IDENTIFYING BIOLOGICALLY MEANINGFUL GROUPS OF GENES IN THE ANDROGEN RECEPTOR NETWORK

Marina Popovscaia¹, Christian Baumgartner¹, Michael Handler¹ and Helmut Klocker²

¹ Institute of Electrical and Biomedical Engineering, UMIT- The Health and Life Sciences University, Hall in Tirol, Austria
² Department of Urology, Division of Experimental Urology, Innsbruck Medical University, Innsbruck, Austria

marina.popovscaia@umit.at

Abstract: Androgen receptor is a transcription factor that plays a crucial role in the development of prostate cancer. Androgen receptor cofactors and androgen response genes are tightly connected to the androgen receptor transcriptional activity. In our study we built and analyzed a network which consist only of these groups of genes. In the constructed network we could identify a number of gene modules which potentially can be used as new biomarkers candidates for prostate cancer.

Keywords: prostate cancer, androgen receptor network, androgen receptor cofactors, androgen response genes.

Introduction

Prostate cancer (PCa) is one of most frequent oncological diseases in men in Western industrialized countries. Up to date PCa is curable when detected at the early organ confined stages. The identification of PCa in early stages is therefore one of the top priorities of modern cancer research.

A fundamental role in the development of PCa, as well as in the growth, differentiation and maintenance of healthy prostate, play androgens - steroid hormones responsible for male reproductive function and behavior [1]. Effects of the androgens are mediated by the specific transcription factor - androgen receptor (AR). In its inactive form AR is located in the cytoplasm of the cell in complex with heat-shock proteins. Upon ligand (androgen) binding AR homodimerizes and translocates in the nucleus where it binds to the specific sequences on the DNA known as androgen response elements (AREs) [1]. It may also recruit cofactors – proteins that either enhance (i.e. coactivators) or reduce (i.e. corepressors) AR transactivation (see Fig. 1).

The aim of our work was to explore in detail the function of the complex "AR, AR cofactors and ARE containing genes" and to investigate the possibility of using one gene or groups of genes as biomarker(s) for PCa detection. For this purpose we constructed a network, further addressed as "AR-network", based specifically on these three groups of genes. Using various techniques we detected a number of modules in the AR-network, ranked them and investigated their ability to differentiate between cancer and benign tissue samples.

Methods

Construction of AR-network. For our study we initially chose 1913 genes: AR, 162 genes known to be AR cofactors and 1752 genes reported to be androgen controlled (containing AREs). The AR-network was inferred from microarray datasets GDS3111 and GDS2782 for LNCaP cell lines treated with dihydrotestosterone (DHT), downloaded from GEO (Gene Expression Omnibus) database. Both datasets had been generated using the platform GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array). After normalization and data preprocessing a total number of 1692 genes were chosen for the AR-network construction. The C3NET method [3] which is based on mutual information (1) was used for inferring the network

\[ I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)}. \] (1)

Identification of gene modules. Groups of genes were identified using hierarchical clustering.

Validation of results. In order to find out which of the gene groups might be biologically meaningful, we, first, ranked them based on the enrichment analysis of gene ontology terms. Second, we investigated the ability of highly ranked gene clusters to function as biomarker candidates and determine their discriminatory ability when comparing PCa versus non-PCa cases. For this step we used a SVM (support vector machine) classifier. Validation was performed on the GDS1439 and GSE32982 datasets containing expression data of benign, clinically localized and metastatic
prostate cancer tumors/tissues. Datasets were downloaded from GEO.
All computational work was performed in R-project statistical programming language.

Results

Upon construction of the AR-network using mutual information we were able to identify more than 60 gene modules of various dimensions. After ranking, we have chosen 18 modules for further investigation. One third of them could discriminate between PCa and non-PCa cases with various accuracy levels (Tab. 1).

As an example, we present here module No.6 (Fig. 2) which showed the highest accuracy (74.2%) of predicting PCa and non-PCa cases in the validation sets. AR cofactors are marked in green and ARE genes – in blue. One can immediately see that in this AR-network module AR cofactors tend to be the hubs centers.

<table>
<thead>
<tr>
<th>Modules</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68.7%</td>
</tr>
<tr>
<td>2</td>
<td>72.6%</td>
</tr>
<tr>
<td>3</td>
<td>61.5%</td>
</tr>
<tr>
<td>4</td>
<td>66.1%</td>
</tr>
<tr>
<td>5</td>
<td>63.7%</td>
</tr>
<tr>
<td>6</td>
<td>74.2%</td>
</tr>
</tbody>
</table>

Table 1: Validation of the gene modules.

Enrichment analysis was performed to significantly (p-value < 0.05) associate Gene Ontology (GO) annotations to the genes in module No.6 using Genecodis web–based tool [4, 5]. The following GO terms were enriched in our group of genes: protein binding (Molecular function (MF)), pathways in cancer (KEGG pathways (KEGG)), focal adhesion (KEGG), small cell lung cancer (KEGG), p53 signaling pathway (KEGG), melanoma (KEGG), cell differentiation (biological (BP)), axon guidance (BP), intracellular protein transport (BP), cell migration (BP), protein export from nucleus (BP) and others (Fig. 3).

Discussion

Using our proposed discovery strategy we were able to construct and analyze the AR-network with respect to AR cofactors and ARE containing genes. Our aim was to identify single genes or biologically meaningful groups of genes which might be used as new biomarker candidates for PCa prediction. We could define a number of gene modules which showed significant accuracy of distinguishing between PCa and non-PCa cases in the validation sets. These promising results together with GO associations of the genes will help us to construct hypotheses for further validation of the results in the wet-laboratory.

Acknowledgement

This work was supported by the Oncotyrol Center for Personalized Cancer Medicine within the scope of the Austrian COMET program (project 1.3.1).

Bibliography


