12-5-2013

Impact of Noise on Molecular Network Inference

Radhakrishnan Nagarajan
University of Kentucky, rnagarajan@uky.edu

Marco Scutari
University College London, United Kingdom

Impact of Noise on Molecular Network Inference

Notes/Citation Information
Published in *PLOS One*, v. 8, issue. 12, e80735.

Digital Object Identifier (DOI)
http://dx.doi.org/10.1371/journal.pone.0080735

OPEN ACCESS Freely available online

PLOS ONE

Impact of Noise on Molecular Network Inference

Radhakrishnan Nagarajan[1]*, **Marco Scutari**[2]

1 Division of Biomedical Informatics, Department of Biostatistics, University of Kentucky, United States of America, **2** UCL Genetics Institute, University College London, London, United Kiingdom

Abstract

Molecular entities work in concert as a system and mediate phenotypic outcomes and disease states. There has been recent interest in modelling the associations between molecular entities from their observed expression profiles as networks using a battery of algorithms. These networks have proven to be useful abstractions of the underlying pathways and signalling mechanisms. Noise is ubiquitous in molecular data and can have a pronounced effect on the inferred network. Noise can be an outcome of several factors including: inherent stochastic mechanisms at the molecular level, variation in the abundance of molecules, heterogeneity, sensitivity of the biological assay or measurement artefacts prevalent especially in high-throughput settings. The present study investigates the impact of discrepancies in noise variance on pair-wise dependencies, conditional dependencies and constraint-based Bayesian network structure learning algorithms that incorporate conditional independence tests as a part of the learning process. Popular network motifs and fundamental connections, namely: (*a*) common-effect, (*b*) three-chain, and (*c*) coherent type-I feed-forward loop (FFL) are investigated. The choice of these elementary networks can be attributed to their prevalence across more complex networks. Analytical expressions elucidating the impact of discrepancies in noise variance on pairwise dependencies and conditional dependencies for special cases of these motifs are presented. Subsequently, the impact of noise on two popular constraint-based Bayesian network structure learning algorithms such as Grow-Shrink (GS) and Incremental Association Markov Blanket (IAMB) that implicitly incorporate tests for conditional independence is investigated. Finally, the impact of noise on networks inferred from publicly available single cell molecular expression profiles is investigated. While discrepancies in noise variance are overlooked in routine molecular network inference, the results presented clearly elucidate their non-trivial impact on the conclusions that in turn can challenge the biological significance of the findings. The analytical treatment and arguments presented are generic and not restricted to molecular data sets.

Citation: Nagarajan R, Scutari M (2013) Impact of Noise on Molecular Network Inference. PLoS ONE 8(12): e80735. doi:10.1371/journal.pone.0080735

Editor: Alberto de la Fuente, Leibniz-Institute for Farm Animal Biology (FBN), Germany

Received July 8, 2013; **Accepted** October 7, 2013; **Published** December 5, 2013

Funding: RN acknowledges support from Kentucky Center for Clinical and Translational Science, UL1TR000117. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rnagarajan@uky.edu

Introduction

Identifying associations and network structures from observational data sets obtained across a given set of entities is a challenging problem and of great interest across a spectrum of disciplines including molecular biology [1–8]. While the molecular entities of interest are represented by the *nodes*, their associations are represented by the *edges*. Such networks can prove to be convenient abstractions of the underlying pathways and signalling mechanisms across distinct phenotypes and disease states. [1,2,7]. They can reveal interesting characteristics including repetitive structures, dominant players, community structures and generative mechanism [9–11] that can assist in developing meaningful interventions.

Molecular data obtained from biological systems may or may not have explicit temporal information. While the former explicitly captures the evolution of the molecular activity as a function of time (*dynamic*), the latter represents a snapshot of the biological activity in a given window of time (*static*). Dynamic data sets are rare and challenging to generate since they demand controlling a number of factors. Static data sets in conjunction with multiple independent realizations are relatively easier to generate. Their prevalence may also be attributed to the tradition of generating replicate measurements in molecular biology in order to demonstrate reproducibility of the findings. Prior studies on static data sets used pairwise dependency measures to capture the associations between a given set of molecules in the form of *relevance networks* [1]. The underlying hypothesis being that correlated genes are likely to be co-regulated or functionally related [12]. However, pairwise dependency measures by definition are symmetric measures resulting in *undirected graphs*. It is also known that the dependency between a given pair of genes may not necessarily be direct and possibly mediated by other gene(s). This possibly motivated the choice of conditional dependencies as opposed to pairwise dependencies for molecular network inference. Subsequently, probabilistic approaches such as Bayesian network structure learning techniques that model the conditional dependencies across a larger number of variables in an automated manner were proposed to infer molecular networks from static data sets [3,6,7]. The resulting networks of constraint-based structure learning are typically in the form of *directed acyclic graphs* (DAGs) or *partially directed acyclic graphs* (PDAGs). While DAGs have directed edges, PDAGs have directed as well as undirected edges and accommodate the presence of *equivalent classes* [13,14]. Constraint-based structure-learning algorithms by their very nature do not accommodate the presence of cycles and feedback between the molecules of interest which is an inherent limitation. They have nevertheless proven to be useful approximations of

pathways and signalling mechanisms [6,7,13]. The DAGs (PDAGs) may also reveal possible *causal relationships* between the nodes under certain implicit assumptions [15].

Of interest, is to note that these molecular data sets are inherently noisy [16,17,18]. Noise and its variation across molecular entities may have contributions from several factors including stochastic mechanisms coupled to the systems dynamics, sensitivity and precision of the measurement device, variations in abundance of specific molecules, preferential binding affinities and experimental artefacts that are an outcome of the estimation process [7,19,20,21]. While identifying the source of noise is a challenging problem in its own merit, understanding its impact on network inference procedure is especially critical in order to avoid identification of spurious associations. In a recent study, we elucidated the non-trivial impact of noise and auto-regulatory feedback on networks inferred using Granger causality tests. The results were established on multivariate time series generated using gene network motifs modelled as vector auto-regressive processes (VAR) [22], as well as those inferred from cell-cycle microarray temporal gene expression profiles [23,24]. The present study investigates the impact of noise on pair-wise correlation, partial correlation and constraint-based structure learning algorithms by considering static data sets generated from linear models of popular *network motifs* and publicly available molecular expression data [7]. Network motifs are repetitive atomic structures that have been found to be prevalent across more complex networks [9]. In the present study, we consider three popular three-node motifs, namely: *common-effect*, *three-chain* and the *coherent type-I feed-forward loop (FFL)* [9,25,26]. The *common-effect motif* and the *three-chain motif* represent the *convergent* and *serial connection* respectively. These connections comprise the *fundamental connections* in Bayesian networks [27]. Furthermore, the conditional independence relationships represented by these motifs are usually among the first to be examined in any constraint-based structure learning algorithm justifying their choice. Common-effect motif is also an essential ingredient in identifying *equivalent classes* and PDAGs [13]. The coherent type-I FFL has been shown to persist across a number of organisms including E. Coli and S. Cerivisiae [25,26]. Of interest, is to note that three-chain and common-effect motifs are an integral part of a type-I coherent FFL. Analytical expressions for large discrepancies in noise variance on pairwise (*correlation coefficient*) and conditional dependencies (*partial correlation*) are investigated. The impact of such discrepancies on constraint-based Bayesian network structure learning is also investigated. Finally, the presence of significant discrepancies in noise variance and its impact on network inference from experimental molecular expression profiles [7] is investigated.

Methods and Results

Prior to investigating the impact of noise on the constraint-based Bayesian network structure learning algorithms, its impact on pairwise and conditional dependencies across the three network motifs is investigated.

2.1 Pairwise and Conditional Dependencies

Network Motif Parameters. In the following discussion, (x_t,y_t,z_t) represent the molecular expression of the three genes (x,y,z) respectively in a small time window $(T,T+t)$. The terms $(\epsilon_t,\eta_t,\delta_t)$ represent zero-mean, unit-variance uncorrelated noise attributed to inherent uncertainties and artifacts prevalent in molecular expression studies. Parameter$(\alpha>0)$ represents the transcriptional coupling strengths between the genes and is constrained to be equal across the genes, since the impact of variations in α on pairwise and conditional dependencies is expected and not the goal of the present study. Discrepancies in the noise variances across the nodes are represented by parameters $\gamma_i>0$, $i=1,2$.

Case 1: Common-effect network motif. The common-effect network motif (*v*-structure) [13] is a fundamental connection, Fig. 1a, discussed widely within the context of Bayesian network structure learning algorithms. For this motif, z is regulated by x and y given by the linear model,

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha & \alpha & 0 \end{bmatrix} . \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \gamma_1.\eta_t \\ \gamma_2.\delta_t \end{bmatrix} \tag{1}$$

The correlation coefficients are given by

$$\begin{aligned} \rho_{xy} &= \frac{E(xy)}{\sigma_x\sigma_y} = 0 \\ \rho_{xz.y} &= \frac{E(xz)}{\sigma_x\sigma_y} = \frac{\alpha}{\sqrt{\left(\alpha^2+\alpha^2\gamma_1^2+\gamma_2^2\right)}} \\ \rho_{yz} &= \frac{E(yz)}{\sigma_y\sigma_z} = \frac{\alpha\gamma_1}{\sqrt{\alpha^2+\alpha^2\gamma_1^2+\gamma_2^2}}, \gamma_1 \neq 0 \end{aligned} \tag{2}$$

The partial correlations are given by

$$\begin{aligned} \rho_{xy.z} &= -\frac{\alpha^2\gamma_1}{\sqrt{\left(\alpha^2\gamma_1^2+\gamma_2^2\right)}\sqrt{\left(\alpha^2+\gamma_2^2\right)}} \\ \rho_{xz.y} &= \frac{\alpha}{\sqrt{\left(\alpha^2+\gamma_2^2\right)}} \\ \rho_{yz.x} &= \frac{\alpha\gamma_1}{\sqrt{\left(\alpha^2\gamma_1^2+\gamma_2^2\right)}} \end{aligned} \tag{3}$$

For large noise limit at $z(\gamma_2\rightarrow\infty)$ with finite noise at $y(\gamma_1 \ll \gamma_2)$, the correlation coefficients are given by

$$\begin{aligned} \lim_{\gamma_2\rightarrow\infty} \rho_{xy} &= 0 \\ \lim_{\gamma_2\rightarrow\infty} \rho_{xz} &= 0 \\ \lim_{\gamma_2\rightarrow\infty} \rho_{yz} &= 0 \end{aligned} \tag{4}$$

The partial correlations are given by

$$\begin{aligned} \lim_{\gamma_2\rightarrow\infty} \rho_{xy.z} &= 0 \\ \lim_{\gamma_2\rightarrow\infty} \rho_{xz.y} &= 0 \\ \lim_{\gamma_2\rightarrow\infty} \rho_{yz.x} &= 0 \end{aligned} \tag{5}$$

For large noise limit at y($\gamma_1 \to \infty$) with finite noise at z($\gamma_2 \ll \gamma_1$), the correlation coefficients are given by

$$\begin{aligned} \lim_{\gamma_1 \to \infty} \rho_{xy} &= 0 \\ \lim_{\gamma_1 \to \infty} \rho_{xz} &= 0 \\ \lim_{\gamma_1 \to \infty} \rho_{yz} &= 1 \end{aligned} \tag{6}$$

The partial correlations are given by

$$\begin{aligned} \lim_{\gamma_1 \to \infty} \rho_{xy.z} &= \frac{-\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}} \\ \lim_{\gamma_1 \to \infty} \rho_{xz.y} &= \frac{\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}} \\ \lim_{\gamma_1 \to \infty} \rho_{yz.x} &= 1 \end{aligned} \tag{7}$$

Remark 1. *Correlation coefficient estimates reveal significant pairwise dependencies across* (x,z) *and* (y,z) *in contrast to* (x,y) *resulting in the undirected graph* $x-z, y-z$. *As expected, conditioning the marginally independent nodes* (x,y) *on* z *renders them dependent* ($i.e. \rho_{xy.z} \neq 0$).

- *(i)Large noise limit at the common-effect node* z($\gamma_2 \to \infty, \gamma_1 \ll \gamma_2$)*: Pairwise as well as conditional dependencies vanish (4, 5) challenging any reliable conclusion on the network structure in the large noise limit when* ($\gamma_2 \to \infty, \gamma_1 \ll \gamma_2$) *preventing any reliable inference of the network. More importantly, conditioning on the common-effect node at large noise levels did not render* x *and* y *dependent as expected (5).*
- *(ii)Large noise limit at one of the causes* z($\gamma_2 \to \infty, \gamma_1 \ll \gamma_2$)*: Pairwise dependencies* (x,y) *as well as* (x,z) *disappear (6). Interestingly, conditional dependencies* $\rho_{xy.z}$ *and* $\rho_{xz.y}$ *are equal in magnitude with opposite signs and function of* γ_2 *(7). Pairwise as well as conditional dependencies* ρ_{yz} *and* $\rho_{yz.x}$ *have maximal values of unity in the large noise limit at* ***y***.

Case 2. *Three-chain network motif*

Consider the three-chain network motif [9], Fig. 1b, where y mediates the activity between (x,z) given by the linear model

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{bmatrix} . \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \gamma_1 . \eta_t \\ \gamma_2 . \delta_t \end{bmatrix} \tag{8}$$

The correlation coefficients are given by

$$\begin{aligned} \rho_{xy} &= \frac{E(xy)}{\sigma_x \sigma_y} = \frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}} \\ \rho_{xz} &= \frac{E(xz)}{\sigma_x \sigma_z} = \frac{\alpha^2}{\sqrt{\alpha^2(\alpha^2+\gamma_1^2)+\gamma_2^2)}} \\ \rho_{yz} &= \frac{E(yz)}{\sigma_y \sigma_z} = \frac{\alpha\sqrt{\alpha^2+\gamma_1^2}}{\sqrt{\alpha^2(\alpha^2+\gamma_1^2)+\gamma_2^2)}} \end{aligned} \tag{9}$$

The partial correlations are given by

$$\begin{aligned} \rho_{xy.z} &= \frac{\alpha\gamma_2}{\sqrt{(\alpha^2\gamma_1^2+\gamma_2^2)(\alpha^2+\gamma_1^2)}}, \gamma_2 \neq 0 \\ \rho_{xz.y} &= 0 \\ \rho_{yz.x} &= \frac{\alpha\gamma_1}{\sqrt{(\alpha^2\gamma_1^2+\gamma_2^2)}}, \gamma_1 \neq 0 \end{aligned} \tag{10}$$

For large noise limit at z($\gamma_2 \to \infty$) with finite noise at y($\gamma_1 \ll \gamma_2$), the correlation coefficients are given by

$$\begin{aligned} \lim_{\gamma_2 \to \infty} \rho_{xy} &= \frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}} \\ \lim_{\gamma_2 \to \infty} \rho_{xz} &= 0 \\ \lim_{\gamma_2 \to \infty} \rho_{yz} &= 0 \end{aligned} \tag{11}$$

The partial correlations are given by

$$\begin{aligned} \lim_{\gamma_2 \to \infty} \rho_{xy.z} &= \frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}} \\ \lim_{\gamma_2 \to \infty} \rho_{xz.y} &= 0 \\ \lim_{\gamma_2 \to \infty} \rho_{yz.x} &= 0 \end{aligned} \tag{12}$$

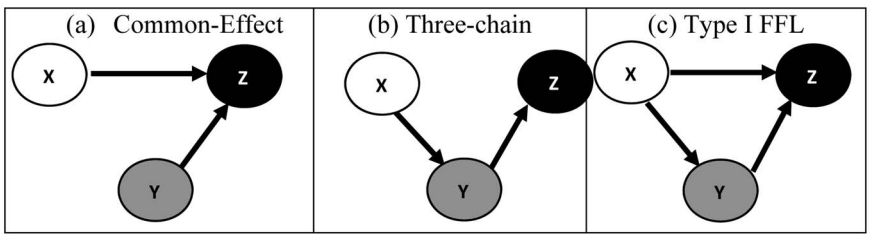

Figure 1. Popular three-gene network motifs: common-effect, three-chain and coherent type-I feed-forward loop are shown in (a), (b) and (c) respectively.
doi:10.1371/journal.pone.0080735.g001

For large noise limit at y($\gamma_1 \to \infty$) with finite noise at z($\gamma_2 \ll \gamma_1$), the correlation coefficients are given by

$$\lim_{\gamma_1 \to \infty} \rho_{xy} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{xz} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{yz} = 1 \quad (13)$$

The partial correlations are given by

$$\lim_{\gamma_1 \to \infty} \rho_{xy.z} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{xz.y} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{yz.x} = 1 \quad (14)$$

Remark 2. *Correlation coefficient estimates reveal significant pairwise dependencies across* (x,y)*,* (y,z) *and* (x,z) *resulting in the undirected graph* $x-y, y-z, x-z$*. As expected, conditioning the marginally dependent nodes* (x,z) *on* y *renders them independent* ($i.e. \rho_{xz.y} = 0$)*. This result is immune to the choice of the linear model parameters and reflects possible directed acyclic graph of the form* $x \to y \to z$*.*

- *(i). Large noise limit at the node* z ($\gamma_2 \to \infty$)*: Pairwise dependencies (11),* ($\rho_{xy}, \rho_{xz}, \rho_{yz}$) *are identical to the conditional dependencies in (12),* ($\rho_{xy.z}, \rho_{xz.y}, \rho_{yz.x}$)*. Of interest is to note that pairwise dependencies* ρ_{xy} *and conditional dependency* $\rho_{xy.z}$ *have identical non-zero magnitude.*
- *(ii). Large noise limit at the node* y ($\gamma_1 \to \infty$)*: Pairwise dependencies (13),* ($\rho_{xy}, \rho_{xz}, \rho_{yz}$) *are identical to those of conditional dependencies (14),* ($\rho_{xy.z}, \rho_{xz.y}, \rho_{yz.x}$) *similar to what was observed for* ($\gamma_2 \to \infty$)*. However, in contrast to* ($\gamma_2 \to \infty$)*, pairwise* (ρ_{yz}) *and conditional dependencies* $\left(\rho_{yz.x}\right)$ *are identical with a maximum value similar to that of the common-effect network motif. Also, pair-wise dependencies* ($\rho_{xy}, \rho_{xz}, \rho_{yz}$) *(13) are identical to those obtained for the common-effect motif (6) failing to distinguish these two structures.*

Case 3. *Coherent Type-I feed-forward loop network motif*

Consider the coherent type-I feed-forward loop [25,26], Fig. 1c, where the expression of y is regulated by x whereas those of z is regulated by x as well as z given by the linear model

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \alpha & 0 & 0 \\ \alpha & \alpha & 0 \end{bmatrix} . \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \gamma_1 . \eta_t \\ \gamma_2 . \delta_t \end{bmatrix} \quad (15)$$

The correlation coefficients are given by

$$\rho_{xy} = \frac{E(xy)}{\sigma_x \sigma_y} = \frac{\alpha}{\sqrt{\alpha^2 + \gamma_1^2}}$$
$$\rho_{xz} = \frac{E(xz)}{\sigma_x \sigma_z} = \frac{\alpha^2 + \alpha}{\sqrt{\alpha^4 + 2\alpha^3 + \alpha^2 + \alpha^2\gamma_1^2 + \gamma_2^2}}$$
$$\rho_{yz} = \frac{E(yz)}{\sigma_y \sigma_z} = \frac{\alpha^3 + \alpha^2 + \alpha\gamma_1^2}{\sqrt{\alpha^2 + \gamma_1^2}\sqrt{\alpha^4 + 2\alpha^3 + \alpha^2 + \alpha^2\gamma_1^2 + \gamma_2^2}} \quad (16)$$

The partial correlations are given by

$$\rho_{xy.z} = \frac{\alpha(\gamma_2^2 - \alpha\gamma_1^2)}{\sqrt{\alpha^2\gamma_1^2 + \gamma_2^2}\sqrt{\alpha^2\gamma_2^2 + \alpha^2\gamma_1^2 + \gamma_1^2\gamma_2^2}}$$
$$\rho_{xz.y} = \frac{\alpha\gamma_1}{\sqrt{\alpha^2\gamma_2^2 + \alpha^2\gamma_1^2 + \gamma_1^2\gamma_2^2}}, \gamma_1 \neq 0$$
$$\rho_{yz.x} = \frac{\alpha\gamma_1}{\sqrt{\alpha^2\gamma_1^2 + \gamma_2^2}}, \gamma_1 \neq 0 \quad (17)$$

For large noise limit at z($\gamma_2 \to \infty$) with finite noise at y($\gamma_1 \ll \gamma_2$), the correlation coefficients and partial correlations are given by

$$\lim_{\gamma_2 \to \infty} \rho_{xy} = \frac{\alpha}{\sqrt{\alpha^2 + \gamma_1^2}}$$
$$\lim_{\gamma_2 \to \infty} \rho_{xz} = 0$$
$$\lim_{\gamma_2 \to \infty} \rho_{yz} = 0 \quad (18)$$

The partial correlations are given by

$$\lim_{\gamma_2 \to \infty} \rho_{xy.z} = \frac{\alpha}{\sqrt{\alpha^2 + \gamma_1^2}}$$
$$\lim_{\gamma_2 \to \infty} \rho_{xz.y} = 0$$
$$\lim_{\gamma_2 \to \infty} \rho_{yz.x} = 0 \quad (19)$$

For large noise limit at y($\gamma_1 \to \infty$) with finite noise at z($\gamma_2 \ll \gamma_1$), the correlation coefficients and partial correlations are given by

$$\lim_{\gamma_1 \to \infty} \rho_{xy} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{xz} = 0$$
$$\lim_{\gamma_1 \to \infty} \rho_{yz} = 1 \quad (20)$$

The partial correlations are given by

$$\lim_{\gamma_1 \to \infty} \rho_{xy.z} = \frac{-\alpha}{\sqrt{\alpha^2 + \gamma_2^2}}$$
$$\lim_{\gamma_1 \to \infty} \rho_{xz.y} = \frac{\alpha}{\sqrt{\alpha^2 + \gamma_2^2}}$$
$$\lim_{\gamma_1 \to \infty} \rho_{yz.x} = 1 \quad (21)$$

Remark 3. *Correlation coefficient estimates reveal significant pairwise dependencies across* (x,y)*,* (y,z) *and* (x,z) *indicating a possible undirected*

Table 1. Pair-wise and conditional dependencies across the three network motifs in the asymptotic noise limits.

$\gamma_1 \to \infty$	ρ_{xy}	ρ_{xz}	ρ_{yz}	$\rho_{xy.z}$	$\rho_{xz.y}$	$\rho_{yz.x}$
Common-Effect	0	0	1	$\frac{-\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}}$	$\frac{\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}}$	1
Three-Chain	0	0	1	0	0	1
Type I FFL	0	0	1	$\frac{-\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}}$	$\frac{\alpha}{\sqrt{(\alpha^2+\gamma_2^2)}}$	1
$\gamma_2 \to \infty$	ρ_{xy}	ρ_{xz}	ρ_{yz}	$\rho_{xy.z}$	$\rho_{xz.y}$	$\rho_{yz.x}$
Common-Effect	0	0	0	0	0	0
Three-Chain	$\frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}}$	0	0	$\frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}}$	0	0
Type I FFL	$\frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}}$	0	0	$\frac{\alpha}{\sqrt{\alpha^2+\gamma_1^2}}$	0	0

doi:10.1371/journal.pone.0080735.t001

graph of the form $x-y-z$. Unlike the three-chain, conditioning (x,z) *on* y *does not render them independent.*

- *(i). Large noise limit at* $z(\gamma_2 \to \infty)$*: Pairwise dependencies (18) and conditional dependencies (19) are identical to those obtained for the three-chain motif (11, 12) failing to distinguish these structures for relatively large noise variance at* z*, Table 1.*
- *(ii). Large noise limit at* $y(\gamma_1 \to \infty)$*: Pairwise dependencies (20) and conditional dependencies (21) are identical to those obtained for the common-effect motif (6), (7) failing to distinguish these structures for relatively large noise variance at* y*, Table 1. Also, the pairwise dependencies for* $(\gamma_1 \to \infty)$*is identical for the coherent Type I FFL, three-chain as well as the common-effect motif.*

2.2 Constraint-based Bayesian Network Structure Learning

Bayesian network structure learning algorithms have been used successfully to infer the associations between a large numbers of variables. Several such algorithms have been proposed in literature, a partial list of contributions include [28,29,30,31,32]. In the present discussion, we focus on constraint-based structure learning algorithms that infer the network structure using tests for conditional independence, namely: the Grow-Shrink (GS) algorithm [30] and the Incremental Association Markov Blanket (IAMB) [31].

GS was the first algorithm that learned the *Markov blanket* of each node as an intermediate step to speed up structure learning process. The Markov blanket $Bl(X)$ of a node X is defined as the set of nodes that makes X independent from all the other nodes in the domain. In a Bayesian network, it is formed by the parents of X, its children, and the other parents of its children [15]. Therefore, the search for the neighbors of each node can be restricted to its Markov blanket, which in most cases contains a limited number of nodes. GS learns Markov blankets using a forward selection (*Growing Phase*) followed by a backward selection (*Shrinking Phase*). Conditional independence tests are performed in order of increasing complexity (i.e. with respect to the number of nodes involved in the test) in order to maximize the overall power of the structure learning algorithm. Markov blankets are then reduced to the corresponding set of neighbors by an additional backward selection. Arc directions are established starting from *v*-structures, which can be identified by the interplay of the causes conditional on their common effect, and then propagated to prevent the formation of further *v*-structures and enforce *acyclicity*. This is achieved using the heuristics described elsewhere [30,33]. IAMB introduces relatively better heuristics to identify Markov blankets while improving on GS by using a forward stepwise regression. However, IAMB in contrast to GS is designed to identify the Markov blanket of each node and not the complete network structure. Essentially, it performs the same task as the first step of GS but the forward stepwise selection in IAMB reduces the number of nodes incorrectly included in the Markov blankets. In the context of Bayesian network structure learning, IAMB is extended to a complete learning algorithm by adding steps 2 to 4 of GS. While both algorithms have been shown to be formally correct, IAMB has been recently supported by more extensive proofs and simulations [34,35]. Of interest is to note that GS as well as IAMB are highly dependent on the ability of the conditional independence tests to correctly identify dependence relationships. In fact, the proofs of correctness of both structure learning algorithms implicitly assume absence of type I or type II errors. Such an assumption can especially be violated in the presence of noise that may accentuate false-positives as well as false-negatives challenging the biological significance of the results. This in turn justifies investigating the impact of discrepancies in noise variance across the nodes on network inference using GS and IAMB. Since the conditional independence tests increase in complexity during the structure learning process across GS and IAMB [36] the present study is restricted to well-established network motifs that are prevalent across more complex structures. The concerns presented across these motifs are expected to be aggravated across more complex network topologies.

Common-effect network motif

For large noise limit at $z(\gamma_2 \to \infty)$ *with finite noise at* $y(\gamma_1 \ll \gamma_2)$*:*

For relatively large noise variance at z, the pairwise as well as conditional dependencies (4, 5) vanish across GS as well as IAMB resulting in an empty network. This happens regardless of the values of $(\rho_{xy.z},\rho_{xz.y},\rho_{yz.x})$ because both GS and IAMB test for significant pairwise dependencies $(\rho_{xy},\rho_{xz},\rho_{yz})$ first and conclude the Markov blankets of x,y and z to be empty sets. As a consequence, none of the nodes have any neighbours resulting in an empty graph.

For large noise limit at $y(\gamma_1 \to \infty)$ *with finite noise at* $z(\gamma_2 \ll \gamma_1)$*:*

For relatively large noise variance at y, GS was able to retrieve a part of the network structure as discussed below. The Markov blankets inferred by GS are as follows:

- *For $Bl(x)$ from (6) we have $x \perp y$, i.e. $\rho_{xy}=0$ and $x \perp z$, i.e. $\rho_{xz}=0$ resulting in $Bl(x)=\varnothing$.*
- *For $Bl(y)$, from (6) we have $y \perp x$, i.e. $\rho_{xy}=0$ and yz, i.e. $\rho_{yz}=1$. As a result, z is added to $Bl(y)$. Also from (7), yx given z, i.e. $\rho_{xy.z} \neq 0$ since $\alpha>0$. Therefore,$x$is added to $Bl(y)$ for suitable values of α resulting in $Bl(y)=\{x,z\}$* characteristic of the motif (1).
- *For $Bl(z)$, from (6) we have $z \perp x$, i.e. $\rho_{xz}=0$ but zy,i.e. $\rho_{yz}=1$. As a result,y is added to $Bl(z)$. Also from (7) $\rho_{xz.y} \neq 0$, since $\alpha>0$. Therefore, a suitable choice of α results in the Markov blanket $Bl(z)=\{x,y\}$ characteristic of the motif (1).*

For IAMB, the conditional independence tests are performed in a different order since the nodes are included in the Markov blankets in decreasing order of association. However, the resulting Markov blankets$Bl(x)$, $Bl(y)$ and $Bl(z)$ are same as those of GS. The impact of discrepancies in noise variance across the nodes on structure learning is especially elucidated by the asymmetry of the Markov blankets $Bl(x)$ and $Bl(y)$ as well as $Bl(x)$and $Bl(z)$. Markov blankets are symmetric by definition, i.e.$x \in Bl(y)$ then $y \in Bl(x)$ and vice versa. However, for the present case we have following asymmetries ($x \in Bl(y)$ while $y \notin Bl(x)$) and ($x \in Bl(z)$ while $z \notin Bl(x)$) violating the definition of Markov blanket. For consistency, a *symmetry correction* [34,35] may be applied either by removing x from $Bl(y)$ and$Bl(z)$, or adding y and z to $Bl(x)$. The latter correction enables faithful reproduction of the motif while the former does not.

Three-Chain network motif

For large noise limit at $z(\gamma_2 \to \infty)$ with finite noise at $y(\gamma_1 \ll \gamma_2)$:z

For relatively larger noise variance at z, the Markov blankets inferred by GS are given as follows:

- *For $Bl(x)$, from (11) we know that xy, i.e. $\rho_{xy} \neq 0$ since $\alpha>0$. For suitable choice of α, we may correctly infer $Bl(x)=\{y\}$. Also, from (11, 12) we have $x \perp z$, i.e. $\rho_{xz}=0$ and $x \perp z|y$, i.e. $\rho_{xz.y}=0$ so $z \notin Bl(x)$. Therefore, the ability to infer $Bl(x)$ depends on α.*
- *For $Bl(y)$, from (11, 12) we have $y \perp z$, i.e. $\rho_{yz}=0$ and $y \perp z|x$, i.e. $\rho_{yz.x}=0$ resulting in either $Bl(y)=\{x\}$or$Bl(y)=\varnothing$ markedly different from $Bl(y)=\{x,z\}$* characteristic of the motif (8).
- *For $Bl(z)$, from (11) we have $z \perp x$, i.e. $\rho_{xz}=0$ and $y \perp z$, i.e. $\rho_{yz}=0$resulting in $Bl(z)=\varnothing$ in contrast to $Bl(z)=\{y\}$ characteristic of the motif (8).*

As in the case of common-effect network motif, reordering of the conditional independence tests in IAMB does not result in Markov blankets different from those inferred by GS. Unlike common-effect motif, no asymmetry between the Markov blankets is observed for the three-chain, since $x \in Bl(y)$ and $y \in Bl(x)$ are established using the same correlation coefficient ρ_{xy}. Given these set of Markov blankets, identifying the correct network structure is impossible. Since for large values of α, both GS and IAMB learn $(x-y,z)$, while for small values of α both GS and IAMB are unable to identify any of the arcs present in the true motif structure. The presence of at most a single arc $x-y$ makes it impossible to infer its direction, since both GS and IAMB use *v*-structures to infer directions and the learned motif structure contains none.

For large noise limit at $y(\gamma_1 \to \infty)$ with finite noise at $z(\gamma_2 \ll \gamma_1)$:

For relatively large noise variance at y, no reliable conclusion of the motif is possible across GS as well as IAMB. The Markov blankets are as follows:

- *For $Bl(x)$, from (13) we have $x \perp y$, i.e. $\rho_{xy}=0$ and $x \perp z$, $\rho_{xz}=0$. As a result, $Bl(x)=\varnothing$ in contrast to $Bl(x)=\{y\}$ characteristic of the motif (8).*
- *For $Bl(y)$, from (13) we have $y \perp x$, i.e. $\rho_{xy}=0$ but yz, i.e. $\rho_{yz}=1$. Even after updating the Markov blanket to $Bl(y)=\{z\}$, the dependence between x and y is obscured by noise as $\rho_{xy.z}=0$. Therefore, the Markov blanket $Bl(y)=\{z\}$.*
- *For $Bl(z)$, from (13) we have that $z \perp x$, i.e. $\rho_{xz}=0$ but zy, i.e. $\rho_{yz}=1$. Also, from (14) we have $x \perp z|y$, i.e. $\rho_{xz.y}=0$. This results in the Markov blanket $Bl(z)=\{y\}$ characteristic of the motif (8).*

In this case, no asymmetry is observed despite the effects of noise. Nevertheless, neither GS nor IAMB was able to learn the motif for relatively large noise variance.

Coherent Type-I Feed-Forward Loop motif

For large noise limit at $z(\gamma_2 \to \infty)$ with finite noise at $y(\gamma_1 \ll \gamma_2)$:

For relatively large noise variance at z, the Markov blankets determined by GS and IAMB are as follows:

- *For $Bl(x)$, from (18), x y i.e. $\rho_{xy} \neq 0$, since $\alpha>0$. Also from (18, 19) we note that $x \perp z$, i.e. $\rho_{xz}=0$ and $x \perp z|y$, i.e.$\rho_{xz.y}=0$. Therefore, z is not included in $Bl(x)$. Thus, GS and IAMB return either $Bl(x)=\varnothing$ or $Bl(x)=\{y\}$ for suitable choice of α in contrast to $Bl(x)=\{y,z\}$ characteristic of the motif (15).*
- *For $Bl(y)$, from (18) xy, i.e.$\rho_{xy} \neq 0$, since $\alpha>0$. Also, from (18, 19) we have $y \perp z$, i.e. $\rho_{yz}=0$ and $y \perp z|x$, i.e. $\rho_{yz.x}=0$. Therefore, z is not included in $Bl(y)$. Thus, GS and IAMB return either $Bl(y)=\varnothing$ or $Bl(y)=\{x\}$ for suitable choice of αas opposed to $Bl(y)=\{x,z\}$* characteristic of the motif (15).
- *For$Bl(z)$, it is impossible to learn the correct Markov blanket $Bl(z)=\{x,y\}$ since$z \perp x$,i.e. $\rho_{xz}=0$ as well as $z \perp y$,i.e. $\rho_{yz}=0$ from (18). As a result, $Bl(z)=\varnothing$.*

In the present case, discrepancy in noise variance does not result in asymmetry in the Markov blankets. Thus, symmetry correction may not alleviate the impact of noise. Possible motif structures corresponding to large discrepancies at z are either an empty structure or $(x-y,z)$. This is problematic for two reasons. First, only one arc out of three is correctly identified and its direction cannot be determined by the learning algorithm. Second, the motif structures above are indistinguishable from those obtained for the three-chain network motif.

For large noise limit at $y(\gamma_1 \to \infty)$ with finite noise at $z(\gamma_2 \ll \gamma_1)$:

For relatively large noise variance y, again neither GS nor IAMB was able to infer the motif. The Markov blankets are given as follows:

- *For $Bl(x)$, from (20) we have $x \perp y$, i.e. $\rho_{xy}=0$ and $x \perp z$, i.e. $\rho_{xz}=0$. This results in Markov blanket $Bl(x)=\varnothing$in contrast to $Bl(x)=\{y,z\}$For $Bl(y)$, from (20) we have $y \perp x$, i.e. $\rho_{xy}=0$. However, y is dependent on z, i.e. $\rho_{yz}=1$. Also, from (21) we have $\rho_{xy.z} \neq 0$, since $\alpha>0$. This results in Markov blanket $Bl(y)=\{x,z\}$ characteristic of the motif (15) for suitable choice of parameter α* characteristic of the motif (15).
- *For $Bl(z)$, from (20) we have $z \perp x$,i.e. $\rho_{xz}=0$. However, z is dependent on y, i.e.$\rho_{yz}=1$. Also, from (21) $\rho_{xy.z} \neq 0$, since $\alpha>0$. These results in turn result in $Bl(z)=\{x,y\}$ for suitably large values of α.*

Asymmetry between the Markov blankets is observed across $Bl(x)$ and $Bl(y)$ as well as between $Bl(x)$ and $Bl(z)$. This can be attributed to the fact that $x \in Bl(y)$ while $y \notin Bl(x)$ and $x \in Bl(z)$ while $z \notin Bl(x)$ for suitably large values of α. Correcting this asymmetry by adding y and z to $Bl(x)$ results in the Markov blankets characteristic of the motif. However, establishing their directions is not possible since the presence of an arc between x and y prevents both GS and IAMB from identifying $x \rightarrow y \leftarrow z$. As a result, all possible configurations of the arcs' directions are probabilistically equivalent resulting in an undirected graph. This is phenomenon is known as the *shielded collider* identification problem and affects all constraint-based learning algorithms [37].

2.3 Simulation Results

In the following discussion, the three gene network motifs are generated using (1, 8, 15) with parameter $(\alpha = 0.5)$ and normally distributed noise. Since the objective is to demonstrate the impact of noise as opposed to the other parameter,$(\alpha = 0.5)$, is held constant across all the simulations. The noise variance at the node x is fixed at unit variance whereas those at $\mathrm{y}(\gamma_1 > 0)$ and $\mathrm{z}(\gamma_2 > 0)$ are varied systematically in order to understand the impact of discrepancy in noise variance on the conclusions. Three distinct cases of noise variances, namely: $(\gamma_1 = 1, \gamma_2 = 1)$, $(\gamma_1 = 10, \gamma_2 = 1)$ and $(\gamma_1 = 1, \gamma_2 = 10)$ are considered. The cases $(\gamma_1 = 10, \gamma_2 = 1)$ and $(\gamma_1 = 1, \gamma_2 = 10)$ correspond to large noise variance limits as discussed under (Cases 1, 2 and 3) whereas $(\gamma_1 = 1, \gamma_2 = 1)$ corresponds to absence of discrepancies in noise variance. The conditional independence tests used in the following discussion is exact t-test for Pearson correlation as implemented in the R package bnlearn [38]. A description of the functions in bnlearn can be found in the accompanying manual with applications to molecular expression profiles in [39].

Results generated using constraint-based structure learning algorithms GS and IAMB were quite similar consistent with their expected behaviour, Section 2.2. Therefore, we discuss only the results from the GS algorithm. The networks were learned across 200 independent realizations of the data (sample size $= 2000$) and Friedman's *confidence* (ψ) [3] was computed for each of the edges. Friedman's confidence essentially represents the percentage of times an edge shows up across networks learnt independently from bootstrapped realizations. In the case of observational data sets, confidences are estimated from networks learned from nonparametric bootstraps of the given empirical sample. In the present study, the underlying model generating the networks is known a priori. Therefore, parametric bootstrap is used where independent realizations of the data were generated from the model in contrast to non-parametric bootstrap [40]. Also, in the present study, confidence estimates of edges known to be present in the given graph a priori essentially represent their *statistical power*. As a rule of thumb [3], edges with confidence at least $(\psi \geq 0.8)$ were deemed significant. In a recent study [41], we proposed a noise floor approach in order to avoid the ad-hoc choice of ψ, and subsequently a statistically motivated approach that estimates optimal ψ from the cumulative distribution of the confidence values [42]. However, in the present study the actual confidence values are presented for enhanced clarity.

Common-effect network motif. The common-effect network motif, Fig. 1a, was generated using (1) with $(\alpha = 0.5)$ and normally distributed noise $(\epsilon_t, \eta_t, \delta_t)$. For finite and equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ at (y,z) the correlation coefficients ρ_{xz}, ρ_{yz} were similar and relatively higher than ρ_{xy} (~ 0) as expected (2), Fig. 2a. In order to investigate the impact of large discrepancies in the noise variances, the noise variance across y was increased relative to $\mathrm{z}(\gamma_1 = 10 \gg \gamma_2 = 1)$. This resulted in small values of ρ_{xz}, ρ_{xy} relative to ρ_{yz}, Fig. 2a and resembled (6) as expected. A similar analysis with $(\gamma_1 = 1 \ll \gamma_2 = 10)$ across y and z resulted in small correlation coefficients across the board similar to (4), Fig. 2a. Therefore, large discrepancies in noise variances across the nodes can have a pronounced effect on the pair-wise dependencies. The corresponding partial correlations for the three choices of noise variance (γ_1, γ_2) are shown in Fig. 2d. For finite equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ at (y,z), the partial correlation $\rho_{xy.z} < 0$ (3) was non-zero in contrast to $\rho_{xy} = 0$, rendering the marginally independent nodes (x,y) dependent. Increasing the noise variance across y relative to z $(\gamma_1 = 10 \gg \gamma_2 = 1)$ resulted in a significant increase in $\rho_{yz.x}$ (7) whereas for $(\gamma_1 = 1 \ll \gamma_2 = 10)$, all the conditional dependencies were rendered negligible (5) preventing any reliable conclusion of the network structure, Fig. 2d. For finite equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ at (y,z), GS was able to faithfully retrieve the structure of the common-effect motif, Fig. 3a. Increasing the noise variance across $\mathrm{y}(\gamma_1 = 10 \gg \gamma_2 = 1)$ relative to z, also retrieved the structure faithfully, Fig. 3c. However, increasing the noise variance on the common effect node $\mathrm{z}(\gamma_1 = 1 \gg \gamma_2 = 10)$ resulted in low confidence values of the edges challenging any reliable inference of the network, Fig. 3b. Thus the magnitude of the noise variance at the nodes can have a pronounced effect on constraint-based structure learning of a common-effect network motif.

Three-chain network motif. The three-chain network motif, Fig. 1b, was generated using (8) with $(\alpha = 0.5)$ and normally distributed noise $(\epsilon_t, \eta_t, \delta_t)$. For finite and equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ at the nodes (y,z)the correlation coefficients $\rho_{xy}, \rho_{xz}, \rho_{yz}$ were significant as expected (9) with ρ_{xz} representing the transitive dependency between x and z, Fig. 2b. In order to investigate the impact of large noise variance, the noise variance on the mediating node y was increased relative to z $(\gamma_1 = 10 \gg \gamma_2 = 1)$. This resulted in small values of ρ_{xy}, ρ_{xz} relative to ρ_{yz} (13) similar to what was observed for the common-effect network motif (6) failing to distinguish these structures. On the other hand, large noise variance on the terminal node z relative to y $(\gamma_1 = 1 \ll \gamma_2 = 10)$ resulted in ρ_{xy} values relatively higher than that of ρ_{xz} and ρ_{yz}, as expected from Fig. 2b. These results clearly demonstrate the non-trivial impact of noise strengths on network inference on pairwise dependencies. Partial correlations $\rho_{xy.z}$ and $\rho_{yz.x}$ for finite equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ were considerably higher than that of $\rho_{xz.y}(\tilde{0})$ as expected, since conditioning on the mediator y should render marginally dependent nodes (x,z) independent. Increasing the noise variance at y relative to z $(\gamma_1 = 10 \gg \gamma_2 = 1)$ and at z relative to y $(\gamma_1 = 1 \gg \gamma_2 = 10)$, rendered the pairwise and conditional dependencies similar. This is reflected by the similar profiles, Figs. 2b and 2e respectively. For finite equal noise variance $(\gamma_1 = 1, \gamma_2 = 1)$ at (y,z), GS was able to faithfully retrieve the underlying undirected graph, Fig. 3d. This is to be expected since the Markov equivalent structure of the three-chain network motif is the undirected graph $(\mathrm{x} - \mathrm{y} - \mathrm{z})$. Increasing the noise variance across the mediator y relative to $\mathrm{z}(\gamma_1 = 10 \gg \gamma_2 = 1)$ resulted in low confidence along $(\mathrm{y} - \mathrm{z})$ preventing any reliable inference of possible association between these nodes, Fig. 3e. Interestingly, increasing the noise variance on the terminal node z relative to $\mathrm{y}(\gamma_1 = 1 \ll \gamma_2 = 10)$ resulted in low confidence along $(\mathrm{x} - \mathrm{y})$ preventing any reliable inference of possible association between these nodes, Fig. 3f.

Coherent Type-I feed-forward loop network motif. The coherent Type-I feed-forward loop network motif, Fig. 1c, was

generated using (15) with ($\alpha = 0.5$) and normally distributed noise ($\epsilon_t, \eta_t, \delta_t$). While one part of the Type-I FFL resembles the common-effect motif (x→z←y), the other part resembles a three-chain (x→y→z), Fig. 1b. For finite and equal noise variance ($\gamma_1 = 1, \gamma_2 = 1$) at (y,z) the pairwise (16) and conditional dependencies (17) were non-zero. Increasing the noise variance across yrelativetoz($\gamma_1 = 10 \gg \gamma_2 = 1$) resulted in pairwise (20) identical to those of the common-effect (6) and three-chain motifs (13) failing to distinguish these network structures. This is reflected by similar profiles across Figs. 2a, 2b and 2c. On a related note, increasing the noise variance across z relative to y($\gamma_2 = 10 \gg \gamma_1 = 1$) resulted in pairwise (18) and conditional dependencies (19) identical to those of the three-chain motif (11, 12) failing to distinguish these two distinct network structures. Similarities in the pairwise and conditional dependencies across these motifs are also reflected by similar profiles between Figs. 2b and 2c and between Figs. 2e and 2f respectively. For finite equal noise variance ($\gamma_1 = 1, \gamma_2 = 1$) at (y,z) GS was able to retrieve the undirected edges (x−y−z), Fig. 3g. Failure to retrieve the exact structure, Fig. 1c, can be attributed to the presence of equivalent classes. Increasing the noise variance across z relative to y($\gamma_1 = 1 \gg \gamma_2 = 10$) resulted in low confidences along (x−z) and (y−z) relative to (x−y) preventing any reliable inference of possible associations along (x−z) and (y−z), Fig. 3h. Thus for these choices of noise variances it is possible the results of GS for Type 1 FFL resembles the structure of the three-chain failing to distinguish them. In contrast, increasing the noise variance at y relative to z($\gamma_1 = 10 \ll \gamma_2 = 1$) resulted in large edge confidence only along y→z and x→z with low edge confidence along (x−y) Fig. 3i preventing any reliable inference of the network structure.

2.4 Application to Molecular Expression Profiles

$$\begin{bmatrix} Plc\gamma_t \\ PIP3_t \\ PIP2_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \alpha_1 & 0 & 0 \\ \alpha_2 & \alpha_3 & 0 \end{bmatrix} . \begin{bmatrix} Plc\gamma_t \\ PIP3_t \\ PIP2_t \end{bmatrix} + \begin{bmatrix} \gamma_0 . \epsilon_t \\ \gamma_1 . \eta_t \\ \gamma_2 . \delta_t \end{bmatrix} \quad (22)$$

In a recent study [7], signalling mechanisms between 11 molecules were inferred from single-cell data using flow-cytometry in conjunction with Bayesian network structure learning algorithms. The resulting network was shown to validate existing associations as well as discovering novel undocumented associations. Of interest, was the sub-network consisting of three molecules ($PIP2, PIP3, Plc\gamma$) weakly connected to the rest of the molecules in the network (see Fig. 3 in [7]). The network structure inferred from the molecular expression data between these three molecules ($PIP2, PIP3, Plc\gamma$) consisted of the following directed edges $PIP3 \rightarrow PIP2$, $Plc\gamma \rightarrow PIP3$ and $Plc\gamma \rightarrow PIP2$. A quick inspection would reveal the resemblance of the relationships between these three molecules (22) to that of coherent Type-I FFL motif (Fig. 1c, Case 3) discussed earlier. The expected and the inferred relationships along with the influence paths for these three molecules can be found in (Table 3, Sachs et al., 2005). While the authors acknowledged that the directionality between ($Plc\gamma \rightarrow PIP3$, *recruitment leading to phosphorylation*) inferred from the data was opposite to that established in the literature [43] (see Supplementary Material, Table I, Sachs et al., 2005), they successfully validated ($PIP3 \rightarrow PIP2$, *precursor-product*) and ($Plc\gamma \rightarrow PIP2$, *direct hydrolysis to IP3*) [44,45] (see Supplementary Material, Table I, Sachs et al., 2005). While several data sets were

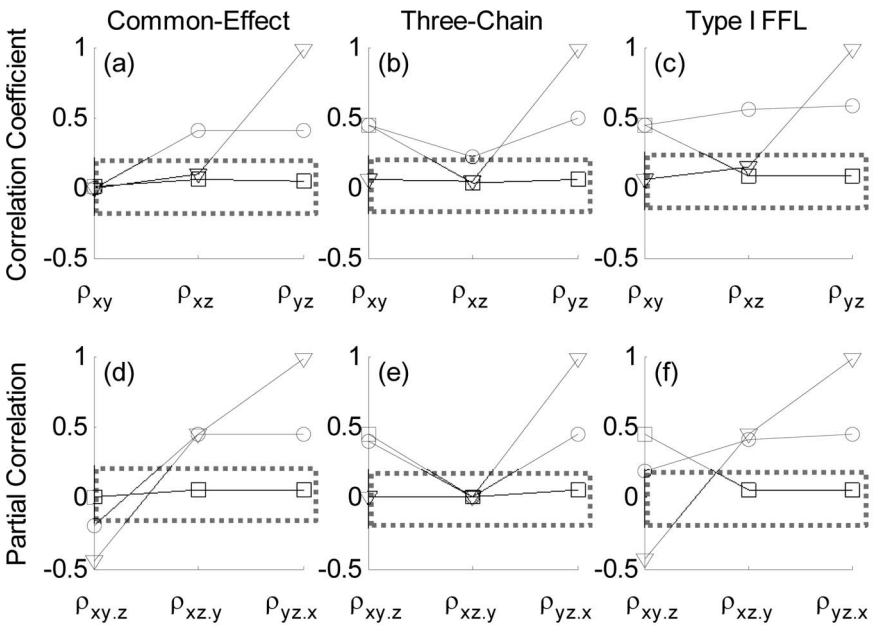

Figure 2. The average correlation coefficient and partial correlation estimates across 200 independent realizations of the common-effect, three-chain and coherent Type I feed-forward loop network motifs for various choices of noise variances (γ_1, γ_2) are shown in (a, d), (b, e) and (c, f) respectively. The *x*-axis labels correspond to the correlation coefficients $(\rho_{xy}, \rho_{xz}, \rho_{yz})$ in (a, b, c) and partial correlations $(\rho_{xy.z}, \rho_{xz.y}, \rho_{yz.x})$ in (d, e, f) respectively. The (circles, squares and triangles) in each of the subplots correspond to noise variances with magnitudes ($\gamma_1 = 1, \gamma_2 = 1$), ($\gamma_1 = 1, \gamma_2 = 10$) and ($\gamma_1 = 10, \gamma_2 = 1$) respectively. The points bounded by the dotted rectangle represent cases that occurred much lesser than 80% of the time as significant ($\alpha^* = 0.001$) across 200 independent realizations.
doi:10.1371/journal.pone.0080735.g002

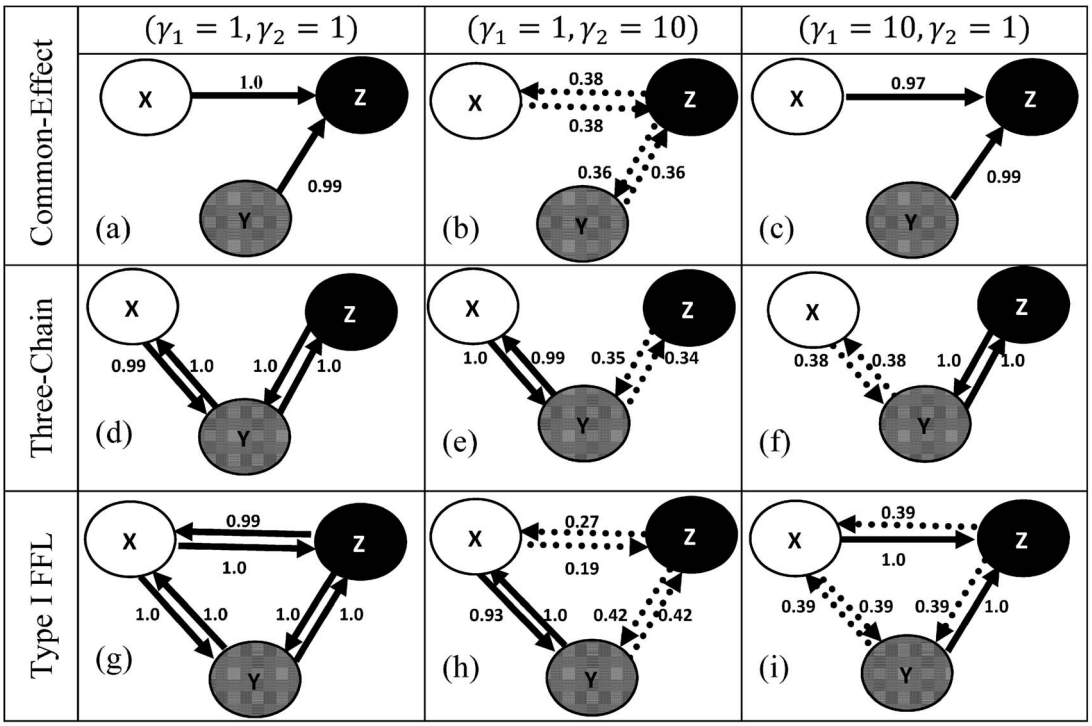

Figure 3. Bayesian networks inferred using Grow-Shrink algorithm along with Pearson correlation ($\alpha^* = 0.01$) for the three-gene network motifs, namely: common-effect (a-c), three-chain (d-f) and coherent type-I feed-forward loop (g-i) for various choices of noise variances: $(\gamma_1 = 1, \gamma_2 = 1)$,$(\gamma_1 = 1, \gamma_2 = 10)$, **and** $(\gamma_1 = 10, \gamma_2 = 1)$**.** The confidences of the edges (ψ) are represented as percentage of the edges that persisted across 200 independent realizations. Edges with $(\psi \geq 0.80)$ are shown by solid arrows whereas others $(\psi \leq 0.80)$ are shown by dotted arrows. Edges with confidence $(\psi \leq 0.05)$ are deemed noisy and excluded for clarity.
doi:10.1371/journal.pone.0080735.g003

investigated in [7], we restrict the present study to the unperturbed data set comprising the expression of $(PIP2, PIP3, Plc\gamma)$ across 853 single cells. Prior to investigating the impact of noise on the network inference between the three molecules, we found the distribution of the expression levels across the single-cells to be positively skewed, indicating large variations in the expression estimates across the cells. Interestingly, we also found the variance in the expression levels proportional to their average value across the molecules $(PIP2, PIP3, Plc\gamma)$. Box-Cox [46] transforms are widely used in literature to minimize the skew in the distribution and suppress non-constant variance as a function of magnitude. In the present study, we used the log-transform which is the limiting case of the classical Box-Cox transform to minimize the skew in the distribution of the expression across these three molecules. Therefore, the results across the raw as well as the log-transformed data are presented.

Three different networks $(\Pi_k, k = 1,2,3)$ were investigated. Π_1 : Network inferred from the given data; Π_2 : Network inferred from data generated from the linear model (22) fit to the given data without any constraints on the model parameters; Π_3 : network inferred from data generated by the linear model fit (22) to the given data with constraint on the noise variance to be equal $(i.e. \gamma_0 = \gamma_1 = \gamma_2)$. The above exercise was repeated for the raw as well as the log-transformed protein expression data and the corresponding edge confidences were estimated. The approach is outlined below.

- **Step 1**: Given the expression $X_{n\times 3}$ of the three molecules across $n = 853$ cells.
- **Step 2**: Generate independent realizations $X^i_{m\times 3}, i = 1 \ldots p$ by resampling $X_{n\times 3}(m < n)$ with replacement. In the present study, we set $(m = 800, p = 200)$. Set each column in $X^i_{m\times 3}$ to zero-mean.
- **Step 3**: Set $i \leftarrow 1$.
- **Step 4**: Infer the network structure from $X^i_{m\times 3}$ using the GS algorithm. Let the resulting network be Π^i_1.
- **Step 5**: Estimate the parameters (i.e. regression coefficients and noise variances $(\gamma_0, \gamma_1, \gamma_2)$ by fitting the linear model (22) to $X^i_{m\times 3}$. Generate $Y^i_{m\times 3}$, using the estimated model parameters and zero-mean i.i.d. noise terms $(\epsilon_t, \eta_t, \delta_t)$ sampled from a log-normally distributed noise to accommodate for the positive-skew in the distribution. Infer the network structure from $Y^i_{m\times 3}$ using the GS algorithm. Let the resulting network be Π^i_2.
- **Step 6**: Generate data $Z^i_{m\times 3}$, using the linear model in Step 5 with the additional constraint on equal noise variance $(i.e. \gamma_1 = \gamma_0; \gamma_2 = \gamma_0)$ in (22). Infer the network structure from $Z^i_{m\times 3}$ using the GS algorithm. Let the resulting network be Π^i_3.
- **Step 7**: Set $i \leftarrow i + 1$.
- **Step 8**: Repeat Steps 4–7 till $i > p$.
- **Step 9**: Estimate the confidences of the edges for each of the networks $(\Pi_k, k = 1,2,3)$.
- **Step 10**: Repeat Steps 1–9 for the log-transformed data with normally distributed noise as opposed to log-normally distributed noise in Steps 5 and 6.

Raw Data. The networks $(\Pi_k, k = 1,2,3)$ inferred using the raw data for the molecules $(PIP2, PIP3, Plc\gamma)$ are shown in

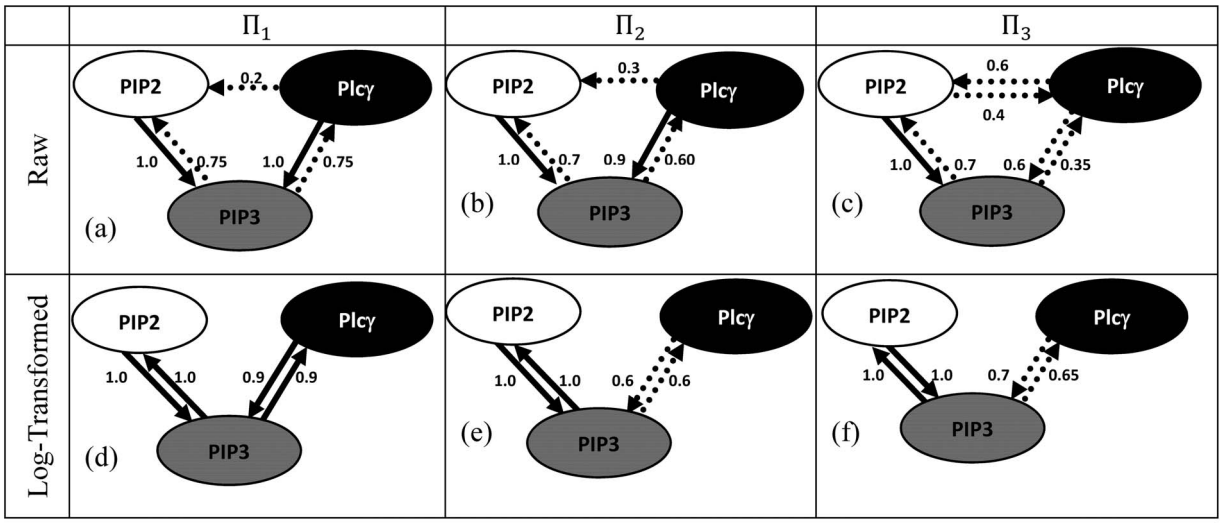

Figure 4. Bayesian networks inferred using Grow-Shrink algorithm from the molecular expression data (PIP2, PIP3, Plcγ) with sample-size 800 and Pearson correlation ($\alpha^* = 0.01$) are shown in (a–f). Confidences estimated from 200 independent bootstrap realizations are shown along the edges. Edges with ($\psi > 0.80$) are shown by solid arrows whereas others ($\psi \leq 0.80$) are shown by dotted arrows. Edges with confidence ($\psi \leq 0.05$) are deemed noisy and excluded for clarity. The edge confidences of the networks (Π_1, Π_2, Π_3) inferred from the raw data are shown in (a), (b) and (c) respectively. Those inferred on the log-transformed data are shown in (d), (e) and (f) respectively.
doi:10.1371/journal.pone.0080735.g004

Figs. 4a–4c respectively. Network structures inferred from the raw data (Π_1, Step 4) and those of the linear model fit (Π_2, Step 5) exhibited considerable similarity as reflected by their edge confidences, Figs. 4a and 4b. The confidence was high along $PIP2 \rightarrow PIP3$ and $Plc\gamma \rightarrow PIP3$, and markedly low along $Plc\gamma \rightarrow PIP2$, Figs. 4a, 4b. Noise variance estimated from the linear model fit (Step 5) of the raw data revealed around a two-fold difference $\left(i.e. \frac{\gamma_2}{\gamma_1} \sim 2.7 \pm 0.3\right)$. Constraining the noise variance to be equal ($i.e. \gamma_1 = \gamma_2 = \gamma_0$) had a marked effect on the resulting network (Step 6) Π_3, Fig. 4c. The edge confidences were considerably high along $PIP2 \rightarrow PIP3$ as seen earlier (Π_1 and Π_2), Figs. 4a, 4b. However, relatively smaller edge confidence along between ($Plc\gamma$,$PIP3$)along either directions, Fig. 4c, in contrast to Figs. 4a or 4b was also observed. More importantly, constraining the noise variance also increased the edge confidences between ($Plc\gamma$,$PIP2$) along either directions in contrast to those shown in Figs. 4a and 4b (i.e. Π_1,Π_2). Thus forcing the noise variance to be equal had a pronounced effect on the inferred network.

Log-transformed Data. In order to minimize the impact of skewness on the conclusions, the entire exercise was repeated on the log-transformed data. The resulting networks along with confidence of the edges are shown in Figs. 4d–4f. The networks ($\Pi_k, k = 1,2$) inferred from the log-transformed data (Π_1, Step 4) and those from data generated on the linear model fit (Π_2, Step 5) along with the edge confidences are shown in Figs. 4d–4e respectively. The noise variance estimates from the linear model fit to the log-transformed data revealed no marked difference $\left(i.e. \frac{\gamma_2}{\gamma_1} \tilde{1}.1 \pm 0.01\right)$ in contrast to what was observed in the raw data. Since there were no marked discrepancies in noise variance, forcing the noise variance to be equal ($i.e. \gamma_1 = \gamma_2 = \gamma_0$) had no profound effect on the resulting network (Π_3, Step 6) Fig. 4f as expected. This was revealed by the similar edge confidences across Π_2 and Π_3. Furthermore, it is important to note that the networks ($\Pi_k, k = 1,2,3$) inferred from the log-transformed data unlike those from raw data, failed to capture any relationship $Plc\gamma$ and $PIP2$.

Discussion

Real-world entities work in concert as a system and not in isolation. Associations between such entities are usually unknown. Inferring associations and network structure from data obtained across the entities is of great interest across a number of disciplines. The recent surge of high-throughput molecular assays in conjunction with a battery of algorithms has facilitated validating established associations while discovering new ones with the potential to assist in novel hypothesis generation. These associations and networks have been shown to capture possible causal relationships under certain implicit assumptions and proven to be useful abstractions of the underlying signaling mechanism. Such an understanding can provide system level insights and often precedes developing meaningful interventions. Several network inference algorithms have been proposed in literature including those that depend on pairwise and conditional dependencies. However, little attention has been given to the impact of possible discrepancies in noise variance across the data obtained across the molecular entities. In molecular settings, such discrepancies can be attributed to several factors including inherent stochastic mechanisms, heterogeneity in cell populations, variations in abundance of the molecules, variation in binding affinities, sensitivity of the measurement device and other experimental artifacts. Understanding the discrepancies in noise variance is critical in order to avoid spurious conclusions and an important step prior to identifying the source of the noise.

The present study clearly elucidated the non-trivial impact of discrepancies in noise variance on associations and network inference algorithms across synthetic as well as experimental data. The impact of large discrepancies in noise variance on associations and network structure inferred from data generated using linear models of popular network motifs and fundamental connections as well as those from experimental protein expression profiles were investigated. Analytical expressions and simulations were presented elucidating the non-trivial impact of noise on three popular molecular network motifs and fundamental connections (common effect, three-chain and coherent Type-I feed-forward loop). It was

shown that discrepancies in noise variance can significantly alter the results of pairwise dependencies, conditional dependencies as well as constraint-based Bayesian network structure learning techniques that implicitly rely on tests for conditional independence. As expected, the discrepancies in noise variances was found to result in markedly different topologies from those of their noise free counterpart challenging reliable inference of the underlying network topology. Such discrepancies were also shown to result in spurious conclusion of similar structures across markedly distinct network topologies. The impact of discrepancies in noise variance were also investigated on publicly available single-cell molecular expression profiles of a sub-network comprising of three molecules (PIP2, PIP3, Plcγ) involved in human T-cell signaling. The sub-network shared considerable resemblance to the coherent Type-I feed-forward loop. The distribution of the raw expression estimates across these three molecules was positively skewed indicating large variations in the expression estimates across the single-cells. Variance about the average expression across the three molecules was found to be markedly different and proportional to their average values. Several factors can contribute to such discrepancies including: abundance of these molecules, antibody binding characteristics, uncertainty due to possible overlap in the wavelengths corresponding to the colors tagged to the molecules. In the present study, a linear model was fit to the molecular expression data. Parameter estimates from the linear model indicated significant discrepancies in the noise variances across the molecules. Adjusting for these discrepancies in the model was shown to significantly affect the edge confidences of the resulting networks, hence the topology. The results were presented on the raw molecular expression data as well as its log-transformed counterpart. As expected, log-transforming the data not only reduced the positive skew of the expression profile but also rendered the noise variance estimates comparable across the molecules. However, the networks inferred using the log-transformed data were considerably different from those inferred on the raw data. While identifying the source of the variation and controlling for the same prior to the network inference may be the long-term goal and a research problem in its own merit, understanding the impact of discrepancies in noise variance is a critical step in this direction. While the present study focused on simple network motifs comprising of three molecules, the concerns are likely to be aggravated across more complex network topologies. The analytical treatment provided in the present study has the potential to be translated across other setting such as genome-wide association studies (GWAS) [47]. Unlike the molecular network motifs investigated in this study, GWAS investigate the impact of causal genes and variants on a given trait or set of traits. Similar to the concerns presented in the present study, discrepancies in biological variances across the traits is not uncommon and can have a pronounced effect in discerning the relationship between the causal and the traits. However, given the intricacies accompanying GWAS studies a more detailed investigation is required.

Author Contributions

Conceived and designed the experiments: RN. Performed the experiments: RN. Analyzed the data: RN MS. Contributed reagents/materials/analysis tools: RN MS. Wrote the paper: RN MS.

References

1. Butte A, Tamayo P, Slonim D, Golub TR, Kohane IS (2000) Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proc Nat Acad Sci USA 7(22): 12182–12186.
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. (2000) Molecular portraits of human breast tumours. Nature 406: 747–752.
3. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian Network to Analyze Expression Data. J Comp Biol 7: 601–620.
4. Gardner TS, di Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301: 102–105.
5. Lucas PJF (2004) Bayesian networks in biomedicine and health-care. Artificial Intelligence in Medicine 30, 201–214.
6. Pe'er D (2005) Bayesian network analysis of signaling networks: a primer, Sci. STKE 281, pl4.
7. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science 308(5721): 523–529.
8. Opgen-Rhein R, Strimmer K (2007) From Correlation to Causation Networks: a Simple Approximate Learning Algorithm and its Application to High-Dimensional Plant Gene Expression Data. BMC Sys Biol 1: 1–37.
9. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network Motifs: Simple Building Blocks of Complex Networks. Science 298: 824–827.
10. Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2): 101–113.
11. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, et al. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotech 27(2): 199–204.
12. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Nat Acad Sci USA 95(25): 14863–8.
13. Verma TS, Pearl J (1991). Equivalence and Synthesis of Causal Models. Uncertainty in Artificial Intelligence 6: 255–268.
14. Pearl J (2000) *Causality: Models, Reasoning, and Inference*, Cambridge University Press.
15. Pearl J (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
16. McAdams HH, Arkin A (1999) It's a noisy business! Genetic regulation at the nanomolar scale. Trends in Genetics 15(2): 65–9.
17. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic Gene Expression in a Single Cell. Science 297(5584): 1183–6.
18. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 6: 451–464.
19. Okoniewski MJ, Miller CJ (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinf 7: 276.
20. Steen HB (1992) Noise, Sensitivity and Resolution of Flow Cytometers. Cytometry 13: 822–830.
21. Welch CM, Elliott H, Danuser G, Hahn KM (2011) Imaging the coordination of multiple signalling activities in living cells. Nat Rev Mol Cell Biol 12: 749–756.
22. Nagarajan R (2009) A note on inferring acyclic network structures using Granger causality tests. Int. J. Biostatistics 5(1): 10.
23. Nagarajan R, Upreti M (2010) Granger causality analysis of human cell-cycle gene expression profiles. Stat Appl Genet Mol Biol 9(1): 31.
24. Nagarajan R, Upreti M (2011) Inferring functional relationships and causal network structure from gene expression profiles. Meth in Enzymol 487: 133–46.
25. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31(1): 64–8.
26. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Nat Acad Sci *USA* 100: 11980–11985.
27. Jensen FV (2001). Bayesian Networks and Decision Graphs. Springer-Verlag.
28. Friedman N, Nachman I, Pe'er D (1999) Learning Bayesian Network Structure from Massive Datasets: The Sparse Candidate Algorithm. Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI-99). 206–215.
29. Spirtes P, Glymour C, Scheines R (2000) *Causation, Prediction and Search*. MIT Press.
30. Margaritis D (2003). Learning Bayesian Network Model Structure from Data. Ph.D. thesis, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA. Available as Technical Report CMU-CS-03–153.
31. Tsamardinos I, Aliferis CF, Statnikov A (2003a) Algorithms for Large Scale Markov Blanket Discovery. In Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference, 376–381.
32. Tsamardinos I, Aliferis CF, Statnikov A (2003b) Time and Sample Efficient Discovery of Markov Blankets and Direct Causal Relations. In KDD '03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 673–678.
33. Meek C (1995). Causal Inference and Causal Explanation with Background Knowledge. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95). 403–410.
34. Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsokos XD (2010a) Local Causal and Markov Blanket Induction for Causal Discovery and Feature

Selection for Classification Part I: Algorithms and Empirical Evaluation. Journal of Machine Learning Research 11: 171–234.
35. Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsokos XD (2010b) Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part II: Analysis and Extensions. Journal of Machine Learning Research 11: 235–284.
36. Tsamardinos I, Aliferis CF, Statnikov A, Brown LE (2003c). Scaling-Up Bayesian Network Learning to Thousands of Variables using Local Learning Techniques. Technical Report DSL 03–02, 2003, DBMI, Vanderbilt University.
37. Castillo E, Gutiérrez J, Hadi AS (1997). *Expert Systems and Probabilistic Network Models*. Springer-Verlag.
38. Scutari M (2010) Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software 35(3): 1–22.
39. Nagarajan R, Lebre S, Scutari M (2013) Bayesian Networks in R: with applications in Systems Biology. Springer-Verlag, NY.
40. Efron B, Tibshirani R (1993). *An Introduction to the Bootstrap*. Chapman & Hall.
41. Nagarajan R, Datta S, Scutari M, Beggs ML, Nolen GT, et al. (2010). Functional Relationships Between Genes Associated with Differentiation Potential of Aged Myogenic Progenitors. Frontiers in Physiology 1(21): 1–8.
42. Scutari M, Nagarajan R (2013) Identifying significant edges in graphical models of molecular networks. Artif Intell Med 57(3): 207–17.
43. Alberts B (2002) *Molecular biology of the cell*, Garland Science, New York.
44. Sofroniew MV, Howe CL, Mobley WC (2001) Nerve growth factor signaling, neuroprotection and neural repair. Ann Rev Neurosci 24: 1217–81.
45. Lee SB, Rhee SG (1995) Significance of PIP2 hydrolysis and regulation of phospholipase C isozymes. Curr Opin Cell Biol 7: 183–9.
46. Box GEP, Cox DR (1964) An analysis of transformations. J Royal Stat Soc Series B 26 (2): 211–252.
47. Li Y, Tesson BM, Churchill GA, Jansen RC (2010) Critical reasoning on causal inference in genome-wide linkage and association studies. Trends in Genet 26(12): 493–498.