Effects of microarray noise on inference efficiency of a stochastic model of gene networks

DANIEL A. CHARLEBOIS[1], ANDRE S. RIBEIRO[2,5], ANTTI LEHMUSSOLA[5], JASON LLOYD-PRICE[1], OLLI YLI-HARJA[5], STUART A. KAUFFMAN[1,3,4]
[1]Institute for Biocomplexity and Informatics, University of Calgary, Alberta, CANADA
[2]Centre for Computational Physics, University of Coimbra, P-3004-516 Coimbra, PORTUGAL
[3]Dept. of Physics and Astronomy, University of Calgary, Alberta, CANADA
[4]Dept. of Biological Science, University of Calgary, Alberta, CANADA
[5]Institute of Signal Processing, Tampere University of Technology, Tampere, FINLAND
dacharle@ucalgary.ca

Abstract: - The appearance of microarray technology led to the development of algorithms to infer the structure underlying the dynamics of gene regulatory networks (GRNs) from gene expression data. Yet, this technique is currently highly noisy, leading to the question of how inferable are GRNs from this data. To answer this question, we use realistic models of GRNs [1, 2, 3] and artificially introduce realistic measurement noise [4] to the resulting time series. We analyze the inference efficiency of IADGRN [5], an inference algorithm of structure and logic, from this data. Since microarrays are attained from multiple identical cell measurements and the effects of averaging their expression levels is still poorly understood, we investigate these effects on the mutual information between pairs of genes (pI) [6, 7] averaged over the time series of multiple identical cells, which are assumed to be initially synchronized. We test noise level effects using two noise models (Rocke and hierarchical error model, Hem). It is found that the Hem model more strongly disrupts correlations for low noise levels. We show that although time series binarization causes information loss, it makes inference more robust to low levels of noise. Also, while indirect interactions are not inferable even for low levels of noise, above a certain relatively small noise ratio even strongly correlated genes lose any significant correlation for both noise models. This shows how inaccurate current inference algorithms are when inferring from data with realistic characteristics.

Key-Words: - Gene Regulatory Networks, Delayed Stochastic Simulation Algorithm, Microarray, Inference.

1 Introduction

One fundamental problem to solve in Systems Biology is to unravel structure and logic of GRNs among thousands of genes and their RNA and protein products [8]. Eventual medical implications include stem cells and regenerative medicine [9] and novel approaches to cancer "differentiation" therapy [10]. To guide such procedures it is essential to understand the GRN, analyze its structure, logic, and dynamics, using a dynamical systems framework [11]. Microarray techniques provide data to solve the inverse problem [12]: given the genes' expression profiles over tissues, conditions or time, infer the structure and logic of the GRN that gives rise to such activity patterns. Recently, several approaches for the reverse engineering of networks from gene expression data have been proposed [12, 13, 14].

One problem with microarray technology is its high noise level [15]. Here, we focus on the effect that such noise has in inference efficiency. We use artificial networks from which temporal series of the dynamics is extracted, without external noise sources, and the effects of microarray noise measurement are artificially added. This allows exact control on how much noise is introduced and how much inference is affected. Since in real experiments the ground truth (i.e. the true underlying structure and logic of the GRN) is not known, only using artificial networks and noise simulation can the effects of noise be correctly characterized.

For this purpose, so far the models of artificial GRNs used are Boolean networks and coupled differential equations [11, 16]. Yet, recent measurements on gene expression and small synthetic GRNs provided evidence of the phenotypic variability due to the stochastic nature of gene expression (for a review see [17]).

When the system has many molecules, its chemical dynamics (i.e. the variations in the concentrations of the chemical species present) can be computed approximately using continuous differential equations. However, in real GRNs,

genes exist only in very small quantities (usually only one or two copies of the same gene); therefore a "mean field" approach is not accurate [2, 17, 18]. For this reason, GRN dynamics should be based on the stochastic simulation algorithm (SSA) [19].

Additionally, while non-delayed reactions correctly model equilibrium states of simple synthetic GRNs, the dynamics of real GRNs are so complex (e.g., due to many feedback loops [20]) that time delays, namely for transcription and translation, should be included [20, 21].

We model GRNs using a general set of reactions [1, 2, 3], where genes interact via their products of expression, and where transcription and translation are modeled as time-delayed events [22]. The dynamics are driven by a generalized delay algorithm that handles multiple distinct delayed output events for each input event [21].

Using this model to generate time series of gene expression from GRN dynamics, we study how microarray noise, artificially generated [5, 23], affects inference by IADGRN [5] (an inference algorithm from time series of gene expression).

Usually, an inference algorithm's efficiency is tested using artificial gene network models to generate gene expression profiles or time series (see e.g., [12, 13, 14]). A few works have considered and modeled artificial array noise [4, 24, 25, 26]. The gene networks models are usually deterministic, namely, either Boolean networks [27] or ODE models [4, 28].

Here we present the first attempt to test the ability of an inference algorithm, using time series of gene expression created by an artificial delayed stochastic model of gene networks that correctly accounts for the internal noise and time delays in the underlying dynamics, to which realistic array noise [4] is added to the time series as well.

We begin by introducing the GRN model. Next, the method for introducing realistic array noise is presented, followed by the inference algorithm. Finally, we present the results and conclusions.

2 Methods

The model of GRN consists of reactions modeling gene transcription and translation, binding of transcription factors (proteins) to gene operator sites, decay of proteins, dimmer formations, and so on (for a detailed description see [1, 2, 3]).

In general, GRNs are here generated from the following reactions: for gene i=1,..,N, basal transcription of promoter Pro_i by RNA Polymerase (RNAp) (Equation 1), activated transcription where operators sites j of gene i, $Pro_{i,j}$, are occupied by one or more activators (Equation 2), translation of RNA by ribosomes (Rib) into proteins (p_i) (Equation 3), repression/unrepression of a gene at operator site j (Equations 4,5, and 6), decay of ribosome binding sites (RBS) and proteins (p_i's) (Equation 7), and proteins polymerization (here limited to dimmers for simplicity) (Equations 8 and 9), that can act as indirect repression or activation. Unless time-delays (τ's) are explicitly represented in the products, all events, reactant depletions, and product appearances, occur instantaneously at t:

$$RNAp + \Pr o_i \xrightarrow{k_{i,bas}} \Pr o_i(\tau_i^1) + RBS_i(\tau_i^1) + RNAp(\tau_i^2) \quad (1)$$

$$\Pr o_{i,op_i} + RNAp \xrightarrow{k_{i,op_i}} \Pr o_{i,op_i}(\tau_i^1) + RNAp(\tau_i^2) + RBS_i(\tau_i^3) \quad (2)$$

$$Rib + RBS_i \xrightarrow{k_{i,RBS}} RBS_i(\tau_i^3) + Rib(\tau_i^4) + p_i(\tau_i^4) \quad (3)$$

$$\Pr o_{i,j} + p_w \xrightarrow{k_{ij,w}} \Pr o_{i,j}p_w \quad (4)$$

$$\Pr o_{i,j}p_w \xrightarrow{k_{w,ij}} \Pr o_{i,j} + p_w \quad (5)$$

$$\Pr o_{i,j}p_z + p_w \xrightarrow{k_{ijz,w}} \Pr o_{i,j} + p_z + p_w \quad (6)$$

$$RBS_i \xrightarrow{k_{dr,i}} \varnothing,\ p_i \xrightarrow{k_{dp,i}} \varnothing \quad (7)$$

$$p_i + p_j \xrightarrow{k_{i,\dim}} p_{i,j} \quad (8)$$

$$p_{i,j} \xrightarrow{k_{i,\mathrm{undim}}} p_i + p_j \quad (9)$$

Protein and RBS degradation are assumed to occur at a constant rate and are modeled as unimolecular reactions. Since genes have multiple operator sites, a promoter "state" is defined by the combined states of its operator sites, i.e., if they have their respective transcription factor bound to it or not. We adopted the following notation: an array is used as the index of a promoter, e.g, $Pro_{i,op(i)}$ where $op(i) = (0,p_w,0)$. In this case, only p_w is bound to operator site 2 of gene i, while the other operator sites are free [18].

SGNSim [18], here used, generates random networks from these reactions. Multiple identical GRNs' time series can be averaged to obtain the effect of measuring the states of multiple cells.

To the resulting averaged time series, we add synthetic noise with realistic characteristics, by emulating the properties of the measurement system [5, 23]. This method has been successfully applied, e.g., when studying the performances of microarray segmentation algorithms [23]. For this study, we selected two different error models implemented in the simulator: the hierarchical error model (Hem) [24] and the Rocke error model (Rocke) [25]. First, the Hem model is defined as:

$$y = x + g_i + c_j + r_{ij} + b_{ijk} + \varepsilon \quad (10)$$

where y is the expression measurement in Log scale from the microarray, x the noise-free expression level, c_j the gene-specific noise, r_{ij} the gene and chip specific noise, b_{ijk} the gene, chip and biological

sample noise, and ε is an independent random noise component. Rocke model is defined as:

$$y = \alpha + xe^{n} + \varepsilon \quad (11)$$

where α is the background noise, n the multiplicative noise term, and ε is an independent random noise component. A more detailed description of the error models can be found in [4].

Since we are interested in the effect of noise in inference efficiency, we increased the noise level gradually starting from a noise-free signal. Instead of only modifying the parameters, we made repeated noise simulations until we achieved a desired noise to signal ratio (NSR). It is far more informative to control the actual noise level, than just modifying parameters without knowing the true effect on the underlying signal. The NSR is defined as a ratio between the power of noise and the power of signal:

$$NSR = \frac{P_{noise}}{P_{signal}} = \frac{\sum n(t)^2}{\sum s(t)^2} \quad (12)$$

where *n(t)* is the simulated noise and *s(t)* the noise free signal. First, we initialized the error models in such a way that both models produce noise with 0.1 noise ratio level. Thereafter, we generated synthetic microarray experiments using both models. We repeated the process by increasing the noise ratio in steps of 0.1. In the Hem model, we increased the noise by modifying the standard deviation of the experimental error term, whereas in the Rocke model we controlled the noise with the standard deviation of the multiplicative noise component. Fig. 2 illustrates the effect of both error models on a noise free signal. Next, both the noise perturbed and initial signal's time series are binarized.

Given the original time series and those affected with experimental noise, we cluster the time series of the number of proteins of each gene into binary clusters, representing "on" and "off'" states [11], for the purpose of inferring the topology and logic of interactions between genes. The k-means clustering algorithm [29] detects these variations by separating the observed quantities in time in two partitions, "high" and "low". Since we want to correlate variations in the protein quantities of inputs with variations of protein quantities of outputs, and because depending on the interaction strength, large variations of inputs can be associated with small variation of outputs (and vice versa), we opted to apply k-means to the time series of each gene independently. One problem when using k-means occurs when a few large outliers exist. For example, in Fig. 2, one large outlier is observed due to experimental noise introduction. On first inspection, one would expect that such an outlier would shift the binarization process towards having more null values than the original time series. Indeed, the binarized time series, after the Rocke noise model was applied, had less 1's than the original one (the original time series binarized has fourteen 1's out of 41 data points while the time series affect by the Rocke model has only eleven 1's). Less expected was the fact that the time series affected by the Hem noise model has only four 1's. Here, the application of the Hem noise model caused a few outliers (Fig. 2), but also some cases where the resulting value was smaller than the original one. This compensated for the outliers and led to the "threshold" between the "on" and "off" classification being lower than in the Rocke case. This only occurred in our examples for large values of NSR (> 0.7). If the reason for information loss due to clustering is only the existence of outliers, it might prove better to use clustering algorithms more robust to outliers than k-means (e.g. k-medoids). In the cases analyzed here, such large outliers were not commonly observed and, as seen, not the main cause of information loss. Thus, we used k-means.

From the binarized time series IADGRN infers the network. The result is compared to the real topology, and efficiency of inference is measured.

As said, to infer from time series of gene expression we used IADGRN [5]. This algorithm is based on ARACNe [14], but is applicable to time series. IADGRN infers the network topology and logic from state transitions of gene expression. These can either be independent state transitions, state transitions from perturbed states of an initial state, or single consecutive time series. Here, only the last case is used. From the time series, the pI is calculated between all pairs of genes (Equation 13), relating the states of a possible input at moment t with the state at $t+1$ of a possible output:

$$I(A_t, B_{t+1}) = H(A_t) + H(B_{t+1}) - H(A_t, B_{t+1}) \quad (13)$$

where $H(x)$ is the entropy of the binary sequence x. If each value of x has a probability p_0 of being equal to 0, and p_1 of being equal to 1, $H(x)$ is given by: $H(x) = -p_0 log(p_0) - p_1 log(p_1)$. The calculated pI must then be compared to a threshold to decide if the correlation is high enough to accept that a directed connection between the two genes exists.

One advantage of using mutual information [14] to infer is that it is a procedure that is far less computationally complex than others because the number of necessary calculations is optimal. Furthermore, algorithms based on mutual information ranking have been shown to be resilient to estimation errors [14].

Here, because we infer from a "continuous" time series and not independent states transitions, and therefore correlations build up in the network, a correct estimation of the spurious correlations is not known. Thus, we assign a threshold value that maximizes inference. The threshold is set at the minimum value necessary for removing all false correlations, leaving only true connections such that these are all inferred. When this is not possible, the threshold is set such that a few false positives are accepted, in order to detect the largest possible number of true connections. This value can vary significantly from network to network, due to sampling time, the system stochasticity and gene correlations due to the topology. Thereafter, IADGRN applies a time dependent data processing inequality test (DPI) [14] to remove indirect correlations. If connections from A to B, B to C, and A to C are inferred, it compares (according to (14)) the pI of each of the three connections and removes the connection from A to C if its pI is lower then either of the other two. A small tolerance τ = 0.05 is introduced to prevent close pI values from being pruned [14].

$$I(A_t,C_{t+1}) \leq min[I(A_t, C_{t+1}), I(A_t, C_{t+1})].(1-\tau) \quad (14)$$

Notice that (14) assumes one has access to the real states of A, B and C. Instead, here we observe states affected by external noise. This can cause some errors. Another problem is that DPI assumes that no other path of influences exist between A and C, than the one passing via B. This might not be true in a network. Regardless, in all numerical simulations the DPI exhibited high efficiency and very few errors (< 3% error rate). Inference accuracy is measured by Recall (15), the fraction of all existing connections that are inferred, and Precision (16), the fraction of inferred connections that were correct. If TP is true positives (correctly inferred connections), TN is the number of true negatives (correctly assigned as non-existing connection), FP is the number of false positives (wrongly assigned connections), and FN is the number of false negatives (connections existing but not detected):

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

Since inputs can act as activators or repressors, we use the Pearson correlation between input and output time series (not binarized) of identified connections to determine whether the input acts as an activator or repressor. Notice that this measure is not used to determine the existence of connections.

3 Results

We begin by analyzing the influence of noise in the simplest cases possible, namely, direct and indirect repression of one gene by another, and direct and indirect activation of one gene by another. This is done to "tune" the sampling times to maximize inference capability and to study the effect of microarray noise on cases where, without external noise, inference is complete and without errors.

The protein that acts as a direct input binds to the promoter region of the output gene. The effect is either inhibiting or activating transcription. If the interaction is indirect, the protein of the input gene either binds to a repressor molecule stopping its effect (indirect activation) or it degrades the protein of the output gene (indirect repression).

To create a signal with non-null entropy, the input gene's protein concentration must be externally driven to vary in time. This variation causes a correlated variation in the output gene's protein concentration, allowing inference of the connection.

In Table 1, we present the results of the measurements of pI between pairs of genes from the time series of their protein concentrations. The results are an average of 1,000 independent experiments of the four different cases.

The effect of noise from both noise models is similar. In both, the correlation between input and output genes decreases as noise increases. However, for low NSR values, the Hem model causes a higher decrease in correlations.

Also visible from Table 1 is the effect of introducing array noise in the time series on inference efficiency. Beyond 0.5 NSR no inference is possible, due to the disruption of the correlation between input and output genes' protein time series.

Interestingly, no self-inputs were identified as FPs for any NSR values in both noise models.

In general, even for low NSR values, indirect interactions are harder to detect than direct ones, since its pI is only slightly above the noise level. Any noise introduced, either using the Hem or Rocke noise models, was enough to destroy the correlation, and no connection was inferred.

The Pearson correlation detected, with 90% accuracy, if inferred connections are activating or repressing. Between genes indirectly connected this measure is small. Given no noise, it is equal to 0.3 (in absolute value) while for direct connections is 0.8. This measurement, in absolute value, is the same whether or not it is an activation or repression.

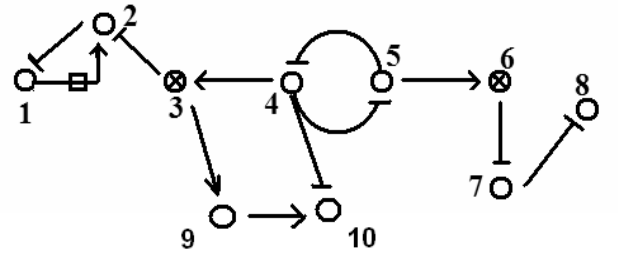

Fig. 1. A GRN. A circle with a cross inside represents genes without basic expression, as opposite to those with. An arrow represents activation, and lines ending with a perpendicular line represent repression. A square in the middle of a line indicates an indirect interaction.

Hem Noise Model				
NSR	pI of TP	pI of FP	Recall	Precision
0.0	0.450	0.175	1.000	1.000
0.1	0.341	0.175	0.720	0.734
0.2	0.286	0.216	0.700	0.638
0.3	0.271	0.210	0.600	0.525
0.4	0.175	0.203	0.025	0.013
0.5	0.142	0.167	0	0
0.6	0.118	0.132	0	0
0.7	0.097	0.112	0	0
0.8	0.063	0.119	0	0
0.9	0.081	0.098	0	0
1.0	0.054	0.064	0	0
Rocke Noise Model				
NSR	pI of TP	pI of FP	Recall	Precision
0.0	0.450	0.175	1.000	1.000
0.1	0.373	0.192	0.950	0.900
0.2	0.286	0.201	0.700	0.600
0.3	0.192	0.186	0.250	0.125
0.4	0.178	0.178	0.200	0.063
0.5	0.204	0.181	0.071	0.063
0.6	0.186	0.172	0	0
0.7	0.153	0.146	0	0
0.8	0.121	0.133	0	0
0.9	0.089	0.098	0	0
1.0	0.073	0.055	0	0

Table 1. pI between input and output gene (TP) and between output and input gene (FP). Threshold set at 0.2. For each NSR value and type of interaction, 1,000 experiments were conducted. The efficiency of inference is measured by Recall and Precision.

Next, we simulate GRNs of 10 genes with various interaction strength values between the genes (rate constants of reactions (4) and (8) were varied from 0.1 to 1, in steps of 0.1). Connectivity was also varied from 1 to 3 (in steps of 0.2). The GRNs' topology is random in all cases. For each data point, 100 networks were generated and the results averaged.

An example of such a randomly generated networks is shown in Fig. 1. From the topology, *SGNSim* [18] defines the corresponding set of chemical reactions from which the dynamics are simulated.

The system state (quantity of proteins at a given t) was sampled each 5,000 s for 100,000 s. Array noise is introduced using the Hem and Rocke noise models. An example of the effects of artificial noise in a time series is shown in Fig. 2. Correlations are more disrupted at high protein concentrations.

The average pI for a given NSR of the networks generated as described is shown in Fig. 3. The effects of noise are very similar to those previously described for 2 gene interactions.

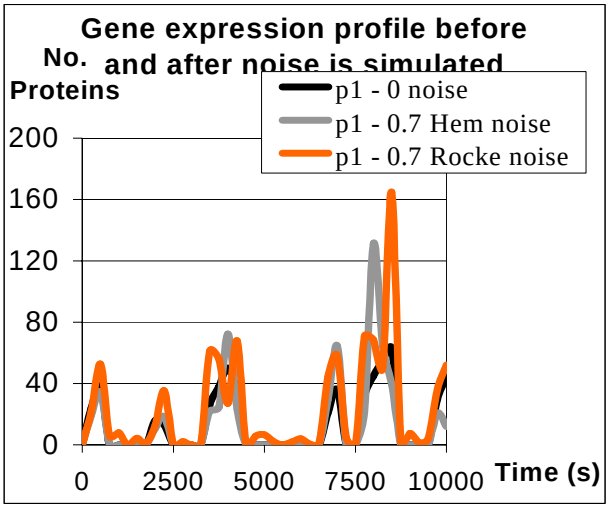

Fig. 2. Time series of an original protein level signal, as well as the same signal affected by both Hem and Rocke noise models, for the case of direct activation.

However, the decrease in pI is now caused not only by the array noise, but also by the indirect correlations between all genes' time series that build up as the network's dynamics evolve. Thus, as expected, inference becomes more difficult as the network size increases (values of pI for these networks is far smaller than those in Table 1). Even for small NSR, inference is not possible. Beyond 0.3 array noise, the average pI is below 0.1 (the threshold value for accepting a connection) making Recall and Precision null.

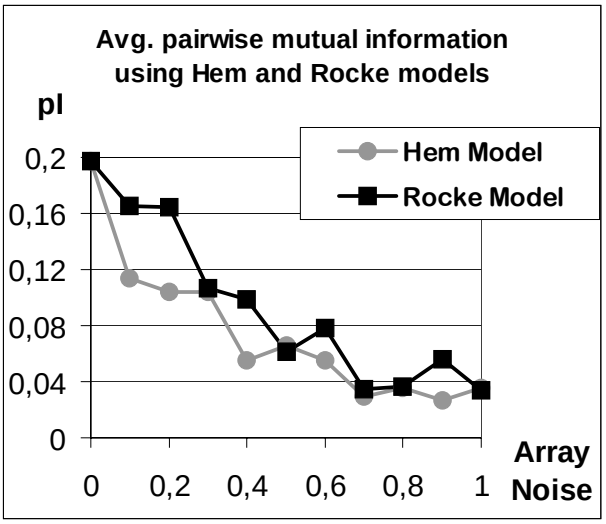

Fig. 3. Network pI, average of 100 independent experiments starting from the same initial state, as the array noise varies, for Hem and Rocke noise models.

The results in Fig. 3 show why due to high correlations between all genes' protein levels in the network, a high threshold must be used, otherwise too many non-existing connections would be incorrectly inferred. We found the best threshold to be 0.36, leading with no noise, to a Recall of 0.2 and Precision of 0.3. When noise was introduced, using either the Hem or Rocke noise models, not a single connection was inferred (above 0.2 NSR). Importantly, no false positives appeared.

It is visible that the pI between genes drops in the Hem model, faster for small values of noise, than in the Rocke model. However, importantly, for more than 40% noise ratio, both noise models caused the signals to become completely uncorrelated.

A higher sampling rate did not improve the inference capability (data not shown), since most of the connected genes are as highly correlated as the non-connected ones, after a very short transient. Removing FPs due to higher sampling frequency also resulted in not inferring TPs.

4 Conclusion

The combined use of a realistic model of GRNs and microarray noise allows efficient evaluation of inference algorithms and provides clues towards developing inference and noise reduction methods, otherwise not possible, since in real experiments noise free signals are so far unattainable and the underlying network structure is unknown. This work is a first look at the ability of a known inference method to infer realistic models of GRNs, affected by noise with realistic characteristics.

Inference of the GRN structure from the dynamics observed in microarrays, might be crucial, for example in the development differentiation therapy to the treatment of cancer, as understanding what changes as the cell progresses from normal to cancerous is essential to provide clues on how to intervene. This has been attempted using other models of GRNs, but have not been successful yet, perhaps because these models are far from the real system dynamics. The use of more realistic models not only demonstrates the necessity of further study, but also may allow for the development of new inference methods using previously unaccounted features of GRN dynamics, such as asynchrony and gene expression level quantization.

References:

[1] AS. Ribeiro, R. Zhu, SA. Kauffman, A General Model for Gene Regulatory Networks with Stochastic Dynamics, *WSEAS International Conference on Mathematical Biology and Ecology*, Miami, USA, Jan. 18-20 2006.

[2] AS. Ribeiro, R. Zhu, SA. Kauffman, A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics, *Journal of Computational Biology*, Vol.13, No.9, 2006, pp. 1630-1639.

[3] AS. Ribeiro, R. Zhu, SA. Kauffman, A General Model for Gene Regulatory Networks with Stochastic Dynamics, *WSEAS Transactions on Biology and Biomedicine*, Vol.3, No.3, 2006, pp. 261-264.

[4] M. Nykter, T. Aho, M. Ahdesmäki, P. Ruusuvuori, A. Lehmussola, O. Yli-Harja, Simulation of microarray data with realistic characteristics, *BMC Bioinformatics*, Vol.7, 2006, pp. 349.

[5] SA. Kauffman, AS. Ribeiro, J. Lloyd-Price, Inferring Dynamical Gene Regulatory Networks Subject to Noise, *3rd Annual Rocky Mountain Bioinformatics Meeting*, Aspen/Snowmass, USA, Dec. 9-11, 2005.

[6] AS. Ribeiro, RA. Este, J. Lloyd-Price, SA. Kauffman, Measuring Information Propagation and Retention in Boolean Networks and its Implications to a Model of Human Organizations, *6th WSEAS International Conference on Simulation, Modeling and Optimization*, Lisbon, Portugal, Sept. 22-24, 2006.

[7] AS. Ribeiro, RA. Este, J. Lloyd-Price, SA. Kauffman, Measuring Information Propagation and Retention in Boolean Networks and its

Implications to a Model of Human Organizations, *WSEAS Transactions on Systems*, Vol.12, No.5, 2006, pp. 2935-2941.

[8] CV. Forst, Network genomics - a novel approach for the analysis of biological systems in the post-genomic era, *Molecular Biology Reports*, Vol.29, No.3, 2002, pp. 265-80.

[9] J. Audet, Stem cell bioengineering for regenerative medicine, *Expert Opinion on Biological Therapy*, Vol.4, No.5, 2004, pp. 631-644.

[10] S. Sell, Stem cell origin of cancer and differentiation therapy, *Critical Reviews in Oncology/Hematology*, Vol.51, No.1, 2004, pp. 1-28.

[11] SA. Kauffman, Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets, *Journal of Theoretical Biology*, Vol.22, No.3, 1969, pp. 437-467.

[12] S. Liang, S. Fuhrman, R. Somogyi, REVEAL: a general reverse engineering algorithm for inference of genetic network architectures, *Pacific Symposium on Biocomputing.*, 1998, pp. 18-29.

[13] AJ. Hartemink, Reverse Engineering Gene Regulatory Networks, *Nature Biotechnology*, Vol. 235, 2005, pp. 554-555.

[14] A. Margolin, C. Nemenman, K. Basso, U. Klein, C. Wiggins, G. Stolovitzky, R. Favera, A. Califano, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, *BMC Bioinformatics*, Vol.7, Suppl. 1, 2006.

[15] Y. Tu, G. Stolovitzky, U. Klein, Qualitative noise analysis for gene expression microarray experiments, *PNAS*, Vol.99, No.22, 2002, pp. 14031-14036.

[16] P. Mendes, W. Sha, K. Ye, Artificial gene networks for objective comparison of analysis algorithms, *Bioinformatics*, Vol.19, Suppl. 2, 2003, pp. II122-II129.

[17] M. Samoilov, G. Price, A. Arkin, From Fluctuations to Phenotypes: The Physiology of Noise, *Science STKE*, No.366, 2006.

[18] AS. Ribeiro, J. Lloyd-Price, SGN Sim, a Stochastic Genetic Networks Simulator, *Bioinformatics*, Vol.23, No.6, 2007, pp. 777-779.

[19] DT. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of Computational Physics*, Vol.22, No.4, 1976, pp. 403-434.

[20] E. Gaffney, N. Monk, Gene expression time delays and turing pattern formation systems, *Bulletin of Mathematical Biology*, Vol.68, No.1, 2006, pp. 99-130.

[21] M. Roussel, R. Zhu, Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression, *Physical Biology*, Vol.3, 2006, pp. 274-284.

[22] MA. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, *Journal of Physical Chemistry A*, Vol.104, No.9, 2000, pp. 1876-1889.

[23] A. Lehmussola, P. Ruusuvuori, O. Yli-Harja, Evaluating the performance of microarray segmentation algorithms, *Bioinformatics*, Vol.22, No.23, 2006, pp. 2910-2917.

[24] H. Cho, J. Lee, Bayesian hierarchical error model for analysis of gene expression data, *Bioinformatics*, Vol.20, No.13, 2004, pp. 2016–2025.

[25] DM. Rocke, B. Durbin, A model for measurement error for gene expression array, *Journal of Computational Biolology*, Vol. 8, 2001, pp. 557–569.

[26] Y. Tu, G. Stolovitzky, U. Klein, Quantitative noise analysis for gene expression microarray experiments, *PNAS*, Vol.99, No.22, 2002, pp. 14031-14036.

[27] AS. Ribeiro, D. Charlebois, J. Lloyd-Price, SA. Kauffman, IADGRN: Inferring gene regulatory networks from time series of genes activity. Increasing the scope of efficiency to more general interaction functions between genes and more complex time series. *6th International Conference of the Canadian Proteomics Initiative*, May 10-12, Edmonton, Canada, 2006.

[28] V. Vyshemirsky, A software tool for Bayesian inference of ODE Biochemical Models, Practical Inference Methods for Mechanistic Modelling of Biological Systems, PIM07, Glasgow, 2007.

[29] JB. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, *5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, 1967, pp. 281-297.