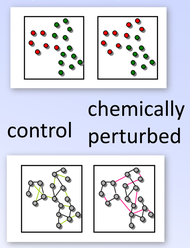
NETWORK-BASED ANALYSIS OF CHEMICAL PERTURBATION EXPERIMENTS

Francesca Mulas[1], Daniel Gusenleitner[1], David Sherr[2], Stefano Monti[1]

[1]Section of Computational Biomedicine, Boston University, Boston, USA
[2]School of Medicine, Boston University, Boston, USA

INTRODUCTION

Methods for the analysis of networks allow identifying groups of tightly connected genes whose activity may be altered during disease progression or due to chemical agents. Connectivity-based comparisons help identify "aggregate changes" that could be missed by standard methods of differential analysis comparing individual genes. In this work we compared networks obtained from wild type liver samples and from samples collected after the exposure to chemicals with varying carcinogenicity and genotoxicity.

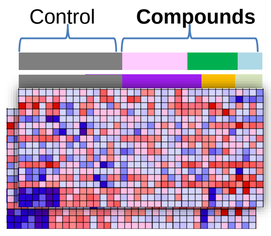


METHODS

Data sets

Two large cohorts of organ-specific **gene expression** profiles from controls and from rats exposed to chemical compounds

DrugMatrix	TG-GATEs
11244 expression profiles	11244 expression profiles
376 chemicals	131 chemicals
Liver, kidney, heart, muscle, hepatocytes	Liver, kidney, hepatocytes

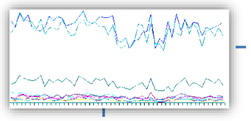


cause **DNA damage**
Genotoxic/ Non genot / NA
Carcinogen / Non carc / NA
cause **cancer**
Chemical classes

Liver samples:
- Control: ≥1000 observations
- Perturbations: ~10-100 observations

Network and modules inference

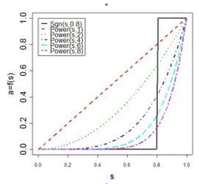
Pearson Correlation between gene expression profiles



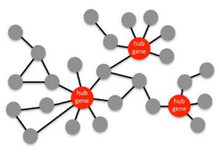
a) CN

b) SFN

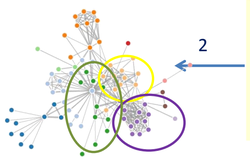
Scale free transformations

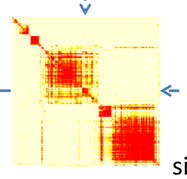


Threshold on correlation to fit a "scale free" topology



Dynamic tree cutting clustering



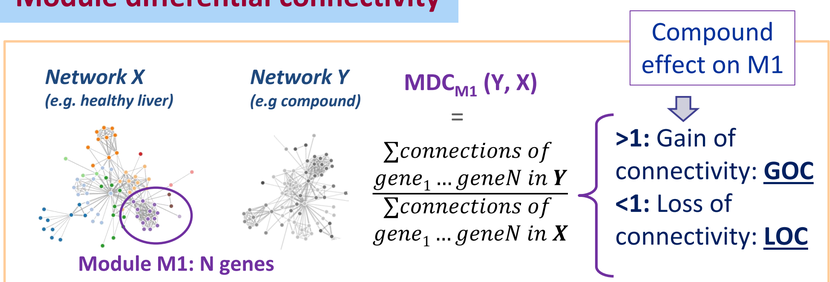


Gene similarity matrix

REFERENCES

[1] Zhang B et al. (2013). Cell 153(3):707-20; [2] Schadt EE et al. (2009). Nat Rev Drug Discov 8: 286–295; [3] Gusenleitner D. et al. (2014). PLoS One. 24;9(7):e102579; [4] Zhang, B. et al. (2005). Stat. Appl. Genet. Mol. Biol. 4, e17.

Module differential connectivity



Validation

Systematic comparison of networks from the same tissue and conditions (liver control). Expected:
- Similar modules composition
- No differential connectivity
  → MDC = 1

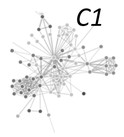
TG-GATEs — Scale free transformations

	Yes (SFN)	No (CN)
Avg. MDCs = 1	✓	✓
MDC variance	0.002286	**0.000214**
Not threshold-dependent	✗	✓

Compounds Analysis

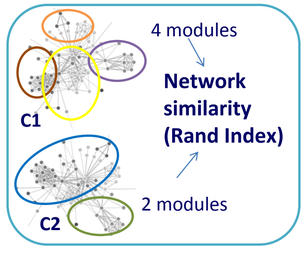
1. Compounds-specific network inference

DrugMatrix
Samples from compound "C1"
CN →

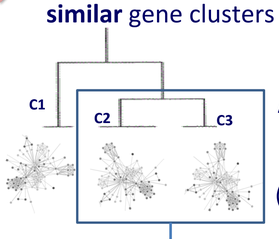


Repeat for C1, .., C62 with at least 10 samples

2. Grouping of similar compounds



Groups of **compounds** with **similar** gene clusters

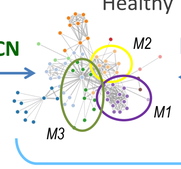


Aggregate C2 and C3 in a single network (increased sample size)

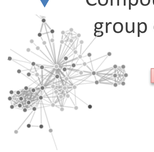
Repeat for all pairs of compounds

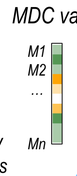
3. Differential analysis

Control samples
DrugMatrix
CN →

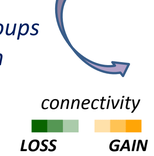


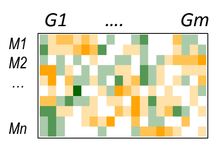
MDC ↔





Repeat for compound groups G1, … , Gm





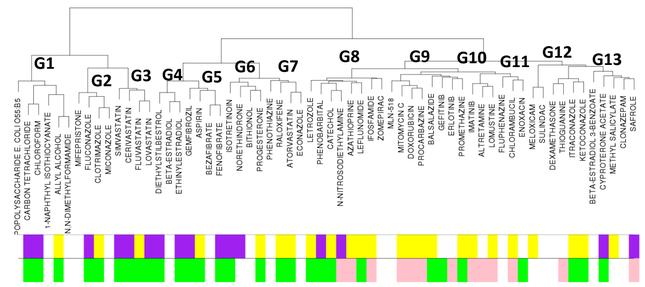
Biological function of gene modules (Enrichment)

RESULTS

2. Grouping of similar compounds

- Good separation of genotoxic and carcinogen compounds based on the similarity of their network structure
- 13 groups of compounds identified (Dynamic Tree Cutting)

Carcinogen
non Carcinogen
Genotoxic
non Genotoxic
NA



- Mostly homogeneous groups in terms of DrugBank actions
- Significance of overlap of drug-interacting proteins within each group (CTD interaction database, p-values: Fisher exact test).

G1: 4.6e-11*, **G2**: 7.7e-11*, **G3**: 6.1e-12*, **G4**: NA, **G5**: 3.6e-11*, **G6**: 5.1e-11*, **G7**: 5e-4, **G8**: 0.04, **G9**: 0.002, **G10**: 1e-6, **G11**: 0.001, **G12**: 3.9e-11*, **G13**: 0.01

* = lower p-value cannot be obtained by chance (permutations)

3. Differential analysis

- 6315 genes, 60 "healthy" gene modules
- Identification of top 3 modules specifically altered by each compounds group (scoring) → GO/KEGG Enrichment

GROUP	MAIN ACTION	FUNCTIONS OF SPECIFICALLY ALTERED GENE MODULES
G1	**SOLVENTS**	Response to chemical stimulus, signaling
G2	**ANTIFUNGALS**	Response to chemical stimulus, development
G3	**STATINS**	Steroid biosynthesis, lipid metabolism
G4	**ESTROGENS**	DNA replication, development
G5	**FIBRATES**	Fatty acid metabolism
G6	**n.c. (estrogen, antifung.)**	mRNA processing, proliferation
G7	**STEROIDS**	Development, transcription
G8	**ANTI-CANCER**	Cell development, apoptosis, signaling
G9	**CHEMOTERAPEUTICS**	Cell. organization, DNA damage response
G10	**ALKILATING- CANCER**	Reg. of transcription, cancer pathways
G11	**n.c (anti-cancer, estrog)**	DNA replication, response to stimulus
G12	**ANTI-INFLAMM/ FUNGAL**	Chemotaxis, immune response, development
G13	**ANTISEPTICS / ESTROGENS**	Development, DNA replication

Pathways actually related to the chemicals' action. Examples of altered connectivity:
- GOC of **lipid metabolism** modules when Statins or Fibrates are used to reduce cholesterol;
- LOC of **cell cycle** modules in response to Chemotherapeutics