**BMC
Bioinformatics**

POSTER PRESENTATION

Open Access

Evaluating four major algorithms for identifying differential regulators in condition-specific transcriptional responses

Hui Yu[1], Zhongming Zhao[1,2,3*]

Background

Identifying molecular regulators underlying condition-specific transcriptional responses is essential for our understanding of their underlying molecular mechanisms. So far, there have been several computational methods developed for this purpose. Specifically, four major algorithms, TFact [1], RIF [2], CSA [3], and DRrank [4], were released one after another. Each of these algorithms has its own features and evaluation strategies. Thus, these methods should be systematically evaluated so that the users can make the most appropriate method selection for their practical application needs.

Materials and methods

In this work, we evaluated the four algorithms using *Escherichia coli* transcription network models and synthetic expression datasets that were generated by SynT-ReN [5] and GeneNetWeaver [6]. Specifically, we developed a simulation-based schema to evaluate each algorithm according to operatively defined, known "differential regulators." In addition, we tested each method's robustness against its key parameter(s) and explored factors that influence algorithm performance in general.

Results

We found that TFact and DRrank stood out as the best methods in terms of both accuracy and robustness. In total, there were seven cases in which one single regulator was artificially perturbed. TFact attained the closest approximation four times, while DRrank had the best

performance the other three times. In another 15 scenarios in which there were perturbations of multiple regulators, DRrank was ranked on average the best algorithm, while TFact had the second-place rank. Based on these observations, TFact and DRrank may each be best applied to different circumstances: TFact is better used for single regulator implication, while DRrank is better for multiple regulators' simultaneous perturbation. In general, we observed that algorithms' performances were negatively correlated with the number of regulation links per target, which may indicate that independent regulators are easier to recover than synergistic ones.

Conclusions

This work represents a preliminary benchmarking evaluation of four major currently-available algorithms for differential regulator identification. According to our simulation-based evaluation paradigm, two of the published algorithms, TFact and DRrank, are more robust than the other two. Further evaluation is needed, such as applying them to specific human disease expression datasets under different conditions or different platforms. Furthermore, these algorithms may be extended to rapidly emerging next-generation sequencing (NGS) data.

Authors' details
[1]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. [2]Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. [3]Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA.

* Correspondence: zhongming.zhao@vanderbilt.edu
[1]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
Full list of author information is available at the end of the article

BioMed Central

References

1. Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin JB: **Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data.** *Nucleic Acids Res* 2010, **38**(11):e120.
2. Reverter A, Hudson NJ, Nagaraj SH, Perez-Enciso M, Dalrymple BP: **Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data.** *Bioinformatics* 2010, **26**(7):896-904.
3. Huang CL, Lamb J, Chindelevitch L, Kostrowicki J, Guinney J, Delisi C, Ziemek D: **Correlation set analysis: detecting active regulators in disease populations using prior causal knowledge.** *BMC Bioinformatics* 2012, **13**:46.
4. Yang J, Yu H, Liu BH, Zhao Z, Liu L, Ma LX, Li YX, Li YY: **DCGL v2.0: An R package for unveiling differential regulation from differential co-expression.** *PLoS One* 2013, **8**(11):e79729.
5. Van den Bulcke T, Van Leemput K, Naudts B, van Remortel P, Ma H, Vershoren A, De Moor B, Marchal K: **SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms.** *BMC Bioinformatics* 2006, **7**:43.
6. Schaffter T, Marbach D, Floreano D: **GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods.** *Bioinformatics* 2011, **27**(16):2263-2270.