Perturbation-based gene regulatory network inference to unravel oncogenic mechanisms

Daniel Morgan[1], Matt Studham[1], Andreas Tjärnberg[2], Bo Lundgren, Fredrik Swartling, Torbjörn Nordling[3], Erik Sonnhammer[1]

[1]*DBB, Stockholm University & SciLifeLab* [2]*Department of Bioinformatics, Linköping University* [3]*Mechanical Engineering, National Cheng Kung University* [4]

Daniel.Morgan@SciLifeLab.se

Motivation: Cancer is known to stem from multiple, independent mutations, the effects of which aggregate to drive the cell into a cancerous state. To understand the complex interplay between affected genes, their gene regulatory network (GRN) needs to be uncovered, revealing detailed insights of regulatorymechanisms. We therefore decided to infer a reliable GRN from perturbation responses of 40 genes known or suspected to have a role in human cancers yet whose regulatory interactions are poorly known.

Results: siRNA knock-down experiments of each gene were done in a human squamous carcinoma cell line, after which the transcriptomic response was measured. From these data GRNs were inferred using several methods, and the false discovery rate was controlled by the NestBoot framework. The best GRN's topology was validated by measuring its ability to predict an independent dataset of the same genes but subjected to double perturbations. It agrees with many known links in addition to predicting a large number of novel interactions, a subset of which were experimentally validated. The inferred GRN captures regulatory interactions central to cancer-relevant processes and thus provides mechanistic insights that are useful for future cancer research.

Background

Cancer is, in part, a progressive, systemic flux of cellular functions driven by the interaction of multiple gene products [7][1] from a more general state of non-cancer. Cancer subtype-specific gene regulatory networks (GRN) encode intracellular dynamics [8], thus understanding them can offer insight into the functional changes driving disease development. Generally, such inference methodologies are designed to exploit certain aspects of the experimental setup, such as pooling among replicates to amplify signal, or make use of prior knowledge [4] [6]. However methods often fall short of guarding against limitations of the experimentation, such as poor estimations of biological variation which can lead to overfitting and potentially contributes to the inconsistencies seen among benchmarks [3]. Methods using systematic perturbation have shown greater accuracy among inference techniques since more information is available to determine regulatory causal mechanism in the system [5]. Assuming a linear dynamical system (LDS) [2] model, once the system has reached a steady-state equilibrium the network can be inferred by , solving a set of linear ordinary differential equations (ODEs).

Methods

Linear Model: $\boldsymbol{Y} = -\boldsymbol{A}^{-1}(\boldsymbol{P} + \boldsymbol{F}) + \boldsymbol{E}$

Inference Method 1: LSCO $\hat{A}_{LS} = X^T X^{-1} X^T y$

Inference Method 2: LASSO $\min_{w \in \mathbb{R}p} \frac{1}{2m} \left\| \boldsymbol{A}_i \boldsymbol{Y}^T + \boldsymbol{P}_i^T \right\|_F^2 + \zeta \|\boldsymbol{A}\|_1$

Inference Method 3: TLSCO $[XY] = USV,$ $\hat{A}_{TLS} = -\frac{VXY}{VYY}$

Algorithm 1 : Leave Out Error Estimation & Balancing

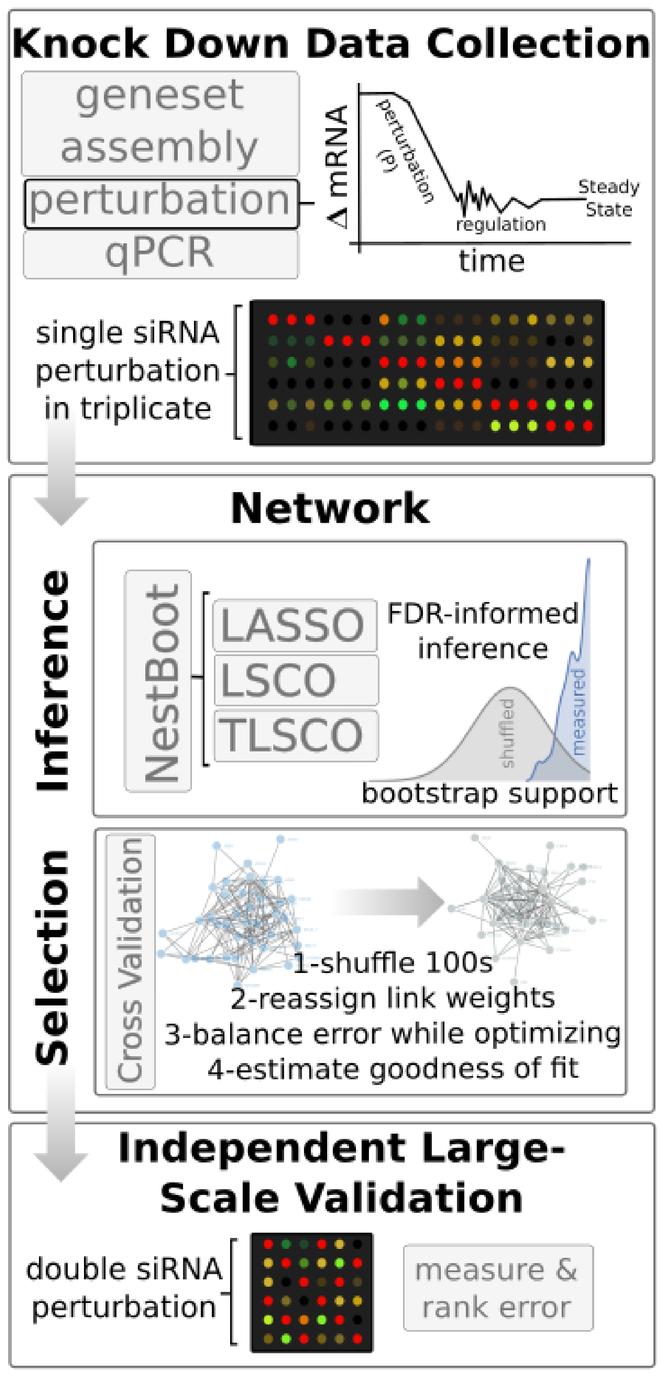
1. *for experiments in dataset Y: remove experimental replicates*
   a. Bisectionally vary w until $Y_{relErr} = P_{relErr}$
      i. define E,F
      ii. solve: minimize $\|E\| + \|F\|$
   b. Return noise free P ($PNF_{LOO} = P - F$)
   c. Estimate F_{LOO} (Eq.4b)
2. ΣwRSS

where

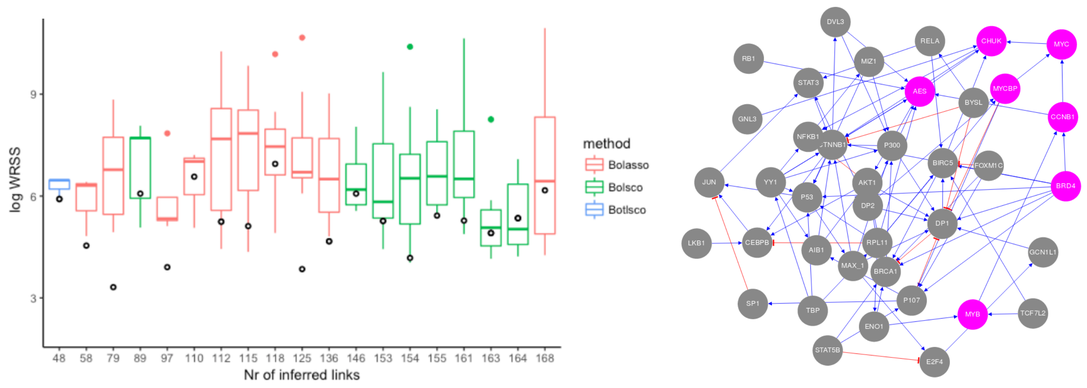
$$E = \mathcal{X}^{-2}(B, Y_m)$$
$$Y_{relErr} = \frac{\|E\|^2}{\|Y_{obs}\|^2}$$

and

$$F = w \|A, Y_m\|_F$$
$$P_{relErr} = \frac{\|F\|^2}{\|P_{obs}\|^2}$$



Results



Conclusions

- A common set of genes was perturbed and measured independently in human squamous carcinoma cell line.
- The training dataset contains genes perturbed and measured three times as experimental replicates, while the validation dataset contains the same genes perturbed in pairs without replicate.
- Taking into account various data properties, the training dataset was used to infer a network of the underlying mechanisms of control.
- This network was able to reproduce its training data in a leave out manner, and whatsmore, it is robust enough to reproduce a separate validation dataset to a degree of accuracy higher than expected by chance.
- In this way, many known links were recovered during the inference, as well as novel links proposed, two of which were verified experimentally.

References

[1] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1):56, 2011.

[2] Timothy S Gardner, Diego Di Bernardo, David Lorenz, and James J Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629):102–105, 2003.

[3] Shun Guo, Qingshan Jiang, Lifei Chen, and Donghui Guo. Gene regulatory network inference using pls-based methods. *BMC bioinformatics*, 17(1):545, 2016.

[4] Anne-Claire Haury, Fantine Mordelet, Paola Vera-Licona, and Jean-Philippe Vert. Tigress: trustful inference of gene regulation using stability selection. *BMC systems biology*, 6(1):145, 2012.

[5] Jessica Minnier, Lu Tian, and Tianxi Cai. A perturbation method for inference on regularized regression estimates. *Journal of the American Statistical Association*, 106(496):1371–1382, 2011.

[6] Fantine Mordelet and Jean-Philippe Vert. Sirene: supervised inference of regulatory networks. *Bioinformatics*, 24(16):i76–i82, 2008.

[7] Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6):1437–1452, 2017.

[8] Adi L Tarca, Roberto Romero, and Sorin Draghici. Analysis of microarray experiments of gene expression profiling. *American journal of obstetrics and gynecology*, 195(2):373–388, 2006.