Optimizing breast cancer survival models based on conventional biomarkers and stromal parameters

Dekker, T.J.A.; Dekker T.J.A.

15

Summarizing discussion

Part I

Despite the increasingly improving prognosis of breast cancer patients [1], much work has to be done to improve the pathological evaluation of these tumors. The parameters and biomarkers determined as part of the breast cancer diagnosis need to be both accurate and precise. Determining the accuracy of a test is often difficult, as this requires validating test results to a gold standard that is often impossible to determine. This gold standard should be the clinical parameter that the marker is supposed to predict, but response to therapy and patient survival for example likely depend on more than just one parameter. Determining testing precision can be more easily done as this in essence is performed by repeating the test and comparing results. Despite this simplicity, this does pose two questions. If a discordant result is found between two tests, which one is correct? Secondly, what kind of variation is acceptable between both test results? Guidelines often cite that reproducibility for biomarker testing should be at least equal to 95%. In the statement of the ASCO/CAP panel regarding the testing reliability of HER2 assessments, Wolff et al. state "It is recommended that to perform HER2 testing, laboratories show 95% concordance with another validated test for positive and negative assay values" [2]. Similar guidelines were also described in the guidelines for hormone receptor testing.

With this in mind, the concordance between core needle biopsies (CNB) and resection specimens regarding estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) was investigated in **chapter 2** of this thesis. Although using CNB for these determinations has certain advantages (superior fixation for instance), the literature remains divided on whether CNB can be reliably used for this purpose. Concordance rates from 61.8 to 99.0% have been published, leading to variable conclusions and recommendations. These studies have all been single-center studies following local protocol for assessing this concordance. We examined the concordance between CNB and resection specimens for a series of patients treated in the Leiden University Medical Center (LUMC) and combined these with cases published in the literature. Pooled analysis included a number of 2622 ER-tested breast carcinomas. Overall concordance between CNB and resection specimens was 93.7%, less than the previously discussed 95%. Considering the substantial number of cases that were ER-negative on CNB while the resection specimen tested ER-positive (2.2% of all cases in the pooled analysis), the recommendation was made to retest tumors on the resection specimen when the CNB result is ER-negative. Regarding HER2, overall concordance was high at 97.8% in a pooled analysis of 646 published cases from 4 studies. Most of the discordant cases were false-positive IHC results on CNB, leading to our

recommendation to verify HER2 overexpression (3+ staining) with in situ hybridization when CNBs are used.

These studies were all performed to compare the concordance of CNB with resection specimens without intervening neoadjuvant chemotherapy. The administration of chemotherapy can induce changes in the expression of hormone receptors and HER2 due to selection of chemotherapy-resistant cells or altered estrogen milieu. A systematic review was performed in **chapter 3** to investigate the concordance between CNB and resection specimens for these markers when neoadjuvant chemotherapy was administered. Neoadjuvant chemotherapy-induced changes regarding ER status were described in 5 prospective studies whereas no significant changes were reported in 8 prospective studies. Studies that reported no changes included lower number of patients, and were thus likely to be underpowered for detecting changes. HER2 amplification (when detected with FISH) was more stable after administration of neoadjuvant chemotherapy regimens that did not include trastuzumab, as significant changes were only found in 1 prospective study, whereas no changes were detected in 6 trials. However, when trastuzumab is administered alongside chemotherapy, 43% of all patients show no HER2 amplification in the residual tumor. These results have led to the suggestion to always retest ER on resection specimens in cases of neoadjuvant chemotherapy and retest HER2 expression when HER-inhibiting drugs are administered.

Research into the quality of ER, progesterone receptor (PR) and HER2 testing has shown cause for concern and has emphasized the need for stringent quality control. In **chapters 4** and **5**, we report the results of tissue micro array (TMA)-study to investigate the reproducibility of hormone receptors (HRs) and HER2 testing. The reproducibility of ER and PR tests performed by pathology laboratories in the Netherlands was investigated in **chapter 4**. TMAs were constructed from tumors that were locally tested as part of routine diagnostics. These TMAs were tested for ER and PR expression, which was compared to the local HR testing result. Whole slides were tested when a discordant result was found between the local test result and the TMA result. Overall concordance between the local and central test result for the ER-tested tumors was 99.0% based on 1569 ER-tested cases. ER discordant cases were due to observer fault in the majority of cases. Regarding PR testing, overall concordance was 94.1% based on 1347 PR-tested cases. The discordances regarding PR status were mostly due to IHC error. Also, we investigated whether these concordant cases were affected by the recommended threshold change to 1% positive cells advocated by the ASCO/CAP. We found that a significant number of cases were concordant when adhering to these guidelines. **Chapter 5** describes the development of a TMA-based

testing method for assessing HER2 reproducibility. Different HER2 testing methods were investigated to assess the optimal testing methods for such a TMA-approach. HER2 status was assessed via SP1, 4B5 antibodies and mono color silver in situ hybridization (SISH) on TMAs and was compared to the HER2 status determined in the local laboratory (whole slides were tested in case of discordance). Overall concordance between the centrally performed tests and the local HER2 tests was 98.0% based on 1008 breast cancers. The most frequent discordant result was a local false-positive result, either due to ISH (*N*=6) or IHC (*N*=6) procedures. All false-negative cases were cases where local IHC-procedures were unable to detect HER2 expression, while this was positive on both IHC and silver in situ hybridization assays when tested centrally.

Previous publications have shown that lymph vascular space invasion (LVSI) can provide prognostic information that is independent of other clinico-pathological parameters. The optimal implementation of lymph vascular space invasion (LVSI) in routine diagnostics of breast cancer remains unknown and requires further investigation in clinical cohorts. In **chapter 6**, we examined quantification of LVSI in a cohort of 358 tumors treated as part of the perioperative chemotherapy trial (POP trial)[3]. We compared two systems for performing LVSI quantification, a cut-off based on the number of LVSI foci in the peritumoral environment or a cut-off based on the multiplication of the number of tumor cells with the number of LVSI foci. The latter quantification system, termed the LVSI tumor burden (LVSI-TB) was superior in detecting patients with an increased risk for disease relapse in a discovery cohort. This observation was validated in a validation cohort. Otherwise low-risk breast cancer patients with high LVSI-TB had a similar prognosis to other high-risk patients, thereby justifying upgrading these patients to high risk. In **chapter 7**, we investigated whether such a method for quantification can be reliably determined among 4 dedicated breast pathologists. These observers were asked to both assess the number of individual LVSI foci as well as perform the quantification as previously described in a set of 60 tumors. Concordance regarding the quantification among all 4 observers was only seen in 77% of all cases, for the remaining cases at least one observer was discordant with the others. Based on the observations by individual observers, we determined a consensus score for quantitative LVSI scoring and assessed the mean sensitivity (83.3%) and specificity (92.8%) of the 4 observers. By implementing a scoring algorithm involving a second observer in cases of LVSI-positive tumors, the chance of minimizing missing LVSI-TB high tumors might be reduced. However, in order to reliably assess the presence of LVSI-TB along the methodology, the use of IHC techniques seems necessary. In **chapter 8**, several aspects of Ki-67 detection using immunohistochemistry were investigated. A number of 105 breast cancer patients treated as part of the MicroarRAy PrognoSTics in Breast CancER (RASTER) study [4] were tested for Ki-67

status using immunohistochemistry. By creating virtual biopsies (reflective of Ki-67 scores that would have been obtained on CNB), the impact of tumor heterogeneity on Ki-67 testing reproducibility was investigated. The overall discordance between CNB and resection specimens was 18.7%. Secondly, the interobserver agreement was found to be 84%, with a kappa score of 0.669. Finally, the concordance between Ki-67 tests between two reference laboratories (namely the NKI-AVL and the LUMC) was tested to discover real-world testing variability of Ki-67 testing on resection specimens. The discordance rate was 12.3%.

Implications for future studies

In this thesis, several aspects concerning the optimization of prognostic breast cancer parameters, specifically HRs, HER2, LVSI and Ki-67, were discussed. These parameters are at this moment the cornerstones of breast cancer risk assessment on which treatment decisions are based. The evidence that has been gathered on the optimization of these parameters in this thesis and in many other works on this matter should be used to improve the standardization of breast cancer diagnostics and treatment. This standardization is essential to guarantee that the impressive results observed in clinical studies from individual centers as part of research projects can be successfully translated to routine patient care. Although this thesis concerns improving the current standard of care, it is foreseeable that novel testing methods will eventually replace current assays, provided that reliable, reproducible results can be assured. These developments are discussed in the following paragraphs.

Alternative biomarker assessments

Although IHC is considered the gold standard for determining HR status, alternative methods for measuring ER and PR expression have been published in the literature [5]. Viale et al. published the concordance of local and centrally performed ER tests with the ER readout performed in the Mammaprint assay [6]. Concordance was found to be high between these scores and the centrally performed IHC scores. Such assays might therefore be preferable to local IHC procedures in cases of underperforming IHC laboratories, thus providing such centers with an alternative. However, might these assays also be preferable to traditional IHC staining in the cases of laboratories that perform well-validated IHC tests? This would be the case if these assays have a higher predictive value of response to estrogen-modulating therapies than the IHC based assays. While this has never been directly examined, there is evidences that suggests that determining ER and PR expression via microarrays might lead to false-positive and false-negative HR results. A small number of ER-positive

cases are mistakenly considered as false-negative with the Oncotype DX assay [7]. Also, HR-negative tumors on IHC might produce HR-positive results due to intermingled HR-positive DCIS lesions and normal ducts leading to false positive results. These results suggest that relying on ER-tests assessed by gene-expression arrays is currently not preferable to traditional IHC assays.

Alternative prognostication strategies

Many studies on additional prognostication strategies have been published. Gene-expression signatures measure the expression of multiple genes and correlate these genetic profiles to risk for disease relapse and patient survival [8, 9]. Will these assays reduce the need for the traditional molecular testing and risk stratification as it is currently performed? The answer to this question is found in their testing reliability and reproducibility, the costs of these multi-gene assays and of course their prognostic power.

Interlaboratory studies of gene-expression assays found high concordance among different testing centers [10, 11]. The analytical reliability of the Oncotype DX assay was found to be high as performing repeat assessments of one sample over multiple days with multiple operators yielded similar results [12]. The Oncotype DX and Mammaprint assays are as of yet performed by one central laboratory but such standardization comes at a cost. The price of one Oncotype DX assay ($4,175) and one Mammaprint assay ($ 4,200) are now an estimated ten-fold of the price for an individual IHC test. However, one should bear in mind that some patients are spared chemotherapy because of these tests, thereby decreasing health-care costs [13].

As is also the case when determining risk stratification by using classical morphology or IHC, tumor heterogeneity can potentially threaten the reproducibility of test results. Whether epithelial tumor heterogeneity might compromise prognostication is currently unknown, but the limited data on this matter suggests that this not the case [14]. Intermingled tissues such as fibroblasts, the extracellular matrix, inflammatory cells, fibroblasts and non-malignant tissues might however theoretically influence test results. Acs et al. described a number of cases where the recurrence score was unexpectedly high based on what is known about the tumor type. These cases generally displayed mitotically active stroma, which might have influenced the recurrence score [15]. This suggests that the RNA extracted from the mixture of tumor cells and the stromal components led to false results.

These results indicate that although repeated measurements of the same samples can be performed with high reproducibility, it is unsure what other factors might cause variation in transcriptional assays. Until this has all been thoroughly investigated, clinicians should be aware that test results might not always be representative of epithelial cancer biology, and as such, critical correlation with the tumor subtype should be performed. One benefit of these analyses is that there is no interobserver variability, which has been shown in this thesis to be a major contributor to false-positive or negative results.

Prognostic power of gene-expression arrays

Compared to other categories of genes incorporated in gene-expression arrays, proliferation-associated genes are generally heavily weighted. This has led some to speculate that gene expression arrays are merely an expensive test for assessing the proliferative status of the tumor. Ki-67 status has been shown to be a strong, but not complete determinant of the Oncotype DX recurrence score [16]. Whether gene-expression arrays are truly superior to risk stratification based on traditional prognostication in this group will ultimately be decided in ongoing clinical trials, most notably the MINDACT [17] and the TAILORx [18] trials.

Several studies have however shown that the recurrence score can be predicted for a relatively high number of breast cancer patients by using conventional parameters [19-21], thereby reducing the number of patients where this assay has added value. Cuzick et al. compared the prognostic performance of the Oncotype DX recurrence score with that of the IHC4 score [22]. The latter is composed of four IHC markers (ER, PR, HER2 and Ki-67). The most informative algorithm for predicting distant metastases was determined and tested in a population of ER-positive patients treated in the Arimidex, Tamoxifen, Alone or in Combination trial (ATAC) trial. The IHC4 score was also added to the clinical score (determined by a formula incorporating nodal status, patient age, tumor size and tumor grade. Remarkably, this parameter had a similar prognostic power to the Oncotype DX [22]. These results suggest that improved implementation of parameters that are already commonplace might provide equally strong prognostic power.

IHC4 score

In an unpublished study, we investigated the prognostic value of the IHC4 score in the ER-positive patients from the POP cohort of node-negative, premenopausal patients. To our knowledge, this parameter has not been assessed in a younger population

of breast cancer patients, nor has it been correlated to the adjuvant therapy recommendations from the 2013 St. Gallen guidelines [23]. The expression of ER, PR, HER2 and Ki-67 was assessed as described previously [24]. The IHC4 score was determined as previously described [22, 25], including a correction of 0.4 for the Ki67 component due to use of non-automated analyses (IHC4=94.7 x (-0.100 ER10 – 0.079PgR10 + 0.586 HER2 + 0.240 ln (1+4 x Ki67)). The IHC4 values were similarly distributed as published by Cuzick et al. [22], with a median of -8.74 and a range of -102.17 to 147.15 (histogram shown in figure 1). The clinical score was determined in essence as previously determined, but without the components for nodal status (as all patients were node-negative), age (as all patients were below 65 in age) and anastrazol treatment (as no patients were treated with hormonal therapy, which resulted in the following formula: clinical score = 100 x (0.930 x (0.497 T1-2 + 0.882 T2-3 + 1.838 T> 3 + 0.559 Gr2 + 0.970 Gr3)). The IHC4 score was added to the clinical score for calculation of the IHC4+score. The tumors were stratified in tertiles according to the clinical score + IH4 score. No statistically significant survival period was found between the second and third tertile regarding overall survival and metastasis-free survival, and for the purpose of further analysis, these were combined as high-risk based on the IHC4+C score. The high-risk IHC4+C score tumors were associated with an increased hazard for metastasis-free survival and overall survival (figure 2AB). We then compared this score to the 2013 St. Gallen guidelines regarding the systematic treatment recommendations [23] which were applied to this dataset. For a number of 277 patients, both the 2013 St. Gallen criteria and IHC4+C score were available. Concordance regarding high and low risk was found in 73.3%. The most common discordant result were the 56 patients who were St. Gallen low-risk while IHC4+C score was considered high-risk. The opposite was true for 18 patients. Survival for these patients and the concordant cases are shown in table 1. These data suggest that both these prognostication schemes might provide complementary information, although this will need to be validated in larger patient cohorts. These results validate the IHC4+C score as a strong predictor for overall survival in this population of ER-positive patients.

The downside of this study was that none of the patients was treated with modern-day hormonal therapies or cytotoxic regimens. Patients were randomized to receive one course of perioperative chemotherapy. No apparent differences were observed when patients were stratified according to whether or not they had received chemotherapy. Due to the lack of current adjuvant therapies, the number of events was high in the studied population. These and other results do indicate that gene-expression arrays might not be necessary for achieving sufficient risk stratification, as long as IHC test are performed properly. Based on what is known on variability of

these IHC markers, stringent quality control should be undertaken to ensure that only quality assured tests are performed.

Targetable alterations

Significant advances are made in the increasing identification of targetable alterations in breast carcinomas and of course other cancers. These alterations include mutated protein kinases or amplified oncogenes that can be inhibited by either kinase-inhibitors or antibodies. These developments might lead to the situation where pathological diagnosis might be more about finding targets (as is of course partly true for ER and HER2) as opposed to risk assessment. Andre et al. presented a study where 423 patients with metastatic breast cancer were subjected to biopsy of the metastatic lesion, which was then analyzed using comparative genomic hybridization and gene sequencing. These techniques were feasible in 67% and 70% respectively. Targetable alterations were found in 46% of these patients [26]. One of the challenges for the coming years will be increase the yield for nucleic acid isolation from FFPE breast cancer biopsies to accommodate testing an increasing number of known mutations or other genetic alterations in order to guide targeted therapies. As our knowledge of cancer increases, more and more different genes are eligible for testing. Whether these can all be tested from core needle biopsies remains to be seen as well as whether such a complex infrastructure can be set up in pathology laboratories outside of academic hospitals while upholding strict demands regarding the quality of testing results [27].

Conclusion

The necessity of decreasing over- and undertreatment of breast cancer patients with chemotherapy has resulted in increasing efforts for finding both prognostic and predictive markers. A large number of cancer patients are treated with adjuvant chemotherapy while no risk of either locoregional or distant disease recurrence is present based on the natural history of their particular tumor. Although numerous papers on prognostic markers have been published on this subject, few are suitable for daily practice. This suitability depends on their reproducibility, prognostic power and technical feasibility. Careful selection of which patients and/or tumor types benefit from the incorporation of which specific biomarkers (and corresponding clinical action) is of vital importance to ensure clinical benefit.

As our knowledge of oncology steadily increases, so does our ability to diagnose, prognosticate and treat malignancy. Although this thesis deals with tumor characteristics,

novel insights regarding host factors (such as pharmacologic genotyping or phenotyping) have also resulted in clinically-applicable findings. Combining these data will lead to increasing personalization of cancer therapy, i.e. treatment tailored to characteristics specific for the individual tumor and patient. The necessity for personalized medicine to improve patient survival and quality of life has been universally accepted. One aspect of this is targeting oncogenic mechanisms that are essential for tumor cell survival thereby increasing treatment efficacy. On the other hand, this should also lead to the omission of unnecessary interventions that are unwarranted based on limited therapeutic response or the benign course of the disease.

Tissue will be the main supplier of information regarding the indication of novel targeted drugs. Strict conventional morphological will remain a vital part of the correct diagnosis of breast tumors and will be further supplemented with tumor-specific molecular tests in the form of IHC, in situ hybridization assays, gene-expression arrays and/or mutational analyses. Accurate testing of these novel markers for therapy poses additional challenges of testing accuracy that will have to be overcome in order to ensure optimal care for breast cancer patients.

Part II

The second part of this thesis discusses investigations concerning the tumor-associated stroma. Several issues were addressed in this section, including the validation of a stromal-derived prognostic parameter, investigation of the molecular content of the tumor-associated stroma and using stromal parameters and pathways for predicting disease progression and response to therapies.

In **chapter 10**, a validation study on the tumor-stroma ratio (TSR) is described. The TSR is an easily applicable, low-cost parameter that previously been shown as an independent predictor of patient survival in breast cancer [28] as well as in colon carcinoma [29] and esophageal carcinoma [30]. For this validation study, H&E-stained sections were investigated from patients that were treated as part of the perioperative chemotherapy trial (POP). The patients that were analyzed were a cohort of premenopausal, node-negative breast cancer patients (*N*=403). The tumor-stroma ratio was determined according to previously published methods [28]. The TSR was associated with disease-free survival in both univariate and multivariate analyses. Importantly, when analyzing this parameter alongside other clinico-pathological parameters, complementary prognostic information was provided. The prognostic power of this parameter was also verified in the subset of triple negative breast tumors in multivariate analysis.

The TSR is based on the morphological evaluation of the stromal compartment. This assessment might be improved by only assessing the amount of 'activated' stroma. We compared the proteomic signals from the extratumoral (supposedly quiescent) stroma with the proteomic signals from the intratumoral (activated) stromal tissues to discover novel markers for stromal activation in **chapter 11**. The proteomic signals were retrieved using matrix assisted laser ionisation mass spectrometry imaging (MALDI MSI). This technique allows for pixel-by-pixel proteomic analysis of frozen breast cancer tissues, which can then be correlated to tissue histology. This allows the user to select areas of interest (in our case intratumoral and extratumoral stroma) and perform direct proteomic comparisons between these. These experiments were performed as part of a multicenter study, with participation of the Helmholtz center in Munich and the Leiden University Medical Center. Three distinct proteomic signals were shown to correlate to the activated, intratumoral stroma in both datasets, among them a cleaved form of thymosin beta 4 and PA28. The localization of this last protein to the intratumoral stroma was verified using immunohistochemistry in an independent set of tumors. By using digital image analysis, upregulation of PA28 in activated stromal cells was confirmed.

MSI can be used to visualize signals in a wide range of masses, including the measurements of the smallest of molecules, metabolites. This application of this technique can be very valuable for studying molecular interactions between epithelial and stromal compartments of breast tumors. For this purpose, we set up a method for the detection of metabolic signatures and analysis of metabolic pathways in cancer in **chapter 12**. Breast cancer tissues were analyzed with the use of the FTICR-mass spectrometer. Data reduction was performed and the data were further refined by excluding all peaks that did not correspond to an entry in the metabocard database. Non-negative matrix factorization was performed to identify metabolic clusters, which were compared to the tissue architecture (tumor epithelium versus tumor stroma) to identify those metabolic clusters that were related to stromal tissue. The individual signals were related to their membership in respective metabolic pathways (according to the metabocard database) and as such, the analysis of metabolic pathways in the tumor-associated stroma was achieved. Potential metabolite-specific suppression might interfere with these investigations. In order to circumvent this problem, 13C-labeled isotopes were added to perform signal normalization. This method can be used in future studies to further characterize metabolic tumor-stroma interactions and can also be used to evaluate the efficacy of drugs that inhibit such interactions.

The TGF-β signaling pathway has been considered as a potent activator of the tumor-associated stroma. Both tumor-promoting and tumor-suppressing capabilities of this pathway are well described, which has translated to contrasting results published in the literature regarding the prognostic effects of TGF-β signaling. In **chapter 13**, we analyzed the expression of the TGF-β receptors, Smad4 and phopshorylated-Smad2 (pSmad2) in a series of 574 breast cancer patients. All these markers were individually related to disease free and overall survival. Because previous studies showed a central role for Smad4 in the functionality of the TGF-β pathway, all markers were investigated in both Smad4 low and high tumors. In the Smad4-high group, no significant relations were identified for the other markers with disease-free survival. However, in the Smad4 low group, strong prognostic power was found for the TGF-β receptors and pSmad2. Similar observations were detected regarding overall survival.

TGF-β has also been shown to influence the alignment of stromal tissues. Preclinical studies have shown that the orientation and organization of the stroma can hinder diffusion, thus possibly decreasing the effectivity of chemotherapy. In **chapter 14**, we investigated whether stromal organization might influence the effectivity of neoadjuvant chemotherapy administered as part of the Neozotac trial. This trial investigated the efficacy of bisphosphonates added to neoadjuvant TAC chemotherapy [31]. Stromal organization was assessed by drawing vectors parallel to stromal bundles observed within the tumor. The standard deviation of the angle of these vectors was analyzed as a measure for the organization of the stroma. In highly organized stroma, the stromal fibers are relatively aligned and the standard deviation would be relatively low in these cases. In case of disorganized, haphazardly aligned stromal fibers, this standard deviation is relatively high. These measures were positively related to both lymph node metastases *and* response to chemotherapy. This implies that tumors with aligned stroma are more likely to benefit from chemotherapy and also have an *increased* chance of developing locoregional metastases. This stromal organization was also related to active TGF-β signaling in the tumor-associated stroma (assessed as pSmad2-positive stromal cells).

Implications for future studies

The tumor-associated stroma has been clearly shown to provide essential contributions to tumor growth, progression and possibly initiation [32, 33]. Considering the role that the tumor-associated stroma plays in both local and distant spread of cancer cells as well as regulating the flow of drug molecules towards the tumor cells, stroma might (and quite possibly should) play a substantial role both in assessing patient prognosis and predicting response to therapy. The challenge for further research will

be combining the knowledge that we have gathered to enable a whole evaluation of the tumor-associated stromal tissues in each individual patient. Key will be assessing both the chance of distant relapse based on the stromal signature as well as determining whether the tissue microenvironment is amendable to pharmacotherapeutic intervention. Several clinical scenarios might be envisioned where a stromal-based prognostication scheme might prove useful. For instance, this might work to prevent the administration of unnecessary chemotherapy (e.g. in the case of a tumor where the ECM might so impenetrable for tumor cells that the risk of metastases is slim to none). Alternatively, the tumor-associated stroma might select patients that benefit from chemotherapy even though this was otherwise not warranted (e.g. an apparently low-grade tumor that has developed a stromal environment that actively contributes to either tumor cell dissemination or is so vulnerable to drug penetration that it warrants neoadjuvant chemotherapy prior to surgery). Although drug response is of course likely to be reliant on multiple additional factors, such as genomic mutations, pharmacodynamics, pharmacogenetics, and mechanisms that lead to multidrug resistance (MDR), stromal factors should be included in this conversation.

Quantitative stromal parameters

The prognostic significance of the TSR has been demonstrated in multiple studies in multiple tumor types. In this thesis, we have presented a validation study of this parameter in a cohort of breast cancer patients [34]. These results have also been validated in an independent cohort [35]. Clinical implementation will depend on identifying groups of breast tumors where this parameter has added value. The ER-negative/HER2-negative seems the most promising subgroup. Gene-transcriptional assays have modest prognostic significance in this subgroup [36], while stromal parameters have often shown the strongest prognostic significance in this group of tumors. Also, seemingly contrasting results have been published on the prognostic impact of this parameter in ER-positive tumors [37], although this was most likely due to differences in methodology [38].

Although reasonable interobserver kappa statistics have been reported in previous clinical studies, no interobserver studies have assessed the concordance between practicing pathologists. It is therefore unknown whether clinical implementation of this parameter would lead to acceptable reproducibility and standardization. In order to resolve this issue, such interobserver studies should be performed or reliable automated analysis should be developed. Studies have previously presented data regarding image-analysis based discrimination of the tumor stroma from the tumor epithelium. Bianconi et al. distinguished tumor epithelium from tumor stroma using

a classifier implementing support vector machines (SVM), nearest neighbour rule (1-NN) and naïve Bayes rule (NB) with good accuracy.

At the morphological level, sizeable differences in stromal appearance between different tumor areas can be observed. This might be reflective of variable degrees of stromal activation throughout the tumor and associated differences in functionality and outcome. Quantitative stromal measures might therefore be more powerful when strictly considering activated stromal tissues. In this thesis we presented results on the possible role of PA28 as a marker for activated stroma, but did not investigate whether increased expression of this marker in the stroma is correlated to clinical outcome. In an unpublished study, we investigated PA28 expression in a set of 50 invasive breast cancer patients and found a relationship between increasing histoscores of PA28 with disease relapse (figure 3). These data suggest that quantitative estimation of stromal activation can indeed provide strong prognostic information.

Besides using an individual IHC marker for assessing stromal activation, alternative procedures for detecting stromal activation have been described in the literature. Beck et al. distinguished morphological characteristics of high-risk and low-risk stromal tissues while using a digital pathology system [39]. Although this study employed TMA tissue cores, this technique might be incorporated into the tumor-stroma ratio parameter in selecting high-risk stromal areas. In this thesis, proteomic profiles detected with MALDI MSI were used to identify proteomic signatures of stromal activation. Only intratumoral with extratumoral stromal tissues were compared and no attempt was made to detect proteomic heterogeneity among activated stromal tissues. However, significant (and prognostically relevant) heterogeneity has been detected using MALDI MSI in otherwise morphologically similar epithelial components of breast cancers [40]. Similarly, these techniques might also be employed to identify such stromal heterogeneity and correlate these differences to both stromal morphology and functionality.

Stromal organization

Differences in stromal ECM arrangement have been observed in the tumor-associated stroma compared to physiologic stromal tissues. The peritumoral connective tissue has been shown to be more wavy compared to the intratumoral and juxtatumoral stroma [41]. Rearrangement into straight aligned ECM bundles might promote the occurrence of metastases [42]. TGF-β signaling has been related to this process in data from our group, which is presented in this thesis. Another possible contributing factor is syndecan-1, which has been shown to lead to parallel fiber organization

and increased tumor cell migration [43]. Caveolin-1, previously described as a strong prognostic factor in breast cancer, has also been related to stromal organization [44]. Data from our group has shown that tumors with highly aligned collagen bundles are more likely to present with lymph node metastases. Additionally, tumors with such aligned collagen are also more likely to respond to chemotherapy. These observations suggest that the same mechanism that contributed to the initial tumor cell dissemination might also sensitize the tumor to chemotherapy. These observations share some similarities with the tumor epithelium. High expression of markers Ki-67 and low expression of estrogen receptor (ER) have both been shown to be associated with disease progression. At the same time, tumors with high expression of these markers also have higher chance of undergoing complete pathological response (pCR) after neoadjuvant chemotherapy (which has been termed the "triple negative paradox" [45]).

Stromal organization in clinical studies has been assessed via multiple methods. Second harmonic generation (SHG) microscopy can be employed to enhance the contrast of collagen fibers with adjacent tissues to enhance visibility of these structures. This enhanced contrast allows for automated analysis that seems difficult to obtain when stromal alignment is analyzed with the use of H&E-stained slides [46]. However, we have shown that reasonable agreement can be achieved when H&E slides are analyzed in combination with freely available image analysis, which increases the feasibility of implementing this methodology into clinical practice. Differences in methodology and relative heterogeneity in study designs currently hamper the evidence for implementation of stromal organization into clinical practice despite encouraging preclinical results.

TILs

An international working group has recently issued a guideline aimed at increasing the uniformity for determining tumor-infiltrating lymphocytes (TILs) [47]. This working group states that most evidence exists for intra-stromal TILs as opposed to the intratumoral lymphocytes. Intratumoral lymphocytes are more difficult to assess and generally seem to correlate to the presence of stromal lymphocytes anyway. For assessing the stromal-TIL%, the percentage of stromal areas that contain TILs should be assessed over the entire intratumoral stromal area. No evidence exists that justifies scoring TIL hotspots or simply scoring the tumor border area. As the guidelines correctly states, no studies have been performed that have assessed the intra- and interobserver variance that exists for this assessment. Despite this, the published

guidelines at least provide a framework on which future studies can be designed and performed in order to provide the strongest evidence for these parameters.

However, the most important question remains, is there evidence for altering clinical decision making based on the presence of TILs? In the HER2 subgroup, the presence of stromal TILs can be used to predict the response to trastuzumab therapies. However, considering the integral role that trastuzumab currently plays in the treatment of HER2-positive breast tumors, and the adverse prognosis that this group of tumors has without therapy of this antibody, omitting this therapy in HER2-positive breast tumors seems ill-advised at this point.

Targeting the tumor-associated stroma

Targeting components of the tumor-associated stroma has been investigated in both pre-clinical and clinical trials. Due to the multifaceted role of CAFs in cancer progression, targeting these cells seems an attractive option for improving breast cancer care. Another attractive aspect of targeting the stromal compartment of breast tumors is its perceived genetic stability (thereby decreasing the risk of creating resistant cell populations) and homogeneity of stromal cells.

The most commonly tested treatment strategy is inhibition of angiogenesis, which has been investigated in numerous trials that have generally shown modest results [48, 49].

Fibroblast activation protein α (FAP) is expressed by tumor-associated fibroblasts [50] and has also been studied for targeted therapies. Vaccination for this protein in preclinical cancer models has been shown to reduce tumor growth and reduce the collagen density within breast tumors, thereby increasing tumor uptake of chemotherapy [51]. The findings have led to clinical trials exploring the efficacy of sibrotuzumab, an antibody directed at the FAP protein [52]. No objective responses were observed in one phase I trial in 26 cancer patients. However, analysis of the distribution of the administered antibody showed that it only distributed to the tumor area without accumulating in other organs [53]. This selectivity and proteolytic activity of FAP itself has been exploited by one study, which administered a prodrug activated by the enzymatic activity of this marker [54]. Xenograft models showed strong inhibition of tumor growth while minimal systemic toxicity occurred, as FAP is not expressed in other cell types.

Although FAP remains the most commonly investigated target, other options for normalizing tumor-associated stromal functionality have been described. In this thesis, we showed a relationship between the stromal alignment and active phospho-Smad2 staining, which indicates that active TGF-β signaling leads to aligned stromal network thereby increasing tumor susceptibility to chemotherapy. Inhibiting this signaling pathway would thus theoretically reduce chemotherapy vulnerability. However, the opposite result has been found in a preclinical study [55]. Liu et al. showed that TGF-β blockade *improved* the distribution of cytotoxic agents in murine breast cancer models in part by reducing collagen type I content [55]. Matrix orientation was not accounted for in this study and it would have been interesting to assess this parameter before and after treatment with TGF-β inhibiting therapies. The effect seen in this study might be attributed to the stabilizing effect that these therapies have on newly formed blood vessels thereby improving perfusion of the tumor. However, these results do illustrate that thorough studies on the peritumoral ECM structure might be used to identify the ideal recipients of TGF-β inhibiting medications. Multiple strategies for inhibiting this pathway have been described in phase I and phase II studies which have included various solid tumors [56]. To our knowledge, no trials have specifically included breast cancer patients. Attention should be given to results published in chapters 13 and 14 of this thesis, in order to predict which patients are most likely to benefit from these treatment strategies.

Conclusion

When translating stromal biomarkers into clinical practice, a combination of both quantitative and qualitative stromal markers might provide strong prognostication schemes for ER-negative breast cancer. Published series regarding stromal parameters are mostly from retrospective studies studying a single biomarker. Randomized trials specifically designed to measure biomarker applicability are rare and have not yet been performed for stromal biomarkers. The highest level of evidence for a stromal biomarker is therefore unavailable. Evidence from retrospective series should guide further studies to clarify which stromal markers have strong clinical merit. Ideally, future studies will incorporate multiple stromal biomarkers in large prospective series to identify clinical scenarios where a stromal-based classification system has additional value to the traditional epithelial-based risk stratification.

The potential benefits of targeting components of the stromal tissues have mostly been shown in preclinical studies, while modest results have been shown in clinical trials. Although the perceived homogeneity is often cited as one of the most attractive aspects of targeting the stroma, significant differences in stromal quantity,

molecular marker expression and ECM orientation have been shown in many studies. Predicting which patients are to benefit from which stromal interventions will require careful investigation in order to decide which patients will receive most benefit from stromal-targeted therapies.

References

1. Berry DA et al.: Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005, 353:1784-1792
2. Wolff AC et al.: American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 2007, 131:18-43
3. Clahsen PC et al.: Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer. A European Organization for Research and Treatment of Cancer Breast Cancer Cooperative Group Study. J Clin Oncol 1996, 14:745-753
4. Drukker CA et al.: A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int J Cancer 2013, 133:929-936
5. Roepman P et al.: Microarray-based determination of estrogen receptor, progesterone receptor, and HER2 receptor status in breast cancer. Clin Cancer Res 2009, 15:7003-7011
6. Viale G et al.: High concordance of protein (by IHC), gene (by FISH; HER2 only), and microarray readout (by TargetPrint) of ER, PgR, and HER2: results from the EORTC 10041/BIG 03-04 MINDACT trial. Ann Oncol 2014, 25:816-823
7. Kraus JA et al.: Semi-quantitative immunohistochemical assay versus oncotype DX(®) qRT-PCR assay for estrogen and progesterone receptors: an independent quality assurance study. Mod Pathol 2012, 25:869-876
8. Paik S et al.: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004, 351:2817-2826
9. van de Vijver MJ et al.: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002, 347:1999-2009
10. Denkert C et al.: Decentral gene expression analysis for ER+/Her2- breast cancer: results of a proficiency testing program for the EndoPredict assay. Virchows Arch 2012, 460:251-259
11. Ach RA et al.: Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. BMC Genomics 2007, 8:148
12. Cronin M et al.: Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. Clin Chem 2007, 53:1084-1091
13. Klang SH, Het al: Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed health-care organization. Value Health 2010, 13:381-387
14. Barry WT et al.: Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome. J Clin Oncol 2010, 28:2198-2206
15. Acs G et al.: Comparison of Oncotype DX and Mammostrat risk estimations and correlations with histologic tumor features in low-grade, estrogen receptor-positive invasive breast carcinomas. Mod Pathol 2013, 26:1451-1460

16. Sahebjam S et al.: Ki 67 is a major, but not the sole determinant of Oncotype Dx recurrence score. Br J Cancer 2011, 105:1342-1345
17. Cardoso F et al.: Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol 2008, 26:729-735
18. Sparano JA: TAILORx: trial assigning individualized options for treatment (Rx). Clin Breast Cancer 2006, 7:347-350
19. Allison KH et al.: Routine pathologic parameters can predict Oncotype DX recurrence scores in subsets of ER positive patients: who does not always need testing? Breast Cancer Res Treat 2012, 131:413-424
20. Ingoldsby H et al.: Prediction of Oncotype DX and TAILORx risk categories using histopathological and immunohistochemical markers by classification and regression tree (CART) analysis. Breast 2013, 22:879-886
21. Geradts J et al.: The oncotype DX recurrence score is correlated with a composite index including routinely reported pathobiologic features. Cancer Invest 2010, 28:969-977
22. Cuzick J et al.: Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol 2011, 29:4273-4278
23. Goldhirsch A et al.: Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013, 24:2206-2223
24. Clahsen PC et al.: p53 protein accumulation and response to adjuvant chemotherapy in premenopausal women with node-negative early breast cancer. J Clin Oncol 1998, 16:470-479
25. Barton S et al.: Assessment of the contribution of the IHC4+C score to decision making in clinical practice in early breast cancer. Br J Cancer 2012, 106:1760-1765
26. Andre F et al.: Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/ UNICANCER). Lancet Oncol 2014, 15:267-274
27. Tursz T, Bernards R: Hurdles on the road to personalized medicine. Mol Oncol 2015, 9:935-939
28. de Kruijf EM et al.: Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. Breast Cancer Res Treat 2011, 125:687-696
29. Huijbers A et al.: The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. Ann Oncol 2013, 24:179-185
30. Courrech Staal EF et al.: Reproducibility and validation of tumour stroma ratio scoring on oesophageal adenocarcinoma biopsies. Eur J Cancer 2011, 47:375-382

31. Charehbili A et al.: Addition of zoledronic acid to neoadjuvant chemotherapy does not enhance tumor response in patients with HER2-negative stage II/III breast cancer: the NEOZOTAC trial (BOOG 2010-01). Ann Oncol 2014, 25:998-1004
32. Ostman A, Augsten M: Cancer-associated fibroblasts and tumor growth – bystanders turning into key players. Curr Opin Genet Dev 2009, 19:67-73
33. Orimo A, Weinberg RA: Stromal fibroblasts in cancer: a novel tumor-promoting cell type. Cell Cycle 2006, 5:1597-1601
34. Dekker TJ et al.: Prognostic significance of the tumor-stroma ratio: validation study in node-negative premenopausal breast cancer patients from the EORTC perioperative chemotherapy (POP) trial (10854). Breast Cancer Res Treat 2013, 139:371-379
35. Gujam FJ et al.: The relationship between the tumour stroma percentage, clinicopathological characteristics and outcome in patients with operable ductal breast cancer. Br J Cancer 2014, 111:157-165
36. Wirapati P et al.: Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008, 10:R65
37. Downey CL et al.: The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. Br J Cancer 2014, 110:1744-1747
38. Mesker WE et al.: Comment on: The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. Br J Cancer 2014
39. Beck AH et al.: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011, 3:108ra113
40. Balluff B et al.: De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. J Pathol 2015, 235:3-13
41. Brabrand A et al.: Alterations in collagen fibre patterns in breast cancer. A premise for tumour invasiveness? APMIS 2015, 123:1-8
42. Provenzano PP et al.: Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med 2006, 4:38
43. Yang N et al.: Syndecan-1 in breast cancer stroma fibroblasts regulates extracellular matrix fiber organization and carcinoma cell motility. Am J Pathol 2011, 178:325-335
44. Goetz JG et al.: Biomechanical remodeling of the microenvironment by stromal caveolin-1 favors tumor invasion and metastasis. Cell 2011, 146:148-163
45. Carey LA et al.: The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 2007, 13:2329-2334
46. Bredfeldt JS et al.: Automated quantification of aligned collagen for human breast carcinoma prognosis. J Pathol Inform 2014, 5:28
47. Salgado R et al.: The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol 2015, 26:259-271

48. Miller K et al.: Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med 2007, 357:2666-2676
49. Robert NJ et al.: RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2-negative, locally recurrent or metastatic breast cancer. J Clin Oncol 2011, 29:1252-1260
50. Garin-Chesa P et al.: Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers. Proc Natl Acad Sci U S A 1990, 87:7235-7239
51. Loeffler M et al.: Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake. J Clin Invest 2006, 116:1955-1962
52. Scott AM et al.: A Phase I dose-escalation study of sibrotuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. Clin Cancer Res 2003, 9:1639-1647
53. Scott AM et al.: A Phase I dose-escalation study of sibrotuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. Clin Cancer Res 2003, 9:1639-1647
54. Brennen WN et al.: Targeting carcinoma-associated fibroblasts within the tumor stroma with a fibroblast activation protein-activated prodrug. J Natl Cancer Inst 2012, 104:1320-1334
55. Liu J et al.: TGF-beta blockade improves the distribution and efficacy of therapeutics in breast carcinoma by normalizing the tumor stroma. Proc Natl Acad Sci U S A 2012, 109:16618-16623
56. Hawinkels LJ, Ten Dijke DP: Exploring anti-TGF-beta therapies in cancer and fibrosis. Growth Factors 2011, 29:140-152

Figure 1. Distribution of the IHC4 score in 331 premenopausal, node-negative breast cancer patients.

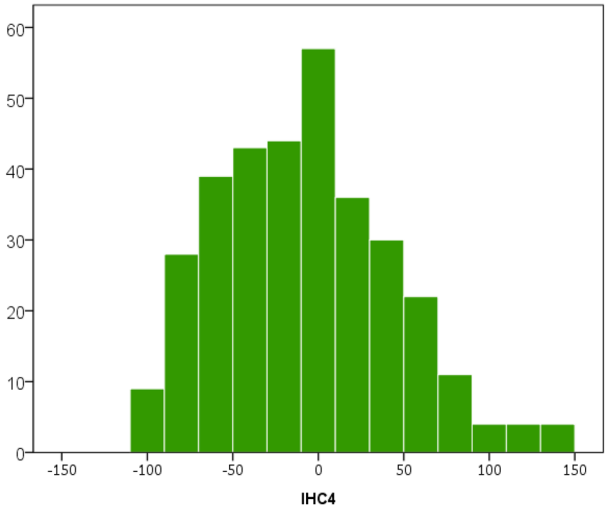

Figure 2. Metastasis-free (A) and overall survival (B) according to the lower and upper two tertiles of the IHC4+C score. Green = lowest tertile, blue = intermediate tertile, red = upper tertile.

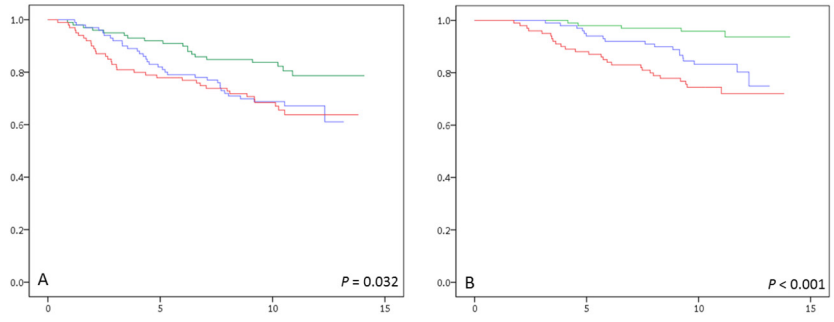

Figure 3. Stromal expression of PA28 in breast cancer. (A) An example of a breast carcinoma that displays PA28 expression in fibroblastic and immune cells of the breast cancer stroma. (B) An example of a breast carincoma that shows little PA28 expression, although some fibroblastic and immune cells are positive for this marker. (C) Disease free survival (in months) of 50 invasive breast cancer patients according to expression level of PA28 in the tumor-associated stroma (*P*=0.021). Green = low expression, blue = intermediate expression, red = high expression.

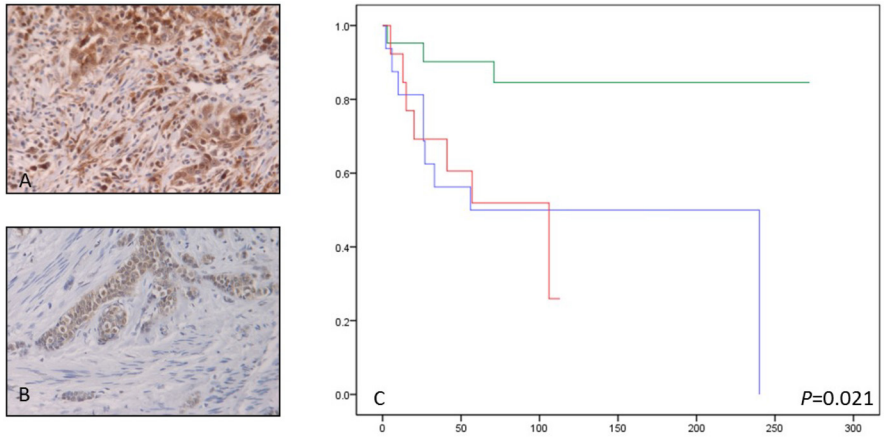

Table 1. Using the St. Gallen guidelines and IHC4+C score risk assessments in combination to identify patients at low risk (LR) and high risk (HR) for the occurrence of metastases and/or death.

Subgroup	10-year MFS	Mean MFS, in years (95% CI)	Hazard for metastasis (95% CI)	10-year OS	Mean OS, in years (95% CI)	Hazard for death (95% CI)
StG LR / IHC4+C LR (*N*=80)	86%	12.5 (11.7-13.3)	1.000	99%	13.9 (13.7-14.1)	1.000
StG LR / IHC4+C HR (*N*=56)	80%	11.8 (10.8-12.8)	1.320 (0.620-2.807)	91%	13.098 (12.5-13.7)	3.631 (0.704-18.718)
StG HR / IHC4+C LR (*N*=18)	78%	10.6 (9.0-12.1)	1.428 (0.470-4.343)	83.3%	11.2 (10.0-12.3)	7.757 (1.294-46.509)
StG HR / IHC4+C HR (*N*=123)	67%	10.1 (9.3-10.9)	2.370 (1.298-4.326)	76%	11.3 (10.7-11.9)	12.059 (2.888-50.342)