Regulation network analysis of testicular seminoma at various stages of progression

J.-J. Sha, Y.-H. Dong, D.-M. Liu, J.-J. Bo, Y.-R. Huang, Z. Li and P. Ping

Department of Urology, Renji Hospital,
Shanghai Jiaotong University School of Medicine,
Shanghai, China

Corresponding author: P. Ping
E-mail: pingpingwinter@hotmail.com

Received August 2, 2012
Accepted November 30, 2012
Published March 11, 2013
DOI http://dx.doi.org/10.4238/2013.March.11.11

ABSTRACT. Testicular seminoma has become the most common solid malignancy in young men, especially in the 20s group. We obtained the gene expression profile of human testicular seminoma cells from NCBI, identified the differentially expressed genes of testicular seminoma cells of different stages, and constructed the regulation networks of different stages of testicular seminoma using bioinformatics methodology. Forty differentially expressed genes of testicular seminoma cells of different stages were identified. These genes and pathways are apparently involved in the progression of testicular seminoma.

Key words: Testicular seminoma; Differentially expressed genes; Regulation network
INTRODUCTION

Testicular germ cell tumors (TGCT) develop from pre-malignant intratubular germ cell neoplasias and can be histologically classified into seminomas and non-seminomas (Masters and Koberle, 2003). Testicular germ cell tumors account for only 1% of all cancers in men; however, they are the most common malignancy in the 15- to 35-year-old age group (Cooper et al., 2008). The incidence of testicular germ cell tumor increased from 4.1 cases per 100,000 men in the general population in 1975 to 6.6 cases per 100,000 in 2005 (Walsh et al., 2009). A seminoma consists of sheets of cells with clear cytoplasm and is relatively homogeneous. According to the American National Cancer Data Base, seminomas account for 56% of all testicular germ cell tumors (Steele et al., 1999), and testicular seminoma has become the most common solid malignancy in young men, especially in their twenties. However, the pathogenesis of testicular seminoma remains undetermined although some epidemiological data have implicated both environmental and genetic factors (Chevalier et al., 2010).

In recent years, various attempts have been made to understand the molecular mechanisms of testicular seminoma. To date, several genes and chromosomal aberrations have been identified that are considered diagnostic markers for testicular germ cell tumors. Somatic mutations of the KIT gene occur in approximately 5% of all testicular germ cell tumors. Of these, Coffey et al. (2008) demonstrated that only seminomas contain activating mutations of the KIT gene. Numerous data exist for the expression of different stem cell markers in seminoma, such as OCT4, NANOG and PIWIL1 (Qiao et al., 2002; Honecker et al., 2004; Ezeh et al., 2005). Additional germ cell marker genes for seminoma such as MCFD2, BOB1 and PROM1 have been described by Gashaw et al. (2007). The most consistent chromosomal aberration in testicular germ line tumors is the presence of an isochromosome of the short arm of chromosome 12, as previously reviewed (Looijenga et al., 2003). These investigations have led to new insights into the pathogenesis and predisposition of this tumor entity.

Testicular cancer is categorized as being in one of three stages, but the size of the tumor in the testis is irrelevant to staging. In broad terms, testicular cancer is staged as follows: in stage 1, the cancer remains localized to the testis; in stage 2, the cancer involves the testis and metastasis to retroperitoneal and/or para-aortic lymph nodes; and in stage 3, the cancer involves the testis and metastasis beyond the retroperitoneal and para-aortic lymph nodes.

In this present study, we analyzed the regulation network of different stages of testicular seminoma using bioinformatics methods. We aimed to identify the differentially expressed genes and regulatory relationships relevant to tumorigenesis, and to gain more insight into the mechanisms of testicular seminoma.

MATERIAL AND METHODS

Affymetrix microarray data

The transcription profile of GSE8607 was downloaded from GEO, a public functional genomics data repository, which is based on the Affymetrix GPL8300 platform data (Affymetrix Human Genome U95 Version 2 Array). A total of 40 pure seminoma samples, including 22 samples of tumor stage 1 (pT1), 14 samples of tumor stage 2 (pT2) and 4 of tumor stage 3 (pT3) as well as three normal testicular samples, were used for analyses.
Regulation data

TRANSFAC (http://www.gene-regulation.com/pub/databases.html) has become the basis for a complex platform for the description and analysis of gene regulatory events and networks. The TRANSFAC database contains data on transcription factors and their experimentally proven binding sites and regulated genes (Wingender, 2008).

TRED (Transcriptional Regulatory Element Database) (http://rulai.cshl.edu/TRED/) has been built to increase the needs of an integrated repository for both cis- and trans- regulatory elements in mammals (Jiang et al., 2007). TRED did the curation for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation currently focuses on target genes of 36 cancer-related TF families.

Combining the two regulation datasets, a total of 7234 regulatory relationships between 376 TFs and 2653 target genes were collected.

Pathway data

Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). The ‘pathway’ database records networks of molecular interactions in the cells and of their variants specific to particular organisms.

Differentially expressed gene (DEG) analysis

We preprocessed the data of profile GSE8607, using the Affy package in R (Team, 2011). After deleting the unqualified probes, the classic \( t \)-test method was used to identify the differentially expressed genes between the three tumor stages (PT1, PT2, and PT3) and normal testicular samples. To circumvent the multi-test problem which could induce too many false positive results, the BH method (Benjamini, 1995) was used to adjust the raw P-values into false discovery rate (FDR). The DEGs only with a fold change greater than 1.5 and FDR less than 0.05 were selected.

Co-expression analysis

For demonstrating the potential regulatory relationship, Pearson’s correlation coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The regulatory relationships whose absolute PCC was greater than 0.65 were considered significant.

Pathway-enrichment analysis

The functional annotation and analysis of large lists of genes in our network was facilitated using DAVID (Database for Annotation, Visualization and Integrated Discovery) (Huang da et al., 2009) bioinformatics resources. DAVID consists of an integrated biological knowledge base and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists.
Regulation network construction

Using the regulation data that were collected from the TRANSFAC database and the TRED database, we matched the relationships between differentially expressed TFs and their differentially expressed target genes. We built the regulatory networks using Cytoscape (Shannon et al., 2003).

RESULTS

DEG selection

To determine the differentially expressed genes between the three stages of tumor (PT1, PT2, and PT3) and normal testicular samples, we used the publicly available microarray dataset GSE8607 from GEO. A total of 40 seminoma samples, including 22 samples of tumor stage 1 (pT1), 14 samples of tumor stage 2 (pT2) and 4 of tumor stage 3 (pT3), as well as three normal testicular samples, were used for microarray analyses. The differentially expressed genes only with a fold change greater than 1.5 and FDR less than 0.05 were selected. We selected 2411 genes as DEGs between PT1 and normal testicular samples (1055 high expression and 1356 low expression), 2404 DEGs between PT2 and normal testicular samples (1080 high expression and 1324 low expression), and 2635 DEGs between PT3 and normal testicular samples (1328 high expression and 1307 low expression) (Table 1). There were 2156 overlapping genes between PT1 and PT2, 1654 overlapping genes between PT1 and PT3, and 1628 overlapping genes between PT2 and PT3. In total, 1548 overlapping genes were selected as DEGs from GSE8607 (Figure 1).

<table>
<thead>
<tr>
<th>Stage</th>
<th>PT1</th>
<th>PT2</th>
<th>PT3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hExp</td>
<td>lExp</td>
<td>hExp</td>
</tr>
<tr>
<td>PT1</td>
<td>0</td>
<td>-</td>
<td>921</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>PT2</td>
<td>921</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1235</td>
<td>-</td>
</tr>
<tr>
<td>PT3</td>
<td>584</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1070</td>
<td>0</td>
<td>1038</td>
</tr>
</tbody>
</table>

Table 1. Number of differentially expressed genes in each stage.

hExp = high expression; lExp = low expression; PT1 = tumor stage 1; PT2 = tumor stage 2; PT3 = tumor stage 3.

Regulation network in 3 stages of testicular seminoma

The regulation network was built according to the correlation coefficient between TFs and their target genes. We obtained three networks of each stage, respectively. In PT1 network, there were 84 regulatory relationships between 92 nodes, and the hub nodes were JUN, CEBPA, NFkB1, and SP1. In PT2 network, there were 116 regulatory relationships between 128 nodes, and the hub nodes were E2F1, JUN, CREB1, SP1, TFAP2A, and CEBPA. In PT3 network, there were 121 regulatory relationships between 145 nodes, and the hub nodes were CEBPA, JUN, ETS1, TP 53, and SP1. As can be seen, the overlapping TFs of the 3 networks were SP1, JUN and CEBPA, suggesting that these three TFs play important roles in testicular seminoma.
Specific regulation network construction in 3 stages

We measured the similarity of every two networks by comparing the overlapping degree of nodes in the two networks (Hong et al., 2011). The similarity of PT1 and PT2 regulation networks was 0.2865, the similarity of PT1 and PT3 was 0.1850, and the similarity of PT2 and PT3 was 0.1717. The similarity of the first 2 networks was higher, while they both showed low similarity to PT3. By comparing the network of each stage, we built a common network between PT1 and PT2, which contained 27 regulatory relationships between 38 nodes (Figure 2). The hub TFs of this network included CEBPA, ESR1 and JUN. However, there were only 7 common regulatory relationships between PT1 and PT3, and only 5 regulatory relationships between PT2 and PT3, and these relationships were unable to construct networks.

We also built a specific network of each stage, respectively. The PT1 specific network had 32 regulatory relationships between 43 nodes (Figure 3), and its hub nodes were NFKB1, JUN and CEBPB. In the PT2 specific network, there were 63 regulatory relationships between 85 nodes (Figure 4), and the hub nodes were CREB1, E2F1, TFAP2A, and SP1. We obtained 96 regulatory relationships between 118 nodes in the PT3 specific network (Figure 5), and the hub nodes were CEBPA, ETS1, JUN, SP1, and TP53. Compared to the other two networks, some new relationships related to CEBPA, TP53, ETS1, and SP1 were found in this network. Interestingly, the TF TP53 and ETV4 were only found in the PT3 specific network, suggesting that TP53 plays a critical role in the development of tumor stage 3.
Figure 2. Network between PT1 and PT2. By comparing the network of each stage, we built a common network between PT1 and PT2, which contains 27 regulatory relationships between 38 nodes.

Figure 3. PT1-specific network. The PT1-specific network has 32 regulatory relationships between 43 nodes.
Figure 4. PT2-specific network. In PT2-specific network there are 63 regulatory relationships between 85 nodes.

Figure 5. PT3-specific network. In PT3-specific network there are 96 regulatory relationships between 118 nodes.
Pathway-enrichment analysis of the specific regulation network

To gain further insight into the function of genes scored in each specific regulation network, we used DAVID (Huang da et al., 2009) to identify the significant pathways related to the DEGs. The enriched pathways in the PT1 specific regulation network included “pathways in cancer”, “NOD-like receptor signaling pathway” and “ErbB signaling pathway”. The enriched pathways in the PT2 specific regulation network included “B cell receptor signaling pathway”, “MAPK signaling pathway” and “TGF-beta signaling pathway”. The enriched pathways in the PT3 specific regulation network included “cell cycle”, “MAPK signaling pathway”, “Toll-like receptor signaling pathway”, “Wnt signaling pathway”, “p53 signaling pathway,” and so on (Table 2).

<table>
<thead>
<tr>
<th>Class</th>
<th>Term</th>
<th>Description</th>
<th>P</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT1 special</td>
<td>hsa05200</td>
<td>Pathways in cancer</td>
<td>2.03E-05</td>
<td>0.019978</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05212</td>
<td>Pancreatic cancer</td>
<td>4.37E-04</td>
<td>0.428525</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa04621</td>
<td>NOD-like receptor signaling pathway</td>
<td>0.003675</td>
<td>3.556873</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05211</td>
<td>Renal cell carcinoma</td>
<td>0.005178</td>
<td>4.977659</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05220</td>
<td>Chronic myeloid leukemia</td>
<td>0.00628</td>
<td>6.008108</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05200</td>
<td>Pathways in cancer</td>
<td>3.21E-07</td>
<td>3.45E-04</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05215</td>
<td>Prostate cancer</td>
<td>1.00E-04</td>
<td>0.107947</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05221</td>
<td>Acute myeloid leukemia</td>
<td>1.25E-04</td>
<td>0.134687</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa04662</td>
<td>B cell receptor signaling pathway</td>
<td>4.23E-04</td>
<td>0.454303</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa04010</td>
<td>MAPK signaling pathway</td>
<td>0.001768</td>
<td>1.885431</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05200</td>
<td>Pathways in cancer</td>
<td>4.85E-13</td>
<td>5.50E-10</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05212</td>
<td>Pancreatic cancer</td>
<td>1.05E-05</td>
<td>0.01189</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05220</td>
<td>Chronic myeloid leukemia</td>
<td>1.43E-05</td>
<td>0.016158</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa05221</td>
<td>Acute myeloid leukemia</td>
<td>2.24E-05</td>
<td>0.02542</td>
</tr>
<tr>
<td>PT1 special</td>
<td>hsa04110</td>
<td>Cell cycle</td>
<td>9.68E-05</td>
<td>0.109629</td>
</tr>
</tbody>
</table>

Table 2. Database for annotation, visualization and integrated discovery-enrichment analysis (only list top 5 of each stage).

PT1 = tumor stage 1; PT2 = tumor stage 2; PT3 = tumor stage 3; FDR = discovery rate.

DISCUSSION

Testicular seminoma as a whole is uncommon, accounting for approximately 1% of all malignancies in men but shows increased incidence in younger men. The incidence of testicular cancer has doubled in the past 40 years (Horwich et al., 2006). This increase suggests that critical changes in environmental factors contribute to the development of these tumors. The disease can progress from stage 1, localized cancer in the testicle, to stage 3, where the cancer has spread to distant parts of the body in just a few months. In recent years, many genes have been described as being associated with the pathogenesis of testicular seminoma in the early tumor stage pT1 (Okada et al., 2003). However, not so many studies have focused on the gene expression changes of different stages (pT1 to pT3) of testicular seminoma. In the present study, we analyzed the regulation network of different stages of testicular seminoma using bioinformatics methods. We aimed to identify the differentially expressed genes and regulatory relationships relevant to tumorigenesis, and to gain more insight into the mechanisms of testicular seminoma development.

According to our regulation network results, we could find that many TFs closely related to testicular seminoma were linked by our method. Among them, SP1, JUN and CEBPA...
appeared in all the 3 networks, suggesting that these genes are conserved genes in the development of testicular seminoma. Some of them have proved to be related to testicular cancer in previous studies. TP53 and ETV4 were only found in the PT3 specific network, suggesting that TP53 and ETV4 play a critical role in the development of tumor stage 3. This interactive relationship could be discussed as follows based on previous reports.

SP1, a member of the Sp protein family, is overexpressed in a variety of cancers including gastric, colorectal, pancreatic, testicular, and lung cancers. Previous studies showed that Sp proteins mediate a number of genes involved in cell proliferation, survival and angiogenesis. SP1 is a zinc finger protein and is believed to play an important role in the transcription of many genes involved in cancer that have an abundance of GC boxes in their promoter region (Hocker et al., 1998; Li et al., 2004; Wang et al., 2007). The zinc finger region of SP1 transcription factors binds to GC or GT-box elements present in the promoters of a number of male germ cell target genes that are developmentally expressed during spermatogenesis (Thomas et al., 2007). Early on, SP1 was proven as a marker for testicular cancer by Lange et al. (1980).

JUN encodes a protein that is highly similar to the viral protein, which interacts directly with specific target DNA sequences to regulate gene expression. A recent study suggests that the cyclic change of the JUN protein levels is significant in the proliferation and apoptosis of glandular epithelial cells. The persistent stromal expression of JUN protein may prevent stromal cells from entering apoptosis during the late secretory phase (Udou et al., 2004). A recent report has linked JUN to other genes including SMAD4, a TGF-β signaling pathway component, implicating the association of JUN with aggressive neoplasms (Atfi et al., 1997).

CEBPA (CCAAT enhancer binding protein alpha) is a bZIP transcription factor that can bind as a homodimer to certain promoters and enhancers. It belongs to the CEBP family and plays a crucial role in controlling the differentiation and proliferation of various cell types. Although mutated CEBPA is primarily observed in AML, it has been found in rare cases to be mutated in myelodysplasia syndromes, lung tumors and prostate tumors. Downregulation of CEBPA has also been observed in blast crisis chronic myeloid leukemia, lung cancer, breast cancer, and liver cancer.

From the three specific regulation networks from our results, we could find that TP53 and ETV4 only exist in the specific regulation network of stage 3, suggesting that these two TFs play a critical role in the progression of testicular seminoma.

TP53 (also known as p53) is a tumor suppressor gene that encodes the tumor suppressor protein. Tumor suppressor p53 is a DNA-binding protein that responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. p53 mutation is the most commonly described molecular event associated with tumorigenesis, as shown in various studies. It can also be frequently indicated by loss-of-heterozygosity which can be picked up by genotyping analysis (Mao et al., 2007). Bártková et al. (1991) have demonstrated that altered expression of the p53 protein is a unifying feature of the majority of invasive male germ-cell tumors and that the change resulting in high levels of p53 appears to be a relatively early step in human testicular cancer pathogenesis.

However, there are also some controversial opinions on the role of p53 in the development of testis cancer. Some authors hold the opinion that the dysfunction of p53 may not play an important role in the development of germ cell cancers of the testis (Peng et al., 1993; Lutzker, 1998). Our results suggest that TP53 plays a role in the development of testicular seminoma.

ETV4 (ETS variant gene 4) is known to regulate the expression of certain matrix me-
talloproteinases (Kaya et al., 1996; Hida et al., 1997). ETV4 has been shown to transactivate multiple MMP genes and to play an important role in tumor invasion, in breast and gastric cancers (Higashino et al., 1995; Yamamoto et al., 2004). Our results are in line with a previous study by Gashaw et al., (2007) who confirmed that ETV4 represents a candidate gene that may be involved in the tumor progression of seminoma by an activation of MMP2 and ADAM15, leading to degradation of the extracellular matrix and migration of seminoma cells.

A deeper understanding of transcription factors and their regulated genes remain an area of intense research activity for the future. Our regulation network was useful in investigating the complex interacting mechanisms of transcription factors and their regulated genes in testicular seminoma. We also found some tumor stage-dependent genes related to different stages of testicular seminoma; besides, many pathways such as pathways in cancer, B cell receptor signaling pathway and MAPK signaling pathway were linked to testicular seminoma using our method. However, further experiments are still needed to confirm our findings.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES


Analysis of testicular seminoma


