Genes Involved in the Seminoma Testicular Cancer: A Bioinformatic Study

Párraga-Álava Jorge Antonio[1]

[1] *Universidad de Santiago de Chile, Departamento de Ingeniería Informática, Avenida Ecuador #3659. Santiago, Chile.*

Abstract: This study aims to identify genes differentially expressed between: (1) normal tissue samples and tissue samples with testicular cancer, and (2) progress stages of testicular seminoma (cancer). In this context, data for the experiments were obtained from the Repository of the National Center for Biotechnology Information (NCBI). On them a cleaning process and pre-processing was performed through the elimination of null or missing values, then, and in order to perform dimensionality reduction of data on them, statistical test t-student was applied to establish genes with capacity to discriminate between the different conditions. Next, Significance Analysis of Microarrays (SAM) was carried out to identify genes that were differentially expressed. Thereby a set of 40 differentially expressed genes in normal samples and cancer samples; and 11 genes in the case of the stages of the disease were identified, all of them were subjected to a functional biological analysis. Finally, it was evidenced that genes *HSPA2*, *SPINK2* and *POU5F1P3* were coincident with previous studies in terms of being labeled as genes responsible for seminoma testicular cancer.

Keywords: Bioinformatics, genes, differential expression, seminoma, testicular cancer, SAM.

Genes Involucrados en el Cáncer Testicular Seminoma: Un Estudio Bioinformático

Resumen: Este estudio tuvo como objetivo identificar genes diferencialmente expresados entre: (1) muestras de tejido normal y muestras de tejido con cáncer testicular, y (2) etapas de progresión del seminoma testicular (cáncer). En este contexto, los datos para realizar los experimentos fueron obtenidos del Repositorio del Centro Nacional de Información Biotecnológica (NCBI). Sobre ellos se realizó un proceso de limpieza y el pre-procesamiento a través de la eliminación de los valores nulos o datos faltantes, luego, y con el objetivo de realizar reducción de dimensionalidad de los datos, se aplicó el test estadístico t-student para establecer genes con capacidad discriminatoria entre las distintas condiciones. Enseguida, se aplicó el Análisis de Significancia de Microarreglos (SAM) para identificar los genes que estaban expresados diferencialmente. Con ello se identificó un conjunto de 40 genes diferencialmente expresados en muestras normales y muestras con cáncer; y 11 genes en el caso de las etapas de dicha enfermedad, todos ellos fueron sometidos a un análisis biológico funcional. Por último, se evidenció que los genes *HSPA2, SPINK2 y POU5F1P3* coinciden con estudios previos en cuanto a ser catalogados como genes responsables del cáncer testicular tipo seminoma.

Palabras clave: Bioinformática, genes, expresión diferencial, seminoma, cáncer testicular. SAM.

1. INTRODUCTION

High-throughput technology is rapidly becoming the standard method for measuring genomic information. The rapid advent of these technologies along with reduced costs have helped detailed profiling of gene expression levels, impacting almost every field in life sciences and is now being adopted for clinical use (Berger et al., 2010). Microarray technology is one which has been at the forefront of the study of simultaneous changes in gene expression and quantifying it through the entire genome (Syrenne et al., 2012).

Among these studies are highlighted analysis of transcriptional behavior of biological systems, grouping states of ill or healthy cells, and identifying of genes whose patterns of expression differ according to phenotype or experimental condition, this difference allows infer changes as a result of treatment, disease or other causes. Because, this task is widely used in fields such as pharmaceuticals, in view of facilitating the development of medicines to the suppression or activation of genes related to various types of cancer. As a result, in this study an analysis of gene expression on microarray data will perform to identify genes differentially expressed between normal tissue samples and tissue samples with testicular cancer, and at different stages.

Germ cell tumors of the testis (TCGT) represent between 90 and 95 % of neoplasms of male gonada. The remaining 5 %

jorge.parraga@usach.cl
Received: 25/03/2015
Accepted: 03/05/2016
Published: 30/09/2016

are testicular tumors of germ cells. The main way to detect this cancer usually begins with a physical exam, which seeks swelling, tenderness or any foreign lump in the testicles; however, the diagnosis is confirmed or rejected through blood tests or ultrasound (Moreno et al., 2014). The two main types of this cancer are seminoma and non seminoma, which grow, spread and they are treated differently: seminomas are tumors with slow growth, usually occur in the testes in men between 30 and 40 years old, but it can spread to lymph nodes (Bray et al., 2006). Meanwhile, non seminomas tend to grow and spread more quickly than seminoma, usually it occurs to men in the last teens and early 30's (Aparicio et al., 2014). The incidence of this cancer has increased in recent decades in many countries; reasons for this increase are unknown because the risk factors for disease are poorly treated (McGlynn and Trabert, 2012). In Latin America the situation follows the global trend, and in the case of Chile, as Vidal (2014) suggests, this type of cancer occurs especially in patients between 20 and 40 years, with an incidence rate -7 out of every 100 000 men- which is high compared to other Latin American countries.

The testicular seminoma is one of type of cancers that can be cured by chemotherapy, in fact over 20 years this treatment was able to confront this disease, as despite presenting in advanced stages, today it is treatable and recoverable with a five-year survival, nearly 90 % of cases (Tandstad et al., 2009). For this reason, the search for genes differentially could help to identify early genes causing of testicular seminoma and thereby, improve treatment for patients and people at risk of suffering it.

This work is organized as follows. In section 2, we present a brief review of similar works available in the literature. In section 3, we present the methodology used to carry out the identification of differentially expressed genes, detailing the processes for analysis and data manipulation performed. Section 4, is devoted to a biological and graphical analysis of the results obtained. Final conclusions and guidelines for further work are presented in section. 5.

2. RELATED WORK

In recent years there have been some researches performed on testicular cancer, all of them with different approaches; however, they have tried usually to find which genes, or protein sequences are influencing development or not of this pathology. In Turnbull et al. (2010), a study on the genome of tumor testicular was conducted, genotyping 298 782 SNPs (Single Nucleotide Polymorphism) in 979 individuals affected in the United Kingdom. They showed that three loci on chromosomes 5, 9, and 12; corresponding to genes *TERT*, *DMRT1*, and *ATF7IP* were associated with testicular germ cell cancer. Similarly, in Lau et al. (2010) an extensive analysis was carried out on *TCL1* protein expression in 63 cases in California, United States, which were distributed as follows: 23 seminomas, 14 embryonal carcinomas, 4 teratomas, 2 sac tumors vitelino, and 20 mixed germ cell tumors. Authors of this study concluded that *TCL1* protein was expressed in 20 of 23 cases of pure seminoma and non seminoma in areas mixed germ cell tumors. Conversely, all histological types of germ cell tumors nonseminomatous, including embryonal carcinoma, tumor yolk sac, teratoma, and choriocarcinoma showed no evidence expression of *TCL1* in their pure forms or as components mixed in germ cell tumors. Meanwhile, the review article in Sheikine et al. (2012) offers a summary of current knowledge in genetics underlying development, progression and chemoresistance of TGCT, in it is clear that *KIT, TP53, KRAS/BRAF* and *NRAS* are the most frequently mutated genes in TGCT and involved in its pathogenesis. Finally, the study in Gashaw et al. (2005) exposes a process of differential expression of variants of certain genes on patients with testicular cancer. In the same, *POU5F1P3, CETN1, SPINK2, HSPA2* stand out as highly expressed genes in tissues with testicular seminoma.

3. METHODOLOGY

To carry out this study, the pipeline shown in Figure 1 was followed. In the first stage, data pre-processing was performed in order to eliminate lost and missing data. Later, statistics test t-student was conducted to establish genes with capacity to discriminate between different conditions and to achieve a reduction of data dimensionality. Next, differentially expressed genes were determined using SAM. Finally, functional biological analysis was done.

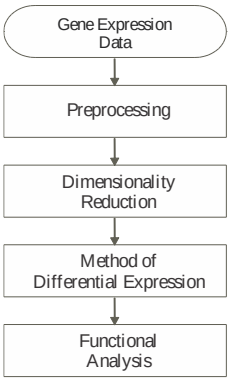


Figure 1. Pipeline for identifying differentially expressed genes.

3.1. Dataset

The dataset with gene expression data "testicular seminoma progression" was used; it was obtained from NCBI, and it includes expressions level in 12 580 probes with 43 samples of testicular tissue, which are divided into:

- Normal, (samples 1-3): 3 high quality testicles samples of patients between 34 and 67 years, with normal spermatogenesis, after vasectomy.
- Stage 1 (pT1), (samples 4-25): 22 tumor samples corresponding to patients between 25 and 56 years.
- Stage 2 (pT2), (samples 26-39): 14 tumor samples corresponding to patients between 27 and 46 years.
- Stage 3 (pT3), (Samples 40-43): 4 tumor samples corresponding to patients between 23 and 49 years.

According to American Joint Committee on Cancer (AJCC) (Cuccurullo and Mansi, 2011): the stage 1 (pT1) is limited to the testis and epididymis without vascular/lymphatic invasion; tumor may invade tunica albuginea but not the tunica vaginalis. The stage 2 (pT2) is limited to the testis and epididymis with vascular/lymphatic invasion, or tumor

extending through the tunica albuginea with involvement of the tunica vaginalis tumor. The Stage 3 (pT3) the tumor invades the spermatic cord with or without vascular invasion.

3.2. Pre-processing

Once the gene expression dataset with 12 580 probes in 43 samples was obtained, it was necessary to clean them to eliminate missing, repeated data or any other typical variation resulting from the application of microarrays technology. Thereby, probes that exhibited those problems were separated from the dataset. In order to visualize the behavior of expression level and determine the need for data cleaning, a set of 13 genes was taken, which according to literature are involved in testicular cancer. *DMRT1* and *ATF7IP* were not included in the dataset, thus expressions level of the remaining 11 genes are shown in Figure 2 and Figure 3.

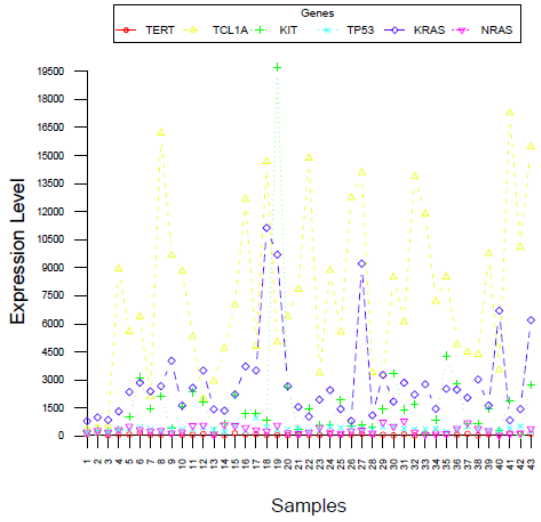


Figure 2. Expression of TERT, TCL1A, KIT, TP53, KRAS and NRAS in all samples of the dataset.

Figure 2 shows that gene *TERT* presents similar expression levels over all samples. The gene *TCL1A*; however, had low expression in normal testicular tissue samples (1-3 samples), but extremely high values in cancer tissue samples (4-43 samples). For its part, expressions of *KIT*, *NRAS* and *TP53* were homogeneous throughout the microarray. The expression of the gene *KRAS* was variable because in normal samples it has a low expression, while in cancer samples, in some cases it increases its expression, but decreases in others. Meanwhile, figure 3 shows clearly that genes *BRAF*, *CETN1*, *SPINK2* and *HSPA2* had higher expression levels in normal testicular tissue (1-3 samples) and low levels in cancerous tissue (4-43 samples). The reverse situation occurred with gene *POU5F1P3*, which had low expression in normal samples and high levels in cancer samples. Once the behavior of the expression level of genes was verified, it was necessary to identify and eliminate sources of variation in the database that had no differences in expression. Therefore, those genes containing values "null" in their expression were deleted. In the case of repeated genes, the expression level was averaged over all samples.

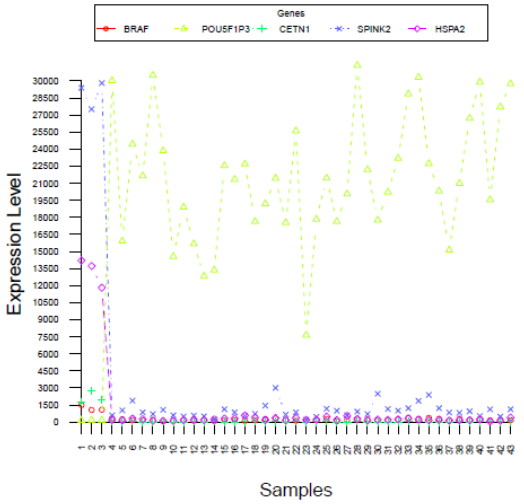
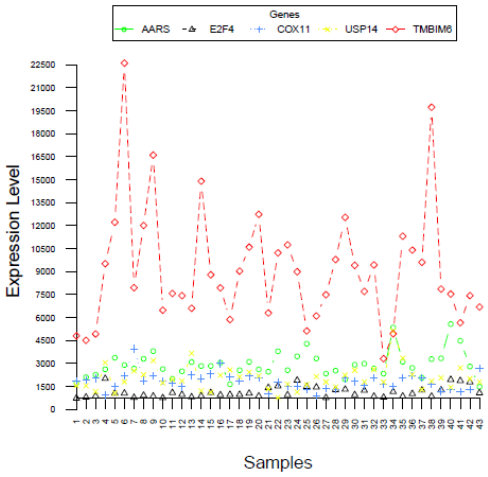
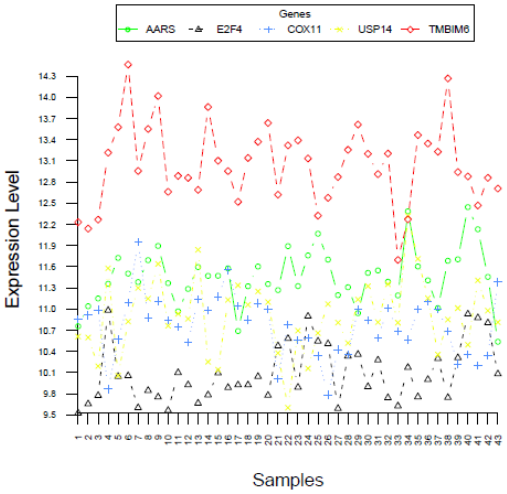


Figure 3. Expression of BRAF, POU5F1P3, CETN1, SPINK2 and HSPA2 in all samples of the dataset.

Furthermore, and in order to facilitate the calculation of differentially expressed genes, data were transformed to logarithmic scale ($\log_2$), as shown in Figure 4.



a) Expression of AARS, E2F4, COX11, USP14 and TMBIM6.



b) Expression in Log_2 of AARS, E2F4, COX11, USP14 and TMBIM6.

Figure 4. Level expression of some genes scaled to $\log_2$.

3.3. Dimensionality Reduction

In order to perform the reduction of data dimensionality, genes with expression levels that could be discriminated between the types of samples were identified. This discrimination was done through a test t-student test, considering the following hypothesis:

Hypothesis Test

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$
$$H_1 : \mu_1 \neq \mu_2 \neq ... \neq \mu_k$$

In hypothesis test, μ_k corresponds to expression level of each gene in condition *k*, with *k={1,2}* for normal and cancer, and *k={1,2,3}* for stages of seminoma. The value of α considered for analysis was 0,05. So, if a gene has a *p-value* < α it is then differentially expressed in the microarray, therefore it has discriminatory ability and is maintained in the dataset. Otherwise, genes are deleted. Such test was carried out using the library of *R* called *stats* (Wickham, 2015). Based on previous test and the preprocessing stage, thousands of genes were discarded due to having missing, "null", or *p-value > 0,05*, which means that they were expressed in a similar manner along the microarray and, therefore, their behavior patterns do not have capacity to discriminate between classes. In this way, it was possible to reduce the dimensionality of the dataset to 5 380 genes candidates to be differentially expressed under diverse conditions.

3.4. Method for Differential Expression

To identify differentially expressed genes under certain conditions was used SAM (Significance Analysis of Microarrays) (Tusher et al., 2001). This method was chosen because it was exclusively designed for this type of genetic data. It operates as follows:

1. For each gene *i*, compute *d-value*. This is the *observed d-value* for gene *i*, and calculated by equation (1).

$$d(i) = \frac{\bar{X}_i - \bar{Y}_i}{s_i + s_0} \quad (1)$$

Where:

- $\bar{X}_i$ is the mean gene expression *i* in the class *X*.
- $\bar{Y}_i$ is the mean gene expression *i* in the class *Y*.
- s_i is "gene-specific scatter" or standard deviation of gene *i* in repeated expression measurements.
- s_0 parameter to ensure that constant variance *d(i)* is independent of gene expression.

2. Rank the genes in ascending order of their *d-values*.
3. Randomly shuffle the values of the genes between class *X* and *Y*, such that the reshuffled class *X* and *Y* respectively have the same number of elements as the original class *X* and *Y* (Figure 5).
4. Compute the *d-value* for each randomized gene. The mean of all these values constitute the permutation d-value expected for each gene.
5. Definitely, *observed d-value* and *expected d-value* are used to determine the significant difference according to a threshold delta (Δ).
6. Plot the *observed d-values* vs the *expected d-values.*

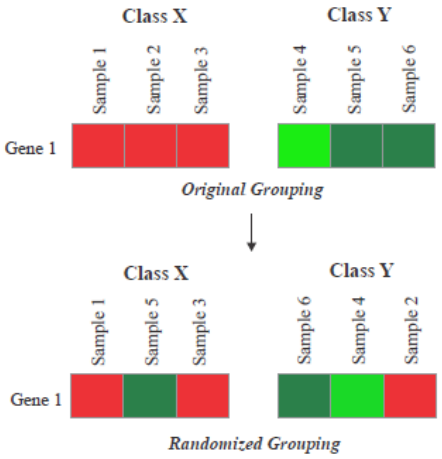


Figure 5. Permutation example. Green tones indicates reduced expression, red tones indicates increased expression

In order to carry out SAM, the library of *R* called *siggenes* (Schwender, 2011) was used considering two versions of this method. The first corresponds to normal and cancer samples, due to having this number of classes, the design of SAM to use is *Unpaired Two-Class*, where genes whose mean expression level is significantly different between two classes are identify as differentially expressed. The second version corresponds to cancer samples for stage 1 (pT1), stage 2 (pT2), and stage 3 (pT3), i.e. *Unpaired Multi-Class*, which selects genes whose mean expression is different across more than two classes of samples.

4. RESULTS AND DISCUSSION

Once was performed the implementation of SAM with two types of design: SAM *Unpaired Two-Class* and *Multi-Class*, the following results were obtained. For both designs, various values for parameters thresholds (Δ) and FDR were set, visualizing how they affect the amount of differentially expressed genes that can be obtained. Such values are shown in Figure 6 and Figure 7.

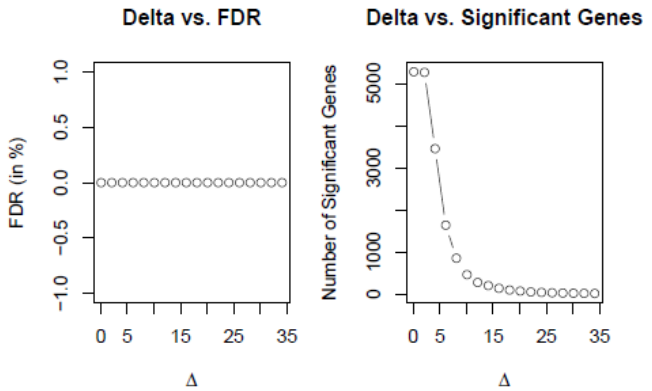


Figure 6. Possible significant genes and FDR with different delta (Δ) values for normal and cancer samples.

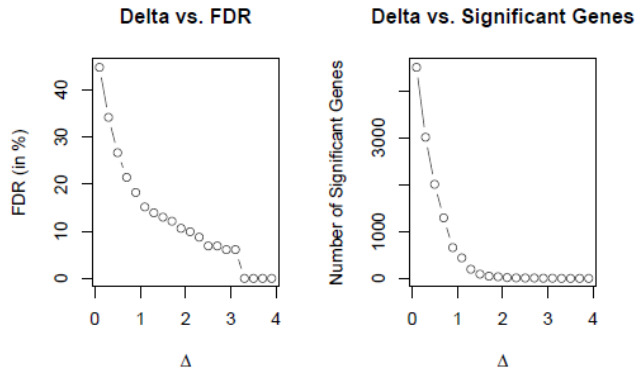


Figure 7. Possible significant genes and FDR with different delta (Δ) values for cancer stages samples.

Moreover, main values of these for first design (normal and cancer classes), are shown in Table 1.

Table 1. Summary of First Design (SAM Unpaired Two-Class).

n	Delta (Δ)	called	FDR
1	1,3	5 297	0
2	5,5	1 980	0
3	9,8	555	0
4	14,0	208	0
5	18,3	85	0
6	22,6	40	0
7	26,8	22	0
8	31,1	6	0
9	35,3	2	0
10	39,6	0	0

In Table 1, column delta (Δ) indicates a threshold for significance that sets the limits of values from which genes are significant based on differential expression between sets of samples. The third column corresponds to the number of genes that are significant according to the given threshold. Column FDR (False Discovery Rate) shows the percentage of genes that were erroneously classified as differentially expressed. In this table is possible to observe that a maximum of 5 297 differentially expressed genes were found with threshold (Δ) of 1.3. While a minimum of 2 differentially expressed genes were obtained with threshold (Δ) of 35,3. In both cases, no gene was identified erroneously as differentially expressed. The threshold (Δ) of 22,6 is the most appropriate value because it locates **40 significant genes,** which is a quantity large enough to biological functional analysis (Huang et al., 2008).

The result of using this threshold (Δ) is shown in Figure 8, where green circles over the dotted line correspond to significant positive genes, namely that the mean expression of cancer class is greater than mean expression of normal class; while, green circles below the dotted line correspond to significant negative genes, namely that mean expression of normal class is greater than mean expression of cancer class.

In Figure 8, it can be observed that gene *POU5F1P3* is the unique that has positive significance (lonely green circle over the dotted line), that is to say it was overexpressed in microarray. Opposite situation occurs with the remaining 39 genes that have negative significance (green circles over the dotted line), namely their expressions are suppressed data.

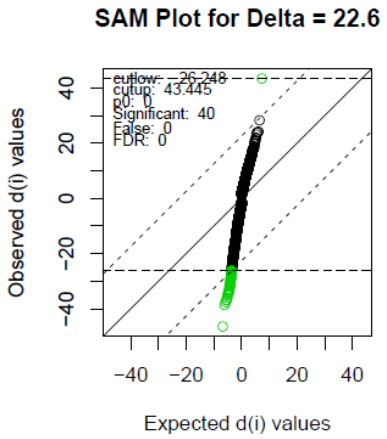


Figure 8. Significant genes or differentially expressed (green circle) with delta (Δ)=22,6 for normal and cancer samples.

Similarly, main results for second design (cancer stages classes) are shown in Table 2.

Table 2. Summary of Second Design (SAM Unpaired Multi-Class).

n	Delta (Δ)	called	FDR
1	0,1	4 500	0,4479
2	0,4	2 424	0,3010
3	0,8	932	0,1970
4	1,1	437	0,1516
5	1,4	113	0,1383
6	1,8	38	0,1144
7	2,1	17	0,0991
8	2,4	10	0,0754
9	2,8	6	0,0659
10	3,1	4	0,0611

In Table 2 can be observed that a maximum of 4 500 differentially expressed genes were found with threshold (Δ) of 0.1, but with FDR very high of 44,79%. While a minimum of 4 differentially expressed genes were obtained with threshold (Δ) of 3,1. The threshold (Δ) of 2,1 is the most appropriate value because it locates **17 significant genes,** which is a quantity large enough to biological functional analysis (Huang et al., 2008) and with low FDR, only 9,91%. The result of using this threshold (Δ) is shown in Figure 9, where there are only green circles over the dotted line which correspond to significant positive genes.

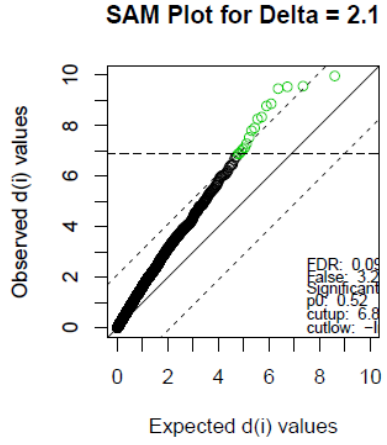


Figure 9. Significant genes or differentially expressed (green circle) with delta (Δ)=2,1 for cancer stages samples.

In Figure 9, it can be observed that 11 genes were differentially expressed and all of them have positive significance (green circles over the dotted line) and thus, they are overexpressed according to their level expression in dataset.

The list of differentially expressed genes for both design is shown in Table 3, on which a functional analysis about their participation in biological processes and coincidence with previous studies was performed.

Table 3. Differentially expressed genes in seminoma testicular (cancer).

	Normal vs Cancer	**Cancer Stages**
Gene	*GAPDHS, AQP5, ARL4A, ANKRD7, AKAP4, PRM2, LDHC, LRP8, PIAS2, SPINK2, ACRV1, ZPBP, PRM1, SMCP, CCIN, CRISP2, PGAM2, DNAH7, CT62, CCT6B, TNP1, TSSK2, ODF1, ART3, LOC81691, SOCS7, SPA17, HSPA2, GSTM3, GK2, TP53TG5, DYRK3, ACR, ODF2, CRAT, IZUMO4, PRKAR2A, ZMYND10, POU5F1P3, TEKT2.*	*ZNF165, GSR, PP14571, ZNF189, STK17B, STAR, PRKAA1, SLC12A2, ARNT2, NRIP1, APOD, GK3P, ATN1, KIAA0020, TMEM251, ROR1, PGRMC1.*
Biological processes	Sexual reproduction, multicellular organism reproduction, reproductive process in a multicellular organism, spermatogenesis, male gamete generation, gamete generation, spermatid development, spermatid differentiation, germ cell development.	Acetylation, catalytic activity, phosphoprotein, polymorphism information online, variant sequence of mitochondria, phosphoprotein, alternative splicing, splicing variant, phosphoprotein, biological regulation, metabolic process, response to stimulus.
Coincidence	*HSPA2, SPINK2, POU5F1P3.*	None.

In this table, it is possible to observe the **official gene symbol** of 40 differentially expressed genes between the normal vs cancer samples, and 11 between cancer stages. In addition, main biological processes in which they participate are shown, information was obtained using functional annotations through DAVID (Database for Annotation, Visualization and Integrated Discovery) (Huang et al., 2008) and GO (Gene Ontology) (Ashburner et al., 2000), (GO Consortium, 2015). Biological processes in which they participate are many; however, in the case of normal vs cancer, we can infer that they are closely related to male reproduction due to the presence of terms like spermatide, sexual or gamete. Meanwhile, for cancer stages, it draw the attention that processes in which genes participate are not closely related to aspects of reproduction and sexuality human. In fact, they are linked to different processes at the chemical, cellular or molecular level.

Finally, contrasts with studies of Turnbull et al. (2010), Lau et al. (2010), Sheikine et al. (2012), and Gashaw et al. (2005) allows to identify matches with the genes *HSPA2*, *SPINK2* and *POU5F1P3* about being differentially expressed between normal and cancer tissues samples. However, in the case of the evaluation of cancer stages, there are not matches with previous works, likely because such studies were conducted in differential expression analysis considering only genes in healthy and diseased samples. Thereby, biological functions associated with these three genes are exposed. The gene *SPINK2* (also called *Serine protease inhibitor Kazal-type 2*) has high importance in the processes of proliferation of male germ cells, spermatogenesis, development of seminiferous tubules, gonads and fertilization The gene *HSPA2* (also called *Heat shock 70kDa protein* 2, isoform CRA\a) participates in processes involved in positive regulation of protein phosphorylation, and male meiosis I spermatid development. Finally, gene *POU5F1P3* (also called *POU domain transcription factor OCT4-pg3*) is a fundamental part of the regulation of *DNA* transcription and *DNA* binding (Binns et al. 2009).

5. CONCLUSIONS AND FURTHER WORKS

The application of SAM on datasets allowed to identify genes differentially expressed for testicular seminoma in different amounts and with a high degree of accuracy by having very low rates of false discoveries (FDR). A total of 40 genes for samples of testicular tissue with cancer were listed as differentially expressed, while 11 did for samples of testicular tissue in different cancer stages. Moreover, thanks to a comparison with the literature, it was possible to identify three genes matches (*HSPA2*, *SPINK2* and *POU5F1P3*), whose biological functions were analyzed observing that they were related to various processes of male germ cells.

Most genes identified have not been reported in previous studies, which does not mean they have no connection with testicular cancer. Thus, a deeper biological level analysis is needed to draw conclusions. Furthermore, the fact that the articles cited in the literature and the studies used in this work correspond to databases of patients from a certain country or region causes that the results are specific to their study population. Therefore, further analysis is required to generalize the results, such that the list of differentially expressed genes be applicable to any man, regardless of race or place of origin.

ACKNOWLEDGMENT

The author gratefully acknowledges the Secretariat for Higher Education, Science, Technology and Innovation (SENESCYT) of Ecuador for its financial support through a doctoral scholarship.

REFERENCES

Aparicio, J., Maroto, P., García del Muro, X., Sánchez-Munoz, A., Guma, J., Margelí, M., . . . Arranz, J. A. (2014). Prognostic factors for relapse in stage I seminoma: a new nomogram derived from three consecutive, risk-adapted studies from the Spanish Germ Cell Cancer Group (SGCCG). *Annals of Oncology, 25*(11), 2173-2178.

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J. M., . . . others. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics, 25*(1), 25-29.

Berger, M., Levin, J., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., . . . Sougnez, C. (2010). Integrative analysis of the melanoma transcriptome. *Genome research, 20*(4), 413-427.

Binns, D., Dimmer, E., Huntley, R., Barrell, D.,… Apweiler, R. (2009). QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, 25(22), 3045-3046.

Bray, F., Ferlay, J., Devesa, S., McGlynn, K., & Moller, H. (2006). Interpreting the international trends in testicular seminoma and nonseminoma incidence. *Nature Reviews Urology*, *3*(1), 532-543.

Cuccurullo, V., & Mansi, L. (2011). AJCC Cancer Staging Handbook: from the AJCC Cancer Staging Manual. *European Journal of Nuclear Medicine and Molecular Imaging*, *38*(2), 408-408.

Gashaw, I., Grummer, R., Klein-Hitpass, L., Dushaj, O., Bergmann, M., Brehm, R., . . . Kliesch, S. (2005). Gene signatures of testicular seminoma with emphasis on expression of ets variant gene 4. *Cellular and Molecular Life Sciences CMLS*, *62*(1), 19-20.

GO Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, *43*(D1), D1049-D1056.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, *4*(1), 44-57.

Lau, S., Weiss, L., & Chu, P. (2010). TCL1 Protein Expression in Testicular Germ Cell Tumors. *American Journal of Clinical Pathology*, *133*(1), 762-766.

McGlynn, K., & Trabert, B. (2012). Adolescent and adult risk factors for testicular cancer. *Nature Reviews Urology*, *9*(1), 339-349.

Moreno, A., Domínguez, A., Alpuente, C., Hernando, A., Torres, J., & Cabrera, J. (2014). Características de la forma de presentación del cáncer de testículo en hospitales públicos de la Comunidad Autónoma de Madrid. *Actas Urológicas Españolas*.

Schwender, H. (2011). Siggenes: Multiple Testing using SAM and Efron Empirical Bayes Approaches. *R package version*.

Sheikine, Y., Genega, E., Melamed, J., Lee, P., Reuter, V., & Ye, H. (2012). Molecular genetics of testicular germ cell tumors. *American Journal of Cancer Research*, *2*(2), 153-167.

Syrenne, R., Shi, W., Stewart, C., & Yuan, J. S. (2012). Omics platforms: Importance of twenty-first century genome-enabled technologies in seed developmental research for improved seed quality and crop yield. *Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield*, 43-57.

Tandstad, T., Dahl, O., Cohn-Cedermark, G., Cavallin-Stahl, E., Stierner, U., Solberg, A., & Langberg. (2009). Risk-adapted treatment in clinical stage I nonseminomatous germ cell testicular cancer: the SWENOTECA management program. *Journal of Clinical Oncology*, *27*(13), 2122-2128.

Turnbull, C., Rapley, E., Seal, S., Pernet, D., Renwick, A., Hughes, D., . . . Easton, D. (2010). Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nature Genetics*, *42*(7), 604-607.

Tusher, V., Tibshiraniand, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, *98*(9), 5116-5121.

Vidal, I. (2014). El cáncer que afecta principalmente a hombres jóvenes en Chile. *Centro Oncológico Oncovida*.

Wickham, H. (2015). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing.*

Jorge Párraga Álava was born in Calceta, Ecuador, on May 28, 1987. He received the Master degree in systems management from the Army Polytechnic School, Quito, Ecuador in 2013. Currently, he is working toward the Ph. D. degree in engineering sciences with a major in informatics at University of Santiago, Chile. His research interests include multi-objective clustering techniques and their application to gene expression data.