Discovery of Genes Implied in Cancer by Genetic Algorithms and Association Rules

Alejandro Sánchez Medina[1], Alberto Gil Pichardo[1],
Jose Manuel García-Heredia[2], and María Martínez-Ballesteros[3](✉)

[1] University of Sevilla, Seville, Spain
alesanmed@alum.us.es, albgilpic@alum.us.es
[2] Department of Vegetal Biochemistry and Molecular Biology,
University of Seville, Seville, Spain
jmgheredia@us.es
[3] Department of Computer Science, University of Seville, Seville, Spain
mariamartinez@us.es

Abstract. This work proposes a methodology to identify genes highly related with cancer. In particular, a multi-objective evolutionary algorithm named CANGAR is applied to obtain quantitative association rules. This kind of rules are used to identify dependencies between genes and their expression levels. Hierarchical cluster analysis, fold-change and review of the literature have been considered to validate the relevance of the results obtained. The results show that the reported genes are consistent with prior knowledge and able to characterize cancer colon patients.

Keywords: Data mining · Association rules · Gene expression · Cancer

1 Introduction

The word cancer refers to a set of different complex diseases, characterized by an uncontrolled and pathogenic growth of cells as a tumor, driving to death without medical treatment. Due to the increase of life expectancy, cancer can be considered as the 21^{st} century's disease. In fact, around one third of the total population will develop cancer during their lifetime [1]. This high percentage shows the importance of the development of adequate techniques for diagnosis and treatment of the different cancers.

The origin of cancer is mainly genetic, due to the accumulation of mutations in cells during lifetime. This enables them to become independent of the extracellular matrix, losing adherent properties and allowing them to metastasize to other tissues. During the period 1989–2008, around 15 % of all deaths due to malignant tumours were caused by colorectal cancer [2].

Data Mining has been used in cancer context for the diagnosis and prognosis of breast cancer [3]. This technique can help to detect it earlier reducing the mortality risk in patients. Indeed, data mining has been used in breast cancer to predict the survival rate of patients [4]. This allows that patients with lower survival rate may change their habits in order to increase it. Association Rules (hereafter AR) mining is a popular methodology in the field of Data Mining to discover significant and apparently hidden relations among attributes in a subspace of the dataset instances. In fact, the work presented in [5] proposes the application of fuzzy AR to find potential associations among prognostic factors in breast cancer. The well-known Apriori algorithm was applied to mine the AR between clinical factors and good survival outcomes in operation in [6].

The expression microarray technology [7] is composed of chips that show us the expression of a large number of genes of the patient, being impossible to perform a particular and individual study and statistical analysis. The goal of bioinformatics is to find genes involved in cancer progression. This is achieved by using normal and cancerous tissue samples.

In this work we propose a study to identify gene expression patterns in colon cancer microarray data using Quantitative Association rules (hereafter QAR). In particular, we have applied a multi-objective genetic algorithm, henceforth called CANGAR, to discover QAR with the aim to find genes probably associated with cancer according to their expression levels. The resulting genes of CANGAR may be studied in order to find out if they are implied or related in any way with a specific type of cancer.

The rest of the paper is structured as follows. Section 2 summarizes the main concepts of AR and quality measures in addition to the fold-change measures used to select the genes to be analyzed. Section 3 thoroughly describes the CANGAR approach to identify genes highly related to colon cancer patients. Section 4 presents and discuss the results obtained in the analysis proposed. Finally, Sect. 5 summarizes the conclusions drawn from the analysis performed.

2 Preliminaries

This section provides a brief description of QARs and quality measures, in addition to the fold-change measure commonly used in microarray analysis.

2.1 Quantitative Association Rules and Quality Measures

AR are implications like $X \Rightarrow Y$ where $X \bigcap Y = \emptyset$. But there are a subtype of AR that allows also to set an interval of membership for the attributes, those are called QAR [8]. For example, a QAR could be numerically expressed as:

$$GeneA \in [1, 2] \; and \; GeneB \in [3, 4] \Rightarrow GeneC \in [2, 3]$$

The left-side and the right-side of an AR are known as antecedent and consequent, respectively. In this example, *GeneA* and *GeneB* belong to the antecedent and *GeneC* belongs to the consequent.

Several probability-based measures exist to evaluate the generality and reliability of AR (and QAR) [9]. The most frequently measures used are support and confidence [10]. Table 1 summarizes the description and the mathematical definition of measures used in this work [11]. Note that the classical algorithms need to obtain some frequent item-sets before generating AR that satisfy minimum support and confidence thresholds. This is undoubtedly the most costly task in seeking AR. One of the most important algorithms for generating frequent item-sets, is the Apriori algorithm [12], which is based on the fact that if a set is frequent, all of it subsets will be frequent too, and the same in the other way.

The CANGAR algorithm does not require the generation of frequent item-set, since the intervals of QAR are evolved through the evolutionary process of CANGAR until those rules with best measures are obtained. Note that n(AB) is the number of occurrences of the rule $A \Rightarrow B$ in the dataset and $|D|$ is the total number of instances in the dataset.

Table 1. QAR quality measures

Measure	Equation
$Sup(A \Rightarrow B)$	$\frac{n(AB)}{\|D\|}$
$Conf(A \Rightarrow B)$	$\frac{Sup(A \Rightarrow B)}{Sup(A)}$
$Lift(A \Rightarrow B)$	$\frac{Sup(A \Rightarrow B)}{Sup(A)Sup(B)}$
$Conviction(A \Rightarrow B)$	$\frac{1-Sup(B)}{1-Conf(A \Rightarrow B)}$
$Gain(A \Rightarrow B)$	$Conf(A \Rightarrow B) - Sup(B)$
$Certainty\,factor(A \Rightarrow B)$	• If $Conf(A \Rightarrow B) \geq Sup(B)$: $\frac{Gain(A \Rightarrow B)}{1-Sup(B)}$ • If $Conf(A \Rightarrow B) < Sup(B)$: $\frac{Gain(A \Rightarrow B)}{Sup(B)}$
$Leverage(A \Rightarrow B)$	$Sup(A \Rightarrow B) - Sup(A)Sup(B)$
$Accuracy(A \Rightarrow B)$	$Sup(A \Rightarrow B) + Sup(A \Rightarrow B)$
$Coverage(A \Rightarrow B)$	$\frac{Sup(A \Rightarrow B)}{Sup(B)}$
$Netconf(A \Rightarrow B)$	$\frac{Sup(A \Rightarrow B) - Sup(A)Sup(B)}{Sup(A)(1-Sup(A))}$
$Yule's\,Q(A \Rightarrow B)$	$\frac{x-y}{x+y}$ s.t $x = Sup(A \Rightarrow B)(1 - Sup(B))$ $-Sup(A) + Sup(A \Rightarrow B)$ $y = (Sup(A) - Sup(A \Rightarrow B))(Sup(B)$ $-Sup(A \Rightarrow B))$

2.2 Differentially Expressed Genes in Microarray Data

In order to know if the change of expression of a specific gene is significant, cancer samples must be compared with normal samples using a parameter called fold change (FC). In that way, at least two different microarrays experiments must be performed. For the dataset used in this work, normal (control) samples come from the same patient, but from healthy tissue close to tumoral tissue. Expression of up to 20000 genes is determined by the mRNA level of each gene and measured by fluorescence differences. These differences must be translate from fluorescence intensity to real values, using some housekeeping genes (genes that are supposed to have a constant expression) to normalize the expression.

Higher mRNA expression are connected with a higher hybridization and, consequently, to a higher fluorescence that show higher expression values. In order to estimate the FC we have applied the following function:

$$FC = log_2(average(C)) - log_2(average(N)) \tag{1}$$

Where C is the expression gene set in cancer samples and N refers to the expression gene set in control samples.

Depending on FC value, we have two different gene behaviors in cancer:

- If FC>0, a gene is up-regulated, showing a higher expression in cancer cells
- If FC<0, a gene is down-regulated, showing a lower expression in cancer cells.

FC values can be analyzed using statistic inference or descriptive statistics. The first simply selects the genes with highest FC and with statistical significance. Descriptive statistics groups genes with similar expression patterns, allowing selecting a parameter to group genes with same behavior. Both due to the experiment and cancer biology, the previous analysis can be inappropriate to extract rules for this disease. In that way, small changes in gene expression of a specific gene can produce a great impact in cancer evolution. This makes necessary to use different parameters to define the role of genes in cancer. As consequence, we have designed a parametrized genetic algorithm to discover QAR able to deal with the expression levels of the genes.

3 Description of CANGAR Algorithm

This section describes the CANGAR algorithm used to mine QAR from microarray data of colon cancer patients. CANGAR is a multi-objective algorithm based on the well-known NSGA-II [13] that discovers QAR in continuous datasets. CANGAR provides rules with adaptive intervals, whereas the classical algorithms, such Apriori, requires a previous discretization of data. The search of the most suitable intervals is addressed by an evolutionary process in which the intervals are evolved to discover high quality QAR according to the objectives optimized.

The main features of CANGAR algorithm are presented in the following sections. Section 3.1 shows the pseudocode of CANGAR algorithm based on the NSGA-II algorithm. Section 3.2 describes the codification of the individuals of the population to represent a QAR. Section 3.3 details the genetic operators used to deal with the aforementioned individual representation.

3.1 Algorithm Pseudocode

In this section the pseudocode of CANGAR algorithm to obtain QAR is presented in Algorithm 1. Although the evolutionary scheme is based on the NSGA-II algorithm, the inherit scheme has been modified and new features have been added. For instance, an external population (P_E) is considered to include all the non-dominated solutions found. Furthermore, the population is restarted when the percentage of evolved individuals is lesser than a minimum threshold (Et).

The pseudocode is divided in four procedures. The main procedure is *Algorithm* (lines $7 - 17$). This procedure performs a number of executions (num_{ex}), in which a set of Re rules are obtained to compose the final set of best rules (Bc) found by the algorithm. *Evolve* procedure (lines $18 - 31$) represents the evolving process of the population. This procedure obtains the external population (P_E) when the generations limit (Gl) is reached. In each generation, the new population obtained is sorted and ranked according to the level of non-domination. Then, the external population (P_E) is updated with the non-dominated solutions that compose the first Pareto Front (P_{r0}).

Procedure *Next-generation* (lines $32 - 47$) is devoted to build the next generation of the population. To fulfil this goal, the offspring population (D) is obtained by genetics operators. After that, these individuals and the current population (P_g) are merged into a new population (T). This new population (T) is sorted and ranked to obtain different fronts according to the level of non-domination of the individuals (F). The next generation (P_{g+1}) is composed of the N best individuals selected from F. Finally, the procedure *Update-external-population* (lines $48 - 51$) updated the external population (P_E) taking into account the non-dominated solutions after performing the union between R and P_E.

3.2 Individual Representation

As the implemented algorithm seeks to do data mining with QAR, the chromosomes must represent those rules. To achieve that, the chromosomes include the following properties:

- **Antecedent**: Set of genes representing the rule antecedents.
- **Consequent**: Depending on the desired rules to be reported, two type of consequents have been considered.

  Type 1 Set of genes representing the rule consequent.

  Type 2 Fixed class. This can be used to fix, for example, if the patient has cancer or not. When the consequent is a gene or a set of them, this property is not used.

  In this work, consequent type 1 can be used to find relationships among genes and consequent type 2 can be used to find which genes are related with a type of patient.

In the context of this work, the genes, referring to the indivisible part of the chromosome, correspond to real genes. Due to the algorithm seeks to find genes implied in cancer, the chromosome genes would be real ones. Each gene has the following properties:

Algorithm 1. CANGAR pseudocode

```
Input 1: Number of executions (num_ex)
Input 2: Population size (N)
Input 3: Generations Limit (Gl)
Input 4: Rules per execution (Re)
Input 5: Evolve threshold (Et)
Output: Best rules found

procedure ALGORITHM(num_ex, N, Gl, Re, Et)
    $Bc \leftarrow \emptyset$
    while Executions $\leq$ num_ex do
        $P_0 \leftarrow$ initial-population(N)
        $Fp \leftarrow \emptyset$
        $Fp \leftarrow evolve(P_0, Gl, Et)$
        $I \leftarrow$sort-by-distance($individuals(Fp)$)
        $Bc \leftarrow Bc \cup I[1 : Re]$
    end while
    return $Bc$
end procedure

procedure EVOLVE($P_g$, Gl, Et)
    $P_E \leftarrow \emptyset$
    while not $Gl$ do
        $P_g \leftarrow$ next-generation($P_g$, Et)
        $Pr = (Pr_0, Pr_1, ...) \leftarrow$non-dominated-sort($P_g$)
        $P_E \leftarrow$ update-external-population($Pr_0$)
        if percentage-evolved($P_g$) $< Et$ then
            $P_g \leftarrow$ initial-population(N)
        end if
    end while
    $P_E \leftarrow$ filter($P_E$)
    $P_E \leftarrow$ rankings($P_E$)
    return $P_E$
end procedure

procedure NEXT-GENERATION($P_g$, Et)
    $D \leftarrow descendants(P_g)$
    $T \leftarrow D \cup P_g$
    $F = (F_0, F_1, ...) \leftarrow$non-dominated-sort($T$)
    $i \leftarrow 0$
    $P_{g+1} \leftarrow \emptyset$
    while $|P_{g+1}| + |F_i| < N$ do
        $P_{g+1} \leftarrow P_{g+1} \cup F_i$
        $i \leftarrow i + 1$
    end while
    if $|P_{g+1}| < N$ then
        $F_i \leftarrow$ sort-by-distance($F_i$)
        $P_{g+1} \leftarrow P_{g+1} \cup F_i[1 : |P_{g+1} - N|]$
    end if
    return $P_{g+1}$
end procedure
```

```
procedure UPDATE-EXTERNAL-POPULATION(R)
    $P_E \leftarrow P_E \cup R$
    filter-dominated($P_E$)
end procedure
```

- **Attribute**: Index referring to the column of this gene in the original dataset. This index is used for not to have to deal with strings, and get the gene name only once, at the end.
- **Lower Limit**: Real number showing the lower limit of the expression interval of the gene.
- **Upper Limit**: Real number showing the upper limit of the expression interval of the gene.

The lower and upper limits turns the AR into QAR, because they do not only link a gene with a state (normal or cancer) or other genes but also they link genes and their expression interval, with a state or with other genes and their expression interval.

3.3 Genetic Operators

Genetic operators are devoted to evolve the population for, starting with parents, generate better children. There are selection mechanisms to filter those individuals who are not good enough and apply the operators only to the best ones. In this work, two genetic operators, the mutation and crossover operator, have been defined to deal with the aforementioned individual representation.

Mutation Operator. Three mutation policies have been implemented.

- Generalisation of rules. The first one is for generalise rules. The policy remove a gene from the antecedent or consequent as long as the condition for minimum number of antecedents or consequents is satisfied.
- Specialisation of rules. The second one is a policy to specialise rules. It works the same way as the generalise policy, but adding a gene to the antecedent or consequent.
- Interval bounds. The last one do not add nor remove any gene from the rule, but takes one (in either the antecedent or the consequent) and mutates it's interval. This is made by taking a random point inside the current interval and generating a new one around this point.

Crossover Operator. The chosen crossover operator is based on taking a gene of each parent and give it to one child alternatively. To do that, the limit is taken as the number of genes of the biggest parent. Then, the genes are iterated one by one. In the first iteration, the gene of the first parent is given to the first child and the gene of the second parent is given to the second child. In the next one, the gene of the first parent is given to the second child and the gene of the second parent is given to the first child. The same process is done in every iteration.

If a gene can not be given to the corresponding child in one of those iterations (because that child already has that gene, for example) this iteration would be marked as invalid. Thus, the assignment regulation will not be reversed in the next iteration. Note that the interval bounds of the individuals are taken from the inherited gene of the parents in the crossover operator.

4 Experimentation

This section presents and analyzes the results obtained by CANGAR to discover relationships among set of genes related with patient that suffer colon cancer according to their expression levels. The cancer colon dataset used in the experimentation is presented in Sect. 4.1. Then, the parametrization of the algorithm is also described in Sect. 4.2. The results obtained in the experimental study are provided in Sect. 4.3.

4.1 Cancer Dataset

This work has used a public microarray gene expression dataset of patients with colon cancer referred as GSE21510 in the National Center for Biotechnology Information repository (NCBI) [14], using Robust Multi-Array Average (RMA) algorithm to standardize microarray data. The dataset is composed of 148 patients and 54675 genes expression profiles per patient. In order to check if the behavior of the database is the expected, we analyzed several known genes connected with colon cancer such as KLF4 [15] and BMI1 [16]. Once validated, we used this dataset applying the CANGAR algorithm to obtain QARs with genes highly related with cancer patients.

4.2 Parameter Settings

The CANGAR algorithm has several parameters to satisfy the user needs that are described as follows. The rule size can be controlled by the maximum number of attributes in the antecedents and the consequents. In this work, the minimum and maximum number of attributes in antecedents and consequents have been 1 and 4, respectively. The consequent can be a fixed class (typically *normal* or *cancer*) or a set of genes (type 1 and 2, respectively). Consequent type 1 have been used in this work. The rules obtained by CANGAR can be also filtered by setting minimum thresholds for several quality measures. For instance, a minimum support threshold of 0.1 and a minimum confidence threshold of 0.8 have been considered.

Other parameters are the maximum population size (100), the number of generations (100), the mutation rate (0.15) according to the three mutation policies (0.5 for the generalizing mutation, 0.1 for the specializing mutation and 0.4 for the interval bounds mutation) and the crossover rate (0.6). There is an option to avoid the algorithm stagnating using a threshold to set the minimum share of the population that have to evolve. In this case, a minimum percentage of individuals to evolve of 0.1 has been used. The objectives to be optimized by CANGAR regarding other previous studies [17] are Leverage, Netconf and Certainty Factor. Finally, the algorithm has be executed 5 times and the number of rules selected per execution has been 10.

4.3 Results

Several experiments have been performed using the aforementioned dataset and parameter settings to obtain a potential set of genes related with colon cancer. The CANGAR algorithm has been applied to obtain a set of QAR being fixed patient type (healthy or cancer) as consequent part. We are highly interested in those rules in which the patients type refers to patients suffering from colon cancer disease. Then, we have selected the top 100 of the most frequent genes that appear more than 20 % in the rules obtained by CANGAR. The selected genes have been validated by hierarchical cluster analysis, statistical and biological significance validation techniques and literature mining.

Hierarchical Cluster Analysis. A hierarchical cluster analysis has been conducted to cluster patients and genes with similar behaviour. This method is applied to assess the capability of the genes obtained by CANGAR to classify between healthy and cancer patients according to the changes in the expression levels (down-regulated or up-regulated). Alternatively, this hierarchical cluster analysis can be used to validate the changes of the expression levels detected by the intervals of the rules obtained by CANGAR.

Figure 1 displays the heatmap of control and cancer patients according to the top genes under study after applying the hierarchical cluster analysis. Note that data have been scaled and centered. The results are plotted as dendrograms. Spearman correlation and Pearson correlation have been used to cluster the columns (patients) and rows (genes), respectively. The resulting tree has been cut at specific height (1.5) with its corresponding clusters highlighted in the heatmap color bar.

In general, four groups can be observed according to the patient type and gene expression levels. Control and cancer patients are separated in different groups as can be observed in the X-axis of the heatmap (from left to right, respectively). The down-regulated genes and up-regulated genes for cancer patients are concentrated at the top-right and bottom-right parts of the heatmap, respectively. The more differences exist between the groups of genes and patients, the higher is its importance in cancer.

Biological Relevance. Using CANGAR algorithm we have found more than 90 altered genes in colon cancer samples. These genes have been classified in different groups regarding its main function in cancer [18] as can be observed in Table 2. The aforementioned functions are described as follows:

- **Cellular Proliferation/Death:** This is the main function involved in tumor transformation, thus these genes have been deeply studied in cancer. Here we

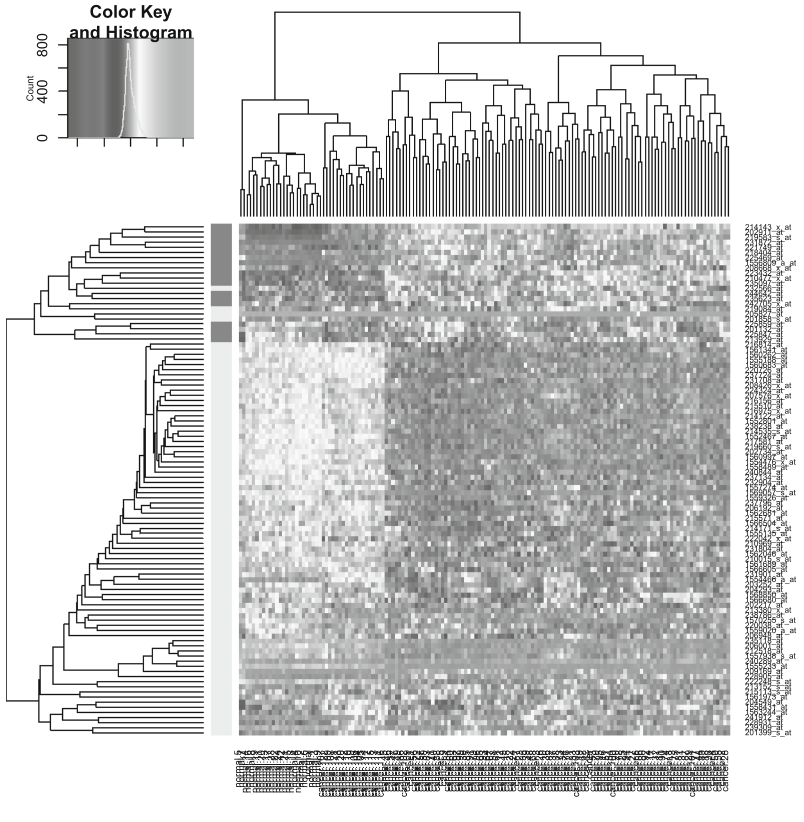

Fig. 1. Hierarchical cluster analysis of top genes.

find genes that take action in the cell cycle, like APC (Anaphase Promoter Complex) which is essential for mitosis.

- **Metastasis:** Invasion is one of the main characteristics of cancer cells, becoming independent from its environment and entering in blood circulation to reach other tissues. Here we can find genes related with cytoskeleton or matrix stabilization, like ACTN1 or ITIH2.
- **Angiogenesis:** cancer cells require high levels of glucose and oxygen, making necessary the develop of new blood vessels. Two genes (MED28 and TNFSF15) involved in angiogenesis have been identified using CANGAR algorithm.
- **Gene Expression Regulators:** cancer progression is characterized by important changes in gene expression, due to alterations in ribosome synthesis (RPL13A) or histone modification (PRDM9, HIST1H3H). In that way, we found downregulation of histone HIST1H3H, involved in heterochromatization, a process that turn-off gene transcription.
- **Other Component:** Many genes of the list are unknown or poorly characterized. In this way, the lack of information about the role in cancer of many of the found genes converts them in potential targets for new research.

Table 2. Functional classification of the genes found by CANGAR

Metastasis	Cellular Proliferation/ Death	Angiogenesis	Regulation of genetic expression	Other components
CCDC87	WWOX	MED28	MBNL1	CLIP 1
ACTN1	MAPK6	TNFSF15	PRDM9	ITSN2
ITIH2	MPP3		HNRNPLL	ARNT
PRP14	GFRA4		IKZF1	CLN8
MMP14	GPR137		RPL13A	CHST1
PPP1R9A	WNT5A		MIEF1	SLC30A9
GPM6	RXRA		RBP1	SLC01B1
MYCN	APC		TCF3	ANKRD43
CEP350	STK38		HIST1H3H	ANKRD46
	CDC34		EIF3CL	RNF10
	R653		DNAZC13	PAM
	TNK1		HELQ	ACTR1A
	TRRAIP		SPOCD1	
	USP40		MIEF1	
	JAG1		PWRN2	
			HIST2H2AA3	
			RPL13A	

*Up-regulated genes with FC higher than 1.5 have been highlighted in light grey
*Down-regulated genes with FC lower than 1.5 has been highlighted in dark grey

5 Conclusions

In this work, we have presented the discovery of quantitative association rules in gene expression microarrays of cancer. In particular, the multi-objective evolutionary algorithm named CANGAR has been applied to discover hidden relations among genes according to their expression levels. Then, we identified the most frequent genes appearing in the rules in order to provide a subset of genes with potencial prognosis role in cancer. To validate the results, a hierarchical cluster analysis has been applied using the reported set of genes to validate the capability of characterization between control and cancer patients of them. Furthermore, the genes have been classified according their main function. As a future work, we intend to improve the validation of the results, to test the implication of the reported set of genes in different types of cancer and apply the CANGAR algorithm in other microarray datasets.

Acknowledgments. The financial support from the Spanish Ministry of Science and Technology, projects TIN2011-28956-C02-02 and TIN2014-55894-C2-1-R, and from the Junta de Andalucia, P11-TIC-7528 and P12-TIC-1728, is acknowledged.

References

1. Ellis, L., Woods, L.M., Estve, J., Eloranta, S., Coleman, M.P., Rachet, B.: Cancer incidence, survival and mortality: explaining the concepts. Int. J. Cancer **135**(8), 1774–1782 (2014)

2. López-Abente, G., Aragonés, N., Pérez-Gómez, B., Pollán, M., García-Pérez, J., Ramis, R., Fernández-Navarro, P.: Time trends in municipal distribution patterns of cancer mortality in spain. BMC Cancer **14**(1), 1–15 (2014)
3. Kharya, S.: Using data mining techniques for diagnosis and prognosis of cancer disease. CoRR abs/1205.1923 (2012)
4. Sarvestani, A., Safavi, A., Parandeh, N., Salehi, M.: Predicting breast cancer survivability using data mining techniques. In: 2nd International Conference on Software Technology and Engineering (ICSTE) 2010, vol. 2, pp. 227–231 (2010)
5. Lopez, F., Cuadros, M., Cano, C., Concha, A., Blanco, A.: Biomedical application of fuzzy association rules for identifying breast cancer biomarkers. Med. Biol. Eng. Comput. **50**(9), 981–990 (2012)
6. Tang, J.Y., Chuang, L.Y., Hsi, E., Lin, Y.D., Yang, C.H., Chang, H.W.: Identifying the association rules between clinicopathologic factors and higher survival performance in operation-centric oral cancer patients using the apriori algorithm. Biomed. Res. Int. **2013**, 7 (2013)
7. Slonim, D.K., Yanai, I.: Getting started in gene expression microarray analysis. PLoS Comput. Biol. **5**(10), e1000543 (2009)
8. Martínez-Ballesteros, M., Troncoso, A., Martínez-Álvarez, F., Riquelme, J.C.: Improving a multi-objective evolutionary algorithm to discover quantitative association rules. Knowl. Inf. Syst. 1–29 (2015)
9. Geng, L., Hamilton, H.: Interestingness measures for data mining: a survey. ACM Comput. Surv. **38**(3), 1–42 (2006)
10. Martínez-Ballesteros, M., Martínez-Álvarez, F., Troncoso, A., Riquelme, J.C.: Quantitative association rules applied to climatological time series forecasting. In: Corchado, Emilio, Yin, Hujun (eds.) IDEAL 2009. LNCS, vol. 5788, pp. 284–291. Springer, Heidelberg (2009)
11. Martínez-Ballesteros, M., Troncoso, A., Martínez-Álvarez, F., Riquelme, J.: Obtaining optimal quality measures for quantitative association rules. Neurocomputing **176**, 36–47 (2016)
12. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. SIGMOD Rec. **22**(2), 207–216 (1993)
13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
14. Tsukamoto, S., Ishikawa, T., Iida, S., Ishiguro, M., Mogushi, K., Mizushima, H., Uetake, H., Tanaka, H., Sugihara, K.: Clinical significance of osteoprotegerin expression in human colorectal cancer. Clin. Cancer Res. **17**(8), 2444–2450 (2011)
15. Hu, R., Zuo, Y., Zuo, L., Liu, C., Zhang, S., Wu, Q., Zhou, Q., Gui, S., Wei, W., Wang, Y.: Klf4 expression correlates with the degree of differentiation in colorectal cancer. Gut Liver **5**(2), 154 (2011)
16. Kreso, A., van Galen, P., Pedley, N.M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N., Moon, Y.C., Gibson, L., Wang, Y., Leung, C., Iscove, N.N., Arrowsmith, C.H., Szentgyorgyi, E., Gallinger, S., Dick, J.E., O'Brien, C.A.: Self-renewal as a therapeutic target in human colorectal cancer. Nat. Med. **20**(1), 29–36 (2014)
17. Martínez-Ballesteros, M., Martínez-Álvarez, F., Lora, A.T., Riquelme, J.C.: Selecting the best measures to discover quantitative association rules. Neurocomputing **126**, 3–14 (2014)
18. Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: the next generation. Cell **144**(5), 646–674 (2015)