LINPS: A database for cancer cell-specific perturbations of biological networks

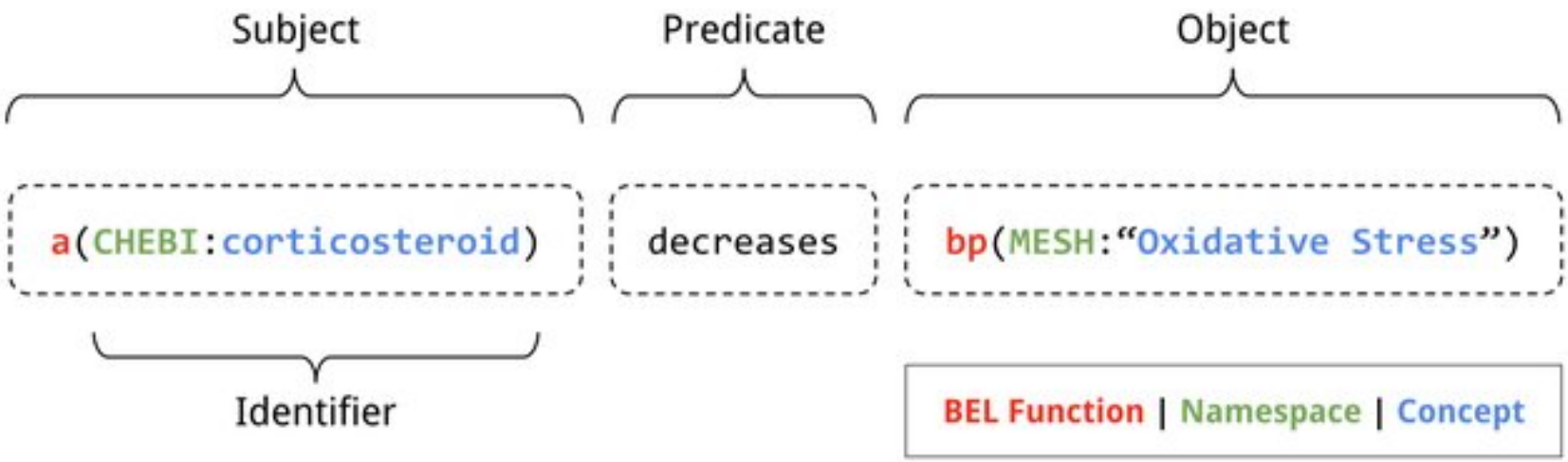
Mahmoud Ahmed

Gyeongsang National University School of Medicine

April 22, 2022

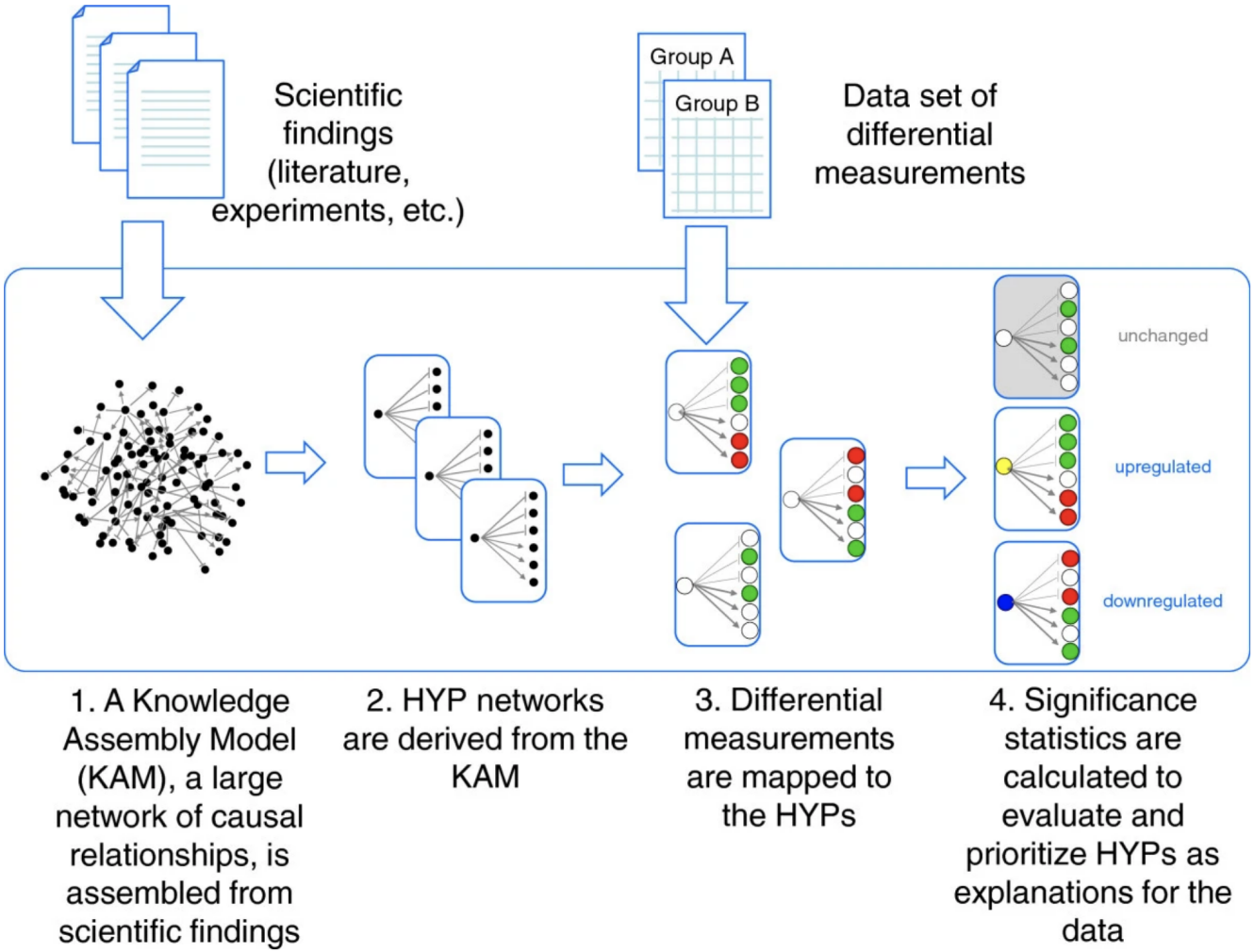
Outline

Biological Expression Language (BEL)



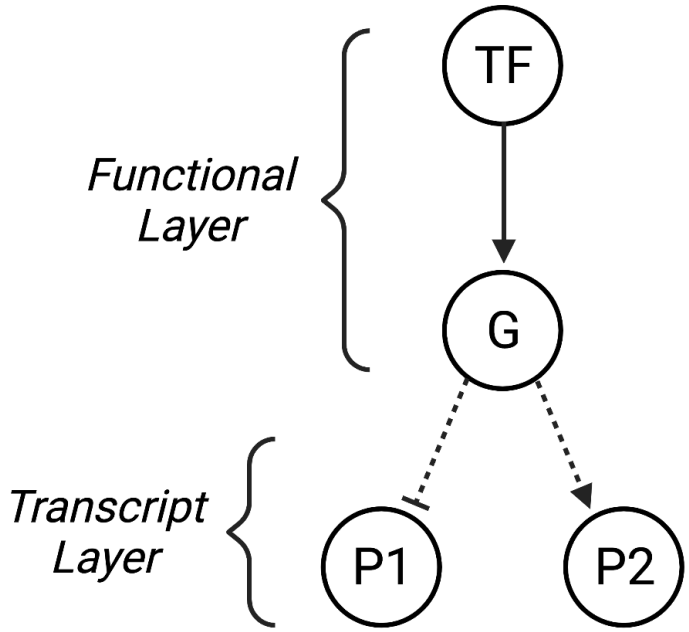
(Mubeen, 2018)

Reverse Reasoning



(Catlett et al., 2013)

Using reverse reasoning to model drug effects



Drug Perturbations (fold-change)

	D 1	D 2
+1	0	-1
-1, +1	0, +1	+1, -1

p(TF) --> r(G)
act(p(G)) -| EXP(P1)
act(p(G)) -> EXP(P2)

The activity of the functional layer is inferred from the changes in the transcript layer.

Datasets

Table 1: Description of the data sources.

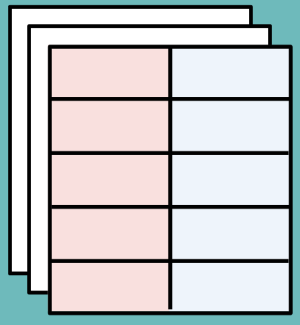
	Cells			Perturbations			Networks		
Desc.	Gene expression profiles of cell lines from different cancer tissue and metadata about the cell phenotypes.			Control vs. treatment gene expression profiles of cancer cell lines and metadata about the treatments.			Two-layers causal biological networks of cellular processes.		
Content		N	Example		(N)	Example		(N)	Example
	Tissue	12	Breast/Skin	MOA	530	NSAID/Anti-ER	Family	5	IPN/CPR
	Cell	70	MCF7/A375	Drug	1938	Asprin/Tamoxifen	Model	8	Apoptosis
Source	LINCS (Koleti et al., 2018)						CBN	(Boué et al., 2015)	

Workflow

1 **Modeling**

Differential Expression

LIMMA: FC & PVal

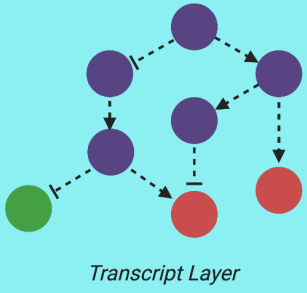


TRT vs. CTR

2 **Scoring**

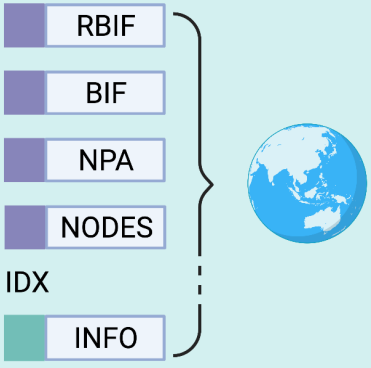
Network Pertubation

NPA: Coefs, CI & PVal



3 **Packaging**

Database tables & Indicies



Scoring the networks with drug perturbations

- Assume the function of an entity (e.g. a transcription factor) is reflected in the downstream gene expression.
- We construct a network of two layers to describe a pathway.
  - A functional layer encodes the causal relations between entities.
  - A transcription layer with all the nodes downstream from each node in the backbone.
- Network perturbation amplitude (NPA) models the perturbation in the nodes of a biological network as the changes in expression of their known downstream nodes (Martin et al., 2014).
- This analysis was applied using an R package with the same name (Martin et al., 2019).

Network Perturbation Amplitudes

From (Martin et al., 2019)

- A constraint optimization problem.
- Differential values should be close to each other (smooth) and equal to the fold-change β in the transcript layer (V_0).
- Differential values are calculated by solving:

$$min_{f \in l^2(V)} \sum_{x \to y} (f(x) - \sigma(x \to y).f(y))^2$$

where $\sigma(x \to y)$ is the sign of the edge $x \to y$, such that

$$f\Big|_{V_0} = \beta$$

- NPA is sum over the edges of the functional layer

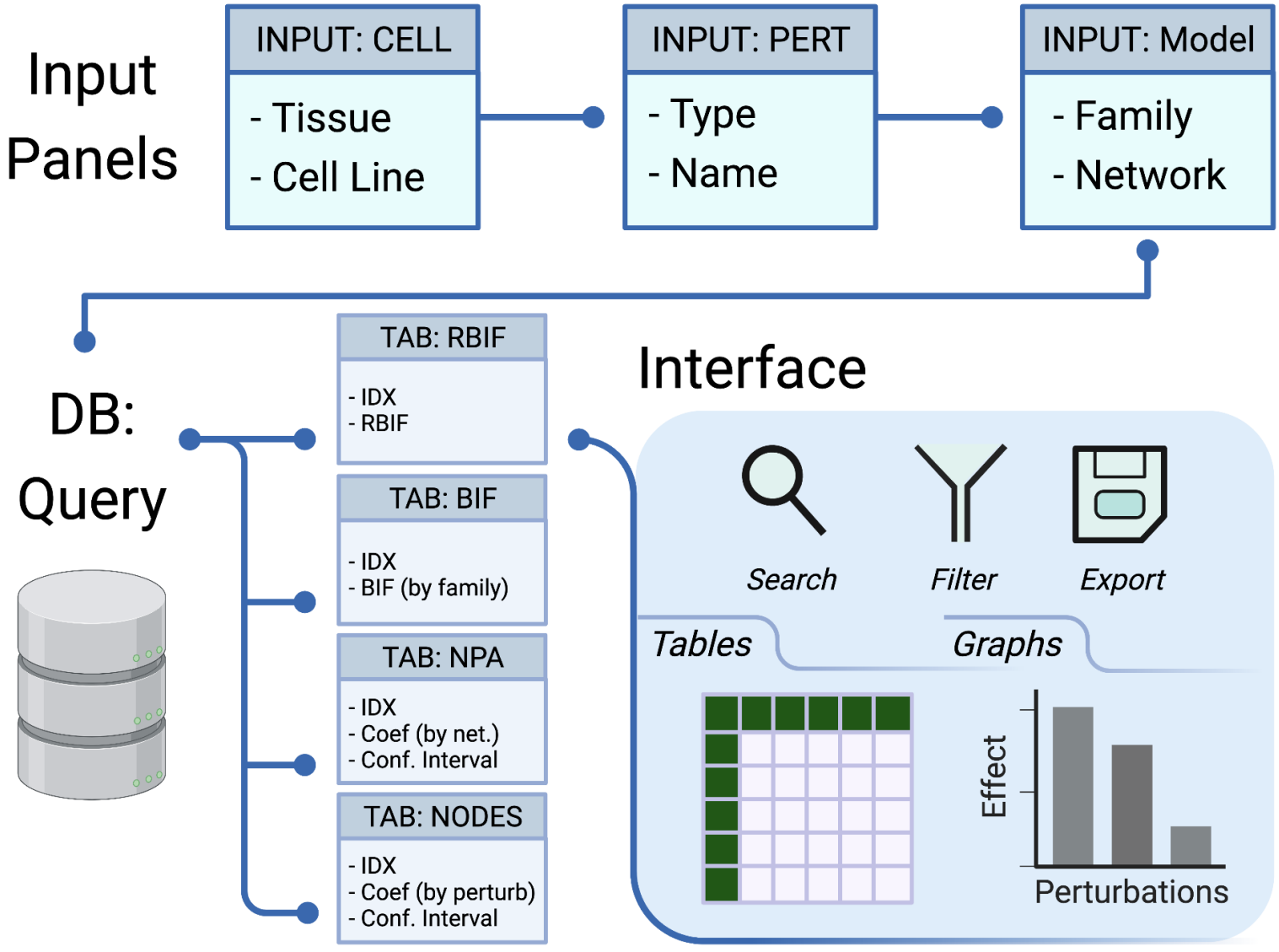
$$NPA = \frac{1}{|E|} \sum_{e \ in \ E} (f(e_0) + \sigma(e) f(e_1))^2$$

where $|E|$ is the size of the edges E, f is the solution of the problem, and e_0 and e_1 are the start and the end.

Interpretation

- Network Perturbation Amplitude (**NPA**) is sum of the differential values of all edges in the functional layer. The amplitude of the perturbation of the network as a whole.
- Leading **nodes** are the nodes with differential values that contribute the most to the NPA scores ($> 80\%$).
- Biological Impact Factor (**BIF**) is the weighted sum of the scores of the significantly perturbed networks (when multiple).

Database & Web interface description



Source-code and reproduciblity

- The analysis, scoring, and packaging was conducted mainly in R and using Bioconductor packages.
- The software environments were packaged into Docker images and made available at `https://hub.docker.com/repository/docker/bcmslab/linps` and `https://hub.docker.com/repository/docker/bcmslab/linpsapp`.
- The source code to build this image, the database, and the web application is open source (GPL-3) and is available at `https://github.com/BCMSLab/LINPS` and `https://github.com/BCMSLab/LINPSAPP`.
- Shiny app: `https://bcmslab.shinyapps.io/LINPSAPP/`

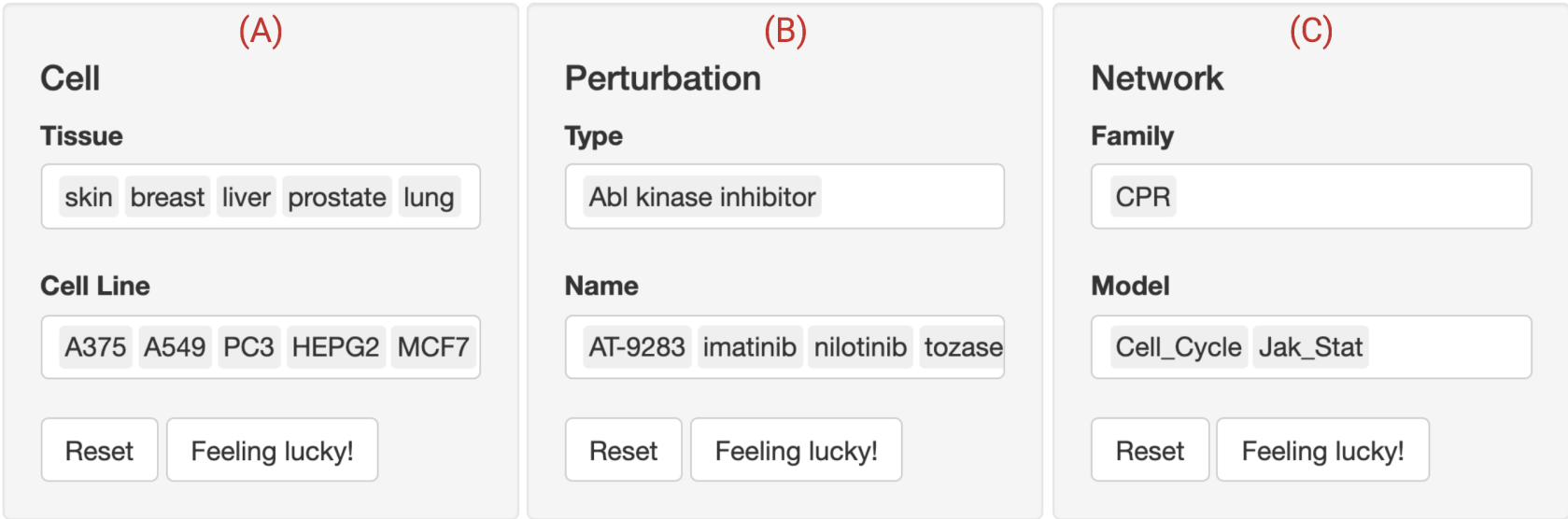
R/Bioconductor Packages

- Annotation and metadata
  - slinky
  - **GEOquery**
  - **org.Hs.eg.db**
- Data management
  - tidyverse & reshape2
  - DBI & RSQLite
  - **SummarizedExperiment**
- Analysis
  - **limma**
  - **NPA** & NPAModels
- Visualization
  - Plotly
  - igraph & visNetwork
- Interface
  - shiny & shinyBS
  - DT

Use case: tyrosine kinase inhibitors arrest the cell cycle in skin cancer cells

- Tyrosine kinase inhibitors target the Abl-Bcr kinase, a chimeric oncogene of the Abelson the Abl and Bcr gene at chromosome 9 and 22 common in chronic myelogenous leukemia (CML).
- The first drug to be developed was Imatinib. Other followed to overcome its resistance and limitations.
- Secondary and complementary mechanisms of action were reported to benefit patients with other pathologies.
- Here, I show a novel mechanisms of the Abl tyrosine kinase inhibitors by studying the effect of the drug treatment on cell cycle in cellular contexts other than their primary targets.

Choosing appropriate cells, biological networks and drugs inputs



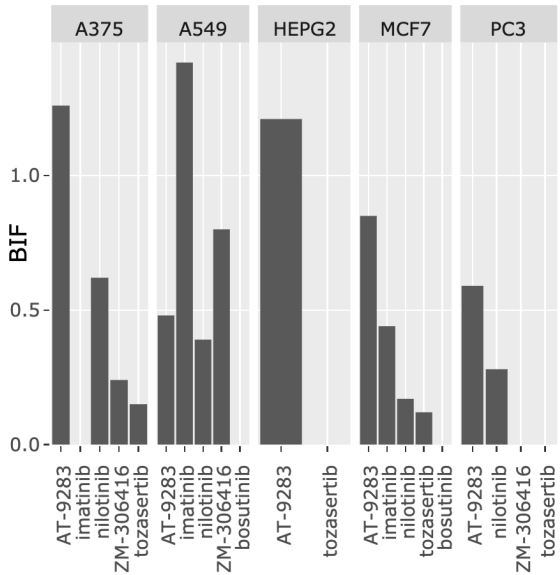
(A) Choose cell/s (Tissue/cell line)

(B) Choose drug/s (MOA/name)

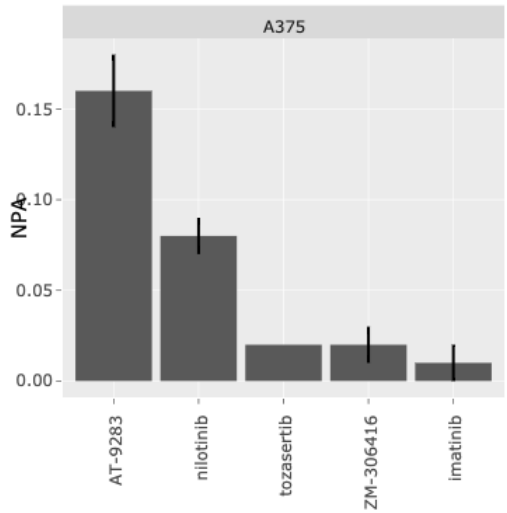
(C) Choose pathway (Family/network)

Impact factors, perturbation amplitudes & node contributions output

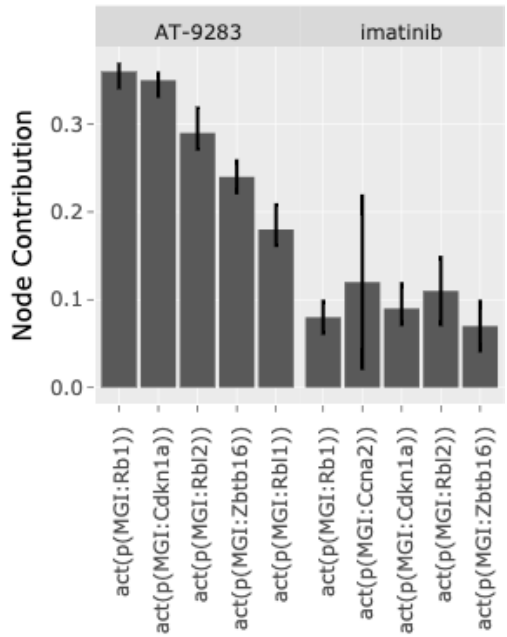
(A)



(B)



(C)



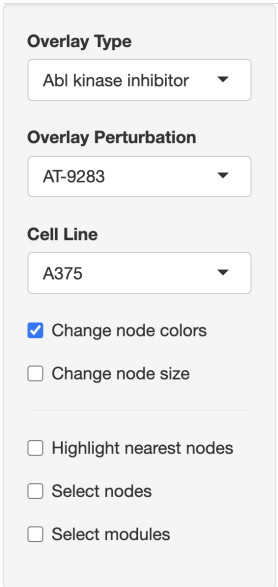
(A) Biological impact on both pathways in different cells

(B) Impact of each drug on cell cycle in A753

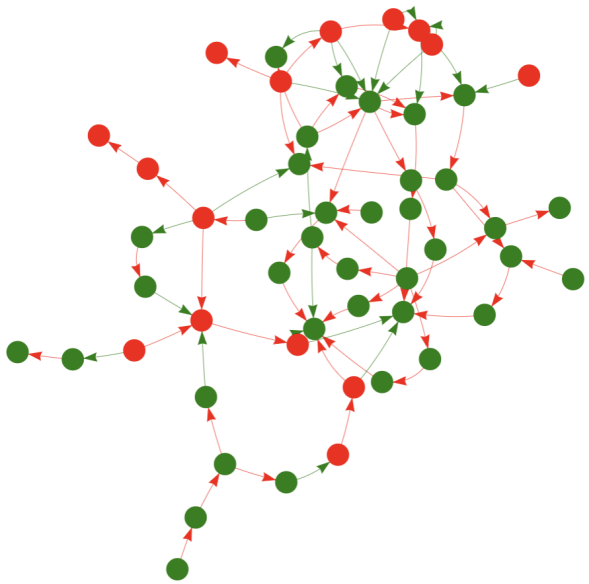
(C) Most contributing nodes to the cell cycle perturbation

Analyzing the perturbed network

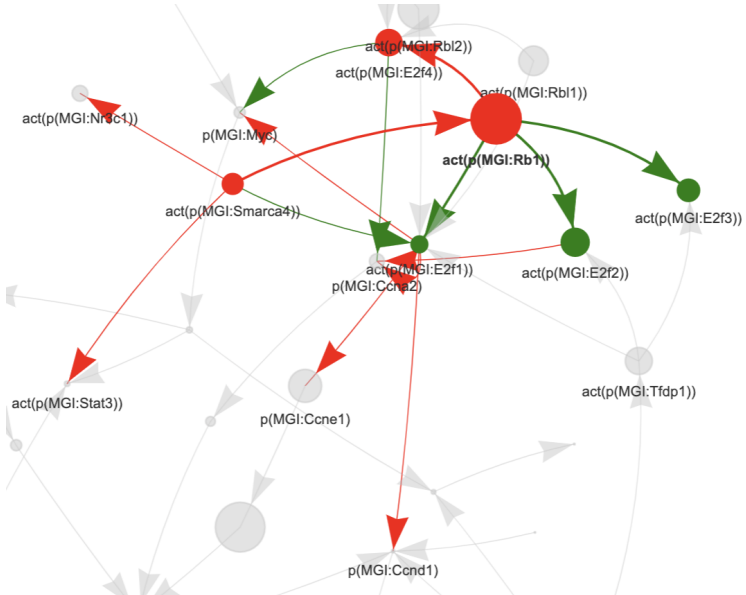
(D)



(E)



(F)



(D) Input panel

(E) The cell cycle overlaid with the effect of AT-9283 in A375

(F) A highlighted cluster of leading nodes

How to interpret the output

- Causal network encodes network nodes and the directed edges between them either manually or extracted from the literature.
- NPA method uses the gene expression of all the nodes connected to an entity to infer its activation or inhibition.
- Considering the two facts, we could determine that a particular node is activated, inhibited or unchanged to produce a specific function.
- Together, the **cell cycle** was arrested in response to **AT-9283** as it negatively regulates the E2F transcription factors through RB1. This effect was present to a lesser extent with other Abl inhibitors and was specific to the **A375** cell line.

Limitations

- This analysis is only valid insofar as the underlying data (gene expression and causal links) reflects the true biology of the conditions and biological functions.
- At this moment, it is not possible to create or customize the scores to other networks or datasets through the web interface.
- Due to resources limitations the relative impact of the drugs was calculated relative to others in the same category (mechanism of action).
- Finally, the computation were applied to cell lines independently due to the large size of the datasets.

Thank, *you*

References

Boué, S., Talikka, M., Westra, J. W., Hayes, W., Di Fabio, A., Park, J., Schlage, W. K., Sewer, A., Fields, B., Ansari, S., Martin, F., Veljkovic, E., Kenney, R., Peitsch, M. C., and Hoeng, J. (2015). Causal biological network database: A comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database*, 2015.

Catlett, N. L., Bargnesi, A. J., Ungerer, S., Seagaran, T., Ladd, W., Elliston, K. O., and Pratt, D. (2013). Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. *BMC bioinformatics*, 14(1):340.

Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D. J., Turner, J. P., VidoviÄ ‡, D., Forlin, M., Kelley, T. T., D'Urso, A., Allen, B. K., Torre, D., Jagodnik, K. M., Wang, L., Jenkins, S. L., Mader, C., Niu, W., Fazel, M., Mahi, N., Pilarczyk, M., Clark, N., Shamsaei, B., Meller, J., Vasiliauskas, J., Reichard, J., Medvedovic, M., Ma'Ayan, A., Pillai, A., and Schürer, S. C. (2018). Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: Integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Research*, 46(D1).

Martin, F., Gubian, S., Talikka, M., Hoeng, J., and Peitsch, M. C. (2019). NPA: An R package for computing network perturbation amplitudes using gene expression data and two-layer networks. *BMC Bioinformatics*, 20(1):1–9.

Martin, F., Sewer, A., Talikka, M., Xiang, Y., Hoeng, J., and Peitsch, M. C. (2014). Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models. *BMC Bioinformatics*, 15(1).