MicroRNAs Analysis by Hypothesis Finding Technics

Andrei Doncescu[1,2], Katsumi Inoue[1], and Anne Pradine[2]

[1] National Institute of Informatics Japan
[2] Cancer Institute of Toulouse, France

Abstract. The cell is an entity composed of several thousand types of interacting proteins. Our goal is to comprehend the cancer regulation mechanisms using the microRNAs. MicroRNAs are present in almost all genetic regulatory networks acting as inhibitors targeting mRNAs. In this paper, it is shown how the Artificial Intelligence description method functioning on the basis of Inductive Logic Programming can be used successfully to describe essential aspects of cancer mechanisms. The results obtained show new microRNAs markers for melanoma metastasis.

1 Introduction

New technologies have been developed developed to measure the expression level of thousands of genes simultaneously. These genomic-scale snapshots of gene expression (i.e. how much each gene is "turned on") are creating a revolution in in biology. Genes encode proteins, some of which in turn regulate other genes. Understanding genetic and metabolic networks is of the utmost importance. These networks control essential cellular processes and the production of important metabolites in microorganisms, and modeling such networks from model organisms will drive applications to other less characterized organisms, which have a high biotechnological potential.

The study of signaling events appears to be a key to the research, biological, pharmacological and medical. The spread of these types of signals are not changing the behavior of proteins on three levels: regulation of the activity, interaction and expression. The three levels are synchronized in a strong momentum that leads to changes in protein activity. Since a decade signaling networks have been studied using analytical methods based on the recognition of proteins by specific antibodies. Parallel DNA chips (microarrays) are widely used to study the co-expression of candidate genes to explain the etiology of certain diseases, including cancer.

From the standpoint of Artificial Intelligence cells are sources of information that include a myriad of intra and extra cellular signals that as the ultimate goal of optimal output describing metabolic proteins. Diseases and cancer in particular can be seen as a pathological alteration in the signaling networks of the cell.

The study of gene networks poses problems identified and studied in Artificial Intelligence over the last twenty years. It is clear identified that the reasoning have to handle incomplete, uncertain, revisable, contradictory and multiple sources of information. The logical representation of signaling pathways is not complete: biological experiments provide a number of protein interactions but certainly not all. On the other hand the conditions and difficulty of some experiments lead to obtain data which are not always accurate. Some data may be very wrong and must be corrected or revised in the future. Finally the information coming from different sources and experiences and can be contradictory. From this observation, generally for most real human activities, it must, however, reason and make decisions. It is the goal of logics of uncertainty and in particular of nonmonotonic logics.

The logical approach provides an intuitive method to provide explanations based on the expressivity of relational language. For example, logic can represent biological networks such as gene regulation, signaling transduction, and metabolic pathways. Unlike other approaches, this method lets us introduce background theory, observations and hypotheses within a common declarative language. It also provides the basis for the three main forms of inference, i.e., deduction (prediction), abduction (explanation) and induction (generalization).

In a quarter of a century of automatic demonstration, propositional calculus has been largely neglected in favor of more substantial logics, such as predicate calculus or certain kind of non-classical logic. In the case of propositional calculus, the algorithms of automatic demonstration, be they of syntactical or semantic nature, are characterized by their great simplicity. It is thus easy to analyze their limits and weak point, and to work towards an improvement in their performance.

Another fundamental asset of propositional calculus is its decidability : we can implement algorithms which will give a result within a finite and reasonable time, even for non-trivial examples. In first-order logic, such algorithms cannot exist.

1.1   Causality and classical inference

If the inference of classical logic $A \rightarrow B$ or $A \dashv B$ is fully described formally, with all the "good" logic properties (tautology, not contradiction, transitivity, contraposition, modus ponens, ...), a description of the properties of causality is less simple. Causality can not be seen as a classical logic relation. In this paper, it is described and use a very simple form of causality necessary and probably sufficient for the application to the cell.

To provide the causal links between our relations cause and blocks in a classical language (propositional calculus or first order logic) it is necessary to do two things :

1. Describe the internal characteristics of relations and causes
2. Describe the links between these relations and classical logic

Taking into account the aspect of discovery (abduction, field production) the hypothesis theory is applied. The default logic raises a number of theoretical and practical problems. To solve these problems we will use the Hypotheses Theory, wich is a nonmonotonic logic. A fragment of this logic is close to stable model. One advantage of hypotheses theory is that the information is described in a classic bimodal logic. It will then be possible to use this formalism to describe the set of all information, uncertain or not. This logic will also allow the use a consequence finding algorithms to treat abduction.

2 Integrating induction and abduction in CF-induction

Both *induction* and *abduction* are ampliative reasoning, and agree with the logic to seek hypotheses which account for given observations or examples. That is, given a background theory B and observations (or positive examples) E, the task of induction and abduction is common in finding a hypothesis H such that

$$B \wedge H \models E, \tag{1}$$

where $B \wedge H$ is consistent. There are several discussions on the difference between abduction and induction in the philosophical and pragmatic levels. On the computational side, induction usually involves *generalization*, while abduction gives *minimal explanations* for individual observation.

Inverse entailment (IE) is a logically principled way to compute abductive and inductive hypotheses H in (1) based on the logically equivalent transformation of the equation (1) to

$$B \wedge \neg E \models \neg H. \tag{2}$$

The equation (2) says that, given B and E, any hypothesis H *deductively* follows from $B \wedge \neg E$ in its negated form. The equation (2) is seen in literature, e.g., [12] for abduction and [15] for induction. The equation (2) is useful for computing abductive explanations of observations in abduction. This is because, without loss of generality, in abduction E is written as a ground atom, and each H is usually assumed to be a conjunction of literals. These conditions make abductive computation relatively easy, and *consequence-finding* algorithms [12] can be directly applied.

In induction, however, E can be clauses and H is usually a general rule. Universally quantified rules for H cannot be easily obtained from the negation of consequences of $B \wedge \neg E$. Then, Muggleton [15] introduced a *bridge formula* U between $B \wedge \neg E$ and $\neg H$:

$$B \wedge \neg E \models U, \quad U \models \neg H.$$

As such a bridge formula U, Muggleton considers the conjunction of all unit clauses that are entailed by $B \wedge \neg E$. In this case, $\neg U$ is a clause called the *bottom clause* $\perp(B, E)$. A hypothesis H is then constructed by generalizing a sub-clause of $\perp(B, E)$, i.e., $H \models \perp(B, E)$. This method with $\perp(B, E)$ is adopted

in Progol, but it has turned out that it is incomplete for finding hypotheses satisfying (1).

In [13], Inoue proposed a simple, yet powerful method to handle inverse entailment (2) for computing inductive hypotheses. The resulting method called *CF-induction* does not restrict the bridge formula U as the set of literals entailed by $B \wedge \neg E$, but consider the *characteristic clauses* [12] of $B \wedge \neg E$, which obviously generalizes the method of the bottom clause. CF-induction then realizes sound and complete hypothesis finding from *full clausal theories*, and not only definite clauses but also non-Horn clauses and integrity constraints can be constructed as H.

In most previous inductive methods including Progol [15], there are syntactical restrictions such that: (i) each constructed hypothesis in H is usually assumed to be a single Horn clause, (ii) an example E is given as a single Horn clause, and (iii) a background theory B is a set of Horn or definite clauses. From the viewpoint of applications, these restrictions are due to the easiness for handling such formulas. An extension to multiple non-Horn clauses in B, E, and H is, however, useful in many applications.

First, an extension allowing *multiple clauses* in either a hypothesis H or an observation E is essential in applications of abduction. In fact, recent work on abductive inference in metabolic pathways [16] uses an independent abductive procedure to obtain a set of literals that explain an observation. In general, there are multiple missing data to account for an observation. This abductive inference is independently computed in [16] not by Progol, and the inductive process for generalization takes place by Progol only after abductive hypotheses have been obtained. On the other hand, CF-induction can be used to compute abductive explanations simply by taking the bridge formula U as a deduced clause. CF-induction thus integrates induction and abduction from the viewpoint of inverse entailment through consequence-finding [13].

Second, an extension to *non-Horn clauses* in representation of B, E, H is also useful in many applications. For example, indefinite statements can be represented by disjunctions with more than one positive literals, and integrity constraints are usually represented as negative clauses. The clausal form is also useful to represent causality. For example, when we want to represent *inhibition* of reaction in a causal pathway network, positive and negative literals with the predicate like `inhibited` can be used in the premise of each causal rule, which results in a non-Horn clause (we will see an example afterwards). Again, the inductive machinery of CF-induction can handle all such extended classes.

Third, introducing *multiple, non-Horn clauses* in a hypothesis H is an unifying extension that combines the first and second extensions. In this case, a hypothesis H forms a *theory*, which can also account for multiple observed data at once.

In our application to cancer analysis, given the background theory of network structures of a pathway and observations, we need a hypothesis H that explains the behavior of the metabolic system. In principle, such a hypothesis consists

of multiple non-Horn clauses each of which represents a causal relation. CF-induction is thus particularly useful for this type of applications.

Here the structure of representation is based on the specification of CF-induction program, which is compatible with the consequence-finding program SOLAR [19] and the TPTP format for theorem proving. SOLAR is a Java implementation of the tableaux variant of SOL resolution [12].

For example, the input clauses can be described as

```
input_clause(axiom1,    bg, [-p(X), -q(X), r(X)]).
input_clause(example1, obs, [r(a)]).
production_field([predicates(pos_all), length < 3]).
```

Here, `axiom1` and `example1` are ID names of clauses, and `bg` and `obs` represent background knowledge and observation, respectively. The axiom1 means $\rceil p(x) \vee \rceil q(x) \vee r(x)$. Each clause is represented as a list of literals. The predicate `production_field` indicates the *production field* of SOLAR, and this example allows it to generate consequences consisting of less than 2 positive literals. In this way, a production field can be used to specify an *inductive bias* in CF-induction. There is other meta information to control deduction in SOLAR such as the search strategy and the depth limit.In this case, CF-induction produces the abductive hypothesis:

```
Hypotheses: [ [p(a)], [q(a)] ]
```

The current CF-induction program has several *generalizers*, which, given a set T of clauses, produce a set S of clauses such that $S \models T$. These basic generalizers include *anti-instantiation*, *reverse Skolemization*, *Plotkin's least generalization*, and *dropping literals* (see [13]).

Therefore, CF-induction is related to top-down decision tree learning algorithm generating a set of rules in the form of predicate logic clauses which can be used to separate the classes.

2.1   Signaling Pathway Representation

In this paper, it is used only a propositional representation. In practice the detailed study of interactions will be asked to represent increases or decrease some biological quantities which could be protein concentration. It therefore falls outside the scope of propositional but the basic problems are the same. To represent a change in concentration is for example possible by using predicates such as "increased" or "decrease" [3].

To describe interactions between genes/proteins we use a language L of classical logic (propositional). The proposition A (resp. $\neg A$) says that A is true (false). We are in a logical framework, so it is possible to represent almost everything in a natural way. Interactions between genes/proteins is a very simple form of causality.

3 MicroRNAs

MicroRNAs are small RNA molecules that were discovered in the 1990s in animals and plants, and which play an essential role in controlling gene expression. In human being, more than 500 microRNAs have been identified, and we now know that their dysfunction is associated with several diseases, including cancer. The microRNAS play an important role of not specific inhibition in many circumstances of the cell life, like chromatin clock control and have a big influence on many metabolic controls of functions like energy systems, cell cycle and defense systems against pathogens.

MicroRNAs are present in almost all genetic regulatory networks acting as inhibitors targeting mRNAs, by hybridizing at most one of their triplets, hence acting as translation factors by preventing the protein elongation in the ribosome. MicroRNAs act as source nodes in the interaction graph of genetic regulatory networks, which are made of elements, the genes, in interaction through the protein they express and control important cell or tissue functions like proliferation, differentiation, energy systems maintenance, and more generally homeostasis [3,4].

3.1 MicroRNAs Identification in Melanoma

The microRNAs appear increasingly as crucial actors in oncogenesis (miR-21 in breast cancer) or as a tumor suppressor. Furthermore, the expression profiles of microRNAs in solid tumors of different origins have been made using prognostic values.

In 2008 the circulating microRNA were revealed for the first time in serum and plasma, with very different profiles between healthy donors (which have a similar expression profile) and patients with breast cancer, lung , prostate with specific expression patterns. In addition, these microRNAs have a high stability since they form complexes with lipids or lipoproteins, allowing the resistance to the activity of RNase and DNase. They can also withstand harsh experimental conditions such as high temperature or high pH variation. These data therefore show the circulating microRNAs (noninvasive) could be novel plasma biomarkers in oncology.

Melanoma is a cancer of the skin or mucous membranes, developed at the expense of melanocytes. In most cases, it develops first on the skin but it is common to find melanoma of the eye (choroidal melanoma), mucous membranes (mouth, anal canal, vagina), or even more rarely internal organs. The incidence of this disease is increasing worldwide. In France, it was evaluated in 2010 at approximately 7-9 new cases per year per 100 000 individuals with a mortality rate ranging from 1.2 to 1.5 individuals.

For this study, we collected 29 subjects having metastatic melanomas and 5 healthy volunteers. The micro-arrays are spotted with microRNA of human, murine, and viral control. In all, there were 4608 different microRNA spots and 2000 different microRNA.

In order to highlight the existence of circulating microRNA biomarkers in melanoma, a technique was developed for extracting microRNAs from the plasma of healthy donors. This lets us collect a larger amount purified RNA from the serum). The expression profile of microRNA was obtained using by a technique of monochrome chip. These chips were spotted (labeled with fluorescent HY3) with probes modified with bases LNA (Locked Nucleic Acid home Exiqon). The goal was to optimize the specificity and sensitivity of probes (annealing temperature adapted to detect microRNA with a low percentage of GC). Each microRNA was represented by two different spots. These chips were analyzed using image analysis software GenePix Pro 6.1.0.4 (Axon Instruments).

3.2 MicroRNAs expressed between 2 populations : healthy subjects and patients with melanoma

Using Bayesian approach (Limma) [17] , the original human microRNA differentially expressed between the two groups of subjects were identified (Fig.1).

Name	logFC	AveExpr	t	P.Value	adj.P.Val	Log-odds
hsa-miR-323-5p/mmu-miR-323-5p/rno-miR-323*	-0,48376	1,41123	-12,52041	0,00000	0,00164	4,83161
hsa-miR-1290	-0,08370	1,87079	-4,86814	0,00002	0,01032	2,35283
hsa-miR-1283	0,27753	1,44350	8,80186	0,00005	0,01768	2,31022
hsa-miR-302a*	-0,48901	1,43369	-9,76053	0,00018	0,03215	0,96393
hsa-miR-299-3p	0,05898	1,66613	4,41086	0,00008	0,02171	0,95292
hsa-miR-191*	0,15563	1,46111	5,31796	0,00033	0,04484	0,60907
hsa-miR-1246	-0,06650	1,85606	-4,16537	0,00016	0,03215	0,22045
hsa-miR-766	0,05295	1,69399	4,11137	0,00020	0,03215	0,05787
hsa-miRPlus-F1221	0,05284	1,61874	3,91631	0,00035	0,04484	-0,52358

Fig. 1. The different microRNAs expressed between the 2 populations (healthy subjects and patients with melanoma).

4 Discretization of Continous Values

Discretizing time series is a research domain on its own and many works [6, 23] have been conducted recently. Our practical problem is that we want to have a statistically relevant (unsuppervised) discretization for N chemical compounds concentrations over time. For that purpose, we compute an appropriate number of levels (5) in regard to a Bayesian score such as Bayesian Information Criterion [24].

We use continuous (Gaussian) hidden Markov models with parameter tying, which means that each chemical compound has a corresponding HMM but all the N Gaussian HMM share the same parameters (means and covariances), to share the same discrete outputed levels between the different compounds of one experiment. This relevant discretized levels of concentration are computed through expectation maximisation with maximum a posteriori [3, 7] . The level of microRNA expression are represented as : *very low, low, medium, high, very high*

4.1 Analysis of microRNAs expressed in melanoma

The highly aggressive character of melanoma makes it an excellent model to probe the mechanisms underlying metastasis, the process by which cancer cells travel from the primary tumor to distant sites in the body. Therefore, the goal of this analysis is to find out a microRNA signature in the case of aggressive melanoma. This signature is represented by a logical clause including the relevant microRNA, age and Breslow index. One hundred thirsty microRNAs have been identified and have been selected for modelling but only 51 has at least five valid replicas. After identifying 23 microRNAs, which already have extensively been studied a file containing the among of these microRNA compared with the healthy persons was created. The 23 microRNA candidates to melonama signature are :   *hsa-let-7b, hsa-miR-17, hsa-miR-18a, hsa-miR-20a, hsa-miR-21, hsa-miR-34a, hsa-miR-130a, hsa-miR-141, hsa-miR-143, hsa-miR-145, hsa-miR-146a, hsa-miR-152, hsa-miR-155, hsa-miR-185, hsa-miR-191, hsa-miR-200c, hsa-miR-221, hsa-miR-222, hsa-miR-338-3p, hsa-miR-1246, hsa-miR-1290, hsa-miR-2110, miR-27b.*

The representation contains information about microRNAs (23), Breslow index, age, relapse and metastasis or die.

A logical predicate *expressed* for each patient has been defined. Each patient is characterized by 23 microRNAs. An important aspect of the representation is to consider the clause as background in CF-induction (bg).

```
Example1: [input_clause(patient11,bg,[hsamiR20a(patient1,medium),
hsamiR21(patient1,medium),
...
hsamiR34a(patient1,low),hsamiR222(patient1,medium),hsamiR3383p(patient1,low)]). ]
```

Concerning the representation of Breslow Index we decided to introduce a predicate *breslow*

```
input_clause(patient17, obs, [breslow(patient1,medium)]).
```

The age of patients has been discretized according with the next empirical information provided by doctors (low: 1-30, medium: 31-60 and high: 61 and more) and the clause is similar to Breslow index. Relapse is an important clinical parameter related to the number of days that the patient kept his "health". It is given as low (1-500 days), medium (501-1000 days) and high (more of 1001 days).

The predicate *relapsed(X,Y)* contains 2 atoms : the first one is the patient and the second parameter gives information about the number of days considered "healthy".

The expert knowledge is represented by the microRNA well known in melanoma development or invasive cancer.

```
input_clause(bg1, bg, [-hsamiR20a(X,high),
-hsamiR21(X,high), ...
-hsmaiR34a(X,low),
relapsed(X,high)]).
```

Basically, melanomas, expressions of miRNA-20a, miRNA-106a, miRNA-17, miRNA-21, and miRNA-34a are significantly up-regulated, while miRNA-145 and miRNA-204 expression are significantly down-regulated. In our case we have considered the microRNA-34a down-regulated. microRNA-34a maps to a chromosome 1p36 region that is commonly deleted and it has been found to act as a tumor suppressor through targeting of numerous genes associated with cell proliferation and apoptosis. The patients analyzed have advanced ages and our statistical analysis showed a low expression in this case.

4.2 Results in microRNA signature

The goal of our logical model is to identify the patients who have a fast evolution of cancer. We have used CF-Induction to obtain diagnosis rules based on these predicates. CF-Induction searches for a set of st-order predicate logic clauses that distinguish between two classes, one which is presented to the learner as positive and negative examples. CF-Induction is independent of the ordering of the positive examples. Another important aspect is the possibility to check whether B and H are consistent or not.

The most noteworthy result is the discover of a signature in the case of aggressive melanoma represented by 5 microRNAs : 20a,21,34a, 222 and 333-3p. This results has been obtained from the analysis of the number of examples explained by these microRNA and compared with the the results published by PubMed. The last two microRNAs are not well-known having a strong oncogene or anti-oncogene regulation.

The idea was to produce all predicates : beginquote

```
production_field(
[
predicates(all)
]
).
```

```
[relapsed(patient25, low), -hsamiR20a(patient25, medium)]
```

```
[relapsed(patient25, low), -hsamiR21(patient25, medium)]
[relapsed(patient25, low), -hsamiR34a(patient25, low)]
[relapsed(patient25, low), -hsamiR222(patient25, medium)]
[relapsed(patient25, low), -hsamiR3383p(patient25, low)]
```

Patient 25 was born in 1947 and died in 2011. The screening was done during stage iV of his cancer.

Using CF-induction with the production field :

```
production_field(
  [
    length <= 8,
    term_depth < 3,
  predicates([hsamiR222(_,_),hsmaiR3383p(_,_),age(_,_),relapsed(_,_)])
  ]
).
```

we have obtained the next result.

```
B & H is consistent

Hypotheses:
[hsamiR21(patient25, low), hsamiR3383p(patient25, low),
hsamiR222(patient25, verylow), hsamiR20a(patient25, low),
-age(patient25, high)]
```

This means the patient's age explains the observation of 4 microRNA: microRNA 21,20a, 222 and 338-3p. Therefore if we corroborate the 2 results, the microRNA 34a is a potential marker of melanoma in the case of older people. Basically microRNA 34a, maps to a chromosome 1p36 region that is commonly deleted and it has been found to act as a tumor suppressor through targeting of numerous genes associated with cell proliferation and apoptosis. A similar result has been obtained for the patient 24, who is born in 1932 and he is still living. Her screening was done during stage III of her cancer. It seems a strong correlation may exist between age and the microRNA 34a.

Another interesting result concerns patient 17, who has a very low concentration of microRNA 338-3p. At first glance, this could be considered a cancer marker, but the next results :

```
[relapsed(patient17, high), -age(patient17, medium)]
[hsamiR3383p(patient17, veryverylow), -age(patient17, medium)]
```

show that in this case age is the most important parameter of melanoma evolution. This person was born in 1962; the melanoma was identified for the first time in 2006. 6 years later she remains in phase IV.

5 Conclusion

Understanding genetic and metabolic networks is of the utmost importance. With the development of DNA microarrays, it is possible to simultaneously analyze the expression of up to thousands of genes and to construct gene networks based on inferences over gene expression data. We have found in this study that 5 microRNAs : 20a,21,34a, 222 and 333-3p are an important indicator of melanoma evolution. The microRNA 333-3p is not known as significantly down-regulated in the case of aggressive melanoma. Another surprising result is the correlation between age and microRNA 34a. Therefore, we are confident that future work will let us appreciate the complexity of micro-RNAs in the case of the invasive melanoma and going further to find out a signature of homeostasis. Homeostasis, the subtle balance between proliferation, differentiation and cell death of an organism is a complex phenomenon still little understood. Researchers are pointing more and more the important role of micro-RNA in this phenomenon.

References

1. Bandiera, S., Rberg, S., Girard, M., Cagnard, N., Hanein, S., Chrtien, D., Munnich, A., Lyonnet, S., Henrion-Caude, A., 2011. Nuclear Outsourcing of RNA Interference Components to Human Mitochondria. PLoS ONE 6, e20746.
2. Barbarotto E, Schmittgen TD, Calin GA. MicroRNAs and cancer: profile, profile, profile. Int J Cancer 2008; 122 (5): 969-77.
3. Beal M. J.: Variational Algorithms for Approximate Bayesian Inference. PhD. Thesis, Gatsby Computational Neuroscience Unit, University College London (2003).
4. A. Doncescu,Y. Yamamoto, K. Inoue. Biological Systems Analysis using Inductive Logic Programming *The 2007 IEEE International Symposium on Bioinformatics and Life Science Computing BLSC07*, Niagara Fall, Ontario, Canada, 2007.
5. Cummins JM, Velculescu VE. Implications of micro-RNA profiling for cancer diagnosis. Oncogene 2006; 25 (46): 6220-7.
6. Geurts P.: Pattern extraction for time-series classification. Proceedings of PKDD 2001, 5th European Conference on Principles of Data Mining and Knowledge Discovery, LNAI 2168, 115-127 (2001).
7. Gauvain J.-L. and Lee C.-H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Transactions on Speech and Audio Processing, Vol.2, Issue 2, pp.291-298 (1994).
8. Ji S., Krishnapuram B., and Carin L.: Variational Bayes for continuous hidden Markov models and its application to active learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.28, Issue 4, pp.522-532 (2006).
9. R. King, K. Whelan, F. Jones, P. Reiser, C. Bryant, S. Muggleton, D. Kell, and S. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427:247–252, 2004.
10. Kitano H.: Systems Biology Toward System-level Understanding of Biological Systems Kitano. In Science Vol. 295. no. 5560, pp. 1662-1664 (2002).
11. K. Inoue: Induction as consequence finding. Machine Learning, 55:109–135, 2004.
12. K. Inoue. Linear resolution for consequence finding. *Artificial Intelligence*, 56:301–353, 1992.
13. K. Inoue. Induction as consequence finding. *Machine Learning*, 55:109–135, 2004.

14. R. J. Mooney Integrating abduction and induction in machine learning. In Working Notes of the IJCAI97 Workshop on Abduction and Induction in AI, 37–42 (1997).

15. S. Muggleton. Inverse entailment and Progol. *New Gen. Comput.*, 13:245–862, 1995.

16. A. Tamaddoni-Nezhad, R. Chaleil, A. Kakas and S. Muggleton. Application of abductive ILP to learning metabolic network inhibition from temporal data. Machine Learning, 64:209–230, 2006.

17. Smyth, G. K. (2005). Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York*, pages 39720.

18. Nabeshima H., Iwanuma K., and Inoue K.: SOLAR: A Consequence Finding System for Advanced Reasoning. Proceedings of the 11th International Conference TABLEAUX 2003, Lecture Notes in Artificial Intelligence, Vol. 2796, pp. 257-263, Springer (2003).

19. H. Nabeshima, K. Iwanuma and K. Inoue. SOLAR: a consequence finding system for advanced reasoning. *Proc. Eleventh Int. Conf. Automated Reasoning with Analytic Tableaux and Related Methods*, *Proc. TABLEAUX 2003*, LNAI 2796, pages 257–263, Springer, 2003.

20. Schultz J, Lorenz P, Gross G, Ibrahim S, Kunz M. MicroRNA let-7b targets important cell cycle molecules in malignant melanoma cells and interferes with anchorage-independent growth. Cell Res 2008; 18 (5): 549-57.

21. Yamamoto Y., Inoue K.,Doncescu A. : ntegrating abduction and induction in biological inference using CF-induction Elements of Computational Systems Biology by Huma M. Lodhi and Stephen H. Muggleton Wiley Book

22. Y. Yamamoto, K. Inoue, A. Doncescu : "Abductive Reasoning in Cancer Therapy" IEEE AINA 2009, May 26-29, Bradford

23. Keogh, E. and Lin, J. and Fu, A.: HOT SAX: efficiently finding the most unusual time series subsequence. 5th IEEE International Conference on Data Mining (2005).

24. Schwarz G.: Estimating the dimension of a model. Annals of Statistics, Vol.6, No.2, pp.461-464 (1978).