www.roavs.com EISSN: 2223-0343
RESEARCH OPINIONS IN ANIMAL & VETERINARY SCIENCES

Research article

Simple gene regulatory network of immune system candidate genes in dairy cattle

Mostafa Ghaderi-Zefrehei[*1], Mustafa Muhaghegh Dolatabady[1] and EbrahimRowghani[2]

[1]Department of Animal science, School of Agriculture, YasoujUnivesity, Yasouj, Iran
[2]Department of Animal sciences, College of Agriculture, Darab Branch, Islamic Azad University, Darab, Iran

Article history
Received: 24 Dec, 2015
Revised: 28 Jan, 2016
Accepted: 3 Feb, 2016

Abstract

Given a sample space with many random variables defined on it, it makes lots of scene to think of some kind of association among them. Generally, sample space due to biological experiment is thought to create many random variable with biologically motivated inter or intra-relationship. In this study, our sample space contained mRNA profile of 6 immune system candidate genes: *Interfron Gamma (IFN-γ), Tomur Necrotic Factor alpha (TNF-α), Granolocyt-Macrofage Colony Stimulating Factor (GM-CSF), Interleukine-2 (IL-2), Interleukine-6 (IL-6) and Interleukine-8 (IL-8)* in which their transcriptome were obtained. These genes were due to dairy cattle in different biological states e.g. the early, middle, late of lactation and cows suffering from mastitis. Using Bayesian Network (BN), we concluded that first, many genes showed independent behaviour in terms of regulation (there were not any wired regulation among them over different biological states) second, the mode of regulation changed across different biological states. In mastitis state, *IL-8* shown to be regulator of *TNF-a*, *IL-8* was regulator of *GM-CSF* in early lactation, in middle of lactation, *IL-2* shown to be regulator of *IFN- γ* and in late of lactation *IL-8* turned out to be regulator of *GM-CSF*. This research revealed that the mode of regulation of candidate genes was not identical over different biological states. Overall, we showed that aforementioned immune system candidate genes should be seen in the biological context which they are functioning. Therefore, if the objective is to tackle with mastitis using drug targeting studies or in genetic selection, it is more relevant to pay close attention to *IL-8* as it is predicted to show the mode of regulator in mastitis state in dairy cattle. However, this gene was also the regulator of other different genes in across different biological states.

Keywords: Bayesian network; immune system; dairy cattle; gene regulatory network

To cite this article: Ghaderi-Zefrehei M, MM Dolatabady and E Rowghani, 2015. Simple gene regulatory network of immune system candidate genes in dairy cattle. Res. Opin. Anim. Vet. Sci., 5(12): 499-506.

Introduction

Bayes theorem is generally defined for both continuous and discrete random variables, though a mixed one could also be derived. Suppose we have two genes X and Y in which their expression profile have been measured in a random sample of cows. Based on general axiom of probability theory, we could say that x and y are independent (e.g. no regulation mode exists between them) if our calculation would end up as:

***Corresponding author:** Mostafa Ghaderi-Zefrehei, Department of Animal science, School of Agriculture, YasoujUnivesity, Yasouj, Iran;E-mail: mghaderi@yu.ac.ir

$f_{X,Y}(x \cap y) = f_Y(y) f_X(x)$. It means both of these two genes are probabilistically biologically independent. However, independency can hardly be seen in the context of sample space which is due to biological experiment. We may take a next step to find the probability that one gene would function conditional on the state of other gene (conditional probability). A tricky thing to keep in the mind is that we always condition on known phenomenon, a matter which most time people would have narrow ideas about it. For example:

$$f_{X/Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

is conditional probably in which we are conditioning the function of gene X based on the fact that gene Y is already functioning (know phenomenon). The value of $f_{X/Y}(x/y)$ would be $0 \le f_{X/Y}(x/y) \le f_X(x)$ therefore we would never expect the value of $f_{X/Y}(x/y)$ would be more than $f_X(x)$. We may get the $f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ in which now we are looking to find the functioning state of gene Y given gene X by doing very simple algebraic rearrangement:

$$f_{X/Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad f_{X/Y}(x/y) f_Y(y) = f_{X,Y}(x,y)$$

$$f_{Y/X}(y/x) f_X(x) = f_{X,Y}(x,y)$$

$$f_{X/Y}(x/y) = \frac{f_X(x) f_{Y/X}(y/x)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx} = \frac{f_{Y/X}(y/x) f_X(x)}{\int_{-\infty}^{\infty} f_X(u) f_{Y/X}(y/x)du}$$

Since $\int_{-\infty}^{\infty} f_X(u) f_{Y/X}(y/x)du$ shall end up to 1 (it is called normalizing constant) we could write up simply $f_{X/Y}(x/y) \approx \Upsilon f_{Y/X}(y/x) f_X(x)$ which is called Bayes theorem (law). This theorem shall play an immense role in area of probabilistic inferences especially in Bayesian network.

Understanding a Bayesian Network in a Discretized Form

Understanding BN in compact way using continuous random variables is a formidable task for non-computational people. We give the following topology of BN due to 4 discrete random variables (we write down $p(X)$ instead of $f(x)$). Figure 1 echoes all details about this system. By probability law, the prior information of all genes can be written as follows (ON means genes function and OFF means gene not function):

$$P(Y = ON) = P(Y = ON \mid X = ON)P(X = ON)$$
$$+P(Y = ON \mid X = OFF)P(X = OFF)$$
$$= (0.9)(0.4) + (0.8)(0.6) = 0.84$$
$$P(Y = ON) = P(Y = ON \mid X = ON)P(X = ON)$$
$$+P(Y = ON \mid X = OFF)P(X = OFF)$$
$$= (0.9)(0.4) + (0.8)(0.6) = 0.84$$
$$P(Z = ON) = P(Z = ON \mid Y = ON)P(Y = ON)$$
$$+P(Z = ON \mid Y = OFF)P(Y = OFF)$$
$$= (0.7)(0.84) + (0.4)(0.16) = 0.652$$
$$P(W = ON) = P(W = ON \mid Z = ON)P(Z = ON)$$
$$+P(W = ON \mid Z = OFF)P(Z = OFF)$$
$$= (0.5)(0.652) + (0.6)(0.384) = 0.5384$$

Which these results can be in the left part of Figure 1. By helps of Bayes law we can have:

$$P(Z = ON \mid W = ON) = \frac{P(W = ON \mid Z = ON)P(Z = ON)}{P(W = ON)}$$
$$= \frac{(0.5)(0.562)}{0.5384} = 0.6096$$
$$P(Y = ON \mid X = ON) = 0.9$$
$$P(Z = ON \mid X = ON) = P(Z = ON \mid Y = ON, X = ON)$$
$$P(Y = ON \mid X = ON) + P(Z = ON \mid Y = OFF, X = ON)$$
$$P(Y = OFF \mid X = ON) = P(Z = ON \mid Y = ON)P$$
$$(Y = ON \mid X = ON) + P(Z = ON \mid Y = OFF)$$
$$P(Y = OFF \mid X = ON) = (0.7)(0.9) + (0.4)(0.1) = 0.67$$
$$P(W = ON \mid X = ON) = P(W = ON \mid Z = ON, X = ON)$$
$$P(Z = ON \mid X = ON) + P(W = ON \mid Z = OFF, X = ON)$$
$$P(Z = OFF \mid X = ON) = P(W = ON \mid Z = ON)$$
$$P(Z = ON \mid X = ON) + P(W = ON \mid Z = OFF)$$
$$P(Z = OFF \mid X = ON) = (0.8)(0.67) + (0.6)(0.33) = 0.734$$

We could see that to compute the probability of gene Y is ON when gene W is also ON, we need to Bayes' theorem .e.g.

$$P(Y = ON \mid W = ON) = \frac{P(W = ON \mid Y = ON)P(Y = ON)}{P(W = ON)}$$

We cannot yet complete this computation because we do not know $P(W = ON \mid Y = ON)$. We can, obtain this value in the manner shown when we discussed downward propagation e.g.

$$P(Y = ON \mid W = ON) = P(W = ON \mid Z = ON)$$
$$P(Z = ON \mid Y = ON) + P(W = ON \mid Z = OFF)$$
$$P(Z = OFF \mid Y = ON)$$

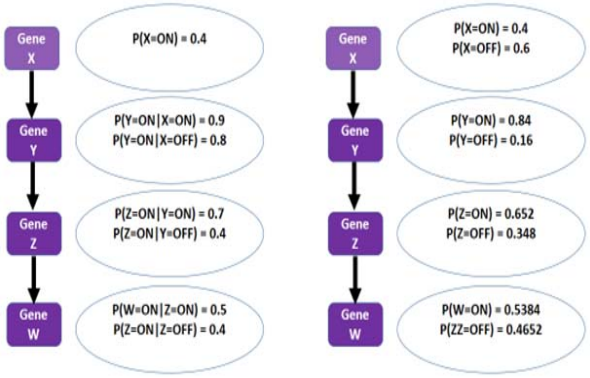

Fig. 1: Illustration of simple discretized Bayesian network with its prior information

After doing this, computation, also computing $P(W=ON \mid Y=OFF)$ (because gene X will need this letter value) and then determining $P(Y=ON \mid W=ON)$, we pass $P(W=ON \mid Y=ON)$ and $P(W=ON \mid Y=OFF)$ to X.

We then could compute: $P(W=ON \mid X=ON)$ and $P(X=ON \mid W=ON)$ in sequences as follow:

$$P(W=ON \mid X=ON) = P(W=ON \mid Y=ON)$$
$$P(Y=ON \mid X=ON) + P(W=ON \mid Y=OFF)$$
$$P(Y=OFF \mid X=ON)$$

It shows that any inquiry can be obtained by having a topology of a BN and in this way, Bayes theorem plays an extremely pivotal role. In the context of gene expression probabilistic modelling, many approaches have been developed. The so-called relevance networks were introduced by Butte and Kohane (2003) and graphical Gaussian models by Scheafer and Strimmer (2005). By extending BN, Segal et al. (2003) came up with module network, which could be seen as one of the best studies in which mathematical prediction formalism met unknown real biological regulation, e.g. many predicted regulations using module network, proved to be biologically real in wet lab work. Adriano and Husmeier (2007) introduced an intelligent way to study the integration of biological prior information (e.g. motif data, KEGG and etc.) into the inference of gene regulatory networks. Out of many findings, they showed that in the course of including two sources of prior knowledge with similar distribution, the marginal posterior distributions of the associated hyper-parameters are similar whereas when prior knowledge are different, higher weight should be assigned to the prior source of information that is more consistent with the data generation process. This findings scale up the role of relevant and irrelevant prior data in learning Bayesian network. Nurul et al. (2011) showed that BN neural network outperformed than resilient back-propagation in terms of accuracy for translation initiation sites (TIS) classification. In the context of classical dairy cattle breeding, our literature review turned out that occurrence of diseases has fetched the BN formalism than other feature of cattle production data. However, Zhang (2011) came up with a BN to model automatic and interactive image segmentation for cow image data. This idea worked well in practical scenario. In the most classical studies which have been done for understanding cytokines genes like interleukins/interferons adapted full reductionism view (Bała et al., 2004; Heiser et al., 2015). Due to their immense importance in immune system, jointly analyzing them using a robust method like BN is promising. As it came out, BN formalism permits construing regulatory relationship among genes. In this study, mRNA profile of 6 immune system candidate genes: Interfron Gamma (IFN-γ), Tomur Necrotic Factor alpha (TNF-α), Granolocyt-Macrofage Colony Stimulating Factor (GM-CSF), Interleukine-2 (IL-2), Interleukine-6 (IL-6) and Interleukine-8 (IL-8) were used to learn BN in different biological states of Iranian Holstein dairy cattle. By our knowledge, this is the first study undertaking such modelling in this theme.

Materials and Methods

Animals

Eighteen healthy dairy heifers were grouped according to their lactation stages (6 at 7-10 days, 6 at 140-150 and 6 at 290-295 days after parturition). In addition to the healthy cows, four dairy heifers with clinical mastitis were also included in this experiment. The selection criteria in healthy heifers was SCC< 350,000/ml for early lactation stage and SCC<100,000/ml for middle and late lactation stages.

RNA extraction

In healthy dairy heifers, one litre of milk sample representing all four quarters was collected in sterile tubes. The milk samples from heifers with mastitis were collected from the quarter with clinical mastitis immediately after the onset of clinical signs and before drug treatment. The milk samples were centrifuged for 20 min at 1500 g at 4°C. The cell pellet was washed in PBS pH 7.4 twice and centrifuged for 20 min at 4°C and 220 g. The pellets were lysed with 500 µl PBS- EDTA and kept at -40 until RNA extraction. Total RNA was isolated using Denazist kit according to the manufacturer's protocol. The extracted RNA samples were treated with DNase I (Cinnagen) to remove DNA contamination. The quality of extracted RNA was assessed by electrophoresis on 1% agarose gel.

cDNA Synthesis and Real-Time PCR

Synthesis of first strand cDNA was performed with random hexamer primers (Takapozist) and Accu Power ® Rocket Script TMRT PreMix kit (Bioneer) according to the manufacture's instructions. The final volume was adjusted to 50 µl with RNase free water. The amplified cDNA samples were then stored at -20 °C until use in real-time PCR. The primers used for the gene expression evaluation and the β-actin gene were used as endogenous reference for the calculation of dCp (Table 1). Real-time PCR was performed using CFX96 (BIORAD, USA) and Hot Taq Eva Green qPCR kit (Cinnagen) according to their instructions. All reactions were performed in duplicate. Amplification conditions were 95°C for 15 min; 50 cycles of 94°C for 15 s, 60°C for 30 s, and 72°C (depending on the product length, 5 s per 100 bp). Then, all samples were submitted to analysis of the dissociation curve in order to confirm the absence of nonspecific products and primer dimmers (melting curve by 95°C for 5 s, 65°C for 15 s, and 95°C for 0 s).

Learning BN

For learning Bayesian network, we used windows interface of SMILE (Structural Modeling, Inference, and Learning Engine), e.g. GeNIe (https://dslpitt. org/genie/). This versatile program is a fully different platforms portable library of C++ classes implementing graphical decision-theoretic methods like BNs). Many decision-systems laboratory around the world, have used this program to solve their hard-probabilistic decisions. We provided right format of our data to be compatible with GeNIe's data requirement. For every run, we save our results. The crucial point with GeNIe is a possibility to assign prior information in neatly way. It shall be so helpful in cases that a researcher is aware of some biological functions of the genes, proteins, etc, that (s) he is interested in learning some parameters.

Results and Discussion

Descriptive statistics of gene expression value based on cross point (CP) for 6 candidate genes across four biological states were listed in Table 1. Implicitly not explicitly, it can be seen that IFN- had high level of expression in comparing with other genes. In other words, this gene could be seen as most functional genes among these states. If we just take the information provided in Table 1, it is quite obvious that aforementioned gene show high amount of variability than other genes. However, when it does come to gene expression analysis, we need much rigor statistical measures to trade off gene expression values across time points or samples. We could find differentially expressed genes across these biological states, but it was not our immediate objective in this study. However, Kadota and Shimizu (2011) provided through references for finding differentially expressed genes.

By looking at Table 2, diverse gene association can be seen. We took a parametric approach (Pearson correlation) to compute these values. A successful and very promising of Pearson-based association analysis (gene network) was introduced in the context of gene expression. Both the strength and mode of gene associations changed (not for all genes) among biological states. For example, it turned out that GM-CSF gene (which shown to be a gene regulator in our study (Fig. 2) functions in different biological levels (Fig. 3). Put our results in quantitative genetic theory, we loosely can say that there are quite low associations across these genes using our data had almost negative and strongest association with IL-8 over early and mid-lactation (-0.87 and -0.95, respectively). But the mode of this association was changed positively in late lactation (0.42). In overall, getting interpretation among genes over these states remained cumbersome and complex. In other words, the mode and magnitude association of genes are different in different biological states. This might underscore this reality that gene's functions likely are affected in biological states in which they are functioning.

The results of Bayesian network can be seen in Figure 2. It could be seen that for each biological states e.g. mastitis, early, middle and late lactation a Bayesian network has been learned. Like Table 2 which shows association of genes in different states is different, here this discrepancy is echoed much better. What we could see here is not only this discrepancy but also the regulation mode among genes. General to all biological sates, we could see that a very simple regulation mode predicted between genes. In mastitis state, IL-8 could be seen as regulator of TNF-a, but there were not any other wiring regulatory links with other genes. Factorization of learned BN for mastitis could be written as: P(IFN-γ, TNF-α, GM-CSF, IL-2, IL-6, IL-8)=P(IFN-γ)P(GM-CSF)P(IL-2)P(IL-6)P(IL-8)P(TNF-α|IL-8). In the early and late lactation, IL-8 gene turned out to be regulator gene too. However, its target (regulate gene) in two aforementioned states was gene GM-CSF. Factorization of learned BN for these two biological states should be similar e.g. P(IFN-γ, TNF-α, GM-CSF, IL-2, IL-6, IL-8)=P(IFN-γ)P(TNF-α)P(IL-2)P(IL-6)P(IL-8)P(GM-CSF|IL-8). In the middle lactation, the regulation scenario is changed. As it can be seen in Figure 2, IL-2 is predicted to be regulator of IFN-γ. It says that by manipulating gene IL-2 it is expected that the function of IFN-γ gene undergone functional disturbance. For this biological sate, factorization of learned BN would be shown as follows P(IFN-γ, TNF-α, GM-CSF , IL-2, IL-6, IL-8)=P(GM-CSF)P(TNF-α)P(IL-2)P(IL-6)P(IL-8)P(IFN-γ|IL-2). By seeing these four BNs (and their factorizations) we

Table 1: Descriptive statistics of expression value of immune system candidate gene across different biological states

Biological state	Genes	Mean	Variance	StdDev	Min	Max
Mastitis	IFN-	33.0	0.6	0.8	32.3	34.0
	TNF-a	27.9	0.1	0.3	27.7	28.3
	GM-CSF	35.0	0.7	0.9	34.0	36.1
	IL-2	34.5	0.4	0.6	33.8	35.4
	IL-6	33.6	0.1	0.3	33.2	33.9
	IL-8	33.8	1.6	1.3	32.1	35.0
Early lactation	IFN-	40.70	10.41	3.23	37.32	45.43
	TNF-a	38.89	6.62	2.57	35.70	42.11
	GM-CSF	40.07	1.87	1.37	38.00	41.86
	IL-2	42.26	1.01	1.01	40.80	43.94
	IL-6	32.74	0.19	0.44	32.26	33.39
	IL-8	38.46	2.89	1.70	37.04	41.55
Mid lactation	IFN-	41.25	4.63	2.15	37.57	43.56
	TNF-a	43.60	1.13	1.06	41.75	44.70
	GM-CSF	37.55	0.60	0.77	36.69	38.64
	IL-2	39.96	1.70	1.30	38.74	41.99
	IL-6	33.22	0.37	0.61	32.73	34.01
	IL-8	37.12	0.47	0.69	36.20	38.00
Late lactation	IFN-	42.72	4.14	2.03	39.00	45.11
	TNF-a	38.63	2.22	1.49	37.51	40.99
	GM-CSF	38.53	5.47	2.34	35.17	41.64
	IL-2	40.67	3.85	1.96	38.19	43.70
	IL-6	32.75	0.28	0.53	32.17	33.73
	IL-8	38.48	1.62	1.27	37.00	40.06

Table 2: Pearson correlation expression value of immune system candidate gene across different biological states

Biological state		IFN-	TNF-a	GM-CSF	IL-2	IL-6	IL-8
Mastitis	IFN-	-					
	TNF-a	-0.1	-				
	GM-CSF	0.4	0.3	-			
	IL-2	-0.2	-1.0	-0.4	-		
	IL-6	-0.9	0.4	-0.2	-0.1	-	
	IL-8	0.2	-1.0	-0.2	0.9	-0.5	-
Early lactation	IFN-	-					
	TNF-a	-0.08	-				
	GM-CSF	0.71	-0.37	-			
	IL-2	0.58	0.48	0.14	-		
	IL-6	0.25	0.30	0.18	0.31	-	
	IL-8	-0.36	0.26	-0.87	-0.08	-0.06	-
Mid lactation	IFN-	-					
	TNF-a	0.53	-				
	GM-CSF	0.31	-0.49	-			
	IL-2	-0.74	-0.70	0.05	-		
	IL-6	0.34	0.07	0.38	0.23	-	
	IL-8	-0.24	0.48	-0.95	0.08	-0.15	-
Late lactation	IFN-	-					
	TNF-a	-0.29	-				
	GM-CSF	-0.58	-0.09	-			
	IL-2	0.83	-0.75	-0.42	-		
	IL-6	-0.76	0.32	0.03	-0.68	-	
	IL-8	-0.55	0.17	0.42	-0.36	0.16	-

could see that small amount of complex regulation might exist between these genes, that is, many genes shown up to function independently. Therefore, it makes sense to use some general tools to find the association of these genes with other genes not included in this study. We used online gene network Gene Mania (http://genemania.org/) to find genes that are related to our set genes in this study. We also used web-based interface http://www.pathwaycommons.org/ that contains a set of publicly accessible pathway information from multiple organisms to find the associations of our immune system candidate genes with each other and other genes biochemical reactions. GeneMANIA program is an online program and has

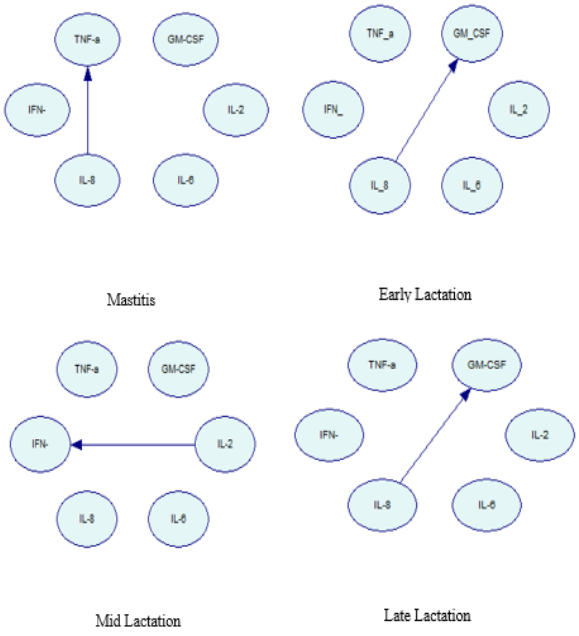

Fig. 2: Topology of learned Bayesian network of immune systems candidate genes across different biological states.

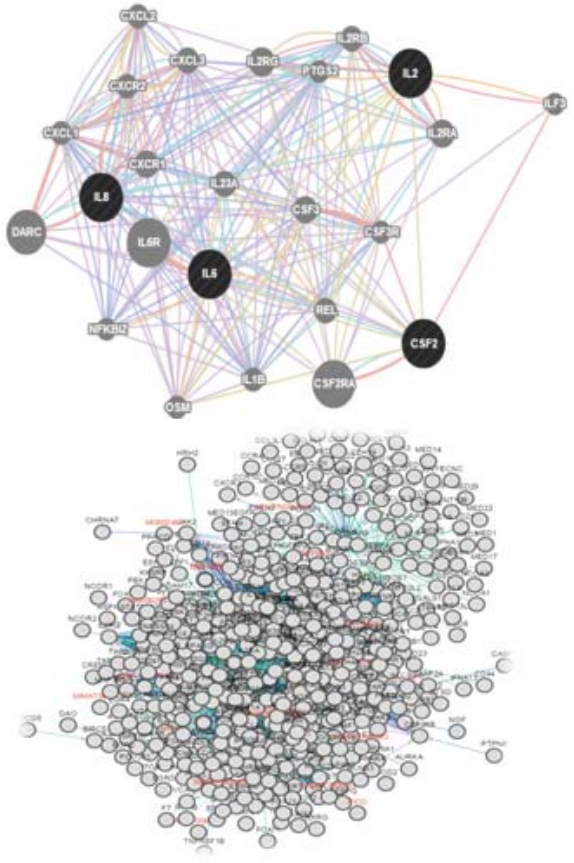

Fig. 3: Results of GeneMANIA (up) and pathway commons (down) for immune systems candidate genes

large ensemble of functional association bucolical information. This program is also accessible in different way but one of right way to find it is to delve in BioGRID (http://thebiogrid.org/). GeneMANIA provides wide list of motivated information e.g. physical interactions, co-expression, predicted, pathway, co-localization, genetic interactions, and shared protein domains just by providing some right gene names (accessions).

Results of both program underpinned that regulation of these are quite complex depending on the view we are looking at them. For example, it turned out that GM-CSF gene (which shown to be a regulate gene in our study (Fig. 2) and is indicated in right panel of Figure 3 (CSF2) functions in different biological levels. Put our results in quantitative genetic theory, we loosely can say that there are quite low associations across these genes using our data.

In other words, if the idea is to put some of these genes in breeding plan, care must be taken to select right ones. Our data could loosely say that indirect selection of these genes is hardly likely to improve the performance of other genes as no regulated wiring was observed in our study. It is generally tempting to think that structure of BN would encode the causality structure across genes. There are so many biological players hampering learned BN out of empirical data. Formal BN adopts acyclic graph with no arrow from children (regulated genes) to parents (regulator genes), representing lack of feedback loops in gene interactions (pathways). To circumvent this restriction, temporal extension called of BN e.g. dynamic Bayesian networks (DBN) was introduced (Zou and Conzen, 2005) permitting loops (feedbacks) by unrolling over time points. Chen et al. (2008) used systems biology approach to construct the gene regulatory network of systemic inflammation. They used DBN in accompany with data mining algorithms and they show a great improvement in analyzing the systemic inflammation. The screened up genes such as NF-kB, TNF-α, RELA, as highly connected hubs of the signalling transduction. Therefore, it is expected if these genes are inactivated by any means (mutation or disease), the inflammatory gene network shall culminated to eventual collapse of the system. Also addressed following genes IL1A, IL1B, IL1R, IL6, TNFA, IL17, IL8, IL1R, TLR4 and TNFR to be vital for the inflammatory response because they were more tightly connected in inflammation than in normal conditions. Recently new probabilistic formalism e.g. continuous time BN was applied to reconstruction gene network of Th17 cell (Acerbi Enzo et al., 2014). In both computationally motivated themes like small and big dataset, these methods outperformed that state-of-art modelling e.g. BN and DBN. Interleukin genes families have been deeply investigated (Pahl, 1999; Hoai et al., 2010).

Kitano and Oda (2006) and Hoai et al. (2010) addressed that Interleukin-1 alpha (IL1A) and Interleukin-1 beta (IL1B) biologically function via their receptor (IL1R) to induce gene expressions. This biological function involved in a protein loop (feedback) production in inflammatory phenomenon. Hoai et al. (2010) has worked out BN with application to family of cytokine IL15. Surprisingly, the application of BN in modelling different features of dairy cattle is not eye-catching (Morota et al., 2012). There should be some reason for it such as difficulty of interpretation of the results, lack of adequate data and etc. Morota et al. (2012) investigated the usefulness of BN for linkage disequilibrium (LD) analysis of milk protein in Holstein breed (Lindstrom et al., 2013). It is found that BN captured several genetic markers that were inter-related in a complicated scheme. Also the results shown that LD-based Bayesian network was capable to infer the associations between genetic markers in a systematic way and provides much precise of big picture of LD than general classical way of LD detection. Lindstrom et al. (2013) proposed a BN based method for understanding epidemic disease outbreak models in the United States using partially observed cattle movement data. The proposed model had this capability to scale up a full network based on Bayesian inference. It is quite likely to have a BN network with unobserved but vital edge in biological paradigms too. We may argue this matter can be talked with by Yu et al. (2009) method. However, BN learning requires much data and knowledge. In our case the dataset was scare. There are some methods which are developed to learn effective Bayesian network. Yu et al. (2009) developed a two-layered BN based diagnostic disease model that addressed such caveats e.g. dealing with uncertain knowledge in the context of existing narrow data. Similar to Lindstrom et al. (2013) which used BN formalism in disease data, Jehan et al. (2009) applied BN to create herd specific based model for existing lameness causing diseases in 50 Danish dairy herds. It is shown that BN, adjusted probability distributions of existing the disease in a given herd systematically a way that couple of sources of information, that is, population, herd and cow level information combined and the uncertainty in inference on disease probability was effectively predicted. The challenges imposed by the large number of variables but the small number of sample points were described, and a variety of computational strategies for addressing these challenges were outlined. To date, BNs have been successfully inferred for microarray data from yeast and for flow cytometry data from human immune system cells, but not for gene expression data from mammalian or oncological sources. Computational inference of Bayesian network structures from high-throughput data is difficult, but new computational methods are making it feasible to automatically deduce robust interactions between variables. The application of these methods to high-throughput biological data sets will help us to understand the nature of the altered biological interactions that lead to and occur in many diseases.

Conclusions

This paper surveyed computational learning BN model, or at least very general features of such models, from non-high-throughput immune system candidate gene expression in dairy cattle. BN formalism had capability to reveal neat and minuscule topology of immune system candidate genes. In our study, we did not have the problem of so-called curse of dimensionality. When tackling with many variables (genes) in gene expression data, many computational issues either in positive or negative side could spring up. One problem would be deriving high-dimensional and efficient search space search algorithms, which still is a hot area of research in computer science and machine learning communities. We claim that putting side-information e.g. those immune system candidate genes which turned out to be regulator (parent) in high-dimensional immune microarray BN learning, would reduce the search space learning, since by having relatively small number of candidate genes as parents, we can restrict our search to learn a network in which candidate genes be its parents culminating in having smaller search space. However, in this scenario many networks should be learned to give a reasonable explanation of real-world problem.

References

Acerbi E, Teresa Z, Narang V, Stella F (2014) Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation. BMC Bioinformatics 15: 387.

Adriano VW, Husmeier D (2007) Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge. Mol Biol 6: 167-174.

Bała G, Swincow G, Odrowaz-Sypniewska G, Czerwionka-Szaflarska M (2004) Interleukin 13 and 18 serum concentration in children with cow milk hypersensitivity. Medycyna Wieku Rozwojowego 8: 33-42.

Butte AS, Kohane IS (2003) Relevance networks: a first step toward finding genetic regulatory networks within microarray data. In: Springer. Chapter: Statist Biol Health, pp: 428-446.

Chen BS, Yang SK, Lan CY, Chuang YJ (2008) A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. BMC Med Genom 30: 31:46.

Heiser A, McCarthy A, Wedlock N, Meier S, Kay J, Walker C, Crookenden MA, Mitchell MD, Morgan S, Watkins K, Loor JJ, Roche JR (2015) Grazing dairy cows had decreased interferon-γ, tumor necrosis factor, and interleukin-17, and increased expression of interleukin-10 during the first week after calving. J Dair Sci 98: 937-946.

Yu H, Chen G, Liu D (2009) in IFIP International Federation for Information Processing, Volume (Boston: Springer), pp. 1001–1009. by domain knowledge. Based on the noisy-model, the conditional probability 294, Computer and Computing Technologies in Agriculture II, Volume 2, eds. Li D, Chunjiang Z.

Hoai N, Gerard R, Leray P, Yannick J (2010) Differential study of the cytokine network in the immune system: An evolutionary approach based on the Bayesian networks. The 2nd Asian Conference on Intelligent Information and Database Systems (ACI-IDS), Mar, Hue City, Vietnam, Hal-00656723.

Jehan FE, Soren O, Kristensen A (2009) Estimation of probability for the presence of claw and digital skin diseases by combining cow- and herd-level information using a Bayesian network. Prev Vet Med 92: 89-98.

Kadota K, Shimizu K (2011) Evaluating methods for ranking differentially expressed genes applied to microArray quality control data. BMC Bioinformatics 12: 227-239.

Kitano H, Oda K (2006) Robustness trade-offs and host-microbial symbiosis in the immune system. Mol Syst Biol 2: 237-245.

Lindstrom T, Grear DA, Buhnerkempe M, Webb C, Miller RS, Portacci K, Wennergren U (2013) A Bayesian approach for modeling cattle movements in the United States: scaling up a partially observed network. PLoS One 8: e53432.

Morota G, Valente BD, Rosa GJ, Weigel KA, Gianola D (2012) An assessment of linkage disequilibrium in Holstein cattle using a Bayesian network. J Anim Breed Gen 129: 474-87.

Nurul A, Herman NS, Burairah H (2011) Comparative Analysis using Bayesian Approachto Neural Network of Translational Initiation Sitesin Alternative Polymorphic Context. Bioautomation 15: 251-260.

Pahl HL (1999) Activators and target genes of Rel/NF-kappaB transcription actors. Oncogene 18: 6853-6866.

Scheafer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statist Applic Gen Mol Biol 4: 432-451.

Segal E, Shapira, M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Gen 34: 166-176.

Zhang L (2011) Bayesian Network Model for Automatic and Interactive Image Segmentation. *IEEE T.* Image Proc 20: 2582-2594.

Zou M, Conzen SD (2005) A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 21: 71-9.