

Study of Meta-Analysis Strategies for Network Inference using Information-Theoretic Approaches

Ngoc Cam Pham*, Benjamin Haibe-Kains†, Pau Bellot‡, Gianluca Bontempi§ and Patrick E. Meyer*

*Bioinformatics and Systems Biology (BioSys) Unit, Université de Liège, Belgium

†Princess Margaret Cancer Centre, University Health Network, Toronto, Canada

‡Image Processing group, Technical University of Catalonia, Spain

§Machine Learning Group, Interuniversity Institute of Bioinformatics in Brussels (IB)², Université Libre de Bruxelles, Belgium

Abstract—Reverse engineering of gene regulatory networks (GRNs) from gene expression data is a classical challenge in systems biology. Thanks to high-throughput technologies, a massive amount of gene-expression data has been accumulated in the public repositories. Modelling GRNs from multiple experiments (also called integrative analysis) has; therefore, naturally become a standard procedure in modern computational biology. Indeed, such analysis is usually more robust than the traditional approaches focused on individual datasets, which typically suffer from some experimental bias and a small number of samples.

To date, there are mainly two strategies for the problem of interest: the first one (“data merging”) merges all datasets together and then infers a GRN whereas the other (“networks ensemble”) infers GRNs from every dataset separately and then aggregates them using some ensemble rules (such as ranksum or weightsum). Unfortunately, a thorough comparison of these two approaches is lacking.

In this paper, we evaluate the performances of various meta-analysis approaches mentioned above with a systematic set of experiments based on *in silico* benchmarks. Furthermore, we present a new meta-analysis approach for inferring GRNs from multiple studies. Our proposed approach, adapted to methods based on pairwise measures such as correlation or mutual information, consists of two steps: aggregating matrices of the pairwise measures from every dataset followed by extracting the network from the meta-matrix.

Keywords—Meta-analysis, gene regulatory networks, systems biology, gene expression, mutual information.

I. INTRODUCTION

One of the most long-standing challenges in Systems Biology is the development of methods, which are able to construct the complete set of regulatory interactions of a cell. The regulating circuitry, also called gene regulatory network (GRN), can then be used by bio-medical experts to understand key mechanisms in cells. Thanks to high-throughput technologies, a large amount of transcriptome data is now available through public repositories (e.g. NCBI GEO [1], ArrayExpress [2]), providing opportunities to study the GRNs of many organisms. In the last decade, a variety of methods have been proposed in an attempt to address this *reverse engineering* problem. The methods can

be classified into several categories [3], such as: regression-based, pairwise similarity (mutual information, correlation,...), Bayesian networks or even ensemble approaches (that combine several different approaches). Among those, mutual information (MI) based methods, such as CLR [4], ARACNE [5], MRNET [6], [32] and so on, gather more and more attention due to their capability to deal with up to several thousands of variables in the presence of a limited number of samples [6]. Generally, these methods first estimate a pairwise mutual information (i.e. a non-linear dependency measure) between all pairs of genes, resulting in a mutual information matrix (MIM). Afterwards, indirect interactions are eliminated from the MIM by the different methods and thus a GRN is inferred.

Since a single dataset has typically a small sample size (usually less than 200 observations) and suffers from potential experimental biases, classical analytical tools show their limits in unravelling reliably underlying interactions. By contrast, integrative analysis of multiple studies is able to increase significantly the statistical power and thus is becoming a standard procedure in modern computational biology [7]. However, the question of integrating all the data consistently and efficiently raises new questions [8].

Meta-analysis strategies are increasingly used for detecting differentially expressed genes from microarray data [17]. In those *meta-analysis* approaches, each single dataset is analysed separately and then the final results are combined [18]. Several strategies have been proposed in order to perform meta-analysis on expression data. For instance, a meta-analysis of public gene expression data and clinical data was conducted by using the concept of “coexpression” modules to reveal various results of previous gene expression studies in breast cancer [19], [20]. In another research [21], Hong et al. developed a Bioconductor package RankProd that allow researchers to do meta-analysis under two experimental microarray conditions to identify differentially expressed genes. Moreover, in [22] a new method for knowledge extraction from RNA-seq cancer data utilizing many classification models named CAMUR showed its efficacy over state of the art algorithms based on a single classification

model. If the problem of detecting differentially expressed genes across several studies has been intensively studied, it is, however, not yet the case when it comes to constructing GRNs.

To deal with this challenge of meta-network inference, there have been plenty of proposed methods, which can be divided into two main categories: *data merging* and *network ensemble*. In the *data merging* approach, datasets are integrated at the expression level into a unique dataset, from which GRNs are inferred [9]–[11]. However, one of the major problem of this approach is the removal of batch effects. Indeed, the use of different platforms, and different methodologies by different research groups introduce statistical biases (batch effects) that can lead to incorrect conclusions [12]. For example, it is known that normalization techniques, such as RMA [13], consisting in re-scaling gene expression values at the probe intensity level for Affymetrix data [14], is not able to remove batch effects. Consequently, batch removal methods, like COMBAT [15], is typically used before merging data [12]. The approach consists in merging GRNs from different datasets, for example by weighting gene-gene interactions according to their average rank in each network [3]. This approach rooted in the "wisdoms of crowds" concept, which was first introduced in the DREAM5 challenge and then further developed by [16] with the Top*k*Net algorithms to produce consensus networks.

In this paper, we also introduce a new meta-analysis strategy to build a consensus network. The new strategy consists in aggregating matrices of pairwise mutual information with each being estimated from a gene expression dataset to produce a meta-matrix, from which a GRN is inferred using classical information-theoretic network inference algorithms. Additionally, the paper presents the first thorough experimental comparison of these three "meta" approaches for the reconstruction of networks, namely data merging, network ensemble and coexpression matrices aggregation. The performances of these three sets of methods are evaluated using synthetic datasets from the standard Bioconductor netbenchmark package.

II. METHODS

A. State-of-the-art

Mutual information is a non-linear measure of dependency between two variables (genes) X and Y , defined as follow

$$I(X, Y) = \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

where $p(x, y)$ is the joint probability distribution of X and Y , and $p(x)$ and $p(y)$ are the marginal probability distributions of X and Y , respectively.

This dependency measure has been used for reconstructing networks by several methods such as CLR, ARACNE or

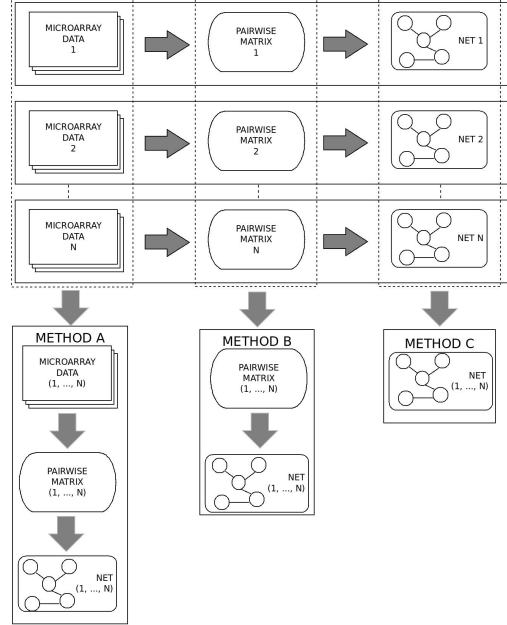


Figure 1: Meta-network strategies: assembling datasets, pairwise matrices or networks

MRNET. The Context Likelihood or Related network (CLR) method [4] creates an edge between each pair of genes i and j if the combined z-score of the mutual information between them is above a given threshold, where the combined z-score is defined as:

$$c_{ij} = \sqrt{c_i^2 + c_j^2} \text{ with } c_i = \max(0, \frac{M_{ij} - \mu_{M_i}}{\sigma_{M_i}}) \quad (2)$$

in which, μ_{M_i} and σ_{M_i} are the mean and standard deviation of the empirical distribution of the mutual information of gene i .

The Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) [5], relies on the "Data Processing Inequality" (DPI) which removes the edge with the weakest mutual information, in every triplet of genes.

Finally, the Minimum Redundancy NETWORKS (MRNET) [6], [32] method reconstructs a network using the feature selection technique known as Minimum Redundancy Maximum Relevance (MRMR) [23]. The minimum redundancy criterion makes the implicit assumption that variables with redundant information to the most relevant variables are indirect links.

Using those three information-theoretic network inference techniques, available in the Bioconductor Minet package, we will compare three meta-analysis approaches detailed in the next sections.

B. Data merging - A methods

A straightforward approach for performing integrative analysis of multiple studies is combining all datasets together and then analysing the merged dataset. These method,

named *data merging* and denoted here with the letter (A), were widely used in [9]–[11] to reconstruct large-scale GRNs because of their simplicity. However, since high dimensional data often suffers from unwanted biases, a variety of techniques can be used to correct for these non-biological variations. We present in the following two classical scaling methods typically used to assemble datasets, and one batch-effect-removal method.

1) *Normalization: BMC(A1) and z-score (A2)*: Let X be a matrix $X^{m \times n}$ denoting the dataset of gene expression values. In this matrix, columns represent samples and rows represent genes, and x_{ij} represents the expression value of gene i in sample j of dataset X . In [24], a normalization technique named BMC (equation 3) was applied for merging breast cancer datasets.

$$\hat{x}_{ij} = x_{ij} - \bar{x}_i \quad (3)$$

Similarly, the z-score normalization [25] described by equation 4 can be also performed.

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_{x_i}} \quad (4)$$

2) *Batch effects removal: COMBAT(A3)*: Expression datasets mostly come from different platforms and laboratories, causing the so-called 'batch effects'. Consequently, batch removal methods, like COMBAT (also known as Empirical Bayes) [15], is often used to detect and remove this inevitable variation. COMBAT, which was shown to outperform other commonly used batch removal methods in some specific scenarios [26], uses estimations for the LS (location-scale) parameters (e.x. mean and variance) for each gene independently [27]. The gene, afterwards, is adjusted to meet the estimated model.

C. Networks ensemble - C methods

The second category of approaches for integrative analysis (networks ensemble - denoted with the letter (C) in this paper) first constructs every single transcriptional networks independently before combining them to produce a so-called community network [3]. In general, combining networks consists in two distinct steps: transformation and aggregation [28]. Indeed, before assembling networks, a network-normalization step is also performed because it is common to observe networks that exhibit different distribution of edge weights.

Let e_{ij} be the weight of an edge between gene i and gene j and $t_n(e_{ij})$ be the normalized value for e_{ij} in the network n . In the next subsections, we discuss three different combinations of network transformation and aggregation.

1) *No normalization and aggregation with the sum rule(C1)*: In this approach, all networks are simply combined together using the sum rule

$$a_{NO}(e_{ij}) = \sum_{n=1}^N e_{ij} \quad (5)$$

with $a_{NO}(e_{ij})$ denoting the aggregated score in the ensemble network.

2) *Normalization with z-score and aggregation with the sum rule(C2)*: Z-score transformation, the well-know normalization strategy for merging data mentioned in the previous section, can also be applied for ensemble networks. It is worth emphasizing that datasets are normalized by variables/columns whereas networks are edge lists normalized "globally", i.e. as a single vector of weights.

$$t_n(e_{ij}) = s_n(e_{ij}) \quad (6)$$

Here, $s_n(e_{ij})$ denotes the z-score transformation of the edge e_{ij} . After normalization, all the network can be combined using the sum rule.

$$a_Z(e_{ij}) = \sum_{n=1}^N t_n(e_{ij}) \quad (7)$$

3) *Median method (C3)*: In [16] the median value was introduced for aggregating consensus networks. This method assigns the median value among N values representing the confidence score of a specific edge in N different networks.

$$a_M(e_{ij}) = \text{median}\{t_1(e_{ij}), \dots, t_N(e_{ij})\} \quad (8)$$

D. Coexpression matrices aggregation approaches - B methods

Our new category of meta-analysis approaches (denoted with the letter (B) in this paper) aggregates mutual information matrices rather than data or networks. The idea behind assembling pairwise matrices is that, although expression data typically shows high variability due to differences in technology, samples, labels, etc., pairwise dependency measures between genes should be much less variant (i.e. dependent variables, such as a regulating variable and its regulated counterpart, should remain dependent in every platform/experiment/dataset even if ranges of values differ greatly). Thus, to infer a network from various expression data, our approach consists in combining mutual information matrices (MIMs) estimated independently from each dataset. Then a GRN network is inferred from the aggregated MIMs. In the following subsections, we compare three methods to assemble matrices of pairwise measure.

1) *Random-effects model (B1)*: It should be noted that the problem of combining MIMs across multiple datasets can be framed in the context of a meta-analysis of correlation coefficients [29]. Hunter and Schmidt [30] introduced a single random-effects method based on untransformed correlation coefficients, at which datasets are weighted simply by the sample sizes on which each effect size (the estimated MIM) is based. Our first weighting schema (method B1 in this paper), described by equation 9, utilizes this random-

effects method, but using MI instead of correlations.

$$\hat{I}_{RE}(X; Y) = \frac{\sum_{k=1}^N n_k I(X_k; Y_k)}{\sum_{k=1}^N n_k} \quad (9)$$

where $I(X_k; Y_k)$ is the MI between two variable X_k and Y_k in the study k and n_k is the number of samples of study k .

The idea is simply that effect sizes based on large samples will be more precise than those based on small samples.

2) *Median method (B2)*: One of the major issue of B1 is that the quality of datasets used in meta-analysis is not explicitly taken into account. Indeed, inclusion of a noisy datasets is likely to weaken statistical power [17]. Thus, an alternative schema for combining MIMs across heterogeneous studies, namely method B2 can be proposed. Method B2 is explained by formula 10, in which the aggregated MI of a gene pair X and Y is the median value of all MI values between them across all studies.

$$\hat{I}_M(X, Y) = \text{median}(I(X_1, Y_1), I(X_2, Y_2), \dots, I(X_N, Y_N)) \quad (10)$$

3) *Internal quality control index (B3)*: In [17], six quantitative quality control measures have been proposed for the inclusion/exclusion of gene expression studies used for the meta-analysis. Among these measures, the internal quality control index will be used, as method B3 for assembling matrices. Let the similarity between two studies m and n be defined as

$$r_{mn} = \text{spcor}((I_{mij}; 1 \leq i \leq j \leq G), (I_{nij}; 1 \leq i \leq j \leq G)) \quad (11)$$

In which r_{mn} is the Spearman's rank correlation of the pairwise correlation structure between study m and n and G represents the total number of genes in the studies. The dissimilarity (or distance) between study m and n is defined as $d_{mn} = (1 - r_{mn})/2$. For a given study k , a weight $-w_k$ will be granted as the fraction between the sum of distances between study k - D_k^* to all other studies and the sum of pairwise distances between all studies excluding the study k - $D_k^\#$ with

$$D_k^* = \{d_{kn}\}_{1 \leq n \leq N; n \neq k} \text{ and } D_k^\# = \{d_{mn}\}_{1 \leq m \neq n \leq N; m \neq k; n \neq k} \quad (12)$$

Afterwards, the MI between two variables (genes) X and Y is aggregated by the following equation:

$$\hat{I}_{IQC}(X; Y) = \frac{\sum_{k=1}^N w_k I(X_k; Y_k)}{\sum_{k=1}^N w_k} \quad (13)$$

III. EXPERIMENTS

A. Data collection and benchmarking process

There are two tasks one needs to consider in order to validate networks: 1) defining a "gold standard" - which is a

set of true regulations describing the underlying interaction network, 2) selecting quantitative measures to statistically assess the quality of inferred networks. Typically, the first task is addressed by collecting well-known regulations mined from literature with strong supporting evidences. However, those regulations just cover a small part of the underlying network and hence cannot be an ideal reference network to thoroughly compare methods. Hence the latter approach is often completed by *in-silico* experiments.

In this paper, *in silico* benchmarks are selected from every one of the 4 biological networks and artificially generated datasets coming from the Netbenchmark Bioconductor package [31]. The characteristics of the 4 networks are shown below in the table I.

Each large dataset will be split into 6 sub-datasets with a number of experiments ranging between 30 to 300 (a number chosen randomly in order to simulate real case scenario). For example, in Figure 3, an original dataset is split into 6 sub-datasets with the following number of samples: 50, 100, 150, 120, 70 and 190. Additionally, two extremely noisy studies are added, both with a large sample size for each (between 280 and 300). Those datasets allow to test the sensitivity of meta-network methods to datasets that should typically be excluded. Indeed, a few biological studies dating back to the beginning of the microarray technology have very little information and are typically excluded from meta-analysis studies. In order to make the network inference problem

Table I: Networks used in the paper

Network	Name	Topology	Experiments	Genes	Edges
<i>SynTreN</i> ₃₀₀	S1	E. coli	800	300	468
<i>SynTreN</i> ₁₀₀₀	S2	E. coli	1000	1000	4695
<i>GNW</i> ₁₅₆₅	G1	E. coli	1565	1565	7264
<i>GNW</i> ₂₀₀₀	G2	Yeast	2000	2000	10392

more challenging and realistic, noise and transformations of data are added. In particular, we define three levels of data-distortion: *i*) Level 1: An independent lognormal noise call "global" noise, with intensity between 20 and 50%, is added to the first 6 datasets. The standard deviation of this noise (σ_{Global}) is the same for the whole data set and is a percentage ($\kappa_g\%$) of the mean variance of all the genes in the dataset ($\bar{\sigma}_g$). It is defined as follows: $\sigma_{Global; \kappa_g\%} = \bar{\sigma}_g \frac{U(0.8\kappa, 1.2\kappa)}{100}$. *ii*) Level 2: In addition to the global noise, a normally distributed "local" noise with intensity also ranging between 20 and 50%, is added. This is an additive Gaussian noise with zero mean and a standard deviation ($\sigma_{Local(g)}$) that is around a percentage ($\kappa\%$) of the gene standard deviation (σ_g). Therefore, the Signal-to-Noise-Ratio(SNR) of each gene is similar. The local noise standard deviation can be formulated as follows: $\sigma_{Local(g); \kappa\%} = \sigma_g \frac{U(0.8\kappa, 1.2\kappa)}{100}$ where $U(a, b)$ is a uniform distribution between a and b . *iii*) Level 3: In addition to the two previous noises, each sub-dataset can be transformed using a randomly chosen

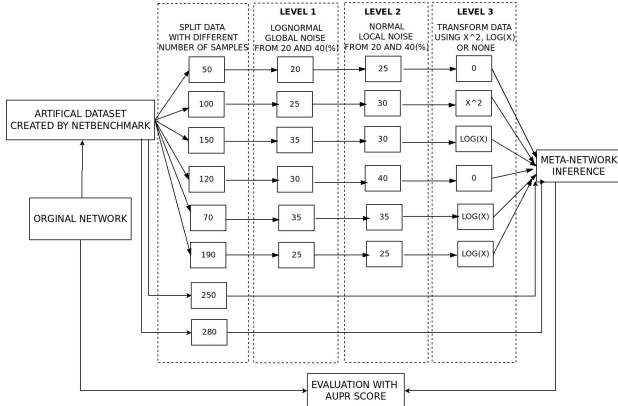


Figure 2: Framework for data collection, network prediction and validation

non-linear transformation such as x^2 or $\log(x)$. This random data transformation is not really meant to be realistic but rather to allow us to better assess the behaviour of each meta-method when faced with extreme distortion. It is worth emphasizing that the two non-informative studies remain unchanged across all experiments. A flowchart of this process is illustrated in Figure3.

B. Network prediction and validation

The schema for network prediction and validation is also illustrated in Figure3. Initially, all methods (three A, B and C, totalling nine) are used to construct a consensus GRN from the split datasets. All methods are assessed on 12 challenges (three levels of distortion for four datasets). Finally, the process is repeated for the three information-theoretic inference methods, hence totalling 36 challenges. This is done to make sure that our analysis is not method specific. In this study, the Area Under Precision Recall (AUPR) of each GRN is calculated for all methods in each challenge. Due to the randomization of various experimental parameters (noise intensity, number of samples), 10 repetitions are made. Finally, the average of the ten AUPR values, for each method on each challenge, is reported.

IV. RESULTS AND DISCUSSION

In this section, we present the experimental results of all presented methods for reconstructing GRNs from multiple expression datasets (Table II). For the A family of methods, it can be observed that normalization using z-score transformation (A2) is better than BMC (A1). This conclusion is true for all three network inference algorithms used in this paper, namely MRNET, ARACNE and CLR. Another striking feature is that batch effect removal methods like COMBAT (A3) is able to increase significantly the robustness of network inference algorithms. The results reinforce the idea that normalization alone can not remove batch effects. In the

case of method C, there are no clear differences between C1, C2 or C3, except when using ARACNE-C3.

Interestingly, we can clearly observe that all three C methods outperform all three A methods suggesting that assembling networks is better than merging datasets. This could be explained by the fact that gene expression values are very dissimilar in various experiments due to our simulated batch effects (i.e. datasets with different global and local noise). However, the particular combination CLR - A3 offers an exception to this observation. It also should be noted that assembling mutual information matrices (B methods) surpasses the two other well-known strategies (A and C) for all datasets under every different levels of distortion, in particular for MRNET and CLR. Experimental results also suggest that MRNET benefits the most from meta-analysis and CLR appears to be the most robust. This suggests that while CLR might be a better strategy for analyzing individual datasets, MRNET might be a better choice when multiple datasets are available. Although ARACNE appears to be much worse than the two other techniques, that is mainly due to a bad recall (though not visible with AUPR numbers, its precision remains quite competitive). Finally, in the B family of methods, it appears that combining MIM using random effect model (B1) is better than the two other strategies, the median method (B2) and the internal quality control index (B3).

V. CONCLUSION

In the present paper, we have proposed a framework for evaluating the different strategies for inferring GRNs from multiple expression datasets. To the best of our knowledge, this is the first systematic evaluation of the two state-of-the-art strategies for the problem of interest, namely *data merging* and *networks ensemble*. Furthermore, we present a new, but promising approach for methods based on coexpression matrices. Indeed, our set of experiments strongly suggest that assembling matrices of pairwise dependencies is a better strategy for network inference than the two commonly used ones. However, there exists many different methods of data and network assembly, as well as experimental conditions that have still to be tested in order to gain a complete understanding of the problem of meta-network inference. Moreover, as mentioned earlier, a large amount of under-exploited transcriptome data of model organisms is now available through public repositories. Thus, additionally to testing new ensemble methods, future works include the use of the best strategy to reconstruct large-scale GRNs of these model organisms.

Table II: Area under PR-Curves (the higher the better) for 9 methods on 4 datasets with 3 levels of increasing data-distortion.

MRNET		A1	A2	A3	C1	C2	C3	B1	B2	B3
<i>S1</i>	Level 1	0.094	0.129	0.121	0.144	0.131	0.129	0.139	0.119	0.135
	Level 2	0.093	0.124	0.115	0.135	0.125	0.126	0.135	0.119	0.137
	Level 3	0.103	0.111	0.103	0.126	0.117	0.120	0.122	0.110	0.118
<i>S2</i>	Level 1	0.013	0.016	0.017	0.041	0.038	0.031	0.056	0.034	0.052
	Level 2	0.014	0.017	0.018	0.028	0.027	0.024	0.034	0.023	0.031
	Level 3	0.015	0.017	0.017	0.027	0.026	0.023	0.032	0.024	0.031
<i>G1</i>	Level 1	0.057	0.103	0.132	0.141	0.140	0.144	0.173	0.148	0.164
	Level 2	0.053	0.099	0.124	0.123	0.129	0.131	0.158	0.132	0.138
	Level 3	0.045	0.078	0.110	0.116	0.121	0.123	0.149	0.124	0.133
<i>G2</i>	Level 1	0.031	0.051	0.079	0.110	0.111	0.109	0.139	0.109	0.123
	Level 2	0.025	0.047	0.071	0.096	0.096	0.097	0.127	0.100	0.118
	Level 3	0.033	0.045	0.075	0.094	0.095	0.096	0.125	0.097	0.111
Mean		0.048	0.070	0.082	0.098	0.096	0.096	0.116	0.095	0.108
ARACNE										
<i>S1</i>	Level 1	0.026	0.039	0.033	0.111	0.114	0.024	0.066	0.046	0.055
	Level 2	0.025	0.042	0.034	0.092	0.098	0.020	0.058	0.037	0.046
	Level 3	0.033	0.024	0.026	0.089	0.094	0.017	0.040	0.033	0.042
<i>S2</i>	Level 1	0.002	0.002	0.002	0.028	0.037	0.006	0.037	0.013	0.028
	Level 2	0.002	0.002	0.002	0.015	0.024	0.002	0.012	0.007	0.012
	Level 3	0.002	0.002	0.002	0.012	0.023	0.002	0.011	0.006	0.009
<i>G1</i>	Level 1	0.036	0.070	0.098	0.134	0.138	0.090	0.144	0.116	0.131
	Level 2	0.028	0.060	0.087	0.113	0.124	0.075	0.128	0.108	0.114
	Level 3	0.027	0.051	0.077	0.111	0.121	0.071	0.123	0.098	0.107
<i>G2</i>	Level 1	0.012	0.027	0.054	0.109	0.114	0.062	0.134	0.098	0.116
	Level 2	0.011	0.026	0.049	0.092	0.100	0.047	0.115	0.081	0.099
	Level 3	0.013	0.024	0.049	0.094	0.101	0.047	0.114	0.080	0.100
Mean		0.018	0.031	0.043	0.083	0.091	0.039	0.082	0.060	0.072
CLR										
<i>S1</i>	Level 1	0.128	0.145	0.144	0.137	0.136	0.139	0.143	0.143	0.142
	Level 2	0.129	0.146	0.144	0.137	0.137	0.139	0.145	0.142	0.144
	Level 3	0.137	0.142	0.142	0.137	0.136	0.138	0.143	0.141	0.143
<i>S2</i>	Level 1	0.037	0.043	0.044	0.041	0.041	0.041	0.043	0.043	0.043
	Level 2	0.033	0.042	0.044	0.041	0.041	0.041	0.044	0.043	0.043
	Level 3	0.038	0.042	0.043	0.041	0.042	0.041	0.045	0.043	0.043
<i>G1</i>	Level 1	0.089	0.151	0.161	0.138	0.138	0.130	0.160	0.159	0.150
	Level 2	0.072	0.140	0.150	0.129	0.128	0.117	0.148	0.147	0.130
	Level 3	0.067	0.121	0.140	0.126	0.125	0.110	0.145	0.143	0.123
<i>G2</i>	Level 1	0.046	0.085	0.101	0.092	0.092	0.087	0.112	0.113	0.106
	Level 2	0.036	0.076	0.090	0.081	0.080	0.073	0.100	0.100	0.089
	Level 3	0.048	0.072	0.091	0.083	0.083	0.077	0.101	0.102	0.092
Mean		0.072	0.100	0.108	0.099	0.098	0.094	0.111	0.110	0.104

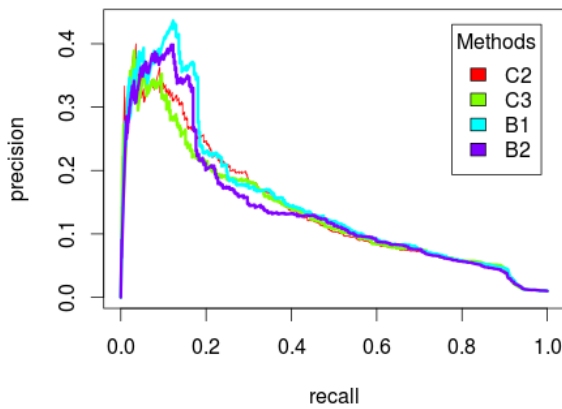


Figure 3: PR-Curves of method C2, C3, B1 and B2 on dataset S1 at level 2 of data distortion

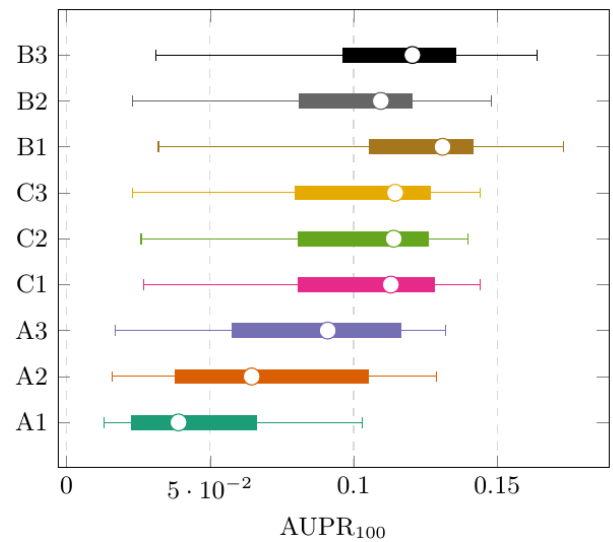


Figure 4: Boxplots for presented methods using MRNET

REFERENCES

- [1] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: Ncbi gene expression and hybridization array data repository," *Nucleic acids research*, vol. 30, no. 1, pp. 207–210, 2002.
- [2] A. Brazma, H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, P. Kemmeren, G. G. Lara, *et al.*, "Arrayexpress—a public repository for microarray gene expression data at the ebi," *Nucleic acids research*, vol. 31, no. 1, pp. 68–71, 2003.
- [3] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, M. Kellis, J. J. Collins, G. Stolovitzky, • *et al.*, "Wisdom of crowds for robust gene network inference," *Nature methods*, vol. 9, no. 8, pp. 796–804, 2012.
- [4] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner, "Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles," *PLoS biology*, vol. 5, no. 1, p. e8, 2007.
- [5] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Faveira, and A. Califano, "Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC bioinformatics*, vol. 7, no. Suppl 1, p. S7, 2006.
- [6] P. E. Meyer, F. Lafitte, and G. Bontempi, "Minet: An open source r/bioconductor package for mutual information based network inference," *BMC Bioinformatics*, vol. 9, 2008.
- [7] K. G. Kugler, L. A. Mueller, A. Graber, and M. Dehmer, "Integrative network biology: graph prototyping for co-expression cancer networks," *PLoS One*, vol. 6, no. 7, p. e22843, 2011.
- [8] J. Taminau, S. Meganck, C. Lazar, D. Steenhoff, A. Coletta, C. Molter, R. Duque, V. de Schaezen, D. Y. W. Solís, H. Bersini, *et al.*, "Unlocking the potential of publicly available microarray data using insilicodb and insilicomerger r/bioconductor packages," *BMC bioinformatics*, vol. 13, no. 1, p. 335, 2012.
- [9] W. Huber, A. Von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron, "Variance stabilization applied to microarray data calibration and to the quantification of differential expression," *Bioinformatics*, vol. 18, no. suppl 1, pp. S96–S104, 2002.
- [10] V. Belcastro, V. Siciliano, F. Gregoretti, P. Mithbaokar, G. Dharmalingam, S. Berlingieri, F. Iorio, G. Oliva, R. Polishchuck, N. Brunetti-Pierri, *et al.*, "Transcriptional gene network inference from a massive dataset elucidates transcriptome organization and gene function," *Nucleic acids research*, vol. 39, no. 20, pp. 8677–8688, 2011.
- [11] P. Adler, R. Kolde, M. Kull, A. Tkachenko, H. Peterson, J. Reimand, and J. Vilo, "Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods," *Genome biology*, vol. 10, no. 12, p. R139, 2009.
- [12] J. T. Leek, R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, and R. A. Irizarry, "Tackling the widespread and critical impact of batch effects in high-throughput data," *Nature Reviews Genetics*, vol. 11, no. 10, pp. 733–739, 2010.
- [13] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.
- [14] L. Dyrskjöt, M. Kruhøffer, T. Thykjaer, N. Marcussen, J. L. Jensen, K. Møller, and T. F. Ørntoft, "Gene expression in the urinary bladder a common carcinoma in situ gene expression signature exists disregarding histopathological classification," *Cancer Research*, vol. 64, no. 11, pp. 4040–4048, 2004.
- [15] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.
- [16] T. Hase, S. Ghosh, R. Yamanaka, and H. Kitano, "Harnessing diversity towards the reconstructing of large scale gene regulatory networks," *PLoS Comput Biol*, vol. 9, no. 11, p. e1003361, 2013.
- [17] D. D. Kang, E. Sibille, N. Kaminski, and G. C. Tseng, "Metaqc: objective quality control and inclusion/exclusion criteria for genomic meta-analysis," *Nucleic acids research*, vol. 40, no. 2, pp. e15–e15, 2012.
- [18] A. Ramasamy, A. Mondry, C. C. Holmes, and D. G. Altman, "Key issues in conducting a meta-analysis of gene expression microarray datasets," *PLoS medicine*, vol. 5, no. 9, p. e184, 2008.
- [19] P. Wirapati, C. Sotiriou, S. Kunkel, P. Farmer, S. Pradervand, B. Haibe-Kains, C. Desmedt, M. Ignatiadis, T. Sengstag, F. Schutz, *et al.*, "Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures," *Breast Cancer Res*, vol. 10, no. 4, p. R65, 2008.
- [20] C. Desmedt, B. Haibe-Kains, P. Wirapati, M. Buyse, D. Larsimont, G. Bontempi, M. Delorenzi, M. Piccart, and C. Sotiriou, "Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes," *Clinical Cancer Research*, vol. 14, no. 16, pp. 5158–5165, 2008.
- [21] F. Hong, R. Breitling, C. W. McEntee, B. S. Wittner, J. L. Nemhauser, and J. Chory, "Rankprod: a bioconductor package for detecting differentially expressed genes in meta-analysis," *Bioinformatics*, vol. 22, no. 22, pp. 2825–2827, 2006.
- [22] V. Cestarelli, G. Fiscon, G. Felici, P. Bertolazzi, and E. Weitschek, "Camur: Knowledge extraction from rna-seq cancer data through equivalent classification rules," *Bioinformatics*, p. btv635, 2015.
- [23] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

- [24] A. H. Sims, G. J. Smethurst, Y. Hey, M. J. Okoniewski, S. D. Pepper, A. Howell, C. J. Miller, and R. B. Clarke, "The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets—improving meta-analysis and prediction of prognosis," *BMC medical genomics*, vol. 1, no. 1, p. 1, 2008.
- [25] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using z score transformation," *The Journal of molecular diagnostics*, vol. 5, no. 2, pp. 73–81, 2003.
- [26] C. Chen, K. Grennan, J. Badner, D. Zhang, E. Gershon, L. Jin, and C. Liu, "Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods," *PLoS one*, vol. 6, no. 2, p. e17238, 2011.
- [27] C. Lazar, S. Meganck, J. Taminau, D. Steenhoff, A. Coletta, C. Molter, D. Y. Weiss-Solís, R. Duque, H. Bersini, and A. Nowé, "Batch effect removal methods for microarray gene expression data integration: a survey," *Briefings in bioinformatics*, p. bbs037, 2012.
- [28] P. Bellot Pujalte, P. J. Salembier Clairon, A. Oliveras Vergés, and P. Meyer, "Study of normalization and aggregation approaches for consensus network estimation," in *IEEE SSCI 2015: 2015 IEEE Symposium Series on Computational Intelligence; 7-10 December 2015, Cape Town, South Afrika*, pp. 1–6, Institute of Electrical and Electronics Engineers (IEEE), 2015.
- [29] K. Wang, M. Narayanan, H. Zhong, M. Tompa, E. E. Schadt, and J. Zhu, "Meta-analysis of inter-species liver co-expression networks elucidates traits associated with common human diseases," *PLoS Comput Biol*, vol. 5, no. 12, p. e1000616, 2009.
- [30] F. L. Schmidt and J. E. Hunter, *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications, 2014.
- [31] P. Bellot, C. Olsen, P. Salembier, A. Oliveras-Verges, and P. E Meyer, "Netbenchmark: A bioconductor package for reproducible benchmarks of gene regulatory network inference," *BMC bioinformatics*, 16.1 (2015): 1.
- [32] P. E Meyer, L. Kontos, and G. Bontempi, "Biological network inference using redundancy analysis," *Bioinformatics Research and Development*, pp. 16–27. Springer Berlin Heidelberg, 2007.