BMC
Cancer

**TECHNICAL ADVANCE**                                                                 **Open Access**

# Decentral gene expression analysis: analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test

Ralf Kronenwett[1*], Kerstin Bohmann[1], Judith Prinzler[2], Bruno V Sinn[2], Franziska Haufe[1], Claudia Roth[1], Manuela Averdick[1], Tanja Ropers[1], Claudia Windbergs[1], Jan C Brase[1], Karsten E Weber[1], Karin Fisch[1], Berit M Müller[2], Marcus Schmidt[3], Martin Filipits[4], Peter Dubsky[5], Christoph Petry[1], Manfred Dietel[2] and Carsten Denkert[2]

## Abstract

**Background:** EndoPredict (EP) is a clinically validated multianalyte gene expression test to predict distant metastasis in ER-positive, HER2-negative breast cancer treated with endocrine therapy alone. The test is based on the combined analysis of 12 genes in formalin-fixed, paraffin-embedded (FFPE) tissue by reverse transcription-quantitative real-time PCR (RT-qPCR). Recently, it was shown that EP is feasible for reliable decentralized assessment of gene expression. The aim of this study was the analytical validation of the performance characteristics of the assay and its verification in a molecular-pathological routine laboratory.

**Methods:** Gene expression values to calculate the EP score were assayed by one-step RT-qPCR using RNA from FFPE tumor tissue. Limit of blank, limit of detection, linear range, and PCR efficiency were assessed for each of the 12 PCR assays using serial samples dilutions. Different breast cancer samples were used to evaluate RNA input range, precision and inter-laboratory variability.

**Results:** PCR assays were linear up to $C_q$ values between 35.1 and 37.2. Amplification efficiencies ranged from 75% to 101%. The RNA input range without considerable change of the EP score was between 0.16 and 18.5 ng/µl. Analysis of precision (variation of day, day time, instrument, operator, reagent lots) resulted in a total noise (standard deviation) of 0.16 EP score units on a scale from 0 to 15. The major part of the total noise (SD 0.14) was caused by the replicate-to-replicate noise of the PCR assays (repeatability) and was not associated with different operating conditions (reproducibility). Performance characteristics established in the manufacturer's laboratory were verified in a routine molecular pathology laboratory. Comparison of 10 tumor samples analyzed in two different laboratories showed a Pearson coefficient of 0.995 and a mean deviation of 0.15 score units.

**Conclusions:** The EP test showed reproducible performance characteristics with good precision and negligible laboratory-to-laboratory variation. This study provides further evidence that the EP test is suitable for decentralized testing in specialized molecular pathological laboratories instead of a reference laboratory. This is a unique feature and a technical advance in comparison with existing RNA-based prognostic multigene expression tests.

**Keywords:** Breast cancer, Prognostic multigene expression test, Analytical validation, PCR, Pathology

* Correspondence: Kronenwett@sividon.com
[1]Sividon Diagnostics GmbH, Nattermannallee 1, 50829, Cologne, Germany
Full list of author information is available at the end of the article
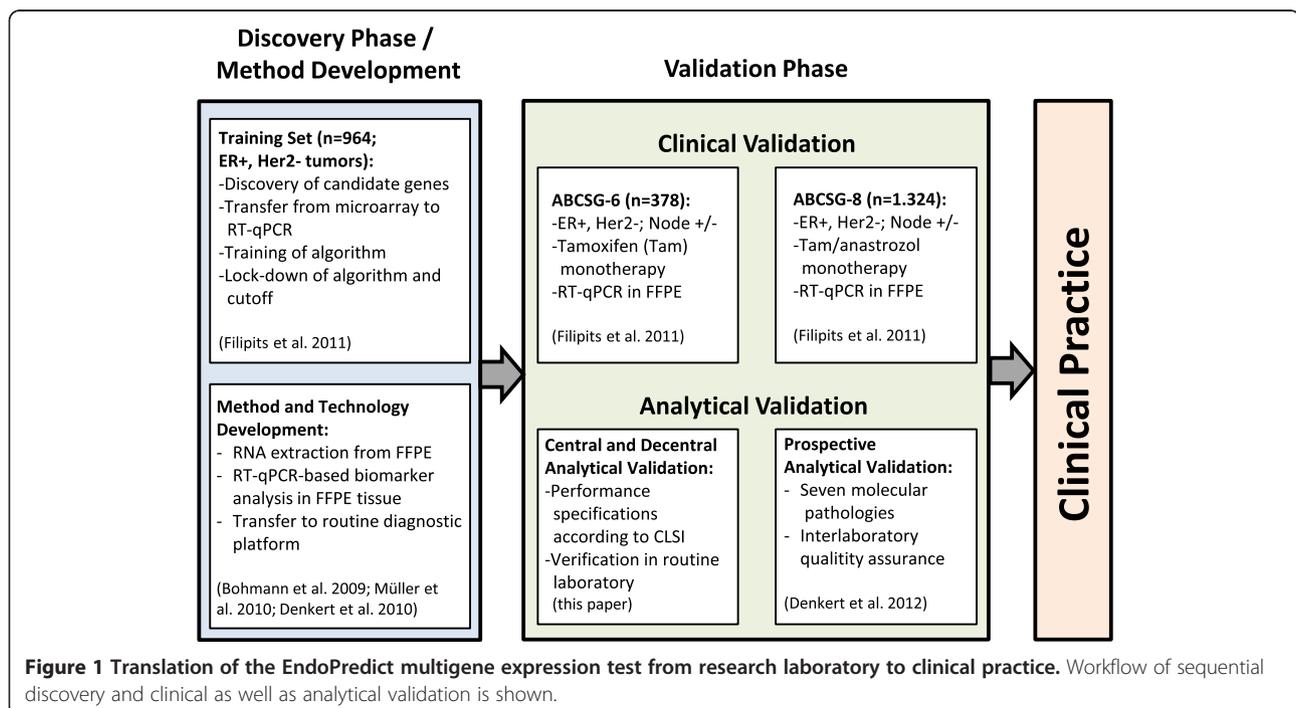
**BioMed** Central

## Background

EndoPredict (EP) is a multigene assay which predicts the risk of distant metastasis in ER+/HER2- breast cancer and identifies a subgroup of patients who have an excellent prognosis if treated with endocrine therapy alone [1]. The test is based on the assessment of expression of 8 informative genes, 3 reference genes, and one gene to measure the presence of genomic DNA in RNA from formalin-fixed, paraffin-embedded (FFPE) tissue from biopsies or surgical specimens using reverse transcription-quantitative real-time PCR (RT-qPCR) [1-3]. Relative gene expression levels are used to calculate the EndoPredict score (EP score) ranging from 0 to 15. Patients with a score below or equal to 5 are classified as low risk for distant recurrence under endocrine therapy, those with a score above 5 as high risk.

Translation of the EP test from research laboratory to clinical practice covered the necessary steps for development of a laboratory test (Figure 1). This included method development for standardized RNA extraction from FFPE tissue [4-6] and transfer of RT-qPCR assays to a certified routine diagnostic platform [2,7] as well as a discovery phase with biomarker identification and training of an algorithm in a multicenter cohort [1]. Following discovery the pre-defined, locked-down EP score was clinically validated in two separate cohorts from the two randomized clinical trials ABCSG-6 (n=378) and ABCSG-8 (n=1324) [1]. Moreover, it has been shown that the EP score provided prognostic information on the risk of distant metastasis of breast cancer

patients beyond clinic-pathological parameters such as ki-67 and quantitative ER immunohistochemistry [1].

Besides EndoPredict, other prognostic multigene expression tests for patients with breast cancer like MammaPrint [8], Oncotype DX [9,10], or PAM50 [11,12] are commercially available. However, all these different tests can only be performed in reference laboratories. In contrast, EndoPredict is suitable for decentralized testing in specialized molecular pathological laboratories as recently shown in a prospective analytical proficiency testing program with seven different molecular pathological laboratories [2].

The aim of this study was a comprehensive analytical validation of the EP test to complete development before implementation in clinical practice. Analytical validation of multianalyte assays is still a challenge, as these types of assays require a more complex evaluation of the performance characteristics compared with single analyte assays in order to assure reliable performance in the clinical routine. Adequate performance evaluation includes the control of the process from the acquisition of the tumor samples and isolation of the RNA to the assessment of each single analyte as well as the combination of the single results to a comprehensive score by an algorithm. Moreover, guidelines for analytical validation of multianalyte genomic assays are rare. Here, we analytically validated the EndoPredict multianalyte gene expression assay according to the adapted guideline MM17-A of the Clinical and Laboratory Standards Institute (CLSI) addressing the analytical validation of nucleic acid-based qualitative and



**Figure 1 Translation of the EndoPredict multigene expression test from research laboratory to clinical practice.** Workflow of sequential discovery and clinical as well as analytical validation is shown.

semiquantitative multiplex assays [13]. Moreover, the performance characteristics of the assay were verified in an independent molecular pathological laboratory to confirm that the test meets its specifications when used in a routine diagnostic laboratory.

## Methods

### Reference and Testing Materials

Nucleic acids for test development and validation were selected based on the specific purpose of the respective analytical performance characteristics to be tested. Sample material was comparable to the specimen used in the clinical testing, i.e. DNA-free total RNA or genomic DNA from FFPE tissue which is fragmented to nucleic acid pieces by formalin-fixation [5,14]. Details about reference nucleic acids are described in supplemental data (see Additional file 1). In brief, for assessment of limit of detection (LoD), linear range, and efficiency of the single PCR assays large pools of control RNA and control DNA from different FFPE tumor blocks were generated and used for the experiments [15]. For the precision studies, three tumor specimens classified by the EP as low risk, high risk or close to the decision point were selected. For the correlation study, ten tumor samples were chosen with EP scores spanning the larger part of the full score range. These ten tumor samples were different to the ten samples used in the recently published EndoPredict proficiency testing [2]. This study was carried out in compliance with the Helsinki Declaration and was approved by the Ethics Committee of the Charité Hospital (Ref. No. EA1/139/05, Amendment 2008).

As positive controls of RT-qPCR assays a standardized reference RNA (Stratagene qPCR Human Reference Total RNA, Agilent Technologies, Böblingen, Germany) and Human Genomic DNA (Roche Applied Bioscience, Mannheim, Germany) were tested on each plate.

### Isolation of RNA and DNA

Total RNA and DNA was extracted from FFPE tissue sections (10 μm) using a fully-automated silica-coated magnetic bead-based method in combination with a liquid handling robot (VERSANT Tissue Preparation System, Siemens Healthcare Diagnostics, Eschborn, Germany) as published previously [4-6].

The mean of $C_q$ (quantification cycle) values of the EP reference genes *RPL37A*, *CALM2* and *OAZ1* was used as surrogate marker for mRNA yield following isolation. Concentration of total RNA was assessed using the QUANT-iT RIBOGREEN assay (Life Technologies, Darmstadt, Germany). For assessment of contamination with residual DNA in RNA preparations, an *HBB* gene-specific quantitative PCR was performed. Samples were considered to be substantially free of DNA when $C_q$ values above 38

were detected. In case of DNA contamination samples were manually re-digested by DNase I treatment.

### Gene expression analysis using RT-qPCR

Expression of 8 genes-of-interest (*AZGP1, BIRC5, DHCR7, IL6ST, MGP, RBBP8, STC2, UBE2C*) and three reference genes (*CALM2, OAZ1, RPL37A*) as well as the amount of residual genomic DNA (*HBB*) were assessed by the EndoPredict assay (Sividon Diagnostics, Cologne, Germany) as previously described [1,2]. This assay is configured on a 96-well plate containing primers and FAM/TAMRA-labeled hydrolysis probes dried into the wells. Functional details about genes, data base accession numbers and sequences of primers and probes were published previously [1]. Gene expression was assessed by one-step RT-qPCR using the SuperScript III PLATINUM One-Step Quantitative RT-PCR System with ROX (Life Technologies, Darmstadt, Germany) according to manufacturer's instructions in a VERSANT kPCR Molecular System (Siemens Healthcare Diagnostics, Eschborn, Germany) with 30 min at 50℃, 2 min at 95℃ followed by 40 cycles of 15 sec at 95℃ and 30 sec at 60℃. 20 μl reaction mix containing buffer, nucleotides, 4.5 mM $Mg^{2+}$, enzymes and 1 μl sample RNA, respectively, were added to each well. The gene-specific reverse PCR primers were used as primers for reverse transcription. Since the *HBB*-specific assay did not target mRNA sequences the RT-qPCR protocol as described above could be used all the same.

For calculation of the EP score genes were measured in triplicates. This is mandatory to control for PCR imprecision and to enable outlier removal [1,2]. $C_q$ values were calculated by the VERSANT kPCR Molecular System software using amplification-based thresholds following baseline correction according to manufacturer's instructions. Detection of outliers, relative expression levels of each gene-of-interest ([GOI]; $\Delta C_q(GOI) = 20 - C_q (GOI) + [C_q (CALM2) + C_q (OAZ1) + C_q (RPL37 A)]/3$) as well as EP scores were calculated as described previously using a web-based implementation to process analytical PCR results into EP scores which can be found at: http://forschung.medizin.uni-mainz.de/epreport/ [1].

### Assessment of limit of blank (LoB), limit of detection (LoD), linear dynamic range and PCR efficiency

LoB, defined as the 5%-percentile of the distribution of $C_q$-values measured in a blank sample without analyte, was calculated as described in supplemental data (see Additional file 1) [16]. LoD was defined as the amount of the reference RNA or DNA at which the $C_q$ value is below the LoB with a probability of 95%. Since an absolute quantification of the 12 different targets in total RNA or DNA from FFPE tumors was not possible LoD was referred to the fold-dilution of the reference nucleic acid and to the respective $C_q$ value as a surrogate for the

amount of the individual analytes. For assessment of LoD and linear dynamic range, four independent series of 20 gravimetrically controlled serial 1:2 dilutions ($\log_2$) were generated from a pooled RNA sample (DNA sample for *HBB* PCR) from FFPE tissue resulting in 21 different concentrations [16-18]. Details about the dilution series and assessment of LoD and linear range are described in supplemental data (see Additional file 1). For each single PCR assay the linear dynamic range was determined by fitting a linear, quadratic, or cubic model. A maximum deviation from linearity of 1 Ct value was accepted. After assessing the linear dynamic range, the PCR efficiency was calculated by $E = (2^{-1/m}-1) \times 100\%$ where m is the slope of the linear regression model.

### Assessment of precision
The precision experiment was designed according to CLSI guidelines [13] and evaluated following ISO 5725-2 and NCCLS EP5-A2 [19,20]. The following variables were included: day (n=11), day time (n=2), PCR instrument (n=4), position of sample on 96-well EndoPredict plate (n=2), lot of EndoPredict plate (n=4), lot of enzyme/master mix (n=2), and operator (n=3). The experiment was performed during 28 calendar days including a 5 working days familiarization period at the beginning. Three different RNAs were used as test samples: One sample from a tumor with a low EP score (2.4), one with a high EP score (13.5) and one at the decision point between low and high risk (4.9). RNA was isolated from several sections and pooled for each tumor to have sufficient RNA for the whole precision experiment. In addition to the test samples, one quality control sample (Stratagene qPCR Human Reference Total RNA) was analyzed in each run.

For verification in the laboratory of Charité as a representative routine laboratory an abbreviated precision experiment with fewer variables was performed: day (n=5), day time (n=2), position of sample on 96-well EndoPredict plate (n=2), lot of enzyme/master mix (n=2), and operator (n=2).

Variable noise, replicate noise and total noise was calculated using univariate N-way analysis of variance (ANOVA) and indicated as standard deviations [18] as described in detail in supplemental data (see Additional file 1).

### Statistics
For EP scores 95% confidence intervals (CI) were calculated as described [1]. For comparison of EP test results between two different laboratories Pearson correlation coefficient ($R^2$) was calculated and agreement of measurements were analyzed as described by Bland & Altman [21].

## Results
### Limit of blank, limit of detection, linear dynamic range and PCR efficiency
For each of the 12 genes the analytical performance of the RT-qPCR assays was assessed. For a type I error of 5% the LoB was at a $C_q$ value of 40 for all genes (Table 1). The LoD of the 12 assays ranged from $C_q$ 35.1 to 37.2 (Table 1; supplemental data Figure 1 [see Additional file 1]).

All 11 RNA-specific assays were linear up to dilutions between $2^{-9}$ and $2^{-16}$ corresponding to $C_q$ values between 35.1 and 37.2 (Table 1; supplemental data Figure 2 [see Additional file 1]). Amplification efficiencies ranged from 76% to 101% with a mean efficiency of 88% (Table 1). The DNA-specific *HBB* PCR assay was linear up to a $C_q$ value of 35.3 (dilution: $2^{-8}$) and had an efficiency of 75%.

### Input range
For a multigene expression test it is essential to determine the acceptable range of input RNA within which the assay yields accurate results for all variants tested [13]. For that purpose a set of six breast cancer samples with different EP scores ranging from 2.5 to 11.5 were selected. Following RNA isolation different amounts of sample per reaction were assessed by the EndoPredict assay. The average of the $C_q$ values of the three reference genes ($C_q$-ARG) was used as surrogate for mRNA input. Although an increase of the 95% CI was observed above $C_q$-ARGs of 26 the EP score did not significantly change within an RNA input range of $C_q$-ARG between 20.5 and 28 (Figure 2A). Analysis of the individual genes showed that *STC2*, *IL6ST*, and *BIRC5* were the first analytes to drop out as the RNA amount was decreased (data not shown). In order to calibrate the $C_q$-ARG values to total RNA concentrations the amount of total RNA in a set of 45 samples was assessed (Figure 2B). The range of input RNA concentration without considerable change of the EP score was between 0.16 and 18.5 ng/µl corresponding to a about 100-fold difference.

### Precision
Precision of the multigene expression assay was evaluated under various stipulated operating conditions including day, day time, PCR instrument, position of the sample on the EndoPredict plate, plate lot, reagent lots, and operators and using three different test RNA samples from FFPE breast cancer tissue with a low EP score (2.4), a high EP score (13.5), and an EP score close to the decision point (4.9). In total, 160 EndoPredict tests (Figure 2C) were performed consisting of 5270 measured $C_q$ values of the RNA-specific PCR assays and 1280 $\Delta C_q$ values of the 8 genes-of-interest. The overall variability (standard deviation [SD]) of the EP scores was 0.15 (Table 2), which is 1% of the total EP score range

**Table 1 LoB, LoD, linear dynamic range, PCR efficiency of the 12 PCR assays included in EndoPredict**

| Gene | LoB | LoD [$C_q$ value] | Linear range [$\log_2$ dilution step] | Linear range [$C_q$ value] | Efficiency [%] |
|---|---|---|---|---|---|
| AZGP1 | 40.0 | 35.6 (34.2 - 36.4) | −13.1 to 0 | 35.6 to 20.5 | 81.9 (80.3 - 83.6) |
| CALM2 | 40.0 | 35.4 (34.3 - 36.0) | −14.0 to 0 | 35.4 to 21.6 | 101.4 (99.8 - 103.2) |
| BIRC5 | 40.0 | 36.3 (35.4 - 36.9) | −9.1 to 0 | 36.3 to 26.7 | 93.3 (90.5 - 96.2) |
| DHCR7 | 40.0 | 36.3 (35.4 - 36.8) | −10.9 to 0 | 36.3 to 24.5 | 89.9 (87.6 - 92.3) |
| IL6ST | 40.0 | 36.8 (35.7 - 37.5) | −11.5 to 0 | 36.8 to 23.3 | 80.7 (78.8 - 82.7) |
| MGP | 40.0 | 37.2 (35.2 - 38.2) | −13.9 to 0 | 37.2 to 20.2 | 76.3 (74.4 - 78.2) |
| OAZ1 | 40.0 | 36.6 (35.5 - 37.2) | −12.9 to 0 | 36.6 to 22.6 | 89.0 (87.6 - 90.4) |
| RBBP8 | 40.0 | 35.6 (34.7 - 36.1) | −9.4 to 0 | 35.6 to 25.9 | 96.3 (93.2 - 99.6) |
| STC2 | 40.0 | 35.1 (34.2 - 35.7) | −9.9 to 0 | 35.1 to 24.0 | 85.2 (82.6 - 87.9) |
| UBE2C | 40.0 | 36.0 (34.9 - 36.7) | −10.1 to 0 | 36.0 to 24.4 | 83.3 (81.1 - 85.7) |
| RPL37A | 40.0 | 36.0 (34.5 - 36.7) | −16.4 to 0 | 36.0 to 19.0 | 94.5 (92.9 - 96.1) |
| HBBV2 | 40.0 | 35.3 (33.6 - 36.2) | −7.6 to 0 | 35.3 to 25.9 | 75.4 (70.1 - 81.6) |

95% confidence intervals are indicated in brackets. For linear range a maximum deviation from linearity of 1 Ct value was accepted.

demonstrating robustness and high reproducibility of the test. Interestingly, the major part of the total noise (SD 0.14) was caused by the replicate-to-replicate noise of the PCR assays (repeatability) and was not associated with different operating conditions (reproducibility). The same was true for the variations of $C_q$ or $\Delta C_q$ values which showed overall standard deviations (total noise) of 0.20 and 0.12, respectively (Table 2). Repeatability and reproducibility of the individual gene-specific PCR assays are summarized in supplemental data Tables 1&2 (see Additional file 1).

### Verification of performance characteristics in an independent laboratory

The performance characteristics of the EndoPredict assay were verified in a routine laboratory at the Charité in Berlin to confirm that the test performs to specifications also in a routine diagnostic laboratory. The parameters verified were efficiency of the single PCR assays, precision, input range, and analytical accuracy with respect to reference values.

For assessment of linear range and efficiency two independent series of seven 1:2 dilutions of the reference RNA pools and DNA from FFPE tissue were generated. Each nucleic acid concentration was assessed four times. The 11 RNA-specific assays were linear over the whole range of concentrations analyzed (dilutions up to $2^{-7}$), the *HBB* assay up to dilution step $2^{-6}$ (Table 3). On average the efficiencies of the RNA assays were 84% and ranged from 78% to 98% which was within the pre-specified reference limits (Table 3). The efficiency of the *HBB* assay was 79%. Assessing the EP scores in each dilution step showed stable values down to an input RNA of $C_q$-ARG of 28 verifying the results of the studies at Sividon (Figure 3A).

Moreover, precision of the EndoPredict test was verified assessing the impact of the day, day time, the position of the sample on the 96-well EndoPredict plate, the reagent lot, and the operator on the reproducibility of the assay. The two RNAs from the tumors with the low and the high EP score were analyzed using the EndoPredict test resulting in 659 $C_q$ values, 160 $\Delta C_q$ values and 20 EP scores. The total variation (standard deviation) of the EP scores was 0.18 (Table 4) and thus almost identical to the variation of the EP scores generated at Sividon (Table 2). In the Charité laboratory variable noise induced by operating conditions had a similar impact on total noise as replicate noise. Standard deviations of $C_q$ values and $\Delta C_q$ values were 0.24 and 0.14 and therefore similar to those at Sividon.

Finally, the analytical accuracy of the EndoPredict assay performed in the Charité laboratory was examined. For that purpose, ten breast tumor samples were selected and the EP scores were determined at Sividon. These pre-determined scores ranging from 3.3 to 11.0 were used as reference values. Five of the cases were very close to the predefined cutoff of the EP score. The pre-specified aim for the verification study at Charité was that the difference between the EP score at Charité and the reference EP score was below 1.0 EP units for at least 9 of 10 samples. Charité received a 10 μm tissue section of each of the ten tumors, isolated RNA and performed the EndoPredict test. The aim of this verification study was achieved as the largest deviation from the reference value was 0.36 score units with a mean deviation of 0.15 (Figure 3B). Using the cutoff value of 5 to classify a sample in low or high risk of distant metastasis the concordance of classifications between Charité and Sividon was 90%. The discrepant sample was very close to the cutoff value (EP scores 5.04 vs 4.99). Moreover, an
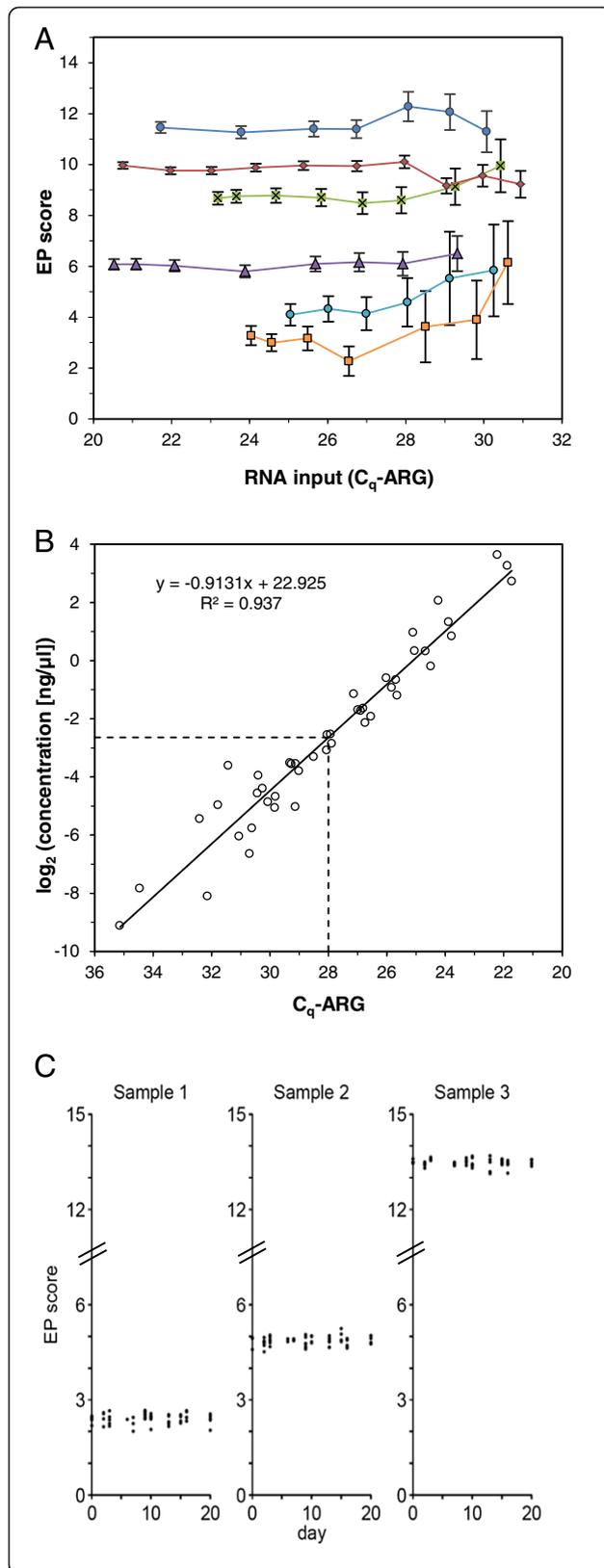
**Figure 2 RNA input range and reproducibility of EndoPredict.**
(**A**) EP scores depending on amount of input RNA. RNAs from 6 different FFPE samples were diluted and EP scores were assessed dependent on RNA input ($C_q$-ARG as surrogate marker). 95% confidence intervals (CI) of EP scores calculated from the noise model are indicated. (**B**) Correlation between $C_q$-ARG and total RNA concentration assessed by RIBOGREEN assay. Lower RNA input limit is indicated by dotted lines. (**C**) Reproducibility of 160 EP scores assessed in three different RNA samples (low risk, close to the decision point, high risk) over time (11 different working days distributed over 21 calendar days). Individual EP measurement results are indicated by dots.

excellent Pearson correlation coefficient of 0.995 ($R^2$) was found.

## Discussion

In this study, we showed by means of a defined analytical validation and verification process developed according to the CLSI guidelines that the RT-qPCR-based EndoPredict multianalyte gene expression test is a robust test that can be performed reproducibly and accurately. The resulting performance characteristics therefore meet the requirements needed for a diagnostic test. Moreover, we verified that a comparable performance with respect to assay efficiency, precision, and accuracy can also be achieved in a routine molecular diagnostic laboratory. In addition, this study provides the specifications for analytical verification of EndoPredict in molecular pathological laboratories.

Successful clinical validation of the EndoPredict score in two large clinical trials was published previously [1] resulting in a level of evidence of 1B according to the classification for prognostic biomarkers that has been

**Table 2 Overall variabilities and variabilities of the EP scores, $C_q$ values, and normalized $\Delta C_q$ values of all genes**

| | Standard deviations | | |
|---|---|---|---|
| | $C_q$ | $\Delta C_q$ | EP |
| **variables** | | | |
| day | 0.024 | 0.013 | 0.006 |
| day time | < 0.001 | 0.016 | <0.001 |
| PCR instrument | 0.027 | 0.009 | 0.037 |
| sample position | 0.028 | 0.015 | <0.001 |
| plate lot | 0.014 | 0.011 | <0.001 |
| reagent lot | 0.085 | 0.027 | <0.001 |
| operator | 0.003 | 0.014 | 0.043 |
| **total variable noise** | **0.097** | **0.042** | **0.057** |
| **replicate noise** | **0.176** | **0.111** | **0.136** |
| **total noise** | **0.201** | **0.119** | **0.147** |

Different operating conditions (variable noise, replicate noise, and total noise) are indicated as standard deviations of the $C_q$ values, $\Delta C_q$ values, and the EP scores.

**Table 3 Verification of linear dynamic range and PCR efficiency at the molecular-pathological laboratory at the Charité**
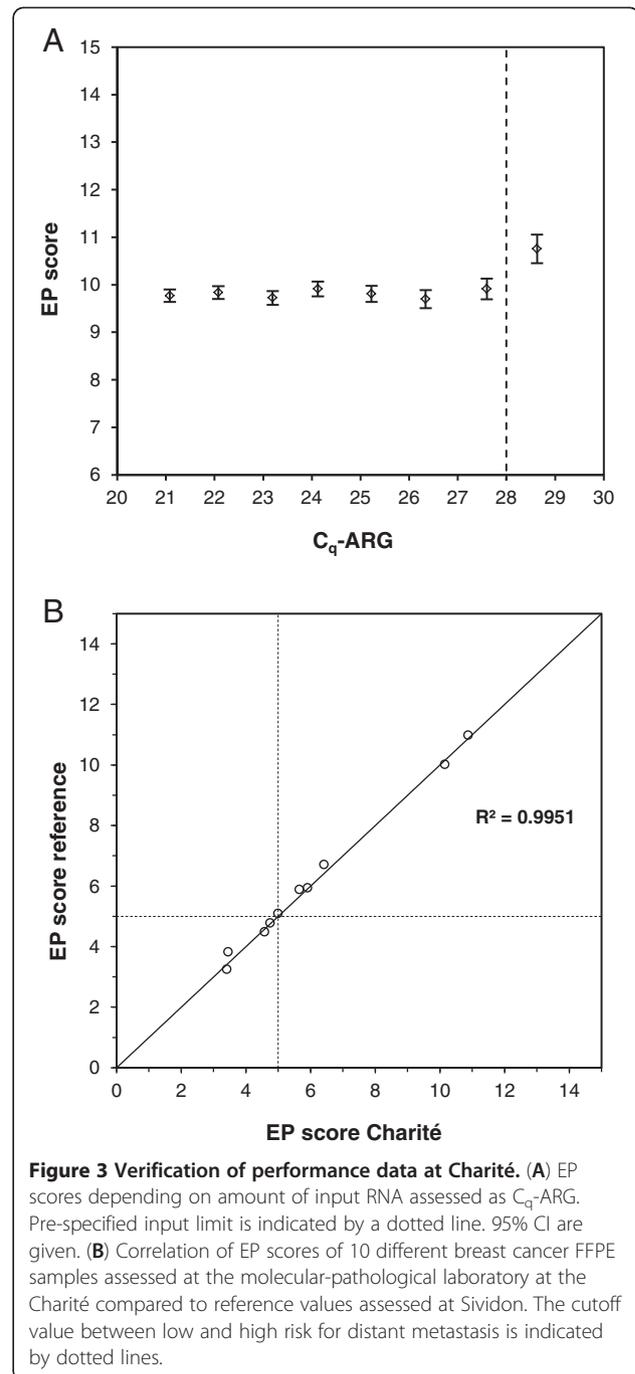
| Gene | Linear range [log$_2$ dilution step] | Linear range [C$_q$ value] | Efficiency [%] |
|---|---|---|---|
| AZGP1 | −7 to 0 | 28.5 to 20.5 | 82.7 (79.1 - 86.7) |
| CALM2 | −7 to 0 | 28.7 to 21.5 | 97.6 (95.7 - 99.7) |
| BIRC5 | −7 to 0 | 34.2 to 26.6 | 89.7 (86.4 - 93.2) |
| DHCR7 | −7 to 0 | 32.4 to 24.4 | 83.2 (79.2 - 87.7) |
| IL6ST | −7 to 0 | 31.7 to 23.3 | 78.0 (74.4 - 81.9) |
| MGP | −7 to 0 | 28.9 to 20.0 | 72.3 (68.1 - 77.2) |
| OAZ1 | −7 to 0 | 30.4 to 22.6 | 86.7 (83.6 - 90.1) |
| RBBP8 | −7 to 0 | 33.5 to 25.8 | 87.6 (84.4 - 91.0) |
| STC2 | −7 to 0 | 32.1 to 23.9 | 81.2 (77.7 - 85.0) |
| UBE2C | −7 to 0 | 32.7 to 24.3 | 78.1 (74.4 - 82.2) |
| RPL37A | −7 to 0 | 26.6 to 18.7 | 85.3 (83.1 - 87.5) |
| HBBV2 | −6 to 0 | 32.7 to 25.6 | 79.0 (74.7 - 83.9) |

95% confidence intervals are indicated in brackets. For linear range a maximum deviation from linearity of 1 Ct value was accepted.

proposed by Simon et al. [22]. The clinical validation studies which were performed within a prospective-retrospective design showed that the test predicted distant metastasis in patients with primary ER-positive, HER2-negative breast cancer and provided significant prognostic information beyond standard parameters to assess the risk of metastasis. Both, clinical and analytical validations of the EndoPredict multigene assay now fulfill the recently published recommendations for translating omics-based tests from research laboratory to clinical practice [23].

The difficulty of a comprehensive analytical validation of a highly complex genomic multianalyte assay such as EndoPredict is that there are no evaluation guidelines to cover all relevant aspects. Therefore, we specified a validation program for both, the 12 single analyte assays and the combined multigene assay. Where applicable, established guidelines of the CLSI as well as the MIQE (Minimum Information for Publication of Quantitative Real-Time PCR Experiments) guidelines were followed [13,18]. Therefore, the analytical validation process conformed to the standard assay validation formats as close as reasonably possible.

Besides EndoPredict, other prognostic multigene expression tests for patients with breast or other cancers like MammaPrint [8], Oncotype DX Breast Cancer or Colon Cancer Assays [9,10], or PAM50 [11,12] are commercially available. Analytical performance characteristics of these tests, which must be performed in reference laboratories, were published only for some tests such as MammaPrint [24] or the two Oncotype DX assays [25,26]. Here, we assessed the analytical performance data of



**Figure 3 Verification of performance data at Charité.** (**A**) EP scores depending on amount of input RNA assessed as C$_q$-ARG. Pre-specified input limit is indicated by a dotted line. 95% CI are given. (**B**) Correlation of EP scores of 10 different breast cancer FFPE samples assessed at the molecular-pathological laboratory at the Charité compared to reference values assessed at Sividon. The cutoff value between low and high risk for distant metastasis is indicated by dotted lines.

EndoPredict and provide further evidence that EndoPredict is suitable for decentralized testing in specialized molecular pathological laboratories which is a unique feature not shown for the other RNA-based multigene expression tests. On one hand this conclusion is supported by the verification of the performance characteristics and the accuracy of the measurements in a routine laboratory. On the other hand, the data of the precision study in two different laboratories showed a total variation (standard deviation) of

**Table 4 Verification of overall variability and variability of the EP scores, $C_q$ values, and normalized $\Delta C_q$ values at the molecular-pathological laboratory at the Charité**

|  | Standard deviations | | |
|---|---|---|---|
|  | $C_q$ | $\Delta C_q$ | EP |
| **variables** |  |  |  |
| day | 0.034 | 0.032 | 0.093 |
| day time | 0.062 | 0.073 | 0.074 |
| sample position | <0.001 | 0.016 | <0.001 |
| reagent lot | 0.182 | 0.031 | <0.001 |
| operator | 0.035 | 0.073 | 0.041 |
| **total variable noise** | **0.198** | **0.114** | **0.125** |
| **replicate noise** | **0.131** | **0.085** | **0.126** |
| **total noise** | **0.238** | **0.142** | **0.178** |

Different operating conditions (variable noise, replicate noise and total noise) are indicated as standard deviations of the $C_q$ values, $\Delta C_q$ values, and the EP scores.

the EP score of 0.15 and 0.18, respectively, which is about 1% of the total EP score range from 0 to 15. This is similar to the variation of the Oncotype DX Breast Cancer Assay performed in one centralized company laboratory [25]. Interestingly, the main factor influencing total noise was not one of the individual variables tested, such as operator, PCR machine, day, day-time, or reagent lots but rather replicate-to-replicate noise which is assumed to be caused by PCR technique-inherent noise. In order to account for PCR-inherent noise the EndoPredict algorithm includes a "noise model" estimating the acceptable variance of replicate noise with respect to the $C_q$ value. On this basis outlier elimination is accomplished [1]; one of its principal requirements is triplicate measurements.

Robust performance of the EndoPredict test in a decentralized setting was also shown in a successful prospective proficiency testing study including seven different molecular pathology laboratories [2]. In this study, 69 out of 70 EndoPredict measurements were within the pre-specified range, 100% of the samples were correctly classified as low or high risk of metastasis, and the total variation (standard deviation) of all measurements was 0.25 units, corresponding to 1.7% of the whole range of the EP score. The slightly larger total variation in comparison to the results from the analytical precision study where large RNA pools were used might be due to the fact that in the proficiency testing the participants received tissue sections and not RNA for testing, thus including variation induced by tumor heterogeneity.

A recent study assessed the variability of Ki-67 immunohistochemistry, which is a standard antibody-based diagnostic test in pathology used for treatment decision making in luminal breast cancer [27]. The authors found standard deviations of Ki-67 results obtained by 15

pathologists on centrally stained slides of three breast carcinomas ranging from 21.7% to 24.1%. Interestingly, even clear guidelines how to assess Ki-67 could not improve variability.

Although our results might suggest a higher reproducibility of the PCR-based test using standardized instruments and reagents, it is important to know that the pathological laboratories involved in this technical verification study as well in the proficiency testing of EndoPredict were highly experienced in molecular work. Therefore, the results might be different in laboratories with less molecular diagnostic experience and ongoing quality control by periodical round robin tests might be reasonable.

A critical issue for accurate results from any diagnostic assay is the use of optimal sample material. For this validation study optimal FFPE tumor material was selected by experienced pathologists on the basis of adjacent HE-stained tissue slides. Hence, in order to obtain high-quality EndoPredict results an expert pathological evaluation of the tissue specimens is mandatory. Moreover, a highly standardized method for RNA isolation as used in this validation study and in the EndoPredict proficiency testing is necessary. In this study, the validation of the RNA extraction method was not a primary aim. The silica-coated magnetic beads-based method used and recommended for EndoPredict was thoroughly validated in previous studies showing a reliable analysis of RNA transcript levels by RT-qPCR in FFPE tissue [4-6]. Finally, the PCR platform used in this validation study was validated and CE-marked for diagnostic purposes by the manufacturer. It also supports robust performance of the EndoPredict assay. Therefore, change of isolation method, enzymes, and PCR platform might alter performance characteristics of this multianalyte assay.

## Conclusions

In conclusion, this study provides the analytical performance characteristics of the EndoPredict breast cancer prognosis multigene expression assay. They can be used as a reference for analytical verification of the test in molecular pathological laboratories. Moreover, the study shows a robust and reliable performance of the test and provides conclusive evidence that RT-qPCR-based quantitative multigene expression analysis of FFPE tissue samples is feasible in a decentral setting in molecular pathological laboratories. This is a major technical advance in comparison with existing prognostic multigene expression tests which are performed in central reference laboratories. Together with the clinical validation studies and the prospective analytical proficiency testing program with seven different molecular pathological laboratories these results provide the basis for the application of EndoPredict as a test to assess prognosis under endocrine therapy in clinical decision making in a decentralized environment.

## Additional file

**Additional file 1: Supplemental Data:** File contains detailed descriptions of methods as well as supplemental tables and figures which could not be included in the main manuscript.

### Abbreviations

EP: Endopredict; ER: Estrogen receptor; FFPE: Formalin-fixed, paraffin-embedded; RT-qPCR: Reverse transcription quantitative real-time PCR; LoD: Limit of detection; $C_q$: Quantification cycle; GOI: Gene-of-interest; LoB: Limit of blank; $C_q$-ARG: Average $C_q$ value of reference genes; CI: Confidence interval.

### Competing interests

### Author's contributions

RK, KB, JCB, KEW, MS, MF, PD, CP, MD, CD were involved in conception and design of the study as well as in analysis or interpretation of data. JP, BVS, FH, CR, MA, TR, CW, KF, BMM acquired, analyzed and interpreted data. All authors have been involved in drafting or critically revising the manuscript and approved the manuscript.

### Acknowledgements

### Author details

[1]Sividon Diagnostics GmbH, Nattermannallee 1, 50829, Cologne, Germany. [2]Institute of Pathology, Charité Hospital, Campus Mitte, Berlin, Germany. [3]Department of Gynecology and Obstetrics, University of Mainz, Mainz, Germany. [4]Department of Medicine I, Medical University of Vienna, Vienna, Austria. [5]Department of Surgery, Medical University of Vienna, Vienna, Austria.

### References

1. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, Dietze O, Greil R, Jelen A, Sevelda P, *et al*: **A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors.** *Clin Cancer Res* 2011, **17**(18):6012–6020.
2. Denkert C, Kronenwett R, Schlake W, Bohmann K, Penzel R, Weber KE, Hofler H, Lehmann U, Schirmacher P, Specht K, *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**(3):251–259.
3. Muller BM, Brase JC, Haufe F, Weber KE, Budzies J, Petry C, Prinzler J, Kronenwett R, Dietel M, Denkert C: **Comparison of the RNA-based EndoPredict multigene test between core biopsies and corresponding surgical breast cancer sections.** *J Clin Pathol* 2012, **65**(7):660–662.
4. 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):1719–1727.
5. Muller BM, Kronenwett R, Hennig G, Euting H, Weber K, Bohmann K, Weichert W, Altmann G, Roth C, Winzer KJ, *et al*: **Quantitative determination of estrogen receptor, progesterone receptor, and HER2 mRNA in formalin-fixed paraffin-embedded tissue–a new option for predictive biomarker assessment in breast cancer.** *Diagn Mol Pathol* 2011, **20**(1):1–10.
6. Hennig G, Gehrmann M, Stropp U, Brauch H, Fritz P, Eichelbaum M, Schwab M, Schroth W: **Automated extraction of DNA and RNA from a single formalin-fixed paraffin-embedded tissue section for analysis of both single-nucleotide polymorphisms and mRNA expression.** *Clin Chem* 2010, **56**(12):1845–1853.
7. Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, *et al*: **Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.** *J Clin Oncol* 2010, **28**(1):105–113.
8. Veer LJ v 't, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, *et al*: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**(6871):530–536.
9. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, *et al*: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**(27):2817–2826.
10. Gray RG, Quirke P, Handley K, Lopatin M, Magill L, Baehner FL, Beaumont C, Clark-Langone KM, Yoshizawa CN, Lee M, *et al*: **Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer.** *J Clin Oncol* 2011, **29**(35):4611–4619.
11. 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):1160–1167.
12. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J, *et al*: **A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer.** *Clin Cancer Res* 2010, **16**(21):5222–5232.
13. CLSI: *Verification and validation of multiplex nucleic acid assays; approved guideline. CLSI document MM17-A*. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
14. Paska C, Bogi K, Szilak L, Tokes A, Szabo E, Sziller I, Rigo J Jr, Sobel G, Szabo I, Kaposi-Novak P, *et al*: **Effect of formalin, acetone, and RNAlater fixatives on tissue preservation and different size amplicons by real-time PCR from paraffin-embedded tissue.** *Diagn Mol Pathol* 2004, **13**(4):234–240.
15. CLSI: *Use of external RNA controls in gene expression assays; approved guideline. CLSI document MM16-A*. Wayne, PA: Clinical and Laboratory Standards Institute; 2006.
16. NCCLS: *Protocols for determination of limits of detection and limits of quantitation; approved guideline. NCCLS document EP17-A*. Wayne, PA: NCCLS; 2004.
17. NCCLS: *Evaluation of the linearity of quantitative measurement procedures: a statistical approach; approved guideline. NCCLS document EP6-A*. Wayne, PA: NCCLS; 2003.
18. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, *et al*: **The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.** *Clin Chem* 2009, **55**(4):611–622.
19. International Organisation of Standardization: *ISO 5725-2. Accuracy (trueness and precision) of measurement methods and results – Part 2: basic method for the determination of repeatability and reproducibility of a standard measurement method. British Standards Institution BS ISO 5725-2*. Geneva: International Organization of Standardization; 1994.
20. NCCLS: *Evaluation of precision performance of quantitative measurement methods; approved guideline-second edition. NCCLS document EP5-A2*. Wayne, PA: NCCLS; 2004.
21. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1**(8476):307–310.
22. Simon RM, Paik S, Hayes DF: **Use of archived specimens in evaluation of prognostic and predictive biomarkers.** *J Natl Cancer Inst* 2009, **101** (21):1446–1452.
23. National Research Council: *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington, DC: The National Academies Press; 2012.
24. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, *et al*: **Converting a breast cancer microarray signature into a high-throughput diagnostic test.** *BMC Genomics* 2006, **7**:278.
25. Cronin M, Sangli C, Liu ML, Pho M, Dutta D, Nguyen A, Jeong J, Wu J, Langone KC, Watson D: **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**(6):1084–1091.

26. Clark-Langone KM, Sangli C, Krishnakumar J, Watson D: **Translating tumor biology into personalized treatment planning: analytical performance characteristics of the Oncotype DX Colon Cancer Assay.** *BMC Cancer* 2010, **10**:691.

27. 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.