# Combination of data driven and literature based gene regulatory networks for TF activity estimation

Intern: Anis Mansouri
Supervised by: Aurelien Dugourd

June 16, 2021

## Summary

The objective of this work is to try to improve the Transcription Factors Activity Estimation (TFAE). TFAE functions (e.g. viper()) requires two inputs: an expression set and a gene regulatory network (regulons). Previously, we were using DoRothEA regulons [1] containing TFs regulating their gene targets according to a given mode of regulation (MOR): 1 for upregulation and -1 for downregulation. Here, we have substituted these discrete MOR by continuous ones, which were inferred from the expression set using WGCNA and ARACNE. Seemingly, using these hybrid regulons (DoRothEA regulons with an inferred MOR) increases the performance of TFAE.

## Methods

### Inference of the gene co-expression networks

First, WGCNA and ARACNE were used to infer gene co-expression networks (GCN) from the DoRothEA Benchmark Expression Set (DBES). DBES contains the expression levels of 22066 genes from bulk, under 124 conditions (samples), each sample corresponds to a perturbation (overexpression or knockdown) of a given TF (see DoRothEA Benchmark Meta data). Secondly, spurious gene-gene interactions were omitted by Soft Thresholding and/or Topological Overlap Measure:

### Soft Thresholding:

This approach amplifies the disparity between strong and weak gene-gene interaction weights by raising the weights to a power. Doing so, the weakest weights become negligible compared to the strong ones [2].

**Topological Overlap Measure:**

This approach reinforces the interactions between gene pairs having a high topological overlap (many shared neighbours), and eliminates interactions between gene pairs sharing few neighbours [3].

## Construction of the hybrid regulons

We matched the processed GCN with DoRothEA regulons and extracted only the common TF-Target pairs and their inferred MOR (continuous weight).

## Benchmark of the TFAE

We run the DecoupleR Benchmark Workflow in order to assess the performance of TFAE using hybrid regulons.

# Results

Figure 1 and 2 show the TFAE performances (Area Under the ROC Curve) after using the original DoRothEA regulons (discrete MOR) and the hybrid ones (hybrid_dorothea: continuous MOR). Seemingly, WGCNA's hybrid regulons outperform ARACNE's ones and result in higher performances than the original DoRothEA's regulons. In addition, using TF-Target pairs having the highest confidence level (dorothea_A and hybrid_dorothea_A) seems to improve the estimation.

DecoupleR Benchmark Workflow runs five TFAE functions: run_mean(), pscira(), scira(), viper() and gsva(). These latter result in more or less different performances, and only the lowest (min) and the highest (max) AUROC are showed in the previous figures (the remaining AUROC are available via this link)

## Performances (Area Under the ROC curve) of TF-Activity estimation using DoRothEA/hybrid_DoRothEA (WGCNA)

| Interactome | AUROC (min/max) |
| --- | --- |
| dorothea_A | 0.7/0.76 |
| dorothea_ABC | 0.65/0.71 |
| dorothea_ABCDE | 0.61/0.69 |

| Interactome | AUROC (min/max) |
| --- | --- |
| hybrid_dorothea_A | 0.67/0.73 |
| hybrid_dorothea_ABC | 0.64/0.69 |
| hybrid_dorothea_ABCDE | 0.61/0.64 |

| Interactome | AUROC (min/max) |
| --- | --- |
| hybrid_dorothea_A | 0.76/0.81 |
| hybrid_dorothea_ABC | 0.71/0.76 |
| hybrid_dorothea_ABCDE | 0.67/0.71 |

Figure 1: TF-Activity estimation performance using DoRothEA regulons (top table) and hybrid regulons inferred by WGCNA: spurious interactions eliminated by a soft thresholding method only (left table) or soft thresholding followed by the topological overlap method (right table), respectively. _A, _ABC, _ABCDE: Confidence levels of the TF-Target interactions. min/max: lowest and highest AUROC among the five TFAE functions

## Performances (Area Under the ROC curve) of TF-Activity estimation using DoRothEA/hybrid_DoRothEA (ARACNE)

| Interactome | AUROC (min/max) |
| --- | --- |
| dorothea_A | 0.7/0.76 |
| dorothea_ABC | 0.65/0.71 |
| dorothea_ABCDE | 0.61/0.69 |

| Interactome | AUROC (min/max) |
| --- | --- |
| hybrid_dorothea_A | 0.66/0.71 |
| hybrid_dorothea_ABC | 0.64/0.69 |
| hybrid_dorothea_ABCDE | 0.61/0.66 |

| Interactome | AUROC (min/max) |
| --- | --- |
| hybrid_dorothea_A | 0.67/0.71 |
| hybrid_dorothea_ABC | 0.63/0.70 |
| hybrid_dorothea_ABCDE | 0.60/0.67 |

Figure 2: TF-Activity estimation performance using DoRothEA regulons (top table) and hybrid regulons inferred by ARACNE: spurious interactions eliminated by a soft thresholding method only (left table) or soft thresholding followed by the topological overlap method (right table), respectively. _A, _ABC, _ABCDE: Confidence levels of the TF-Target interactions. min/max: lowest and highest AUROC among the five TFAE functions

# Scripts and hybrid regulons

The hybrid regulons and R scripts are available on: [https://github.com/saezlab/Hybrid_Regulon_Based_TF_Act_Estimation](https://github.com/saezlab/Hybrid_Regulon_Based_TF_Act_Estimation). Further technical details are included as comments in the scripts.

# References

[1]  Christian H. Holland et al. "Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data". In: *Genome Biology* (2020). DOI: 10.1186/s13059-020-1949-z.

[2]  Peter Langfelder and Steve Horvath. "WGCNA: an R package for weighted correlation network analysis". In: *BMC bioinformatics* 9.1 (2008), pp. 1–13.

[3]  Lin Song, Peter Langfelder, and Steve Horvath. "Comparison of co-expression measures: mutual information, correlation, and model based indices". In: *BMC bioinformatics* 13.1 (2012), pp. 1–21.