1

RECENT DEVELOPMENT IN METHODOLOGY FOR GENE NETWORK PROBLEMS AND INFERENCES

Sung W. Han and Hua Zhong

Division of Biostatistics, School of Medicine, Department of Population Health, New York University, New York, NY, USA

1.1 INTRODUCTION

The cell inside of a human body is similar to a manufacturing system producing an appropriate protein that functions according to the specific organ or the part of the body to which it belongs. The nucleus centered at the cell contains the DNA sequence, which is a designed map for the human body. Each time the cell produces a protein, it duplicates a certain part of the DNA sequence and generates mRNA sequences. This is called a transcription process. After leaving the nucleus, the mRNA is attached to a ribosome, and the ribosome interprets the code in mRNA. This is called a translation process. After interpretation, the ribosome generates a sequence of amino acids; then it is folded into a certain type of protein.

The manufacturing system from DNAs to proteins sometimes malfunctions due to the DNA damage, which is known to be a main cause of cancers, also called malignant neoplasms [1, 2]. The DNA damage can occur naturally, but the damage can also be caused by two groups of agents: (i) exogenous agents such as radiation, smoke [3], ultraviolet light [4], and viruses [5]; and (ii) endogenous agents such as diet [6] and macrophages/neutrophils [5]. Such DNA damage leads to epigenetic alteration

Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, First Edition.
Edited by Hui Yang and Eva K. Lee.

for DNA repair genes, which play the key roles in preventing cancer cell growth. Reducing the DNA repair gene expression (DNA repair deficiency; [7]) or switching off the function of the DNA repair gene, called silence, finally leads to the development of cancers. For example, MGMT is the DNA repair gene, and most types of colorectal cancers have reduced MGMT expression ([8–11], and [12]). The following are other examples of proteins corresponding to DNA repair genes [1].

- BRCA1 and BRCA2 (breast cancer genes 1 and 2) for breast and ovarian cancers.
- ATM (ataxia telangiectasia mutated) for leukemia and breast cancers.
- XPC (xeroderma pigmentosum) for skin cancers.
- p53 (Li–Fraumeni syndrome) for sarcoma, leukemia, breast, lung, skin, pancreas, and brain cancers.

In addition, the miRNA (micro RNA) outside of the nucleus is known to have an effect on the DNA repair gene because it can reduce the expression of DNA damage response genes or repair genes [1]. For example, miRNA-155 is overly expressed in colon cancers, and it is known to reduce the expression of MLH1, a DNA repair protein [13].

For finding the mechanism of cancer development, understanding the causal relationship in transcriptional regulatory networks is important, and the related inference is often based on the gene network problem. The examples of the application of the network problem are in gene expression analysis or gene–gene expression networks [14–19], protein–protein interaction analysis [20, 21], phenotype networks utilizing gene expression information [22–24], and causal networks linking gene expression and metabolic change [24].

The probabilistic graphical modeling is a popular approach to find causal relationships between variables in cell signal pathways or gene networks [25]. In this chapter, the graphical models are assumed to be directed acyclic graphs (DAGs), in which all the edges are directed edges and contain no cycles [26]. Since the estimation of DAGs is computationally very challenging, we cannot simply apply approaches that are used to estimate undirected graphs [27–29]. First, DAGs with the same set of conditional independence are not identifiable from observational data alone [26]; this is called observational equivalence. Second, the number of possible DAGs exponentially increases as the number of nodes increases [27]. Third, in gene network problems, the number of genes is much larger than the sample size, which is called high-dimensional data.

The DAGs with conditional probability distribution for each child node given its parents are called Bayesian networks. The comprehensive review about learning Bayesian network is in Buntine [30, 31], Heckerman [32], Neapolitan [33], and Daly et al. [34]. Apart from cancer gene problems, the Bayesian network is used in broad applications such as ecology [35, 36], neuroscience [37, 38], distributed sensor networks for change detection, and diagnosis [39–41].

The main approaches to estimate the Bayesian networks are as follows: (i) a score-and-search approach through the space of Bayesian network structures,

(ii) a constraint-based approach that uses conditional independencies identified in the data, and (iii) a hybrid approach. A score-and-search approach is to find a structure corresponding to a good score function value [42] and use a heuristic algorithm to find the solution. The examples of this approach are in Daly et al. [34]. A constraint-based approach is to use a statistical test of conditional independence on the data. One of the efficient methods is the PC algorithm [43]. In high-dimensional contexts, Kalisch and Buhlmann [44] proposed the PC algorithm with a reasonable computational time [43] and proved consistency for sparse DAGs. Hybrid search strategies including the above-mentioned two criteria have also been proposed such as in Tsamardinos et al. [45], where the method used is a Max–Min Hill-Climbing (MMHC) algorithm. The methods mentioned have been successfully proposed to estimate DAGs with a small to moderate number of nodes.

For the score-and-search approach, a network is identified by maximizing a certain score function [31, 33, 42, 46], and several heuristic search algorithms are then developed to find a high score [27, 34]. To overcome high dimensionality in gene expression data, the L1-penalized method or lasso approach has been recently developed. Meinshausen and Buhlmann [28] theoretically show that the neighborhood of a node corresponding to a conditional dependence set can be obtained by a lasso problem, and it is efficient for high-dimensional DAGs. For DAGs, Shojaie and Michailidis [29] used the L1-penalized likelihood with a structural equation model to estimate directed graphs with a known variable order and found that such a problem was transformed into separable subproblems with lasso penalty. Huang et al. [47] used a penalized linear regression that imposes penalties to the coefficient values as well as to acyclic constraints. Fu and Zhou [48] used an adaptive lasso-based score function when the variable order is unknown. However, their objective function without the acyclic constraint is nonconvex, which makes finding the optimal solution infeasible. Han et al. [49] proposed the adaptive lasso-based score function, and it demonstrated superior performance to other methods when the network has a hub structure. In this chapter, we overview the approach based on the lasso-type score function for gene network problems in high-dimensional data.

1.2 BACKGROUND

We explain the basic theoretical background in probabilistic graphical modeling or Bayesian networks. Let us have p random variables, $Y_1, Y_2, \ldots, Y_p$, and the variables have causal relationships with each other. The variables and relationships in probabilistic distribution need to be mapped to p nodes, V, and edge sets, $E(\subset V \times V)$. In other words, the separation in a graph needs to be mapped to the independence in probability [50].

In probabilistic graphical modeling, the d-separation (directed separation) is an important concept described by Pearl [26]. The definition of d-separation is complicated, but it implies the following argument. Suppose we have three node sets V_1, V_2, and V_3. We define that V_2 is a d-separate between V_1 and V_3 if one of the conditions is satisfied:

- All edges between V_1 and V_2 inflow from V_1 to V_2, and all edges between V_2 and V_3 inflow from V_2 to V_3.
- All edges between V_1 and V_2 inflow from V_2 to V_1, and all edges between V_2 and V_3 inflow from V_3 to V_2.
- All edges between V_1 and V_2 inflow from V_2 to V_1, and all edges between V_2 and V_3 inflow from V_2 to V_3.

For all disjoint subsets of V_1, V_2, and V_3, we state that the probability distribution P is faithful to the graph G if the following condition is satisfied.

V_1 and V_3 are independent given V_2 if and only if V_1 and V_3 are d-separated given V_2.

Based on the d-separation, we can express the probability distribution by using the Markov property. The probability distribution $f(Y)$ is represented by

$$f(Y) = \prod_{i=1}^{p} f(Y_i|\text{Pa}(Y_i))$$

where $\text{Pa}(Y_i)$ is a set of parents for Y_i.

Another important issue in probabilistic graphical model is observational equivalence. The example of observational equivalence is in Figure 1.1. The three cases in Figure 1.1a–c are not distinguishable based on observational data. They are said to be in one equivalence class. However, based on the data, the case in Figure 1.1d can be distinguished from the other three cases. We say that this case has a v-structure. Such equivalence class causes multiple solutions with the same score function values if we apply the score-and-search approach to estimate a DAG. To show all equivalence classes, the complete partial DAG (cpDAG) can be used, which can be implemented by the "essentialGraph()" function in R package [51].

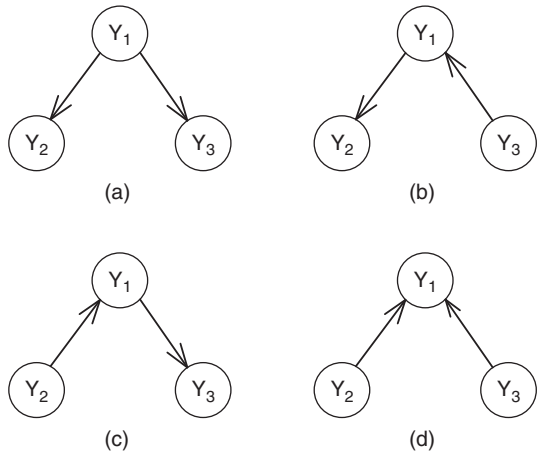

Figure 1.1 Examples of observational equivalence.

1.3 GENETIC DATA AVAILABLE

The technology in recent decades has allowed genome-wide monitoring of DNA and RNA levels on thousands of samples [52]. For example, The Cancer Genome Atlas (TCGA) project seeks to provide a comprehensive landscape of genetic and genomic alternations by profiling DNA copy number, mRNA expression, and miRNA expression for about 20 cancer types. The Genotype-Tissue Expression (GTEx) project studies human gene expression regulation and its relationship to genetic variation. The goal of these projects is to understand global regulation including genetic (from DNA to RNA), transcriptional (among mRNAs), and posttranscriptional processes determining normal cell physiology. Thus, the analysis of these regulatory layers can provide a useful picture of the underlying processes.

Open sources of gene expression data including TCGA and GTEx, or pathway information are in the following link:

- The Cancer Genome Atlas (TCGA) Data (https://tcga-data.nci.nih.gov/tcga/): This data set provides high-level sequence analysis of the tumor genomes and clinical information. The TCGA data consist of several types and levels as follows: Copy Number Variation (Low Pass DNASeq), Copy Number Variation (SNP Array), DNA Methylation, Expression Protein, METADATA, miRNASeq, RNASeqV2, and Somatic Mutation.
- The GTEx project (http://www.broadinstitute.org/gtex/): The objective of this project is to accumulate the comprehensive data of gene expression across multiple tissues in the human body.
- modENCODE Project (Model Organism ENCyclopedia Of DNA Elements) (http://www.genome.gov/modencode/): This project aims to create the data of a comprehensive encyclopedia of genomic functional elements in the model organisms.
- Pathway Interaction Database (http://pid.nci.nih.gov/index.shtml): This has the maps of biomolecular interactions and cellular processes organized into human signaling pathways. It was a collaborative work between Nature Publishing Group (NPG) and National Cancer Institute (NCI).
- The DREAM5 Network Inference Challenge (http://wiki.c2b2.columbia.edu/dream/index.php/D5c4): This website provides gene expression data, which have been obtained from microorganisms.

1.4 METHODOLOGY

In this section, we explain a recently developed method based on the lasso-type score function. The first two sections describe the model and the score function for the graphical model to estimate the gene network problem, and the next two sections explain technical details.

1.4.1 Structural Equation Model

We express the genes by random variables. Denote by Ψ the $n \times p$ data matrix, where n is the sample size and p is the number of the variables. We assume that an edge is directed, so (j, i) is not in E if (i, j) belongs to E. The causal relationship of random variables in a DAG can be represented by the structural equation model [26, 29, 53]. Let Z_i be a latent variable, which is assumed to follow independent normal distributions. Then, the structural equation model that represents the relationship is

$$Y_i = \sum_{j \in \mathrm{Pa}_i} c_{ij} Y_j + Z_i \tag{1.1}$$

where c_{ij} is a causal effect from a parent j to a child i. Z_i s are latent variables. Denote $Y = [Y_1, Y_2, \ldots, Y_{\mathrm{p}}]^T$, and $Z = [Z_1, Z_2, \ldots, Z_{\mathrm{p}}]^T$. Here, we assume that the latent vector Z follows the multivariate normal distribution, $\mathrm{MN}(0, \Gamma)$, where $\Gamma = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2]^T$. Under the unknown variable order, we represent coefficients of Y_j s, c_{ij}, by the coefficient matrix C, where

$$C = \begin{pmatrix} 0 & c_{12} & \cdots & c_{1(p\text{-}1)} & c_{1p} \\ c_{21} & 0 & \cdots & c_{2(p\text{-}1)} & c_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{(p\text{-}1)1} & c_{(p\text{-}1)2} & \cdots & 0 & c_{(p\text{-}1)p} \\ c_{p1} & c_{p2} & \cdots & c_{p(p\text{-}1)} & 0 \end{pmatrix}$$

Thus, c_{ij} is the (i, j)th entry of C, and Equation 1.1 can be rewritten by $Y = CY + Z$. In addition, if the variable order is partially known, the blockwise matrix can be used [53], which is represented by

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & & 0 & 0 & \cdots & 0 & 0 \\ c_{1U_1} & 0 & \cdots & 0 & 0 & & 0 & c_{12} & \cdots & c_{1(p\text{-}1)} & c_{1p} \\ 0 & c_{2U_1} & \cdots & 0 & 0 & & c_{21} & 0 & \cdots & c_{2(p\text{-}1)} & c_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{(p\text{-}1)U_{p\text{-}1}} & 0 & & c_{(p\text{-}1)1} & c_{(p\text{-}1)2} & \cdots & 0 & c_{(p\text{-}1)p} \\ 0 & 0 & \cdots & 0 & c_{pU_p} & & c_{p1} & c_{p2} & \cdots & c_{p(p\text{-}1)} & 0 \end{pmatrix}$$

Then, the corresponding structural equation model is

$$Y_i = c_{iU_i} U_i + \sum_{j \in \mathrm{Pa}_i} c_{ij} Y_j + Z_i$$

where U_i is an upper-level variable, which regulates the down-level variable Y_i. The example of the upper-level variable is DNA copy number or miRNA, which regulates mRNA.

1.4.2 Score Function Formulation

In the gene network problem, the number of gene expressions is very large, for example, over 10,000, but the sample size is relatively small such as 200–300. Thus, the estimation of gene networks is related to the variable selection problem in high-dimensional data. The well-known method for variable selection uses a discrete manner. One approach is a best subset selection. However, if the number of variables becomes large, the subset selection is computationally infeasible [54]. In addition, due to the discreteness of the approach, the result of the subset selection is unstable [55, 56]. Similarly, most discrete-based approaches may show instability in finding a solution. However, the L1-penalized linear regression leads to a continuous search, and it gives stable and robust estimation [54, 57]. Thus, for high-dimensional data, the lasso approach proposed in Fu and Zhou [48] and [53] is a proper way. In addition, they use adaptive lasso, which can break the equivalence class, and give a high probability of estimating a correct solution among the solutions in the equivalence class. In this section, we discuss a lasso-type approach as a recent development of methodology.

The lasso-type score function can be derived from the L1-penalized log likelihood, which is

$$-\frac{2}{n}\log\prod_{i=1}^{p} f(Y_i|\mathrm{Pa}(Y_i), C) + \mathrm{Pe}(C)$$

where $\mathrm{Pa}(Y_i)$ is the set of parent variables for Y_i, and $\mathrm{Pe}(C)$ is a penalty function for the coefficient matrix C. Based on the L1-penalized log likelihood, Fu and Zhou [48] proposed the score function, which is

$$\min_{C}\sum_{k=1}^{p}\left[\log\left(\frac{1}{n}\|\Psi_k - \Psi C_k\|^2\right) + \lambda\sum_{j=1}^{p} w_{kj}|c_{kj}|\right] \tag{1.2}$$

subject that the estimated graph is acyclic. In the score function shown as Equation 1.2, Ψ is a $n \times p$ data matrix and Ψ_k is a data vector as the kth variable. C_k is a coefficient vector, $[c_{k1}, c_{k2}, \ldots, c_{k(p\text{-}1)}, c_{kp}]^T$, which is a column vector representing the kth row in matrix C. Zou [54] mentioned that the original lasso without the weight term does not provide consistent estimates. Thus, they suggested the adaptive lasso, which used the weights for the coefficient term c_{kj}. Zou [54] suggested $w_{kj} = \left(\frac{1}{|\beta_{kj}|}\right)^{\delta}$, where β_{kj} is the least square estimate from the ordinary least square, while Fu and Zhou [48] suggested $w_{kj} = \min\left(\frac{1}{|\beta_{kj}|}, 10^4\right)^{\delta}$.

Han et al. [49] proposed another type of score function:

$$\min_{C}\sum_{k=1}^{p}\left[\frac{1}{n}\|\Psi_k - \Psi C_k\|^2 + \lambda\sum_{j=1}^{p} w_{kj}|c_{kj}|\right] \tag{1.3}$$

subject that the estimated graph is acyclic. The notations are the same as in Equation 1.2. The score function in Equation 1.3 is the equivalent form of Shojaie and Michailidis [29], which provides the score function with a known variable order. For the value of the weights, Han et al. [53] used the scaled weight, $w_{kj} = \min\left(\frac{0.0001}{|\beta_{kj}|}, 1\right)^{\delta}$. Han et al. [53] represented the acyclic constraint by the optimization formulation, which is

$$|c_{kj}| \leq T_{kj} \tag{1.4}$$

$$T_{kj} = 0 \text{ or } 1 \tag{1.5}$$

and

$$\begin{gathered} T_{i_1 i_2} + T_{i_2 i_1} \leq 1 \text{ for all } i_1 \text{ and } i_2 \\ T_{i_1 i_2} + T_{i_2 i_3} + T_{i_3 i_1} \leq 2 \text{ for all } i_1,\ i_2,\ \text{and } i_3 \\ \vdots \\ T_{i_1 i_2} + T_{i_2 i_3} + \cdots + T_{i_p i_1} \leq p\text{-}1 \text{ for all } i_1,\ i_2, \ldots, i_p \end{gathered} \tag{1.6}$$

The score functions proposed by both Fu and Zhou [48] and Han et al. [53] have an advantage. The adaptive lasso score function can break observational equivalence, which expresses the same score function values [51]. If the structure of the graph is a hub network, there exist multiple solutions that give the same objective function value, called score equivalence [51]. Most score-and-search approaches such as penalized likelihood or constraint-based approach such as the PC algorithm cannot distinguish the solutions in the equivalence class since they give the same score function values or the same p-value, respectively. However, the adaptive lasso provides a different score function value to each solution even though they are in the same equivalence class. Furthermore, it gives a high probability of selecting a correct solution among the equivalence class especially in the hub network [53].

However, the score function from Fu and Zhou [48] has several disadvantages in comparison with that from Han et al. [53]. First, the score function from Fu and Zhou [48] in Equation 1.2 used the residual sum of the square from the penalized likelihood to estimate the variance of latent variables, which becomes a log form in the score function. The mean square error from the penalized linear regression is known to be a biased estimate of the variance of the latent variables if the penalty is large. Thus, from the various simulation studies, Han et al. [53] showed that when the penalty is large, the score function from Equation 1.3 gave a higher true positive of edges than the score function from Equation 1.2 given the same false positive, which indicates that the score function from the former gives better performance than the score function from the latter. However, as the penalty becomes small, the performance of the two methods becomes similar. The performance of Equation 1.2 is sensitive to the value of the penalty parameter, but that of Equation 1.3 is robust. Based on the simulation studies with $p = 100$, $n = 500$, and density $= 2$ (average number of parent nodes per child node), the receiver operating characteristic (ROC) curves based on true positive

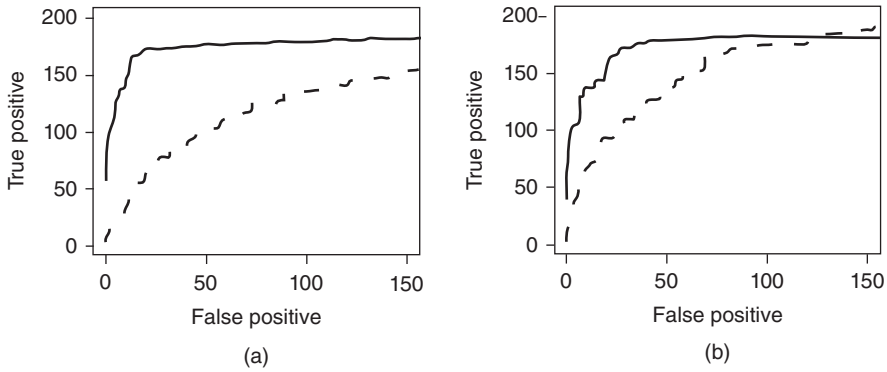

Figure 1.2 Receiver operating characteristic (ROC) curves from the adaptive lasso methods: (a) random network and (b) hub network.

versus false positive is in Figure 1.2a and b. The black solid line indicates the curves from the score function in Equation 1.3, and the gray dash line indicates the curves from the score function in Equation 1.2.

Second, the score function in Equation 1.2 is nonconvex. Thus, Fu and Zhou [48] can find only a local optimal solution at best. Both the nonconvexity of the score function and the acyclic constraints make finding a good solution very hard. In addition, Fu and Zhou [48] proposed a block coordinate descent (BCD) algorithm to obtain a solution from Equation 1.2 to overcome such nonconvexity, but they did not justify the quality of the solution. However, the score function in Equation 1.3 is convex, and it can be transformed into the quadratic programming (QP) problem, which guarantees a global optimal solution. Han et al. [53] implemented the DIST (discrete improving search with TABU list) algorithm to find the solution from Equation 1.3, which is called the cLasso method. They showed that the method based on Equation 1.3 requires smaller computational time than that based on Equation 1.2.

Apart from normal distributions, the lasso framework is easily extended to other distributions. In particular, due to recent technology, some gene expressions such as RNA sequences are recorded in discrete count data, and Poisson log-normal distribution can be used for the RNA sequence measurements. By the assumption of the log-normal data, Han et al. [53] represented the observed compounded Poisson data by

$$Y_k \sim \text{Poisson}(\exp(\sigma_k X_k + \mu_k))$$

where σ_k and μ_k are the standard deviation and average related to the marginal normal variable. In this case, we only observe the count data Y_k, and X_k is treated as unobserved data. Han and Zhong [58] proposed a penalized likelihood score function as follows. Let X be an unobserved normal vector with the X_k in the kth entry. Based on Bayes theorem,

$$P(C|\mathrm{Y}) = \int P(C, X|\mathrm{Y})dX \propto \int P(Y|X, C) \times P(X|C)dX$$

The penalized log likelihood can be represented by

$$-\frac{2}{n}\sum_{j=1}^{n}\log\int \exp[h(X|C,y_j)]dX + \mathrm{Pe}(C)$$

where $h(X|C,y_j) = \log P(\,y|X,C) + \log P(XC)$, and $\mathrm{Pe}(C)$ is a penalty function in terms of the coefficient matrix C. Finding a solution based on the score function is quite complicated. Based on several approximations, they first transformed the score function to the lasso framework, and then searched the solution by using two iterative optimization procedures based on two groups of parameters.

1.4.3 Two-Stage Learning

Since searching for the solution in entire space takes a lot of computational time, much literature uses the two-step learning technique. The first step is to find the potential parents of each child by estimating an undirected graph/skeleton/Markov blanket, and the second step is to identify directionality (or parents). Tsamardinos et al. [45] proposed the MMHC algorithm, which estimates the skeleton by the constraint-based method and identifies the directionality by a score-and-search algorithm. Schmidt et al. [59] used the lasso regression to estimate an undirected graph, and then used a permutation approach based on swapping adjacent variable orders to identify directionality. Neto et al. [22] first estimated an undirected graph, and then estimated a DAG without an acyclic constraint based on the likelihood ratio between one direction and the opposite direction of each edge. Pellet and Elisseeff [60] used feature selection algorithms [61] to estimate a Markov blanket and then identified a directionality based on v-structured patterns. Other examples of the two-stage learning approaches are the Sparse Candidate (SC) algorithm [62] and the Grow-Shrink algorithm [63]. The two-step learning approach is also called a hybrid algorithm in Nagarajan et al. [50].

Huang et al. [47] mentioned that the two-stage procedure has high risk of misidentification of the true parents. If the first stage missed a true parent, the parent is not considered again as a possible parent in the second stage. Thus, they proposed a combined score function, which is

$$\min_{C}\sum_{k=1}^{p}\left[\frac{1}{n}\|\Psi_k - \Psi C_k\|^2 + \lambda_1\sum_{j=1}^{p}|c_{kj}| + \lambda_2\sum_{k\neq j}|c_{kj}\times P_{jk}|\right]$$

where P is a $p\times p$ matrix with 1 in the (i,j) entry if there is a direct path from X_i to X_j, otherwise 0. The acyclic constraint, $\sum_{k\neq j}|c_{kj}\times P_{jk}|$, is plugged into the score function with a second penalty, which is similar to Lagrangian relaxation. They used the BCD algorithm to estimate the network structure.

Han et al. [53] proposed an alternative approach. First, without the acyclic constraints, they minimized the objective function (score function) to find the infeasible solution. They set the infeasible solution as the potential parent set per child. Then,

they added the acyclic constraints back to the problem and found the feasible solution. Unlike other hybrid approaches, Han et al. [53] used only one model and score function, but split the solution search algorithm by making it the two-step procedure. Han et al. [53] argued that their approach is more stable that the existing two-stage procedure.

1.4.4 Further Issues

Robustness of the estimation in Bayesian networks for gene network problems is an important issue. The performance of Bayesian networks also depends on the network structure. Han et al. [53] showed that the method with the score function based on Equation 1.3 is robust in terms of the network structure, but the performance of the pc-type algorithm is very sensitive to the network structure. Another issue related to the robustness is that the data are sometimes not complete and have missing parts. Ramoni and Sebastiani [64] discussed robust learning for Bayesian networks with missing data. In addition, most approaches for learning Bayesian networks show point estimator for the causal relations. Thus, several papers have studied how much confidence can be placed in the network estimate [65–67]. Finally, to estimate the correct structure, all variables with causal relationships are assumed to be included before learning.

Another issue in the estimation of the structure of DAGs based on the lasso problems is how to choose the penalty. To minimize average prediction errors, the cross-validation (CV) is suggested since it gives an asymptotic optimal estimate of the penalty. However, the CV does not lead to a consistent model selection for the lasso-type penalty [68]. For the consistency of the model selection, the Bayesian Information Criterion measures are suggested [69, 70]. Tibshirani [57] showed that the lasso gives a stable and correct estimate. Proper tuning parameters need to be selected for a consistent model selection [54, 71, 72]. How to select the best penalty parameter in the network problem is still an open question.

1.5 SEARCH ALGORITHM

In this section, we discuss solution search algorithms to minimize the score function.

1.5.1 Global Optimal Solution Search

The straightforward way to find a global optimal solution is to use total enumeration, which is an exhaustive search in terms of edges by permuting an entire combination of directed edges and selecting the combination that gives the maximum of the score function. For example, suppose that the number of variables is p, then the total number of edges we need to consider is $p(p–1)/2$. Each edge can take three options: one direction, the opposite direction, or an empty edge. Then, the total number of combinations we need to consider is $3^{(p(p–1)/2)}$. However, such approaches are computationally not feasible even when the number of variables is small.

We may use some optimization technique to estimate a DAG based on the traveling salesman problem (TSP) [73, 74]. The straightforward approach [53] to find a global optimal solution is as follows:

- Step 1: Solve the objective function (1.3) without the acyclic constraints.
- Step 2: If cycles exist, add the constraints corresponding to the cycles. Then, solve it again.
- Step 3: If the solution does not have any cycle, it is an optimal solution. Otherwise, go to Step 2.

However, to solve the problem with the above-mentioned approach is also computationally inefficient. For example, the asymmetric TSP [73, 74] presented computational difficulty for long decades [75]. Thus, directly solving the gene network problem requires heavy computational time if p increases.

Another approach described by Han et al. [53] is to find upper and lower bounds of the solution based on the branch-and-bound technique and sub-tour elimination, which are used in the TSP. Since it is almost impractical to include all possible cycles in the formulation in the solution approach, we add acyclic constraints if needed. During the branch-and-bound algorithm, the intermediate solution can be infeasible since it violates the acyclic constraints. For the minimization problem, the solution should be lower bound since it is the optimal solution under the relaxed conditions. Then, we can obtain upper bound of the solution that breaks all cycles. Han et al. [53] implemented the branch-and-bound technique to find the optimal solution, but the computational time is over 2 h even when $p = 20$. Thus, in high-dimensional data, using a heuristic algorithm or meta-heuristic algorithm is preferable.

1.5.2 Heuristic Algorithm for a Local Optimal Solution

To find the solution for the problem with nonconvexity of the acyclic constraints, the ad hoc or rudimentary solution search algorithms have been proposed. Shojaie and Michailidis [29] proposed permutation of variable order since known order indicates directionality between two genes and automatically satisfies acyclic constraints. Such one-time permutation relies on random chance to obtain the corrected order, so it is not a reliable approach. Schmidt et al. [59] swapped adjacent ordered variables and searched the best score function values. Fu and Zhou [48] used BCD algorithm. They started the algorithm from the empty network and added edges or reverse directionality of edges step-by-step while avoiding a cycle. However, they did not consider a leaving edge, which is a necessary step to find a good solution. In the optimization area, a greedy algorithm or Hill-Climbing algorithm is commonly used, which considers adding an edge, reversing an edge, or removing an edge.

Most ad hoc or heuristic algorithms only guarantee a local optimal solution due to acyclic constraints. Han et al. [53] proposed a meta-heuristic algorithm based on a discrete improving search with a TABU list, which is called DIST. In the DIST algorithm, at each iteration, a leaving edge is put in the TABU list and it is not considered as an entering edge in the whole round. Han et al. [53] compared the solution from

the DIST algorithm with a global optimal solution when $p = 10$ or 20, and he showed that the solution from the DIST algorithm is close to the global optimal solution. The description of the DIST algorithm is as follows:

- Select an entering edge among all unselected edges, which gives the most improvement in the objective function.
- Select a leaving edge among all selected edges, which breaks the cycle and gives the least harm in the objective function. To find any cycle, use forward and backward Breadth-First Search (BFS) algorithms. Any confirmed leaving edges are put into the TABU list.
- Update the intermediate solution as well as the objective function values.
- Do the above-mentioned three steps repeatedly until all edges are searched. We call it a round. After each round, empty the TABU list and restart the next round.
- If there is no improvement in the objective function value in two consecutive rounds, stop the algorithm. Otherwise, keep running the next round.

1.6 PC ALGORITHM

The well-known alternative method to estimate Bayesian networks is the PC algorithm [43]. Kalisch and Buhlmann [44] applied the PC algorithm to estimate the DAGs with high-dimensional data. It first estimates an undirected graph and removes edges iteratively based on the conditional independence test. This method requires a condition of faithfulness, which is mentioned in Section 1.2. Under the faithfulness condition, the d-separation in a graph is equivalent to the conditional independence relationship. In the multivariate normal distribution, the conditional independence can be derived from partial correlation. Kalisch and Buhlmann [44] used Fisher's Z-transform statistics to test the significance of a partial correlation, which is defined by

$$Z(X, Y|Z) = \frac{\sqrt{n\text{-}\#(Z)\text{-}3}}{2} \log \frac{1 + \rho_{XY|Z}}{1\text{-}\rho_{XY|Z}}$$

If the Z-transform statistics is greater than a certain threshold, it rejects $c_{kj} = 0$, and it adds an edge (i,j) in the graph. After estimating the skeleton, they extended it to a complete partial DAG.

The PC algorithm has been used in high-dimensional data [76–79], and it is known to be computationally feasible to estimate sparse network with the large number of variables. There is an R package in the software, "pcalg()" [44]. Since the performance of those methods is order dependent, which means that they are sensitive to the true variable order, Colombo and Maathuis [80] developed the PC-stable algorithm.

Han et al. [53] compared the method based on the adaptive lasso score function with the PC-stable algorithm by simulation studies. The results showed that the former outperforms the latter in terms of true positive given the same false positive in most scenarios except for the random network with high density. Under the hub

network, which is a main structure in gene networks, the adaptive lasso approach performs better than the PC-stable method. Han et al. [53] found that the performance of the PC-stable method is sensitive to the network structure, and as the percentage of isolation of parents increases, the performance of the PC-stable method decreases. Thus, for the gene network problem, which is related mostly to a hub network, the adaptive lasso might be preferable to the PC-stable method.

The computational complexity of the PC algorithm is approximately $O(p^q)$, where q is the maximal size of neighbors and p is the number of variables. On the other hand, the computational complexity of the cLasso method is bounded by $O(\max(nk^2p, np^2))$, where k is the maximum number of parent candidates and n is the sample size. As the complexity shows, if p or k is moderate such as 2, 3, or above, the complexity of the PC algorithm is higher than that of the cLasso method. Han et al. [53] also verified the complexity difference based on the simulation study. Under the random network and $d = 1$, the computational time of the PC algorithm is similar to that of the cLasso method. However, under the hub network or the random network with $d = 2$, the computational time of the PC algorithm is much higher than that of the cLasso method. Especially, under the hub network, from which many possible parents can be derived, with $d = 2$, the computational time of the PC algorithm is much higher than that of the cLasso method.

1.7 APPLICATION/CASE STUDIES

We discuss three examples of the data-driven analytics for gene network problems. In the first example, which is based on open data set of melanoma skin cancers in TCGA Data Portal, we describe how to construct gene networks under partially known variable order and obtain inferences for clinical purposes. In the second example based on Cancer Cell Line Encyclopedia (CCLE) data, we show how to build the network under unknown variable order. In the third example based on flow cytometry data of protein expressions, we describe some tutorials of how to use R softwares to estimate the network.

1.7.1 Skin Cutaneous Melanoma (SKCM) Data from the TCGA Data Portal Website

The incidence rate of melanoma skin cancer increased recently with a lifetime risk of 1 in 50 [81, 82]. The patients with metastatic tumors survive about 7 months on average [83]. Although several therapies such as BRAF mutant kinase inhibitors [84, 85] or CTLA-4 inhibitors [86, 87] have been developed, the results of therapy have not been successful. Fleming et al. [88] found that based on array-based screening, about 18 miRNAs out of 358 miRNAs are potential predictors of recurrence. They revealed that based on statistical analysis of TCGA data, the signature miRNAs regulate functions to melanoma biology such as the immune signaling pathway. In this section, we describe how to construct gene pathways or networks under partially known structure based on Skin Cutaneous Melanoma (SKCM) data from the TCGA Data Portal website (https://tcga-data.nci.nih.gov/tcga/). The TCGA Data Portal provides a

platform of data sets containing clinical information as well as genomic data and sequence analysis of various tumors.

The description of the data structure is as follows. The TCGA data consist of several data types (so-called platform types) and levels. The data types are as follows: CNV (Low Pass DNASeq), CNV (SNP Array), DNA Methylation, Expression Protein, METADATA, miRNASeq, RNASeqV2, and Somatic Mutations. Each type has the data in four different levels at most. The data in level 1 are raw data or signals, and the data in level 2 are processed data. The level 3 data are segmented or interpreted with normalization, and the level 4 data are summary or regions of interest data. For our purpose, we use level 3 data.

After downloading the data sets, the file, "file_manifest.txt," showed the main information of data type and sample ID/barcode as well as the corresponding file name containing genomic data. For example, in the file, "file_manifest.txt," there are three columns: "Platform Type," "Sample," and "File Name." It is useful to note that in the sample barcode, the first 10 letters, for example, "TCGA-BF-A1PU," indicate a patient barcode. If we want to find miRNA and mRNA data for a certain sample, say "TCGA-BF-A1PU-01A," we need to search the file name in the "File Name" column, which is matched to a certain platform type and sample barcode. For example, for miRNA data, the file name matching to "miRNASeq" and "TCGA-BF-A1PU-01" is "TCGA-BF-A1PU-01A-11R-A18V-13.mirna.quantification.txt." In the miRNA data file, there are raw count data and normalized count data for each miRNA ID, and we used the normalized count data. The detailed description of the structure of all data sets is in https://tcga-data.nci.nih.gov/tcga/tcgaDataType.jsp.

The clinical information for each sample barcode is in the following folder and file: Clinical>Biotab>biospecimen_sample_skcm.txt. Each tumor sample is matched to a normal sample. The number at the end of the sample barcode is the key to understand the tumor source. For example, the ending number of the sample barcode is 1A for a primary tumor and 06A for a metastatic tumor. The clinical information for each patient is in the folder and file: Clinical>Biotab>clinical_patient_skcm.txt.

We extracted the data of miRNA and mRNA corresponding to primary or metastatic tumors. The number of miRNAs is 1046, and the number of mRNAs is 20,531. The sample size in the miRNA data set is 329, but the sample size in the RNASeqV2 (mRNA) data set is 325. Thus, we removed four samples in the miRNA data that did not match the data set of mRNA. Next, we took a log transformation of each data set and then standardized them by centering and scaling. In the miRNA data, 200 miRNAs have zero values except for at most one value. In the mRNA data, 405 mRNAs have zero values except for at most one value. Thus, after removing unnecessary miRNAs and mRNAs, the total numbers of remaining miRNAs and mRNAs are 846 and 20,126, respectively.

Given the data set, we tried to estimate the gene network with the causal relationship from miRNA to mRNA. We used the lasso-type score function in Equation 1.3. Since we are interested in the causal relationship between the miRNA group and the mRNA group, we use a block matrix for C in Equation 1.3. The penalty parameter is selected based on empirical study in Fleming et al. [88]. The estimated gene network is shown in Figure 1.3, which is drawn by network() and plot() functions

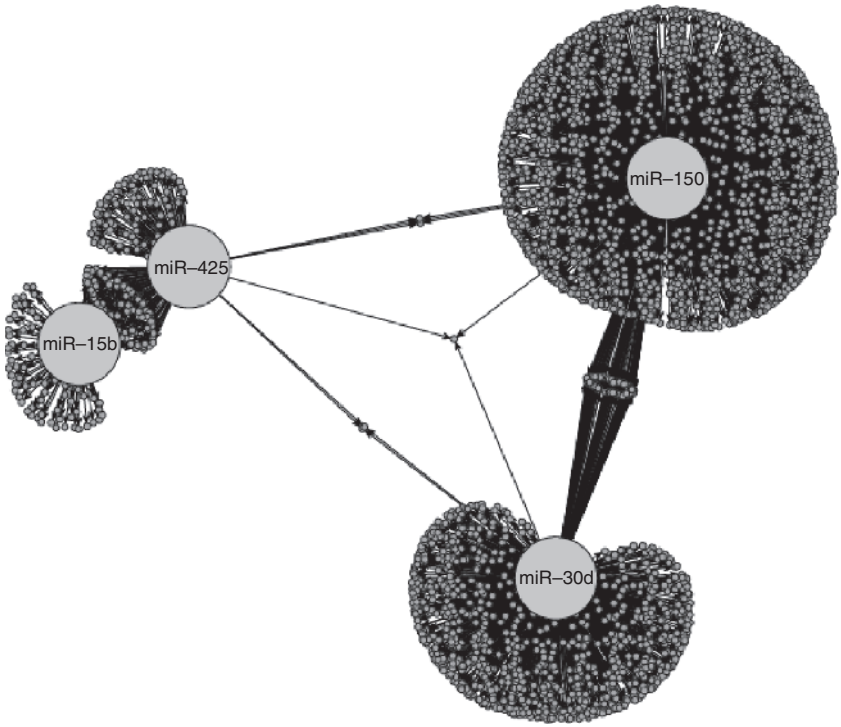

Figure 1.3 Estimated graph: the small dots indicate mRNAs and the large dots indicate miRNAs.

in R software. In this figure, the four large dots indicate miRNA and the small dots indicate mRNAs. Based on the result, the miRNA-150 has a broad effect on many mRNAs, so we included the network related to miRNA-150. We also include the network plot with clinically important miRNA such as miRNA-30d, miRNA-425, and miRNA-15b [88]. After deciding the miRNAs, we selected the mRNAs with significant coefficients to the corresponding miRNAs, the absolute values of which are greater than 0.1. The miRNA-150 was causally related to 1360 mRNAs with high coefficients. A number of selected mRNAs with respect to the other three important miRNAs (miRNA-30d, miRNA-425, and miRNA-15b) are 185, 5, and 15, respectively.

From the result of gene network estimation, we draw inferences for clinical purposes. Since the miRNA-150 gives the main impact for most mRNA, we make a grouping for patients based on mRNAs affected by the miRNA-150. We draw a heat map as shown in Figure 1.4. In the heat map, the rows indicate patient samples and the columns indicate mRNA expression. As shown at the left side in the heat map, the patients can be clustered by two groups such as a low-level group (bottom part) and a high-level group (upper part) based on hierarchical cluster analysis. For each clustered group, we estimate Kaplan–Meier survival curves that are shown in Figure 1.5. The gray dash line indicates the high level group, and the black solid line indicates

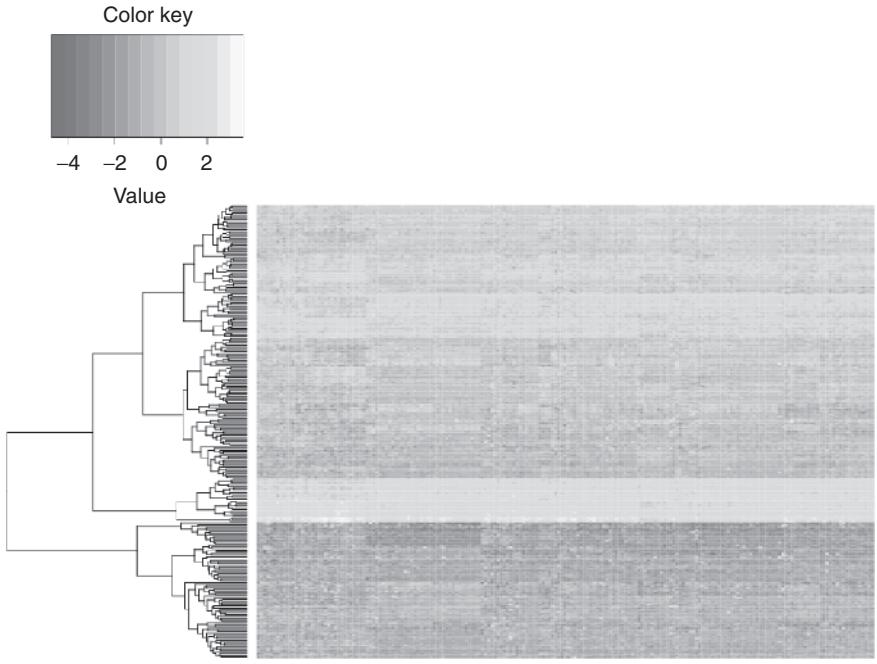

Figure 1.4 Heat map: the rows indicate patient samples and the columns indicate mRNA expression.

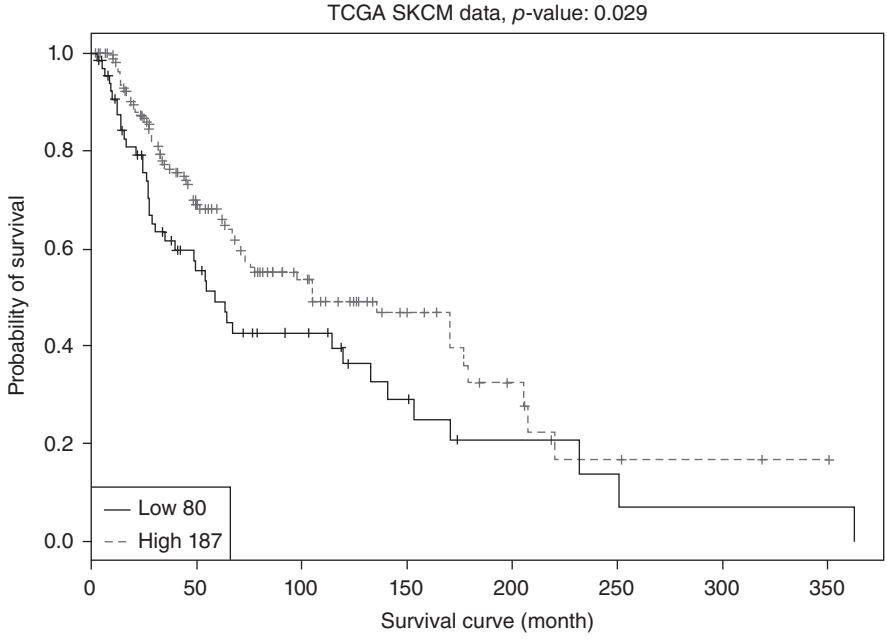

Figure 1.5 Survival curve: the gray dash line indicates the high-level group, and the black solid line indicates the low-level groups.

the low level groups. The p-value based on the log-rank test is 0.029, which is less that 0.05 (significance level in biostatistics). This shows that the high-level group has a better survival rate than the low-level group with the statistical significance.

1.7.2 The CCLE (Cancer Cell Line Encyclopedia) Project

It is widely known that p53 is an important gene as a tumor suppressor, which can prevent cancer growth. This p53 controls genes, whose function is DNA repair, cell cycle control, and apoptosis. Under normal conditions, the protein level of p53 drives MDM2 transcription factor, and as a feedback loop, MDM2 regulates the protein levels of p53. MDM2 can block or impair p53 pathway activity. Zhong et al. [89] explained the development of gene expression signature, which is predictive of the response to MDM2 antagonist therapy. Barretina et al. [90] discuss finding a predictive model for anticancer drug sensitivity from CCLE data.

The CCLE contains cancer-related genomic data from 947 human cancer cell lines with 36 tumor types, which are characterized by genomic technology platforms. This data can be downloaded from http://www.broadinstitute.org. To identify mRNAs as the cell line responses to the MDM2, we need to estimate the gene networks. By using a partial model of Equation 1.3, we extracted 16 genes, which have roles in different processes surrounding MDM2. These genes were BRCA1, DDB2, and XPC involved in DNA repair; FDXR in P53-related process; SESN1 and ZMAT3 in cell growth; CCNG1 and CDKN1A in cell cycle arrest; and BAX, EP300, PIDD, RPS27L, and TP53 involved in apoptosis. To estimate the gene network, we applied the method in Equation 1.3 to the data with a penalty parameter as suggested in Han et al. [53]. The estimated network is plotted in Figure 1.6.

1.7.3 Cellular Signaling Network in Flow Cytometry Data

In this section, we provided a tutorial of the usage in R packages to estimate a DAG in a gene network problem. The common example used to explain a network problem is cell signaling in protein expression as in Sachs et al. [91]. Intracellular multicolor flow cytometry can generate observational data of cell signaling molecules [92, 93], and by flow cytometry, we can measure the protein expression and protein modification states such as phosphorylation [93–95]. Based on this finding, Sachs et al. [91] demonstrated that cell signaling data can be used for inferring causal relationships. The multivariate flow cytometry data used in Sachs et al. [91] are collected to identify the effects of different conditions on the intracellular signaling networks of human primary naive CD4+ T cells, which are downstream of CD3, CD28, and LFA-1 activation. Eleven phosphorylated proteins were measured: PKC, PIP2, PLGG, PIP3, AKT, JNK, P38, ERK, MEK, RAF, and PKA.

In this section, we show the examples of R codes for four different estimates for the cellular signaling network and the corresponding network plots, which are shown in Figure 1.7. The following is the R code to show the estimated network plot from Sachs et al. [91].

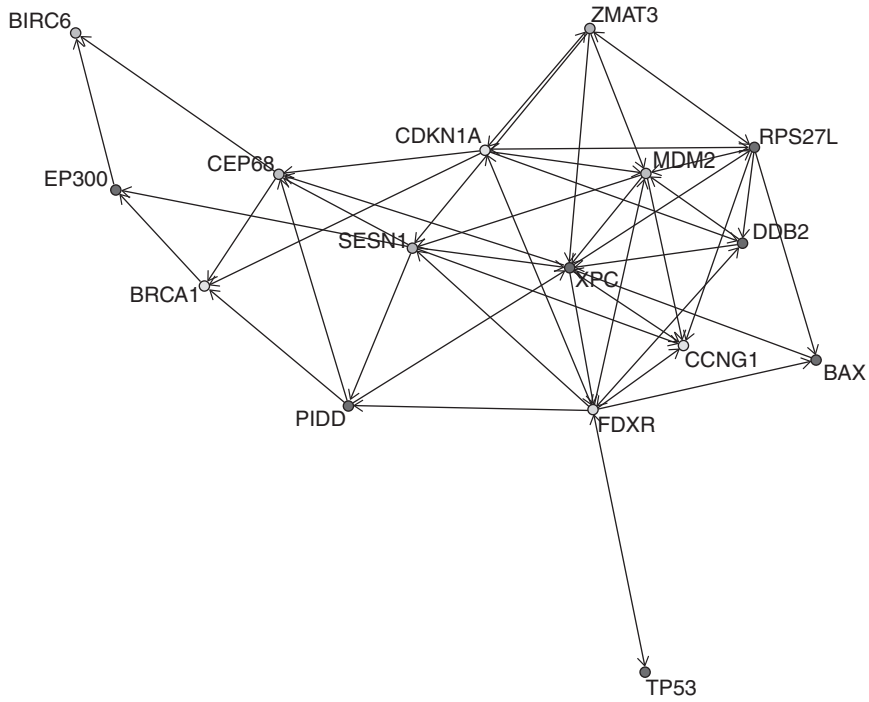

Figure 1.6 Estimate structure in CCLE data with 16 genes, which have roles in different processes surrounding MDM2.

```
> library(network)
> estmated.net <- network(t(estimate_Sachs))
> main.name=paste("(a) Estimate from Sachs et al. [91]",sep="")
> plot(estmated.net,
+ displaylabels=TRUE,
+ boxed.labels=FALSE,
+ mode="circle",
+ vertex.cex=3,
+ arrowhead.cex=2,
+ label.cex=1.0,
+ pad=0.15,
+ label.pos=6,
+ cex.main=1.2,
+ main=main.name)
```

In order to show network plots, "library(network)" needs to be called. "estimate_Sachs" indicates the structure matrix based on the estimated network from Sachs et al. [91]. The estimated network plot is in Figure 1.7a. Next, the R code to show how to obtain the estimated network structure from the cLasso method [53] is as follows:

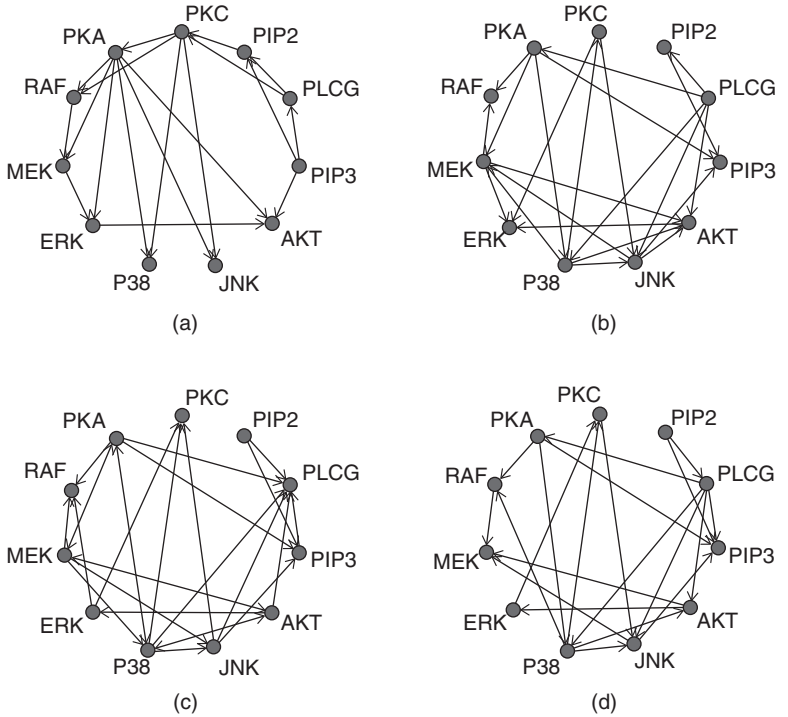

Figure 1.7 Four different estimates for the cellular signaling network. (a) Estimate from Sachs et al. [91]; (b) estimate by cLasso; (c) estimate by PC algorithm; (d) estimate by MMHC algorithm.

```
> delta<-0.15
> cLasso.fit <- cLasso(x.stand=Data.protein.expression,lambda=0.7,
search.index=0,gamma_weight=delta)
> estimate_T_cLasso <- cLasso.fit$est_T
```

The delta in the above-mentioned code is the value of the power for calculating the weight, w_{kj}, in Equation 1.3. The R codes and tutorials will be released as an R package, the name of which is tentatively library(DAGLasso). The estimated network plot is in Figure 1.7b. Then, the next two R codes are the examples of the PC algorithm and the MMHC algorithm. The estimated network plots are in Figure 1.7c and d, respectively. The details of the R codes are in the tutorials provided in the R packages, library(pcalg) and library(bnlearn).

```
> library(pcalg)
> alpha_PC<-0.01
> data_size<-dim(Data.protein.expression)[1]
> data_dim<-dim(Data.protein.expression)[2]
> indepTest <- gaussCItest
> suffStat <- list(C = cor(Data.protein.expression), n = data_size)
```

```
> pc.fit <- pc(suffStat, indepTest, data_dim, alpha_PC)
> g<-pc.fit@graph
> estimate_T_pc_algorithm <-t(as(g,"matrix"))
> library(bnlearn)
> alpha_MMHC<-0.01
> mmhc.fit<-mmhc(Data.protein.expression,alpha=alpha_MMHC)
> g<-mmhc.fit$arcs
> estimate_T_mmhc_algorithm <-matrix(as.numeric(g),ncol=2,
nrow=dim(g)[1])
```

1.8 DISCUSSION

This chapter discusses the recent development of model-based methodologies for estimating cancer gene networks based on the score-and-search approach with the adaptive lasso-based score function, which has been a very important problem in genomic projects for decades. We overviewed the background of Bayesian networks and the available genetic data. We explored structural equation models and the lasso-type score function formulation, which is a recently developed approach in gene network problem. We also discussed an optimization problem to find the solution for the network estimation problem and described the application of the method for data-driven analytics.

1.9 OTHER USEFUL SOFTWARES

There are several software tools visualizing the pathways and networks.

- Pathway Studio (http://www.elsevier.com/online-tools/pathway-studio): Pathway Studio is made from Elsevier Life Science Solutions, and it is a tool for biological decision support. It shows protein–protein interaction and target–drug interaction.
- Ingenuity Pathway Analysis (IPA) (http://www.ingenuity.com/products/ipa): IPA is developed from QIAGEN, and it is a tool to model and analyze the complex biological systems. This tool gives several functions such as causal network analysis, upstream regulator analysis, and downstream effects analysis.

There are also other packages in the R software for estimating DAGs of the Bayesian network under different data types with various techniques. Following are some of the packages that are briefly explained.

- The package, “bnlearn” [96], implements score-based, constraint-based, and hybrid algorithms for learning the structure of Bayesian networks. It supports parameter learning via maximum likelihood and Bayesian estimators and provides inference. This function can be applied to both discrete and continuous data such as Gaussian data.

- The package, "deal" [97] , implements a heuristic search algorithm and defines priors to estimate the Bayesian network in both discrete and continuous data. It can also simulate data sets from a given structure.
- The package, "pcalg" [44], implements the PC algorithm and the extended algorithms for causal structure learning and causal effect estimation.
- The package, "catnet" [98], implements a maximum likelihood-based method to estimate a Bayesian network for categorical data.

ACKNOWLEDGMENTS

Research is supported by NIH-1-R21 GM110450-01.

REFERENCES

[1] Bernstein C, Prasad AR, Nfonsam V, Bernstei H. DNA damage, DNA repair and cancer. In: Clark C, editor. *New Research Directions in DNA Repair*. InTech; 2013. p 413–465.

[2] Bernstein C. 2009. DNA damage and cancer, SciTopics. Available at http://www.scitopics.com/DNA_Damage_and_Cancer.html. Accessed 2014 May 6.

[3] Cunningham FH, Fiebelkorn S, Johnson M, Meredith C. A novel application of the Margin of Exposure approach: Segregation of tobacco smoke toxicants. Food Chem Toxicol 2011;49(11):2921–2933.

[4] Kanavy HE, Gerstenblith MR. Ultraviolet radiation and melanoma. Semin Cutan Med Surg 2011;30(4):222–228.

[5] Handa O, Naito Y, Yoshikawa T. Redox biology and gastric carcinogenesis: The role of *Helicobacter pylori*. Redox Rep 2011;16(1):1–7.

[6] Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol 2011;85(8):863–871.

[7] Malkin D. Li-Fraumeni syndrome. Genes Cancer 2011;2(4):475–484.

[8] Halford S, Rowan A, Sawyer E, Talbot I, Tomlinson I. O^6-methylguanine-methyltransferase in colorectal cancers: Detection of mutations, loss of expression, and weak association with G:C>A:T transitions. Gut 2005;54(6):797–802.

[9] Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J, Houlihan PS, Krouse RS, Prasad AR, Einspahr JG, Buckmeier J, Alberts DS, Hamilton SR, Issa JP. MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst 2005;97(18):1330–1338.

[10] Psofaki V, Kalogera C, Tzambouras N, Stephanou D, Tsianos E, Seferiadis K, Kolios G. Promoter methylation status of hMLH1, MGMT, and CDKN2A/p16 in colorectal adenomas. World J Gastroenterol 2010;16(28):3553–3560.

[11] Amatu A, Sartore-Bianchi A, Moutinho C, Belotti A, Bencardino K, Chirico G, Cassingena A, Rusconi F, Esposito A, Nichelatti M, Esteller M, Siena S. Promoter CpG island hypermethylation of the DNA repair enzyme MGMT predicts clinical response to dacarbazine in a phase II study for metastatic colorectal cancer. Clin Can Res 2013;19(8):2265–2272.

[12] Mokarram P, Zamani M, Kavousipour S, Naghibalhossaini F, Irajie C, Moradi Sarabi M, Hosseini SV. Different patterns of DNA methylation of the two distinct O6-methylguanine-DNA methyltransferase (O6-MGMT) promoter regions in colorectal cancer. Mol Biol Rep 2013;40(5):3851–3857.

[13] Valeri N, Gasparini P, Fabbri M, Braconi C, Veronese A, Lovat F, Adair B, Vannini I, Fanini F, Bottoni A, Costinean S, Sandhu SK, Nuovo GJ, Alder H, Gafa R, Calore F, Ferracin M, Lanza G, Volinia S, Negrini M, McIlhatton MA, Amadori D, Fishel R, Croce CM. Modulation of mismatch repair and genomic stability by miR-155. Proc Natl Acad Sci USA 2010;107(15):6982–6987.

[14] Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. J Comput Biol 2000;7:601–620.

[15] Keller MP, Choi YJ, Wang P, Davis DB, Rabaglia ME, Oler AT, Stapleton DS, Argmann C, Schueler KL, Edwards S, Steinberg HA, Neto EC, Kleinhanz R, Turner S, Hellerstein MK, Schadt EE, Yandell BS, Kendziorski CM, Attie AD. A gene expression network model of type 2 diabetes establishes a relationship between cell cycle regulation in islets and diabetes susceptibility. Genome Res 2008;18:706–716.

[16] Zou M, Conzen SD. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 2005;21:71–79.

[17] Imoto S, Kim S, Goto T, Miyano S, Aburatani S, Tashiro K, Kuhara S. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. J Bioinf Comput Biol 2003;1:231–252.

[18] Beal MJ, Falciani F, Ghahramani Z, Rangel C, Wild DL. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics 2005;21:349–356.

[19] Werhli AV, Grzegorczyk M, Husmeier D. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks. Bioinformatics 2006;22:2523–2531.

[20] Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M. A Bayesian networks approach for predicting protein–protein interactions from genomic data. Science 2003;302:449–453.

[21] Tu Z, Keller MP, Zhang C, Rabaglia ME, Greenawalt DM, Yang X, Wang IM, Dai H, Bruss MD, Lum PY, Zhou YP, Kemp DM, Kendziorski C, Yandell BS, Attie AD, Schadt EE, Zhu J. Integrative analysis of a cross-loci regulation network identifies app as a gene regulating insulin secretion from pancreatic islets. PLoS Genet 2012;8:e1003107.

[22] Neto EC, Ferrara CT, Attie AD, Yandell BS. Inferring causal phenotype networks from segregating populations. Genetics 2008a;179:1089–1100.

[23] Neto EC, Keller MP, Attie AD, Yandell BS. Causal graphical models in systems genetics: A unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. Ann Appl Stat 2010;4:320–339.

[24] Ferrara CT, Wang P, Neto EC, Stevens RD, Bain JR, Wenner BR, Ilkayeva OR, Keller MP, Blasiole DA, Kendziorski C, Yandell BS, Newgard CB, Attie AD. Genetic networks of liver metabolism revealed by integration of metabolic and transcriptomic profiling. PLoS Genet 2008;4:e1000034.

[25] Markowetz F, Spang R. Inferring cellular networks: A review. BMC Bioinf 2007;8 (Suppl 6):S5.

[26] Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press; 2009.

[27] Robinson R. Counting unlabeled acyclic digraphs. In: Little CHC, editor. *Combinatorial Mathematics V: Proceedings of the Fifth Australian Conference, Held at the Royal Melbourne Institute of Technology*. Berlin: Springer; 1977. p 28–43.

[28] Meinshausen N, Buhlmann P. High-dimensional graphs and variable selection with the Lasso. Ann Stat 2006;34:1436–1462.

[29] Shojaie A, Michailidis G. Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. Biometrika 2010;97:519–538.

[30] Buntine WL. Operations for learning with graphical models. J Artif Intell Res 1994;2:159–225.

[31] Buntine WL. A guide to the literature on learning probabilistic networks from data. IEEE Trans Knowl Data Eng 1996;8:195–210.

[32] Heckerman D. A tutorial on learning with Bayesian networks. Technical report MSR-TR-95-06, Microsoft Research; 1995.

[33] Neapolitan RE. *Learning Bayesian Networks*. Series in Artificial Intelligence. Prentice Hall; 2004.

[34] Daly R, Shen Q, Aitken S. Learning Bayesian networks: Approaches and issues. Knowl Eng Rev 2011;26:99–157.

[35] Marcot BG, Holthausen RS, Raphael MG, Rowland M, Wisdom M. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. For Ecol Manage 2001;153:29–42.

[36] Borsuk ME, Stow CA, Reckhow KH. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecol Model 2004;173:219–239.

[37] Rajapakse JC, Zhou J. Learning effective brain connectivity with dynamic Bayesian networks. Neuroimage 2007;37:749–760.

[38] Li JN, Wang ZJ, Palmer SJ, McKeown MJ. Dynamic Bayesian network modeling of fMRI: A comparison of group-analysis methods. Neuroimage 2008;37:749–760.

[39] Li J, Jin J. Optimal sensor allocation by integrating causal models and set-covering algorithms. IIE Trans 2010;42(8):564–576.

[40] Liu K, Shi J. Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network. IIE Trans 2013;45(6):630–643.

[41] Liu K, Zhang X, Shi J. Adaptive sensor allocation strategy for process monitoring and diagnosis in a Bayesian network. IEEE Trans Autom Sci Eng 2014;11(2):452–462.

[42] Broom BM, Subramanian D. Computational methods for learning Bayesian networks from high-throughput biological data in Bayesian inference for gene expression and proteomics marina. In: Do K-A, Muller P, Vannucci M, editors. *Bayesian Inference for Gene Expression and Proteomics*. New York, NY: Cambridge University Press; 2006.

[43] Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press; 2000.

[44] Kalisch M, Buhlmann P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. J Mac Learn Res 2007;8:613–636.

[45] Tsamardinos I, Brown L, Aliferis C. The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn 2006;65:31–78.

[46] Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn 1995;20:197–243.

[47] Huang S, Li J, Ye J, Fleisher A, Chen K, Wu T, Reiman E, the Alzheimer's Disease Neuroimaging Initiative. A sparse structure learning algorithm for gaussian Bayesian network identification from high-dimensional data. IEEE Trans Pattern Anal Mach Intell 2013;35(6):1328–1342.

[48] Fu F, Zhou Q. Learning sparse causal gaussian networks with experimental intervention: Regularization and coordinate descent. J Am Stat Assoc 2013;108:288–300.

[49] Han SW, Chen G, Belousov A, Essioux L, Zhong H. Estimation of sparse directed acyclic graphs through a penalized likelihood method for gene network inference. Submitted to Journal of the American Statistical Association; 2013.

[50] Nagarajan R, Scutari M, Lebre S. *Bayesian Networks in R: with Applications in Systems Biology*. Springer; 2013.

[51] Chickering DM. Learning equivalence classes of Bayesian-network structures. J Mach Learn Res 2002;2:445–498.

[52] The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–1068.

[53] Han, SW, Chen, G, Belousov, A, Essioux, L and Zhong, H, Estimation of sparse directed acyclic graphs through a penalized likelihood method for gene network inference, Technical Report; 2014.

[54] Zou H. The adaptive LASSO and its oracle properties. J Am Stat Assoc 2006;101:1418–1429.

[55] Breiman L. Better subset regression using the nonnegative garotte. Technometrics 1995;37:373–384.

[56] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 2001;96:1348–1360.

[57] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B 1996;58:267–288.

[58] Han SW, Zhong H. Estimation of directed acyclic graphs under Poisson log-normal distribution. Submitted; 2013.

[59] Schmidt M, Niculescu-Mizil A, Murphy K. Learning graphical model structure using l1-regularization paths. AAAI 2007;7:1278–1283.

[60] Pellet J-P, Elisseeff A. Using markov blankets for causal structure learning. J Mach Learn Res 2008;9:1295–1342.

[61] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–1182.

[62] Friedman N, Nachman I, Péer D. Learning Bayesian network structure from massive datasets: The 'Sparse Candidate' algorithm. Proc. 15th Conf. Uncertainty in Artificial Intelligence; 1999a.

[63] Margaritis D, Thrun S, Bayesian network induction via local neighborhoods. Proc. Conf. Advances in Neural Information Processing Systems; 1999.

[64] Ramoni M, Sebastiani P. The use of exogenous knowledge to learn Bayesian networks from incomplete databases. In Proceedings of the Second International Symposium on Advances in Intelligent Data Analysis, Reasoning about Data (IDA '97), Lecture Notes in Computer Science, 1280, 537–548, Springer; 1997.

[65] Friedman N, Goldszmidt M, Wyner A. Data analysis with Bayesian networks: a bootstrap approach. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI-99), Prade, H., and Laskey, K. (eds), Morgan Kaufmann, 196–205; 1999b.

[66] Peng H, Ding C. Structure search and stability enhancement of Bayesian networks. In Proceedings of the Third IEEE International Conference on Data Mining (ICDM 2003), Wu, X., Tuzhilin, A., and Shavlik, J. (eds), IEEE Computer Society, 621–624; 2003, doi: 10.1109/ICDM.2003.1250992.

[67] Holness GF. A direct measure for the efficacy of Bayesian network structures learned from data. In Proceedings of the Fifth International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2007), Lecture Notes in Artificial Intelligence 4571, 601–615, Springer; 2007.

[68] Wang H, Li B, Leng C. Shrinkage tuning parameter selection with a diverging number of parameters. J R Stat Soc B 2009;71:671–683.

[69] Shao J. An asymptotic theory for linear model selection. Stat Sin 1997;7:221–264.

[70] Shi P, Tsai CL. Regression model selection. a residual likelihood approach. J R Stat Soc B 2002;64:237–252.

[71] Fan J, Peng H. Nonconcave penalized likelihood with a diverging number of parameters. Ann Stat 2004;32:928–961.

[72] Wang H, Li G, Tsai CL. Regression coefficient and autoregressive order shrinkage and selection via the lasso. J R Stat Soc B 2007;69:63–78.

[73] Bellmore M, Nemhauser GL. The traveling salesman problem: A survey. Oper Res 1968;16:538–558.

[74] Burkard RE. Travelling salesman and assignment problems: A survey. Ann Dis Math 1979;4:193–215.

[75] Junger M, Reinelt G, Rinaldi G. The traveling salesman problem. In: Ball M, Magnanti T, Monma CL, Nemhauser G, editors. *Handbook on Operations Research and Management Sciences, Networks*. Amsterdam: North-Holland; 1995. p 225–330.

[76] Kalisch M, Fellinghauer BAG, Grill E, Maathuis MH, Mansmann U, Buhlmann P, Stucki G. Understanding human functioning using graphical models. BMC Med Res Methodol 2010;10(1):14.

[77] Nagarajan R, Datta S, Scutari M, Beggs M, Nolen G, Peterson C. Functional relationships between genes associated with differentiation potential of aged myogenic progenitors. Front Physiol 2010;1:1–8.

[78] Stekhoven DJ, Moraes I, Sveinbjörnsson G, Hennig L, Maathuis MH, Bühlmann P. Causal stability ranking. Bioinformatics 2012;28(21):2819–2823.

[79] Zhang X, Zhao XM, He K, Lu L, Cao Y, Liu J, Hao JK, Liu ZP, Chen L. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. Bioinformatics 2012;28(1):98–104.

[80] Colombo D, Maathuis MH. Order-independent constraint-based causal structure learning. Technical Report, arXiv:1211.3295v2; 2013.

[81] National Cancer Institute. 2013. SEER Stat Fact Sheets: Melanoma of the Skin. Available at http://seer.cancer.gov/statfacts/html/melan.html. Accessed 2013 Jan 29.

[82] Garbe C, McLeod GR, Buettner PG. Time trends of cutaneous melanoma in Queensland, Australia and Central Europe. Cancer 2000;89:1269–1278.

[83] Bedikian AY, Millward M, Pehamberger H, et al. Bcl-2 antisense (oblimersen sodium) plus dacarbazine in patients with advanced melanoma: The Oblimersen Melanoma Study Group. J Clin Oncol 2006;24:4738–4745.

[84] Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;364(26):2507–2516.

[85] Solit D, Sawyers CL. Drug discovery: How melanomas bypass new therapy. Nature 2010;468(7326):902–903.

[86] Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 2010;363(8):711–723.

[87] Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. N Engl J Med 2011;364(26):2517–2526.

[88] Fleming NH, Silva I, Miera EV-S, Brady B, Han SW, Hanniford D, Wang J, Shapiro RL, Hernando E, Zhong J, Osman I. Serum-based miRNAs in the prediction and detection of recurrence in melanoma patients, Submitted to Cancer; 2014.

[89] Zhong H, Chen G, Jukofsky L, Geho D, Han SW, Birzele F, Bader S, Himmelein L, Cai J, Alvertyn Z, Rothe M, Essioux L, Burtscher H, Middleton SA, Rueger R, Chen L-C, Dangl M, Nichols G, Pierceall WE. MDM2 antagonist clinical response association with a gene expression signature in acute myeloid leukemia. Br J Haematol 2015;171:432–425.

[90] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603–607.

[91] Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science 2005;308:523–529.

[92] Herzenberg LA, Parks D, Sahaf B, Perez O, Roederer M, Herzenberg LA. The history and future of the fluorescence activated cell sorter and flow cytometry. Clin Chem 2002;48(12):1819–1827.

[93] Perez OD, Nolan GP. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. Nat Biotechnol 2002;20(2):155–162.

[94] Perez OD, Mitchell D, Jager GC, South S, Murriel C, McBride J, Herzenberg LA, Kinoshita S, Nolan GP. Leukocyte functional antigen 1 lowers T cell activation thresholds and signaling through cytohesin-1 and Jun-activating binding protein 1. Nat Immunol 2003;4(11):1083–1092.

[95] Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud Ø, Gjertsen BT, Nolan GP. Single cell profiling of potentiated phospho-protein networks in cancer cells. Cell 2004;118(2):217–228.

[96] Scutari M. Learning Bayesian networks with the bnlearn R package. J Stat Softw 2010;35:1–22.

[97] Bottcher SG, Dethlefsen C. Deal: A package for learning Bayesian networks. J Stat Softw 2003;8:1–40.

[98] Balov, N, Salzman, P. 2013. catnet: Categorical Bayesian network inference, http://cran.rproject.org/web/packages/catnet/catnet.pdf. Accessed 2016 Feb 1.