

Bayesian Networks Learning for Gene Expression Datasets

Giacomo Gamberoni¹, Evelina Lamma¹, Fabrizio Riguzzi¹,
Sergio Storari¹, and Stefano Volinia²

¹ ENDIF-Dipartimento di Ingegneria, Università di Ferrara, Ferrara, ITALY
{ggamberoni;elamma;friguzzi;sstorari}@ing.unife.it

² Dipartimento di Biologia, Università di Ferrara, Ferrara, ITALY
s.volinia@unife.it

Abstract. DNA arrays yield a global view of gene expression and can be used to build genetic networks models, in order to study relations between genes. Literature proposes Bayesian network as an appropriate tool for develop similar models. In this paper, we exploit the contribute of two Bayesian network learning algorithms to generate genetic networks from microarray datasets of experiments performed on Acute Myeloid Leukemia (AML).

In the results, we present an analysis protocol used to synthesize knowledge about the most interesting gene interactions and compare the networks learned by the two algorithms. We also evaluated relations found in these models with the ones found by biological studies performed on AML.

1 Introduction

From DNA microarray experiments, we can obtain a huge amount of data about gene expression of different cell populations. An intelligent analysis of these results can be very useful and important for cancer research.

An important field of interest in this area is the discovering of genetic networks, intended as a synthetic representation of genetic interactions. Similar problems are often studied in other fields, and several techniques were developed in order to learn interaction networks from examples. One of the most used approach for this type of problems is the Bayesian Network one.

Bayesian networks are suitable for working with the uncertainty that is typical of real-life applications. These are robust models and usually maintain good performance also with missing or wrong values.

A Bayesian network is a directed, acyclic graph (DAG) whose nodes represent random variables. In Bayesian networks, each node is conditionally independent of any subset of the nodes that are not its descendants, given its parent nodes.

By means of Bayesian networks, we can use information about the values of some variables to obtain probabilities for the values of others. A probabilistic inference takes place once the probability of the values of each node conditioned

to just its parents are given. These are usually represented in a tabled form, called Conditional Probability Tables (CPTs).

Applying this theory to the Bioinformatic field, we aim to build a network in which nodes represent different genes or attributes of a biological sample. Studying several samples related to a particular pathology, we build up a network that represents the probabilistic relations between genes and attributes. This model may be useful for biologists, because it highlights these interactions in a synthetic representation.

Techniques for learning Bayesian networks have been extensively investigated (see, for instance [12]). Given a training set of examples, learning such a network is the problem of finding the structure of the direct acyclic graph and the CPTs associated with each node in the DAG that best match (according to some scoring metric) the dataset. Optimality is evaluated with respect to a given scoring metric (for example, description length or posterior probability [6, 12, 20]). A procedure for searching among possible structures is needed. However, the search space is so vast that any kind of exhaustive search cannot be considered, and a greedy approach is followed.

In the literature, we find two different approaches for learning Bayesian networks: the first one is based on information theory [5], while the second one is based on the search and score methodology [6, 12, 20].

In this paper we use the K2 and the K2-lift algorithm. The K2 algorithm [6] is one of the best known algorithms among those that follows the search and score methodology. K2-lift algorithm is a modified version of K2 algorithm that uses the lift parameter (defined in association rules theory [1]), in order to improve the quality of learned networks and to reduce the computational resources needed .

The paper is structured as follows. Section 2 provides an introduction to Bayesian networks and to algorithms for learning them. In Section 3 we present the experimental domain of the dataset used, presenting the Leukemia dataset used for testing and validation. In Section 3.3 we present the results, comparing K2 algorithm with the K2-lift one; then we report some considerations on the results, from the biological point of view. Related work is mentioned in Section 4. Finally, in Section 5, we conclude and present future work.

2 Bayesian Networks Theory and Learning

A Bayesian network \mathcal{B} is defined as a pair $\mathcal{B} = (\mathcal{G}, \mathcal{T})$, where \mathcal{G} is a directed, acyclic graph and \mathcal{T} is a set of conditional probability tables. \mathcal{G} is defined as a couple $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where \mathcal{V} is a set of nodes $\mathcal{V} = \{V_1, \dots, V_n\}$, representing a set of stochastic variables, and \mathcal{A} is a set of arcs $\mathcal{A} \subseteq \mathcal{V} \times \mathcal{V}$ representing conditional and unconditional stochastic independences among the variables [14, 16]. In the following, variables will be denoted by upper-case letters, for example V , whereas a variable V which takes on a value v , that is $V = v$, will be abbreviated to V .

The basic property of a Bayesian network is that any variable corresponding to a node in the graph \mathcal{G} is conditionally independent of its non-descendants

given its parents; this is called the *local Markov property*. A joint probability distribution $Pr(V_1, \dots, V_n)$ is defined on the variables. As a consequence of the local Markov property, the following decomposition property holds:

$$Pr(V_1, \dots, V_n) = \prod_{i=1}^n Pr(V_i | \pi(V_i)) \quad (1)$$

where $\pi(V_i)$ denotes the set of variables corresponding to the parents of V_i , for $i = 1, \dots, n$.

Once the network is built, probabilistic statements can be derived from it by probabilistic inference, using one of the inference algorithms described in the literature (for example [14, 16]).

Given a training set of examples, learning a Bayesian network is the problem of finding the structure of the direct acyclic graph and the Conditional Probability Tables (CPTs) associated with each node that best match (according to some scoring metrics) the dataset.

2.1 Learning algorithms used

A frequently used procedure for Bayesian network structure construction from data is the K2 algorithm [6]. Given a database D , this algorithm searches for the Bayesian network structure \mathcal{G} with maximal $Pr(\mathcal{G}, D)$, where $Pr(\mathcal{G}, D)$ is determined as described below. Let D be a database of m cases, where each case contains a value assignment for each variable in \mathcal{V} . Let \mathcal{T} be the associated set of conditional probability distributions. Each node $V_i \in \mathcal{V}$ has a set of parents $\pi(V_i)$.

The K2 algorithm assumes that an ordering on the variables is available, and that all structures are a priori equally likely. For every node V_i , it searches for the set of parent nodes $\pi(V_i)$ that maximizes a function $g(V_i, \pi(V_i))$.

K2 adopts a greedy heuristic method. It starts by assuming that a node lacks parents, and then, at every step, it adds the parent whose addition mostly increases the function $g(V_i, \pi(V_i))$. K2 stops adding parents to the nodes when the addition of a single parent does no longer increase $g(V_i, \pi(V_i))$.

K2 is characterized by the insertion of a large number of extra arcs. The extra arc problem of K2 arises especially when the network is characterized by a lot of root nodes (nodes without parents). During network learning, the algorithm tries to add parents to each of these nodes until it maximizes function $g(V_i, \pi(V_i))$. The algorithm will add at least one arc to root nodes because the value of the heuristic for this new structure is always better than the value of the previous structure.

K2-lift [13] is an extension of the K2 algorithm. It uses parameters normally defined in relation to association rules [1].

Association rules represent co-occurrence between events $ant \Rightarrow cons$ in which an event is defined as the association of a value to an attribute, while ant and $cons$ are set of events.

The support of a set of events is the fraction of records that contain all the events in the set.

In K2-lift we focused our attention on rules with one item in the antecedent and one item in the consequent (called one-to-one rules) and on the lift parameter [4] of a rule, a measure of rule interest, computed as follows: $lift = support(ant \cup cons) / (support(ant) \times support(cons))$. The knowledge represented by one-to-one association rules parameters is used to reduce the set of nodes from which the K2 algorithm tries to identify the best set of parents. This reduces the problem of extra arcs.

3 Experiments

In this section we present the experimental dataset (in section 3.1), then, in section 3.2 we describe the analysis protocol used. Finally, in section 3.3 we present and evaluate the results of the performed experiments.

3.1 Dataset

Recent technical and analytical advances make it practical to evaluate quantitatively the expression of thousands of genes in parallel using microarrays (as described extensively in [2, 11, 7]). A microarray experiment consists of measurements of the relative representation of a large number of mRNA species in a set of biological samples. This mode of analysis has been used to observe gene expression variation in a variety of human tumors.

The analyzed dataset, available on-line in the ArrayExpress repository of the European Bioinformatics Institute³, regroups the results of 20 microarray experiments, divided as follows:

- 10 Acute Myeloid Leukemia (AML) samples;
- 10 MyeloDysplastic Syndrome (MDS) samples.

Acute Myeloid leukemia (AML) may develop de novo or secondarily to MyeloDysplastic Syndrome (MDS). Although the clinical outcome of MDS-related AML is worse than that of de novo AML, it is not easy to differentiate between these two clinical courses without a record of prior MDS. Large-scale profiling of gene expression by DNA microarray analysis is a promising approach with which to identify molecular markers specific to de novo or MDS-related AML.

The experiments were performed using Affymetrix⁴ Genechip Human Genome U95Av2 arrays. The Detection algorithm (included in the Affymetrix Microarray Suite Version 5.0) uses probe pair intensities to assign a Present, Marginal, or Absent call. This is a very reliable discretization method that reveals if a probe is expressed or not in the sample. Notice that Bayesian Networks can handle only discrete attributes, so we absolutely need such a discretized expression level.

³ <http://www.ebi.ac.uk/arrayexpress/>, access code E-MEXP-25

⁴ <http://www.affymetrix.com>

The data from M experiments considering N probes, may be represented as a $M \times N$ detection matrix, in which each of the M rows consists of a N -element detection vector for a single sample.

In order to obtain a more readable Bayesian Network, we need to reduce the number of attributes. So we conducted the experiments focusing the attention on specific aspects of the illnesses described in the dataset and, in particular, on the probes of our dataset related to each one of these aspects. Filtering the dataset probes in such way we created new smaller datasets to be analyzed.

In these dataset we kept only the probes related to an aspect. In order to get their names, we used the NetAffx Analysis Center⁵, searching for a term and specifying the GeneChip Array name.

Following the indications of our biologist, we considered two different aspects and subsequently worked on two different datasets.

The first interesting aspect of the pathology under evaluation is the Negative Regulation of Cell Proliferation (GO:0008285): this GO term refers to any process that stops, prevents or reduces the rate or extent of cell proliferation. The initial dataset contains 32 probes related to this GO term so we create a new smaller dataset (named *NRCP_dataset*) composed by 20 rows (cases) and 33 columns (32 probes expression levels and the class attribute).

The second interesting aspect of the pathology under evaluation is the study of the biological process related to Hepatocyte Growth Factor or Substrate (HGF, HGS). The initial dataset contains 9 probes related to this aspect so we create a new smaller dataset (named *HGS/HGF_dataset*) composed by 20 rows (cases) and 10 columns (9 probes expression levels and the class attribute);

3.2 Analysis Protocol

Given a dataset, the analysis protocol followed in our experiments consists of 3 steps :

1. Generate a set of 20 random attribute orderings named SAO_i , with $i = 1, \dots, 20$. The attribute ordering is required by the Bayesian network learning algorithms described in Section 2.1. The generation of the set of SAO_i is necessary because the optimal attribute ordering is unknown in our experiments.
2. For each learning algorithm $La \in \{K2, K2 - lift\}$:
 - (a) For $i=1, \dots, 20$
 - i. Learn the Bayesian network $BN_{La,i}$ by using La on SAO_i
 - ii. Compute the Bayes score $BS_{La,i}$ of $BN_{La,i}$
 - (b) Rank the learned network $BN_{La,i}$ according to their score $BS_{La,i}$
 - (c) Analyze the first five learned networks $BN_{La,i}$ and identify:
 - frequent parent probes
 - probes with frequently the same subset of parent probes (found on more than 3 networks over 5)

⁵ <https://www.affymetrix.com/analysis/netaffx/index.affx>

3. Compare the results achieved by using $K2$ and $K2 - lift$

The analysis performed on the learned networks $BN_{La,i}$ is preliminary and the definition of a more complete methodology is required. For example, the probabilistic relations represented by a Bayesian network consist of both qualitative and quantitative probabilistic meanings, but in our preliminary analysis for each Bayesian probabilistic relation found we consider only the qualitative one.

3.3 Results and discussion

Results of application of the analysis protocol on the $HGS/HGF_dataset$ by using $K2$ -lift are presented in Table 1. In each cell you can find the number of occurrence of a relation between a parent node (on the column) and a child node (in the row), frequent relations are in bold.

Table 1. HGS/HGF K2-lift results

	class	742_at HABP2	829_s_at GSTP1	1095_s_at HGF	1340_s_at HGF	33396_at GSTP1	33887_at HGS	35063_at HGFAC	36231_at MGC17330	40508_at GSTA4
class							3		1	
742_at										
829_s_at	2			4		1	1		1	
1095_s_at	3		1				1			
1340_s_at										
33396_at	1		2	2						
33887_at	2		3	1		1			1	
35063_at										
36231_at				5						
40508_at	3		2							

A graphical representation of the results is presented in Figure 1, notice that the width of the arcs are proportional to the number of relations found (we omitted relations found only once, in order to keep the graph simple).

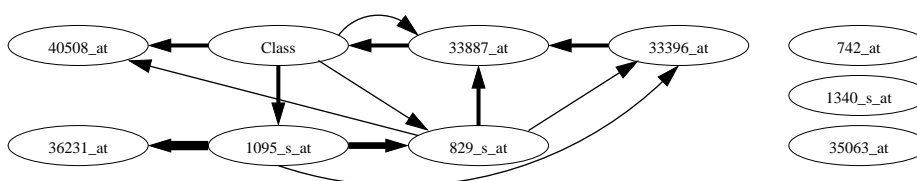


Fig. 1. K2-lift network

The most frequent parents in the obtained networks are the probes 829_s_at ($GSTP1$) with frequency 19.5%, 1095_s_at (HGF) with frequency 29.3% and the $class$ attribute with frequency 26.8%.

Results of application of the analysis protocol on the *HGS/HGF_dataset* by using K2 are presented in Table 2. In each cell you find the number of occurrence of a relation between a parent node (on the column) and a child node (in the row), frequent relations are in bold.

Table 2. HGS/HGF K2 results

	class	742_at HABP2	829_s_at GSTP1	1095_s_at HGF	1340_s_at HGF	33396_at GSTP1	33887_at HGS	35063_at HGFAC	36231_at MGC17330	40508_at GSTA4
class							3		1	
742_at	2			1		1				
829_s_at	2			4		1	1		1	
1095_s_at		3			1					
1340_s_at	2	1		1		1				
33396_at	1	1		2						
33887_at	2		3	1		1			1	
35063_at	2	1		1		1				
36231_at		4			1					
40508_at		3	1		1					

The most frequent parents are the probes *742_at* (*HABP2*) with frequency 25%, *1095_s_at* (*HGF*) with frequency 18.9% and the *class* attribute with frequency 20.8%.

Applying the analysis protocol on the *NRCP_dataset* (complete tables are omitted, due to lack of space) by using K2-lift algorithm, we observed that:

- the most frequent parent probes are *1880_at* (*MDM2*) with frequency 16.9%, *37107_at* (*PPM1D*) with frequency 11.3% and *40631_at* (*TOB1*) with frequency 10.5%;
- there are five frequent probabilistic probe relations composed by:
 - *36136_at* (*TP53I11*) with parents *1880_at* (*MDM2*) and *38379_at* (*GP-NMB*);
 - *36479_at* (*GAS8*) with parent *1880_at* (*MDM2*);
 - *38379_at* (*GP-NMB*) with parent *1880_at* (*MDM2*);
 - *38682_at* (*BAP1*) with parent *38379_at* (*GP-NMB*);
 - *41141* (*PRKRIR*) with parent *37218_at* (*BTG3*).

Applying the analysis protocol on the *NRCP_dataset* by using K2 algorithm we observed that:

- the most frequent parent probes are *1880_at* (*MDM2*) with frequency 27.1%, *34629_at* (*TP53I11*) with frequency 10%, *32568_at* (*BTG3*) with frequency 9.0% and the *class* attribute with frequency 10.5%;
- there are six frequent probabilistic probe relations composed by:
 - *36136_at* (*TP53I11*) with parent *1880_at* (*MDM2*);
 - *36479_at* (*GAS8*) with parent *1880_at* (*MDM2*);
 - *38379_at* (*GP-NMB*) with parent *1880_at* (*MDM2*);

- *38639_at (MXD4)* with parent *1880_at (MDM2)*;
- *38682_at (BAP1)* with parent *38379_at (GPNMB)*;
- *41141_at (PRKRIR)* with parents *37218_at (BTG3)* and *38682_at (BAP1)*.

These results may be evaluated both regards the applied methods and the biological significance.

About the methods, some considerations arise about the difference between the results produced by K2 and K2-lift, and about the adopted analysis protocol.

Analyzing the Bayesian networks learned by K2, we can see that the first probe in the ordering is usually overrepresented as parent of the other probes. So it creates a more connected network that is difficult to evaluate by a biologist. K2-lift learns a more synthetic network which highlights the most interesting probes interactions. About the Bayes score of the learned networks, the score average of the best ones proposed by K2 and K2-lift are similar (-170.14 against -170.97 for the *NRCP_dataset* and -61.37 against -62.38 for the *HGS/HGF_dataset*).

About the adopted analysis protocol, the probabilistic relations found in the datasets consider only the qualitative meaning of such relations. The conditional probability table associated to each frequent relation found can be computed by standard statistic methodologies.

The results propose also some biological considerations. Acute myelogenous leukemia is a heterogeneous disease that appears to evade the normal regulatory controls of tumor suppressor genes (Stirewalt et al, 2000 [19]). Studies in AML have documented mutations in p53 (strictly related to *TP53I11* probe), but these mutations are relatively uncommon, especially compared to their mutational frequency in solid tumors. In addition, expression abnormalities have now been documented in several tumor suppressor genes or related genes including MDM2, p73, Rb, p14(ARF), p15(INK4B), and p16(INK4A). ERBB2 (strictly related to *TOB1* probe, found by K2-lift as a frequent parent) is a receptor protein tyrosine kinase frequently mutated in human cancer. The protein-kinase family is the most frequently mutated gene family found in human cancer and faulty kinase enzymes are being investigated as promising targets for the design of anti-tumour therapies.

Stephens et al (2004) [18] have sequenced the gene encoding the transmembrane protein tyrosine kinase ERBB2 from 120 primary lung tumours and identified 4% that have mutations within the kinase domain; in the adenocarcinoma subtype of lung cancer, 10% of cases had mutations. ERBB2 inhibitors, which have so far proved to be ineffective in treating lung cancer, should now be clinically re-evaluated in the specific subset of patients with lung cancer whose tumours carry ERBB2 mutations. MDM2 (found by K2-lift as a frequent parent) is a target gene of the transcription factor tumor protein p53. Overexpression of this gene can result in excessive inactivation of tumor protein p53, diminishing its tumor suppressor function.

Faderl et al (2000) [8] showed that overexpression of MDM-2 is common in AML and is associated with shorter complete remission duration and event free survival rate. It is striking to note that MDM2 and a TP53 induced protein are so tightly connected in the networks, since MDM2 is probably the most

important protein for regulation of TP53 activity. The ERBB2 and MDM2 interaction is also very revealing and it should be noted that ERBB2 amplification or overexpression can make cancer cells resistant to apoptosis and promotes their growth.

Zhou et al (2001) [22] showed that ERBB2-mediated resistance to DNA-damaging agents requires the activation of Akt, which enhances MDM2-mediated ubiquitination and degradation of TP53.

We then compared our results to the original conclusions from the paper by Oshima and colleagues [15] who generated the datasets we used for analysis. They identified a set of genes associated to the course of the disease and to clinical classification. Furthermore they identified genes which are related to clinical outcome after induction chemotherapy. When compared to their gene lists, we identified novel gene interactions in our analysis, which might be very important for the clinical outcome. TP53 implications are outlined above while HGF has been widely implicated in tumor scattering and invasive growth and is of prognostic importance in AML (Verstovsek et al 2001 [21]). Our findings therefore represented novel potentially informative results related to the AML datasets. Thus it seems that by using the proposed method it is possible to mine useful data from microarray experiments in a previously undescribed way.

4 Related works

A work related to ours is [9]. In it the authors learn causal networks from DNA microarray data with the aim of discovering causal relations between the different genes. A causal network is a network where the parent of a variable are its immediate causes. A causal network can be interpreted as a Bayesian network if we make the Causal Markov Assumption: given the values of a variable's immediate causes, it is independent of its earlier causes.

In order to learn causal networks, the authors make two assumptions: the first is that the unknown causal structure of the domain satisfies the Causal Markov Assumption, the second is that there are no latent or hidden variables. However, from these assumptions it is not possible to distinguish from observations alone between causal networks that specify the same independence properties. Therefore, what is learned is a partially directed acyclic graph (PDAG), where some of the edges can be undirected.

When data is sparse, a single PDAG can not be identified, rather a probability distribution over causal statements is induced. Moreover, the posterior probability is not dominated by a single model. In order to solve this problem, the authors try to identify *features*, i.e., relations between couples of variables. There are two types of features. The first is Markov relations: X is in a Markov relation with Y if Y is in the Markov blanket of X . The second is order relations: X is in an order relation with Y in a PDAG if there is a path between X and Y where all the edges are directed. The aim of the authors is to estimate the posterior probability of the features given the data. Ideally this should be done by sampling networks from the posterior to estimate this quantity. However, this is

a hard problem. Therefore, they resort to a simpler analysis that consists of the bootstrap method: they generate perturbed versions of the data set and learn from them. In this way, they collect many networks that are reasonable models of the data. Then they compute the confidence of a feature as the fraction of networks containing the feature

In order to learn a PDAG from data they use the Sparse Candidate algorithm: a relatively small number of candidate parents for a variable can be identified by means of local statistics (such as correlation). Then the search is performed by picking parents for a variable only from the identified set.

They apply these techniques to DNA microarrays for *S. cerevisiae*. In particular they consider 800 genes whose expression varies over the different cell-cycle stages and 76 gene expression measurements. The learning experiment was conducted using 200-fold bootstrap. The results show that they were able to recover intricate structures even from such small data set. A biological analysis show that the results are well supported by current biological knowledge.

Our approach differs from the one of [9] because we learn a number of networks starting from different orders of the variables rather than from perturbed datasets. Moreover, we compute the confidence of the features only taking into account the best scoring networks according to a Bayesian metric.

Another work very related to ours is [10]. In it the author describe the state of the art in the use of probabilistic graphical models for inferring regulatory networks in cells. Besides the bootstrap approach of [9], the author describe two other studies that are relevant to ours. In the first, the authors examines only the networks in which a small number of regulators explain the expression of all other genes. This simplifies the learning procedure thus leading to statistical and computational advantages. They performed a systematic validation comparing the process and function annotation of the target set of each regulator with the known literature about the regulator. In most cases, they found a match between the annotation and the literature.

In the second study, the genes are divided into module that share a regulatory program. The learning procedure simultaneously identifies the composition of the modules and the regulators for each module. The module approach is in accordance with biological principles that suggest that a regulatory process usually involves many genes at the same time. Moreover, shared regulatory processes require less parameters thus leading to an improvement of the robustness of the model. Finally, the learned networks are easier to interpret, thanks to the module partition. The authors of this study confirmed the obtained results both by comparing the results with the literature and by examining gene expression of knockout strains.

These two studies are alternative to ours and exploit prior biological knowledge in order to target more effectively the gene expression domain, while we used general purpose techniques without exploiting other knowledge besides that contained in the microarray dataset.

5 Conclusions and future works

In this paper we describe the results of experiments conducted applying Bayesian network learning algorithm on microarray datasets. These preliminary results shows that in most of the cases, K2-lift creates a more synthetic network with respect to K2. It is also noteworthy that many relations found confirmation in biological literature.

The analysis protocol used for the result evaluation is very simple and need some enhancements both in complexity and in statistical significance. For these reasons, a bootstrap-based approach will be the subject of further studies. In a more complex protocol, we also need to consider the conditional probability tables and other information associated to the learned networks

In the future we also plan to empirically compare our approach to that of [9] in order to better compare the performances of the two methods. Moreover, we plan to improve our learning process by means of prior biological knowledge, as done in the studies described in [10].

References

1. Agrawal, R., Imielinski, T. and Swami, A.: Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (1993) 207–216
2. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma ,C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J. Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., Staudt, L.M.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403 (2000) 503–511
3. Anderson, D.R., Sweeney, D.J. and Williams, T.A. : Introduction to statistics concepts and applications, Third Edition. West Publishing Company (1994)
4. Berry, J.A. and Linoff, G.S.: Data Mining Techniques for Marketing, Sales and Customer Support. John Wiley & Sons Inc. (1997)
5. Cheng, J., Greiner, R., Kelly, J., Bell, D. and Liu, W.: Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence* 137 (2002) 43–90
6. Cooper, G. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9 (1992) 309–347
7. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. : Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95 (1998) 14863–14688
8. Faderl, S., Kantarjian, H.M., Estey, E., Manshour, T., Chan, C.Y., Rahman El-saied, A., Kornblau, S.M., Cortes, J., Thomas, D.A., Pierce, S., Keating, M.J., Estrov, Z., Albitar, M.: The prognostic significance of p16(INK4a)/p14(ARF) locus deletion and MDM-2 protein expression in adult acute myelogenous leukemia. *Cancer* 89(9) (2000) 1976–82
9. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian Networks to Analyze Expression Data. *Journal of Computational Biology* (7)3/4 (2000) 601–620

10. Friedman, N.: Inferring Cellular Networks Using Probabilistic Graphical Models. *Science* 303 (2004) 799–805
11. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 (1999) 531–537
12. Heckerman, D. and Geiger, D. and Chickering, D.: Learning Bayesian Networks: the combination of knowledge and statistical data. *Machine Learning* 20 (1995) 197–243
13. Lamma, E., Riguzzi, F. and Storari, S.: Exploiting Association and Correlation Rules Parameters for Improving the K2 Algorithm. 16th European Conference on Artificial Intelligence (2004) 500–504
14. Lauritzen, S. L. and Spiegelhalter, D. J.: Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistics Society B* 50 (1988) 157–194
15. Oshima Y., Ueda M., Yamashita Y., Choi Y.L., Ota J., Ueno S., Ohki R., Koinuma R., Wada T., Ozawa K., Fujimura A., Mano H.: DNA microarray analysis of hematopoietic stem cell-like fractions from individuals with the M2 subtype of acute myeloid leukemia. *Leukemia* 17(10) (2003) 1900–1997
16. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
17. Ramoni, M. and Sebastiani, P.: Robust learning with missing data. Technical Report, Knowledge Media Institute, The Open University KMI-TR-28 (1996)
18. Stephens, P., Hunter, C., Bignell, G., Edkins, S., Davies, H., Teague, J., Stevens, C., O’Meara, S., Smith, R., Parker, A., Barthorpe, A., Blow, M., Brackenbury, L., Butler, A., Clarke, O., Cole, J., Dicks, E., Dike, A., Drozd, A., Edwards, K., Forbes, S., Foster, R., Gray, K., Greenman, C., Halliday, K., Hills, K., Kosmidou, V., Lugg, R., Menzies, A., Perry, J., Petty, R., Raine, K., Ratford, L., Shepherd, R., Small, A., Stephens, Y., Tofts, C., Varian, J., West, S., Widaa, S., Yates, A., Brasseur, F., Cooper, C.S., Flanagan, A.M., Knowles, M., Leung, S.Y., Louis, D.N., Looijenga, L.H., Malkowicz, B., Pierotti, M.A., Teh, B., Chenevix-Trench, G., Weber, B.L., Yuen, S.T., Harris, G., Goldstraw, P., Nicholson, A.G., Futreal, P.A., Wooster, R., Stratton, M.R.: Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* 431(7008) (2004) 525–6
19. Stirewalt, D.L., Radich, J.P.: Malignancy: Tumor Suppressor Gene Aberrations in Acute Myelogenous Leukemia. *Hematology* 5(1) (2000) 15–25
20. Suzuki, J.: Learning Bayesian Belief Networks Based on the MDL principle: An Efficient Algorithm Using the Branch and Bound Technique. *IEICE Transactions on Communications Electronics Information and Systems* (1999)
21. Verstovsek S., Kantarjian H., Estey E., Aguayo A., Giles F.J., Manshoury T., Koller C., Estrov Z., Freireich E., Keating M., Albitar M.: Plasma hepatocyte growth factor is a prognostic factor in patients with acute myeloid leukemia but not in patients with myelodysplastic syndrome. *Leukemia* 15(8) (2001) 1165–70
22. Zhou, B.P., Liao, Y., Xia, W., Zou, Y., Spohn, B., Hung, M.C.: HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nat Cell Biol.* 3(11) (2001) 973–82