MicroRNAs Analysis by Hypothesis Finding Technics

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

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

1 Introduction

The study of signaling events appears to be key in biological, pharmacological and medical research. The spread of these types of signals do not change the behavior of proteins : regulation of the activity, interaction and expression. These 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.

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). Deduction has traditionally been used for proving theorems of a given axiom set, but here we need to find new consequences (consequence finding) which is more general than theorem proving. Interestingly, the hypothesis-finding problem (abduction and induction) can be translated into consequence-finding problems, so that we can realize all three modes of inference using a deductive, consequence-finding procedure. We mention the Inductive Inference as a type of reasoning that justifies some modifications from one state of absolute belief to another by adding new information to the initial assumptions that is consistent with it but does not entail it.

2 Integrating induction and abduction in CF-induction

In [9], Inoue proposed a simple, yet powerful method to handle inverse entailment 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* [8] 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 [11], 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. Here the structure of representation is based on the specification of CF-induction program, which is compatible with the consequence-finding program SOLAR [14] and the TPTP format for theorem proving. SOLAR is a Java implementation of the tableaux variant of SOL resolution [8].

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 [9]). 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.

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 defence 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 14 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 miRNA spots and 2000 different microRNA.

In order to highlight the existence of circulating miRNA 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), 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).

3.3 Analysis of microRNAs expressed in melanoma

One hundred thirsty microRNAs have been identified and have been selected for modelling but only 51 has at least five valid replicas. After identifying these 51 microRNAs, a file containing the presence or absence of these miRNA for each individual patient was created.

We have deliberately chosen to consider only the presence or absence of miRNA and not the intensities to limit the effect related to the average quality of the micro-arrays used for the experiment.

A logical predicate *expressed* for each patient has been defined. Each patient is characterized by 51 microRNAs. An atom of this predicate concerns the metastases.

```
Example1: [ [patient1, microRNA1,levelofexpression, 1 (metastases) ]
Example2: [ [patient1, microRNA2,levelofexpression, 1 (metastases) ]
```

The goal of our logical model is to identify the patients who have a fast evolution of cancer. This approach is different from our previous work concerning the proteins marking the breast cancer [17].

We have used the CF-Induction to induce classiation rules based on these predicates. CF-Induction searches for a set of st-order predicate logic clauses

Fig. 2. Analysis of microRNAs .

that distinguish between two classes, one which is presented to the learner as positive and second as negative examples. CF-Induction is independent of the ordering of the positive examples. The results showed that we have 3 groups of patients. First one contains patients 11,13 and 14. This group is considered by our approach as the fastest cancer evolution. The other groups are obtained after the extraction of the mentioned patients from database and reconsidering the problem. The analysis of the patients by RMI indicated that the patients : 6, 11, 13, 14 et 15 had a fast evolution of metastases. The most interesting result is this group of patients has mir182/181 over-expressed but less than hsa-miR-630. Therefore, microRNA-630 and microRNA-182/181 allow melanoma cells to migrate and survive independently, two properties necessary for metastasis.

4 Conclusion

Understanding genetic and metabolic networks is of the utmost importance. These networks control essential cellular processes and the production of important metabolites in microorganisms. 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. One of the secrets of life is certainly in the "homeostasis", the subtle balance between proliferation, differentiation and cell death of an organism. In this complex phenomenon and still unknown, researchers are pointing to the importance of micro-RNA.

We find in this study that miR-630 could be an important indicator of metastases. The other microRNA, the 182, is an well known up-regulatory in human melanoma cell lines and tissue samples. Therefore, we are confident that the feature work concerning the patient signature allows to find the complex of micro-RNAs in the case of the metastasis melanoma.

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. 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.
4. Cummins JM, Velculescu VE. Implications of micro-RNA profiling for cancer diagnosis. Oncogene 2006; 25 (46): 6220-7.
5. 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.
6. Kitano H.: Systems Biology Toward System-level Understanding of Biological Systems Kitano. In Science Vol. 295. no. 5560, pp. 1662-1664 (2002).
7. K. Inoue: Induction as consequence finding. Machine Learning, 55:109–135, 2004.
8. K. Inoue. Linear resolution for consequence finding. *Artificial Intelligence*, 56:301–353, 1992.
9. K. Inoue. Induction as consequence finding. *Machine Learning*, 55:109–135, 2004.
10. 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). luwer Academic Press
11. S. Muggleton. Inverse entailment and Progol. *New Gen. Comput.*, 13:245–862, 1995.
12. 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.
13. 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).
14. H. Nabeshima, K. Iwanuma and K. Inoue. SOLAR: a consequence finding system for advanced reasoning. *Proc. TABLEAUX 2003*, LNAI 2796, pages 257–263, Springer, 2003.
15. 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.
16. 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
17. Y. Yamamoto, K. Inoue, A. Doncescu : "Abductive Reasoning in Cancer Therapy" IEEE AINA 2009, May 26-29, Bradford