Gene expression profiles in prostate cancer: Identification of candidate non-invasive diagnostic markers

L. Mengual\(^a\), E. Ars\(^b\), J.J. Lozano\(^c\), M. Burset\(^a\), L. Izquierdo\(^a\), M. Ingelmo-Torres\(^a\), J.M. Gaya\(^b\), F. Algaba\(^b\), H. Villavicencio\(^b\), M.J. Ribal\(^a\), A. Alcaraz\(^a\)

\(^a\) Laboratorio y Servicio de Urología, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
\(^b\) Laboratorio de Biología Molecular, Servicio de Anatomía Patológica y Servicio de Urología, Fundació Puigvert, Barcelona, Spain
\(^c\) CIBERehd, Plataforma de Bioinformática, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Hospital Clinic, Barcelona, Spain

Received 11 June 2013; accepted 17 July 2013
Available online 13 January 2014

**KEYWORDS**
Prostate cancer; DNA microarrays; Gene expression; Molecular markers; RT-PCR quantitative

**Abstract**

**Objective:** To analyze gene expression profiles of prostate cancer (PCa) with the aim of determining the relevant differentially expressed genes and subsequently ascertaining whether this differential expression is maintained in post-prostatic massage (PPM) urine samples.

**Materials and methods:** Forty-six tissue specimens (36 from PCa patients and 10 controls) and 158 urine PPM-urines (113 from PCa patients and 45 controls) were collected between December 2003 and May 2007. DNA microarrays were used to identify genes differentially expressed between tumor and control samples. Ten genes were technically validated in the same tissue samples by quantitative RT-PCR (RT-qPCR). Forty-two selected differentially expressed genes were validated in an independent set of PPM-urines by qRT-PCR.

**Results:** Multidimensional scaling plot according to the expression of all the microarray genes showed a clear distinction between control and tumor samples. A total of 1047 differentially expressed genes (FDR \(\leq 0.1\)) were identified between both groups of samples. We found a high correlation in the comparison of microarray and RT-qPCR gene expression levels (\(r = 0.928\), \(p < 0.001\)). Thirteen genes maintained the same fold change direction when analyzed in PPM-urine samples and in four of them (HOXC6, PCA3, PDK4 and TMPRSS2-ERG), these differences were statistically significant (\(p < 0.05\)).

**Conclusion:** The analysis of PCa by DNA microarrays provides new putative mRNA markers for PCa diagnosis that, with caution, can be extrapolated to PPM-urines.

\(© \) 2013 AEU. Published by Elsevier España, S.L. All rights reserved.
PALABRAS CLAVE
Cáncer de próstata; Microarrays de ADN; Expresión génica; Marcadores moleculares; RT-PCR cuantitativa

Perfil de expresión génica en el cáncer de próstata: identificación de marcadores candidatos para el diagnóstico no invasivo

Resumen
Objetivo: Analizar los perfiles de expresión génica del cáncer de próstata (CaP) e identificar los genes diferencialmente expresados. Determinar si la expresión diferencial en tejido se mantiene en muestras de orina-posmasaje prostático (PMP).
Material y métodos: Un total de 46 muestras de tejido prostático (36 de pacientes con CaP y 10 controles) y 158 orinas-PMP (113 de pacientes con CaP y 45 controles) se recogieron entre diciembre de 2003 y mayo de 2007. Se utilizaron microarrays de ADN para identificar los genes diferencialmente expresados entre las muestras de tejido tumoriales y las controles. Diez genes fueron seleccionados para la validación técnica de los microarrays en las mismas muestras tisulares mediante PCR cuantitativa (RT-qPCR). Se seleccionaron 42 genes para ser validados en muestras de orina-PMP mediante RT-qPCR.
Resultados: El gráfico de escalado multidimensional mostró una clara separación entre las muestras de tejido tumoriales y las controles. Se han identificado 1.047 genes diferencialmente expresados (FDR ≤ 0,1) entre los 2 grupos. La correlación entre los datos de microarrays y RT-qPCR fue alta (r = 0,928, p < 0,001). Trece genes mantuvieron el mismo sentido de expresión diferencial al ser analizados en orinas-PMP y 4 de ellos (HOXC6, PCA3, PDK4 y TMPRSS2-ERG) mostraron diferencias de expresión estadísticamente significativas entre orinas-PMP tumoriales y controles (p < 0,05).
Conclusión: Existe un perfil de expresión génica diferencial en el CaP. Aunque la extrapolación de la expresión génica obtenida en tejido prostático a orina-PMP se debe realizar con precaución, el análisis del tejido prostático permite la identificación de nuevos biomarcadores para diagnóstico no invasivo del CaP.
© 2013 AEU. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Introduction

Prostate cancer (PCa) is one of the most common male malignancies in our population. During the last 2 decades, the prostate specific antigen (PSA) has been widely used for screening, diagnosis, and follow-up of PCa. Routine use of PSA has been the object of continuous controversy due to its limited specificity, which is derived from the fact that elevated serum PSA levels are produced in a variety of non-neoplastic conditions, such as prostatitis and benign prostatic hyperplasia (BPH). On the contrary, a substantial number of men with PSA in the normal range (<4 ng/ml) have PCa. Currently, the diagnosis of PCa is made by means of prostate biopsy guided by transrectal ultrasound, which does not detect approximately 20–30% of the cases. Therefore, there is a need for additional specific markers for more accurate and early detection of PCa, which allows for a reduction in the number of unnecessary prostate biopsies. Furthermore, it would be of great interest that these markers could be determined non-invasively, like in urine or serum samples.

The development of new technologies of medium–high yield for the analysis of gene expression, such as DNA microarrays or new platforms based on quantitative real-time PCR (RT-qPCR, reverse transcription quantitative PCR), has had a significant effect on the discovery of new molecular markers for diagnosis and prediction of disease progression in several types of cancer, including PCa. In addition, the detection of biomarkers for PCa in non-invasive samples has already been described in the literature, but so far no biomarker has replaced the routine use of PSA as a screening and follow-up method of PCa.

In this study we used the DNA microarray technology to identify the differentially expressed genes in PCa. Furthermore, we analyzed a selection of these genes in samples of post-prostatic massage urine (PPM urines) of an independent cohort of patients in order to assess the correlation between gene expression in the tissue sample and the PPM urine and identify potential new molecular markers for the non-invasive diagnosis of PCa.

Materials and methods

Patients and samples

The samples used in this study were collected between December 2003 and May 2007 in the Fundación Puigvert and the Clinic Hospital of Barcelona. This study was approved by the ethics committee and all patients and controls included in it were duly informed of its objectives before being included and they signed the informed consent sheet.

In order to identify the differentially expressed genes in PCa, 36 PCa tissue samples of different tumor stages and Gleason scores were collected and 10 samples of prostate tissue from patients with BPH without evidence of malignancy in the prostate (supplementary material). The tumor samples were grouped into 3 categories: (1) pT2G7 samples (15 pT2G7); (2) pT3-4 samples, independent of Gleason score (5 T3G7, 1 T3G8, 7 T3G9 and 1 T4G7); and (3) Gleason samples >8, independent of tumor stage (7 T3G9).
In order to correlate the gene expression between the tumor and the PPM urine samples, the first portion (30–50 ml) of urine was collected after prostate massage of 125 patients with PCa and 55 controls. Of the total urines collected, we excluded 12 of PCa patients (10%) and 10 of control individuals (18%) because they did not have enough amount of RNA from prostate cells and/or the RNA obtained was of very poor quality (value of quantification of the cycle [Cq] KLK3 > 31 and/or B2M > 26 prior to the pre-amplification of target genes, see below). The clinical characteristics of the 113 patients with PCa and the 45 controls finally included are shown in Table 1.

The urine samples were processed as previously described. The obtained cell pellet was homogenized in TRIzol (Invitrogen, Carlsbad, CA) and frozen at –80°C until processed.

Preparation of ribonucleic acid

The tissue samples were mechanically homogenized in TRIzol (Invitrogen, Carlsbad, CA) and subsequently extraction of ribonucleic acid (RNA) was performed following the manufacturer’s instructions. The RNA obtained was purified according to the ‘‘Clean-up protocol’’ of the RNA extraction kit RNeasy (Qiagen, Valencia, TX). The RNA of the samples of PPM urine was obtained following the supplier’s recommendations (Invitrogen, Carlsbad, CA). The RNA obtained was quantified in a spectrophotometer.

Processing of the microarray and data analysis

Each category of PCa to be analyzed (pT2G7, pT3-4 independent of Gleason, Gleason score > 8 independent of stage and controls) includes 2 or 3 groups of samples containing 3–5 individual samples, with an equimolar amount of each individual sample in the group. Thus, 11 groups of samples were analyzed: 3 of pT2G7; 3 of pT3-4, independent of Gleason; 2 of Gleason 8, independent of tumor stage, and 3 of controls.

In order to determine the gene expression profiles of the 11 groups of samples, DNA microarrays (Affymetrix Human Genome U133 Plus 2.0) were used, which were processed and analyzed as described in the supplementary material. Genes with a corrected p value (false discovery rate [FDR]) ≤ 0.1 were considered differentially expressed.

Reverse transcription and quantitative polymerase chain reaction

In order to technically validate the gene expression data from microarrays, 10 genes from the microarrays (AMACR, ERG, FOLH1, HPN, KLK2, NKX3.1, ORM2, PCA3, PSA, KLK4) that were analyzed by RT-qPCR were randomly selected in the same RNA samples used for microarray hybridization (supplementary material).

In order to correlate the gene expression between prostate tumors and PPM-urines, 42 target genes were selected to be analyzed in 158 independent urine samples of patients with PCa and controls. These genes were selected based on a p value < 0.05 and/or level of expression change ≥ 1.5 (fold change [FC]) in the microarray experiments and some genes described in the literature were also added as markers of PCa in PPM urine (additional material).

cDNAs were synthesized from 100 ng in PPM urine samples using the High-Capacity cDNA Archive Kit (Applied Biosystems, Foster City, USA, henceforth referred to as AB), following the instructions of the manufacturer, except that the final volume of the reaction was 50 µL. In order to discard those PPM urines that do not contain enough prostate cells,


Table 1  Clinical and pathological features of the patients (n = 113) and controls (n = 45) of which PPM urine was obtained.

<table>
<thead>
<tr>
<th>Patients with PCa</th>
<th>Average ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>62.6 ± 0.6</td>
</tr>
<tr>
<td>Glandular weight (g)</td>
<td>39.9 ± 2.5</td>
</tr>
<tr>
<td>Tumor (%)</td>
<td>11.6 ± 1.3</td>
</tr>
<tr>
<td>PSA (ng/ml)</td>
<td>7.7 ± 0.3</td>
</tr>
<tr>
<td>Levels</td>
<td>No. patients (%)</td>
</tr>
<tr>
<td>PSA (ng/ml)b</td>
<td></td>
</tr>
<tr>
<td>0–4</td>
<td>7 (6)</td>
</tr>
<tr>
<td>4–10</td>
<td>80 (74)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>22 (20)</td>
</tr>
<tr>
<td>Gleasonb</td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>7 (7)</td>
</tr>
<tr>
<td>&gt;7</td>
<td>78 (72)</td>
</tr>
<tr>
<td>≥8</td>
<td>23 (21)</td>
</tr>
<tr>
<td>Stagec</td>
<td></td>
</tr>
<tr>
<td>pT2a</td>
<td>31 (31)</td>
</tr>
<tr>
<td>pT2b</td>
<td>1 (1)</td>
</tr>
<tr>
<td>pT2c</td>
<td>55 (54)</td>
</tr>
<tr>
<td>pT3a</td>
<td>12 (2)</td>
</tr>
<tr>
<td>pT3b</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Treatment</td>
<td>103 (91)</td>
</tr>
<tr>
<td>Control patients</td>
<td></td>
</tr>
<tr>
<td>Average ± SD</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>60.1 ± 14.8</td>
</tr>
<tr>
<td>Average PSAd</td>
<td>3.5 ± 0.6</td>
</tr>
<tr>
<td>No. controls (%)</td>
<td>29 (64)</td>
</tr>
<tr>
<td>BPH</td>
<td>11 (24)</td>
</tr>
<tr>
<td>Prostatitis</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Healthy</td>
<td>2 (4)</td>
</tr>
</tbody>
</table>

CRI0: cryotherapy; SD: standard deviation; RP: radical prostatectomy.

b Data available for 109 patients.

c Data available for 108 patients.

d Data available for 101 patients.

d Data available for the patients with BPH.
or that do not meet minimum quality criteria, prior to the pre-amplified cDNA, we individually analyzed by means of quantitative PCR KLK3 (prostate content marker) and B2MG (RNA quality marker) using the specific trials for these genes (AB) (Table 15 of the additional material) and standard PCR conditions described by the supplier. We excluded those samples with Cq values of KLK3 > 31 and/or B2M > 26.

For the valid samples (n = 158), the cDNA was multiplex pre-amplified with 48 genes (Table 15 of the additional material) and then the 48 genes were amplified using TaqMan Array (TA) in an ABI PRISM 9700 HT (AB).

Data normalization was performed with the B2MG endogenous control in the case of the tissues and with the average of the Cq of KLK2 and KLK3 in the case of PPM urines. The differences in expression levels were analyzed using Student’s ‘t’-test. Significance was defined as p values < 0.05. Pearson’s correlation for the comparison of microarray data and RT-qPCR (additional material) was used.

Results

Identification of differentially expressed genes in prostate cancer

Fig. 1 shows that there is a clear difference between the gene expression patterns of the samples of tumor and control prostate tissue.

Comparison of gene expressions between tumor and control tissue allowed us to identify 1047 differentially expressed genes (FDR ≤ 0.1) between both groups; 375 and 672 overexpressed and underexpressed genes in PCa, respectively (Table 25 of the additional material).

Technical validation of microarray results

A high correlation of gene expression was observed in the comparison of the data of microarrays and RT-qPCR for the 10 selected genes (r = 0.942; p < 0.001) (Fig. 2).

Correlation of gene expression between tumors and post prostatic massage urines

Of the 42 target genes analyzed by RT-qPCR in an independent series of PPM urines, 13 maintained the same sense of differential expression as tissue samples, and in 4 of them (HOXC6, PCA3, PDK4 and TMPRSS2-ERG) the differences in expression between tumor and control PPM urines were statistically significant (p < 0.05) (Fig. 3).

It should be noted that the fusion of ERG and TMPRSS2 genes could be evaluated in PPM urine samples, but the microarrays used do not contain the probe used for the TMPRSS2-ERG fusion gene. However, there is evidence that the overexpression observed for the ERG oncogene in PCa is due, in over 90% of the samples studied, to its fusion with TMPRSS2, suggesting that the fusion is the major cause of its overexpression.24 That is why, the values of ERG expression have been considered in microarrays for comparison with the TMPRSS2-ERG fusion gene in PPM urine.

Figure 1 Patterns of gene expression in PCa. (A) Multidimensional scaling graph of tumor samples (red) and controls (blue) based on the intensity values of the 18,960 genes contained in the microarray. The distances between the samples in the 3D graph are proportional to their differences in the mRNA expression. (B) Heat map showing the 50 most differentially expressed genes in the microarrays between the groups of samples of tumors and controls (n = 11). Blue bar: control samples; red bar: tumor samples. The red pixels correspond to an increase in mRNA expression, while green pixels indicate a decrease in the levels of mRNA of the gene. The lines represent individual genes and the columns the groups of experimental samples.
they could be useful for developing a non-invasive diagnostic method for PCa.

In this paper, for the first phase of global screening of genes, prostate tissue has been used to identify these non-invasive diagnostic markers of PCAs. This was because a high quantity and quality of RNA was required for such studies, and because we wanted to make sure that genes expressed in prostate cells were selected. On the contrary, in this first phase, we used the strategy of grouping samples with the same histopathologic features to dilute the interindividual variability, maintaining the expression patterns of the individual samples grouped.25,26

The microarray results show that there is a clear difference between the patterns of gene expression in tumor samples and controls, as it had already been described above.3-8 It is interesting to emphasize that many of the genes identified in our study are consistent with those described in previous studies,6,6,7 despite the different cohorts studied and the different technologies used.

In order to correlate the gene expression of the tissue samples and the PPM urine ones and identify potential new molecular markers for the non-invasive diagnosis of PCAs, a selected group of genes was analyzed in an independent cohort of samples of PPM urine. Most of the studies aimed at identifying transcripts in PPM urine for the diagnosis of PCa use as control group individuals with suspected PCa who underwent prostate biopsy with negative results. However, it is known that 20–30% of these individuals will end up being diagnosed with a PCa.3 To avoid this limitation, in the present study, we used as control group subjects clearly not affected with PCAs. It is also important to mention that the studies of quantification of gene expression using RT-qPCR in PPM urines usually use the PSA (KLK3) to normalize the expression of the target genes, since this gene is expressed exclusively and at similar levels in both normal prostate cells and in tumor cells.9,11 In this work, in order to minimize the possibility of error in the normalization, we used a second endogenous control, KLK2, given that its level of expression is highly correlated with KLK3 and it also presents exclusive expression in the prostate.27

Of the 42 target genes tested in our series of PPM urines, 4 of them (HOXC6, PCA3, PDK4 and TMPRSS2-ERG) maintained in the PPM urines the same sense of significant differential expression observed in tumors. One might think that the urine samples used do not contain enough prostate cells so that the gene expression profiles of the tumor can reproduce. However, we have discarded this hypothesis because we found, by means of PSA (KLK3) quantification, that all the PPM urines were obtained from sufficient amount of prostate cells. Additionally, we also found, by analyzing the amount of B2MG, that the RNA was of sufficient quality. The most likely cause of the limited correlation between tissue and PPM urine samples is that the high content of urethelial cells in PPM urine is masking the gene expression of prostate cells, making it possible to detect only those genes with a greater change in expression.

In the last decade, a trial for the detection of PCa based on the detection of the PCA3 gene in PPM urine has been developed and approved by the FDA. Although this trial has a relatively high diagnostic accuracy (AUC = 0.717),9 this test is not routinely used in clinical practice, since it does not reach the levels of sensitivity and specificity for the

**Figure 2** Correlation between the gene expression data obtained by means of the microarray technique and the RT-qPCR for 10 selected genes.

**Figure 3** Comparison of the gene expression obtained in prostate tissue by means of the microarray technique and in PPM urines by means of RT-qPCR. *FDR ≤ 0.1 in microarrays and p ≤ 0.05 in RT-qPCR.

**Discussion**

The development of a non-invasive test based on the detection of molecular markers in PPM urine capable of diagnosing PCa would be of great interest in the field of oncological urology. In this study, the analysis of the characteristic patterns of gene expression using DNA microarrays in 3 representative histological categories of prostate tumors has been carried out, making it possible to generate a list of potential new molecular markers. Subsequently, we have confirmed that some of the biomarkers identified in prostate tissue samples can be determined in PPM urine samples, so
diagnosis of PCa. Given the heterogeneity of PCa, it seems unlikely that a single marker can detect all types of PCa. Conversely, it is logical that the combination of several markers in a multiplex trial could provide an improvement in the diagnostic accuracy of PCa. In fact, different groups have analyzed different combinations of genes in PPM urine,\textsuperscript{10,12-14,28} in order to improve the diagnostic performance of PCA3 alone, improving its accuracy, but without reaching values that allow for clinical application of these trials. Interestingly, of the 4 genes identified in this study as potential noninvasive biomarkers of PCa, 2 of them, TMPRSS2-EGR and PCA3, had already been described by different groups.\textsuperscript{10,12,15,16} The other 2 identified genes, HOXC6 and PDK4, had not been previously analyzed in PPM urine of patients with PCa. HOXC6 is a member of the HOX gene family and its overexpression had been previously described in tissue samples of PCa.\textsuperscript{29} Contrarily, PDK4, which encodes a protein involved in the metabolism of carbohydrates,\textsuperscript{30} had not previously been linked to the PCa.

The generation of a list of biomarker candidates in prostate tumors, and the subsequent demonstration that these can be evaluated in PPM urine samples, represents a first step toward the development of a panel of urine markers for the non-invasive diagnosis of PCa. We are currently working on the generated list of differentially expressed genes in PCa in order to develop a panel of genes for detecting PCa in PPM urine.

Conclusions

Analysis by means of DNA microarrays of tumor prostate tissue provides biomarker candidates for the diagnosis of PCa in PPM urine. The genes identified in this study may be useful for the development of a non-invasive diagnostic method of PCa. Future studies will help to select genes of interest.

Funding

This work was funded by FINA BIOTECH Laboratories and the Spanish Association of Urology (FIU 2009 to AA).

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank all the patients included in this study who have voluntarily given their samples and the doctors, nurses and residents of the Fundació Puigvert and the Clinic Hospital of Barcelona for their collaboration in collecting urine samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2013.12.002.

References

Gene expression in prostate cancer


