<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Meta-analysis on gene regulatory networks discovered by pairwise Granger causality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Tam, GHF; Hung, YS; Chang, C</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 7th IEEE International Conference on Systems Biology (ISB 2013), Huangshan, China, 23-25 August 2013. In Conference Proceedings, 2013, p. 123-128</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/186741">http://hdl.handle.net/10722/186741</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>IEEE International Conference on Systems Biology Proceedings. Copyright © IEEE.; ©2013 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Meta-Analysis on Gene Regulatory Networks Discovered by Pairwise Granger Causality

Gary Hak Fui Tam and Yeung Sam Hung
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong, China
htam@eee.hku.hk, yshung@eee.hku.hk

Chunqi Chang
School of Electronic and Information Engineering
Soochow University
Suzhou, Jiangsu Province, China
cqchang@suda.edu.cn

Abstract—Identifying regulatory genes partaking in disease development is important to medical advances. Since gene expression data of multiple experiments exist, combining results from multiple gene regulatory network discoveries offers higher sensitivity and specificity. However, data for multiple experiments on the same problem may not possess the same set of genes, and hence many existing combining methods are not applicable. In this paper, we address this problem using a number of meta-analysis methods and compare their performances. Simulation results show that vote counting is outperformed by methods belonging to the Fisher’s chi-square (FCS) family, of which FCS test is the best. Applying FCS test to the real human HeLa cell-cycle dataset, degree distributions of the combined network is obtained and compared with previous works. Consulting the BioGRID database reveals the biological relevance of gene regulatory networks discovered using the proposed method.

Keywords—gene regulatory network; meta-analysis; multiple experiments; pairwise Granger causality; Fisher’s chi-square test

I. INTRODUCTION

Gene regulatory network (GRN) discovery detected gene-gene interactions from gene expression data [1]–[11]. Genes identified to play roles in disease development are potential targets of future drugs [2]–[4]. Since DNA microarray technology has been extensively applied to various medical problems, data of multiple experiments (datasets) concerning the same problem often exist. Combining results from multiple experiments offers higher statistical power and the discovered network is more reliable, but it is not a trivial task [1]. Particularly, data for multiple experiments may not contain the same set of genes. For example, in the widely studied human HeLa cell-cycle dataset [12], out of a total of 1134 periodic genes, the number of genes contained in experiments 1–3 are only 828, 828 and 1099, respectively [11]. Although experiments 1 and 2 both have “828” genes, they correspond to slightly different gene subsets.) Therefore, many combining methods (e.g. [1], [8]) assuming the same gene set are not applicable. Meta-analysis (MA) provides a direct solution to the problem of how these experiments may be combined.

Previously, MA on gene expression data focused on detection of differentially expressed genes, e.g. see [13], [14]; while applications to GRN discovery are relatively fewer. Nevertheless, recent studies [15]–[17] demonstrate that MA is a powerful tool for combining multiple results for GRN discovery. However, to our knowledge, we have not seen a paper comparing MA methods in the usual situation that multiple experiments do not have the same set of genes. In this paper, we apply several most commonly adopted MA methods to this problem and compare their performance. First, the selected MA methods are evaluated by synthetic data. Then, the best performing MA method is applied to the HeLa dataset.

II. PAIRWISE GRANGER CAUSALITY

Granger causality (GC) [18] has a number of merits over other time-series analysis methods [19], and hence is adopted here for GRN discovery. Consider two time series X and Y, if Y can help predict the future of X, then Y “Granger-causes” X. Suppose the time series of these two variables have data length T, denoting their values at time t by X(t) and Y(t) (t=1,2,...,T), respectively, they can be modelled by a bivariate autoregressive (AR) model:

\[
X_t = \sum_{i=1}^{p} a_{ij} X_{t-i} + a_{2j} Y_{t-i} + \epsilon_t,
\]

\[
Y_t = \sum_{i=1}^{p} a_{2i} X_{t-i} + a_{22j} Y_{t-i} + \xi_t,
\]

where \( p \) is the model order, \( a_{ij} \) (\( i,j = 1,2 \)) are coefficients of the model, and \( \epsilon_t \) and \( \xi_t \) represent residuals. The coefficients can be estimated by ordinary least squares, and the GCs between X and Y can be detected by F tests [10], [11].

For a system with \( n \) variables, the above modelling can be applied to each pair of variables, which is referred to as pairwise GC (PGC). An \( n \)-variable network has a total of \( M = n(n-1) \) possible directed edges, and each edge will be assigned a \( p \)-value by PGC. However, not all \( p \)-values are valid, since some bivariate models may not fit to data well and they may give rise to false discoveries easily. These badly fitted models are excluded by model validation, hence performance of PGC can be increased [10]. The valid \( p \) values then undergo Benjamini-Hochberg false discovery rate (FDR) controlling procedure [20], where we need to provide a FDR level \( q \) to be controlled, and the procedure will return a threshold on \( p \)-value.
Those edges with \( p \)-values \( \leq \) the threshold are taken as detected GCs that constitute the discovered network. Further details can be found in [10] and [11].

### III. META-ANALYSIS METHODS

In GRN discovery, a gene is referred to as a variable, so “gene” and “variable” may be used interchangeably in literature. We formulate the problem as follows. Suppose we have data of \( k \) experiments on a total of \( n \) genes. Each experiment \( E_i \) (\( i = 1, 2, \ldots, k \)) contains \( n_i \) genes, which is a subset of the \( n \) genes. Under the null hypothesis that gene \( Y \) does not Granger-cause \( X \), for each of these possible edges, we count the number of times (votes) this edge is included in some network \( G_i \), of \( n \) genes. In the following, we consider methods belonging to two families of commonly adopted MA methods [14]-[17]: vote counting and Fisher’s chi-square methods.

#### A. Vote Counting

\( G_{MA} \) has \( M = n(n - 1) \) possible edges. In vote counting (VC), for each of these possible edges, we count the number of times (votes) this edge is included in some \( G_i \). Those edges with votes \( \geq \) a certain threshold constitute \( G_{MA} \). If we only require the number of votes \( \geq 1 \), then \( G_{MA} \) includes all edges in \( G_i \). We refer to this network as VC1. Similarly, \( G_{MA} \) composed of edges with votes \( \geq 2 \) is referred to as VC2, etc.

#### B. Fisher’s Chi-Square Methods

According to [21], given \( p \)-values \( p_i \) (\( i = 1, 2, \ldots, \kappa \)) from \( \kappa \) independent studies, the Fisher’s chi-square (FCS) test statistic can be computed as:

\[
\chi^2_F = -2\sum_{i=1}^{\kappa} \ln p_i.
\]

Regarding our GRN discovery problem, if an edge (say, from gene \( Y \) to gene \( X \)) has \( \kappa \) \( p \)-values from \( G_{i} \) (\( \kappa \leq k \)), its \( \chi^2_F \) can be computed by (2). Under the null hypothesis that gene \( Y \) does not Granger-cause \( X \), this \( \chi^2_F \) has a chi-square distribution with \( 2\kappa \) degrees of freedom. Hence, a new \( p \)-value \( p_F \) for combining \( \kappa \) studies can be obtained. After computing \( p_F \) for all potential edges, again we apply Benjamini-Hochberg FDR control (often at another FDR level \( q \)) to fix a \( p \)-value threshold and edges with \( p_F \) below the threshold will constitute \( G_{MA} \). We term this method as FCS test (FCST).

Alternatively, \( \chi^2_F \) can also be used to rank the potential edges, e.g. as in [15]. We term this approach as FCS Ranking (FCSR). To compare with FCST, suppose the \( G_{MA} \) obtained by FCST consists of \( H_F \) edges, we also take \( H_F \) edges with largest \( \chi^2_F \) to be another \( G_{MA} \), which is referred to as FCSR1. To compare with VC2, we take the same number of edges as VC2 from the FCSR to be a \( G_{MA} \) – denoted as FCSR2.

Since the FCSR approach does not have a natural threshold, we derive one as follows. Suppose \( G_i \) has \( H_i \) edges (remark: \( H_i \) comes from Benjamini-Hochberg FDR control at level \( q \)), we scale up \( H_i \) by the ratio of all possible edges (the search space in GRN discovery) between \( G_i \) and \( G_{MA} \):

\[
H' = \frac{1}{k} \sum_{i=1}^{k} \frac{n(n-1)}{n_i(n_i-1)} H_i.
\]

\( H' \) is then rounded to nearest integer, which is the number of edges to be taken. The resulted \( G_{MA} \) is denoted as FCSR3.

It should be remarked that FCS methods are not restricted to PGC. They can also be applied to other GRN discovery methods, as long as every discovered edge has a \( p \)-value. VC does not even require a \( p \)-value.

### IV. SIMULATIONS

#### A. Synthetic Data

To generate synthetic data, we adopt an 8-variable AR model (i.e. \( n = 8 \)), which is composed of two smaller models of 5 and 3 variables [22], [23]:

\[
\begin{align*}
x_i(t) &= 0.95\sqrt{2} x_i(t-1) - 0.9025 x_i(t-2) + w_i(t) \\
x_i(t) &= 0.5 x_i(t-2) + w_i(t) \\
x_i(t) &= -0.4 x_i(t-3) + w_i(t) \\
x_i(t) &= -0.5 x_i(t-2) + 0.25\sqrt{2} x_i(t-1) + 0.25\sqrt{2} x_i(t-1) + w_i(t) \\
x_i(t) &= -0.25\sqrt{2} x_i(t-1) + 0.25\sqrt{2} x_i(t-1) + w_i(t) \\
x_i(t) &= 0.8 x_i(t-1) - 0.5 x_i(t-2) + 0.4 x_i(t-1) + w_i(t) \\
x_i(t) &= 0.9 x_i(t-1) - 0.8 x_i(t-2) + w_i(t) \\
x_i(t) &= 0.5 x_i(t-1) - 0.2 x_i(t-2) + 0.5 x_i(t-1) + w_i(t)
\end{align*}
\]

The corresponding regulatory network is shown in Fig. 1. \( w_i \) and initial values of \( x_i \) (\( j = 1, 2, \ldots, 8 \)) are independent Gaussian noises of zero mean and unit variance. This 8-variable model has order \( p = 3 \). Starting with some initial values, we iterate (4) to get long time series of length 140 \( \times \) 3 . For each of the 140-long segments, the first 100 time points (transient) are dropped and the subsequent 40 time points are taken as experimental data (i.e. \( T = 40 \)). Hence, data of 3 experiments are obtained (\( k = 3 \)).

To simulate the situation that multiple experiments may not contain the same set of genes, we purposely exclude data of a
few variables in each experiment, where 0–3 variables may be excluded randomly. Since the average exclusion is 1.5 variables, on average each experiment contains \( n_i = 6.5 \) variables. We also require that the union of variables in these 3 experiments should be 8, i.e. the 3 experiments together should contain all the 8 variables in (4).

### B. Performance Measures

Here we describe some performance measures for a discovered network if ground truth is known. Suppose a discovered network has \( H \) edges, \( e \) of them are true positives (TPs), then precision \( P = e / H \). Denote the number of edges of the ground truth network by \( L \), then recall \( R = e / L \). The mean compromising \( P \) and \( R \) is \( F_i = 2PR/(P+R) \). \( P \), \( R \) and \( F_i \) all range from 0 to 1, and they are different if \( H \neq L \). In extreme cases, a 1-edge discovered network can have \( P = 1 \) (if this edge is a TP) but a low \( R \); a discovered network consists of all \( M \) possible edges can have \( R = 1 \) but a low \( P \). But \( F_i = 1 \) requires the discovered network matches the ground truth exactly. Thus in the following, we focus more on \( F_i \).

### C. Results and Discussion

Using the synthetic data as generated above, we apply PGC with model validation to each experiment \( E_i \) separately to obtain discovered network \( G_i \) (\( i = 1, 2, 3 \)). We assume ground truth is unknown, so the model order of each PGC application is estimated by Akaike information criterion (AIC) [11], [24]. For the Benjamini-Hochberg FDR control, \( q \) is fixed at 0.05.

When the 3 \( G_i \)'s are ready, MA methods are applied to them. Each MA method gives a \( G_{\text{MA}} \), and its performance measures can be calculated by comparing with the ground truth network in Fig. 1. Since combining 3 experiments offers a much higher statistical power, the new \( p \)-values \( p_i \) in FCST are much smaller than those in \( G_i \). Hence, \( q_f = 0.001 \) is used for Benjamini-Hochberg FDR control in FCST.

The above exercise is repeated 30 times with different initial values for \( x_i \). Results of these 30 runs are shown in Table I. Besides the MA methods, we also show the results of \( G_i \) in the first row – labelled by “1-Expt”, because each \( G_i \) only comes from a single experiment. It turns out that \( \approx 6.6 \) (\( \pm 1.1 \)) variables are included in each experiment. The model order estimated by AIC \( \approx 3.3 \) (\( \pm 1.1 \)). The average number of edges in \( G_i \) is 5.0, and \( G_i \)'s have average \( F_i \) of 0.50.

From Table I(a), MA methods generally give higher \( F_i \) than 1-Expt. VC1 contains too many edges (\( H = 10.6 \) is substantially larger than \( L = 7 \)), so \( R \) is high, but \( P \) is even lower than 1-Expt case. VC2 has high \( P \). But it is already a bit too strict that \( H \) is substantially smaller than \( L \), so results in low \( R \), making that \( F_i \) is lower than VC1’s. VC3 is really strict that very few edges can pass this criterion – in average \( H \approx 0.7 \) only. However, it is amazing that all edges passing this criterion are correct (\( P = 1 \))! Yet, since \( H \) is too small, VC3 is not useful in practice.

FCST and FCSR1 give the same and the highest (best) \( F_i = 0.69 \) in these 30 runs. Discovered edges by FCST generally have \( p \)-values < 0.0006. FCSR2 should be compared with VC2: though \( H \) are the same, FCSR2 gives higher measures for \{TP, \( P \), \( R \), \( F_i \)\}, meaning that the ranking ordered by \( X_i^2 \) (through \( p \)-values from single experiment) is better than simple vote counting. FCSR3 turns out also to give high \( F_i \).

Although FCST and FCSR1 give exactly the same results in the 30 runs, they are conceptually different. The same results here can be understood as follows. \( H \) of VC1, VC2 and FCST are 10.6, 3.7 and 6.2, respectively. That means the thresholding of FCST usually lies in region 1, i.e. \( \kappa = 1 \). With the same \( \kappa \), the two criteria FCST and FCSR1 are the same. Nevertheless, we have repeated similar exercise as above 100 times with larger number of experiments \( k = 10 \), then \( F_i \) of FCST is a bit higher than FCSR1’s, where the difference is 0.006. Their discovered networks differ in 9 out of 100 runs, where FCST is better in 7 runs and worse in 2 runs. It means that FCST and FCSR1 give the same discovered network most of the time. However, there are still some situations that they give different discovered networks, and FCST often works better in such cases. This reveals that the degrees of freedom in the chi-square test also count. Since \( X_i^2 \) is only a test statistic, and FCST rigorously returns a \( p \)-value, FCST is preferred conceptually. Moreover, there is lack of guidance for setting a threshold for FCSR. Therefore, among the MA methods listed in Table I, we conclude that FCST is the best. FCST will be applied to real data in next section.

| TABLE I. SIMULATION RESULTS OF SINGLE EXPERIMENT AND META-ANALYSIS METHODS |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| (a) Mean | (b) Standard Deviation |
| \( H \) | \( TP \) | \( P \) | \( R \) | \( F_i \) | \( H \) | \( TP \) | \( P \) | \( R \) | \( F_i \) |
| 1-Expt | 5.0 | 3.3 | 0.65 | 0.46 | 0.50 | 3.1 | 1.7 | 0.27 | 0.24 | 0.20 |
| VC1 | 10.6 | 5.4 | 0.55 | 0.78 | 0.63 | 3.4 | 1.0 | 0.16 | 0.14 | 0.11 |
| VC2 | 3.7 | 3.3 | 0.84 | 0.47 | 0.58 | 2.1 | 1.4 | 0.18 | 0.19 | 0.17 |
| VC3 | 0.7 | 1.8 | 1.00 | 0.25 | 0.39 | 1.0 | 0.8 | 0.00 | 0.11 | 0.14 |
| FCST | 6.2 | 4.6 | 0.79 | 0.65 | 0.69 | 2.2 | 1.0 | 0.17 | 0.15 | 0.11 |
| FCSR1 | 6.2 | 4.6 | 0.79 | 0.65 | 0.69 | 2.2 | 1.0 | 0.17 | 0.15 | 0.11 |
| FCSR2 | 3.7 | 3.5 | 0.89 | 0.50 | 0.61 | 2.1 | 1.3 | 0.15 | 0.19 | 0.15 |
| FCSR3 | 7.3 | 4.8 | 0.69 | 0.69 | 0.67 | 2.3 | 1.0 | 0.15 | 0.14 | 0.10 |
We have also repeated the simulation of 3 experiments using different data lengths \( T = 20, 30 \) and 40 in \( E_1 \), \( E_2 \) and \( E_3 \), respectively. This exercise leads to the same conclusion as before: FCST remains the best.

V. META-ANALYSIS ON REAL HE.LA DATA

The human HeLa dataset [12] contains time-series gene expression data from cell division cycle experiments using cDNA microarrays. Most GRN discovery studies on this dataset (e.g. [5]-[7]) concentrated on analyzing the 1134 periodic genes using data of experiments 1–3. In a previous work [11], for the sake of comparison, we have applied PGC with model validation also to the 1134 periodic genes in experiments 1–3. Discovered networks \( (G_i) \) consisting thousand of edges are obtained. A brief summary is given in Table II. \( E_3 \) yields substantially more edges than \( E_1 \) and \( E_2 \) because \( E_3 \) contains more genes and longer series which offer higher statistical power for GC detection.

Now, we apply FCST to the 3 \( G_i \)’s. Since \( p \)-values from PGC are generally larger in HeLa real data (one order of magnitude higher than those in synthetic data), we relax \( q_F \) to 0.01 such that the number of discovered edges in \( G_{MA} \) is more reasonable and not too small. The resulted \( G_{MA} \) contains a total of 29146 edges, with \( p \)-values \( (p_F) \) ranging from 0.0066 to \( 6.0 \times 10^{-11} \), and their geometric mean is \( 9.3 \times 10^{-4} \). It turns out that these 29146 edges include all edges with \( VC \geq 2 \). \( E_1 \), \( E_2 \) and \( E_3 \) contribute 2301, 3878 and 24001 edges, respectively, and they have some overlaps.

Since the ground truth network is unknown for this real dataset and hence previous measures \( \{ TP, P, R, F_1 \} \) cannot be computed, we compute degree distributions instead [25], which also provides insight to the structure of large networks [25], [26]. The discovered network \( G_{MA} \) involves 1101 genes, where 1088 have in-degree \( > 0 \) and 1007 have out-degree \( > 0 \). The (in-/out-) degree distributions of the 1101 genes are plotted in Fig. 2, which shows similar power-law decay as in [5] and [7]. Denoting degree by \( d \), the degree distribution in Fig. 2(c) has a decay approximately as \( d^{-2} \). The degree exponent 2 is consistent with our previous work [11], but it is smaller than 2.7 in [5] which used \( E_1 \) only.

Table III shows the top 10 genes with maximum (in-/out-) degrees. Network hubs (genes with high degrees) shown in Table III(c) generally have higher out-degrees than in-degrees, meaning that they act as sources rather than recipients in interactions. This agrees with our previous works [10], [11]. Genes 3.UBE2C, 5.CDC2, 10.FLJ10468, 16.TOP2A, 23.KNSL5, 26.CDC2, 42.DJ616B8.3, 87.USF1 were also found to have high degrees in [6].

![Fig. 2. Degree distributions of the 1101 genes involved in discovered network by FCST for HeLa dataset. Excess kurtosis = kurtosis – 3. So, skewness and excess kurtosis are both zero for Gaussian distribution.](image-url)
Remarkedly, FCST can also be applied to networks obtained from Table III, 5.CDC2 is the network hub with highest degree. 484 genes are connected to it in our discovered network. CDC2 is known as cyclin-dependent kinase 1, with official symbol CDK1. According to the Entrez Gene record on CDC2 [28], “the protein encoded by this gene is a member of the Ser/Thr protein kinase family”. This protein is “essential for G1/S and G2/M phase transitions of eukaryotic cell cycle”. In our discovered network, many interactions with CDC2 are documented in the BioGRID database [29]. For example, using low-throughput experiment, Qi et al. [30] found that BUB1 activity was enhanced by CDK1-mediated phosphorylation. This is consistent with the discovered edge “S.CDC2 → 15.BUB1”. The interaction “5.CDC2 ⇔ 17.CKS2” is confirmed by 2 papers [31], [32], where [32] mentioned that mammalian CKS2 bound CDK1 and participated in cell-cycle control. Kong et al. [33] found that “Cyclin F (CCNF) regulates the nuclear localization of cyclin B1 (CCNB1) through a cyclin-cyclin interaction”, which involved “a complex composed of CDC2 and a B-type cyclin”. This agrees well with our discovered edges involving these 3 genes, as drawn in Fig. 3.

![Fig. 3. Discovered edges involving genes 5.CDC2, 11.CCNF and 50.CCNB1.](image)

From Table III, 5.CDC2 is the network hub with highest degree. 484 genes are connected to it in our discovered network. CDC2 is known as cyclin-dependent kinase 1, with official symbol CDK1. According to the Entrez Gene record on CDC2 [28], “the protein encoded by this gene is a member of the Ser/Thr protein kinase family”. This protein is “essential for G1/S and G2/M phase transitions of eukaryotic cell cycle”. In our discovered network, many interactions with CDC2 are documented in the BioGRID database [29]. For example, using low-throughput experiment, Qi et al. [30] found that BUB1 activity was enhanced by CDK1-mediated phosphorylation. This is consistent with the discovered edge “S.CDC2 → 15.BUB1”. The interaction “5.CDC2 ⇔ 17.CKS2” is confirmed by 2 papers [31], [32], where [32] mentioned that mammalian CKS2 bound CDK1 and participated in cell-cycle control. Kong et al. [33] found that “Cyclin F (CCNF) regulates the nuclear localization of cyclin B1 (CCNB1) through a cyclin-cyclin interaction”, which involved “a complex composed of CDC2 and a B-type cyclin”. This agrees well with our discovered edges involving these 3 genes, as drawn in Fig. 3.

### VI. Conclusion and Discussion

We have applied MA methods to combine GRN discoveries (by PGC) from data of multiple experiments that may not have the same set of genes. Simulation results show that vote counting is outperformed by FCS methods, among which FCST is the best. Applying FCST to the HeLa dataset, degree distributions of the combined network is obtained and compared with previous works. The gene CDC2, which plays important roles in cell-cycle regulation, is found to be the network hub with highest degree. Many of our discovered edges are documented in the BioGRID database. It should be remarked that FCST can also be applied to networks obtained by GRN discovery methods besides PGC, as long as $p$-values of all discovered edges are available.

A potential improvement to the above scheme is to consider network inference under a single model which optimizes the network structure by consulting data of all multiple experiments simultaneously, such as the two methods in [1] and [8]. However, these methods need to be modified for the present situation that multiple experiments do not possess the same set of genes. This is a non-trivial task and demands further efforts.

### Acknowledgment

The work described in this paper is partially supported by Hong Kong SAR Research Grants Council (Project No. HKU762111M) and CRCG of the University of Hong Kong. We would like to thank the anonymous reviewers for comments that help improve the paper.

### References


