Clinical Chemistry 59:6
000–000 (2013)

Establishment and Validation of Analytical Reference Panels for the Standardization of Quantitative *BCR-ABL1* Measurements on the International Scale

Helen E. White,[1,2] John Hedges,[3] Israel Bendit,[4] Susan Branford,[5] Dolors Colomer,[6] Andreas Hochhaus,[7] Timothy Hughes,[8] Suzanne Kamel-Reid,[9] Dong-Wook Kim,[10] Vijay Modur,[11†] Martin C. Müller,[12] Katia B. Pagnano,[13] Fabrizio Pane,[14,15] Jerry Radich,[16] Nicholas C.P. Cross,[1] and Emmanuel Labourier[2*]

BACKGROUND: Current guidelines for managing Philadelphia-positive chronic myeloid leukemia include monitoring the expression of the *BCR-ABL1* (breakpoint cluster region/c-abl oncogene 1, non-receptor tyrosine kinase) fusion gene by quantitative reverse-transcription PCR (RT-qPCR). Our goal was to establish and validate reference panels to mitigate the interlaboratory imprecision of quantitative *BCR-ABL1* measurements and to facilitate global standardization on the international scale (IS).

METHODS: Four-level secondary reference panels were manufactured under controlled and validated processes with synthetic Armored RNA Quant molecules (Asuragen) calibrated to reference standards from the WHO and the NIST. Performance was evaluated in IS reference laboratories and with non–IS-standardized RT-qPCR methods.

RESULTS: For most methods, percent ratios for *BCR-ABL1* e13a2 and e14a2 relative to *ABL1* or *BCR* were robust at 4 different levels and linear over 3 logarithms, from 10% to 0.01% on the IS. The intraassay and interassay imprecision was <2-fold overall. Performance was stable across 3 consecutive lots, in multiple laboratories, and over a period of 18 months to date. International field trials demonstrated the commutability of the reagents and their accurate alignment to the IS within the intra- and interlaboratory imprecision values of IS-standardized methods.

CONCLUSIONS: The synthetic calibrator panels are robust, reproducibly manufactured, analytically calibrated to the WHO primary standards, and compatible with most *BCR-ABL1* RT-qPCR assay designs. The broad availability of secondary reference reagents will further facilitate interlaboratory comparative studies and independent quality-assessment programs, which are of paramount importance for worldwide standardization of *BCR-ABL1* monitoring results and the optimization of current and new therapeutic approaches for chronic myeloid leukemia.

Residual disease in patients with chronic myeloid leukemia undergoing therapy with tyrosine kinase inhibitors is assessed by measuring the quantity of transcripts of the *BCR-ABL1*[17] (breakpoint cluster region/c-abl oncogene 1, non-receptor tyrosine kinase) fusion gene in peripheral white blood cells. This analysis is carried out with analytically sensitive molecular tests based on quantitative reverse-transcription PCR (RT-qPCR)[18] technology *(1–3)*; however, the wide array of preanalytical and analytical methods used worldwide and the lack of consensus guidelines have led to large variation in quantitative *BCR-ABL1* measurements, which hinder interlaboratory comparative studies, patient portability, and harmonized definition of treatment response. In 2005, standardization of *BCR-ABL1*

[1] National Genetics Reference Laboratory Wessex, Salisbury District Hospital, Salisbury, UK; [2] Faculty of Medicine, University of Southampton, Southampton, UK; [3] Asuragen, Inc., Austin, TX; [4] Department of Hematology and Hemotherapy, Medical School at University of São Paulo, São Paulo, Brazil; [5] Department of Molecular Pathology, SA Pathology and the University of Adelaide, Adelaide, Australia; [6] Unitat d'hematopatologia, Hospital Clinic, Institut d'investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain; [7] Abteilung Hämatologie/Onkologie, Klinik für Innere Medizin II, Universitätsklinikum Jena, Jena, Germany; [8] Department of Hematology, SA Pathology and the University of Adelaide, Adelaide, Australia; [9] Department of Pathology, University Health Network, Toronto, Ontario, Canada; [10] Cancer Research Institute, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea; [11] Novartis Oncology, Cambridge, MA; [12] III Medizinische Klinik, Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Germany; [13] Hematology and Hemotherapy Center, University of Campinas, Campinas, Brazil; [14] Division of Hematology, University of Naples Federico II, Naples, Italy; [15] CEINGE, Biotecnologie Avanzate, Naples, Italy; [16] Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

* Address correspondence to this author at: Asuragen, Inc., 2150 Woodward St., Ste. 100, Austin, TX 78744. Fax 512-681-5201; e-mail elabourier@asuragen.com.

† Current address: HTG Molecular Diagnostics, Tucson, AZ.

Received October 2, 2012; accepted February 11, 2013.
Previously published online at DOI: 10.1373/clinchem.2012.196477

[17] Nonstandard abbreviations: RT-qPCR, quantitative reverse-transcription PCR; IRIS, International Randomized Study of Interferon vs STI571; MMR, major molecular response; CF, conversion factor; ARQ, Armored RNA Quant; CP, correction parameter; LOA, limits of agreement.

[18] Human genes: *ABL1*, c-abl oncogene 1, non-receptor tyrosine kinase; *BCR*, breakpoint cluster region.

reporting across tests and laboratories was established through a common international scale (IS) anchored to the baseline *BCR-ABL1* expression levels of the IRIS study (International Randomized Study of Interferon vs STI571) *(4, 5)*. Achieving a 3-logarithm reduction from the standardized baseline level (defined as 100% IS) corresponds to the so-called major molecular response (MMR or 0.1% IS) and is associated in some studies with improvements in the long-term response to tyrosine kinase inhibitors and progression-free survival *(6, 7)*. In addition, the degree of molecular response at early time points has recently been shown to be predictive of overall survival *(8, 9)*.

Because of the prognostic value of MMR, a key goal of the IS standardization effort was to eliminate—or correct—the relative differences between MMR ascertainment and response rates across laboratories. The evaluation of representative sets of clinical specimens spanning at least 3 logarithms of detectable residual disease in both local and IS reference laboratories has provided an important proof of principle by significantly improving MMR concordance rates in about 50% of *BCR-ABL1* testing laboratories *(10, 11)*. This approach also suffers from some obvious limitations, however. For example, both conversion factor (CF) establishment and validation require a set of 30–40 clinical samples *(10)*, which can make the overall process lengthy and costly. Because there is no mechanism for continuous monitoring of performance, local and IS reference laboratories must either revalidate their CF each time a local method parameter is changed or risk drifting from the IS over time. Perhaps one of the most limiting features of the CF conversion process is that without sharing and testing of a common set of reference samples on a global scale, each sample exchange is an independent performance assessment relative to a single IS reference method.

It is likely that the comparability of *BCR-ABL1* results would be improved by the availability of independent, broadly accessible external quality-assessment programs, such as the scheme recently launched by the United Kingdom National External Quality Assessment Service *(12–14)*. White blood cell lysates *(11, 15)* and leukemia cell lines *(14, 16)* have been successfully used in international comparative studies across *BCR-ABL1* testing platforms and assay designs. In 2009, the Expert Committee on Biological Standardization of the WHO approved the first WHO International Genetic Reference Panel for quantification of *BCR-ABL1* mRNA by RT-qPCR *(17, 18)*. This 4-level panel consists of freeze-dried e14a2-positive K562 cells diluted in a background of *BCR-ABL1*–negative HL-60 cells, with assigned reference values corresponding to the mean IS percent ratios obtained by repeat testing in multiple IS-standardized laboratories *(18)*. Unfortunately, the supply of WHO primary standards is limited, and their availability has been restricted to manufacturers of secondary reference standards *(17, 18)*. Armored RNA Quant (ARQ) (Asuragen) is a mature technology that may be well suited for the development of such secondary reference materials. These stable and commutable synthetic molecules have been extensively validated with multiple qualitative and quantitative tests for infectious diseases, and their performance has been fully evaluated against recognized biological international standards *(19–21)*. More recently, their potential for use in oncology applications has also emerged *(22, 23)*. The aim of the broad collaborative study we describe was to develop robust synthetic ARQ reference panels calibrated to the mean IS percent ratios of the WHO primary standards in order to support the ongoing worldwide effort to standardize *BCR-ABL1* quantitative measurements on the IS.

Materials and Methods

ARQ PREPARATIONS

Sequences of interest were cloned into pCR2.1 with the TA Cloning Kit (Life Technologies) and pairs of primers specific for each target (see the Data Report of Prototype Field Trial Evaluation and Fig. 1 in the online Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue6). Inserts were subcloned in pT7T3–18-Kan (Asuragen) for RNA synthesis by in vitro transcription and subsequent packaging, as previously described *(24)*. The sequences of the 4 synthetic transcripts and an overview of the validated ARQ manufacturing and QC processes are shown in Figs. 1 and 2 in the online Data Supplement. The final concentrations of the packaged RNAs were measured by quantifying the phosphate groups within each transcript (see Supplemental Data Protocols in the online Data Supplement). Large-scale bulk ARQ preparations were formulated in buffer (10 mmol/L Tris, pH 7.0, 100 mmol/L NaCl, 1 mmol/L $MgCl_2$) and stored at 2 °C–8 °C. Panels formulated as described in the online Supplemental Data Protocols were stored below −15 °C.

EXPERIMENTAL DESIGN AND DATA ANALYSES

Protocols and additional information on the experimental design for each Results section below are provided in the online Supplemental Data Protocols under their respective headings. All data analyses assumed a normal-like distribution of results after $\log_{10}$ transformation of the percent ratios. Correction parameters (CPs) were calculated as described in the instructions for use of the ARQ IS calibrator panels. Correlation between results was assessed by least squares linear re-

gression analyses and calculation of the corresponding slopes, standard errors of the slope, coefficients of determination (square of the Pearson product–moment correlation coefficient, or r^2), and standard errors of the residuals (root square of the sum of squared residual differences divided by the degrees of freedom). Differences between results were assessed by estimating the mean bias with the Bland–Altman method and calculating the 95% limits of agreement (LOA) *(10, 25)*. All *P* values were calculated at the 95% confidence level.

Results

ESTABLISHMENT OF COMMUTABLE SYNTHETIC MATERIALS

To generate a reproducible and lasting source of reference material, synthetic and nuclease-resistant *ABL1*, *BCR*, and *BCR-ABL1* transcripts were prepared on a large scale with the validated ARQ manufacturing process (see Figs. 1 and 2 in the online Data Supplement). Transcripts were designed to contain the exonic sequences most commonly targeted by *BCR-ABL1* RT-qPCR assays (Fig. 1A; see Fig. 1 in the online Data Supplement). Four-level pilot panels containing a variable number of copies of e13a2 or e14a2 fusion transcripts in a constant background of control transcripts were evaluated in a large international field trial across 29 laboratories in 15 countries (see Report of Prototype Field Trial Evaluation in the online Data Supplement). This pilot study demonstrated the compatibility of the synthetic panels with independent RT-qPCR methods with 7 different RNA-extraction protocols and 13 different platforms and highlighted their potential as commutable reference materials.

DEVELOPMENT OF PROTOTYPE CALIBRATOR PANELS

On the basis of the results of the pilot study and current recommendations for IS standardization *(4, 5, 10, 11)*, 4-level ARQ panels were formulated to cover a representative interval of percent ratios from >1% to <0.1% and were calibrated to the mean IS percent ratios of the WHO reference panel (Fig. 1B). With this design, testing of the secondary ARQ IS calibrator panels in local laboratories would theoretically enable indirect comparison to the WHO primary standard, derivation of analytical CPs, and correction of local percent ratios to IS percent ratios if required (raw local percent ratio × CP = IS_{Cal} percent ratio, where IS_{Cal} is the IS calibrator panel; Fig. 1B). Preliminary evaluation of the prototype panels showed that the ARQ calibrators could be processed with different RNA-extraction methods optimized for the quantitative recovery of nucleic acids or be added directly to the reverse transcription reaction after a short heat-denaturation step (see Fig. 3 in the online Data Supplement). Any variation was streamlined and removed during the preanalytical steps by performing all subsequent analyses with the simple and rapid heat-lysis protocol. Repeat testing in 2 IS reference laboratories showed linear and reproducible detection of both e13a2 and e14a2 over 3 logarithms of IS percent ratios (Fig. 1C). Accurate panel formulation and compatibility with the IS were confirmed for the *ABL1* control transcript by the RT-qPCR method recommended by the Europe Against Cancer program *(26, 27)* (Fig. 1C, left panels) and for the *BCR* control transcript with the method used since 2001 to maintain the MMR reference value of the IRIS trial *(2, 10, 28)* (Fig. 1C, right panels). However, synthetic ARQ *BCR-ABL1* cDNA was inefficiently amplified with a third established qPCR primer/probe design *(3, 11, 15)*, producing a loss of calibrator 4 detection with or without RNA extraction and with different RT protocols (data not shown).

ANALYTICAL VALIDATION

To fully characterize the e13a2 and e14a2 ARQ IS calibrator panels, we conducted a series of analytical studies over a 2-year period. Duplicate analyses performed on 192 measures from 12 independent runs indicated that 95% of the paired percent ratios were within a 1.92-fold difference from each other with the expected increase in imprecision for lower percent ratios (Table 1 and Fig. 2A). An interassay imprecision study of the mean of triplicate test results from 8 runs over 6 months showed the same pattern, with CVs of 14%, 23%, 28%, and 37% for calibrators 1–4 and a 95% LOA of 1.61-fold overall (Table 1 and Fig. 2B). In both studies, the 4 calibrators covered 3 logarithms of percent ratios with a precisely linear fit and no significant difference between the *ABL1* and *BCR* control transcripts (Table 1; see Fig. 4 in the online Data Supplement). Stability studies confirmed that each calibrator was reproducibly detected after 3, 6, 12, or 18 months of storage, with percent ratios all within the imprecision established for each calibrator level at time 0 (Fig. 2C). Finally, the reproducibility of the overall process was validated with 3 independent consecutive lots of e13a2 and e14a2 calibrators manufactured over a 12-month period and with each lot being tested in 4 independent runs (Fig. 2D). The same interval of percent ratios was observed with a precisely linear relationship between the 4 calibrator levels across 3 logarithms; similar levels of imprecision were observed for the 3 lots (18%–43% for the level-specific CV; Table 1). The evaluation of 2 independent lots in 2 IS-standardized laboratories 6 months apart also confirmed accurate and linear formulation across 3 logarithms of IS percent ratios (Fig. 1C and data not shown).

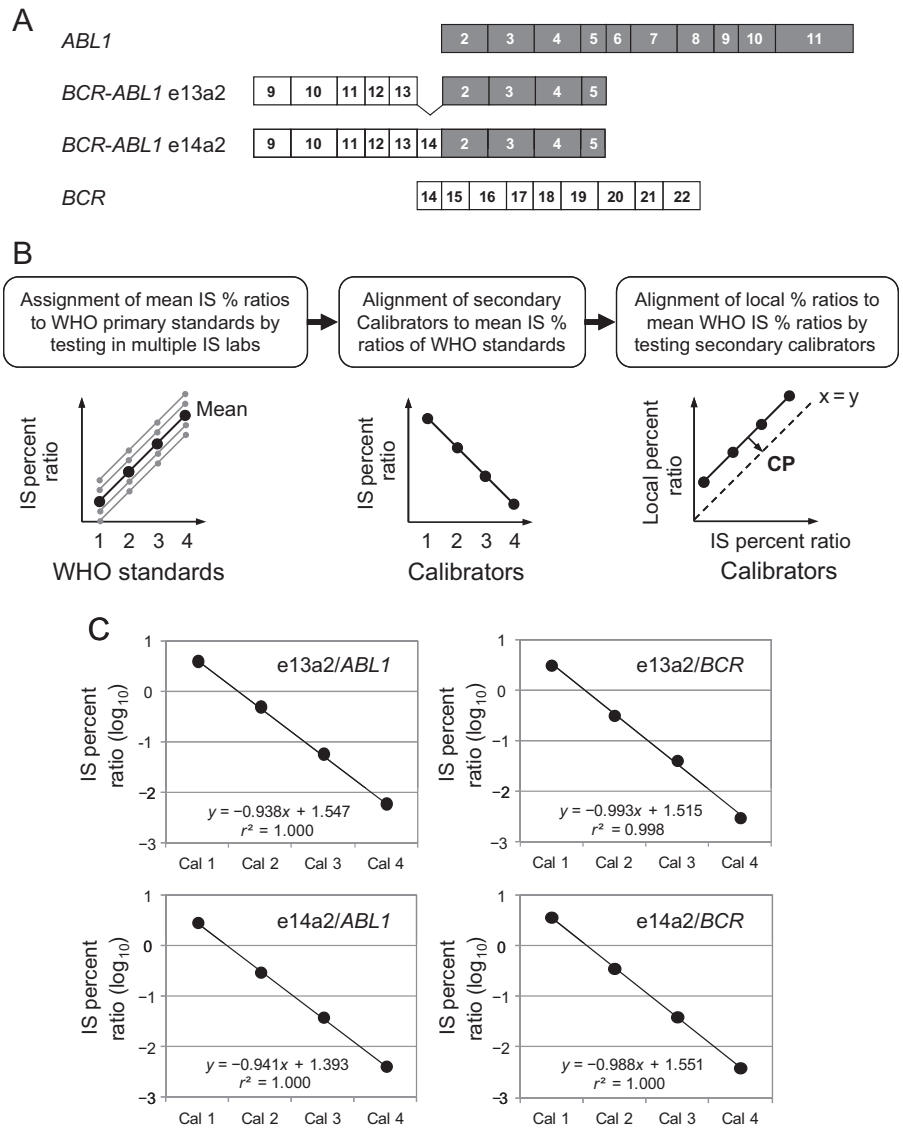

Fig. 1. Synthetic calibrator panels for IS standardization.

(A), Schematic representation of the 4 synthetic transcripts packaged in ARQ molecules. Exon numbers according to the reference sequences NM 005157.3 (*ABL1*) and NM 004327.3 (*BCR*) are indicated in color-coded boxes. *ABL1*, nucleotide (nt) 60 of exon 2 to nt 93 of exon 11; *BCR*, nt 42 of exon 14 to nt 148 of exon 22; *BCR-ABL1*, nt 90 of *BCR* exon 9 to nt 78 of *ABL1* exon 5. (B), Process to align a local-laboratory RT-qPCR method (local percent ratio) to the mean reference values of the WHO primary standards through secondary synthetic calibrators. (C), Compatibility with IS-standardized methods. The graphs show the mean IS percent ratios obtained by testing in 3 independent runs of the 4-level e13a2 (top) or e14a2 (bottom) prototype panels with the heat-lysis protocol and 2 IS-standardized methods *(2, 10, 26–28)* using either *ABL1* (left) or *BCR* (right) as the endogenous control transcript. Cal, calibrator panel.

VALIDATION IN IS-STANDARDIZED LABORATORIES

To validate the accuracy of the ARQ IS calibrator panels, we conducted a second international field trial in 7 countries across 8 laboratories with either *ABL1* (n = 4) or *BCR* (n = 4) control transcripts and various platforms and assay designs *(2, 26–29)*. The e13a2 and e14a2 panels were tested in 3 independent runs at each site (192 measures overall), and the nominal IS percent ratios assigned to each calibrator level were compared directly with the IS percent ratios determined with in-

Table 1. Summary of analytical-characterization studies.

	Intraassay	Interassay	Lot 1	Lot 2	Lot 3
ABL1					
Measures, n	96	64	32	32	32
Δ Percent ratio, $\log_{10}$	3.02	3.06	2.97	3.07	2.97
Slope	0.987	−1.014	−0.999	−1.022	−0.996
SES[a]	0.013	0.015	0.017	0.024	0.015
r^2	0.984	0.986	0.991	0.983	0.993
SER	0.144	0.138	0.110	0.154	0.095
BCR					
Measures, n	96	64	32	32	32
Δ Percent ratio, $\log_{10}$	3.01	3.00	2.97	3.00	2.96
Slope	0.999	−0.995	−0.998	−1.000	−0.990
SES	0.013	0.015	0.018	0.028	0.014
r^2	0.984	0.988	0.991	0.976	0.994
SER	0.146	0.127	0.112	0.179	0.087
Overall					
Measures, n	192	128	64	64	64
Percent ratio, $\log_{10}$	3.02	3.03	2.97	3.04	2.96
Slope	0.993	−1.005	−0.999	−1.011	−0.993
SES	0.009	0.011	0.014	0.019	0.012
r^2	0.984	0.986	0.988	0.979	0.992
SER	0.144	0.133	0.124	0.168	0.104
Cal 1 CV, %	7.0	14	18	22	18
Cal 2 CV, %	10	23	27	27	25
Cal 3 CV, %	15	28	29	41	25
Cal 4 CV, %	31	37	39	43	28

[a] SES, standard error of slope; SER, standard error of residuals, Cal, calibration panel.

dependent IS-standardized methods and laboratory-specific CFs (raw local percent ratio × CF = IS percent ratio). The nominal and measured IS percent ratios were tightly correlated across the entire scale for both panels (Table 2). The mean bias between the nominal and measured IS percent ratios for the 4 calibrator levels combined was <2.25-fold in all laboratories but one (e13a2 panel in laboratory 8) and 1.37-fold overall (Fig. 3A). A comparison across laboratories showed no significant trend in the relative bias between IS-standardized methods ($P > 0.05$; Fig. 3B, left panel). The level-specific interlaboratory CVs (18%–52%) were in the same range as previously reported for the 4-level WHO primary standards (14%–57%) *(18)*, with similar bias distributions and overall 95% LOA across 3 logarithms of IS percent ratios (Fig. 3B, right panel). Importantly, the calibrators were reproducibly detected at all levels, thereby enabling the derivation of analytical CPs and correction of the local raw percent ratios with laboratory-specific CPs (raw local percent ratio × CP = IS_{Cal} percent ratio). As expected, the resulting postcorrection bias was null [i.e., 1.00-fold, Fig. 3A (all data combined); see Fig. 5 in the online Data Supplement]. The relative difference between the calculated CP and the validated CF (CP/CF ratios) was 1.04- to 3.06-fold in distinct laboratories, corresponding to a minimum agreement in MMR classification between CP-corrected and CF-converted IS percent ratios of 83.8%–99.4% (see Protocols, Calculation of MMR Accuracy, in the online Data Supplement).

USAGE VALIDATION

To further validate the interlaboratory imprecision and the intended use of the ARQ IS calibrators, we evaluated the synthetic panels in 8 non–IS-standardized laboratories with the same commercial RT-qPCR assay that had previously been shown to be compatible with reporting on the IS *(22)*. Each laboratory followed the

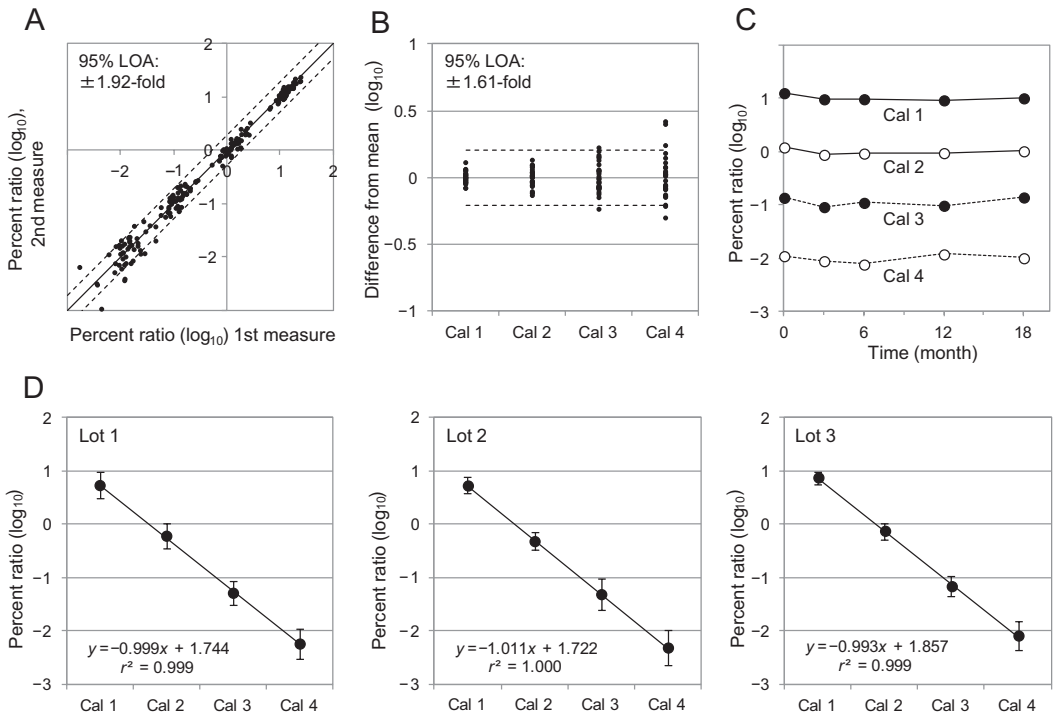

Fig. 2. Analytical-characterization studies.

(A), Correlation between 192 duplicate percent ratio measures. Black line, equality line (i.e., second measure = first measure). Dashed lines, overall 95% LOA. (B), Distribution of percent ratios determined in 8 independent runs for each calibrator level (32 measures per level). Dash lines, overall 95% LOA. (C), Mean percent ratios obtained by testing the same lot of calibrators panels stored at −15 °C to −30 °C over time. (D), Correlation of mean percent ratios (16 measures per level) for 3 independent lots of calibrators. The error bars represent the 95% LOA for each calibrator level and lot. Cal, calibrator panel.

calibrators' instructions for use, which recommended 3 independent runs on different days for each panel (228 measures overall). The linear regression coefficients were similar (mean slope, 1.029; 95% CI, 0.994–1.063), indicating strong correlation and absence of trend in the bias relative to the nominal IS percent ratios (Table 3). Calculations of the mean bias for each laboratory produced CPs ranging from 0.369 to 0.648 with 95% LOA of 1.28- to 2.87-fold. The mean CP was 0.50 (95% CI, 0.43–0.57). After correction of the individual raw percent ratios with either the laboratory-specific CP or the mean CP, the residual mean bias was null in both cases (postcorrection CP close to 1.00, Table 3). Establishment of a preliminary CF by sample exchange (n = 61) in 2 different laboratories relative to 2 independent IS-standardized methods further showed an overall agreement in MMR classification of 91.8% between CP-corrected and CF-converted percent ratios, and 90.2% between CP-corrected and IS reference percent ratios (data not shown).

Discussion

Efficient monitoring of treatment response in chronic myeloid leukemia requires precise and accurate quantification of *BCR-ABL1* expression across a wide linear dynamic interval of IS percent ratios *(1, 4, 30)*. For that reason, interlaboratory IS standardization cannot be achieved with a single reference sample. Both imprecision (intra- and interlaboratory reproducibility) and accuracy (mean bias relative to the IS) should be assessed across a clinically relevant interval of IS percent ratios *(4, 8–11)*. We have demonstrated that 4-level synthetic calibrator panels reproducibly manufactured under controlled processes and directly anchored to the reference values of the WHO primary standards enable reproducible analytical calibration of local RT-qPCR methods within the expected intra- and interlaboratory imprecision of IS-standardized methods. Testing of the ARQ IS calibrator panels generates valuable information at multiple levels: (*a*) assessment of

Table 2. Correlation between nominal IS percent ratios and 8 IS-standardized methods.

	Slope	SES[a]	r^2	SER
e13a2 Panel				
Lab 1 *ABL1*	0.956	0.032	0.989	0.123
Lab 2 *BCR*	1.038	0.038	0.987	0.142
Lab 3 *ABL1*	0.931	0.033	0.987	0.127
Lab 4 *BCR*	1.036	0.033	0.990	0.125
Lab 5 *ABL1*	1.012	0.030	0.991	0.116
Lab 6 *BCR*	1.023	0.040	0.985	0.149
Lab 7 *BCR*	1.007	0.023	0.995	0.086
Lab 8 *ABL1*	0.949	0.040	0.982	0.153
Mean	0.994	0.034	0.988	0.127
Combined	0.992	0.017	0.973	0.187
e14a2 Panel				
Lab 1 *ABL1*	0.902	0.033	0.987	0.134
Lab 2 *BCR*	0.922	0.039	0.982	0.155
Lab 3 *ABL1*	1.004	0.030	0.991	0.122
Lab 4 *BCR*	1.016	0.026	0.994	0.102
Lab 5 *ABL1*	0.876	0.030	0.988	0.121
Lab 6 *BCR*	0.954	0.030	0.990	0.120
Lab 7 *BCR*	0.970	0.022	0.995	0.087
Lab 8 *ABL1*	0.966	0.042	0.981	0.171
Mean	0.951	0.032	0.989	0.127
Combined	0.981	0.022	0.953	0.259
Overall				
Mean	0.972	0.033	0.988	0.127
Combined	0.981	0.015	0.955	0.246

[a] SES, standard error of slope; SER, standard error of residuals.

linearity, analytical sensitivity, and imprecision; (*b*) estimation of the mean bias and corresponding correction parameter, if required; and (*c*) estimation of the correlation between measures across the entire scale. We propose that the ARQ IS calibrator panels may be used as secondary IS reference materials to facilitate interlaboratory comparative studies or external quality-assessment programs.

To ensure the long-term adequacy and integrity of calibration measurements requires reference materials to be stable, homogeneous, traceable, and commutable. Synthetic transcripts have the advantage of being nuclease resistant, reproducible, well characterized, and linked to reference standards from the WHO and the NIST, but they are limited to specific sequence contents. Although all *BCR-ABL1* assays must target the junction between *BCR* exon 13 or 14 and *ABL1* exon 2, unfortunately no global consensus exists on the identity of the control gene. Therefore, the synthetic calibrator panels were designed to be compatible with a majority of laboratories using *ABL1* or *BCR* control transcripts—about 80% of the assays used worldwide *(4)*. To date, the calibrators have been evaluated in about 50 laboratories in 17 different countries on 5 continents, including 10 laboratories that use RT-qPCR methods already standardized to the IS through sample-exchange programs. With the exception of 1 established assay design in which a single reverse primer targeting *ABL1* exon 4 is used to generate long *ABL1* and *BCR-ABL1* e13a2 or e14a2 amplicons *(3, 11, 15)*, all evaluations were successful, with robust detection at all calibrator levels.

The synthetic calibrators are not process controls, and like white blood cell lysate or cell line reference samples, they cannot fully recapitulate the variety of clinicopathologic presentations, biological variation, and sample collection, storage, or shipping conditions encountered in the clinical setting. Direct processing with the heat-lysis protocol, however, decreases the likelihood of RNA loss or degradation during the preanalytical steps and also removes any variation potentially affecting the imprecision or analytical sensitivity of local RT-qPCR methods. It therefore preserves the stability and homogeneity of the calibrators, 2 important characteristics for an analytical reference material. In our study, the direct heat-lysis protocol was compatible with multiplex and singleplex methods, 1- and 2-step protocols, various real-time PCR instruments from multiple manufacturers, manual and automated setups, and commercial and laboratory-developed reagents *(2, 22, 26–29)*, but it could not be evaluated with integrated closed-assay systems such as the GenXpert instrument *(31)*.

Throughout the study, we observed remarkably stable 95% LOA of <2.5-fold across multiple laboratories, assay designs, and platforms. The 95% LOA is a key performance metric because it is a global estimate of method imprecision across the entire interval of percent ratios assessed. Unlike other common metrics such as the CV, 95% LOA are appropriately calculated with normally distributed log-transformed values, provide quantitative information on the same unit as the measurement itself (-fold or logarithmic change), and allow easier interpretation of variation independently from the mean of the distribution *(25)*. In our study, all performance metrics were within the interval expected for the quantitative measurement of percent ratios: from about 10% to <0.01%. For example, the interassay imprecision study produced CVs from 14% to 37% and 95% LOA from 1.21- to 2.09-fold for calibrators 1–4. Without adequate imprecision values, a given RT-qPCR method cannot be accurately aligned to the IS, and its performance cannot be reproducibly

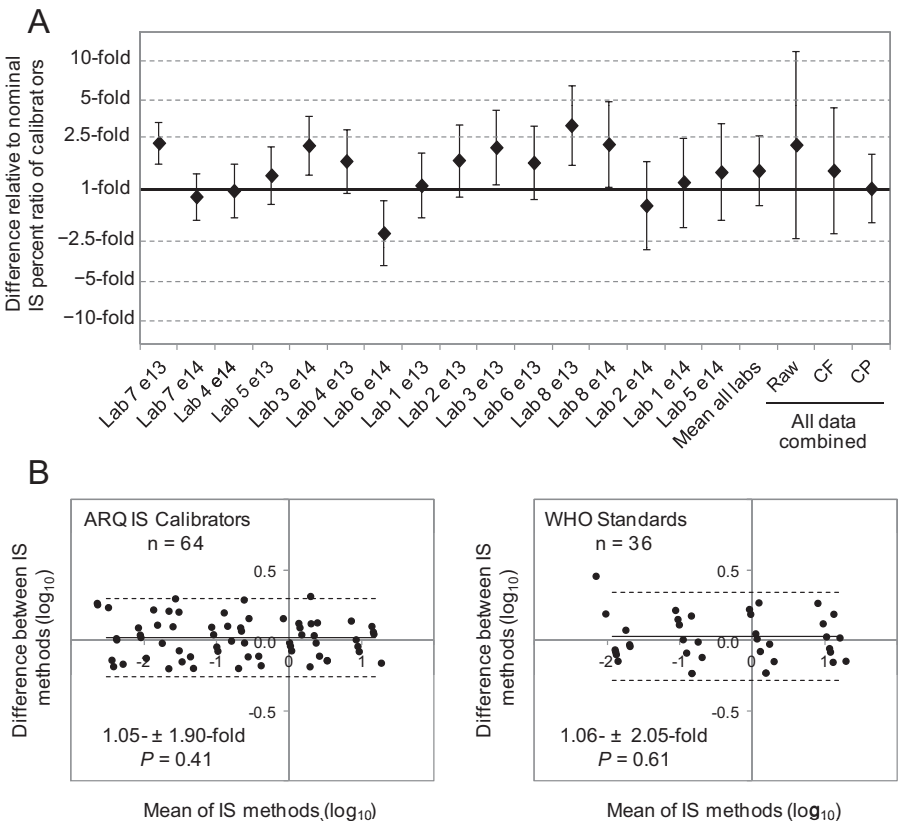

Fig. 3. Evaluation in 8 laboratories using independent IS-standardized RT-qPCR methods.

(A), Relative difference between nominal and measured IS percent ratios for all calibrator levels combined. The relative differences (black diamond) for each laboratory and panel (e13a2 or e14a2) are sorted from the lowest to highest 95% LOA values (error bars). Overall results for all data combined (192 measures) using raw, CF-converted, or CP-corrected percent ratios are also shown on the right. (B), Bias analyses for the ARQ IS calibrator panels evaluated in the present study (left) and for the WHO primary standards (right) previously evaluated in 9 laboratories *(18)*. Each point represents the difference between laboratory-specific IS percent ratios and the reference IS percent ratios (mean IS percent ratios across all laboratories per fusion transcript and control transcript) plotted against the mean of both values. Black lines, mean bias. Dashed lines, 95% LOA. *P*, *P* value associated with the slope of the least squares linear regression.

monitored over time *(5)*. With a demonstrated variability of <2.5-fold over 3 logarithms of values, the synthetic calibrators enabled precise analytical assessments of distinct assay designs. Evaluations in 8 laboratories with the same reagents further highlighted the contribution of other parameters, such as operator, instrument, interpretation of raw data, and intralaboratory variation (95% LOA of 1.28- to 2.87-fold).

Previous studies have shown that very large biases, sometimes >1 logarithm, can be efficiently corrected *(10, 11)*. Similarly, an analytical CP can be calculated to correct any amount of bias, provided the imprecision of the local RT-qPCR method is adequate (95% LOA of <2.5-fold) and there is no trend in the bias (slope of the least squares linear regression analysis). If the slope is between 0.90 and 1.10, the maximum potential error on the measurement of a change of 3 logarithms is 0.3 logarithms, or 2-fold. At MMR (a 3-logarithm change from IS baseline), that would be equivalent to measuring 0.1% on the IS between 0.05% and 0.2% at most, which is within the imprecision of current RT-qPCR methodologies *(10)*. For example, during the international field trial in 8 independent IS-standardized laboratories, the differences between measured IS percent ratios and nominal calibrator IS percent ratios were <2.25-fold for most laboratories, with an overall slope of 0.981. Because imprecision was low in all IS laboratories, 67.2% and 98.4% of the 192 individual measures were within 2- and 4-fold of the nominal IS percent ratio, respectively, well within

Table 3. Summary of evaluation results for 8 laboratories using the same RT-qPCR method.

	Correlation analysis				Bias analysis		
	Slope	SES[a]	r^2	SER	Mean	CP	95% LOA
Results by laboratory							
Lab A	1.020	0.012	0.996	0.067	−0.340	0.457	1.369
Lab B	1.020	0.027	0.991	0.122	−0.289	0.514	1.719
Lab C	1.074	0.035	0.964	0.197	−0.331	0.466	2.537
Lab D	0.996	0.009	0.998	0.055	−0.302	0.499	1.279
Lab E	1.055	0.037	0.967	0.229	−0.433	0.369	2.867
Lab F	0.996	0.014	0.995	0.081	−0.188	0.648	1.435
Lab G	1.093	0.030	0.981	0.178	−0.249	0.563	2.507
Lab H	0.975	0.024	0.994	0.091	−0.326	0.472	1.509
Mean	1.029	0.023	0.986	0.128	−0.307	0.499	1.903
All data combined							
No correction	1.041	0.010	0.983	0.153	−0.303	0.497	2.049
Single CP (0.50)	1.041	0.010	0.983	0.153	−0.002	0.994	2.049
Lab-specific CP	1.038	0.009	0.986	0.138	0.002	1.004	1.913

SES, standard error of slope; SER, standard error of residuals.

the acceptance criteria defined for CF validity *(10, 11)*. The bias in 50% of the laboratories was <1.6-fold, corresponding to an agreement in MMR classification between CP-corrected and CF-converted IS percent ratios of at least 93% for a representative sample set covering 3 logarithms on the IS. Even with a maximum difference of 2.5-fold, only samples generating IS percent ratios between 0.04% and 0.25% would have the potential to be misclassified, within the imprecision of most RT-qPCR methods. The corresponding minimum MMR agreement would be 87% to 92% for samples covering 3 to 5 logarithms, which is within the interval previously reported for IS standardization via sample exchange *(10, 11)*. This expected performance was further confirmed by the results of testing 61 representative clinical samples covering 3.36 logarithms of percent ratios with independent IS standardized methods and the commercial assay. These tests reproducibly generated a CP of 0.50 (agreement in MMR classification of 91.8%, with all discrepant IS percent ratios between 0.061% and 0.236%).

In summary, our comprehensive study demonstrated that synthetic calibrators anchored to the WHO primary standards have the performance characteristics required to facilitate rapid and accurate calibration of *BCR-ABL1* RT-qPCR methods to the IS. Ultimately, we expect the calibrators to reduce the need for *BCR-ABL1* testing laboratories to complete the lengthy and costly process of CF establishment, validation, and periodic revalidation. In addition, the ARQ IS calibrator panels are reproducibly manufactured, broadly available, and commercially distributed worldwide, features that should enable dynamic monitoring of performance over time and further reduce the potential drift of methods already IS standardized. Finally, the rate of patients achieving an at least 4-logarithm decrease on the IS ($MR^{4.0}$) has increased substantially because of the advent of the second-generation tyrosine kinase inhibitors nilotinib and dasatinib *(30, 32)*. These extremely low to undetectable *BCR-ABL1* concentrations must be carefully interpreted in the context of the exact number of *ABL1* cDNA copies present in the qPCR; consequently, the absolute quantity of *ABL1* transcripts should be comparable across testing laboratories *(30)*. Future studies will determine whether the synthetic calibrators whose copy numbers are precisely measured and traceable to a NIST Standard Reference Material can help further the standardization of these deep molecular responses below the MMR.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: J. Hedges, Asuragen, Inc.; V. Modur, HTG Molecular Diagnostics; E. Labourier, Asuragen.
Consultant or Advisory Role: S. Branford, Novartis, Bristol-Myers Squibb, and Ariad; T. Hughes, Novartis, Bristol-Myers Squibb, and Ariad; J. Radich, Novartis, Bristol-Myers Squibb, Pfizer, Ariad, and Incyte; N.C.P. Cross, Novartis.
Stock Ownership: J. Hedges, Asuragen, Inc.; V. Modur, HTG Molecular Diagnostics.
Honoraria: S. Branford, Novartis, Bristol-Myers Squibb, and Ariad; J. Radich, Novartis, Bristol-Myers Squibb, Pfizer, Ariad, and Incyte; N.C.P. Cross, Novartis, Bristol-Myers Squibb, and Asuragen.
Research Funding: Asuragen and Novartis; H.E. White, UK Department of Health; A. Hochhaus, Asuragen.

Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played a direct role in the design of the study, the review and interpretation of data, the preparation of the manuscript, and the final approval of the manuscript.

Acknowledgments: The authors thank the laboratory personnel from each evaluation site for contributing data, and we thank Wailairat Laosinchai-Wolf and Frank Hui from Asuragen for the development and extensive characterization of the ARQ IS calibrator panels. We also thank Vivian Pacheco from Novartis Oncology and Jane Papadaki-Markley from Asuragen for facilitating this international collaborative work.

References

1. Baccarani M, Cortes J, Pane F, Niederwieser D, Saglio G, Apperley J, et al. Chronic myeloid leukemia: an update of concepts and management recommendations of European LeukemiaNet. J Clin Oncol 2009;27:6041–51.
2. Branford S, Hughes TP, Rudzki Z. Monitoring chronic myeloid leukaemia therapy by real-time quantitative PCR in blood is a reliable alternative to bone marrow cytogenetics. Br J Haematol 1999;107:587–99.
3. Emig M, Saussele S, Wittor H, Weisser A, Reiter A, Willer A, et al. Accurate and rapid analysis of residual disease in patients with CML using specific fluorescent hybridization probes for real time quantitative RT-PCR. Leukemia 1999;13:1825–32.
4. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J, et al. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. Blood 2006;108:28–37.
5. Branford S, Cross NC, Hochhaus A, Radich J, Saglio G, Kaeda J, et al. Rationale for the recommendations for harmonizing current methodology for detecting BCR-ABL transcripts in patients with chronic myeloid leukaemia. Leukemia 2006;20:1925–30.
6. Hughes TP, Kaeda J, Branford S, Rudzki Z, Hochhaus A, Hensley ML, et al. Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. N Engl J Med 2003;349:1423–32.
7. Hochhaus A, O'Brien SG, Guilhot F, Druker BJ, Branford S, Foroni L, et al. Six-year follow-up of patients receiving imatinib for the first-line treatment of chronic myeloid leukemia. Leukemia 2009;23:1054–61.
8. Marin D, Ibrahim AR, Lucas C, Gerrard G, Wang L, Szydlo RM, et al. Assessment of BCR-ABL1 transcript levels at 3 months is the only requirement for predicting outcome for patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors. J Clin Oncol 2012;30:232–8.
9. Hanfstein B, Muller MC, Hehlmann R, Erben P, Lauseker M, Fabarius A, et al. Early molecular and cytogenetic response is predictive for long-term progression-free and overall survival in chronic myeloid leukemia (CML). Leukemia 2012;26:2096–102.
10. Branford S, Fletcher L, Cross NC, Muller MC, Hochhaus A, Kim DW, et al. Desirable performance characteristics for BCR-ABL measurement on an international reporting scale to allow consistent interpretation of individual patient response and comparison of response rates between clinical trials. Blood 2008;112:3330–8.
11. Muller MC, Cross NC, Erben P, Schenk T, Hanfstein B, Ernst T, et al. Harmonization of molecular monitoring of CML therapy in Europe. Leukemia 2009;23:1957–63.
12. Clark R, Jack A, Holden L, Barnett D, Reilly JT. BCR-ABL quantitation: To is or not to is, that is the question? [Abstract]. Haematologica 2011;96(Suppl 2):S82.
13. Ramsden SC, Daly S, Geilenkeuser WJ, Duncan G, Hermitte F, Marubini E, et al. EQUAL-quant: an international external quality assessment scheme for real-time PCR. Clin Chem 2006;52:1584–91.
14. Zhang T, Grenier S, Nwachukwu B, Wei C, Lipton JH, Kamel-Reid S. Inter-laboratory comparison of chronic myeloid leukemia minimal residual disease monitoring: summary and recommendations. J Mol Diagn 2007;9:421–30.
15. Muller MC, Saglio G, Lin F, Pfeifer H, Press RD, Tubbs RR, et al. An international study to standardize the detection and quantitation of BCR-ABL transcripts from stabilized peripheral blood preparations by quantitative RT-PCR. Haematologica 2007;92:970–3.
16. Saldanha J, Silvy M, Beaufils N, Arlinghaus R, Barbany G, Branford S, et al. Characterization of a reference material for BCR-ABL (M-BCR) mRNA quantitation by real-time amplification assays: towards new standards for gene expression measurements. Leukemia 2007;21:1481–7.
17. 1st WHO International Genetic Reference Panel for quantitation of BCR-ABL translocation by RQ-PCR. NIBSC code 09/138. Instructions for use. Herfordshire, UK: NIBSC; 2010.
18. White HE, Matejtschuk P, Rigsby P, Gabert J, Lin F, Lynn Wang Y, et al. Establishment of the first World Health Organization International Genetic Reference Panel for quantitation of BCR-ABL mRNA. Blood 2010;116:e111–7.
19. Hietala SK, Crossley BM. Armored RNA as virus surrogate in a real-time reverse transcriptase PCR assay proficiency panel. J Clin Microbiol 2006;44:67–70.
20. Madej RM, Davis J, Holden MJ, Kwang S, Labourier E, Schneider GJ. International standards and reference materials for quantitative molecular infectious disease testing. J Mol Diagn 2010;12:133–43.
21. Walkerpeach CR, Pasloske BL. DNA bacteriophage as controls for clinical viral testing. Clin Chem 2004;50:1970–1.
22. Brown JT, Laosinchai-Wolf W, Hedges JB, Watt CD, Van Deerlin VM, Fletcher L, et al. Establishment of a standardized multiplex assay with the analytical performance required for quantitative measurement of BCR-ABL1 on the international reporting scale. Blood Cancer J 2011;1:e13.
23. Doleshal M, Magotra AA, Choudhury B, Cannon BD, Labourier E, Szafranska AE. Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. J Mol Diagn 2008;10:203–11.
24. Pasloske BL, DuBois DB, Brown DM, Winkler M. Ribonuclease resistant RNA preparation and utilization. United States patent US 7,033,749.
25. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
26. Beillard E, Pallisgaard N, van der Velden VH, Bi W, Dee R, van der Schoot E, et al. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe Against Cancer program. Leukemia 2003;17:2474–86.
27. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. Leukemia 2003;17:2318–57.

28. Branford S, Hughes T. Diagnosis and monitoring of chronic myeloid leukemia by qualitative and quantitative RT-PCR. Methods Mol Med 2006; 125:69–92.

29. Radich JP, Gooley T, Bryant E, Chauncey T, Clift R, Beppu L, et al. The significance of BCR-ABL molecular detection in chronic myeloid leukemia patients "late," 18 months or more after transplantation. Blood 2001;98:1701–7.

30. Cross NC, White HE, Muller MC, Saglio G, Hochhaus A. Standardized definitions of molecular response in chronic myeloid leukemia. Leukemia 2012;26:2172–5.

31. Winn-Deen ES, Helton B, Van Atta R, Wong W, Peralta J, Wang J, et al. Development of an integrated assay for detection of BCR-ABL RNA. Clin Chem 2007;53:1593–600.

32. Kantarjian HM, Baccarani M, Jabbour E, Saglio G, Cortes JE. Second-generation tyrosine kinase inhibitors: the future of frontline CML therapy. Clin Cancer Res 2011;17:1674–83.