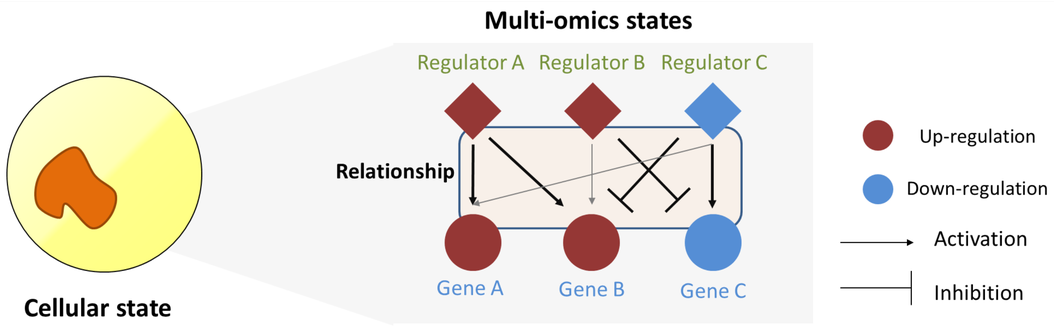
Combinatorial Modeling and Optimization using Iterative RL Search for Inferring Sample-Specific Regulatory Network

Minsik Oh[1], Sangseon Lee[2], Dohoon Lee[3], Sangsoo Lim[3] and Sun Kim[4,5,6]

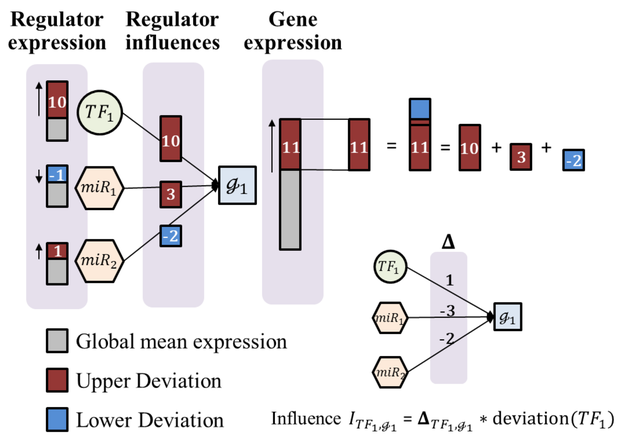
[1]BK21 FOUR Intelligence Computing, Seoul National University, [2]Institute of computer technology, Seoul National University, [3]Bioinformatics Institute, Seoul National University, [4]Department of Computer Science and Engineering Seoul National University, [5]Interdisciplinary Program in Bioinformatics, Seoul National University, [6]Institute of Engineering Research Seoul National University

Introduction & Motivation



Understanding the cell state from the multi-omics data requires identifying the global gene regulation network described by *n-to-m* relationships between regulators and genes.

In order to determine the *n-to-m* relationships, we used transcription factors (TFs) and microRNA as regulators that have been well studied its regulatory mechanisms at the transcriptional level and post-transcriptional level.



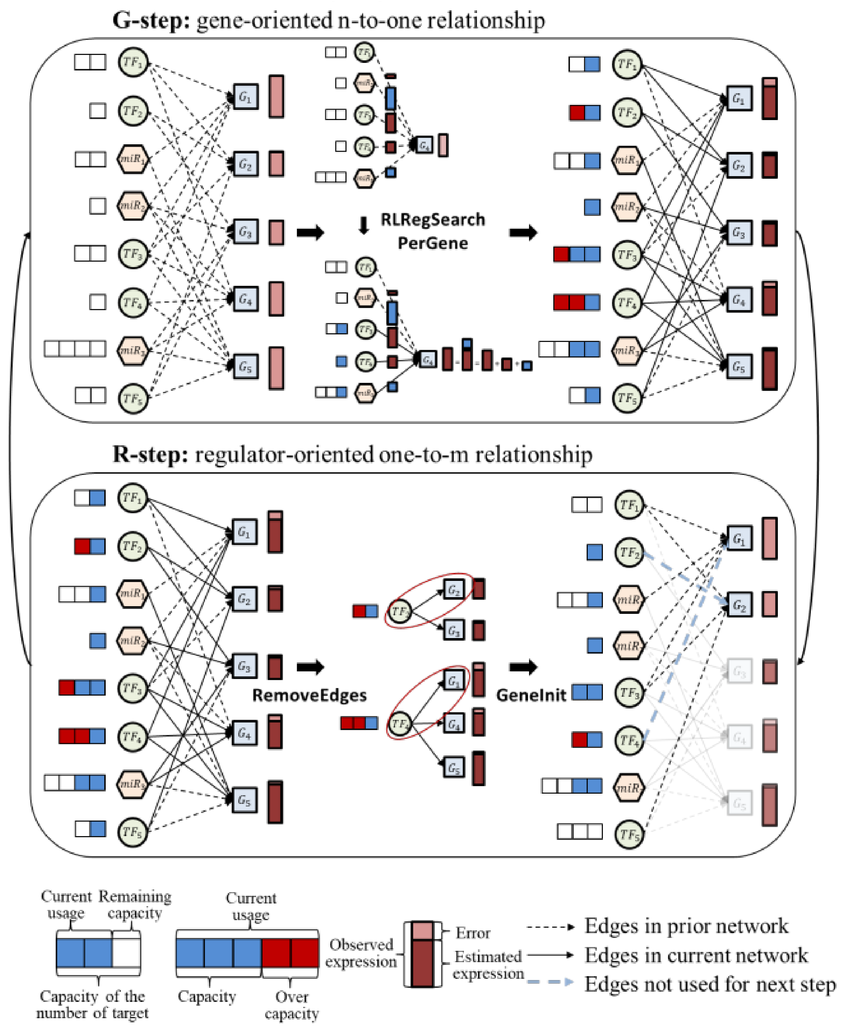
To construct the regulator-gene network, we assumed that the sum of regulator influences determines the deviation of gene expression and each regulator has the capacity of the number of target genes.

The above figure illustrates the example of regulation that G_1 is targeted by TF_1, miR_1, miR_2 and the sum of influences determines the upper deviation of G_1 expression.

From the assumptions, we formulated an objective function for estimating gene expression of cell line from the network. In order to minimize the objective function that is a combinatorial optimization problem, we proposed an iterative search algorithm to find a sample-specific regulatory network.

Method

The overview of proposed search method



Our method consists of two steps. The first step is the gene-oriented step (G-step) to explore n-to-one relationships between regulators and each target gene. In G-step, edges are added to the network using reinforcement learning-based heuristics. The second step is the regulator-oriented step (R-step) to explore one-to-m relationships between each regulator and target genes. In R-step, edges are removed from the network using stochastic selection.

Results

Quantitative evaluation

The quantitative evaluation results for different problem size

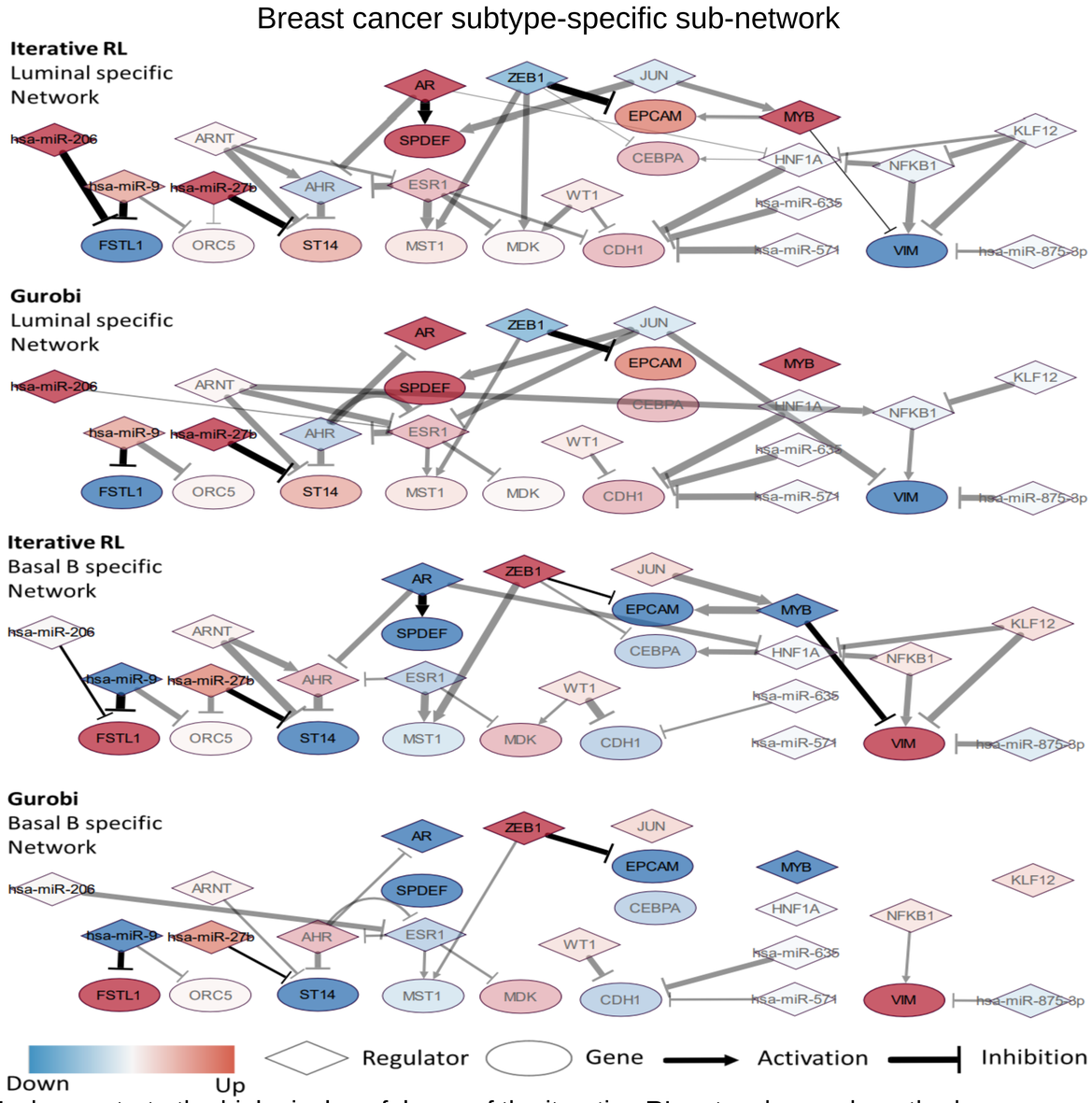
Number of genes	Number of edges	Iterative RL	Gurobi	GA
100	917	0.187 (0.5h)	**0.175 (2h)**	0.656 (1h)
300	2755	0.267 (0.5h)	**0.258 (2h)**	0.748 (2h)
500	4587	0.225 (0.5h)	**0.217 (2h)**	0.717 (3.5h)
1000	9871	0.198 (0.5h)	**0.190 (3h)**	0.723 (8h)
2000	18987	**0.193 (1h)**	0.205 (6h)	0.701 (16h)
3000	28929	**0.193 (1h)**	0.554 (12h)	0.724 (24h)
4000	39307	0.184 (2h)	**0.181 (24h)**	-
5000	49925	**0.180 (2h)**	0.190 (24h)	-
6000	60060	**0.183 (2h)**	0.679 (24h)	-

For the quantitative evaluation of our method, we compared the performance of the iterative RL method with Gurobi and Genetic Algorithm (GA), which are widely used to solve combinatorial optimization problems.

As the complexity of the problem increases, the iterative RL method provided a more accurate gene expression estimation than Gurobi and GA. The iterative RL method provided a promising solution with a reasonable running time. The quantitative evaluation suggests that our method is more suitable for exploring search spaces of gene regulatory networks than previous Gurobi and GA.

Qualitative evaluation

Breast cancer subtype-specific sub-network



To demonstrate the biological usefulness of the iterative RL network search method, we constructed breast cancer subtype-specific (Luminal, Basal B) networks using cancer hallmark genes of CCLE breast cancer cell line data.

Several regulatory relationships in estimated networks have been experimentally verified in previous studies. For instance, Androgen receptor (AR) directly upregulates expression of SAM pointed domain containing Ets transcription factor (SPDEF) that promotes the proliferation and invasion of breast cancer cell lines [1]. Suppression of Tumorigenicity 14 (ST14) is a target gene of miR-27b, which increases cancer progression by decreasing ST14 expression [2]. Vimentin (VIM) is known to be over-expressed in basal breast cancer cell lines. There is an experimental result that the expression of VIM increased when MYB was knocked down in breast cancer cells [3].

Our iterative RL method more detects experimentally validated edges than Gurobi.

Reference

[1] Cao, Lu, et al. "AR–PDEF pathway promotes tumour proliferation and upregulates MYC-mediated gene transcription by promoting MAD1 degradation in ER-negative breast cancer." Molecular cancer 17.1 (2018): 1-15.
[2] Wang, Yanfang, et al. "ST14 (suppression of tumorigenicity 14) gene is a target for miR-27b, and the inhibitory effect of ST14 on cell growth is independent of miR-27b regulation." *Journal of Biological Chemistry* 284.34 (2009): 23094-23106.
[3] Hugo, Honor J., et al. "Direct repression of MYB by ZEB1 suppresses proliferation and epithelial gene expression during epithelial-to-mesenchymal transition of breast cancer cells." Breast cancer research 15.6 (2013): 1-19.