

# Induction of Fuzzy Rules with Artificial Immune Systems in aCGH based ER Status Breast Cancer Characterization

Filippo Menolascina<sup>12</sup>,  
Patrizia Chiarappa<sup>1</sup>,  
Stefania Tommasi<sup>1</sup>,  
Angelo Paradiso<sup>1</sup>  
<sup>1</sup>National Cancer Institute,  
Via F. Hahnemann, 10  
70126, Bari - Italy  
f.menolascina@ieee.org

Roberto T. Alves<sup>3</sup>,  
Myriam Delgado<sup>3</sup>,  
Alex Freitas<sup>4</sup>  
<sup>3</sup>Federal University of Technology of  
Paraná, Av. 7 de Setembro, 3165  
Curitiba – Brazil  
<sup>4</sup>University of Kent, Canterbury, Kent,  
CT2 7NF, UK  
ralves@cpgei.cefetpr.br

Vitoantonio Bevilacqua<sup>2</sup>,  
Giuseppe Mastronardi<sup>2</sup>  
<sup>2</sup>Polytechnic of Bari,  
Via E. Orabona, 4  
70125, Bari - Italy  
bevilacqua@poliba.it

## ABSTRACT

Genomic DNA copy number aberrations are frequent in solid tumours although their underlying causes remain obscure. In this paper we show how Artificial Immune System (AIS) paradigm can be successfully employed in the elucidation of biological dynamics of cancerous processes using a novel fuzzy rule induction system for data mining (IFRAIS) [1]. Competitive results have been obtained using IFRAIS. A biological interpretation of the results, carried out using Gene Ontology, followed the statistical assessment and put in evidence interesting patterns that are currently under investigation.

## Categories and Subject Descriptors

I.5.2 [Pattern Recognition]: Design Methodology – Classifier design and evaluation, feature design and evaluation, Pattern Analysis I.2.6 [Artificial Intelligence]: Learning – Concept Learning and Induction

**General Terms:** Algorithms, Design, Experimentation.

## Keywords

aCGH, AIS, Breast Cancer, Data Mining, IFRAIS, Fuzzy Rules

## 1. INTRODUCTION

Breast cancer (BC) is the most extensively investigated. Recently researchers focused on epigenetic events underlying cancer (amplifications and deletions) through a methodology called aCGH. Using aCGH we are able to map duplications or losses at the gene level; it obvious, then, that investigation of similar events can point out genetic actors in the development of cancer. A comparative study of J48, Ant Miner, GEP and IFRAIS is performed to establish relative and absolute performances of the systems under investigations. IFRAIS' results have been validated using statistical and knowledge driven approaches using Gene Ontology through GO Miner [2].

## 2. MATERIALS AND METHODS

In this study a cohort of 124 patients has been considered. Nucleic acids were extracted from tumour blocks as described elsewhere

[3]. aCGH protocol and image analysis steps followed. Data were pre-processed for normalization and gene selection has been carried out using a consensus scheme based on Student's T-Test, ROC and Kullback-Lieber divergence derived statistical ranking.

## 3. ALGORITHM

### 3.1 IFRAIS

The most important characteristic of IFRAIS (Induction of Fuzzy Rules with an Artificial Immune System) is that it discovers fuzzy classification rules [1]. This fuzzy format to rules is naturally comprehensible to human experts. Nowadays, comprehensible knowledge is essential in real-world data mining problems (e.g. in bioinformatics).

## 4. RESULTS AND CONCLUSIONS

IFRAIS has been used to induce rules describing ER+/- status BC. Eight rules have been obtained [4]: further investigations on these rules put in evidence several interesting pathways and genes have been highlighted whose function, assessed using GO, and role in breast cancer ER status determination is currently under investigation. The contemporary presence of estrogen metabolism related genes, PDGF/VEGF (known to be involved in angiogenesis and vascularisation of tissues) family of endothelial growth factors and inflammatory response pathways pushes the interest for further researches on correlation between ER status and tumour aggressiveness. We can conclude that the experimental pipeline described appears to return results reasonably correlated with processes expected to result highlighted.

## 5. REFERENCES

- [1] Alves, RT., Delgado, MR., Lopes, HS., Freitas, AA. An artificial immune system for fuzzy-rule induction in data mining. (PPSN04), LNCS 3242: 1011–1020, 2004.
- [2] Barry R. Zeeberg, et al, GoMiner: A Resource for Biological Interpretation of Genomic and Proteomic Data. *Genome Biology*, 2003 4(4):R28
- [3] Albertson DG. et al, Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene, *Nature Genetics* 25, 144 - 146 (2000)
- [4] Supplementary Material:  
<http://oncologico.bari.it/laboratorio/bioinformaticsnews/>