Validation of a fully automated HER2 staining kit in breast cancer

Cathy B. Moelans a, Robby E. Kibbelaar b, Marius C. van den Heuvel b, Domenico Castigliego a, Roel A. de Weger a,∗ and Paul J. van Diest a

a Department of Pathology, University Medical Centre Utrecht, Utrecht, The Netherlands
b Pathology Friesland, Leeuwarden, The Netherlands

Abstract. Background: Testing for HER2 amplification and/or overexpression is currently routine practice to guide Herceptin therapy in invasive breast cancer. At present, HER2 status is most commonly assessed by immunohistochemistry (IHC). Standardization of HER2 IHC assays is of utmost clinical and economical importance. At present, HER2 IHC is most commonly performed with the HercepTest which contains a polyclonal antibody and applies a manual staining procedure. Analytical variability in HER2 IHC testing could be diminished by a fully automatic staining system with a monoclonal antibody.

Materials and methods: 219 invasive breast cancers were fully automatically stained with the monoclonal antibody-based Oracle HER2 Bond IHC kit and manually with the HercepTest. All cases were tested for amplification with chromogenic in situ hybridization (CISH).

Results: HercepTest yielded an overall sharper membrane staining, with less cytoplasmic and stromal background than Oracle in 17% of cases. Overall concordance between both IHC techniques was 89% (195/219) with a kappa value of 0.776 (95% CI 0.698–0.854), indicating a substantial agreement. Most (22/24) discrepancies between HercepTest and Oracle showed a weaker staining for Oracle. Thirteen of the 24 discrepant cases were high-level HER2 amplified by CISH, and in 12 of these HercepTest IHC better reflected gene amplification status. All the 13 HER2 amplified discrepant cases were at least 2+ by HercepTest, while 10/13 of these were at least 2+ for Oracle. Considering CISH as gold standard, sensitivity of HercepTest and Oracle was 91% and 83%, and specificity was 94% and 98%, respectively. Positive and negative predictive values for HercepTest and Oracle were 90% and 95% for HercepTest and 96% and 91% for Oracle, respectively.

Conclusion: Fully-automated HER2 staining with the monoclonal antibody in the Oracle kit shows a high level of agreement with manual staining by the polyclonal antibody in the HercepTest. Although Oracle shows in general some more cytoplasmic staining and may be slightly less sensitive in picking up HER2 amplified cases, it shows a higher specificity and may be considered as an alternative method to evaluate the HER2 expression in breast cancer with potentially less analytical variability.

Keywords: HER2, Oracle, HercepTest, CISH

1. Introduction

HER-2/neu is a proto-oncogene located on chromosome 17q21 encoding a 185 kD transmembrane tyrosine kinase receptor protein that is involved in signal transduction [1,15]. HER2 belongs to the human epidermal growth factor receptor (EGFR) family and is amplified in about 15–25% of breast carcinomas causing an increased expression of its protein [13, 17,20]. Patients having this overexpression respond well to treatment with trastuzumab (Herceptin®), a recombinant humanized monoclonal anti-HER2 antibody [5,21]. Since the costs for trastuzumab therapy are high and side-effects are significant, accurate selection of eligible patients for this therapy is crucial. Furthermore, amplification and overexpression of HER2 has also been shown to correlate with poor prognosis [7] and with resistance to conventional adjuvant chemotherapy and tamoxifen [3,4,18,19,24]. For these reasons, testing for HER2 amplification and/or overexpression is currently considered routine practice in clinical pathology laboratories. At present, HER2 status is most commonly assessed by immunohistochemistry (IHC) and/or gene amplification tests such as fluorescence in situ hybridization (FISH) [6,9,14] or chro-
mogenic in situ hybridization (CISH) [22] or multiplex ligation-dependent probe amplification [10–12]. Immunohistochemistry (IHC) is the most commonly used method to assess HER2 protein overexpression. It is a rather easy morphological method which has many advantages like its wide availability, relatively low cost, easy preservation of stained slides, and use of a familiar routine microscope. Disadvantages of IHC include the impact of pre-analytic issues including storage, duration and type of fixation, intensity of antigen retrieval, type of antibody (polyclonal versus monoclonal) [17], nature of control samples and the difficulties in applying a subjective semi-quantitative slide scoring system. For scoring of IHC staining, the 0–3+ visual system developed for the HercepTest (based on a polyclonal anti-HER2 antibody, clone A0485, Dako, Glostrup, Denmark) is widely in use. As there is little difficulty in assigning the 0 and 3+ scores, interpretation is more problematic for the two intermediate levels. For cases scoring 2+ (10–15% of all breast cancers), the concordance with gene amplification by FISH or CISH is barely 25%, and yet a proportion of these 2+ cases are true HER2 amplified tumors. These cases, therefore, require a second line gene amplification test. Because of its central importance in breast cancer therapy selection, standardization of HER2 IHC assays and slide interpretation are of utmost clinical and economical importance. Analytical variability in HER2 IHC testing can be minimized by the use of standardized tests, and by inter-laboratory quality control assessments. A fully automatic IHC staining system similar to the FDA-approved Ventana system (Pathway® anti-HER2 rabbit monoclonal antibody, clone 4B5, Ventana Medical Systems, Tucson, AZ, USA) can improve the specificity, positive predictive value and efficiency of IHC [2] and can thereby produce a more consistent and reproducible result. The present study aimed to examine the suitability of the new Oracle HER2 Bond IHC System (Leica Microsystems, Newcastle, UK, TA9145) for use as an aid in determination of eligibility for trastuzumab therapy. This fully automatic system is intended for use on Leica Microsystems’ Bond-maxTM devices and contains a ready-to-use mouse monoclonal anti-HER2 antibody (clone CB11) and a ready-to-use Compact PolymerTM detection system, both required to complete an immunohistochemical staining procedure for formaldehyde-fixed paraffin-embedded tissues.

2. Materials and methods

2.1. Patient material

Tissue samples of 219 invasive breast cancer patients were retrospectively collected at the Department of Pathology of the University Medical Centre in Utrecht (UMCU) and at Pathology Friesland, The Netherlands. Anonymous use of redundant tissue for research purposes is part of the standard treatment agreement with patients in both hospitals [26]. Biopsies were excluded from this study, and only whole sections were used. Both institutes separately carried out parallel manual and automated IHC stainings on their own tissue samples, using identical protocols and machines. All CISH stainings were performed at the UMCU.

2.2. Immunohistochemistry (IHC)

Manual IHC for HER2 was performed using the HercepTest (Dako, Glostrup, Denmark) according to the manufacturers’ instructions on 4 µm thick sections from neutral buffered formaldehyde fixed tissue blocks. As control, a small tissue array containing a 0, 1+ and 3+ breast tumor sample was taken along on the same slide as the tumor to be analyzed. Negative controls were obtained by omission of the primary antibody.

Automated IHC for HER2 was performed on a Bond-maxTM device using the Oracle HER2 Bond IHC System (Leica Microsystems, Newcastle, UK, TA9145). Staining was performed according to the manufacturers’ instructions on 4 µm thick sections from neutral buffered formaldehyde fixed tissue blocks. Control slides with four (0, 1+, 2+ and 3+ intensity) neutral buffered formaldehyde fixed, paraffin-embedded human breast cancer cell lines are provided to validate staining runs. In each run (i.e., slide tray) 4 tumor samples (primary monoclonal anti-HER2 antibody, clone CB11), 4 negative control samples (primary antibody is replaced by a supplied ready-to-use mouse IgG), a supplied HER2 positive control slide and an in-house positive control slide were analyzed. Also, at the UMCU, a small tissue array containing a 0, 1+ and 3+ breast tumor sample was taken along on the same slide as the tumor to be analyzed.

IHC membrane staining was semi-quantitatively scored as negative (0), weakly positive (1+), equivocal (2+) and strongly positive (3+) according to the DAKO FDA-approved scoring system. Areas with in-
traductal carcinoma were excluded from the evaluation and cytoplasmic staining was ignored. Interpretation of all IHC stainings was done blinded by 1 experienced breast pathologist at the UMCU to exclude inter-observer variability.

2.3 Chromogenic in situ hybridization (CISH)

All CISH assays were run using the Zymed SPoTLight HER2 CISH (Zymed, South San Francisco, CA, USA) kit according the manufacturers’ instructions. CISH was performed on 4 µm thick paraffin sections and scoring was performed according the manufacturers’ guidelines. Briefly, HER2 was scored high-level amplified when large peroxidase-positive intra-nuclear gene copy clusters or >10 individual small signals were present in more than 50% of tumor cells. HER2 was scored low-level amplified when more than 50% of the tumor cells showed 6–10 dots per nucleus, or in the presence of small clusters. Tumors were scored as non-HER2 amplified when tumor cells showed 1–5 dots per nucleus. A positive control was included in each CISH run and consisted of a paraffin section of a case known to be HER2 amplified by CISH. Scoring was done blinded to the IHC results by one experienced observer. Doubtful cases were evaluated together with another experienced observer until agreement was reached.

2.4 Statistics

Results obtained with manual and automated IHC techniques were compared by cross tables, and the concordance percentages and weighted kappa-scores were calculated. Using CISH results as gold standard and regarding low-level amplification as positive, we calculated sensitivity, specificity, and positive and negative predictive values for both IHC techniques.

3. Results

Tissue samples in this study were collected at two hospitals. There were only small differences in scoring (performed at the UMCU) between both hospitals. Of samples originating from the UMCU, 59% were scored 0/1+, 14% were scored 2+ and 27% were scored 3+. Of samples originating from the Pathology Friesland laboratory, 68% were scored 0/1+, 8% were scored 2+ and 24% were scored 3+. Furthermore, at the UMCU and PF, 59% and 68% of samples were scored non-amplified, respectively. Nine percent and 5% were scored low-level amplified, and 31% and 27% of samples were scored CISH-amplified, respectively.

Overall, HercepTest yielded a sharper membrane staining and showed less cytoplasmic and stromal background than Oracle in 37/217 (17%) of the patients, which is illustrated in Fig. 1. Table 1 shows the concordance between the HercepTest and Oracle HER2 staining. Overall concordance between both IHC techniques was 89% (195/219) with a kappa value of 0.776 (95% confidence interval 0.698–0.854) indicating a substantial agreement.

Tables 2 and 3 show 86% concordance between CISH and HercepTest, and 84% between CISH and Oracle. Concordance was highest in HercepTest and Oracle 0/1+ cases, with percentages of 95% and 91.5%, respectively, and in HercepTest and Oracle 3+ cases, both 87%. Of all HercepTest and Oracle 2+ cases, 11/23 (48%) and 16/21 (76%) were amplified by CISH.

Table 4 shows the CISH results on all discrepancies between HercepTest and Oracle. The two “2+ to 0” discrepancies were both not amplified by CISH. Of the ten “2+ to 1+” discrepancies, three were high-level amplified by CISH, 4 were low-level amplified, and 3 tumors did not show HER2 amplification by CISH. The single case that was 1+ by HercepTest and 2+ by Oracle was low-level amplified by CISH. The single case that was 2+ by HercepTest and 3+ by Oracle was high-level amplified by CISH. Of the 10 cases that were 3+ by HercepTest and 2+ by Oracle, 1 was low-level amplified and 9 high-level amplified by CISH.

Of the 24 discrepant cases in this study, 13 were high-level amplified by CISH. Of these 13 cases, HercepTest better reflected gene amplification status than Oracle in 12/13 cases. All the 13 HER2 amplified discrepant cases were at least 2+ by HercepTest, while 10/13 of these were at least 2+ for Oracle. Weighted kappa scores were calculated based on the 24 cases in Table 4. Weighted kappa for HercepTest versus CISH was 0.273 with 95% CI (0.02604–0.51996). Weighted kappa score for Oracle versus CISH was 0.121 with 95% CI (−0.006912–0.31112).

Figure 2 shows two tumors with discrepant HercepTest and Oracle scores. Considering CISH as gold standard, we also calculated sensitivity and specificity of HercepTest and Ora-
cle to detect HER2 overexpression. Sensitivity of HercepTest and Oracle was 91% and 83%, respectively, and specificity was 94% and 98%, respectively. Positive and negative predictive values for HercepTest were 90% and 95%, respectively, and 96% and 91% for Oracle.

4. Discussion

This study aimed to validate the Oracle HER2 Bond IHC System as an alternative to HercepTest for determination of eligibility of breast cancer patients for trastuzumab therapy. The Oracle system is based on a monoclonal antibody and automated staining, and is thereby potentially less liable to analytical variability than the polyclonal antibody and manual staining based HercepTest. Within the study setup of the present paper, one observer scored all cases to exclude inter-observer variability. In practice, these kind of stainings are usually being scored by several pathologists in one lab, inherently introducing inter-observer variability. Especially in 1+ and 2+ cases, several studies have shown marked inter-observer variability [25]. We therefore realize that the reproducibility in the present study is in the optimistic range.

Table 1
Concordance between the manual polyclonal antibody based HercepTest and fully automated monoclonal antibody based Oracle HER2 staining in 219 invasive breast cancers

<table>
<thead>
<tr>
<th></th>
<th>Oracle</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0/1+</td>
<td>2+</td>
</tr>
<tr>
<td>HercepTest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0/1+</td>
<td>140</td>
<td>1</td>
</tr>
<tr>
<td>2+</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>3+</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>152</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 2
Concordance between the manual polyclonal antibody-based HercepTest and chromogenic in situ hybridization (CISH) in 219 breast cancers

<table>
<thead>
<tr>
<th>CISH Total</th>
<th>NA</th>
<th>LA</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>HercepTest 0/1+</td>
<td>134</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2+</td>
<td>6</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>3+</td>
<td>2</td>
<td>5</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>142</td>
<td>14</td>
<td>63</td>
</tr>
</tbody>
</table>

Notes: NA – not amplified; LA – low-level amplified; A – amplified.

Table 3
Concordance between fully automated monoclonal antibody-based Oracle HER2 staining and chromogenic in situ hybridization (CISH) in 219 breast cancers

<table>
<thead>
<tr>
<th>CISH Total</th>
<th>NA</th>
<th>LA</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle 0/1+</td>
<td>139</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2+</td>
<td>1</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>3+</td>
<td>2</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>142</td>
<td>14</td>
<td>63</td>
</tr>
</tbody>
</table>

Notes: NA – not amplified; LA – low-level amplified; A – amplified.

Overall, HercepTest yielded a sharper membrane staining with less cytoplasmic and stromal background than Oracle. For Oracle, this is a slight disadvantage since membrane staining may be more difficult to assess, and this may lead to a higher inter-observer variability as described in other studies comparing CB11 and other anti-HER2 antibodies [8,16,23]. Nevertheless, the clinically relevant 3+ staining is still well recognizable, and 2+/3+ discrepancies were not caused by background staining in the present study. In practice, this theoretical disadvantage is therefore probably not a big problem.

We found a good agreement (89%) between HercepTest and Oracle with a kappa score of 0.78. All cases were tested for HER2 gene amplification by CISH as gold standard. In the vast majority of the discrepancies, Oracle showed a weaker staining than HercepTest. Furthermore, in the 13 discrepant cases that were high-level HER2 amplified by CISH, HercepTest better reflected gene amplification status (by higher IHC score) than Oracle in 12/13 cases. All the 13 HER2 amplified discrepant cases were at least 2+ by HercepTest, while 10/13 of these were at least 2+ for Oracle. Weighted kappa scores were calculated based on the 24 cases and showed that kappa for HercepTest versus CISH was higher than kappa for Oracle versus CISH. This implies that HercepTest better reflected gene amplification status (by CISH) than Oracle in these 24 discrepancies.

Overall, HercepTest and Oracle were 86% and 84% concordant with CISH, respectively. The overall sensitivity of HercepTest was slightly better than that of Oracle, but Oracle had a slightly better specificity. There were 4 patients with HercepTest score 0/1+ but CISH amplification, and 7 patients showed an Oracle score of 0/1+ while CISH showed an amplification. Those patients are clinically most relevant, as the 1+ score would not have triggered a second line amplification test in daily practice, and these patients would not have received Herceptin therapy from which they may have benefited. Larger studies need to further assess the clinical sensitivity of the Oracle system compared to the HercepTest.

Certain pre-analytical factors should result in the rejection of the specimen for IHC evaluation of HER2 status such as fixation longer than 48 h, tissues fixed in fixatives other than neutral-buffered formaldehyde and
the presence of severe edge or crush artefacts in core needle biopsies. Given this last recommendation we decided to exclude biopsies from this validation study. Further validation of the Oracle system for core needle biopsies will be the subject of future research.

IHC assays are appealing for a number of practical perspectives including the lower cost, lower turnaround time compared to gene amplification tests and the adaptability to most pathology laboratories. However, immunohistochemistry is a multi-step diagnostic process that requires specialized training in all aspects of the procedure including the selection of the appropriate reagents and tissue, fixation, processing and interpretation of the staining results. It is thus of the utmost importance that all steps of the process are properly standardized. The development of fully-automated systems can aid in the standardization of the IHC staining process, thereby potentially producing more consistent results.

In conclusion, fully-automated HER2 staining with the monoclonal CB11 antibody in the Oracle kit shows a high level of agreement with manual staining by the polyclonal antibody in the HercepTest. Although Oracle shows in general some more cytoplasmic staining and may be slightly less sensitive in picking up HER2 amplified cases, it shows a higher specificity and may be considered as an alternative method to evaluate the HER2 expression in breast cancer with potentially less analytical variability.

Acknowledgements

We acknowledge Nicolette Kurten and Menarini Diagnostics for providing us with the Oracle HER2
Bond IHC and CISH kits without restrictions.

References


