BMC Cancer

Comparison of immunohistochemistry with PCR for assessment of ER, PR, and Ki-67 and prediction of pathological complete response in breast cancer

Hans-Peter Sinn[1*], Andreas Schneeweiss[2], Marius Keller[1], Kornelia Schlombs[6], Mark Laible[6], Julia Seitz[2], Sotirios Lakis[4], Elke Veltrup[4], Peter Altevogt[3], Sebastian Eidt[5], Ralph M. Wirtz[4,5] and Frederik Marmé[2]

Abstract

Background: Proliferation may predict response to neoadjuvant therapy of breast cancer and is commonly assessed by manual scoring of slides stained by immunohistochemistry (IHC) for Ki-67 similar to ER and PgR. This method carries significant intra- and inter-observer variability. Automatic scoring of Ki-67 with digital image analysis (qIHC) or assessment of MKI67 gene expression with RT-qPCR may improve diagnostic accuracy.

Methods: Ki-67 IHC visual assessment was compared to the IHC nuclear tool (AperioTM) on core biopsies from a randomized neoadjuvant clinical trial. Expression of ESR1, PGR and MKI67 by RT-qPCR was performed on RNA extracted from the same formalin-fixed paraffin-embedded tissue. Concordance between the three methods (vIHC, qIHC and RT-qPCR) was assessed for all 3 markers. The potential of Ki-67 IHC and RT-qPCR to predict pathological complete response (pCR) was evaluated using ROC analysis and non-parametric Mann-Whitney Test.

Results: Correlation between methods (qIHC versus RT-qPCR) was high for ER and PgR (spearman´s $r = 0.82$, $p < 0.0001$ and $r = 0.86$, $p < 0.0001$, respectively) resulting in high levels of concordance using predefined cut-offs. When comparing qIHC of ER and PgR with RT-qPCR of ESR1 and PGR the overall agreement was 96.6 and 91.4%, respectively, while overall agreement of visual IHC with RT-qPCR was slightly lower for ER/ESR1 and PR/PGR (91.2 and 92.9%, respectively). In contrast, only a moderate correlation was observed between qIHC and RT-qPCR continuous data for Ki-67/MKI67 (Spearman's $r = 0.50$, $p = 0.0001$). Up to now no predictive cut-off for Ki-67 assessment by IHC has been established to predict response to neoadjuvant chemotherapy. Setting the desired sensitivity at 100%, specificity for the prediction of pCR (ypT0ypN0) was significantly higher for mRNA than for protein (68.9% vs. 22.2%). Moreover, the proliferation levels in patients achieving a pCR versus not differed significantly using MKI67 RNA expression (Mann-Whitney $p = 0.002$), but not with qIHC of Ki-67 (Mann-Whitney $p = 0.097$) or vIHC of Ki-67 ($p = 0.131$).

Conclusion: Digital image analysis can successfully be implemented for assessing ER, PR and Ki-67. IHC for ER and PR reveals high concordance with RT-qPCR. However, RT-qPCR displays a broader dynamic range and higher sensitivity than IHC. Moreover, correlation between Ki-67 qIHC and RT-qPCR is only moderate and RT-qPCR with MammaTyper® outperforms qIHC in predicting pCR. Both methods yield improvements to error-prone manual scoring of Ki-67. However, RT-qPCR was significantly more specific.

Keywords: Image analysis, Breast cancer, Ki67, mRNA, RT-qPCR, Prediction, Pathologic complete response, neoadjuvant, Immunohistochemistry (IHC), MammaTyper®

* Correspondence: peter.sinn@med.uni-heidelberg.de
[1]Institute of Pathology, University of Heidelberg, Im Neuenheimer Feld 220-221, 69120 Heidelberg, Germany
Full list of author information is available at the end of the article

Sinn *et al. BMC Cancer* (2017) 17:124

Page 2 of 10

Background

The proliferative activity of individual cells is a hallmark of tumor biological aggressiveness and a key determinant of sensitivity to (neo)adjuvant chemotherapy, thus being among the principal factors guiding clinical management in primary breast cancer [1, 2]. The most widely used method to assess proliferation as well as hormone receptor expression is immunohistochemistry (IHC).

Nuclear staining of the nuclear antigen Ki-67 is most widely used as a surrogate for proliferative activity. Ki-67 is present in the cell nucleus throughout all stages of the cell-cycle excluding the resting phase G0 [3]. The recently proposed St Gallen recommendations for the identification of the intrinsic subtypes using surrogate pathologic-based definitions have underlined the value of Ki-67 as a clinical tool in routine clinical practice [2]. Ki-67 is recommended as a valuable factor to distinguish between Luminal A- and B-like tumors, a fundamental distinction in clinical decision-making today. [4–6].

Despite its widespread use, driven by the premise of solving delicate therapeutic dilemmas combined with several advantages such as universal accessibility, easy application and low cost, the assessment of Ki-67, ER and PR is affected by technical and observer-based variabilities of the IHC method [7, 8]. This can be illustrated by observations, that tumors with as little as 1% positive nuclei still respond to anti-hormonal treatment, which indicates that tumor cells lacking nuclear ER staining within in these tumors do have some extend of ER expression rendering them sensitive to estrogen deprivation or estrogen receptor blockade [9]. While the clinical role particularly of ER testing by IHC is well established, the clinical utility of Ki-67 is still controversial [10]. The reason lies in a series of analytical and preanalytical factors, but also in staining interpretation and scoring [11]. Importantly, attempts to reduce the high discordance rates either by means of formal counting quantification methods (as opposed to simple eyeballing) or by training of individuals have not been successful [12].

Despite methodological concerns, overall a strong correlation of Ki-67 with breast cancer outcome is sufficiently supported, particularly by data originating from randomized clinical trials with central review of biomarkers [10]. This has also been shown in the neoadjuvant setting, where higher Ki-67 values are consistently associated with higher rates of pathological complete response (pCR) [13], a finding which reflects the fundamental link between tumor replication fraction and activity of cytotoxic agents. Still it remains difficult to identify a reasonable cut-off to predict pCR [14].

Two techniques which could circumvent the inter- or intra-observer variability of Ki-67 manual microscopic assessment are automated image analysis and reverse transcription quantitative real-time PCR (RT-qPCR). A trained human eye may achieve an excellent understanding of images and patterns, but is less accurate when it comes to quantification. Computer-based vision methods could represent a solution to this problem by offering standardized image processing and reliable quantification [15]. However with regard to the assessment of tumor proliferation, the areas of interest and staining intensities have to be defined, and measured reproducibly. Also, it is still under debate how to deal with areas of increased proliferative activity (hot spots), and if low intensity staining should be taken into account [16]. Therefore, automated estimations of Ki-67 are highly correlated with manual assessments, but it is not yet certain whether or not they can improve prediction and prognostication [17–19].

RT-qPCR has a series of widely acknowledged methodological advantages over IHC, which appear particularly beneficial in the context of reducing the bias of routine Ki67 assessment; it is quantitative by nature with much wider dynamic range, it does not require an experienced eye, and results are not affected by subjective interpretations [20]. Moreover, access to standardized protocols and automation ensures accurate performance and fast turn-around. In recent years highly specific and sensitive techniques have been developed, which allow for fast and efficient extraction of high-quality nucleic acids from FFPE overcoming the challenges posed by fixation and embedding [21].

Validation of the various available methods for the assessment of Ki-67 requires comparative testing preferably in a specifically defined clinical context. In the present study we used the neoadjuvant setting of a phase II trial randomizing patients receiving anthracyclin/taxane based standard treatment between pemetrexed and cyclophosphamide in order to directly compare the assessment of Ki-67 with automatic, quantitative read-out of IHC (qIHC) and the determination of tumor MKI67 mRNA with RT-qPCR on FFPE tissue extracted RNA.

Methods

Study population

Core needle biopsies from 101 out of 105 patients (96,2%) with primary invasive breast cancer, that had been enrolled in the H3E-MC-S080 (NCT00149214, Sponsor: Eli Lilly and Company) neoadjuvant phase II study [1], were obtained. All patients had been diagnosed with operable (T2-T4/N0-2/M0) breast cancer at a single institution (National Center for Tumor Diseases, University-Hospital, Heidelberg) had been randomized to receive sequential anthracycline/taxane-based regimens containing either pemetrexed or cyclophosphamide in combination with epirubicin. A written informed consent for the research

Sinn *et al. BMC Cancer* (2017) 17:124

Page 3 of 10

use of patient biological material was granted at the time of enrolment. The study was approved by the local ethics committee. Complete molecular data (including RT-qPCR data) and clinical follow-up information were available in 83 out of 105 (79%) patients (statistics data set #1). Ki-67, ER, and PR IHC slides were available in 54 (51%) patients for quantitative IHC (statistics data set #2).

Isolation of tumor RNA

For RNA extraction from FFPE tissue, a single 10 µm curl was processed according to a commercially available bead-based extraction method (RNXtract® kit; BioNTech Diagnostics GmbH, Mainz, Germany). In brief, a lysis buffer was used to liquefy FFPE tissue slices while melting of paraffin was carried out in a thermo-mixer. Tissue lysis was accomplished with a proteinase K solution. Thereafter, lysates were admixed with germanium-coated magnetic particles in the presence of special buffers, which promote the binding of nucleic acids. Purification was carried out by means of consecutive cycles of mixing, magnetization, centrifugation and removal of contaminants. RNA was eluated with 100 µl elution buffer and RNA eluates were then stored at −80 °C until use.

Gene expression by RT-qPCR

The MammaTyper® is a molecular in vitro diagnostic tool for the assessment of the gene expression levels of the four cancer biomarkers that are required for the clinical management of breast cancer patients in daily routine clinical practice. Instead of using IHC to assess protein expression of HER2, ERα, PR, and Ki-67, with MammaTyper®, it is possible to measure the mRNA transcripts of the corresponding genes (ERBB2, ESR1, PGR, and MKI67), doing so by using routine FFPE material and by achieving accurate, reproducible and objective results. The gene expression data may be then integrated so as to assign individual samples to a molecular subtype of breast cancer.

The mRNA expression levels of ERBB2, ESR1, PGR, and MKI67 as well as of two reference genes (REF), namely B2M and CALM2, were determined by RT-qPCR, which involves reverse transcription of RNA and subsequent amplification of cDNA executed successively as a 1-step reaction. In MammaTyper®, the 6 assays (assay = primer pair and probe specific for the respective target sequence) are duplexed into three assay mixes, each using a pair of hydrolysis probes labelled with different fluorophores for separate detection of the duplexed assays [22].

Each patient sample or control was analyzed with each assay mix in triplicates. The experiments were run on a Versant kPCR Molecular System (Siemens Healthcare, Erlangen, Germany) according to the following protocol:

5 min at 50 ° C, 20 sec at 95 ° C followed by 40 cycles of 15 sec at 95 ° C and 60 s at 60 ° C and according to MammaTyper® instructions for use 140603-90020-EU Rev 2.0.

Forty amplification cycles were applied and the cycle quantification threshold (Cq) values of MKI67 and the two REF genes for each sample (S) were estimated as the median of the triplicate measurements. These were then normalized against the mean expression of the REF genes and set off against a calibrator (PC), to correct for inter-run variations (ΔΔCq method) (Livak et al. 2001). The final values were generated by subtracting ΔΔCq from the total number of cycles to ensure that normalized gene expression obtained by the test is proportional to the corresponding mRNA expression levels, a method that facilitates interpretation of data and clinicopathological correlations. The various calculation steps are summarized in the following formula:

$$40\text{-}\Delta\Delta Cq(MKI67)S = 40\text{-}((Cq[MKI67]S - meanCq[REF]S) - (Cq[MKI67]pc - meanCq[REF]pc))$$

In 18 patients the MammaTyper® assay failed, because the required amount of RNA was not sufficient for analysis according to pre-specified criteria as described in the instructions for use.

Pathology and Immunohistochemistry

Tumor grading, tumor typing and immunohistochemistry (ER, PR, Ki-67) was performed on the pretreatment core biopsies on all patients. Pathological complete response (pCR) was determined on tumor resection specimens after completion of neoadjuvant chemotherapy, and was defined as no evidence of residual invasive and ductal disease in the breast and lymph nodes (ypT0,ypN0).

Immunohistochemistry was performed according to previously standardized protocols on an automated IHC platform (Dako Techmate 500) with citrate buffer for antigen retrieval [23] and observing the ASCO/CAP guidelines for immunohistochemistry [7]. The following primary antibodies and corresponding dilutions were used (DakoCytomation, Glostrup, Denmark): ER (clone 1D5, 1:100), PR (clone PgR636, 1:100) and Ki-67 (MIB-1, 1:200). Slides were assessed by quantitative image analysis (qIHC) using the Aperio Image Analysis toolbox (Leica Biosystems, Nussloch, Germany). Staining intensity and percentage of positive nuclei were recorded after manually segmenting tumor from adjacent stroma. Tumors with ER/PR Remmele scores greater than 3 or positive nuclei greater than 1% were considered hormone receptor positive.

Sinn *et al. BMC Cancer* (2017) 17:124

Page 4 of 10

Statistical methods

The Spearman correlation coefficient r was used as a measure of the strength and direction of the linear relationship between variables. 2×2 contingency tables were used to calculate positive percent agreement (PPA) and negative percent agreement (NPA) as a measure of agreement: $PPA = 100\% \times a/(a + c)$, $NPA = 100\% \times d/(b + d)$. Receiver Operating Characteristics (ROC) analysis was performed to determine the optimal cut-off for Mamma-Typer® gene and qIHC protein measurements with pCR as the endpoint. ROC analysis instead of comparing odds ratios to take into account the ratios of clinically relevant false positive and false negative determinations and to identify cut points for each method at clinically relevant prerequisites (i.e., detect all responding tumours). ROC analysis has been used to objectively address each method providing different result codings in a non-parametric manner [24]. On the other hand ROC analysis bears the risk of misinterpreting clinical validity when analyzing heterogeneous populations [25]. However, we have analyzed the response to neoadjuvant chemotherapy within controlled, randomized phase II trial which has defined inclusion and exclusion criteria to have the most comparable basic risk situation. As optimal cut-off for the identification of complete response by the methodologies the point of highest sensitivity still retaining 100% specificity was chosen. The p value reported for evaluating the ROC curve tests the null hypothesis that the area under the curve really equals 0.50 as provided by the statistical program used (GraphPad Prism). The non-parametric Mann-Whitney test was used to confirm the statistical significance when comparing responding versus non-responding tumors and box plots were used to illustrate each case of responding and non-responding tumor above and below the cut-off value. Statistical analyses were performed with JMP SAS (SAS Institute, Cary, NC, USA) and Graph Pad Prism software (Version 5.04; Graph Pad Software Inc., La Jolla, CA, USA).
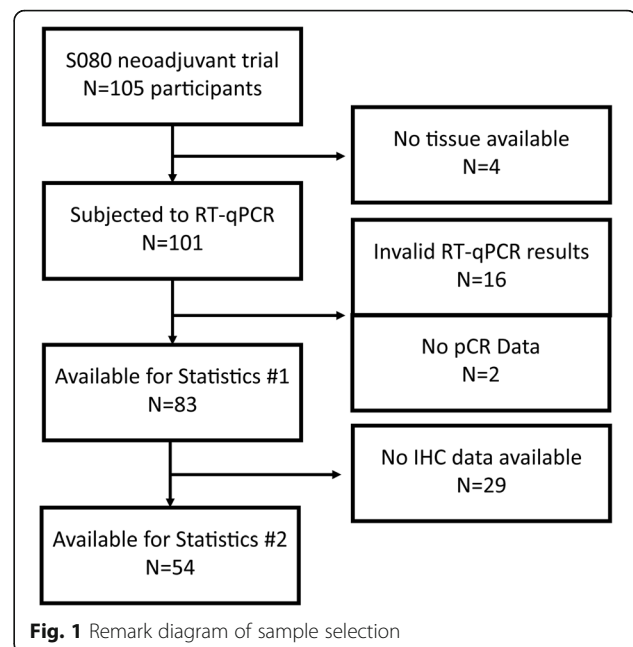
Results

Patient population

Biopsy tissue was available from 101 out of 105 patients. Gene expression analysis by MammaTyper® was successful in 83 biopsy specimens with full clinical data out of a total of 105 trial participants (Fig. 1). 12 patients out of this group had achieved complete pathological remission (pCR). Basic clinicopathological characteristics of statistics data set #2 that includes quantitative IHC data is listed in Table 1.

Comparison of IHC with RT-qPCR based assessment of ER, PR and Ki-67

Comparing qIHC of ER with RT-qPCR of ESR1 demonstrated a good overall agreement of 96,6% (PPA

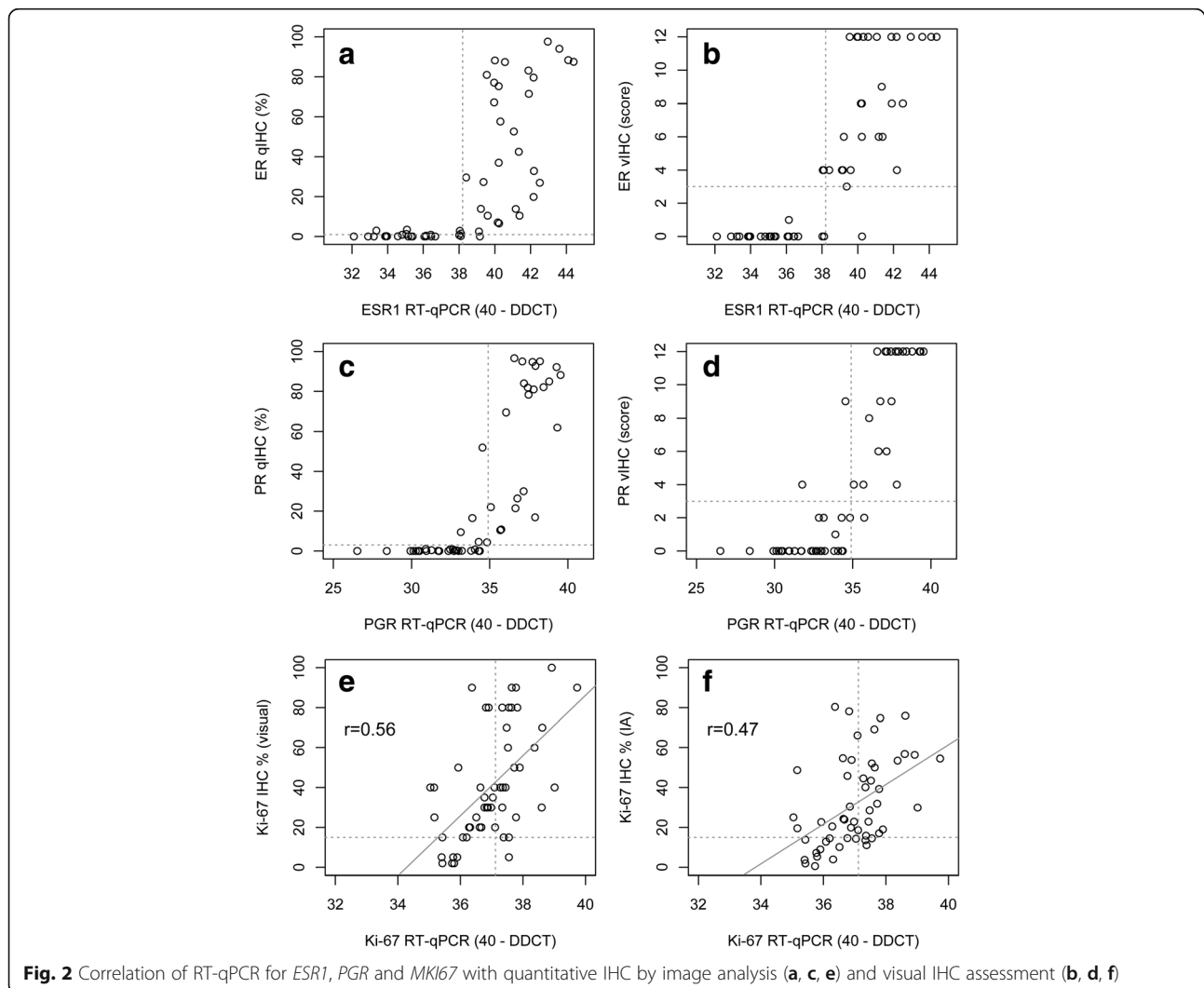**Fig. 1** Remark diagram of sample selection

100%; NPA 92.3%) as well as a good correlation looking at the continuous data (spearman's $r = 0.82$, $p < 0.0001$). Correlation between vIHC of ER and RT-qPCR for ESR1 (spearman's $r = 0.85$, $p < 0.0001$) and between vIHC and qIHC ER (spearman's $r = 0.88$, $p < 0.0001$) was high, too.

Table 1 Basic tumor characteristics

	no-pCR		pCR		
Histological Type					*p*= 0.39
Invasive ductal NOS	37	82.2%	9	100.0%	
Invasive lobular	7	15.6%	0	0.0%	
Other	1	2.2%	0	0.0%	
Histological Grade					*p*= 1.00
Grade 2	15	33.3%	3	33.3%	
Grade 3	30	66.7%	6	66.7%	
ypT Category					*p*= 0.81
ypT0	0	0.0%	9	100.0%	
ypT1	27	50.0%	0	0.0%	
ypT2	9	20.0%	0	0.0%	
ypT3	8	17.8%	0	0.0%	
ypT4	1	2.0%	0	0.0%	
ypN Category					*p*= 0.81
ypN0	27	50.0%	9	100.0%	
ypN+	25	46.3%	0	0.0%	
ypNX	2	3.7%			
HER2 status					*p*= 0.68
Neg	36	66.7%	7	13.0%	
Pos	7	13.0%	2	3.7%	

Overall agreement for PR protein and PGR mRNA expression was 91.4% (PPA 83.3%; NPA 100%) comparing qIHC and RT-qPCR and there was a high correlation for the continuous data ($r = 0.86$, $p < 0.0001$). Correlation between vIHC and RT-qPCR and between vIHC and qIHC was very high, too ($r = 0.88$, $p < 0.0001$ and $r = 0.90$, $p < 0.0001$, respectively). Concordance when comparing visual IHC protein with RT-qPCR RNA expression was good for ESR1 as well as for PGR, although slightly lower compared to the agreement between qIHC and RT-qPCR (OPA 91.2%; PPA 90.9%; NPA 91.7%) and for PGR (OPA 92.9%; PPA 88.0%; NPA 96.9%, respectively). For both, ESR1 and PGR, only 4 cases were discordant, with 3 cases each positive by vIHC and negative by RT-qPCR, while 1 case was negative by vIHC and positive by RT-qPCR (Fig. 2). However, several of these discrepancies could be resolved by using quantitative IHC, as these cases were also discrepant when comparing vIHC with qIHC. Moreover, qIHC could delineate quantitative differences of

hormone receptor expression at the highest Remmele Score value of 12, where vIHC could not resolve expression differences. In addition, at the lower range of expression levels RT-qPCR based assessment could still determine substantial differences of mRNA levels while the IHC based assessment could not detect any protein expression. The inter-gene spearman correlation was moderate for ESR1 and PGR ($r = 0.59$, $p < 0.0001$), while Ki-67 correlated negatively with PGR (-0.37, $p = 0.007$). While the correlation for ESR1 and PGR protein and RNA expression was high when comparing IHC with RT-qPCR results (spearman's $r = 0.82$, $p < 0.0001$ and $r = 0.86$, $p < 0.0001$, respectively), the correlation between MKI67 protein and RNA expression was only moderate (spearman's $r = 0.56$ for vIHC and $r = 0.47$ for qIHC). In contrast the correlation between both methods for Ki-67 assessment by IHC was high ($r = 0.80$, $p < 0.0001$). The median value of Ki-67 proliferation index by image analysis IHC (qIHC) was 23.4%, by conventional visual IHC (vIHC)

Fig. 2 Correlation of RT-qPCR for *ESR1*, *PGR* and *MKI67* with quantitative IHC by image analysis (**a**, **c**, **e**) and visual IHC assessment (**b**, **d**, **f**)

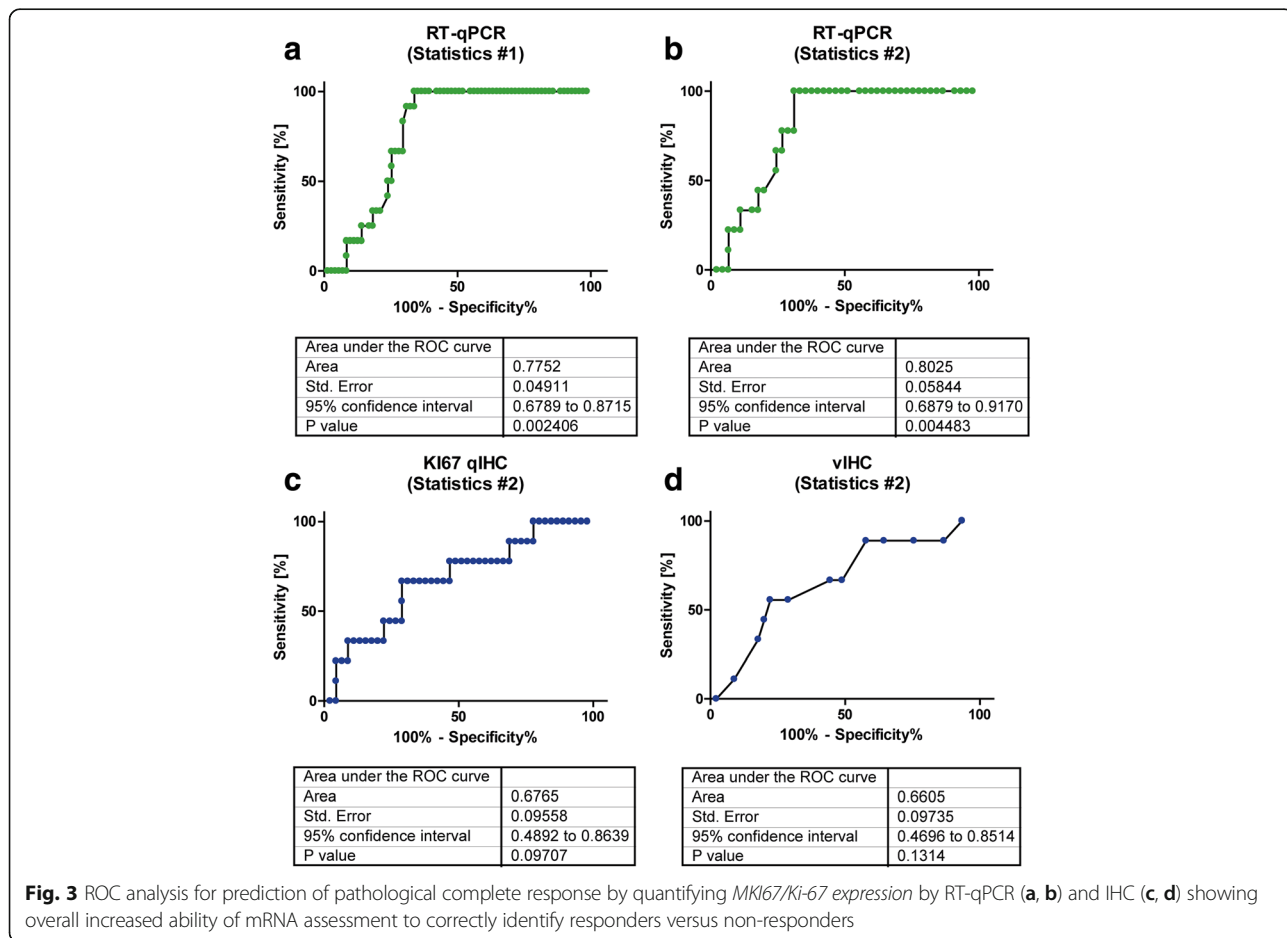Sinn *et al. BMC Cancer* (2017) 17:124
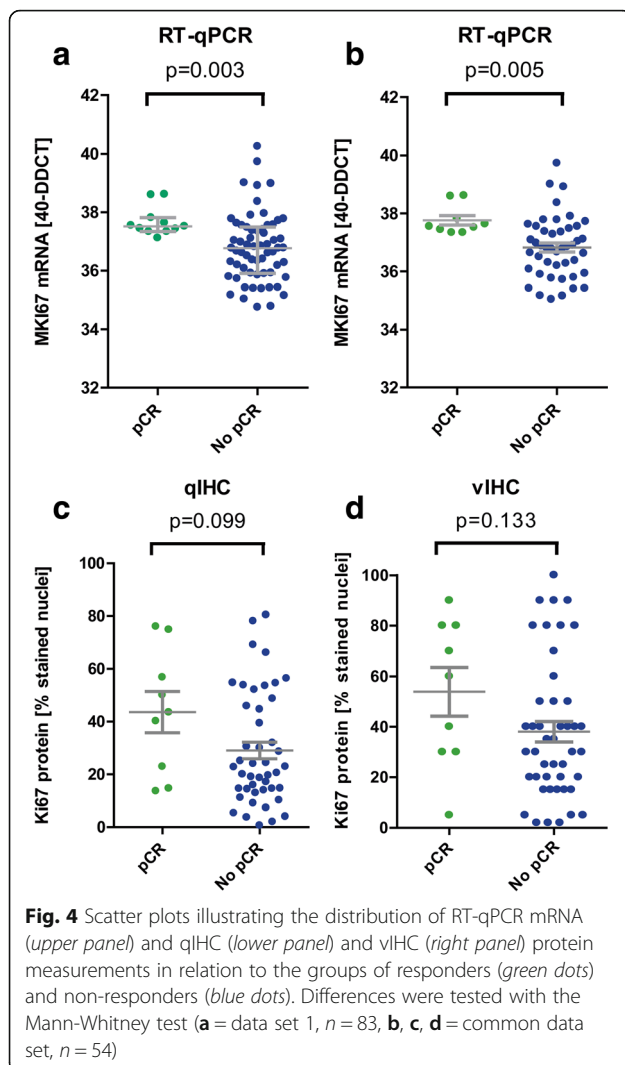
Page 6 of 10

35.0%, and 37.01 for RT-qPCR, clearly reflecting the inclusion criteria of the S080 trial which targeted clinically higher-risk patients. Scatter plot analysis displays the positive correlation between RT-qPCR and visual as well as quantitative IHC assessment in Fig. 4.

Prediction of pCR

To compare the clinical utility we performed a ROC analysis to determine the optimal cutoff for predicting the pCR. The results of the ROC analysis are presented in the graphical plots of Fig. 3. With RT-qPCR, 100% of responders could be detected with a specificity of 68.9% at a 40-ddCT level of 37.31 which almost reflected the median mRNA expression in this cohort (Fig. 3a, b). Conversely, no responder was below RT-qPCR of 37.31 (Fig. 4a). For IHC assessment, it was difficult to determine a reliable cutoff reaching high sensitivity and specificity. With RT-qPCR the area under the curve was 0.78 for the overall cohort and 0.80 for the IHC cohort (Statistics #2) ($p = 0.002$ and $p = 0.004$). For both IHC methods, the ROC was not significant.

Since maximum sensitivity was a pre-requisite, the two methods were compared with respect to specificity, which was found to be substantially higher for MammaTyper® (68.9%) compared to qIHC (22.2%). Using the cut-offs indicated by ROC analysis (37.31 for RT-qPCR, 13.2% for qIHC and 3.5% for vIHC), tumors were characterized as bearing either high or low MKI67 RNA or Ki-67 protein expression, respectively. However, as illustrated in Fig. 4, statistically significant differences between groups were found for the RT-qPCR but not for IHC methods when patients were stratified according to proliferation and pCR (Mann–Whitney $p = 0.003$ and $p = 0,005$ for RT-qPCR and $p = 0.099$ for qIHC and $p = 0.133$ for vIHC). Using the cut-offs obtained by ROC analysis, pCR was observed in 9 of 24 patients (37.5%, $p < 0.001$) with high MKI67 RNA expression but in no patient with low RNA expression. Accordingly for qIHC, pCR was observed in 9 of 44 patients (20.5%, $p = 0.27$) with high Ki-67 labelling index and similarly it was entirely lacking in patients with low proliferation. For RT-qPCR the ROC analysis was also highly significant

a RT-qPCR (Statistics #1)

Area under the ROC curve	
Area	0.7752
Std. Error	0.04911
95% confidence interval	0.6789 to 0.8715
P value	0.002406

b RT-qPCR (Statistics #2)

Area under the ROC curve	
Area	0.8025
Std. Error	0.05844
95% confidence interval	0.6879 to 0.9170
P value	0.004483

c KI67 qIHC (Statistics #2)

Area under the ROC curve	
Area	0.6765
Std. Error	0.09558
95% confidence interval	0.4892 to 0.8639
P value	0.09707

d vIHC (Statistics #2)

Area under the ROC curve	
Area	0.6605
Std. Error	0.09735
95% confidence interval	0.4696 to 0.8514
P value	0.1314

Fig. 3 ROC analysis for prediction of pathological complete response by quantifying *MKI67/Ki-67 expression* by RT-qPCR (**a**, **b**) and IHC (**c**, **d**) showing overall increased ability of mRNA assessment to correctly identify responders versus non-responders

Sinn *et al. BMC Cancer* (2017) 17:124

Page 7 of 10

Fig. 4 Scatter plots illustrating the distribution of RT-qPCR mRNA (*upper panel*) and qIHC (*lower panel*) and vIHC (*right panel*) protein measurements in relation to the groups of responders (*green dots*) and non-responders (*blue dots*). Differences were tested with the Mann-Whitney test (**a** = data set 1, *n* = 83, **b**, **c**, **d** = common data set, *n* = 54)

For the mRNA estimation we used the MammaTyper®, a novel in vitro diagnostic test for breast cancer molecular subtyping. To prove the clinical utility of mRNA based assessment, we compared RT-qPCR with conventional visual assessment as well as digital image analysis based determination at a reference pathology lab in the context of a clinical trial. Moreover, the methods were examined with respect to their ability to predict pCR according to Ki-67 protein or MKI67 mRNA expression levels measured on pretreatment core biopsies. Our results indicate, that, when using RT-qPCR valid cut-offs for mRNA expression, which reliably distinguish between non-responding and responding tumors as determined by pCR (ypT0 ypN0) can be identified.

Pathological complete response has gained wide acceptance as one of the strongest predictors of prolonged survival in the setting of neoadjuvant chemotherapy [26, 27]. Therefore, laboratory assays that can efficiently predict a patient's response to a given preoperative chemotherapeutic combination may serve as tools for individualizing treatment and improving long-term outcomes [28]. As with adjuvant chemotherapy, neoadjuvant regimens also suffer from the fact that substantial therapeutic benefit is restricted only to a fraction of those treated, whereas all patients will experience adverse events because of toxicity [29].

In several neoadjuvant studies Ki-67 protein expression has been investigated in pre-operative biopsies in relation to the response to treatment and in most cases a high Ki-67 proliferation rate was predictive of higher probability of pCR [13]. Fasching et al. analyzed Ki-67 by conventional IHC in core biopsies from 552 patients from a single German institution and showed that a pre-defined 13% cut-off could predict pCR with 94% sensitivity and 36% specificity [30]. Interestingly, our ROC analysis for qIHC requiring 100% sensitivity with the least possible loss on specificity led to an identical cut-off (13.2%) for Ki-67. However, this finding requires careful interpretation, due to differences characterizing the clinical settings between the two neoadjuvant studies and the original work by Cheang [31]. In the latter case, the Ki-67 cut-off was fine-tuned against gene expression profiling in order to distinguish Luminal A from Luminal B tumors in a population containing both high- and low-risk breast cancers, whereas in the neoadjuvant setting the same cut-off was intended to identify the majority of, mainly high-risk, patients that would most likely benefit from preoperative cytotoxic therapy.

Alike what has been repeatedly shown in the adjuvant setting, it appears that the molecular architecture of tumors as defined by the expression of hormone receptors and HER2/neu may act as a modifier of the association between Ki-67 and response to neoadjuvant treatment

when only luminal tumors had been assessed, though the sample size was small in this subset (data not shown).

Discussion

In this study we have validated clinical performance of hormone receptor gene expression by RT-qPCR by comparing predefined cut-offs in a blinded fashion with the current standard of IHC. Furthermore, we have investigated the diagnostic performance of two methods for assessing MKI67 gene expression, namely IHC with computerized quantification of protein and RT-qPCR RNA quantification with the MammaTyper® IVD kit in the setting of pCR prediction. When continuous data were dichotomized to reflect high- and low-MKI67 categories with cut-offs obtained by ROC curve analysis after considering 100% sensitivity, RT-qPCR was significantly more specific than qIHC.

To the best of our knowledge this is the first direct comparison of this kind in the context of a clinical trial.

Sinn *et al. BMC Cancer* (2017) 17:124

Page 8 of 10

and between pCR and long-term outcomes [14, 32]. While 101 tumors were available for analysis, the inclusion of 83 or 54 tumors in this study was not based on a statistical rational but was dictated by the availability of tumor tissue with complete RT-qPCR and qIHC data.

A novel aspect of the present work is the comparison between protein-based and mRNA-based methods for the assessment of tumor proliferation. Our findings highlight the feasibility of using RT-qPCR for the routine assessment of ESR1, PGR & MKI67 in order to assist the selection of breast cancer patients for neoadjuvant treatment. Even though both RT-qPCR and qIHC of MKI67/Ki-67 could be calibrated to maximize negative predictive value, only with the former this was achieved whilst ensuring sufficient specificity, which if validated would signify that MammaTyper® could help a considerable number of patients safely forego unnecessary treatment. These data collectively indicate that MammaTyper®MKI67 RNA was overall more representative of the true proliferation state of the tumor than was computer assisted Ki-67 protein estimation, a finding that is worth validating in larger datasets.

Significant correlations between conventional Ki-67 visual assessment and RT-qPCR have been previously reported [33, 34], indicating a strong biological link between mRNA and protein expression despite methodological variations, as is further indicated by comparable prognostic hazard ratios obtained by both methods [35]. To the best of our knowledge however, our study is the first to compare image analysis with RT-qPCR for the assessment of tumor proliferation with the additional advantage of using material from a randomized clinical trial. The correlation between mRNA and protein was significant but moderate, a finding which may reflect post translational modifications or may be related to the increased dynamic range of RT-qPCR as compared to IHC. Another possible explanation might be that mRNA levels are a reflection of the average gene expression in the entire FFPE slice, whereas IHC may be biased in favor of selected "representative" tumor areas. Even in the case of image analysis systems, inspection of digitalized images and manual identification of tumor areas is necessary before automatic scoring.

Computerized methods have been recommended as a solution to the problem of subjectivity in the visual assessment and scoring of IHC-stained slides. Not surprisingly, Ki-67 scores from image analysis systems are generally in close agreement with those of manual methods because manual scoring for research purposes is customarily performed by a pathologist with long-standing experience in the field [17, 36]. It is worth mentioning, however, that in a routine decentralized setting, digital processing and scoring of slides would probably outperform manual assessment which is prone to considerable subjectivity often not improved upon standardization [37]. Digital analysis yields more reproducible results with regard to staining intensity, by facilitation the definition of low grade staining intensities. Definitive conclusions would require comparisons between all three methods (central versus local versus automatic) performed preferably in the prospective retrospective setting of a large multi-center trial.

Multi-gene molecular signatures have also been tested as a way for predicting pCR in patients with breast cancer [38–40]. However, generalized use of these commercialized assays is limited by their increased cost and the requirement to run in centralized platforms or both. Interestingly, proliferation genes, including MKI67, are often heavily weighted in multi-gene scores which serve as estimators of a patients' risk of developing recurrences. This is perhaps one of the reasons why multi-gene tests do not always prove to be convincingly superior to conventional or less sophisticated methods for tumor risk stratification [35, 41, 42], leading some authors to question their cost-effectiveness [43]. Moreover, for several commercially available tests, neither doctors nor consumers can gain access to the continuous expression data of individual proliferation markers that make up the final risk scores. This restriction overall minimizes the possibility of potentially interesting comparisons between proliferation motifs or scores and single proliferation markers based on RT-qPCR or IHC. Strikingly, our ROC curve analysis of MKI67 $40\text{-}\Delta\Delta Cq$ values for the prediction of pCR displayed performance characteristics that are comparable with those of a 50-gene predictor of tumor recurrence risk developed by supervised training of Cox models [39]. Along these lines, single-gene MKI67 RT-qPCR may be worth considering as a golden means for assessing tumor proliferation due to its unique ability to combine technical advancements and diagnostic accuracy with more affordable pricing.

Conclusions

Image analysis-assisted scoring of ER, PR and Ki-67 IHC and quantification of ESR1, PGR and MIKI67 RNA expression with RT-qPCR both represent promising alternatives to conventional visual estimation and may assist in improving reproducibility and accuracy in the field. However, RT-qPCR assessment of tumor proliferation was overall more accurate than quantitative IHC. This is the first study to compare tumor MKI67 gene expression by RNA and protein assessment in a prospective retrospective neoadjuvant setting. Due to the relatively small sample size, these data should be considered preliminary and worth validating in larger datasets.

Sinn *et al. BMC Cancer* (2017) 17:124

Page 9 of 10

Additional file

Additional file 1: Raw data; qPCR data, quantitative and visual IHC values, pathologic complete response yes/no. (XLSX 42 kb)

Abbreviations
CISH: Chromogenic in situ hybridization; Cq: Quantification cycle; DDFS: Distant disease free survival; ESR1/ER: Oestrogen receptor alpha; ERBB2/HER2; FEC: Fluorouracil & epirubicin, cyclophosphamide chemotherapy; FFPE: Formalin fixed paraffin embedded; GOI: Gene of interest; HR: Hazard ratio; IHC: Immunohistochemistry; MKI67/Ki67: marker of proliferation Ki-67; mRNA: Messenger ribonucleic acid; NPA: Negative percentage agreement; OPA: Overall percentage agreement; OS: Overall survival; PGR/PgR: Progesterone receptor; PPA: Positive percentage agreement; REF: Reference gene; RT-qPCR: Reverse transcription quantitative real time polymerase chain reaction

Availability of data and materials
The dataset has been made available in the Excel file format as Additional file 1.

Authors' contributions
HPS, KS, ML, SL, RMW and FM were involved in the conception of the study. EV performed the RT-qPCR assays; HPS, AS, MK, JS and FM provided study data and materials; HPS, SL, PA and RMW performed the statistical analysis and wrote the statistical plan; HPS, AS, KS, ML, SL, SE, RMW and FM interpreted the data; HPS, SL, RMW and FM drafted the manuscript; all authors read and approved the final manuscript.

Competing interests
RMW and SE are founders of STRATIFYER Molecular Pathology GmbH. SL, EV and RMW are employees of STRATIFYER Molecular Pathology GmbH. KS and ML are employees of BioNTech Diagnostics GmbH. All other authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
All patients had been enrolled in the H3E-MC-S080 (NCT00149214) neoadjuvant phase II study [1] at a single institution (National Center for Tumor Diseases, University-Hospital, Heidelberg). A written informed consent for the research use of patient biological material was granted at the time of enrolment. The research described herein is completely independent from the sponsor of the original study (Eli Lilly and Company). The study was approved by the local ethics committee of the Heidelberg University Hospital.

Author details
[1]Institute of Pathology, University of Heidelberg, Im Neuenheimer Feld 220-221, 69120 Heidelberg, Germany. [2]National Center for Tumor Diseases, University-Hospital Heidelberg, Im Neuenheimer Feld 460, 69120 Heidelberg, Germany. [3]German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [4]STRATIFYER Molecular Pathology GmbH, Werthmannstr. 1c, 50935 Köln, Germany. [5]Department of Pathology, St. Elisabeth-Krankenhaus, Werthmannstr. 1c, 50935 Köln, Germany. [6]BioNTech Diagnostics GmbH, 55131 Mainz, Germany.

References
1. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F, Committee EG. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26 Suppl 5:v8–30.
2. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thurlimann B, Senn HJ, Panel M. Tailoring therapies–improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26(8):1533–46.
3. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. J Immunol. 1984;133(4):1710–5.
4. Braun L, Mietzsch F, Seibold P, Schneeweiss A, Schirmacher P, Chang-Claude J, Sinn P, Aulmann S. Intrinsic breast cancer subtypes defined by estrogen receptor signalling-prognostic relevance of progesterone receptor loss. Mod Pathol. 2013;26(9):1161–71.
5. Maisonneuve P, Disalvatore D, Rotmensz N, Curigliano G, Colleoni M, Dellapasqua S, Pruneri G, Mastropasqua MG, Luini A, Bassi F, et al. Proposed new clinicopathological surrogate definitions of luminal A and luminal B (HER2-negative) intrinsic breast cancer subtypes. Breast Cancer Res. 2014;16(3):R65.
6. Peppercorn J, Perou CM, Carey LA. Molecular subtypes in breast cancer evaluation and management: divide and conquer. Cancer Invest. 2008;26(1):1–10.
7. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. Arch Pathol Lab Med. 2010; 134(6):907–22.
8. de Gramont A, Watson S, Ellis LM, Rodon J, Tabernero J, de Gramont A, Hamilton SR. Pragmatic issues in biomarker evaluation for targeted therapies in cancer. Nat Rev Clin Oncol. 2015;12(4):197–212.
9. Harvey J, Clark G, Osborne C, Allred D. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. J Clin Oncol. 1999;17(5):1474–81.
10. Luporsi E, Andre F, Spyratos F, Martin PM, Jacquemier J, Penault-Llorca F, Tubiana-Mathieu N, Sigal-Zafrani B, Arnould L, Gompel A, et al. Ki-67: level of evidence and methodological considerations for its role in the clinical management of breast cancer: analytical and critical review. Breast Cancer Res Treat. 2012;132(3):895–915.
11. Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, Viale G, Zabaglo LA, Penault-Llorca F, Bartlett JM, et al. An international Ki67 reproducibility study. J Natl Cancer Inst. 2013;105(24):1897–906.
12. Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, Obermann E, Ohlschlegel C, Padberg B, Rakozy C, et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss working group of breast- and gynecopathologists. PLoS ONE. 2012; 7(5):e37379.
13. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 2010;11(2):174–83.
14. Denkert C, Loibl S, Muller BM, Eidtmann H, Schmitt WD, Eiermann W, Gerber B, Tesch H, Hilfrich J, Huober J, et al. Ki67 levels as predictive and prognostic parameters in pretherapeutic breast cancer core biopsies: a translational investigation in the neoadjuvant GeparTrio trial. Ann Oncol. 2013;24(11):2786–93.
15. Di Cataldo S, Ficarra E, Macii E. Computer-aided techniques for chromogenic immunohistochemistry: status and directions. Comput Biol Med. 2012;42(10):1012–25.
16. Voros A, Csorgo E, Nyari T, Cserni G. An intra- and interobserver reproducibility analysis of the Ki-67 proliferation marker assessment on core biopsies of breast cancer patients and its potential clinical implications. Pathobiology. 2013;80(3):111–8.
17. Konsti J, Lundin M, Joensuu H, Lehtimaki T, Sihto H, Holli K, Turpeenniemi-Hujanen T, Kataja V, Sailas L, Isola J, et al. Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer. BMC Clin Pathol. 2011;11:3.

Sinn *et al. BMC Cancer* (2017) 17:124

Page 10 of 10

18. Cass JD, Varma S, Day AG, Sangrar W, Rajput AB, Raptis LH, Squire J, Madarnas Y, Sengupta SK, Elliott BE. Automated quantitative analysis of p53, Cyclin D1, Ki67 and pERK expression in breast carcinoma does not differ from expert pathologist scoring and correlates with clinico-pathological characteristics. Cancers (Basel). 2012;4(3):725–42.

19. Abubakar M, Howat WJ, Daley F, Zabaglo L, McDuffus LA, Blows F, Coulson P, Raza Ali H, Benitez J, Milne R, et al. High-throughput automated scoring of Ki67 in breast cancer tissue microarrays from the Breast Cancer Association Consortium. J Pathol Clin Res. 2016;2(3):138–53.

20. Noske A, Loibl S, Darb-Esfahani S, Roller M, Kronenwett R, Muller BM, Steffen J, von Toerne C, Wirtz R, Baumann I, et al. Comparison of different approaches for assessment of HER2 expression on protein and mRNA level: prediction of chemotherapy response in the neoadjuvant GeparTrio trial (NCT00544765). Breast Cancer Res Treat. 2011;126(1):109–17.

21. Bohmann K, Hennig G, Rogel U, Poremba C, Mueller BM, Fritz P, Stoerkel S, Schaefer KL. RNA extraction from archival formalin-fixed paraffin-embedded tissue: a comparison of manual, semiautomated, and fully automated purification methods. Clin Chem. 2009;55(9):1719–27.

22. Laible M, Schlombs K, Kaiser K, Veltrup E, Herlein S, Lakis S, Stohr R, Eidt S, Hartmann A, Wirtz RM, et al. Technical validation of an RT-qPCR in vitro diagnostic test system for the determination of breast cancer molecular subtypes by quantification of ERBB2, ESR1, PGR and MKI67 mRNA levels from formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer. 2016;16:398.

23. Elsawaf Z, Sinn HP, Rom J, Bermejo JL, Schneeweiss A, Aulmann S. Biological subtypes of triple-negative breast cancer are associated with distinct morphological changes and clinical behaviour. Breast. 2013;22(5):986–92.

24. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882–90.

25. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928–35.

26. Guarneri V, Broglio K, Kau SW, Cristofanilli M, Buzdar AU, Valero V, Buchholz T, Meric F, Middleton L, Hortobagyi GN, et al. Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors. J Clin Oncol. 2006;24(7):1037–44.

27. Cortazar P, Geyer Jr CE. Pathological complete response in neoadjuvant treatment of breast cancer. Ann Surg Oncol. 2015;22(5):1441–6.

28. Berruti A, Generali D, Kaufmann M, Puztai L, Curigliano G, Aglietta M, Gianni L, Miller WR, Untch M, Sotiriou C, et al. International expert consensus on primary systemic therapy in the management of early breast cancer: highlights of the fourth symposium on primary systemic therapy in the management of operable breast cancer, Cremona, Italy (2010). J Natl Cancer Inst Monogr. 2011;2011(43):147–51.

29. Mauri D, Pavlidis N, Ioannidis JP. Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. J Natl Cancer Inst. 2005;97(3):188–94.

30. Fasching PA, Heusinger K, Haeberle L, Niklos M, Hein A, Bayer CM, Rauh C, Schulz-Wendtland R, Bani MR, Schrauder M, et al. Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. BMC Cancer. 2011;11:486.

31. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009;101(10):736–50.

32. von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, Gerber B, Eiermann W, Hilfrich J, Huober J, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol. 2012; 30(15):1796–804.

33. Potemski P, Pluciennik E, Bednarek AK, Kusinska R, Kubiak R, Jesionek-Kupnicka D, Watala C, Kordek R. Ki-67 expression in operable breast cancer: a comparative study of immunostaining and a real-time RT-PCR assay. Pathol Res Pract. 2006;202(7):491–5.

34. Yamamoto S, Ibusuki M, Yamamoto Y, Fu P, Fujiwara S, Murakami K, Iwase H. Clinical relevance of Ki67 gene expression analysis using formalin-fixed paraffin-embedded breast cancer specimens. Breast Cancer. 2013;20(3):262–70.

35. Niikura N, Iwamoto T, Masuda S, Kumaki N, Xiaoyan T, Shirane M, Mori K, Tsuda B, Okamura T, Saito Y, et al. Immunohistochemical Ki67 labeling index has similar proliferation predictive power to various gene signatures in breast cancer. Cancer Sci. 2012;103(8):1508–12.

36. Klauschen F, Wienert S, Schmitt WD, Loibl S, Gerber B, Blohmer JU, Huober J, Rudiger T, Erbstosser E, Mehta K, et al. Standardized Ki67 diagnostics using automated scoring–clinical validation in the GeparTrio breast cancer study. Clin Cancer Res. 2015;21(16):3651–7.

37. Mohammed ZM, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, Doughty JC, Edwards J. Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. Br J Cancer. 2012;106(2):383–8.

38. Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J, Mariani G, Rodriguez J, Carcangiu M, Watson D, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. J Clin Oncol. 2005;23(29):7265–77.

39. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160–7.

40. Straver ME, Glas AM, Hannemann J, Wesseling J, van de Vijver MJ, Rutgers EJ, Vrancken Peeters MJ, van Tinteren H, Van't Veer LJ, Rodenhuis S. The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. Breast Cancer Res Treat. 2010;119(3):551–8.

41. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, Zabaglo L, Mallon E, Green AR, Ellis IO, 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(32):4273–8.

42. Tobin NP, Harrell JC, Lovrot J, Egyhazi Brage S, Frostvik Stolt M, Carlsson L, Einbeigi Z, Linderholm B, Loman N, Malmberg M, et al. Molecular subtype and tumor characteristics of breast cancer metastases as assessed by gene expression significantly influence patient post-relapse survival. Ann Oncol. 2015;26(1):81–8.

43. Williams DJ, Cohen C, Darrow M, Page AJ, Chastain B, Adams AL. Proliferation (Ki-67 and phosphohistone H3) and oncotype DX recurrence score in estrogen receptor-positive breast cancer. Appl Immunohistochem Mol Morphol. 2011;19(5):431–6.