

Inferring Transcriptional Regulatory Relationships Among Genes in Breast Cancer: An Application of Bayes' Theorem

Emmanuel S. Adabor¹, George K. Acquah-Mensah² & Francis T. Oduro¹

¹ Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

² Pharmaceutical Sciences Department, Massachusetts College of Pharmacy and Health Sciences (MCPHS University), USA

Correspondence: Emmanuel S. Adabor, Department of Mathematics, Kwame Nkrumah University of Science and Technology, PMB, Kumasi, Ghana. Tel: 233-28-504-2519. E-mail: healmes@gmail.com

Received: January 9, 2014 Accepted: March 11, 2014 Online Published: March 25, 2014

doi:10.5539/ijsp.v3n2p52 URL: <http://dx.doi.org/10.5539/ijsp.v3n2p52>

The research is supported by Novartis Institute for Biomedical Research/Ghana Biomedical Research Network fellowship

Abstract

The introduction of Deoxyribonucleic acid (DNA) microarray technologies provides a means of measuring the expression of thousands of genes simultaneously. It has generally sought to revolutionize biological research by significantly elucidating biological processes. Gene networks may be inferred from such microarray data. Bayes' theorem, in this work is applied to the problem of inferring new transcriptional regulatory relationships among gene products in Breast Cancer. A compendium of human breast epithelial cell probe level microarray data from the Gene Expression Omnibus (GEO) repository was subjected to the Robust Multiarray Average (RMA) procedure for normalization and background correction. A subset of the resulting expression matrix consisting of the expression values of only relevant probe-set identifiers (IDs) representing the genes of interest in the data were extracted with a LISP code. This subset was supplied to a Bayesian Network inference learning algorithm to unearth new regulatory relationships from the data. Variations in parameters of the learning algorithms resulted in the prediction of at least 10 new relationships among genes in breast cancer. Among these were the direct regulatory signaling relationship between S-phase kinase associated protein 2 (SKP2) and the Cell division cycle 25A (CDC25A) and that between the cyclin-dependent kinases regulatory subunit 1 (CKS1B / CDC28) and E2F transcription factor 3 (E2F3). The identified causal networks are potentially useful for understanding complex drug actions and dysfunctional signaling in breast cancer.

Keywords: DNA microarray, inferring, gene network, bayesian network, normalization

1. Introduction

The process of inferring gene regulatory networks from gene expression data has brought on the need for experts and analytical scientists to develop several algorithms that seek to model cellular networks, protein signaling pathways and genetic data analysis for biological research (Friedman, 2004; Sachs, Perez, Pe'er, Lauffenburger, & Nolan, 2005; Beaumont & Rannala, 2004). The biomedical literature presents a plethora of algorithms that have been proposed to solve the problem of inferring networks. Clustering algorithms are used to visualize and analyze expression data whereby genes with similar expression profiles are grouped into clusters (Eisen, Spellman, Brown, & Botstein, 1998). Co-expressed genes are viewed as being functionally linked to each other at a higher chance.

At the other end of the spectrum of algorithms are the information theoretic approaches. This approach sought to correct the failure of clustering approach to capture more complex statistical dependencies between expression patterns by proposing the Mutual Information (MI) measure. The MI among all pairs of genes are computed, and inference of interaction is established by comparing it to a set threshold value (Butte & Kohane, 2000). However, the MI and the Pearson correlation may yield almost similar results (Stuer et al., 2002). The networks inferred from information theoretic approaches represent statistical dependencies but not necessarily direct causal interactions among genes. The Context Likelihood of Relatedness (CLR) algorithm, Algorithm for the Reconstruction

of Accurate Cellular Networks (ARACNE) and Minimum Redundancy NETWORK (MRNET) algorithms provide refinements to the MI and can distinguish between direct and indirect interaction networks (Faith, Hayete, Thaden, Mogno, & Wierzbowski, 2007; Margolin, Wang, Lim, Kustagi, & Nemenman, 2006; Meyer, Kontos, Lafitte, & Bontempi, 2007).

A non-probabilistic inferring algorithm is the Ordinary Differential Equation (ODE) approach (Gardner, di Bernardo, Lorenz, & Collins, 2003). This deterministic approach models a system of differential equations with each equation describing a gene as a function of other genes. Networks derived from the ODE-based technique show causal interactions. However a large number of experimental data are necessary for good performance. ODE-based algorithms such as Network identification by multiple regression and microarray network identification have been proposed in the literature (Gardner, di Bernardo, Lorenz, & Collins, 2003; Di Bernardo et al., 2005). Probabilistic graphical models such as Bayesian Networks and Gaussian graphical models have also been proposed to solve the problem of inferring gene regulatory networks from microarray data (Friedman, 2004; Pe'er, Nachman, Linial, & Friedman, 2008). The Gaussian graphical model assumes that gene expression values are normal variates within undirected networks. A robust approach to estimate the Gaussian graph for a high dimensional data is further presented in the literature (Ambroise, Chiquet, & Matias, 2009). The undirected nature of Gaussian graphical model makes it undesirable when it competes with other approaches such as Gene Network Inference with Ensemble of trees (Genie3) that predicts direct causal interaction but assumes availability of high dimensional data (Huyhn-Thu, Irrthum, Wehenkel, & Geurts, 2010).

In this work, a network inference algorithm based on the Bayes' theorem, called Bayesian Network, is used to infer transcriptional regulatory relationships among genes in breast cancer. A Bayesian Network (BN) is a probabilistic graph model describing the multivariate probability distribution for a set of variables. Due to their probabilistic nature, BNs can cope with the noise present in microarray measurements, while their graphical nature makes it easy to convey linear and higher order dependencies between genes. The BN models direct causal relationships as a direct acyclic graph (DAG) when the Causal Markov Assumption holds. This assumption can be stated as: a variable X is independent of every other variable conditional on all its causes (Bansal, Belcastro, Ambesi-Impiobato, & Bernardo, 2007). This work further seeks to establish optimal settings for building a Bayesian Network using actual breast cancer data. BN algorithms are furthermore suited for the inferring since they can be extended to dynamic Bayesian networks (DBN) to model time series and data that incorporates feedback (Friedman, Murphy, & Russell, 1998).

2. Method

2.1 Bayesian Networks

A BN, also called Belief Network, is an augmented directed acyclic graph (DAG) represented by the nodes or variables, X_i for $i = 1, 2, \dots, n$ connected by directed edges which denotes relationships among the variables. The nodes denote the genes in a gene network. Figure 1 is an example of a BN describing a gene regulatory network.

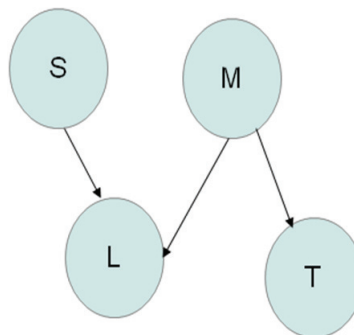


Figure 1. Example of BN

T is regulated by M and L is regulated by S and M. T is conditionally independent of L whereas M and S are independent. Furthermore, M is said to be a parent of T and both S and M are parents of L.

The relationships between variables are described by a joint probability distribution, $P(X_1, X_2, \dots, X_n)$, expressed in terms of the conditional probabilities. Given that the Markov assumption holds, then the resulting $P(X_1, X_2, \dots, X_n)$

is given by Equation (1).

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i)) \quad (1)$$

where $\text{parents}(X_i)$ denotes the set of parents for each variable X_i . This rule is based on the Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2)$$

2.2 Learning BN Structure

The BN procedure involves finding a DAG, G , which best fits the dataset, D . Algorithms for learning BN consist of two components namely a scoring metric and a search technique. The Scoring metric computes how well a proposed network fits the data whereas a search procedure tries to identify network structures with high scores traversing networks.

2.2.1 Scoring Metric

In general, the Bayesian scoring metric that evaluates the network is described as

$$\text{Score}(G : D) = \log P(G|D) = \log P(D|G) + \log P(G) - \log P(D) \quad (3)$$

where D is an assumed multinomial sample. Conditioning the network parameters with Q , then $P(D|G)$ is expressed as

$$P(D|G) = \int P(D|G, Q)P(Q|G)dQ. \quad (4)$$

With these bases, the Bayesian Dirichlet equivalence metric (BDe) was developed to avoid overfitting the data (see Buntine, 1991; Heckerman, Geiger, & Chickering, 1995; Cooper & Herskovits, 1992 for details). The BDe assumes event equivalence property which implies that the scores of two isomorphic belief networks are equal (score equivalence). The sum up of the likelihood equivalence assumption is: the data does not discriminate equivalent structures. Furthermore, the BDe metric requires the prior over parameters $P(Q|G)$ to have Dirichlet prior distributions and a uniform prior $P(G)$. The BDe is used in this work since it discriminates the simple and complex structures just as the Occam's Razor at work.

2.2.2 Searcher Methods

The search space for Bayesian network structures is the set of all possible DAGs that fits the data set. The number of all possible static DAGs given the number of nodes as determined by Robinson (1977) is given in Equation (5):

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) \quad (5)$$

for $n > 2$ where $f(0) = 1$ and $f(1) = 1$.

Searching for all possible networks and generating high scoring networks is an NP-hard problem and therefore heuristic methods have been proposed (Chickering, 1996). Generally, heuristic methods perform any of the following at each iteration: an existing link between variables can be reversed or removed, or new links may be introduced to connect variables that were not connected previously while ensuring an acyclic network. In this work two searchers, namely: Greedy Hill Climbing and Simulated Annealing are used to learn the BN of the breast cancer data.

2.2.3 Greedy Hill Climbing Search

This method initiates with an initial structure also called prior network, considers all possible changes to it and the resulting network (neighbour) with the highest score is then used to initiate another search. If no neighbour has a score higher than current structure, then the algorithm stops with the current structure as optimal. The greedy search may fail to reach the global optimal network as it always stops on a local optimum. To circumvent this limitation of the method, it is advisable to repeat the process several times with different random initial networks and choosing the best local optimum. In this case, the maximum number of restarts or specified period may be set for the search to terminate.

2.2.4 Simulated Annealing Search

This method starts with an initial network, accepts changes to existing network and improves the current network if its score is higher or if its score is lower but with a probability based on a system parameter, T , known as “the temperature”. At the start of the process, a lot of changes are accepted but reduces as the value for T is lowered gradually. This allows enough exploration of the search space so that the problem of terminating at a local optimum is solved. The mechanism of the simulated annealing search is further explored by Heckerman (1995) and Janzura and Nielson (2006).

2.3 Inferring Networks From Data

The process of learning the Bayesian structure from the breast cancer data is shown in Figure 2. A compendium of microarray data was generated from the Gene Expression Omnibus (GEO) record numbers GDS3716, GDS3139 and GSE21947. The GDS3716 data sets consist of data obtained from samples of reduction mammoplasty breast epithelium of patients, samples of histologically normal breast epithelium from prophylactic mastectomy patients, samples of histologically normal breast epithelium from Estrogen Receptor - Negative (ER-) breast cancer patients, and samples of histologically normal breast epithelium from Estrogen receptor - positive (ER+) breast cancer patients. The GSE21947 data sets consist of data obtained from gene expression patterns in histologically normal breast epithelium of breasts containing estrogen receptor positive (ER+) and estrogen receptor negative (ER-) in cancer patients. The GDS3139 data set included samples obtained from epithelium adjacent to a breast tumor and patients undergoing reduction mammoplasty without apparent breast cancer. The data set on breast cancer will enable the results to have direct interpretation in relation to cancer.

The resulting compendium contains the gene expression microarray experiment data from human breast epithelial cells under a variety of conditions. These arrays are based on the Affymetrix Human Genome U133A platforms. Probe level data derived from microarray experiments may harbour obscuring variations (Hartemink, Gifford, Jaakola, & Young, 2001). In all there were 22,283 probe sets and 89 arrays from the human genome U133A after the .CEL files format were downloaded and subjected to Robust Multi-array Average for normalization and background correction (Irizarry et al., 2003). This eliminates the effects of all obscuring variations in the data derived under different conditions. Expression data associated with the relevant probe set IDs representing the genes of interest were extracted with a LISP code. The previously known interactions of these genes in cell cycle were derived using reactome by Joshi-Tope et al. (2005) to set up an initial structure (prior) of 36 relationships (edges) in learning the static BN structures (See Table 1). The genes are also referred to as the variables or nodes of the network.

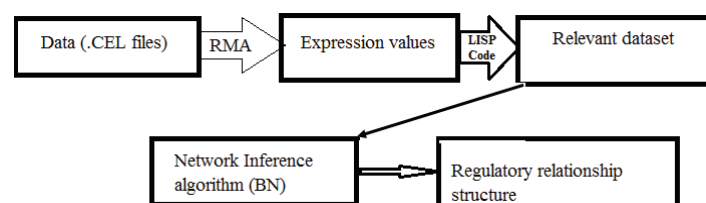


Figure 2. The procedure of inferring relationships from data

Probe level data is subjected to Robust Multi-array average (RMA) to make the data set ready for the Bayesian Network Inferring algorithms. The relevant data set of the genes of interest is extracted from the entire results of the RMA with a LISP code.

Table 1. Initial regulatory relationship

	Source	Target
1	ATM	MDM2
2	ATM	TP53
3	ATR	TP53
4	TP53	CDKN1A
5	MDM2	TP53
6	CHEK1	CDC25A
7	CHEK2	CDC25A
8	CDKN1A	CDK4
9	CDKN1A	CDK6
10	CDKN1B	CDK4
11	CDKN1B	CDK6
12	CDK7	CDK4
13	CDK7	CDK6
14	CDKN1B	CDK1
15	CDKN1B	CDK2
16	CDC25A	CDK2
17	SKP2	CDK1
18	SKP2	CDK2
19	CDKN1A	CDK1
20	CDKN1A	CDK2
21	CKS1B	CDK1
22	CKS1B	CDK2
23	SKP1	CDK1
24	SKP1	CDK2
25	CDC34	CDK1
26	CDC34	CDK2
27	RBX1	CDK1
28	RBX1	CDK2
29	CUL1	CDK1
30	CUL1	CDK2
31	CDK4	E2F1
32	CDK4	E2F2
33	CDK4	E2F3
34	CDK5	E2F1
35	CDK5	E2F2
36	CDK5	E2F3

Each variable at source node (second column) regulates its corresponding target variable (in the third column). The variables are denoted by gene symbols of genes or transcription factor.

3. Results

The results herein are presented in two parts based on the search techniques: Simulated Annealing and Greedy Search techniques. In implementing any of these methods, two different edge change methods namely, All-Local-Moves and the Random-Local-Moves were varied. All-Local-Moves involves composing a list of all available local moves (adding or subtracting of a parent, or the reversing of a parent relationship), given the current state of the network, and then selects the move that yields the highest scoring network. The disadvantage in this approach is that it is computationally expensive compared with experiments with the Random-Local-Move. On the other hand, Random-Local-Move selects a move at random from all possible local moves. In all these operations, the acyclic requirements of networks are satisfied. The computations to derive the networks are manually infeasible and so we implement these concepts of BN structure learning with Banjo (Sladeczek, Hartemink, & Robinson, 2008). All algorithms were made to search for best BN structures over 2 hours since Janzura and Nielson (2006) showed that increasing time for a simulated annealing search does not guarantee an improvement in its outcome. However the visualization of the optimal networks of predictor relationships was developed with Cytoscape (Shannon et al.,

2003).

3.1 Simulated Annealing With All-Local-Moves (SAA)

The optimal network showed 69 predictory relationships indicated by the edges as shown in details in Figure 3. The predicted relationships by SAA included the initial relationships. See Table 2 for details of other observations of results. This method predicted ATM regulated TP53 when initial structure with five dropped edges (including ATM regulates TP53) was used to learn BN structure from the data.

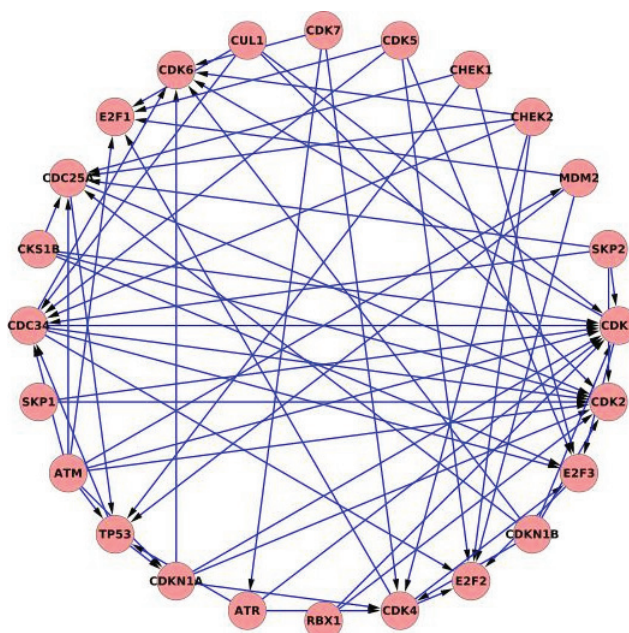


Figure 3. Network generated via simulated annealing with all-local-moves

The gene symbols are the nodes or variables of the network and the regulatory relationships are denoted by the directed edges. All the genes of interest were found to be present in the optimal network with 69 established regulatory relationships (directed edges). For instance, directed edge from CDC34 to E2F3 indicates that CDC34 regulates E2F3.

Table 2. Performance of BN with simulated annealing and greedy search methods

	SAA	SAR	GA	GR
No of Nodes	24	24	24	24
No. of edges	69	85	46	114
Additional edges	33	49	10	78
Scores	-2946.78	-2823.46	-3160.05	-2849.78
Edges recovered	1	2	-	2

The observations are derived from the Bayesian Network inference from a microarray dataset of 24 Variables/genes and 89 arrays. The Edges recovered represent the number of 5 excluded edges prior to Bayesian Network inference from data. SAA is Simulated Annealing with all-local-moves, SAR is Simulated Annealing with random-local-moves, GA is Greedy all-local-moves and GR is Greedy random-local-moves.

3.2 Simulated Annealing With Random-Local-Moves (SAR)

The optimal network comprised of 85 edges including the 36 initial relationships (edges) which is shown in details in Figure 4. See Table 2 for details of other results. SAR method predicted that ATM regulates TP53 and CDK4 regulates E2F3 when initial structure with four dropped edges (including ATM regulates TP53 and CDK4 regulates E2F3) was used to learn BN structure from the data.

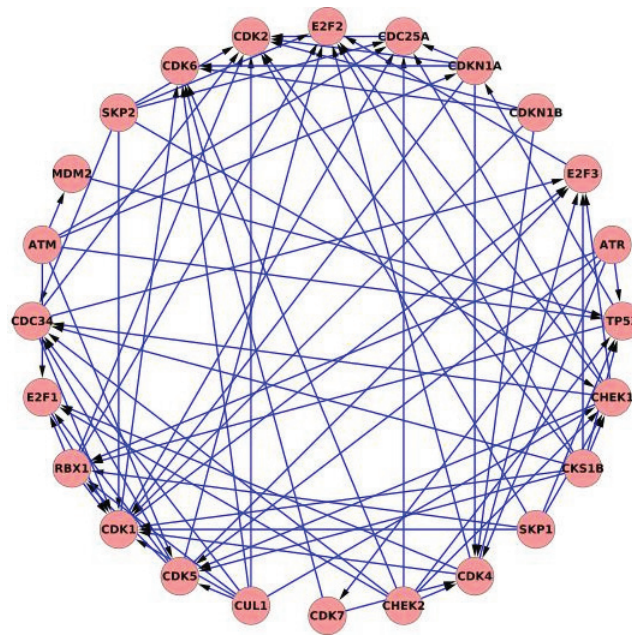


Figure 4. Network generated via simulated annealing with random-local-moves

The gene symbols are the nodes or variables of interest and the regulatory relationships denoted by the directed edges. For example, the gene with symbol TP53 regulates the gene with the symbol CDKN1A.

3.3 Greedy With All-Local-Moves (GA)

The BN structure learning with the Greedy search predicted 46 relationships (edges) in the revealed optimal network (see Figure 5). See Table 2 for details of other GA results. However, GA failed to predict any dropped relationship when five edges of the initial structure were dropped and the remaining structure used for the learning BN with GA.

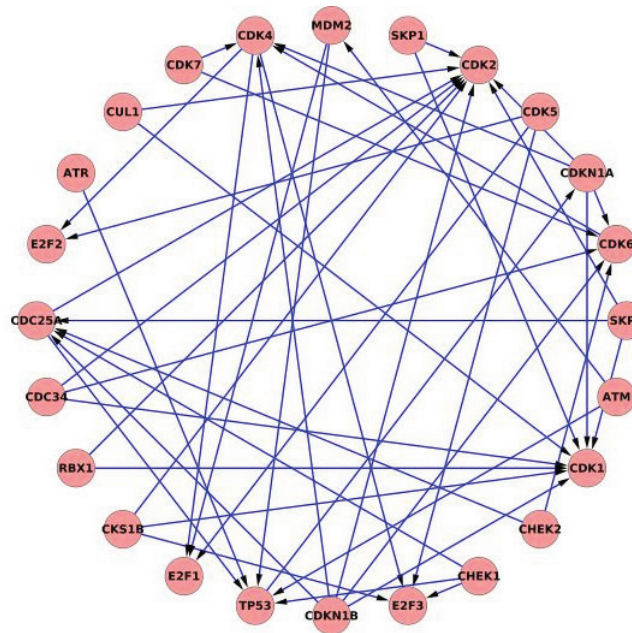


Figure 5. Network generated via greedy search with all-local-moves

The gene symbols are the nodes or variables of interest in the network and the directed edges represent the regulatory relationships. For example gene with the symbol CDKN1B regulates the gene with the symbol CDC25A.

3.4 Greedy With Random-Local-Moves (GR)

The highest number of predicted relationships including the initial relationships was found in the outcome of the BN algorithm with the GR. GR results showed 114 relationships (edges) in the optimal network with details presented in Figure 6. See Table 2 for other details of the GR results. GR recovered ATM regulates TP53, and CDK4 regulates E2F3 as well as CDK4 regulates CDKN1B when five edges of the initial structure (including CDKN1B regulates CDK4, ATM regulates TP53, and CDK4 regulates E2F3) were dropped and the remaining structure used for the learning the BN structure.

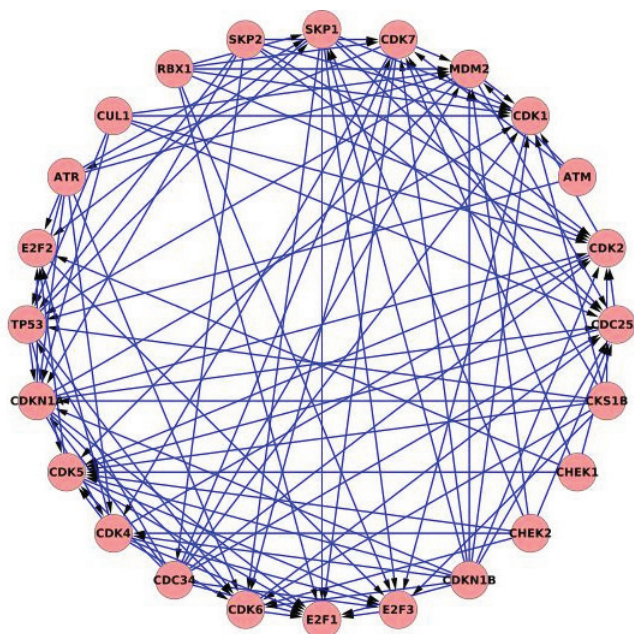


Figure 6. Network generated via greedy search with random local moves

The gene symbols are the nodes or variables of interest and the directed edges represent the regulatory relationships. For example gene with the symbol CDKN1B regulates the gene with the symbol CDC25A.

4. Discussion

All the methods repeated the initial regulatory structure in the optimal BN models. However, all methods predicted different additional edges owing to the different approaches in searching for optimal networks. GR had the highest prediction with 78 new relationships and a BDe score of -2849.78. SAR with Random-Local-moves had the next highest prediction of 49 new relationships to Greedy search with Random-local-moves. Though second in prediction, it had the best optimal network score of -2823.46 indicating that it performed a more exhaustive search to obtain the best network with the best score. SA with all local moves had an optimal network score of -2946.78 after it had predicted 33 new relationships whereas Greedy search with all local moves predicted 10 additional relationships with an optimal score of -3160.05. The scores indicate that BN inference with SAR produces better networks with high scores than the other approaches inferring BN models. Thus given microarray data set, BN inference with SAR will generally infer relationships that best fit the dataset than the other approaches. Furthermore, The GR and SAR being able to recover an equal number of dropped edges imply that they have equally higher chances of predicting true relationships.

Both SA with random local moves and Greedy search with random local moves predicted that ATM regulates TP53 and CDK4 regulates E2F3 when initial structure with five dropped edges (including ATM regulates TP53 and CDK4 regulates E2F3) was used as priors to learn BN. Since the optimal network is selected based on highest scores, it suggests that Simulated Annealing with random local moves will usually perform a more exhaustive search among the four methods used in this work. These results are contrary to the opinion of Chickering (1996) that greedy search outperforms simulated annealing in inferring BN structures. However, it affirms the assertion that there are no clear-cut superior approaches among the existing algorithms as each performs better under different conditions of data (Bansal, Belcastro, Ambesi-Impibato, & Bernardo, 2007). In all situations, the results of

various approaches may be subject to validation by literature and laboratory experimentation.

Results from all four methods above predicted that SKP2 regulates CDC25A and CKS1B regulates E2F3. These discoveries are supported by report in the biomedical literature (Hartman et al., 2009; Zimmermann et al., 2011). Results from SAR and GR showed 50 similar predicted relationships. Prominent among these is SKP2 regulates CDK2. This is supported by literature in Patsoukis et al. (2012). The fact that some of the consensus predicted relationships have been established in existing literature makes for greater confidence in all the other predictions. Figure 7 shows several other equally predicted relationships.

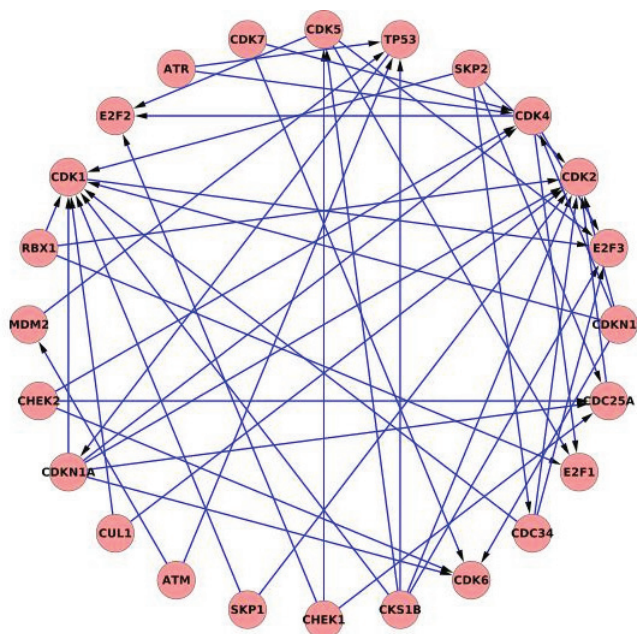


Figure 7. Consensus predictions of greedy and simulated annealing with random-local-moves

The gene symbols are nodes or variables of interest and the regulatory relationships are indicated by the directed edges as predicted by both the Greedy and the Simulated Annealing with random local moves. Common predictions by Bayesian Network structure learning with Greedy and the Simulated Annealing indicate the greater confidence in these prediction. In particular, MDM2 and CKS1B all directly regulate TP53.

The new predicted regulatory relationships (Figure 7) provide further insights into underlying biological processes leading to the breast cancer. Cancer is failure of controls over cellular births (through cell cycle) and deaths. Since the predicted relationships among the genes and transcription factors are involved in cell cycle, the suggested regulatory relationships are key components in elucidating the etiology of breast cancer and consequently identifying new therapies for the disease. In particular, CDC25A is required for progression from G1 to S-phase of the cell cycle as it activates the CDK2. Therefore the inferred novel indirect regulatory relationship between TP53 and CDC25A through CDKN1A is particularly interesting, suggesting the added role of TP53 in cell cycle progression in breast cancer. The predicted regulatory relationships between SKP2 and CDC34, and CDC25A further suggest a possible cause for the dysregulation of the cell cycle genes leading to tumorigenesis. Furthermore, observation of DNAs from different cancer patients generally contains the mutation or deletion of TP53 which identifies its role in tumor suppression. Therefore, the predicted relationships involving TP53, CKS1B and MDM2 are of particular importance since they affirm earlier work which suggested that inhibiting TP53-MDM2 interaction will be fascinating new anticancer agents necessary for activating some TP53 in tumors (Hamzehloie, Mojarrad, Hasanzadeh Nazarabadi, & Shekouhi, 2012).

5. Conclusion

Inferring BN with Simulated Annealing has the best performance based on scores compared to that with greedy search which is biased towards selecting only high scoring networks. In particular, Simulated Annealing with Random-Local-Moves and Greedy with Random-Local-Moves have equal true recovery rates as both recovered two of the dropped or excluded initial relationships (edges). Therefore Simulated Annealing with Random-Local-

Moves and Greedy with Random-Local-Moves are useful for effective inference with BN. Consensus predictions among four different search strategies make for greater confidence in the inference. A regulatory signaling relationship is predicted by consensus between S-phase kinase associated protein 2 and the Cell division cycle 25A. This interaction prediction occurs directly. Another regulatory signaling relationship is predicted by consensus between the CDC28 protein kinase regulatory subunit 1B (CDKN1B) and E2F transcription factor 3. This interaction prediction occurs directly. There is also a consensus predictions from Simulated Annealing with random local moves and greedy with random local moves: a regulatory signaling relationship is predicted by consensus between S-phase kinase associated protein 2 and the cyclin-dependent kinase 2. This interaction prediction occurs directly. That is BN structure learning methods can be valuable in harnessing signaling relationship studies in genes given a sample data.

References

- Ambroise, C., Chiquet, J., & Matias, C. (2009). Inferring sparse Gaussian graphical models with latent structure. *Electronic Journal of Statistics*, 3, 205-238.
- Bansal, M., Belcastro, V., Ambesi-Impibato, A., & Bernardo, D. (2007). How to Infer Gene Networks from expression profiles. *Molecular Systems Biology*, 3, 78. <http://dx.doi.org/10.1038/msb4100120>
- Beaumont, M. A., & Rannala, B. (2004). The Bayesian revolution in genetics. *Nat Rev Genet*, 5, 251-261.
- Buntine, W. (1991). Theory Refinement on Bayesian Networks. In B. D'Ambrosio, P. Smets, & P. Bonissone (Eds.), *Proceedings of the Seventh Annual Conference on Uncertainty in Artificial Intelligence* (pp. 52-60). Los Angeles, USA: University of California at Los Angeles.
- Butte, A. J., & Kohane, I. S. (2000). Mutual Information relevance networks: functional genomic clustering using pairwise entropy measurements. In R. Altman, K. Dunker, L. Hunter, K. Lauderdale, & T. Klein (Eds.), *Proceedings of Fifth Pacific Symposium on Biocomputing* (pp. 418-429). Honolulu, USA: Sheraton Waikiki.
- Chickering, D. M. (1996). Learning Equivalence Classes of Bayesian Network Structures. In E. Horvitz & F. V. Jensen (Eds.), *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence* (pp. 150-157). Oregon, USA: Reed College.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, 309-347.
- Di Bernardo, D., Thompson, M., Gardner, T., Chobot, S., Eastwood, E., Wojtovich, A., . . . , Collins, J. (2005). Chemogenomic profiling on genome-wide scale using reverse-engineered gene networks. *Nature Biotechnology*, 23, 377-383.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster Analysis and Display of genome-wide Expression Patterns. *PNAS*, 95(25), 14863-14868.
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., . . . , Gardner, T. S. (2007). Large-Scale Mapping and Validation of Escherichia Coli transcriptional regulation from a compendium of expression profiles. *PloS Comput Biology*, 5, e8. <http://dx.doi.org/10.1371/journal.pbio.0050008>
- Friedman, N. (2004). Inferring Cellular Networks using probabilistic graphical models. *Science*, 303, 799-805.
- Friedman, N., Murphy, K., & Russell, S. (1998). Learning the structure of dynamic probabilistic networks. In G. Cooper & S. Moral (Eds.), *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence* (pp. 139-147). Wisconsin, USA: University of Wisconsin Business School.
- Gardner, T., Di Bernardo, D., Lorenz, D., & Collins, J. (2003). Inferring Genetic Networks and identifying compound mode of Action via expression profiling. *Science*, 301, 102-105.
- Hamzehloie, T., Mojarad, M., Hasanzadeh Nazarabadi, M., & Shekouhi, S. (2012). The role of tumor protein 53 mutations in common human cancers and targeting the murine double minute 2-p53 interaction for cancer therapy. *Iran J Med Sci.*, 37(1), 3-8.
- Hartemink, A. J., Gifford, D. K., Jaakola, T. S. & Young, R. A. (2001). Maximum likelihood estimation of optimal scaling factors for expression array normalization. *SPIE BiOS*. Retrieved from <http://www.cgs.lcs.mit.edu/pubs/normabs.htm>.
- Hartman, T. R., Nicolas, E., Klein-Szanto, A., Al-Saleem, T., Cash, T. P., Simon, M. C., & Henske, E. P. (2009).

- The role of the Birt-Hogg-Dubé protein in mTOR activation and renal tumorigenesis. *Oncogene*, 28(13), 1594-1604.
- Heckerman, D. (1995). A Bayesian Approach to Learning Causal Networks. In P. Besnard, & S. Hanks (Eds.), *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 285-295). Montreal, Canada: McGill University.
- Heckerman, D., Geiger D., & Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20(3), 197-243.
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE* 5(9), e12776.
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, Normalization and Summaries of High Density Oligonucleotide Array Probe level Data. *Bio-statistics*, 4(2), 249-264.
- Janzura, M., & Nielson, J. (2006). A simulated Annealing-based method for Learning Bayesian Networks from Statistical Data. *International Journal of Intelligent Systems*, 21, 335-348.
- Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., . . . , Stein, L. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, 33(Database issue), D428-32.
- Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., & Nemenman, I. (2006). Reverse engineering cellular networks. *Nature Protocols*, 1, 663-672.
- Meyer, P. E., Kontos, K., Lafitte, F., & Bontempi, G. (2007). Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J Bioinform Syst Biol*, 2007, 79879. <http://dx.doi.org/10.1155/2007/79879>
- Patsoukis, N., Brown, J., Petkova, V., Liu, F., Li, L., & Boussiotis, A. V. (2012). Selective Effects of PD-1 on Akt and Ras Pathways Regulate Molecular Components of the Cell Cycle and Inhibit T Cell Proliferation. *Sci. Signal.*, 5(230), ra46. <http://dx.doi.org/10.1126/scisignal.2002796>
- Pe'er, D., Nachman, I., Linial, M., & Friedman, N. (2008). Using Bayesian Networks to analyze expression Data. *Journal of Computational Biology*, 7, 601-620.
- Sachs, K., Perez, O., Pe'er, D., Lauffenburger D. A., & Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308, 523-529.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., . . . , Ideker, T. (2003). Cytoscape: a software environment for integrated models of bimolecular interaction networks. *Genome Res*, 13, 2498-2504.
- Sladeczek, J., Hartemink, A. J., & Robinson, J. (2008). *Banjo*. Retrieved from <http://www.cs.duke.edu/~amink/software/banjo>
- Stuer, R., Kuths, J., Daub, C. O., Weise, J., Wang, J., & Selbig, J. (2002). The Mutual Information: Detecting and Evaluating Dependencies between variable. *Bioinformatics*, 18(2), 231-240.
- Zimmermann, C., Chymkowitz, P., Eldholm, V., Putnam, C. D., Lindvall, J. M., Omerzu, M., . . . , Enserink, J. M. (2011). A chemical-genetic screen to unravel the genetic network of DC28/CDK1 links ubiquitin and Rad6-Bre1 to cell cycle progression. *PNAS*, 108(46), 18748-18753.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).