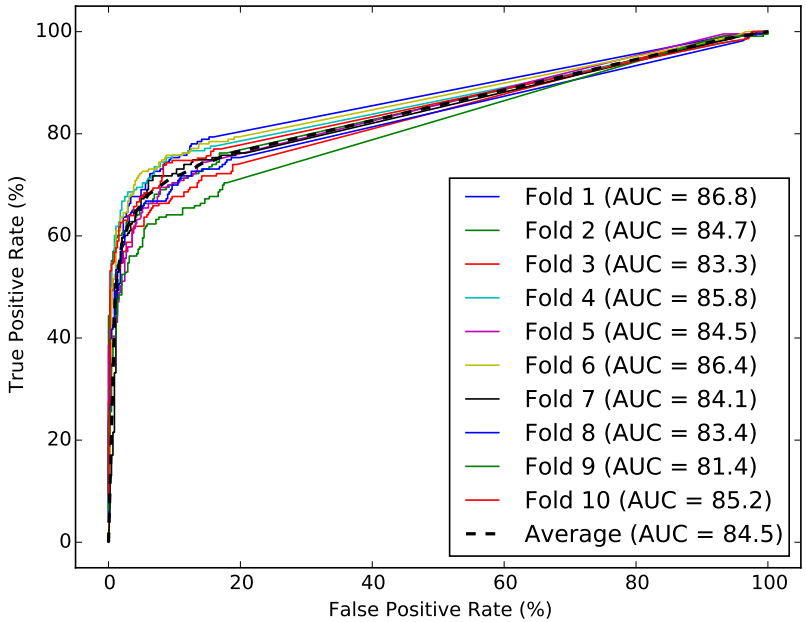
Data set		2-fold cross validation		2-fold disjoint cross validation	
		Fold 1	Fold 2	Fold 1	Fold 2
$(c_1, p_1, +)$	$(c_3, p_1, -)$	$(c_1, p_1, +)$	$(c_2, p_2, +)$	$(c_1, p_1, +)$	$(c_2, p_2, +)$
$(c_1, p_2, -)$	$(c_3, p_2, -)$	$(c_2, p_1, -)$	$(c_1, p_2, -)$	$(c_1, p_2, -)$	$(c_2, p_1, -)$
$(c_2, p_1,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

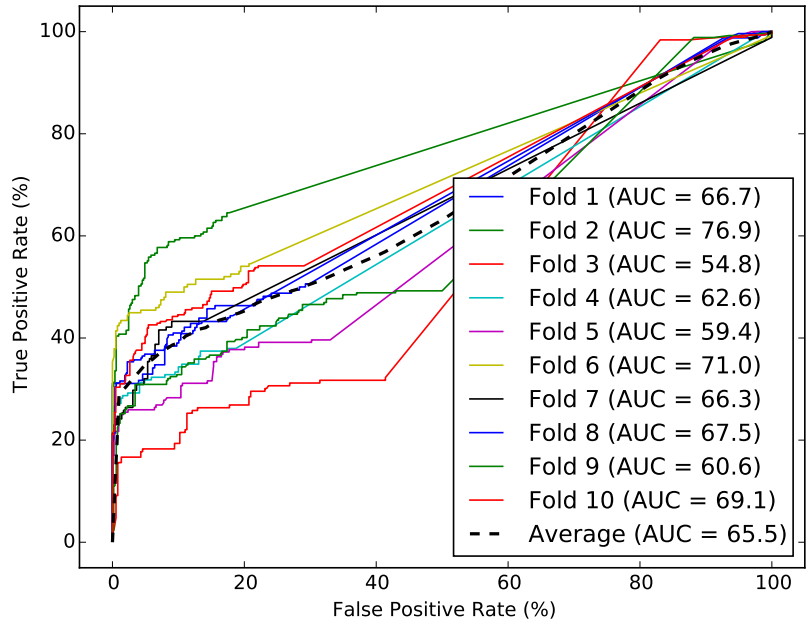
Defining non-overlapping drug groups

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
$i :=$ random($[0, 100]$)
$fold := \{\}$
for each $(c, p, l) \in D$ **do**
  $sum := 0$
  for each $x \in$ characters(c) **do**
    $sum := sum +$ to_integer(x)
  $fold(c, p) :=$ modulo$(sum + i, k)$
return $fold$

Guney, 2017, Pac Symp on Biocomp

Models perform poorly on drugs they have not seen before



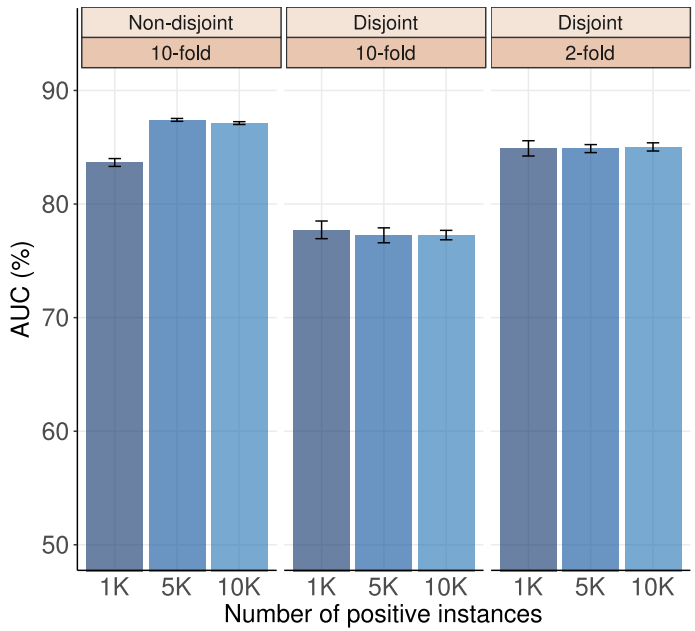


Disjoint folds	Mean AUC (%)	Mean AUPRC (%)
No	84.1 ($\pm$ 0.3)	83.7 ($\pm$ 0.3)
Yes	65.6 ($\pm$ 0.5)	62.8 ($\pm$ 0.5)

Guney, 2017, Pac Symp on Biocomp

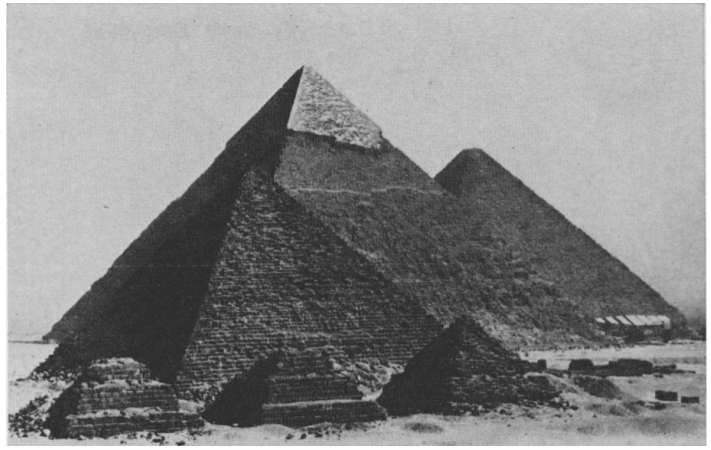
The diversity of the training set has a strong effect on the accuracy

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)



Guney, 2017, Pac Symp on Biocomp

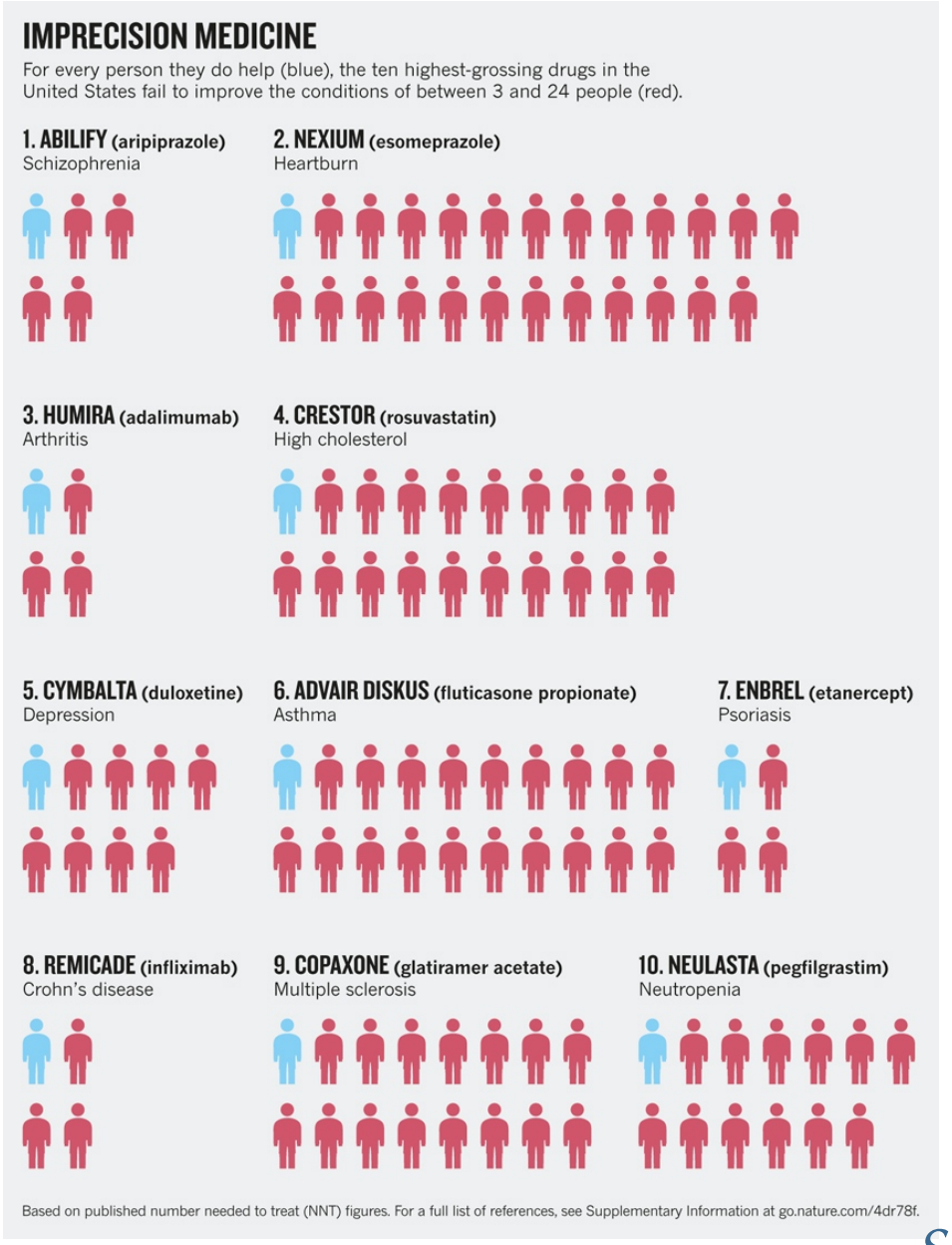
Limitations of similarity-based approaches



- Heterogeneity among disease phenotypes and patients
- Interpretability of the underlying model

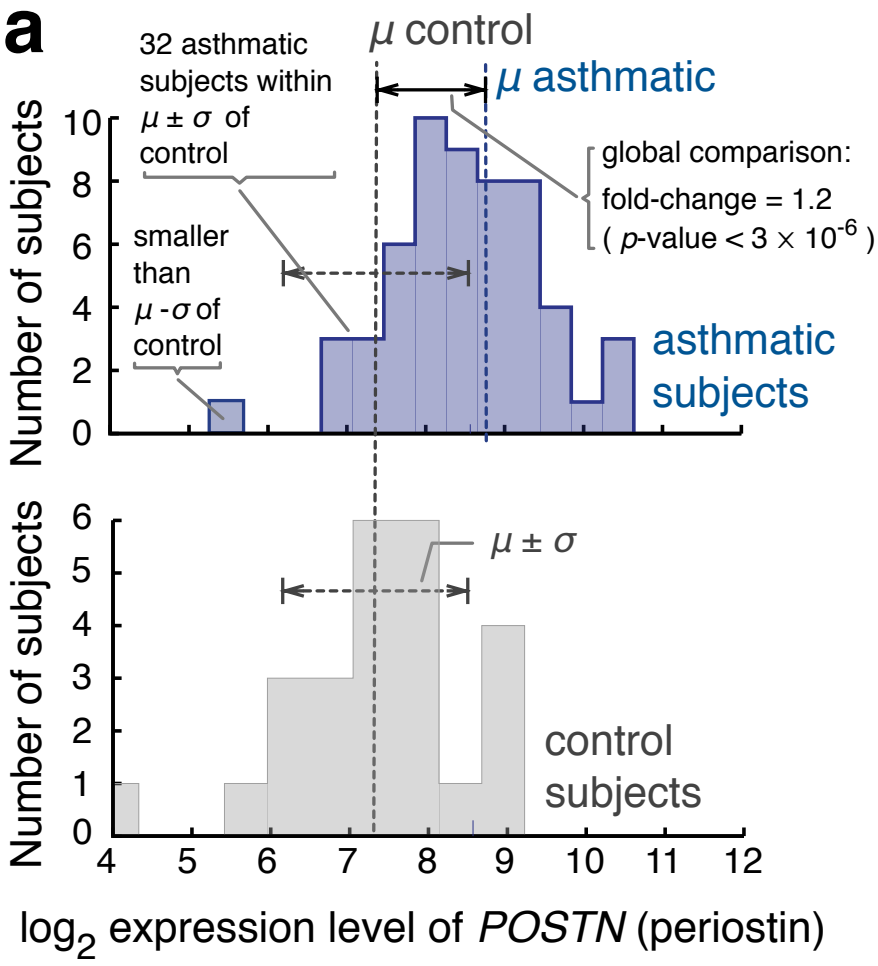
Image from firebox.com

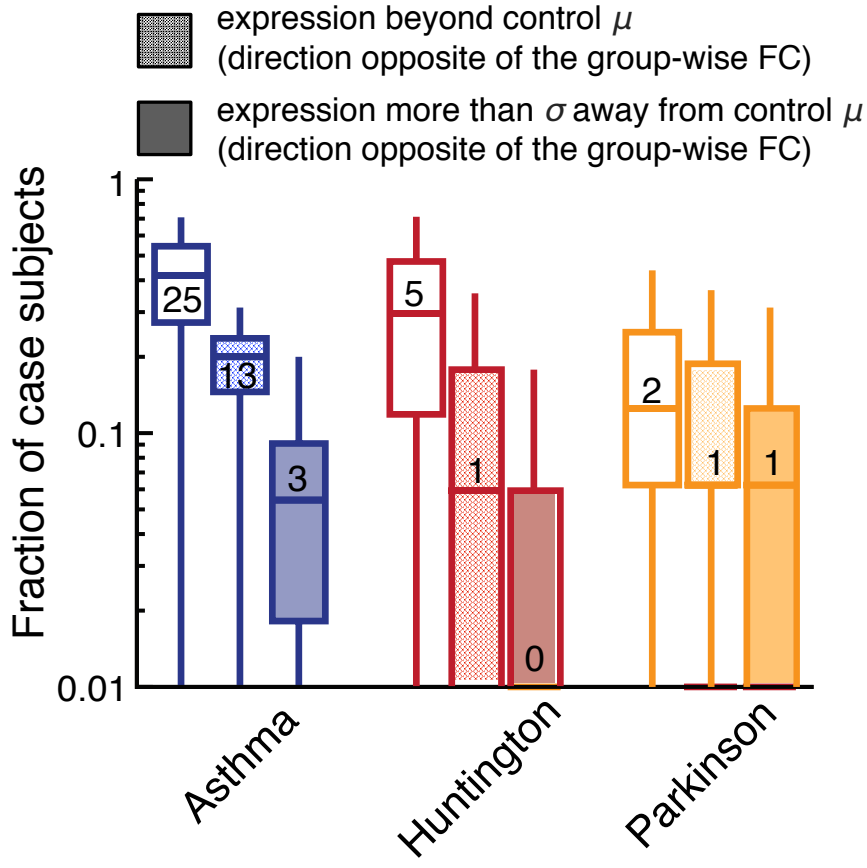
Patient-level heterogeneity



Schork, 2015, Nature

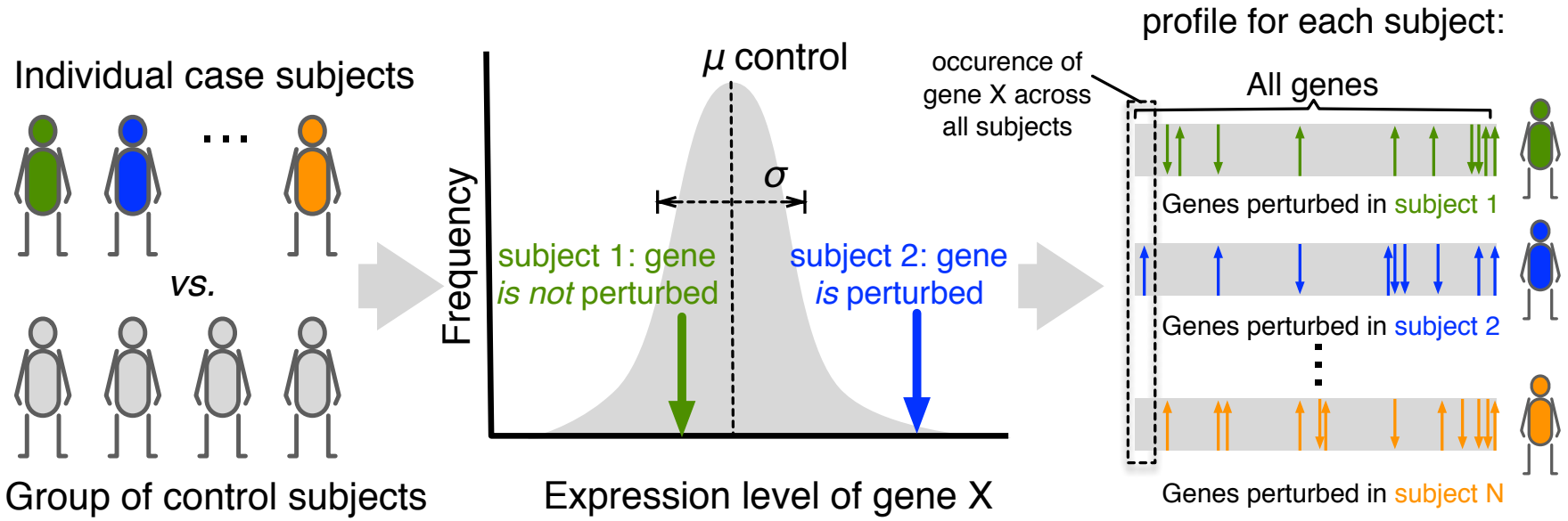
Group-wise differentially expressed genes do not capture transcriptomic heterogeneity





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

PeeP: PErsonalized Expression Profile

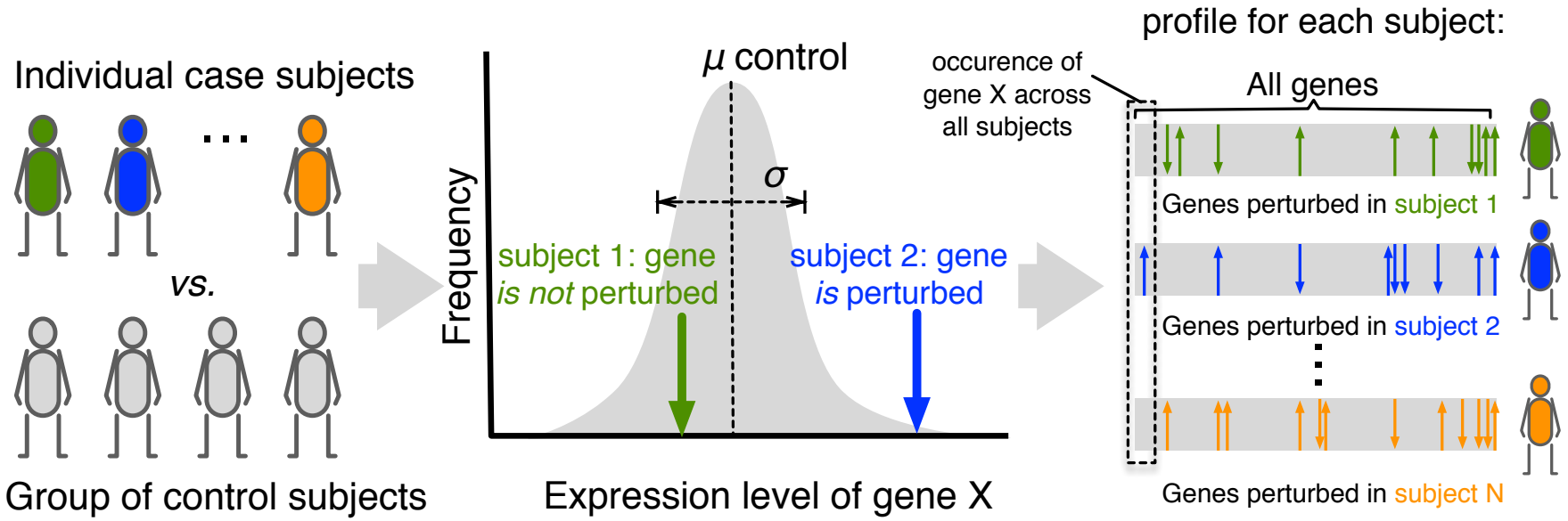


	Samples			
Case	x	x	...	x
Control	c	c	c	c

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

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

PeeP: PErsonalized Expression Profile



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

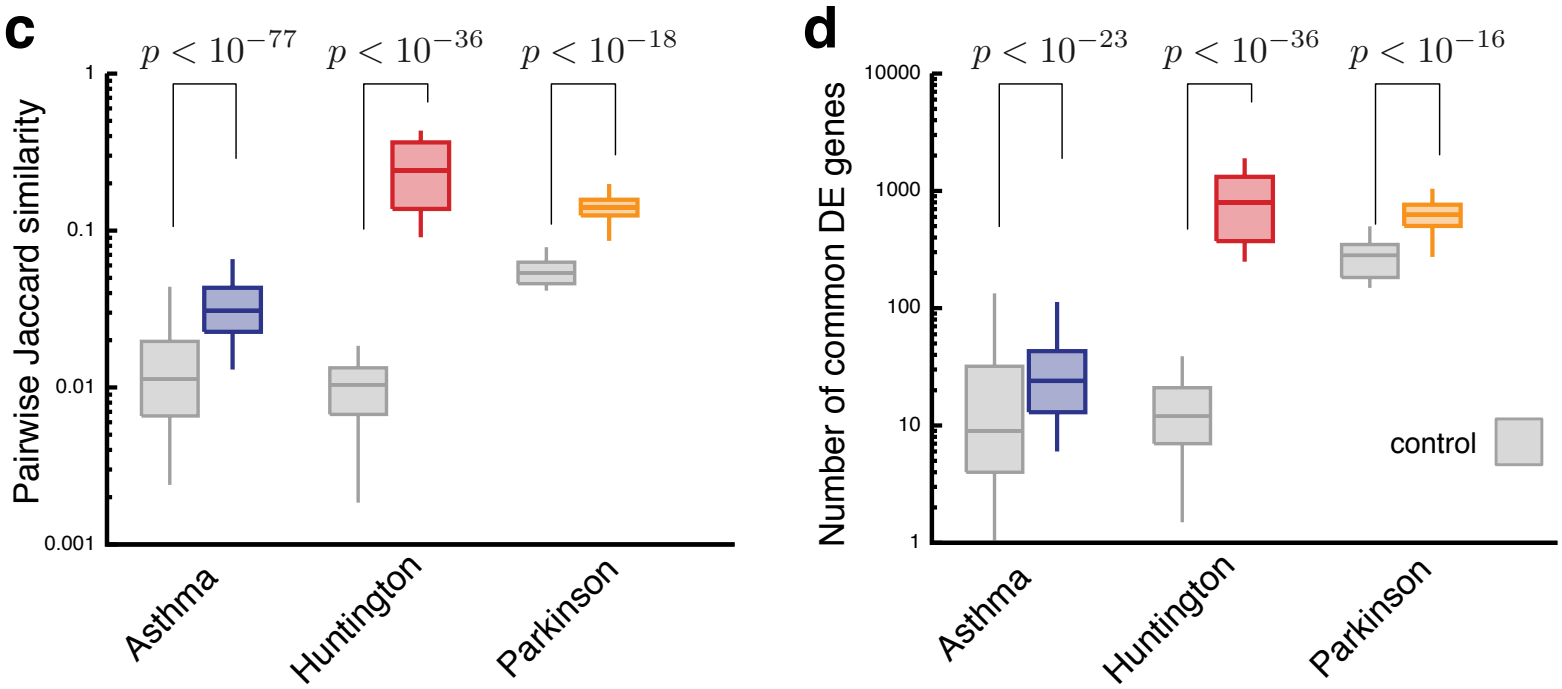
$$z(gene \text{ in } x) = \frac{expression_x(gene) - \mu_c(gene)}{\sigma_c(gene)}$$

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

[Genes that are significantly perturbed in each individual]

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

Quantifying the heterogeneity using PeePs

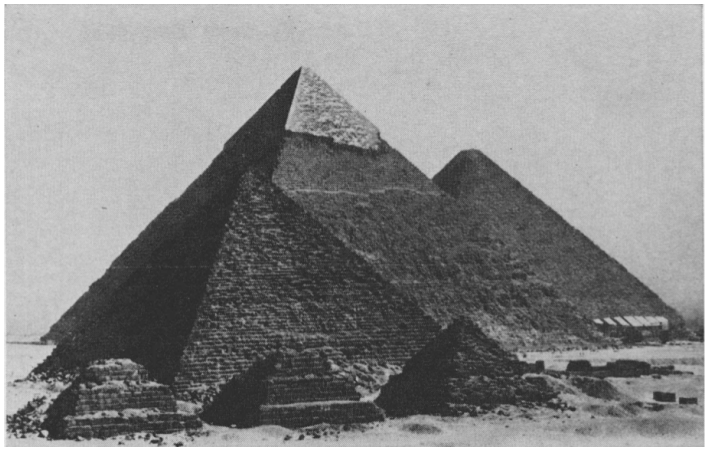


The overlap between PeePs of two individuals with the same disease

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

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

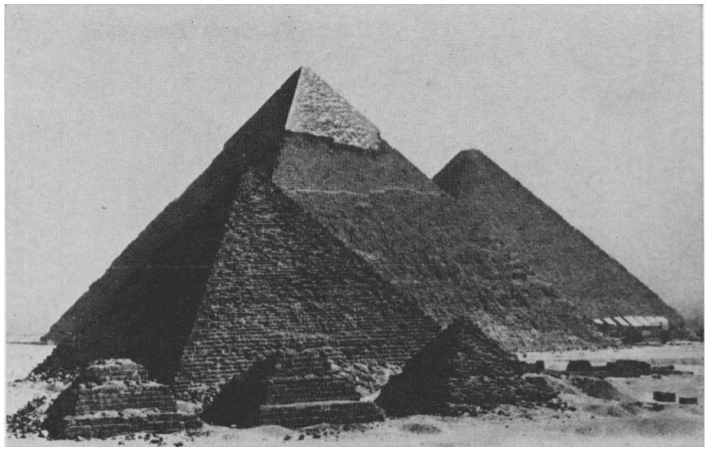
Limitations of similarity-based approaches



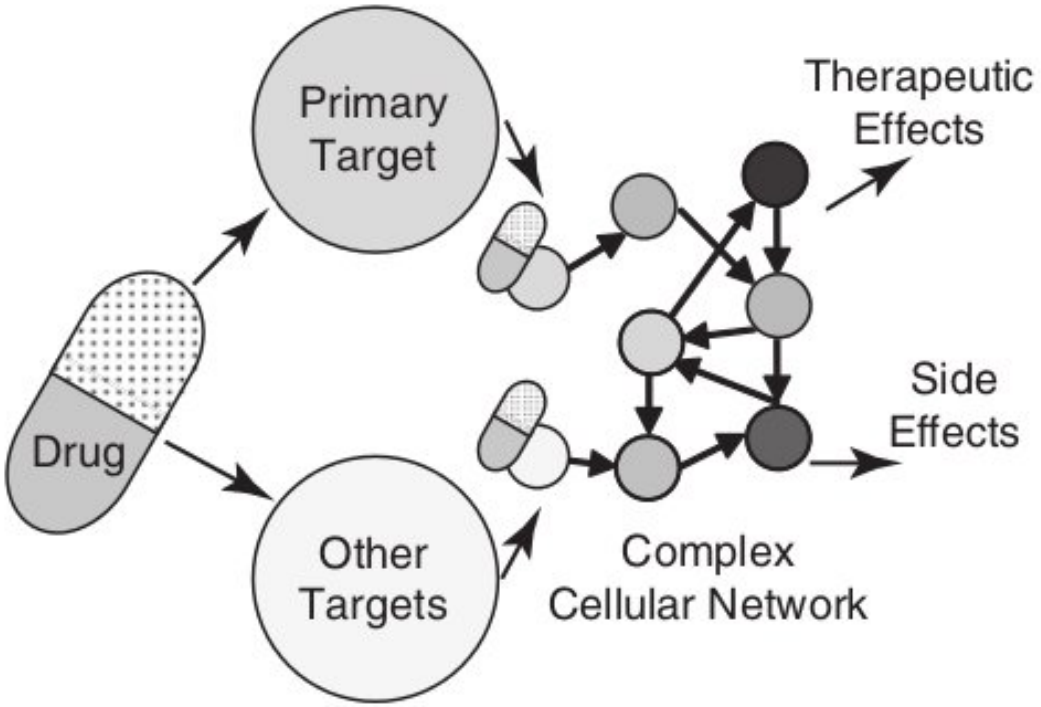
- Heterogeneity among disease phenotypes and patients
- Interpretability of the underlying model

Image from firebox.com

Limitations of similarity-based approaches



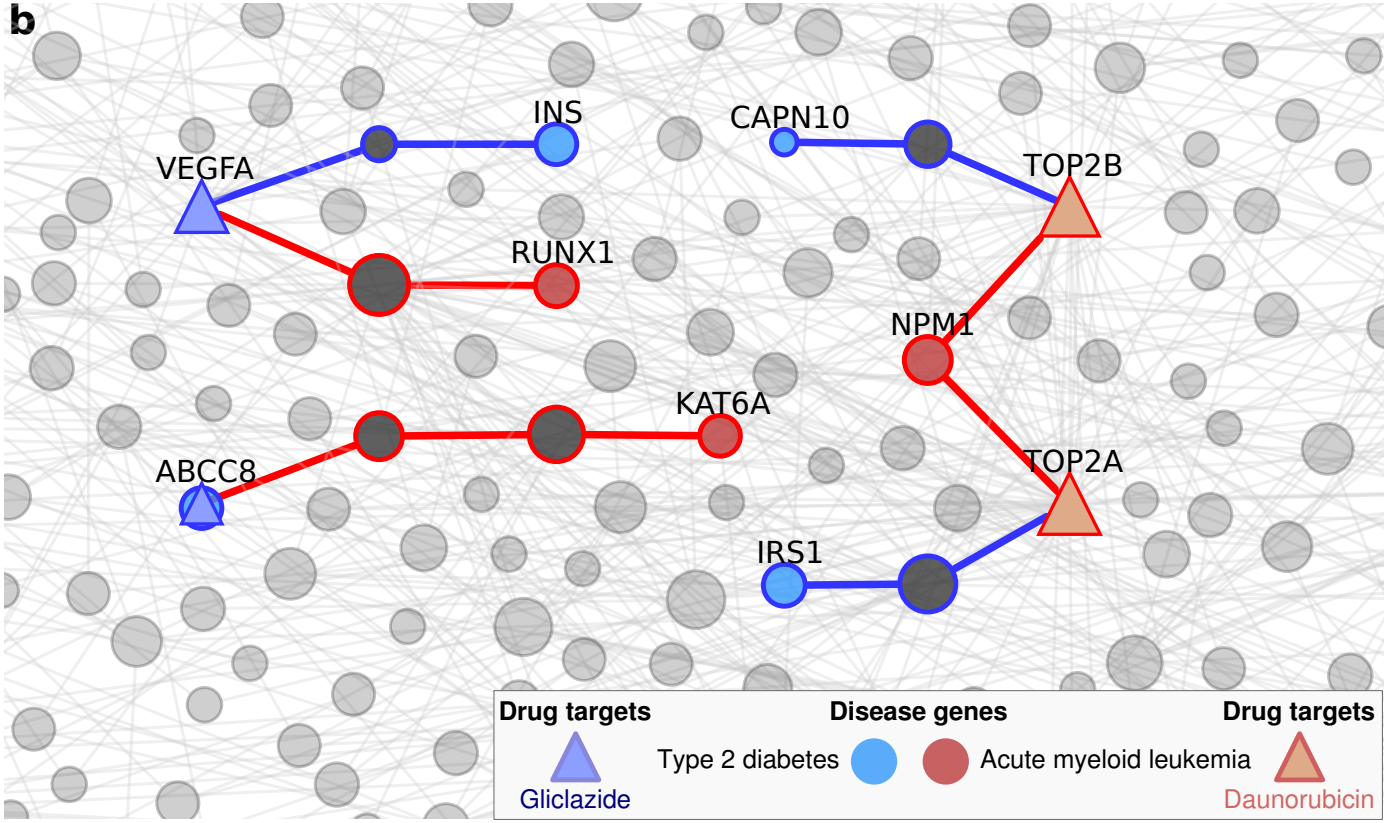
Systems Pharmacology View of Drug Action

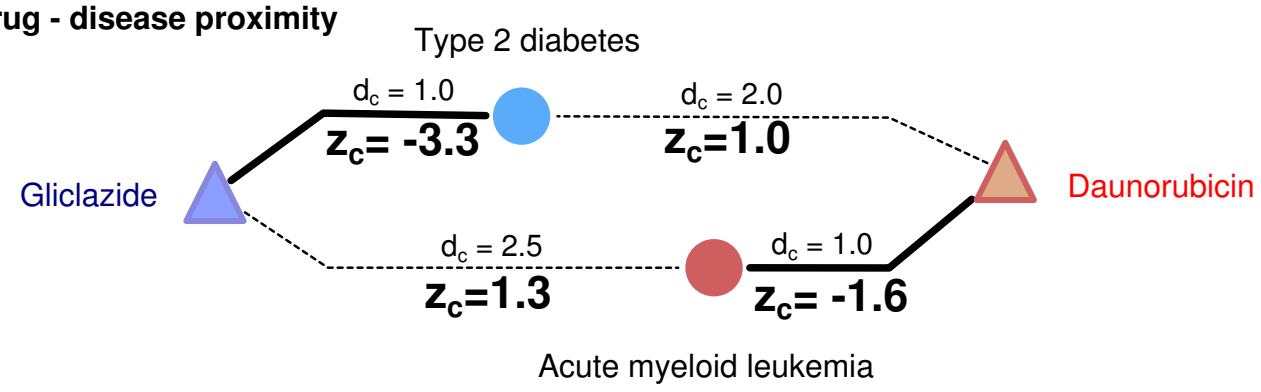


Interpretability of the underlying model

Image from firebox.com

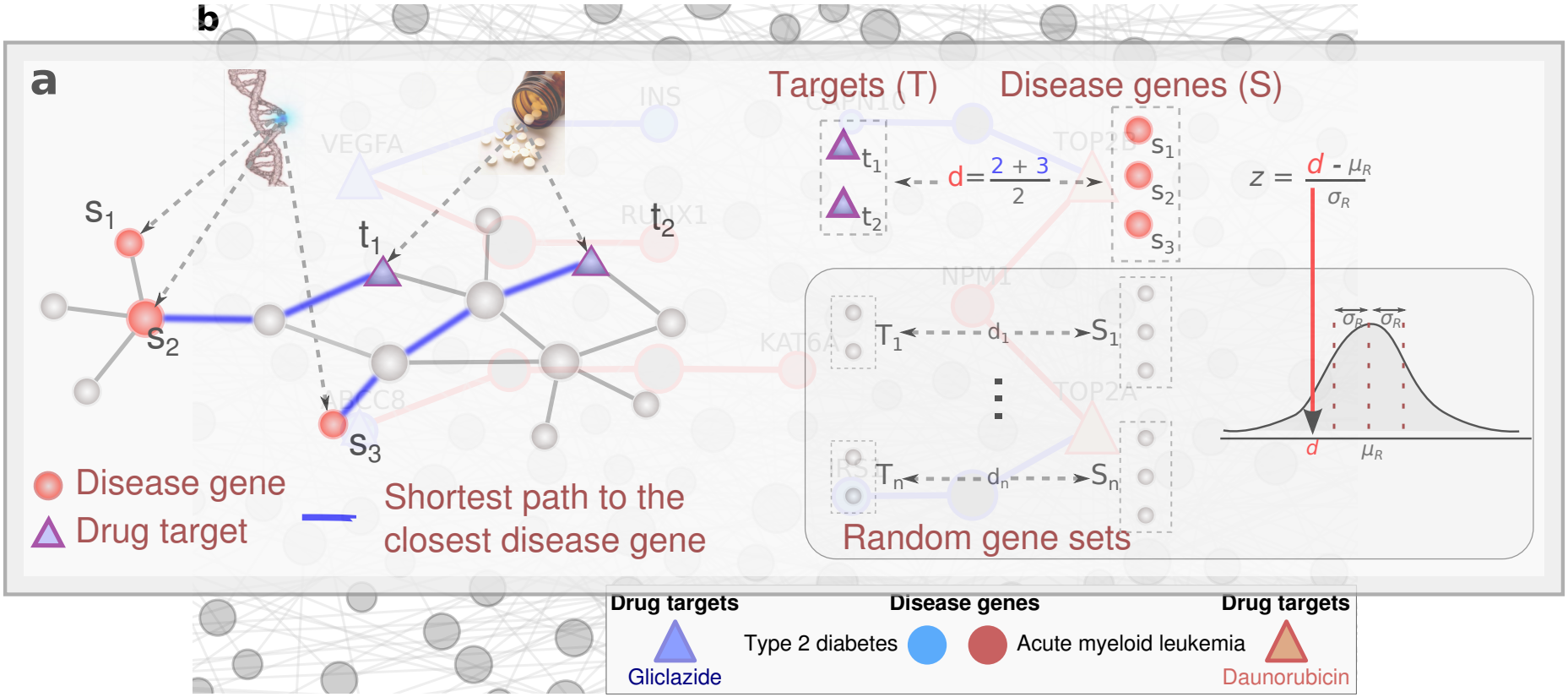
Modeling drug effect via interactome-based proximity



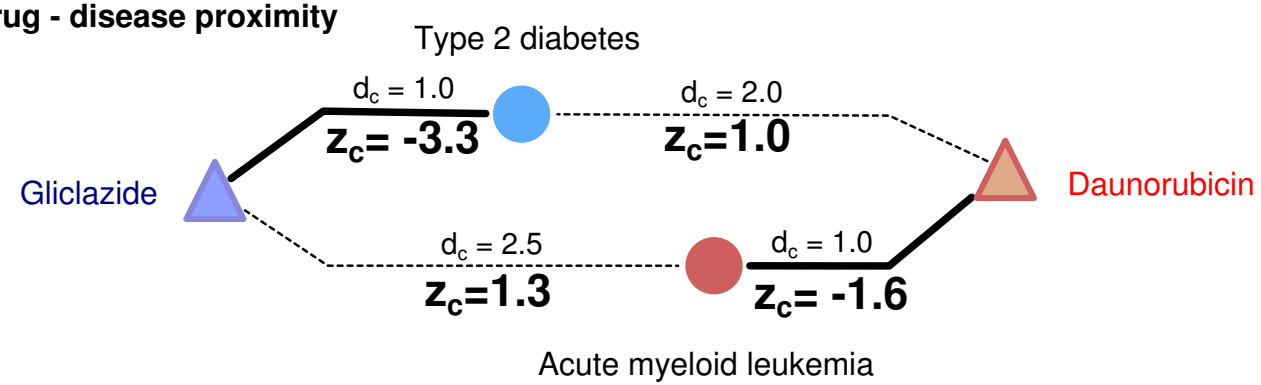


Guney et al., 2016, Nat Comm

Modeling drug effect via interactome-based proximity

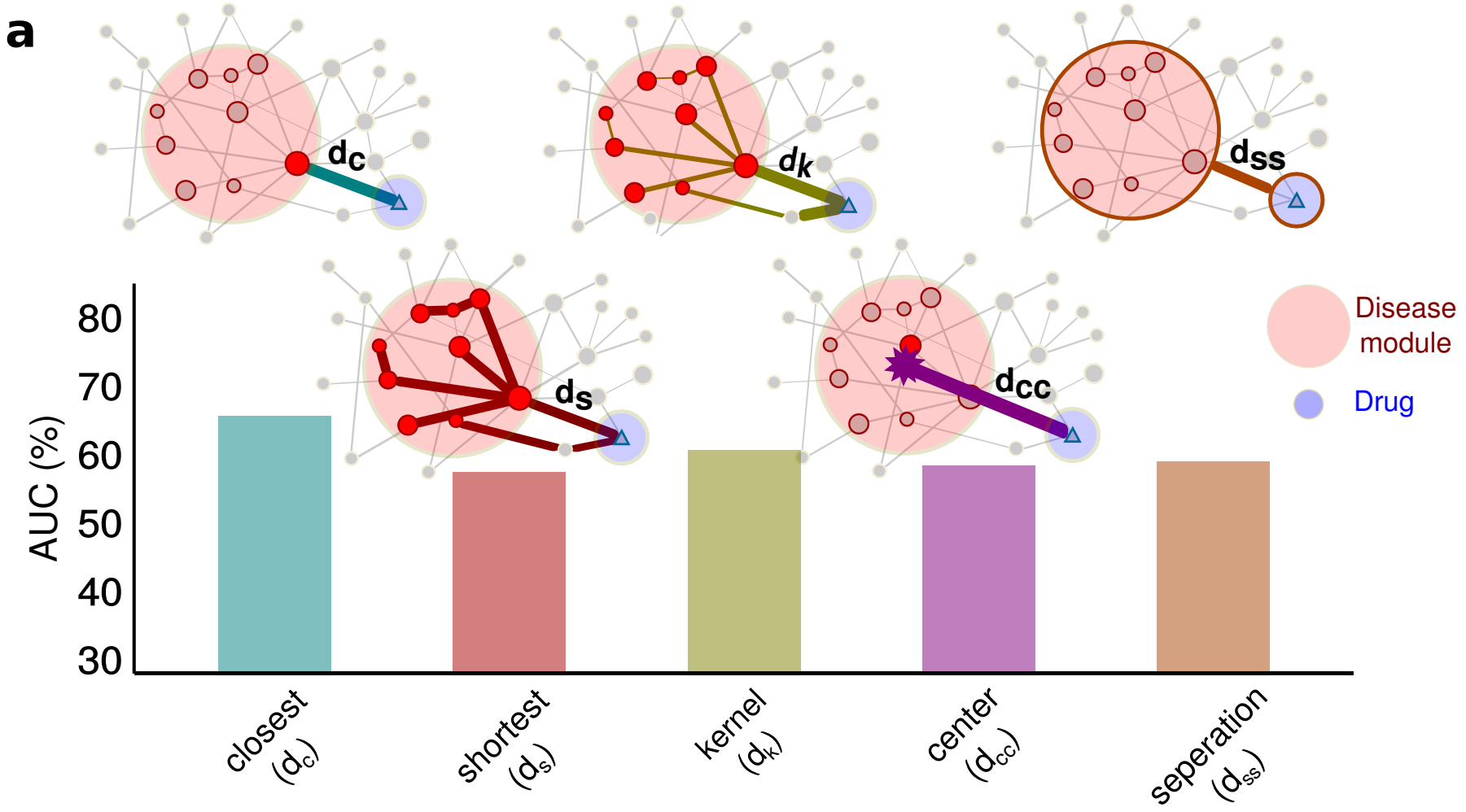


c Drug - disease proximity



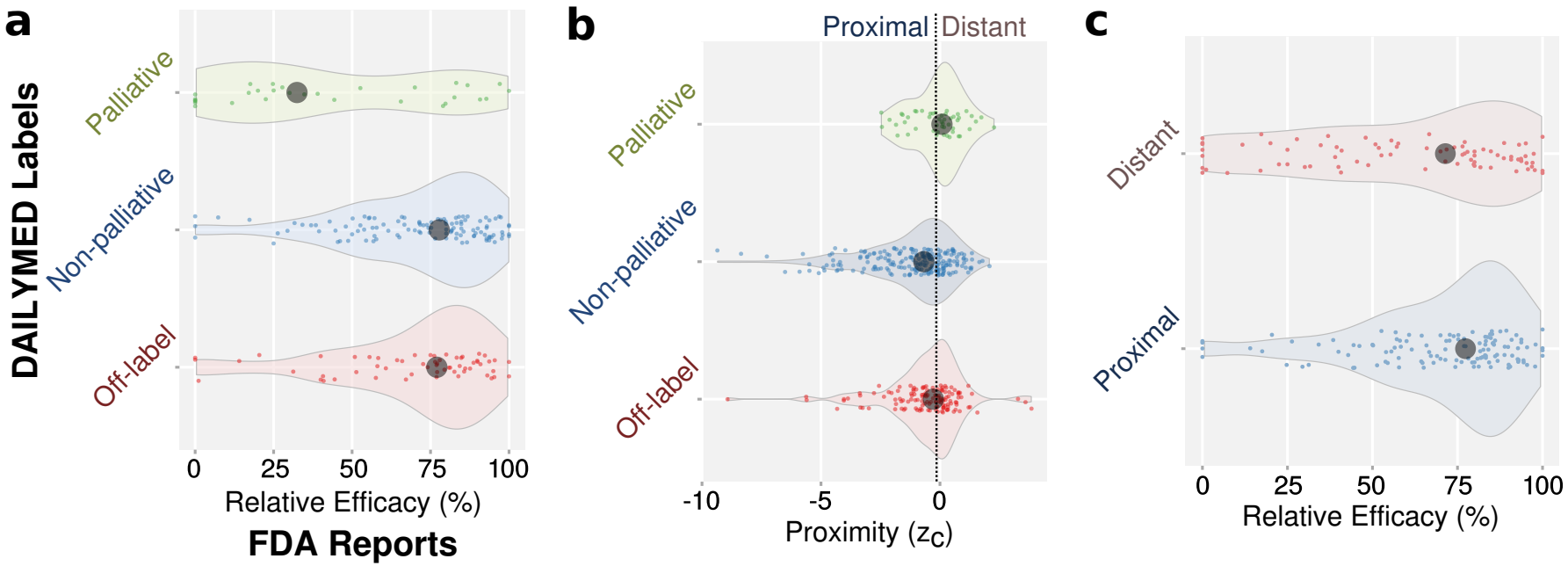
Guney et al., 2016, Nat Comm

Drugs do not target the disease module as a whole



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

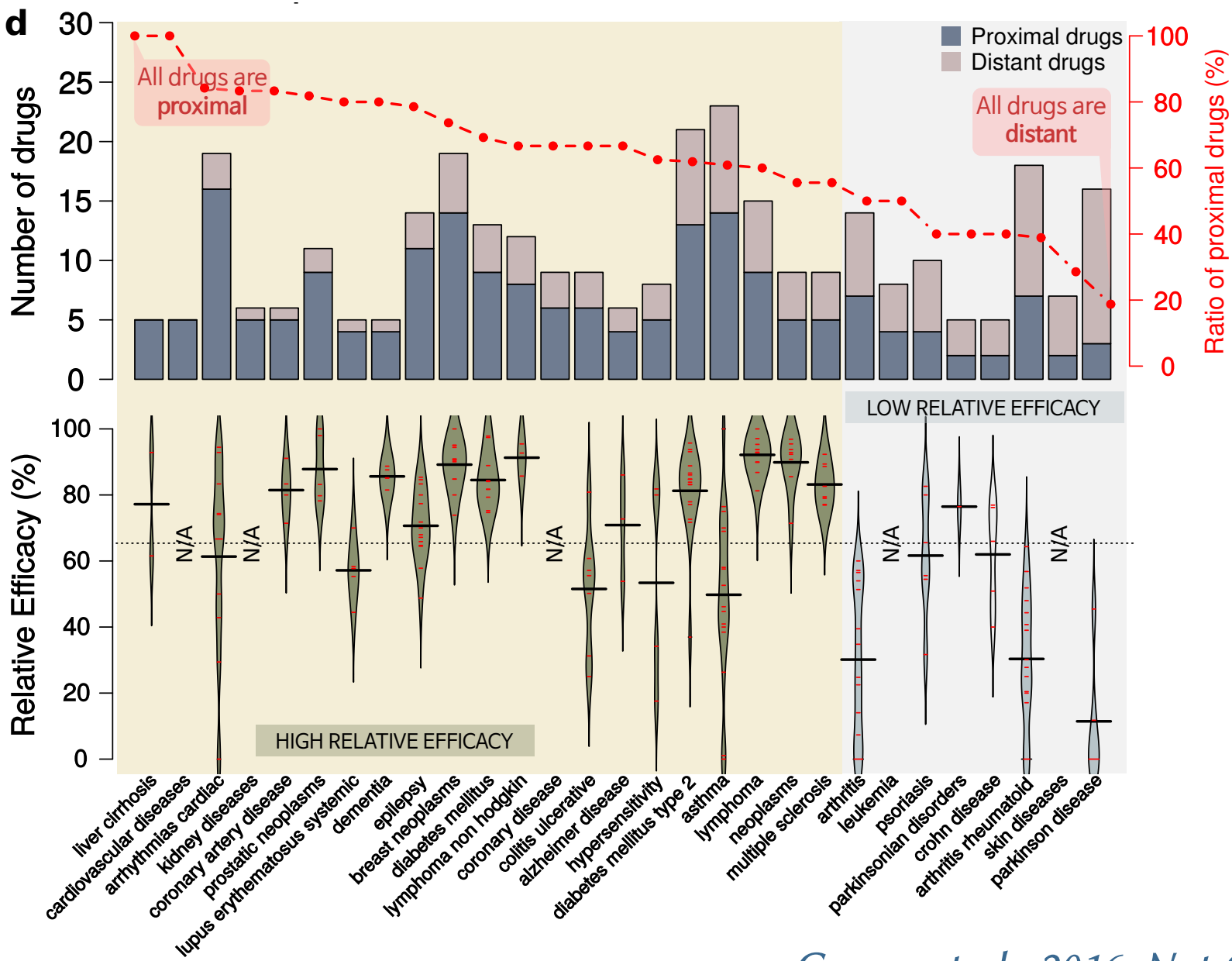
$P = 7.3x10^{-5}$
$P = 7.6x10^{-4}$
(n = 50, 219, 133)

$P = 4.0x10^{-5}$
$P = 0.02$
(n = 50, 219, 133)

$P = 0.04$
(n = 237 vs 165)

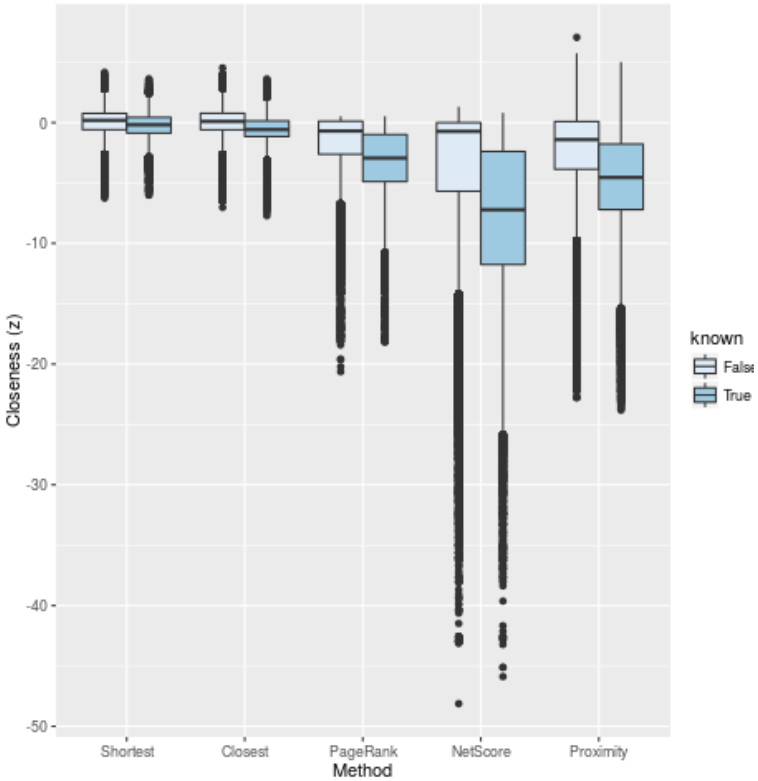
Guney et al., 2016, Nat Comm

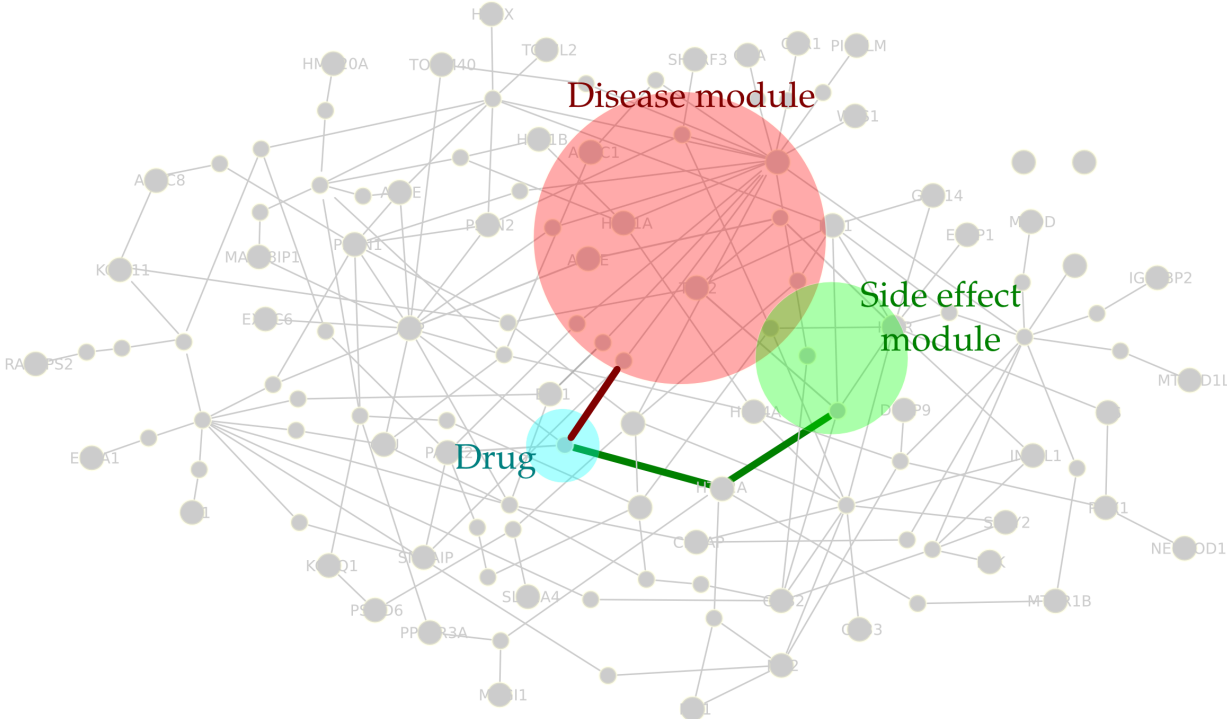
Proximity highlights treatment bottlenecks



Guney et al., 2016, Nat Comm

Predicting the directionality of drug effect using interactome-based modeling

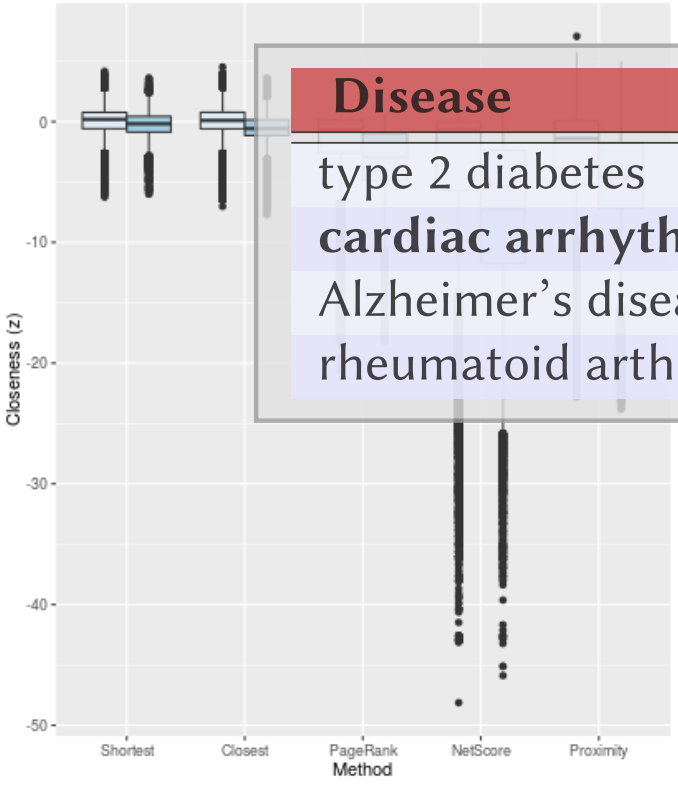




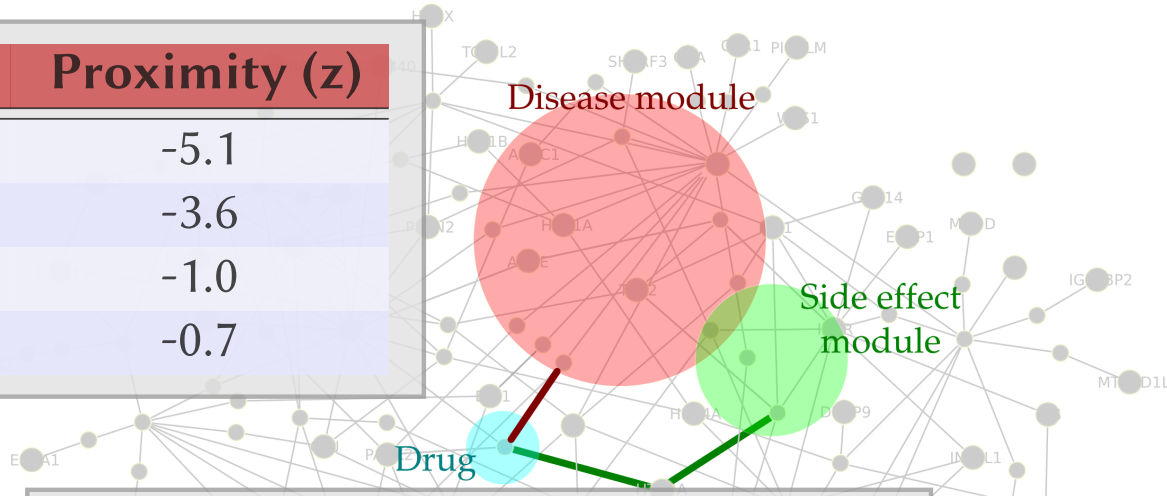
Guney et al., 2016, Nat Comm

Guney, 2017, Workshop on Complex Networks

Predicting the directionality of drug effect using interactome-based modeling



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



Side effect	Proximity (z)
hemolytic anemia	-5.8
arrhythmia	-1.9
vascular disease	-1.6
hypopigmentation	-1.5

Guney et al., 2016, Nat Comm

Guney, 2017, Workshop on Complex Networks

Understanding relationships between diseases using interactome-based modeling

Donepezil (Alzheimer disease)

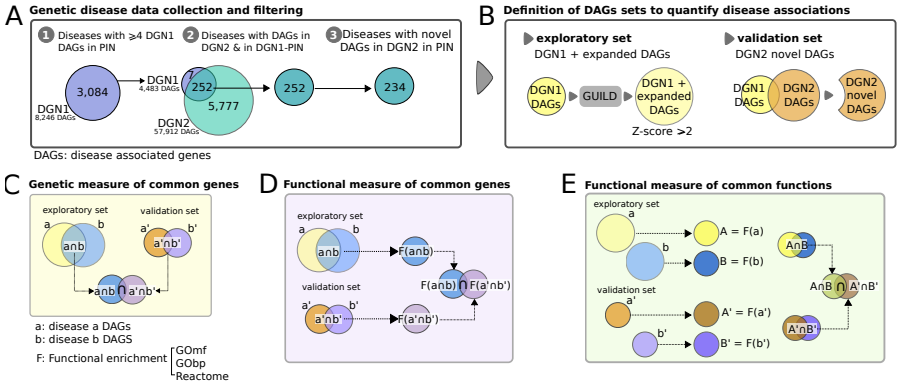
Pathway	n	z
synthesis of phosphatidylcholine	11	-3.3
serotonin receptors	11	-3.3
adenylate cyclase inhibitory pathway	13	-2.2
IL-6 signaling	10	-2.1
the NLRP3 inflammasome	11	-2.1
regulation of insulin secretion by acetylcholine	10	-2.1

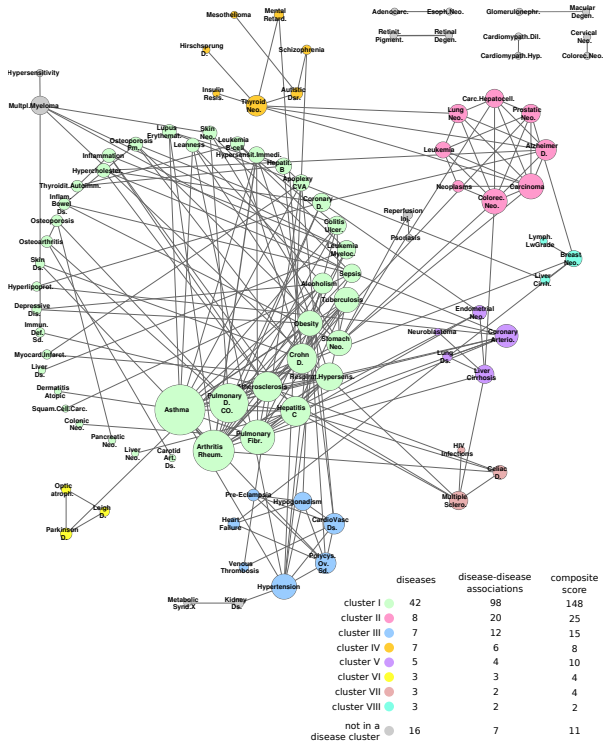
Glyburide (Type 2 diabetes)

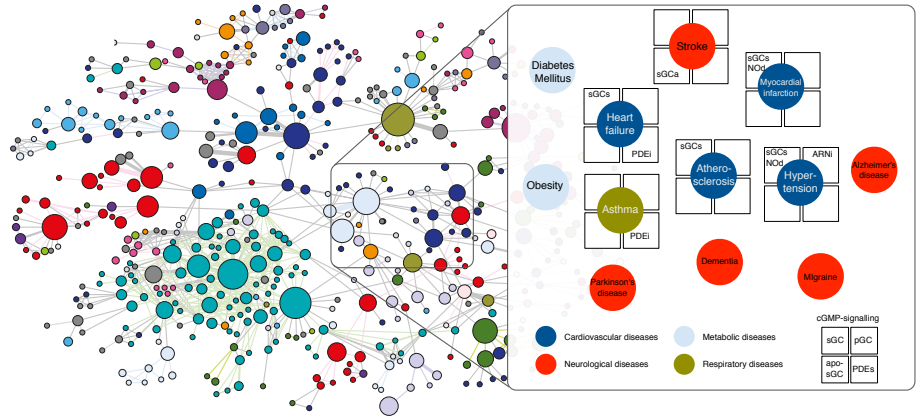
Pathway	n	z
inwardly rectifying K^+ channels	30	-9.0
ABC family proteins mediated transport	22	-8.5
Inhibition of voltage gated Ca^{+2} channels via G beta gamma subunits	25	-4.3
$GABA_B$ receptor activation	38	-4.1
regulation of insulin secretion by acetylcholine	10	-3.3
Na^+/Cl^- dependent neurotransmitter transporters	9	-3.3

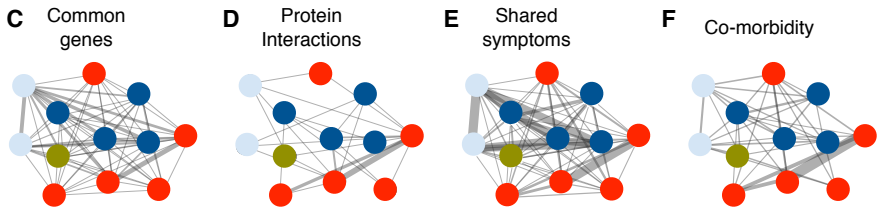
Guney et al., 2016, Nat Comm

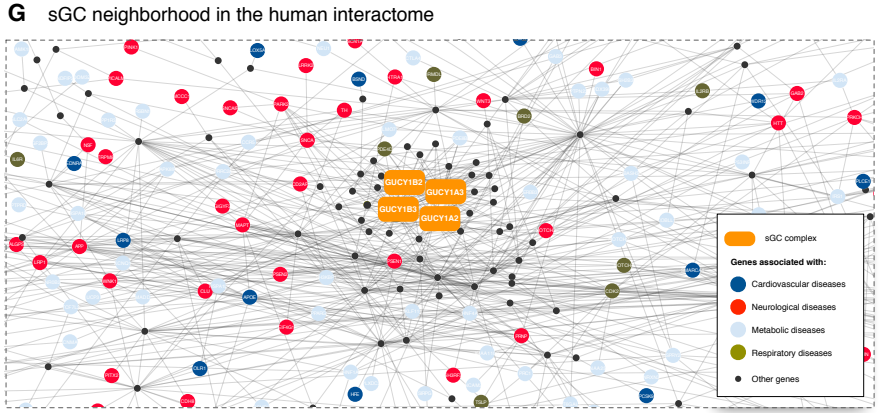
Leveraging disease-disease relationships for drug repurposing









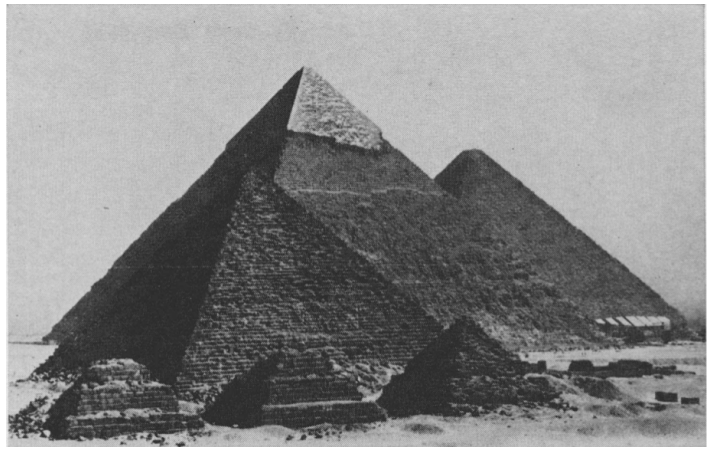


Guney and Oliva, 2014, PLoS ONE

Rubio-Perez et al., 2017, Sci Rep

Langhauser et al., in press

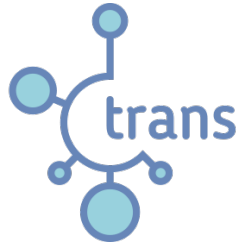
In summary...



Challenge = Opportunity

- Data-driven models are powerful, yet prone to overfitting (especially on small data sets)
- PeePs quantify transcriptomic heterogeneity across patients
- Interactome-based modeling can offer improved interpretability

Acknowledgements



UPF
Baldo Oliva

CeMM
Jörg Menche

NEU
Albert-László Barabási

IRB
Patrick Aloy

Maastricht Uni.
Harald Schmidt

DFCI
Marc Vidal

github.com/emreg00