Multiple Making Sense out of knockout Experiments

Alon Kaufman[1], Martin Kupiec[2], Eytan Ruppin[3].

[1]Interdisciplinary center of Neural Computation, Hebrew University. [2] Department of Molecular Microbiology and Biotechnology, Tel Aviv University.
[3]Schools of Computer Science and Medicine, Tel Aviv University.

1 Motivation

Addressing the up coming challenge of causal system identification in analysis of genetic networks:

- Who does what? Which are the important/non-important genes in the network?
- Characterize the interactions between the elements.
- Functional inference and the underlying properties on the network.
- What functional predictions/hypothesis can be made?

❖Causal analysis requires perturbations (contrasting correlation studies) → Knockouts.

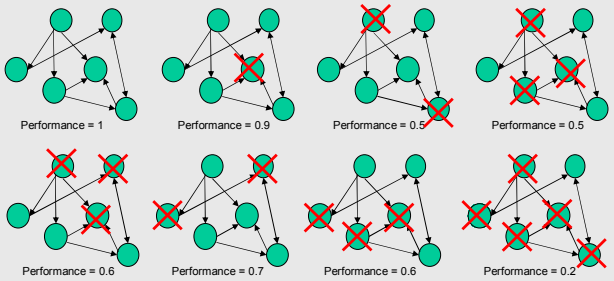
❖Single knockouts are not sufficient for a faithful system identification → Multiple KO.

The following analysis is based on multiple knockout experiments. Showing the results of applying novel analysis methods to the Rad6 pathway network.

- **MSA – Multi-perturbation Shapley Value analysis**
- **FIN – Functional Influence Network**

2 Suggested solution: Multi-knockout Analysis

Input Data: a series of knockout experiments, each with a performance score associated with it regarding the investigated task.



Output:

- Contribution values denoting **the importance of each element** in the network. 4
- Quantification and classification of the **interactions between the elements**. 5
- A reduced performance function, leading to **functional inference**. 6

3 The Rad6 Example

DNA post replication repair (PRR) in the yeast *S. cerevisiae* is dominated by the activity of the Rad6 pathway.

The analyzed data includes a series of 21 multi-knockout experiments.

The performance after each knockout experiment is measured by the relative number of colonies that survive the UV irradiation compared to the wild-type yeast

The genes involved In the experiments are:

- ***ELG1*** ***CTF18*** ***RAD24*** replication factor C like complexes (RLCs)
- ***REV3*** encoding for a specific DNA polymerase ζ
- ***RAD18*** a regulatory gene in the Rad6 pathway

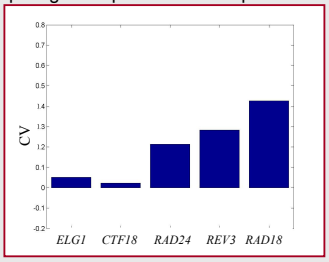
The function and role of the three RLCs is unknown! One possibility is that the RLCs may act similarly to RFC, loading PCNA-related complexes that act as clamps for specific DNA polymerases.

4 Results

The MSA based on the Shapley value theory (Game Theory), enables a correct and fair assignment of contribution values to elements participating in a specific function performed by the investigated system.

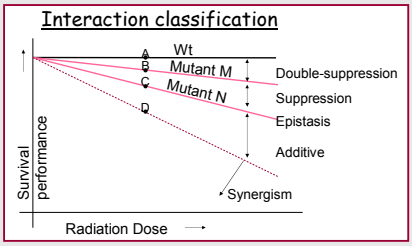
One-dimensional MSA Analysis results:

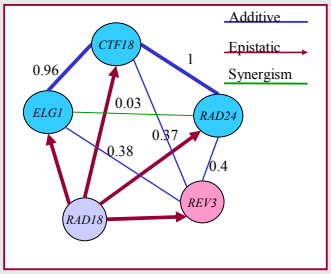
- ✓ All genes play a causal role in PRR
- ✓ RAD18 and REV3; the most important genes:
- ✓The three RLCs play a causal role, although their importance markedly differs.



5 Results-Interactions

In complex systems as genetic networks beside the one-dimension contribution one may be interested in the interactions among the genes. The interaction describe how each gene is depended on the state of the other genes.





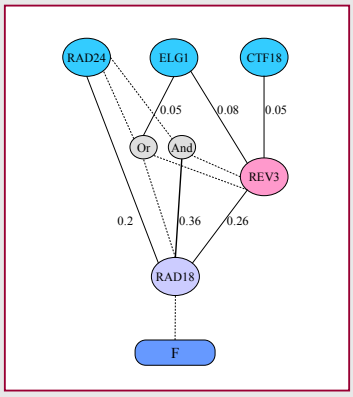
- ✓All three RLC have high additive interactions
- ✓All genes have an epsistatic interaction with RAD18. (RAD18 is crucial for their contributions and an essential gene in the PRR process)
- ✓All three RLC have a moderate additive interaction with REV3

6 Results-Inference

The performance function F(UV sensitivity): linear summation of reduced Boolean terms:

$$F = e \cdot [d \cdot (0.26 + 0.08 \cdot a + 0.05 \cdot b) + c \cdot 0.2 + (c \cap d) \cdot (0.36) + (c \cup d) \cdot (0.05 \cdot a)]$$

$$Where \quad a = ELG\ 1, b = CTF\ 18, c = RAD\ 24, d = REV\ 3, e = RAD\ 18$$



FIN: Functional Influence Network, the dashed lines are connectivity edges the solid lines are the influence edges with a weight assigned to each edge representing the functional influence.

FIN based inference:

There are three main pathways, *RAD18* is an essential gene in all of them:

- ✓ ***REV3-RAD18* pathway**: Suggests that there may be some **additional DNA polymerase loaders** beside those investigated.
- ✓ ***RAD24 –RAD18* pathway**: *RAD24* plays a positive role even without *REV3*. Probably another gene encoding for **another DNA polymerase** involved in PRR. We predict that these additional polymerases will be independent of CTF18 (additive interaction), very dependent on RAD24 and with a moderate dependency on ELG1.
- ✓ ***RAD24 –REV3 –RAD18* pathway**: The pathway includes RAD24 and ELG1 as RLCs that possibly load the DNA polymerase.

7 Conclusions and The Future

- Systematic functional screens are fundamentally changing how biologists identify the molecular components that derive biological processes.
- Multi-perturbation studies are a necessity, and are hence bound to take place, starting in the near future (e.g., RNAi), analysis methods are currently not available.
- The MSA and the FIN form the first systematic approach addressing the challenge of analyzing such experiments in a rigorous manner.
- The methods described are general and not "system specific" suitable for analysis of any type of perturbation studies!!!

If you have multi-knockout/perturbation data and are interested in trying out our methods to analyze your data, please come talk to us!!!
For more details: www.cns.tau.ac.il/~alonk or contact kalon@post.tau.ac.il.