Tissue microarray in prognostic studies on vulva cancer

Fons, G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://ueba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter

Validation of tissue microarray technology in vulvar cancer

Guus Fons
Jacobus van der Velden
Matthé PM Burger
Fiebo JW ten Kate

International Journal of Gynecological Pathology 2009;28:76-82
Summary

The aim of the study was to validate tissue micro array (TMA) for vulvar cancer by comparing immunohistochemical staining results of triplicate core biopsies on TMA with the results of full section analysis. The study material consisted of slides and selected tissue blocks from 40 patients with vulvar cancer. A TMA was constructed with 3 cores/case. Both the TMA and the slides were stained with the same antibodies against COX-2, Caspase-3, epidermal growth factor receptor, p16\textsuperscript{INK\textsubscript{4}}, Cyclin D1, and Ki67. For COX-2, 2 different scoring systems were applied. Agreement in the readings between TMA and slides was expressed in total agreement and $\kappa$. Expression patterns of antibodies can be reproduced on TMA with good reliability ($\kappa$ 0.68 to 0.75) for Ki-67, p16\textsuperscript{INK\textsubscript{4}}, COX-2, Cyclin D1, and epidermal growth factor receptor in comparison with whole slides. For Caspase-3 agreement is only slight with a $\kappa$ of 0.40. The majority of discordant cases for COX-2 and Ki67 were negative on slide and positive on TMA. For epidermal growth factor receptor and Caspase-3 an opposite pattern was found. For COX-2, the use of an alternative scoring system resulted in a decrease of $\kappa$ from 0.68 to 0.21. Agreement between results on TMA and slides depends on the distribution of the protein in the cancer tissue and on the scoring system used.
Introduction

Squamous cell cancer of the vulva is a rare disease that mainly affects elderly women. Tumor diameter, lymph node involvement and vascular space involvement are the most important prognostic factors for disease-specific survival. Nevertheless, a substantial number of cancer-related deaths occur in patients without these risk factors. Knowledge of the biological characteristics of a specific vulvar tumor might help in predicting the prognosis for the individual patient more accurately. One of the strategies to understand more about biological characteristics is the assessment of protein marker expression in tumor tissue. The recent development of tissue micro array (TMA) technology has given the opportunity to perform these analyses on a large scale using archival formalin-fixed, paraffin-embedded tissue. A TMA is made by arraying cores of 0.6mm, which are punched out from a series of original blocks, in a recipient paraffin block. An important question is whether the readings from the tiny cores are representative for a full section because the expression of protein markers may differ regionally in the tumor mass. If the number of cases is very large, 1 core per tumor may suffice. In a few studies with a limited number of cases, the use of 2 to 3 cores/case produced satisfactory results. Three cores/tumor reduces the risk of losing cases because of tissue damage and 3 cores allow a majority decision in the reading. However, many biomarkers are expressed with different intensity and extent in different tissue samples, which will be of influence on the reproducibility of TMA results. This attributes to the difference in reported agreement between results on TMA and full sections in previous reports. Different cut-off values also affect agreement between TMA and slides. So, before any solid conclusions can be drawn from the expression pattern of a specific antigen on TMA in vulvar cancer, the agreement between the expression pattern on slide and TMA needs to be determined.

The objective of this study was to validate the TMA technique for COX-2, Caspase-3, epidermal growth factor receptor (EGFR), p16INK4, Cyclin D1, and Ki67. In a previous study, using a TMA of 50 vulvar cancer specimens, COX-2 overexpression and absent Caspase-3 expression were associated with poor disease-specific survival in a univariate analysis, whereas Ki67, Cyclin D1, and p16INK4 were not. Other investigators found that EGFR expression was related to lymph node metastases in 197 patients with vulvar cancer. Although Ki67, Cyclin D1, and p16INK4 were not of prognostic significance in the test TMA, these markers were used in this validation study as well because they might play a role in vulvar carcinogenesis.

For the purpose of validation, we analyzed agreement between expression patterns in triplicate 0.6 mm core biopsies on TMA and on full sections. The TMA comprised specimens from 40 vulvar cancer patients. Regarding COX-2, we also analyzed the
influence of 2 different scoring systems on the measures of agreement between TMA and full sections.

**Materials and methods**

The study material consisted of slides and selected tissue blocks from 40 consecutive patients with squamous cell cancer of the vulva, treated at the Academic Medical Center in Amsterdam, The Netherlands, between 1999 and 2001. Cases were excluded if the diameter of the tumor was less than 1 cm and/or the depth of invasion less than 2 mm. All histologic specimens were reviewed by 2 of the authors (Fiebo ten Kate en Guus Fons) for histologic type and grade.

Of each tumor, 1 representative hematoxylin and eosin slide was selected. Of the selected blocks 8 slides were cut. Three representative areas of interest with infiltrative carcinoma were encircled on each slide. In the corresponding paraffin block, 0.6 mm cores were punched out. These cores, each 3 to 4 mm high, were then embedded in the donor block using a manually operated TMA device (Beecher Instruments, Silver Springs, MD). The spacing between the cores was 1 mm. The recipient block was sectioned at 4μm and transferred to glass slides. The array consisted of 123 cores (120 tumor cores, 3 cores/case) and 3 control cores of liver, lymph node, and kidney, serving as positive and negative controls. The avidin-biotin method was used for immunostaining. The unstained sections of the TMA and the slides were deparaffinized with xylol, and rehydrated through series of graded alcohol. One section of the array and 1 slide were stained with hematoxylin and eosin. Other sections of TMA and slides were stained with a panel of 6 antibodies (Table 1).

**Immunohistochemistry**

Two observers scored the staining results (Fiebo ten Kate en Guus Fons). Ki67, Cyclin D1, p16INK4 and Caspase-3 staining of tumor cells were read as negative (<10% of cells show staining), weak positive (10-50% of cells show staining) or strong positive (>50% of cells show staining). For COX-2 and EGFR, the intensity of staining was taken into account. Tumor cells were read as negative (<10% of cells show staining), weak positive (10-50% of cells show weak staining or <10% of cells show strong staining) and strong positive (>50% of cells show moderate staining or >10% of cells show strong staining). For COX-2, an alternative scoring system used in esophagus cancer has also been applied and results compared[18]. In the latter, specimens were considered to have strong expression of COX-2 if more than 50% of the carcinoma cells showed expression. Those with less than 50% expression were considered to have a weak COX-2 expression.
The same scoring system was used on both TMA and slides. Score results of cores of 1 tumor were combined into 1 score. If scores of 3 individual cores of 1 tumor differed, the one that occurred most often determined the final score. Nonassessable cores were either lost during processing or contained less than 10% tumor cells. If 2 of 3 cores were not assessable, the case was excluded. If only 2 cores with different scores were available the case, was excluded as well, except for COX-2 and EGFR. For the latter, strong focal expression was considered positive on slide. If on TMA, in case of two assessable cores, one core showed strong expression and the other core did not, the case was considered positive on TMA.

For statistical analysis, scores were dichotomized. Dichotomization with 3 classes can be achieved in 2 different ways. First, by combining negative with weak positive results and second, by combining weak-positive with strong-positive results. For each protein, the combination with the best discriminatory ability was determined.

**Statistical analysis**

We analyzed both complete agreement (also denoted as concordance) and κ, which is a measure of agreement correcting for agreement by chance. For κ < 0.2 agreement is poor, agreement is slight for 0.21 < κ ≤ 0.4, fair for 0.41 < κ ≤ 0.6, good for 0.61 < κ ≤ 0.8, very good for 0.81 < κ ≤ 0.92 and excellent for 0.93 < κ ≤ 1.0 [19]. Calculations were performed with SPSS for Windows 11.5 (SPSS, Inc, Chicago,IL).

**Results**

The results of the readings of full sections are shown in Table 1. Of 120 tumor cores, between 83% and 88% of cores were assessable, depending on the antigen used. Assessment was possible with 3 cores in 65% to 72% of the cases and limited to 2 cores in 20% to 25% of the cases (Table 2).

In 3 cases, slides were not assessable because of inadequate staining repeatedly (1 for Ki76, 1 for Cyclin D1 and 1 for EGFR). In another case (p16\(^{INK4}\)), the scores of the available 2 cores were discordant. This case was excluded as well (Table 3).

Agreement between scoring results on slides and those on cores depended on the scoring system. If a 2 class scoring system was used, due to dichotomization, agreement was better than in a 3 class scoring system (Table 4). Using the 2-class scoring system κ was 0.68 to 0.75 (good) for all proteins tested, except for Caspase-3. For Caspase-3, κ was 0.40 (slight).
Table 1. Characteristics of antibodies used for immunohistochemical staining and scoring results on full sections

<table>
<thead>
<tr>
<th>Supplier, clone</th>
<th>Dilution</th>
<th>Negat. (%)</th>
<th>Weak (%)</th>
<th>Strong (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proliferation marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ki-67 n Dako, MIB1</td>
<td>1:200</td>
<td>5</td>
<td>80</td>
<td>15</td>
</tr>
<tr>
<td>Tumor suppressor gene p16INK4 n/c NeoMarkers, Ab-7</td>
<td>1:100</td>
<td>77</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Oncogenes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin D1 n NeoMarkers SP4</td>
<td>1:100</td>
<td>15</td>
<td>23</td>
<td>62</td>
</tr>
<tr>
<td>EGFR m NeoMarkers, 111.6</td>
<td>1:400</td>
<td>44</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>Apoptosis marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caspase-3 c Cell Signaling Technology, Asp175</td>
<td>1:100</td>
<td>77</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Prostaglandin biosynthesis marker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COX-2 c Cayman Chemical, 160112</td>
<td>1:800</td>
<td>66</td>
<td>13</td>
<td>21</td>
</tr>
</tbody>
</table>

Dako (Carpinteria, CA); NeoMarkers, (Fremont, CA); Cell Signaling Technology (Danvers, MA); Cayman Chemical (Ann Arbor, MI) 1c indicates cytoplasm; EGFR, epidermal growth factor receptor; m, membrane; n, nucleus.

Table 2. Number of cases per number of assessable cores per antigen

<table>
<thead>
<tr>
<th>Antigen</th>
<th>0</th>
<th>%</th>
<th>1</th>
<th>%</th>
<th>2</th>
<th>%</th>
<th>3</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ki-67</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>25</td>
<td>28</td>
<td>70</td>
</tr>
<tr>
<td>p16INK4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>20</td>
<td>29</td>
<td>72</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>22</td>
<td>27</td>
<td>67</td>
</tr>
<tr>
<td>EGFR</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>22</td>
<td>27</td>
<td>67</td>
</tr>
<tr>
<td>Caspase-3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>25</td>
<td>27</td>
<td>67</td>
</tr>
<tr>
<td>COX-2</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>22</td>
<td>26</td>
<td>65</td>
</tr>
</tbody>
</table>

EGFR indicates epidermal growth factor receptor.

Table 3. Characteristics of cases regarding availability for comparison of tissue microarray and slides

<table>
<thead>
<tr>
<th>Antigen</th>
<th>At least 2 cores</th>
<th>Nonassessable slides</th>
<th>2 discordant cores</th>
<th>No. assessable cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ki-67</td>
<td>38</td>
<td>1</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>p16INK4</td>
<td>37</td>
<td>0</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>EGFR</td>
<td>35</td>
<td>1</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>Caspase-3</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>COX-2</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
</tbody>
</table>

EGFR indicates epidermal growth factor receptor.
Discordant cases on TMA and Slides

Cases of which the result on TMA did not agree with the one on slide were analyzed. For COX-2 and EGFR, the majority of the conflicting cases were negative on slide and positive on TMA. For Ki67 and Caspase-3, the opposite pattern was shown. For p16INK4 and Cyclin D1, a pattern could not be distinguished (Table 5).

For Caspase-3, all 8 discordant cases were positive on slide and negative on TMA. In the majority of positive cases, the percentage of positive cells was around 10%. As the number of positive cells is very low and positive cells are not equally distributed in the tumor, the chance for false-negative cases on TMA increases. This has contributed to the difference between the numbers of positive cases on TMA and slides.

COX-2 Scoring Systems

For COX-2, 2 different scoring systems were applied and the results compared. In the first chosen scoring system, intensity of staining was part of the scoring system. Although COX-2 expression is focally, concordance between slides and cores showed a $\kappa$ of 0.68.
An example of focal expression is shown in Figure 1. Two cores and 2 parts of the corresponding slide are shown. The first core (A) shows strong expression. This corresponds well with part of the slide, as is shown in B. The second core of the same case (C) does not show expression, which is corresponding well to another part of the slide (D).

By using the alternative scoring system, in which strength of expression is not taken into account, complete agreement decreased from 89% (31/35) to 84% (27/32) and $\kappa$ decreased from 0.68 to 0.21.

![Figure 1](image.png)

**Figure 1.** Cox-2 expression on cores and slide of one case. (A) Over-expression on core. (B) Over-expression on corresponding part of slide. (C) Absent expression on core. (D) Absent expression on corresponding part of slide.

### Discussion

Expression patterns of antibodies can be assessed on TMA with good reliability ($\kappa$ 0.68-0.75) for Ki-67, p16$^{INK4a}$, COX-2, Cyclin D1, and EGFR in comparison with whole slides. For Caspase-3, agreement is only slight with a $\kappa$ of 0.40. All 8 discordant cases for Caspase-3 were positive on slide and negative on TMA. For Cox-2, 2 different scoring systems were applied. When intensity of staining was part of the scoring system used, $\kappa$ was better (0.68)
than when intensity of staining was not taken into account (0.21). Agreement between results on TMA and slides depended on the scoring system used and the protein tested. For Ki67, agreement between slides and TMA was good (κ 0.68). In a validation study on tissues of 59 patients with human fibroblastic tumors, Ki67 expression on slides was compared with TMA. A κ of 0.87 was found.[8] In the large validation study by Nocito et al. [5] on bladder cancer, a strong correlation was also found between the findings on TMA and slides (p<0.0001) for Ki67.

In the current study, 3 cases of 37 showed nonagreement. In these cases, slides were positive and cores were negative. Background staining, which is more distinct on slides than in small cores, might have contributed to this.

For p16INK4, agreement between TMA and slides was good (92%) with a κ of 0.75. This result was comparable with that in an earlier validation study on lung cancer for p16INK4.[10]

For Cyclin D1, agreement between TMA and slides was also good (86%), κ was 0.68. This result was comparable with that in earlier validation studies. A κ of 0.90 was reported in a validation study on colorectal adenoma and a κ of 0.66 was reported in a validation study on head and neck squamous cell cancer (HNSCC).[9] [11]

For interpreting EGFR expression, many different scoring systems exist, as was shown in a meta-analysis of the predictive value of this protein in colorectal cancer.[20] Despite its subjective nature it is recognized that staining intensity of the membrane should be a component of the EGFR scoring system.[21;22] The combination of intensity and the estimation of the number of positive cells into one score, as used in our study, will theoretically have a negative effect on the intraobserver and interobserver variability. Despite this, a good agreement was found between scores on slides and cores (κ 0.68). This result was better than the agreement reported in the study on validation of tissue array technology in HNSCC (κ 0.41).[11]

Four cases showed a discordant result in our study. In all cases, slides were negative and cores were positive. There were 2 studies on EGFR expression in vulvar epithelium.[23;24] In both studies, based on a limited number of patients, the problem of interpretation of cytoplasmatic staining was discussed. It is not clear whether cytoplasmatic staining influences the scores on slides and TMA in the same direction.

For Caspase-3, agreement between slides and TMA was 78% (29/37) with a κ of 0.40. Caspase-3 expression of more than 10% was observed in only 23% of the cases. This was less than 40% Caspase-3 positive cases in our test set consisting of 50 vulvar cancer specimens on TMA.[12] The total amount of positive cells was very low, both on slides and on TMA, as compared with some other studies on Caspase-3 expression in non-small cell lung cancer and esophageal squamous cell carcinoma.[25;26] In these studies, 60% and 72% of cases were considered positive (more than 25% of cells showed staining).
In a study on Caspase-3 expression in 36 cases with nasopharynx cancer, however, the percentage of Caspase-3-positive tumor cells ranged from <1% to 3%.[27] These differences might have been partly due to the use of different antibodies. All discordant cases (8) were negative on TMA and positive on slides. Two factors might have contributed to this. First, as the number of positive cells is low and positive cells were not equally distributed in the tumor the chance for false-negative cases on TMA increases. Second, tumor-infiltrating lymphocytes showed positive Caspase-3 staining as well. [27] It is tempting to state that these cells are more easily detected in small cores than on whole slides resulting in more false positive cases on slides than on TMA. By the use of another scoring system, in which the number of positive cells is determined per field, these drawbacks are possibly overcome.

A wide range of cytoplasmatic immunoreactivity of COX-2 was described in vulvar cancer. [28] The importance of strong focal expression was incorporated in our scoring system. Despite the focal expression pattern, a good agreement existed between slides and TMA (κ 0.68).

In most studies on COX-2, different scoring systems were used.[18;29;30] If a scoring system is used, in which only the number of positive cells are important for the score, agreement between TMA and slides decreased as κ decreased from 0.68 to 0.21. The magnitude of the difference between these results illustrates the weakness of the strategy. For COX-2, immunostaining was also observed in inflammatory and endothelial cells.[28] This will mainly influence the staining pattern on the whole slide, adding to more positive results on slides than on TMA. In our series, the opposite pattern was found, which could not be explained.

TMA is a reliable tool to investigate protein expression patterns in tumors. The agreement between scores on slides and TMA varies with the antigens tested and the scoring systems used. The best results are found with antigens, which are diffusely expressed in a tumor. Results are especially vulnerable in case of focal expression and a low number of positive cells.
References


(20) Italiano A. Targeting the epidermal growth factor receptor in colorectal cancer: advances and controversies. Oncology 2006; 70(3):161-167.

(21) Goldstein NS, Armin M. Epidermal growth factor receptor immunohistochemical reactivity in patients with American Joint Committee on Cancer Stage IV colon adenocarcinoma: implications for a standardized scoring system. Cancer 2001; 92(5):1331-1346.


