Identification of Genes Associated with Testicular Cancer Aggressiveness

Graciele de Souza Medeiros; Barbara Cardoso de Oliveira; Vinícius Barbosa Parula Fernandes; Vinícius Santos Cardoso; Gabriel Arantes dos Santos; Poliana Romão da Silva; Sabrina Thalita dos Reis

ABSTRACT

Introduction: Testicular germ cell tumors represent approximately 97% of testicular cancers. Histologically, they are classified into seminomas and non-seminomas, having diagnostic and prognostic applicability. Therapeutic success depends on early diagnosis associated with correct staging, the evaluation of biomarkers is important for the correct management of this disease. Objective: To identify genes that may be correlated with prognosis and survival in testicular cancer. Method: Bioinformatics analysis was performed using 137 testicular cancer samples from The Cancer Genome Atlas and 165 normal testicular tissue samples from The Genotype-Tissue Expression. Gene identification and subsequent analyses were performed using GEPIA2. Results: Initially, in relation to gene expression, the 500 genes most significantly associated with overall survival from testicular cancer and the 500 with disease-free survival were evaluated. These two lists were then superimposed and a Venn diagram was constructed showing the 13 genes in common. Of these, only the protein-coding genes were kept, investigating which ones differed significantly from normal tissue in relation to gene expression. Only ATP10A, SAMD14 and PCAL4 showed a statistically significant difference, all of which were under-expressed in testicular cancer. The joint analysis of these genes was even more significant for overall and disease-free survival. Conclusion: Three genes were identified in the analysis in silico which demonstrated significant association of the expression with survival and prognosis of patients with testicular cancer.

Key words: Testicular Neoplasms; Germ Cells; Computational Biology/statistics & numerical data; Prognosis.

RESUMO

Introdução: Os tumores de células germinativas testiculares representam cerca de 97% dos cânceres testiculares. Histologicamente, classificam-se em seminomas e não seminomas, tendo aplicabilidade diagnóstica e prognóstica. O sucesso terapêutico depende do diagnóstico precoce associado ao correto estadiamento, sendo então de grande importância a avaliação de biomarcadores que possam contribuir para o manejo dessa doença. Objetivo: Identificar os genes que podem estar correlacionados com o prognóstico e a sobrevida no câncer testicular. Método: Análise de bioinformática utilizando 137 amostras de câncer testicular do The Cancer Genome Atlas e 165 amostras de tecido testicular normal do The Genotype-Tissue Expression. A identificação dos genes e análises subsequentes foram feitas pelo GEPIA2. Resultados: Inicialmente avaliou-se, em relação à expressão génica, os 500 genes mais associados com a sobrevida global do câncer testicular e o 500 com a sobrevida livre de doença. Em seguida, foi realizada a superposição dessas duas listas e construído um diagrama de Venn mostrando os 13 genes em comum. Destes, mantiveram-se apenas os codificadores de proteína, verificando quais diferiram significativamente do tecido normal em relação à expressão génica. Somente ATP10A, SAMD14 e PCAL4 mostraram diferença com significância estatística, todos subexpressos no câncer testicular. A análise deles em conjunto foi ainda mais significativa para a sobrevida global e livre de doença. Conclusão: Foram identificados nesta análise in silico três genes que demonstraram associação significativa de sua expressão com a sobrevida e o prognóstico dos pacientes com câncer testicular.

Palavras-chave: Neoplasias Testiculares; Células Germinativas; Biologia Computacional/estatística & dados numéricos; Prognóstico.
INTRODUCTION

Testicular cancer (TC) is the most common tumor among men aged 15 to 34-years-old\(^1\). It can present itself as Leydig cell tumor, Sertoli cell tumor, granulosa cell tumor, which are usually benign, but have malign potential, in addition to testicular germ cell tumors (TGCT), which are malignant and correspond to 95% of TC\(^2\). Histologically, TGCT can be classified in seminomas and non-seminomas. This histological distinction is highly relevant for etiology and tumor treatment, since non-seminoma tumors are more prone to metastasis\(^1\).

The risk factors involving TC are related to cryptorchidism – condition which causes the testicle to not descend into the scrotum –, TC-positive family history, ethnicity, age, and infertility\(^1\).

These are considered rare neoplasms, with estimated incidence of about 5 for each 100 thousand individuals\(^3\). TC is usually aggressive, especially for non-seminomatous tumors, however, it has a high healing rate with current therapies\(^4\), with the mortality rate close to 0.26/100,000 in Brazil, showing a growth trend over the last couple of years\(^7\).

Nowadays, TC is diagnosed through serum tumor markers and image exams. When presenting testicular mass, individuals are recommended to take a testicular ultrasound followed by a serum tumor marker test, including α-fetoprotein (AFP), human chorionic gonadotropin (hCG), and lactate dehydrogenase (LDH). However, these markers still present low specificity to TGCT, thus the need for new studies with the aim of finding biomarkers that can help manage TC\(^7\).

As with other neoplasms, therapeutic success depends on early diagnosis associated with correct staging\(^4\), which can be improved by adding molecular markers. In view of this, studies search for specific markers for seminomas and biomarkers that can help stratify the risk related to clinical staging, and guide the therapeutic choice and monitoring, especially in advanced stages of the disease\(^9\). Based on that, the objective of this study is to identify the genes that can be associated to prognosis and survival in TC.

METHOD

Bioinformatics analysis was performed using 137 testicular cancer (TC) samples from The Cancer Genome Atlas (TCGA)\(^9\), and 165 normal testicular tissue samples from The Genotype-Tissue Expression (GTEX)\(^11\). As this was an \textit{in silico} analysis, this study required no approval from the Research Ethics Committee, in compliance with Resolution number 510\(^{12}/2016\) of the National Health Council.

To select the genes associated to survival, data available in the GEPIA2\(^{13}\) tool were used, with genes from the TCGA dataset selection, as overall survival method, later followed by a disease-free survival using the median as the cutoff. In this tool’s expression matrix, the color density in each block represents the median expression value of a gene in a given tissue, normalized by the maximum median expression value in all blocks. In this step, the first 500 genes with significant statistical difference were selected, that is, those that had the smaller \(p\) value. The overall survival and disease-free survival curves were produced in the same tool using the gene expression median as cutoff.

Analysis of the association of gene expression, presence of methylation with histological type, age, race, and stage of the disease was performed by the Ualcan\(^{14}\) tool, using “TCGA” followed by “TGCT – testicular germ cell tumors” as search terms, followed by correlated expression to “Histological type”, “Age”, “Race”, “Cancer individual stages”, and “Methylation” of each gene.

The assessment of genetic alterations in each gene was performed by the tool cBiportal\(^{19}\) using the search term “TESTICULAR” followed by “TCGA, Firehose Legacy”. Thus, the assessment of mutations and supposed changes in the number of GISTIC copies in each gene was performed.

Initially, 100 genes with the expression pattern similar to ATP10A, HPCAL4, and SAMD14 in testicular cancer were selected by the GEPIA2\(^{15}\) tool. Then, the WebGestalt\(^{16}\) tool was used for enrichment analysis, using as basic parameters the organism of interest: \textit{Homo sapiens}, functional database; gene ontology and method of interest: “ORA sample run” and pathway + KEGG of each list.

RESULTS

To select the differentially expressed genes in testicular cancer (TC), the platform GEPIA2 was used. The first 500 genes associated to overall TC survival and the 500 genes associated to disease-free survival were selected. Later, the superimposition of those two lists was performed and a Venn diagram was built with the objective of selecting the genes associated with overall survival and disease-free survival. This analysis found 13 genes significantly associated with the two analyzed survivals: SAMD14, CISD3, HPCAL4, RP11-442G21.2, RN7SL208P, ATP10A, CASC8, HOXA9, ANPEP, CTC-459F4.9, RPL35P1, TMLHE-AS1 and ARPC3P1. After analyzing each one of those genes, those that were coded from proteins resulting in ATP10A, SAMD14, HPCAL4, CISD3, HOXA9, and ANPEP were selected. To choose which genes would be more relevant in TC, the expression
of those six genes was assessed using the GEPIA2 platform in testicular tumors samples and comparing normal testicles samples available in the platform. Three genes have showed to be differentially expressed with statistical significance ($p < 0.05$) in testicular tumors – ATP10A, SAMD14, and HPCAL4 (Figure 1). All genes were subexpressed in TC when compared to the GTEx control.

The analysis included if the expression of those three genes was associated to survival data. This joint analysis was able to significantly predict overall survival ($p = 0.046$) and disease-free survival ($p = 0.013$) (Figure 2).

A comparison of the expression of the three genes was carried out between seminoma and non-seminoma histological type testicular tumors and it was observed that the expression of the three genes was greater in the non-seminoma histological type (Figure 3).

When evaluating the expression of genes with the tumor stage, greater expressions of each gene associated with a specific stage of the disease were observed. The ATP10A gene was more expressed in T1 tumors, HPCAL4, in T2 tumors, and SAMD14, in T3 tumors (Figure 4).

The expression of those genes was further analyzed in different ethnicities and age groups, showing that, regarding race, the ATP10A gene showed greater expression in Asian people (comparison between Caucasian vs. African American with $p = 2.857400E-01$, Caucasian vs. Asian, $p = 8.720400E-01$, and African American vs. Asian $p = 2.466800E-01$); SAMD14 showed greater expression in Caucasian and African Americans (comparison between Caucasians vs. African Americans with $p = 5.935400E-01$, Caucasians vs. Asians, $p = 7.486800E-01$, and African Americans vs. Asians, $p = 9.479600E-01$); and HPCAL4 showed no difference among races (comparison between Caucasians vs. African Americans with $p = 3.880300E-03$, Caucasians vs. Asians, $p = 9.851800E-01$, and African Americans vs. Asians $p = 3.747200E-01$).

In the analysis between age groups, ATP10A had expression inversely proportional to age, that is, it showed greater expression in the age group from 21 to 40 years-old and minor expression in 61 to 80 years-old ($p = 3.420400E-01$), SAMD14 showed greater expression in the age group from 61 to 80 years-old ($p = 4.895400E-01$) and HPCAL4 showed no significant difference among age groups (comparison between 21-40 y vs. 41-60y with $p = 7.732200E-02$, 21-40y vs. 61-80y $p = 1.437890E-03$ and 41-60y vs. 61-80y $p = 7.543600E-01$).
Then, mutation analysis by presence of methylation was performed. No mutation was demonstrated for genes ATP10A (p < 1E-12) and HPCAL4 (p = 6.6634475149773E-12); as to gene SAMD14 (p = 5.028100E-02), only one patient, that is, 0.7% of the sample had a single nucleotide change mutation of the missense type (G397W).

Lastly, an enrichment analysis was carried out on the set of 100 genes with expression similar to that of ATP10A, HPCAL4, and SAMD14 to evaluate and understand the function expressed by each group and consequently of these three genes (Chart 1).

**DISCUSSION**

Testicular cancer (TC) is a clinical challenge, as it represents one of the most common neoplasms that affect the male sex, which includes testicular germ cell tumors (TGCT). TC is classified in two main groups: seminomas and non-seminomas, and, on top of that, it is known that TC emergence depends on genetic factors. Thus, many studies have been conducted to develop new tumoral molecular biomarkers with greater specificity and sensitivity. However, currently, there is a considerably small number of known biomarkers considering TC. Therefore, in this study, a bioinformatics analysis was carried out in search of biomarkers that found three genes, which were shown to be differentially expressed with statistical significance and associated with the aggressiveness of testicular neoplasm. The differentially expressed genes were: ATP10A, SAMD14, and HPCAL4, all sub-expressed in TC.

It is worth noting that there is a minimum availability of studies which sustain TC genetic base, therefore, there are few documented biomarkers for prognosis. The protein-coding genes ATP10A, SAMD14, and HPCAL4 found in this work are differentially expressed in CT and also demonstrate an associative relationship between their expression and survival rate. In this scenario, the relevancy of this study is justified by the fact these genes have not yet been assessed in the literature as molecular markers of TC.

Tumoral markers AFP, bHCG, and LDH are known to be important in the clinical management of TGCT. However, expression is relatively low in TGCT, with only 60% of TCs showing an increase in these molecules at initial diagnosis, which makes their clinical applicability low in specificity and falling short in the follow up and monitoring of advanced disease. In view of this, TGCT-specific biomarkers are needed, since they are essential to guide the choice of treatment, especially when the disease is in an advanced stage.

By analyzing the three genes identified in this study, it was observed that there is a predominantly greater gene expression in non-seminoma types. It is known that there is a differential expression depending on the histological subtype and that non-seminomatous germ cell tumors (NSGCT) are the most likely to cause metastases, which ends up impacting the quality of survival to the disease.

It is also known that TC is usually more common in men aged 15 to 34 years-old, which is in line with the findings of this study, as the greater expression was identified in the ATP10A and HPCAL4 genes between the ages of 21 to 40 years-old, responsible for greater expression in stages 1 and 2, respectively. However, there was a discrepancy regarding SAMD14 gene, which was found in greater expression in the 61 to 80 years age group, which could be associated to the advanced stage of the disease, since it presented greater expression in stage 3. These findings can guide new studies to confirm the correlation of these genes’ expression with a specific stage of the disease in order to validate biomarkers with high prognostic classification potential. Analyses have also shown the expression of these genes regarding race: ATP10A is more expressed in Asians, HPCAL4 in African Americans and Asians, and SAMD14 in Caucasians and
Chart 1. Representation of the functions associated with sets of genes with similar expression to ATP10A, HPCAL4, and SAMD14 in testicular cancer, identified through enrichment analysis using the methodology “Ora Sample Run” and “Pathway + KEGG”.

<table>
<thead>
<tr>
<th>Method</th>
<th>ATP10A</th>
<th>HPCAL4</th>
<th>SAMD14</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ora sample run</strong></td>
<td>Cilium or flagellum-dependent cell motility, sperm motility, cell-cell fusion, syncytium formation, fertilization, microtubule bundle formation, cellular process involved in reproduction in multicellular organism, microtubule-based movement, cilium storage</td>
<td>Cilium or flagellum-dependent cell motility, microtubule bundle formation, neuron migration, glutamate receptor signaling pathway, synaptic transmission, glutamatergic, neural nucleus development, microtubule polymerization or depolymerization, regulation of postsynaptic membrane neurotransmitter receptor 1, sperm motility, regulation of neurotransmitter receptor activity, organization of synapses, differentiation of photoreceptor cells, localization of proteins at the synapse, negative regulation of nervous system development, hindbrain development, regulation of transsynaptic signaling, adult behavior, development of the pancreas, regulation of synaptic structure or activity, chemical, postsynaptic synaptic transmission, regulation of microtubule-based process, differentiation of neurons of the central nervous system, cycle of synaptic vesicles, exocytosis regulated by calcium ions, organization of projection of neurons, microtubule-based movement, vesicle-mediated transport at the synapse, dendrite development, locomotor behavior, negative regulation of cell development, regulation of drug response, columnar/cuboidal epithelial cell differentiation, forebrain development, cognition, regulation of neurological system process, orientation of projection neurons, organization of the cilium, regulation of transporter activity, transport of neurotransmitters, development of axons, negative regulation of organization of cell projection, regulation of development of projection neurons, gliogenesis, location within the membrane, regulation of membrane potential, multicellular organism signaling, developmental maturation, positive regulation of ion transport, developmental growth involved in morphogenesis, vesicle localization, cell fate impairment, potassium ion transport, regulation of neurotransmitter levels, signal release, sensory system development, positive regulation of neurogenesis</td>
<td>Adhesion of II cells via molecular adhesion to the plasmatic membrane</td>
</tr>
<tr>
<td><strong>Pathway + KEGG</strong></td>
<td>-</td>
<td>Glutamatergic synapse</td>
<td>Transendothelial migration of leukocytes, hepatitis C, cell adhesion molecules, tight junction</td>
</tr>
</tbody>
</table>

Note: The chart lists only the functions with statistical significance ($p \leq 0.05$).

African Americans. It is worth mentioning that, according to literature, the incidence of TGCT varies among populations, being greater in individuals with European ancestry than in those with African ancestry.22

Regarding genes, ATP10A was identified as being responsible for producing a flippase-type protein which transports phosphatidylcholine from the external face to the internal face of the plasmatic membrane. Thus, according to the findings of the present study, the gene is differentially expressed in non-seminomatous tumors, given a hypomethylation of its promoter site, with no significant differences when comparing age, race, or...
stages. In this context, when carrying out an enrichment analysis to evaluate the gene pathways, it was observed that there is a large role in the cell-cell fusion pathway and syncytium formation. Deregulation of cell adhesion is known to play a role in the process of transformation, contributing to the metastatic processes\(^3\). Hence, it is possible that the hyper expression of AT10A gene can implicate in a minor adhesion among cells and, thus, in a greater ability of metastasis of neoplastic cells, validating the referenced findings.

It is known that SAMD14 was identified in mast cells from prostate cancer specimens as a mediator of intercellular communication and epithelial interaction in the microenvironment and that a favorable microenvironment, made from directional communication between cells and the extracellular matrix (ECM) is essential for the development of tumorigenesis and metastasis\(^3\). The results of the present study can be correlated with this statement, since, when performing the enrichment analysis, SAMD14 presented the function of adhesion of II-cells via molecular adhesion to the plasma membrane. Furthermore, patients with gastric cancer that present low expression and hypermethylation of SAMD14 promoter have a worse overall survival\(^3\), which can be related to the findings of this study, in which the subexpression of the TGCT gene in comparison to normal tissue was observed. Recent studies have shown that the hyper expression of ATP10A gene can implicate in a minor adhesion among cells and, thus, in a greater ability of metastasis of neoplastic cells, validating the referenced findings.

CONCLUSION
The three protein-coding genes are differentially expressed in testicular cancer (TC) in the non-seminoma histological type and have shown significant association between their expression and survival and, consequently, with the prognosis of patients, being potential biomarkers for TC. However, more detailed studies are needed to validate these genes in TC. 

CONTRIBUTIONS
All the authors have substantially contributed to the study design, acquisition, analysis and interpretation of the data, wording, and critical review. They approved the final version for publication.

DECLARATION OF CONFLICT OF INTERESTS
There is no conflict of interest to declare.

FUNDING SOURCES
None.

REFERENCES
12. Conselho Nacional de Saúde (BR). Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais cujos procedimentos metodológicos envolvem a utilização
Genes Associated with Testicular Cancer Aggressiveness


