Multi-agent System for Obtaining Relevant Genes in Expression Analysis between Young and Older Women with Triple Negative Breast Cancer

Alfonso González-Briones[1,*], Juan Ramos[1], Juan F. De Paz[1] and Juan M. Corchado[1,2]

[1]Biomedical Research Institute of Salamanca/BISITE Research Group,
Calle Espejo, s/n, 37007 Salamanca, Spain,
http://bisite.usal.es

[2]Osaka Institute of Technology, Osaka 535-8585, Japan,
http://www.oit.ac.jp/

Summary

Triple negative breast cancer is an aggressive form of breast cancer. Despite treatment with chemotherapy, relapses are frequent and response to these treatments is not the same in younger women as in older women. Therefore, the identification of genes that cause this difference is required. The identification of therapeutic targets is one of the sought after goals to develop new drugs. Within the range of different hybridization techniques, the developed system uses expression array analysis to measure the expression of the signal levels of thousands of genes in a given sample. Probesets of Gene 1.0 ST GeneChip arrays provide categorical genome transcript coverage, providing a measurement of the expression level of the sample. This paper proposes a multi-agent system to manage information of expression arrays, with the goal of providing an intuitive system that is also extensible to analyze and interpret the results. The roles of agent integrate different types of techniques, statistical and data mining methods that select a set of genes, searching techniques that find pathways in which such genes participate, and an information extraction procedure that applies a CBR system to check if these genes are involved in the disease.

1 Introduction

Triple negative breast cancer (TNBC) is a form of cancer in which none of the three most common types of receptors (estrogen, progesterone, and the HER-2 / neu gene) usually present in other breast cancers. This means that breast cancer cells tested negative for hormone epidermal growth factor receptor 2 (HER-2), estrogen receptor (ER) and progesterone receptor (PR) [1]. For this reason, common treatments such as hormone therapy and drugs that target these three receptors are ineffective. To improve the treatment of this cancer we must attempt to understand the possible role played by genes that affect younger patients and older patients differently. To this end, drugs that enhance the therapeutic efficacy of current methods are currently being sought.

*To whom correspondence should be addressed. Email: alfonsogb@usal.es

There are several techniques that can be used to study genetic variation in patients, such as tissue microarrays, expression arrays (RNA) [2, 3], genomic arrays (DNA) and arrays of microRNAs (miRNAs). Arrays used in the study of expression profiling are cDNA arrays and oligonucleotide chips. Moreover, different types of genomic arrays (DNA) are used, including BAC aCGH, oligo CGH, SNP CGH and aCGH (Comparative Genomic Hybridization) [4]. CGH arrays (aCGH) can compare the DNA of a patient with control DNA and use this information to detect mutations [5, 6] based on the increase, loss or amplifications [7] in different regions of the chromosome. There are new exon arrays that provide accurate assessments of gene expression [8]. Information sources are varied but laboratory personnel usually follow fixed analysis processes that are distributed in sequences, which are in turn executed repeatedly in the search for genes that are considered to be relevant. Therefore, it is necessary to find a system that automates this process so that the work of the laboratory staff is simplified.

There have also been efforts to provide a solution to the main challenges associated with analyzing microarray data, which are: the high amount of data (coming from thousands of genes extracted from few samples); the high complexity of the data; the fact that gene datasets in microarrays are often correlated (either directly or indirectly); and the fact that most gene selection and prediction models emphasize the capacity for effective classification instead of the function of an effective selection. The assumption is that statistical significance is equivalent to biological importance.

There are other investigations which focus their efforts in predicting genes that cause diseases. Thanh-Phuong Nguyne and Tu-Bao Ho have developed a semi-supervised framework in order to find genes and detect possible connections among those that can lead to diseases [9]. They are based on feature extraction, and preprocessing of data and integrate the following resources: Universal ProteinResource (UniProt) [10] Gene Ontology (GO) [11], Pfam [12], InterDom [13], Reactome [14] and expression datasets [15]. Maglietta et al. [16] propose a method from a similar point of view. The target is the selection of genes relevant to a pathology by analysing the tissue expression profiles for two different phenotypic conditions. Statistical techniques are used and the presence of genes in similar studies is verified. Other studies use multiagent systems in order to analyze array data, including a system proposed by Juan F. De Paz et al. [17], where a multiagent system analyses CGH arrays searching for gene gains or losses, which are then represented. This study is more oriented to obtain relevance areas and provide easy access to information but works only with CGH arrays. The present paper proposes a multi-agent system to analyze expression arrays. The main novelty is that the system can learn analysis flows (workflow) while the expression analysis is being performed, thus automating the analysis of expression. During the analysis, services are incorporated in order to carry out the analysis and extraction of information from databases, through which the most relevant genes are selected. Different data mining techniques and databases were used to analyse expression profiles and obtain relevant genes for two different phenotypic conditions. The system was applied to a real case study for the analysis of breast cancer with the aim of analysing differences in this type of cancer with specific regard to the patients age.

This article is organized as follows: section 2 describes the state of the art of expression arrays, Section 3 describes the proposal, and Section 4 presents the results and conclusions.

2 Gene Expression Arrays

Microarrays constitute a widely used tool that measure gene expression [18]. Moreover, this technology has attracted special interest in cancer research [19]. An expression arrays analysis makes it possible to study and compare transcriptomes of different samples. The value of gene expressions in these biochips is determined by the intensity of the hybridization of transcripts with a group of probes [8].

With these qualities, expression arrays become a very useful tool that makes it possible to determine which genes have an altered expression, to compare expressions based on certain parameters, and to diagnose and distinguish subtypes of cancers with similar clinical manifestations, among other things. Different kinds of cancer genes share groups and altered pathways. Array analysis can investigate typical genes, as well as those that are not common to the vast majority of proliferative syndromes [18], existing in more specific forms of the disease. This is one factor that makes arrays a useful diagnostic tool.

Beyond studying the expression of each gene and its degree of responsibility in an alteration, it is vital to understand the expression of these genes and the proteins they encode in the context of signaling pathways [20].

To be able to perform a complete analysis, one of the roles of a multi-agent system is to search in different databases for the pathways taken by the specific genes that are being studied. One mutation in a particular gene can give rise to various effects, even in the same type of tissue [21]. Because of this, the function of the platform is interesting, specifically because the information obtained from the study of a single gene is not representative if, after it has been studied, its relationship to other elements that also influence the signaling pathway is not verified [20].

The main function of this platform is to be able to select the relevant genes for the investigation. There comes a point during the screening process when there is no longer a sufficient number of elements to obtain pathways. It is precisely for this reason, from a research point of view, that it is important to compare the genes obtained from the analysis that best explain the gene alteration (depending on the studied parameter) according to the altered pathways.

Among the most influential resulting gene expression analyses of patient samples in our case study, those that are also therapeutic targets are of particular interest to medical research. One of the great difficulties of the analysis of arrays is to obtain biologically valid conclusions from vast amounts of data [22]. Consequently, one agent from the multi-agent system is responsible for conducting searches in databases known to contain therapeutic targets: TTB (Therapeutic Target Database) [23] and DrugBank [24]. This opens up the possibility of attacking these targets with drugs and conducting pharmacogenomic research after the analysis [18]. This makes it possible to check and directly study the influence of the pathway in the carcinogenic process, in addition to its clinical implications and the search for effective treatments.

3 Multi-agent System

In these studies, users have to work with a large volume of information, which involves the development of programs to improve data analysis systems and to automatically extract information through databases [25].

Our study uses expression arrays that determine the expression of genes to the probes used. This information is taken into account to observe differences that may occur in the same genes with regard to the age factor. Because large amounts of data are handled, it is necessary to develop a system aimed at simplifying the management and analysis of this information, and at automatically extracting information to determine the correlation of these genes in breast cancer.

Distributed analysis of expression data is performed by various laboratory personnel: from chip hybridization to the removal of variations and irelevant information associated with the chips. This study shows a multiagent system specifically designed, with an abstract architecture for this virtual organizations [26], to analyze expression arrays. The functionality of the multi-agent system is divided into layers and roles to perform the analysis, which usually consists of several stages. The first stage is preprocessing, which performs the important task of removing probes without notation and screening the data for the first time. The next stage performs an analysis of the expression probes (Non-parametric statistical test like Mann-Whitney or Kruskal-Wallis), searching for differences with respect to the expression under normal conditions for that gene, or with respect to any specific factor. This test are applied to obtain the genes for testing the proposed statistical hypothesis. In the next stage data mining techniques are applied, allowing the data set studied to be further reduced. When looking for differences between groups of patients, it is important to confirm whether a cluster has been properly formed at the end of this stage, according to the case study. If a suitable result is not obtained, it will be necessary to review the previous step of extracting relevant genes through the data mining techniques. The final stage is initiated once the data set containing the different genes has been identified. This data set will be transferred to a database that checks the implication of these genes in the specific disease being studied to determine whether there is a relationship. The gen notation and pathways in which most of these genes develop their biological activity is obtained.

JADE (Java Agent DEvelopment Framework) was adopted for the design and implementation of the proposed intelligent multi-agent. The architecture is composed of four layers: Analysis, Information Management, Visualization and Workflow. The architecture is shown in Fig. 1.

The workflow layer includes an agent in charge of workflow in the other layers, and of establishing the correct order for the activity of each agent. Workflow analysis collects information about the settings and can repeat the sequences performed above for expression analysis. This aspect makes it possible to automate repetitive analysis tasks for laboratory personnel.

The analysis layer performs microarray analysis tasks as required by the process. This layer consists of several agents that are responsible for implementing the necessary processes and algorithms. An agent applies the Entropy-based filters, which are responsible for finding discrete attribute weights based on their correlation with continuous class attributes. Another agent ap-

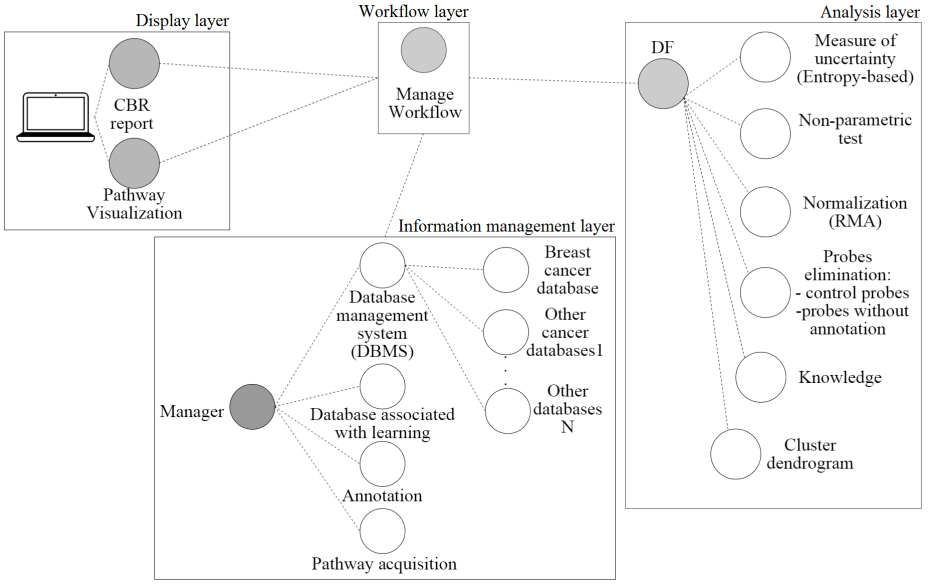

Figure 1: Multi-agent System Arquitecture

plies the nonparametric Mann-Whitney statistical test to two independent samples. One more agent applies a normalization to adjust the signal, which may contain errors caused by technical and / or biological factors. The normalization technique applied is RMA (Robust multiarray Average), which adjusts information on each probeset to a comparable value. If any is over-expressed or under-expressed its genetic value is regulated with regard to the total. Another role performed by a system agent is the elimination of control probes and probes without notation. Finally, another agent produces a dendrogram that allows us to organize data into subcategories (younger and elderly patients) in which the case study is divided. This representation enables a clear representation of the relationship between data grouping.

The information management layer is responsible for confirming whether the genes obtained from the results of the analysis layer are related to the type of cancer that are the focus in the case study. This layer creates a database that collects the learned genes that are related to a cancer that is not contained in the databases used for the visit; this database is also queried in the process of relating genes to the cancer in the case study. This layer also includes an Agent that collects the annotation and other information for each resulting gene and the agent that retrieves the pathways associated with these genes.

The display layer shows the pathway associated with the result as well as the list of the implicated genes. This layer also manages the pathway visualization, showing the pathway associated with the gene list obtained. The gene list is sent using the KEGG Mapper (`http://www.genome.jp/kegg/tool/map_pathway1.html`). A pathway can be disrupted by a mutation on a single gene. The multi-agent system selects the pathways which present several changes in the pathology due to age factor. The selected pathways are shown.The display layer manages the case based reasoning (CBR) system doing the four-step process. Once a relevant case is retrieved, that workflow is reused in the new problem with a previous adaptation as needed to fit the new problem. Later, the revise step is performed to avoid undesired steps in the workflow for the next process: retaining the solution in a database.

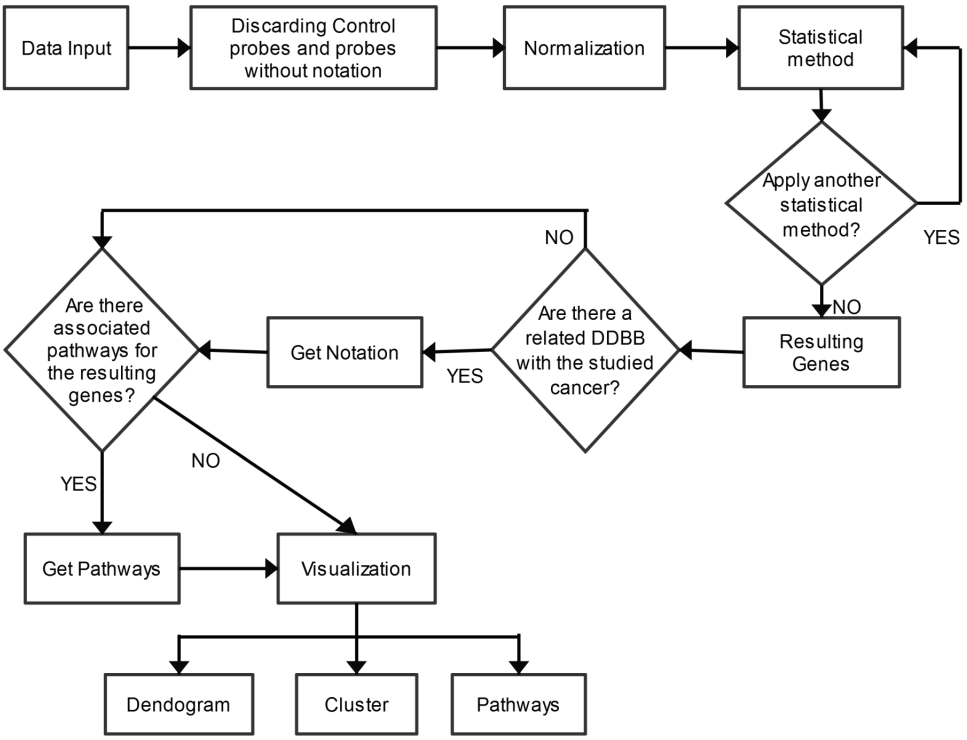

Figure 2: Tipical Workflow

Once an analysis of the input data has been performed, the CBR retrieves a relevant case of previous studies from the memory workflow. The CBR is based on the type of input variable (nominal or categorical qualitative);an analysis is performed for the other variable, the level of expression. Once one of the workflows to reuse is selected, the solution is reviewed and changes are made in the value of the test variable to the hypothesis if considered appropriate,ie the p-value of the used statistical method. If the applied workflow is considered a satisfactory solution, the corresponding activities are carried out with the respective values of the variables to be applied in posterior analysis. This process can be seen in Fig. 2. The implemented workflow can follow the same steps as the typical process of Fig. 3.

4 Results

The case study was performed with 16 samples from patients with triple negative breast cancer provided by the Salamanca Cancer Institute. 8 samples corresponded to younger patients (less than 45 years) and the remaining 8 samples to elderly patients (over 68 years). Additional samples continue to be gathered in order to improve the accuracy of the results.

The technology used to analyze the microarray was Affymetrix, which is based on oligonucleotide chips. The specific chip used was the HuGene-1.0st-v1 chip, which contains 33,297 probes that identify about 23,000 sequenced genes.

The multi-agent system designed in this paper is applied to study gene expression arrays from

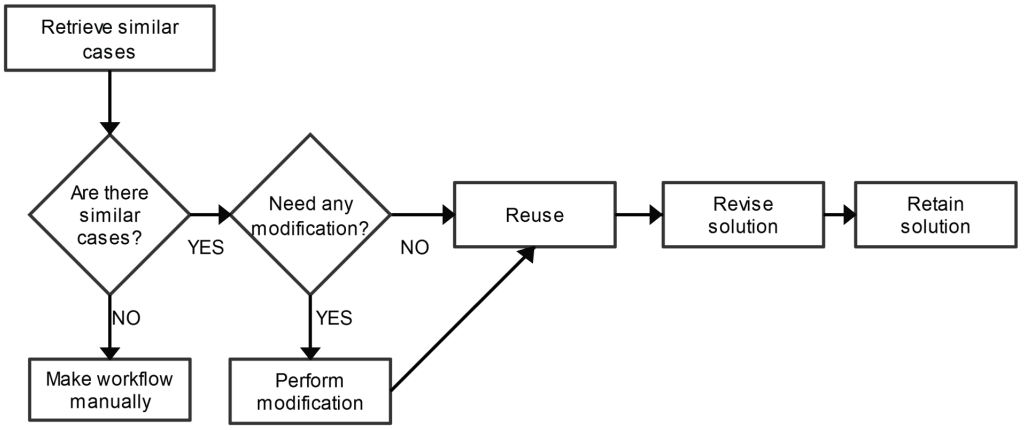

Figure 3: CBR Workflow

the samples of these patients. The goal is to obtain genes that show differences between samples from younger patients and older patients in order to discover why older women respond better to the treatment.

The first step in the analysis of oligonucleotide chips is the process of discarding any control probes and probes without notation. Once these probes are discarded a denormalization process [27] is performed. This process discards the values that deviate from the normal value, and applies the statistic test (Mann-Whitney) to each of the independent samples used in the case study (younger women and older women). In that step a Benjamini and Hochberg FDR multiple test correction is performed and applied on the p-value, calculated based on the two-samples Mann-Whitney samples.

In our case of study, we look for variations that may occur in the expression levels of genes for samples associated with younger women compared with those of the older women. Once applied, we discard all values exceeding a p-value greater than 0.01, i.e., we keep the probes that have a greater interest because of their expression level compared with older youth at statistical level.

In this process of data analysis, once the preprocessing, normalization and application of statistical tests and techniques of data mining are completed, a clustering algorithm is applied. This algorithm provides us with a dendrogram which allows us to check the degree of clustering of the probes with respect to the two samples (probes for younger and older women). Process results are shown in Fig. 4. The next stage is responsible for managing information received from the data obtained at the end of the analysis stage. These genes are contrasted with the corresponding type of cancer data base with which the data are associated (in this case study, the data base is for breast cancer). With this process we discard genes that are not implicated in a certain cancer, focusing on those that are. The pertinent notation and associated pathways are obtained for this final gene set, allowing the access and use of this information. The workflow layer agent learns the execution sequence of tasks, the order in which the agents interact with each other to execute the various processes and algorithms. With this initial learning, the following analysis of expression arrays is performed automatically, so the lab technicians do not have to perform the process manually, which avoids the risk of human error, loss of time, or a

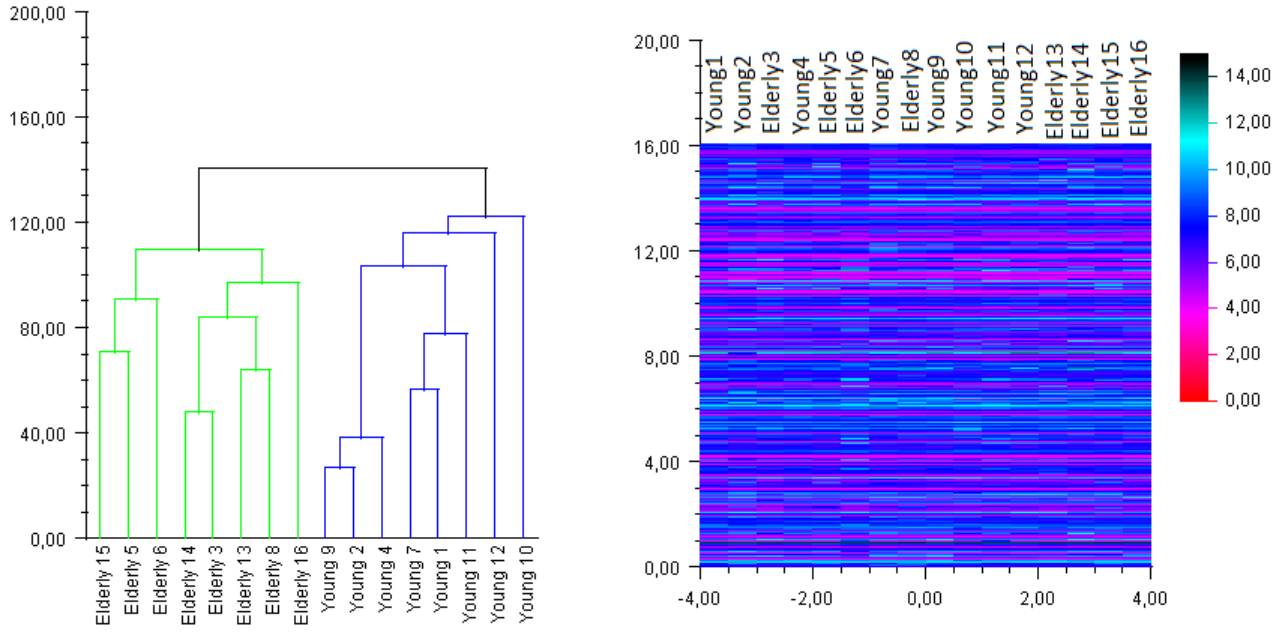

Figure 4: Resulting Dendogram and Heatmap

lower yield.

Once the genes and pathways are shown, an agent from the visualization layer performs a reasoning cycle (CBR). During information from the catalogued gene is obtained from the accurate database G2SBC (Genes-to-Systems Breast Cancer Database - `http://www.itb.cnr.it/breastcancer/`). While contrasting genes with the breast cancer database, these genes are evaluated according to the contrast hypothesis mentioned in part 3. In this way, if our system detects genes which are not in the databases and influence the pathology according to the results, those genes are stored in their own database for future analysis.

Table 1. shows the most important genes obtained in the case study for triple negative breast cancer. These genes are considered by the system as the most important with regards to differences in response to treatment of younger women and older women.

There was no pathway in which two or more selected genes were present. However, this is not surprising since the final number of genes kept is very small in order to provide a manageable quantity for the researcher. Although it is well known that expression varies with aging, there are few simultaneous gene relationships between age and breast cancer at the same time that have been previously described in literature.

We have obtained five relevant genes in TNBC presenting expression changes due to age factor. The Sfrp1 gene (Secreted frizzled-related protein 1) can be interesting in this case because its regulator role of the Wnt pathway which can be seen in Fig. 5, regulates growth and differentiation in different cell types. The protein encoded by this gene has a direct inhibitory effect on Wnt proteins. This ultimately leads to the inhibition of proliferation in vascular cells, and also delays the G1 cycle phase in these same cells [28].

According to our gene expression data, Sfrp1 (represented in Fig. 5 by FRP) decreases with age

Table 1: Resulting altered genes

Altered genes with close relationship to the age factor	Altered genes with no-apparent relation to the age factor
GABRP	CAPN6
SFRP1	SLC6A14
MID1	SCRG1
RARB	BCL11A
ACTG2	PTGS2
	BBOX1
	S100A7
	PRKAA2
	ACPP
	ALCAM
	RND3
	GGH
	PKP2

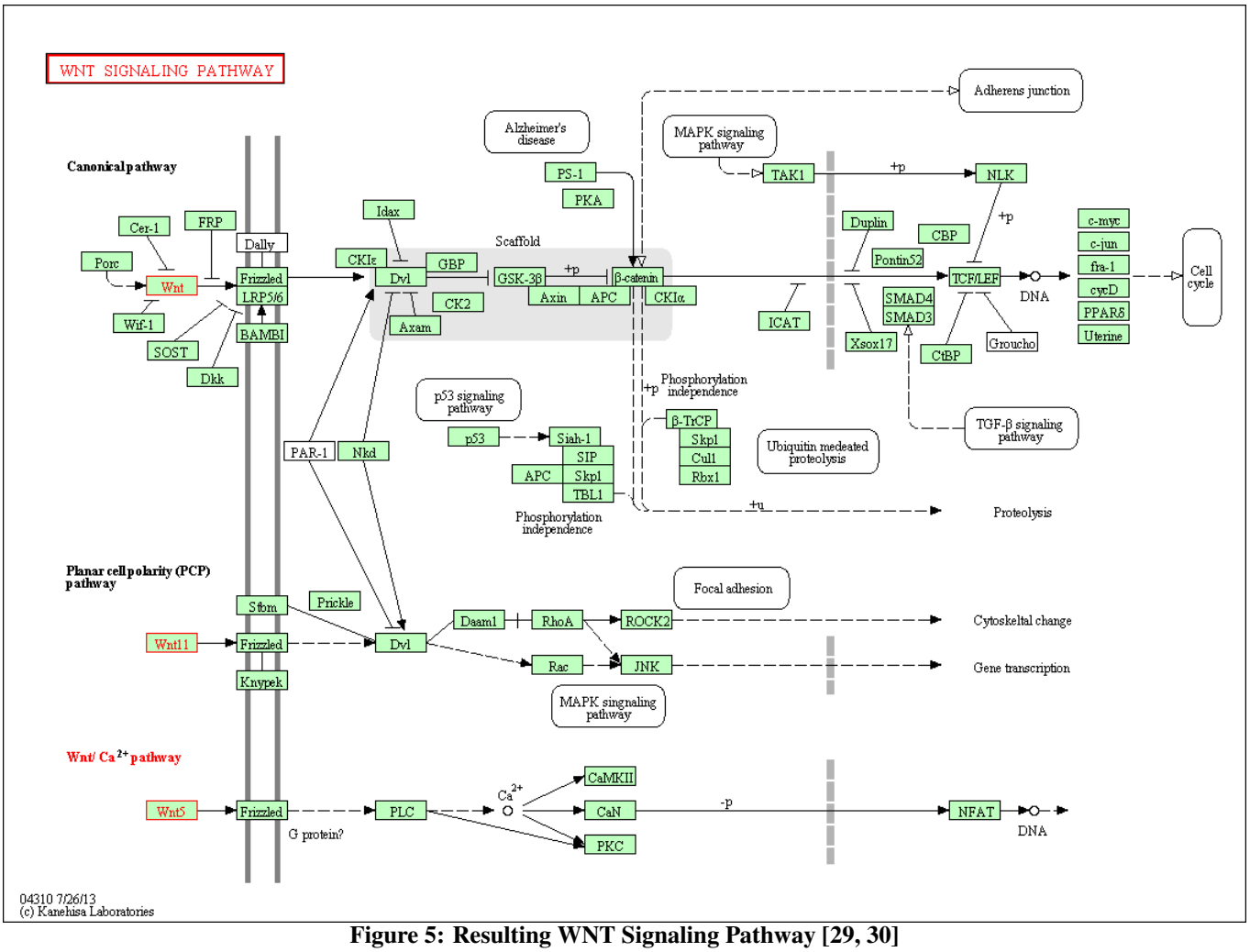

Figure 5: Resulting WNT Signaling Pathway [29, 30]

in our TNBC patient samples. Although SFRP1 loss is most notable with age, this mutation occurs early in tumor development, so damage caused by this impairment is not unique to advanced age patients In fact, it has been suggested that this could be a prognosis marker of patients with breast cancer in early stages [31]. SFRP1 expression has been associated with a greater chance of a positive response to neoadjuvant chemotherapy [32]. Furthermore, it has been found that this loss of mRNA is reflected in a remarkable effect on protein loss. By interfering with this gene expression, it has been found that Wnt pathway becomes active in hepatoma [33]. Therefore, due to its reduced expression levels in the disease, SFRP1 could not exert an inhibitory effect in Wnt in order to control proliferation. However, this is not the only interpretation to date. Bernemann et al. suggest in their study that the invasion of tumor cells is not really regulated by the Wnt pathway in TNBC, but through other pathways [32].

Retinoic acid (active form of vitamin A) binds to RARB, which is involved in cell signaling processes such as morphogenesis, differentiation and cell growth. Loss of retinoic acid receptor in breast cancer has been described previously [34]. This receptor mediates the growth inhibitory effect exerted by retinoic acid, due to its apoptosis-promoting action. For this reason, the underexpression of the RARB gene may be favoring the uncontrolled growth of tumor cells.

The involvement of GABA A receptor subunit pi (GABRP) in breast cancer has not been well defined. However it is noteworthy that it is detectable in many non-neural tissues. Furthermore, it has been observed that with patients in which brain metastasis occurs, there is a considerable increase in GABA A receptor levels so it is also a prognostic marker. Its reinforcing role has been demonstrated in tumor cell lines [35] and it is a potential therapeutic target. In the sample of patients used in this study GABRP expression decreases with age. This result seems consistent with the described lower survival when this gene is overexpressed.

MID1 is part of the tripartite motif. This gene expression was higher in the group of elderly patients in our case study. MID1 is a negative regulator of tumor suppressor PP2A, in the mTOR pathway. Consequently, its overexpression is affecting proliferation, contributing to uncontrolled division.

The multiagent identified GABA A and RARB as registered therapeutic targets, increasing their interest in what regards to pharmacogenomics. However, to better appreciate the usefulness of resulting genes that are themselves therapeutic targets, it is necessary tocheck if the gene in question is over or under-expressed. Although GABA A and RARB are registered as targets, blocking them is not appropriate from a therapeutic point of view, since the alteration consists precisely in an expression decrease. However, the use of drugs on these targets could have great scientific interest in order to study the involvement of each alteration in cancer and its relative importance.

Although MID1 has not been registered as a target yet, several authors have suggested that this is a promising [36] therapeutic target, and it would be interesting to find an appropiate drug. In fact, this gene is more suitable for the subsequent investigation, considering its role and, as mentioned, the tendency toward overexpression in the analyzed data.

5 Conclusions

The developed system enables using patient samples to know if there are differences in the expression level for the proposed gene sets, allowing the system to return the genes that produce differences in the samples with regard to the associated notation and pathways in which are implicated.

This case study looked at the differences in expressions that can occur in female patients younger than 50 years of age compared with those older than 50, since the latter group respond better to the treatments used.

This study is interesting because finding genes that behave differently can lead to new information and the possibility of adjusting treatments for this type of cancer in younger patients.

The multi-agent system is developed in a way that allows new agents to be inserted with new techniques or existing data to be modified for analysis and facilitates data analysis. The system provides access to various databases so different cancer datasets can be introduced. The system uses a CBR that handles all information obtained from the databases and allows the incorporation of new information that may be used in future analysis.

In conclusion, it is noteworthy that genes obtained in this case study from using our tool are suitable for pharmacogenomics research because of their influence on tumor processes in breast cancer.

Acknowledgements

This work has been carried out by the project Sociedades Humano-Agente en entornos Cloud Computing (Soha+C) . SA213U13. Project co-financed with Junta Castilla y León funds.

References

[1] A. S. Lukes, M. F. Kohler, C. F. Pieper, B. J. Kerns, R. Bentley, G. C. Rodriguez, J. T. Soper, D. L. Clarke-Pearson, R. C. Bast and A. Berchuck. Multivariable analysis of dna ploidy, p53, and her-2/neu as prognostic factors in endometrial cancer. *Cancer*, 73(9):2380–2385, 1994.

[2] J. M. Corchado, J. F. D. Paz, S. Rodrguez and J. Bajo. Model of experts for decision support in the diagnosis of leukemia patients. *Artificial Intelligence in Medicine*, 46(3):179 – 200, 2009. URL http://www.sciencedirect.com/science/article/pii/S0933365708001875.

[3] J. F. D. Paz, J. Bajo, V. Vera and J. M. Corchado. Microcbr: A case-based reasoning architecture for the classification of microarray data. *Applied Soft Computing*, 11(8):4496 – 4507, 2011. URL http://www.sciencedirect.com/science/article/pii/S1568494611003048.

[4] B. Ylstra, P. van den IJssel, B. Carvalho, R. H. Brakenhoff and G. A. Meijer. Bac to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array cgh). 34(2):445–450, 2006. URL http://nar.oxfordjournals.org/content/34/2/445.abstract.

[5] D. Pinkel and D. G. Albertson. Array comparative genomic hybridization and its applications in cancer. *Nature genetics*, 37:S11–S17, 2005.

[6] K. K. Mantripragada, P. G. Buckley, T. D. de Sthl and J. P. Dumanski. Genomic microarrays in the spotlight. *Trends in Genetics*, 20(2):87 – 94, 2004. URL http://www.sciencedirect.com/science/article/pii/S0168952503003470.

[7] P. Wang, Y. Kim, J. Pollack, B. Narasimhan and R. Tibshirani. A method for calling gains and losses in array cgh data. *Biostatistics*, 6(1):45–58, 2005. URL http://biostatistics.oxfordjournals.org/content/6/1/45.abstract.

[8] K. Kapur, Y. Xing, Z. Ouyang and W. H. Wong. Exon arrays provide accurate assessments of gene expression. *Genome Biol*, 8(5):R82, 2007.

[9] T.-P. Nguyen and T.-B. Ho. Detecting disease genes based on semi-supervised learning and proteinprotein interaction networks. *Artificial Intelligence in Medicine*, 54(1):63 – 71, 2012. URL http://www.sciencedirect.com/science/article/pii/S0933365711001230.

[10] The universal protein resource (uniprot). *Nucleic Acids Research*, 35(suppl 1):D193–D197, 2007. URL http://nar.oxfordjournals.org/content/35/suppl_1/D193.abstract.

[11] G. O. Consortium. The gene ontology (go) database and informatics resource. *Nucleic Acids Research*, 32(suppl 1):D258–D261, 2004. URL http://nar.oxfordjournals.org/content/32/suppl_1/D258.abstract.

[12] R. D. Finn, J. Tate, J. Mistry et al. The pfam protein families database. *Nucleic Acids Research*, 36(suppl 1):D281–D288, 2008. URL http://nar.oxfordjournals.org/content/36/suppl_1/D281.abstract.

[13] S.-K. Ng, Z. Zhang, S.-H. Tan and K. Lin. Interdom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Research*, 31(1):251–254, 2003. URL http://nar.oxfordjournals.org/content/31/1/251.abstract.

[14] G. Joshi-Tope, M. Gillespie, I. Vastrik et al. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research*, 33(suppl 1):D428–D432, 2005. URL http://nar.oxfordjournals.org/content/33/suppl_1/D428.abstract.

[15] E. T. Dermitzakis. From gene expression to disease risk. *Nature genetics*, 40(5):492–493, 2008.

[16] R. Maglietta, A. DAddabbo, A. Piepoli, F. Perri, S. Liuni, G. Pesole and N. Ancona. Selection of relevant genes in cancer diagnosis based on their prediction accuracy. *Artificial Intelligence in Medicine*, 40(1):29 – 44, 2007. URL http://www.sciencedirect.com/science/article/pii/S0933365706000972.

[17] J. F. De Paz, R. Benito, J. Bajo, A. E. Rodríguez and M. Abáigar. acgh-mas: Analysis of acgh by means of multiagent system. *BioMed research international*, 2015.

[18] U. Nuber. *DNA microarrays*. Taylor & Francis, 2007.

[19] S. Knudsen. *Cancer diagnostics with DNA microarrays*. John Wiley & Sons, 2006.

[20] J. Zhang, L.-Y. Wu, X.-S. Zhang and S. Zhang. Discovery of co-occurring driver pathways in cancer. *BMC Bioinformatics*, 15(1):271, 2014. URL http://dx.doi.org/10.1186/1471-2105-15-271.

[21] B. Vogelstein and K. W. Kinzler. Cancer genes and the pathways they control. *Nature Med*, 10(8):789–799, 2004.

[22] D. J. Lockhart and E. A. Winzeler. Genomics, gene expression and dna arrays. *nature*, 405(6788):827–836, 2000.

[23] B. N. U. of Singapore. Therapeutic targets database, 2015. URL http://bidd.nus.edu.sg/group/cjttd/.

[24] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam and M. Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, 36(suppl 1):D901–D906, 2008.

[25] Y. W. Choon, M. S. Mohamad, S. Deris, R. M. Illias, C. K. Chong, L. E. Chai, S. Omatu and J. M. Corchado. Differential bees flux balance analysis with optknock for ¡italic¿in silico¡/italic¿ microbial strains optimization. *PLoS ONE*, 9(7):e102744, 2014. URL http://dx.doi.org/10.1371%2Fjournal.pone.0102744.

[26] E. Argente, V. Botti, C. Carrascosa, A. Giret, V. Julian and M. Rebollo. An abstract architecture for virtual organizations: The thomas approach. *Knowledge and Information Systems*, 29(2):379–403, 2011. URL http://dx.doi.org/10.1007/s10115-010-0349-1.

[27] N. J. Armstrong and M. A. van de Wiel. Microarray data analysis: from hypotheses to conclusions using gene expression data. *Analytical Cellular Pathology*, 26(5-6):279–290, 2004.

[28] J. Ezan, L. Leroux, L. Barandon, P. Dufourcq, B. Jaspard, C. Moreau, C. Allières, D. Daret, T. Couffinhal and C. Duplàa. Frza/sfrp-1, a secreted antagonist of the wnt-frizzled pathway, controls vascular cell proliferation in vitro and in vivo. *Cardiovascular research*, 63(4):731–738, 2004.

[29] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, 42(Database issue):199–205, 2014.

[30] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28(1):27–30, 2000.

[31] E. Klopocki, G. Kristiansen, P. J. Wild et al. Loss of sfrp1 is associated with breast cancer progression and poor prognosis in early stage tumors. *International journal of oncology*, 25(3):641–649, 2004.

[32] C. Bernemann, C. Hülsewig, C. Ruckert et al. Influence of secreted frizzled receptor protein 1 (sfrp1) on neoadjuvant chemotherapy in triple negative breast cancer does not rely on wnt signaling. *Mol Cancer*, 13:174, 2014.

[33] Y.-L. Shih, C.-B. Hsieh, H.-C. Lai, M.-D. Yan, T.-Y. Hsieh, Y.-C. Chao and Y.-W. Lin. Sfrp1 suppressed hepatoma cells growth through wnt canonical signaling pathway. *International journal of cancer*, 121(5):1028–1035, 2007.

[34] S. E. Singletary, G. L. Robb and G. N. Hortobagyi. *Advanced therapy of breast disease*. PMPH-USA, 2004.

[35] G. M. Sizemore, S. T. Sizemore, D. D. Seachrist and R. A. Keri. Gaba (a) receptor pi (gabrp) stimulates basal-like breast cancer cell migration through activation of extracellular-regulated kinase 1/2 (erk1/2). *Journal of Biological Chemistry*, 289(35):24102–24113, 2014.

[36] A. Köhler, Ü. Demir, E. Kickstein et al. A hormone-dependent feedback-loop controls androgen receptor levels by limiting mid1, a novel translation enhancer and promoter of oncogenic signaling. *Mol Cancer*, 13:146, 2014.