npg

ORIGINAL ARTICLE

Establishment of a standardized multiplex assay with the analytical performance required for quantitative measurement of *BCR–ABL1* on the international reporting scale

JT Brown[1,4], W Laosinchai-Wolf[1,4], JB Hedges[1], CD Watt[2], VM Van Deerlin[2], L Fletcher[3], S Branford[3] and E Labourier[1]

[1]Research and Development, Asuragen Inc., Austin, TX, USA; [2]Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA; [3]Department of Genetics and Molecular Pathology, SA Pathology, University of Adelaide, Adelaide, SA, Australia

Accurate and standardized methods for the quantitative measurement of *BCR–ABL1* are a prerequisite for monitoring of treatment response in t(9;22)-positive leukemia. Here, we describe a novel multiplex assay system based on the proven TaqMan and Armored RNA technologies and optimized for sensitive detection of three *BCR–ABL1* fusion transcripts and *ABL1* in a single reaction. Analytical experiments confirmed the absence of significant competition between the simultaneous amplification reactions and established the sensitivity, linearity and precision of the assay. Comparative studies with 115 clinical specimens resulted in high qualitative and quantitative agreement with independent singleplex laboratory-developed tests routinely used in clinical testing. Direct comparison with a reference laboratory calibrated to the international scale (IS) demonstrated minimal analytical bias between methods and an overall accuracy and precision within the performance range required for quantitative measurement of *BCR–ABL1* on the IS. We conclude that detection of e1a2, b2a2, b3a2 and *ABL1* can be achieved in a multiplex assay format compatible with IS reporting. Further clinical validation of the assay could improve the operational efficiency of clinical laboratories, increase their adherence to current recommendations for b2a2/b3a2 reporting on the IS and provide for the first time an opportunity to standardize e1a2-monitoring results.

Introduction

The development and clinical use of the tyrosine kinase inhibitor imatinib mesylate has fundamentally altered the management of chronic myeloid leukemia patients. Over the past 10 years, a reasonable consensus has been reached in the United States and Europe for treatment and monitoring of residual disease during first-line therapy.[1–4] The International Randomized Study of Interferon vs STI571 (IRIS) and follow-up studies have demonstrated that achieving major molecular response (MMR), or a 3-log decrease in *BCR–ABL1* expression from a standard baseline level, was a key clinical outcome.[4] Sensitive molecular methods to quantify the expression levels of *BCR–ABL1* fusion transcripts have therefore emerged as valuable tools for the assessment of treatment response and detection of relapse.[2,3] As many different preanalytical, analytical and reporting methods are used worldwide, it was proposed in 2005 to harmonize quantitative *BCR–ABL1* results on an international scale (IS) anchored to the standard baseline level from the IRIS trial (100% IS), with MMR corresponding to 0.1% IS.[3] Subsequent international collaborative studies demonstrated that

protocol standardization and establishment and validation of conversion factors (CFs) between field methods and IS Reference Laboratories improved harmonization and MMR concordance rates.[5,6] This approach has been particularly successful in Europe through the establishment of National Reference Laboratories harmonized to the IS in 24 different countries.[7] Collectively, these publications also provided clear recommendations on the optimal preanalytical parameters and analytical performance characteristics, such as sensitivity, linearity, accuracy and precision, that are required for quantitative measurement of *BCR–ABL1* and reporting on the IS.[3,5–7]

A prerequisite for standardization is the use of an appropriate endogenous control gene, most often *ABL1*,[3,8] and reporting of the *BCR–ABL1* to control gene ratio. One limitation of this approach is the difficulty to fully control the efficiency of independent PCR reactions for *BCR–ABL1* and the control gene, and to ensure that unrelated results can be combined in a single-ratio value. The inclusion of additional PCR reactions in every run to test each positive/negative control and to build standard curve(s) for the control gene(s) can also decrease the number of clinical samples tested per batch and reduce the operational efficiency of clinical laboratories. Moreover, the IRIS trial and IS harmonization effort exclusively focused on the *BCR–ABL1* fusion transcripts resulting from the major break point (e13a2 and e14a2, also named b2a2 and b3a2 in this manuscript). Yet, other fusion transcripts resulting from t(9;22), such as e1a2, are clinically relevant. For example, Verma *et al.*[9] recently showed that chronic myeloid leukemia expressing only e1a2, although rare, are associated with inferior outcome and should be closely monitored during therapy with tyrosine kinase inhibitor. In acute lymphoblastic leukemia, where e1a2 represents 70% of t(9;22)-positive cases, correlation between *BCR–ABL1* expression levels and long-term outcome is less clear and much work is still needed to define appropriate *BCR–ABL1* response criteria.[10] Standardization of e1a2 monitoring assays for which MMR and IS have not yet been established would likely be the first step toward addressing these clinical needs. One potential solution to overcome the above limitations would be to detect e1a2, b2a2 and b3a2, as well as an endogenous control in a single-well reaction. In this report, we describe such an assay and show that it can meet the stringent performance characteristics that have been established with the well-characterized singleplex assays currently used in routine clinical testing.

Materials and methods

Samples
Peripheral blood or bone marrow specimens from leukemia patients were collected, processed and archived at two independent sites. Specimens processed at the Hospital of the

Correspondence: Dr E Labourier, Research and Development, Asuragen Inc., 2150 Woodward Street, Ste 100, Austin, TX 78744, USA.
E-mail: elabourier@asuragen.com
[4]These authors contributed equally to this work.

University of Pennsylvania (HUP, method 1, $n = 33$) were tested with the BCR/ABL1 Quant assay at HUP on a standard 7500 Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA). Each specimen had previously been tested with a laboratory-developed test (LDT) consisting of two quantitative singleplex assays for the *ABL1* endogenous control or *BCR–ABL1* major fusion transcripts (b2a2/b3a2). Specimens acquired from an independent clinical laboratory and previously characterized with independent singleplex LDTs for *ABL1* and the major (b2a2/b3a2) or minor (e1a2) *BCR–ABL1* fusion transcripts (method 2, $n = 82$) were tested at Asuragen Inc. (Austin, TX, USA) on a 7500 Fast Dx Real-Time PCR System (Applied Biosystems). All human specimens in this study were de-identified and evaluated according to protocols approved by their respective institutions. No results were reported to physicians or patients or used for treatment decision, and no protected health information or other information identifying patients was released.

Total RNA was purified from translocation-positive leukemic cell lines (SUP-B15/e1a2, BV173/b2a2, or K562/b3a2) or from the t(9;22)-negative leukemic cell line HL-60 using an optimized laboratory-validated method based on the *mir*Vana miRNA Isolation Kit (Ambion, Carlsbad, CA, USA). When indicated, positive cell line RNA was diluted mass-to-mass in a background of negative cell line RNA keeping the concentration of total RNA constant. Samples from the First World Health Organization International Genetic Reference Panel for quantitation of BCR–ABL mRNA, comprising four dilution levels of freeze-dried K562 cells diluted in HL-60 cells, were processed using 600 μl of RLT buffer (Qiagen, Valencia, CA, USA) as recommended in White *et al*.[11] RNAs corresponding to specific fusion transcripts were prepared using standard *in vitro* transcription methods. The concentration and purity of every sample in this study was evaluated at 260 and 280 nm using standard spectrophotometric methods.

Multiplex reverse transcription-PCR
Multiplex reactions were performed in 96-well plates using the BCR/ABL1 Quant kit (for research use only, not for use in diagnostic procedures) according to the instructions for use (Asuragen Inc.). Briefly, the kit consists of two reverse transcription (RT) reagents, three PCR reagents and four calibrators. Up to 5 μl of RNA was reverse transcribed into complementary DNA (25 °C, 10 min; 42 °C, 45 min; 93 °C, 10 min) on a GeneAmp PCR System 9700 (Applied Biosystems) using 8 μl of RT buffer and 2 μl RT enzyme mix in a final volume of 20 μl. Twenty-five percent (5 μl) of the complementary DNA was then amplified by multiplex PCR (37 °C for 15 min, 95 °C for 10 min, followed by 45 cycles of 95 °C for 15 s and 60 °C for 1 min) on a 7500 real-time PCR instrument (Applied Biosystems) using 16.5 μl quantitative PCR buffer, 2.5 μl primer/probe mix and 0.5 μl of provided AmpliTaq Gold. The Armored RNA Quant calibrators were diluted 1:10 with the provided diluent and heat denatured for 5 min at 75 °C immediately before the RT reactions.

Four-point standard curves were generated in triplicate for each color channel in every run (12 reaction wells). As the assay detects e1a2, b2a2 and b3a2 in the FAM channel with similar efficiency, a unique *BCR–ABL1* standard curve was used to quantify all three fusion transcripts. Cycle threshold (Ct) values for each target were determined within the log-linear phase of the amplification curves after setting the appropriate baseline in each color channel of the 7500 using the manual baseline method. Standard curves, Ct values and copy numbers for each tested sample were automatically generated by the 7500 software.

Size fractionation
For capillary electrophoresis (CE) analysis, the PCR products stored in the dark below −15 °C were diluted 1:50 in water and 1 μl of diluted material was combined with 0.5 μl of GeneScan 500 ROX Size Standard (Applied Biosystems) and 13.5 μl of Hi-Di Formamide (Applied Biosystems). After heat denaturation at 95 °C for 2 min, samples were transferred on a cold block, quickly centrifuged and analyzed on a 3130xl Genetic Analyzer equipped with 36 cm POP-7 capillaries (Applied Biosystems) using the following conditions: pre-run = 15 kV for 180 s; temperature = 60 °C; injection = 1.6 kV for 20 s; run = 15 kV for 50 min; and other settings = default. Raw data (.fsa file) were analyzed with the GeneMapper Software V4.0 (Applied Biosystems). The calculated sizes for b2a2, e1a2 and b3a2 amplicons are 90, 119 and 160 bp, respectively, and the observed sizes were 88, 116 and 166 bp, respectively.

Data analysis
Ct values and copy numbers generated by the 7500 software were processed with the BCR/ABL1 Quant XL data analysis tool to automatically calculate percent ratios. Samples with $< 5.10^4$ copies of *ABL1* per RT reaction were flagged for review, and quantitative results were not reported if the *BCR–ABL1* copy number was < 50 copies. Samples with $< 10^4$ copies of *ABL1* per RT reaction were excluded and retested. When appropriate, raw Ct or log (% ratio) were further analyzed in Excel (Microsoft Corp., Redmond, WA, USA) to generate graphic representations and linear regression data. Mean bias, CF, percent agreement and MMR concordance rate were calculated according to Branford *et al*.[5] The percent agreement between methods in Figure 3 was calculated by scoring the qualitative results obtained by CE analysis with the BCR/ABL1 Quant amplicons (major or minor) versus the quantitative results obtained with independent LDTs specific either for the major or minor *BCR–ABL1* fusion transcripts.

To assess analytical precision, two samples covering a 3-log percent ratio range were tested on 5 independent days by three different operators (runs 1 to 3 by operator 1, run 4 by operator 2 and run 5 by operator 3). The medium-high positive sample (1 in 10^2 cell line dilution, expected percent ratio at about 5%) was tested in triplicate and the low positive sample (1 in 10^5 dilution, expected percent ratio at about 0.005%) was tested in quadruplicate. For the low positive sample, where the highest variability is expected and some replicates can be below the assay limit of quantitation (LOQ), the largest outlier from the daily median in each run was excluded so that the three replicates per run would be analyzed. Standard deviations and coefficients of variation were calculated according to approved consensus guidelines for the evaluation of quantitative devices.[12]

Results

Assay design and rationale
To enable both qualitative and quantitative measurement of *BCR–ABL1* in a multiplex format, the three fusion transcripts e1a2, b2a2 and b3a2 are detected by a single TaqMan probe specific for *ABL1* exon a2 in the FAM channel of a real-time instrument (Figure 1). For amplification of e1a2 or b2a2/b3a2, the assay contains two sense primers specific for *BCR* exons e1
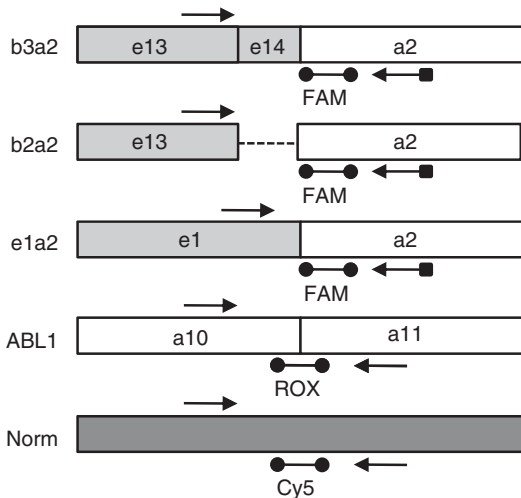
Figure 1 Primers and probes design. The positions of each primer relative to *BCR* (gray box) and *ABL1* (white box) exonic sequences are indicated by arrows. The relative positions of each of the three TaqMan probes, carrying a unique combination of dye and quencher (●) for real-time amplicon detection in the FAM, ROX or Cy5 color channels of the 7500 instrument, are also shown. The reverse primer specific for *ABL1* exon a2 carries a 5′ label (■) to enable subsequent detection of the *BCR–ABL1* amplicons in the FAM channel of a CE instrument. Primer and probe sequences were checked against databases and designed to avoid common polymorphisms such as the T to C substitution in BCR exon 13. Norm (dark gray box) is a synthetic sequence with no significant homology to known genomic sequences.

and e13, respectively. In addition, each *BCR–ABL1* amplicon has a different size and shares an antisense primer specific for *ABL1* exon a2 carrying a 5′ label to enable subsequent size separation by CE. This design does not enable the detection of other rare fusion transcripts, such as e1a3, e13a3 or e14a3. Total *ABL1* (that is, *ABL1* and *BCR–ABL1*) is detected in the ROX channel of the real-time instrument using a pair of primers and a probe specific for the exon 10/exon 11 junction. *ABL1* was chosen as the endogenous control based on its documented expression and stability in leukemia samples.[8,13] Because the assay is multiplexed, four-point standard curves for both *ABL1* and *BCR–ABL1* can be built by testing only four Armored RNA Quant calibrators provided with the kit (Figure 2a). Finally, the assay can also detect a non-human sequence called Norm. This synthetic transcript can be spiked as a nuclease-resistant Armored RNA Quant molecule into test specimens before RNA extraction and be codetected by the assay in the Cy5 channel of the real-time instrument (Figure 1). This optional feature may be used as a qualitative or quantitative process control, for example, to determine the normalized copy number of *BCR–ABL1* per unit of white blood cells; however, this assay feature was not evaluated in this study. We focused exclusively on the key performance metrics required to determine whether the current research tool would be appropriate for the quantitative measurement of *BCR–ABL1* to *ABL1* ratio on the IS.

Evaluation of analytical performance
Analytical performance was first evaluated using synthetic RNA targets prepared by *in vitro* transcription. Specificity was demonstrated by detection of *BCR–ABL1*, *ABL1* and Norm in their respective fluorescence channels (FAM, ROX or Cy5) with no cross-detection between the three color channels of the instrument and no cross-reactivity between the different targets

(data not shown). No Ct value could be measured when testing other synthetic RNAs corresponding to various leukemia fusion transcripts resulting from t(1;19), t(4;11), t(8;21), t(12;21), t(15;17) or inv(16) (data not shown). Serial dilution of e1a2, b2a2, b3a2, *ABL1* and Norm synthetic RNAs from 10^8 to 10^1 copies per RT reaction showed that the assay was linear across 7 logs with similar R^2 and linear regression curves (Figure 2b). Each target was reproducibly detected at 50 copies per RT with >95% positivity (Ct below 40), suggesting a LOQ at least equivalent to 10–15 copies per PCR (25% of the RT reaction input). At lower input, the assay was still linear with >50% positivity at 10 copies per RT reaction, suggesting a limit of detection of at least 2–5 copies per PCR.

Linearity, limit of detection and LOQ were further assessed in a multiplex format by diluting total RNA purified from translocation-positive leukemic cell lines (SUP-B15/e1a2, BV173/b2a2 or K562/b3a2) into a background of total RNA purified from the t(9;22)-negative leukemic cell line HL-60 and keeping the total RNA input constant. The assay output (mean *BCR–ABL1* to *ABL1* ratio) was linear across 6 logs of dilution with similar R^2 and linear regression curves for each fusion transcript (Figure 2c). The individual Ct values, demonstrating linear Ct response and no loss of sensitivity or linearity at low *BCR–ABL1* input, are shown in Supplementary Figure 1. Greater than 95% positivity was obtained at 1 in 10^5 dilution (percent ratio between 0.001 and 0.005%) and >50% positivity was obtained at 1 in 10^6 dilution (percent ratio below 0.0005%). All samples detected by real-time RT-PCR (Ct below 40) were also positive by CE analysis, and no *BCR–ABL1* signal was detected in HL-60 RNA (data not shown). Additional experiments confirmed the assay's LOQ between 500 and 1500 ng RNA input and the absence of non-specific signal with genomic DNA purified from t(9;22)-positive cell lines (data not shown).

Evaluation of analytical precision
Cell line RNA samples were also used to estimate assay precision. First, independent dilutions covering a range of percent ratios from about 10 to 0.005% were tested in duplicate (Figure 2d). All duplicate results were within 3.7-fold of each other, with 85% of the duplicates within 2-fold and the largest differences observed for the lowest percent ratios below 0.02% (Figure 2d). Calculation of the 95% limit of agreement on this set indicated that 95% of the duplicate results were expected to be within plus or minus 2.4-fold of the mean. Analytical precision was further evaluated by testing samples covering a 3-log range on multiple days (see Materials and methods). In five independent runs performed by three different operators, slope, intercept and R^2 for the *BCR–ABL1* and *ABL1* standard curves were highly reproducible, and the expected 3-log difference between the high and low sample levels was observed in every run (Table 1). An estimate of assay precision based on the standard deviation of the daily means and corresponding coefficients of variation was 20.8 and 52.7% for the high and low levels, respectively. The exact within-run variability (repeatability precision or S_r), between-day/run variability (combined day and run precision or S_{dd}) and within-device variability (total precision or S_T) are shown in Table 2.

Comparison with singleplex clinical assays
Assay performance was evaluated with 115 total RNA samples archived at two independent sites, each using their own preanalytical and analytical methods to collect, process and test the samples with singleplex quantitative LDT for *ABL1* and
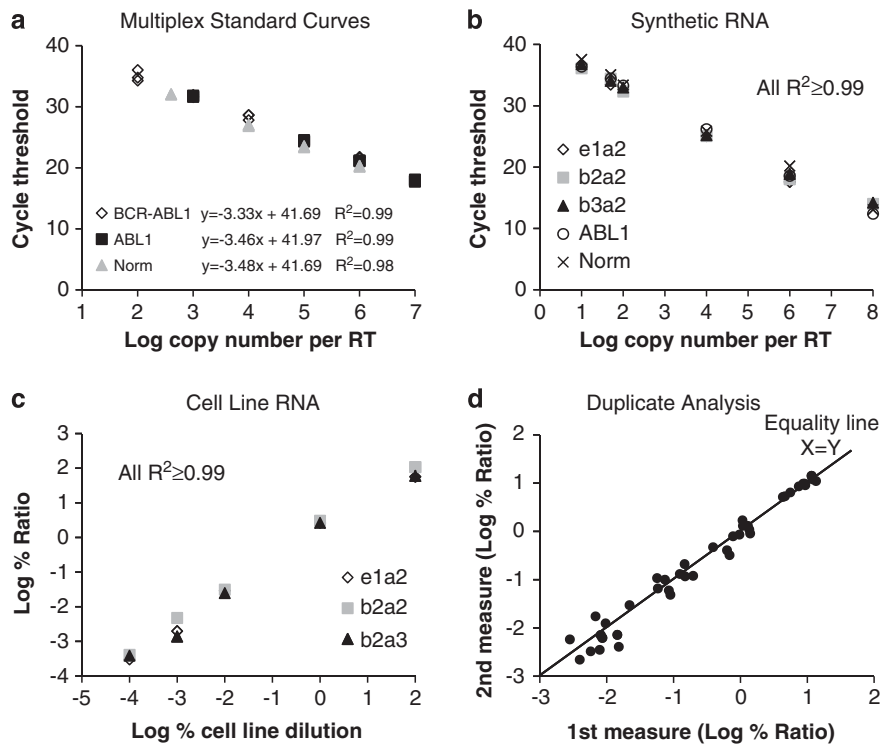
Figure 2 Analytical performance. (**a**) Multiplex standard curves using Armored RNA Quant technology. The four calibrators, each containing various levels of BCR–ABL1 (e1a2), ABL1 and Norm targets and covering 5 logs of linear dynamic range overall, were tested in triplicate. The three resulting four-point standard curves and corresponding R^2 automatically generated by the 7500 instrument's software are shown. (**b**, **c**) Representative examples of analytical sensitivity and linearity with five synthetic RNAs prepared by in vitro transcription or three leukemic cell line RNAs diluted in a background of HL-60 total RNA (1500 ng input). The graphs show the mean values obtained from duplicate testing (10^8, 10^6 or 10^4 copies of synthetic target; cell lines undiluted or diluted 1 in 10^2 or 1 in 10^4), triplicate testing (100 or 50 copies of synthetic target; cell lines diluted 1 in 10^5) or quadruplicate testing (10 copies of synthetic target; cell lines diluted 1 in 10^6). (**d**) Results from duplicate testing using 40 independent cell line RNA dilutions at 1500 ng input. The graph shows the percent ratio for the second measure plotted against the percent ratio of the first measure and the theoretical equality line (first measure = second measure).

Table 1 Results from five independent runs

|	Day 1	Day 2	Day 3	Day 4	Day 5	Mean	% CV
BCR–ABL1							
Slope	−3.41	−3.58	−3.33	−3.35	−3.22	−3.38	3.9
Intercept	41.89	42.89	41.22	41.86	40.87	41.75	1.9
R^2	0.994	0.995	0.991	0.995	0.994	0.994	0.2
ABL1							
Slope	−3.52	−3.41	−3.41	−3.56	−3.5	−3.48	1.9
Intercept	41.1	41.05	41.06	41.84	40.61	41.13	1.1
R^2	0.998	0.998	0.999	0.999	0.999	0.999	0.1
% Ratio							
10^2 Dilution	5.67	5.68	7.23	4.77	4.20	5.51	20.8
10^5 Dilution	0.0077	0.0063	0.0096	0.0018	0.0038	0.0059	52.7
% Ratio difference (log)	2.87	2.95	2.88	3.42	3.04	3.03	7.5

Abbreviation: CV, coefficients of variation.

Table 2 Calculation of assay precision

	Within-run		Between-day/run		Total precision	
	S_r	% CV	S_{dd}	% CV	S_T	% CV
10^2 Dilution	0.84	15.3	1.04	18.9	1.34	24.3
10^5 Dilution	0.0035	74.1	0.0011	24.3	0.0037	78.0

Abbreviation: CV, coefficients of variation.

BCR–ABL1 e1a2 or b2a2/b3a2 (see Materials and methods). The clinical set represented a broad range of specimen type, BCR–ABL1 expression level and fusion transcript identity (Figure 3a). All of the 115 total RNA samples tested in singleton with the BCR/ABL1 Quant assay had acceptable ABL1 copy number (Supplementary Table 1) and generated BCR–ABL1-positive results, with percent ratios ranging over 4 logs (Figure 3b). Quantitative analysis was performed on 103 specimens only, as 12 specimens had no percent ratio
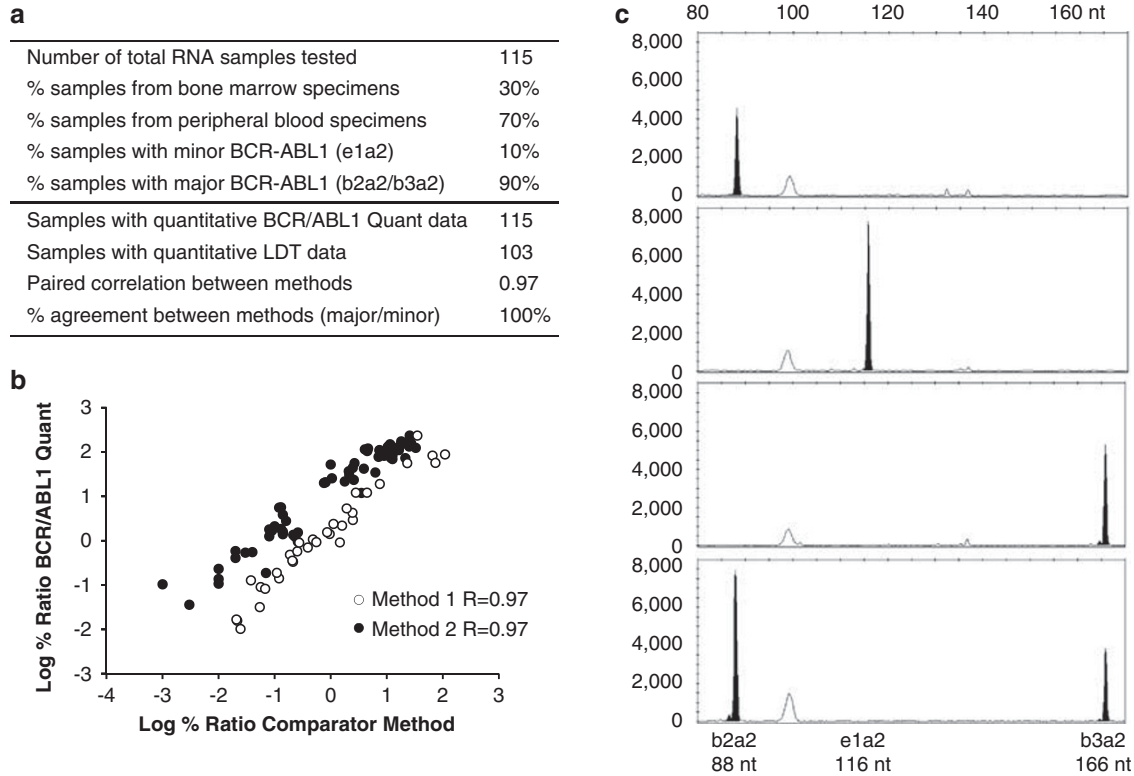
Figure 3 Comparison with existing LDTs. (**a**) Description of sample set and summary of results. (**b**) Quantitative analysis on 103 archived total RNA samples with LDT results ranging from >100 to about 0.001% ratio. Percent ratio and calculated paired Pearson correlation values are shown for each independent method. (**c**) Representative examples of CE traces for *BCR–ABL1* amplicons analyzed in the FAM channel of a 3130xl CE instrument. Nucleotide (nt) size for the expected b2a2, e1a2 and b3a2 fragments are shown.

reported with one of the comparator methods (reported as positive but below the LOQ of the LDT). With the multiplex assay, these 12 specimens were all above the limit of detection (16–123 copies of *BCR–ABL1* per RT, median = 70 copies) and 8 specimens were above the LOQ (57–123 copies per RT). The paired correlation between each LDT and the BCR/ABL1 Quant assay was 0.97 with 95% limits of agreement of plus or minus 3.2-fold for method 1 and 4.0-fold for method 2 (Figure 3b). The PCR amplicons were further resolved by CE analysis to determine the type of fusion transcript(s) detected during the real-time PCR (Figure 3c). There was perfect agreement between the quantitative LDTs and the BCR/ABL1 Quant assay for the identification of minor (e1a2) versus major (b2a2/b3a2) fusion transcripts. All 12 e1a2-positive specimens were correctly identified, and about 15.5% (16/103) of the specimens reported as positive for a major fusion transcript by the LDTs showed coexpression of b2a2 and b3a2 (Figure 3c, bottom panel).

Comparison with the IS
Compatibility of the assay with the IS was evaluated by testing reference materials with known IS percent ratios determined by an international group of IS-standardized laboratories.[11] All four samples tested in triplicate showed robust and reproducible Ct values, with a linear regression curve characteristic of a 10-fold serial dilution (Figure 4a). The calculated mean percent ratios were perfectly correlated (Figure 4b). The slope (1.03) and overall measured fold change (3.09 log) were consistent with the previously reported IS percent ratios, suggesting appropriate assay performance in terms of sensitivity and linearity.

To fully characterize the potential bias between the multiplex assay and the IS, 20 representative mixed cell line RNA samples were tested with BCR/ABL1 Quant and an IS reference method (International IS Reference Laboratory, Adelaide, Australia). Bland and Altman analysis[5] showed that the mean bias between methods was 0.017, corresponding to a preliminary CF of 1.04 (Figure 5a). One hundred percent (20/20) of the corrected percent ratios obtained with the BCR/ABL1 Quant assay were within fivefold of the IS, 95% (19/20) were within threefold and 80% (16/20) were within twofold. The paired correlation between methods was 0.97, with 95% limit of agreement of plus or minus 3.6-fold (Figure 5b). After 14 months, a validation experiment was performed using an independent lot of BCR/ABL1 Quant reagents. If there were no changes between reagent lots and methods, the bias after conversion should be close to 0 and the antilog of the bias (CF) should be close to 1. Results showed a mean bias of −0.019 corresponding to a CF of 0.96, with 95% limit of agreement of plus or minus 4.2-fold after conversion (Figure 5b). A secondary analysis of the mean bias between methods for the pooled 40 samples before conversion also showed an overall CF of 0.99, with 95% limit of agreement of plus or minus 3.9-fold (Figure 5b).

We also examined the assay's ability to accurately classify specimens as MMR positive ($\leqslant$0.1% IS) or MMR negative (>0.1% IS) for the set of 40 samples described above. According to the IS reference method, 57.5% (23/40) of the samples were MMR negative and 42.5% were MMR positive (Figure 5b). The overall agreement between the IS reference method and the BCR/ABL1 Quant assay was 87.5% (35/40; Figure 5b), with a corresponding MMR concordance rate of 74%
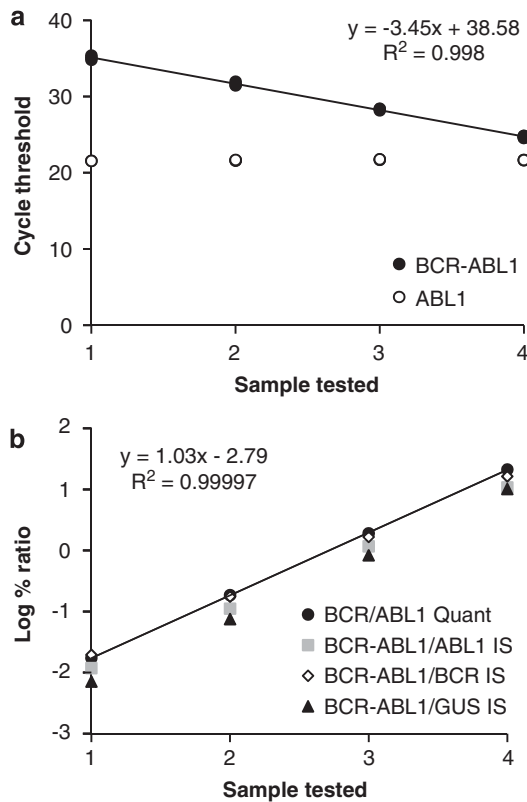
Figure 4 Evaluation of reference materials. (**a**) Ct values obtained for each of the four samples tested in triplicate with the BCR/ABL1 Quant assay (1000 ng input). The linear regression curve and corresponding equation and R^2 for *BCR–ABL1* are shown. (**b**) Correlation between the mean *BCR–ABL1* and *ABL1* percent ratios obtained for each sample with the multiplex assay. The linear regression curve and corresponding equation and R^2 are shown. The mean IS percent ratios reported by White *et al.*[11] using three different endogenous control genes (*ABL1*, *BCR* or *GUS*) are also shown.

(14/19, that is, the number of true MMR positive divided by the sum of true MMR positive, false MMR negative and false MMR positive). A closer examination of the five misclassified samples, two false MMR positives and three false MMR negatives showed little difference between the methods (Figure 5c). All samples misclassified had a BCR/ABL1 Quant IS percent ratio within the 95% limit of agreement between methods, independently of the lot of BCR/ABL1 Quant reagent used.

Discussion

The research assay described here is designed and optimized to enable quantitative measurement of both *ABL1* and *BCR–ABL1* in a single well of a real-time PCR instrument. One feature of multiplex assays that is often considered a daunting technical challenge is the potential loss of sensitivity or linearity when one of the targets is present at low copy number. The efficient detection of *BCR–ABL1* at low copy number in a high background of *ABL1* was demonstrated by a linear Ct response in serial dilution series (Figures 2 and 4, Supplementary Figure 1) and a high correlation with different singleplex assays including an IS reference method (Figures 3 and 5). Further, parallel reactions with the multiplex assay or an assay containing only the *BCR–ABL1*-specific primers and probe resulted in similar Ct values (data not shown). A significant competition between the *BCR–ABL1* and *ABL1* amplification reactions in the

multiplex assay can therefore be ruled out. Serial dilutions of well-characterized cell lines have been shown to generate a range of percent ratios similar to leukemic specimens[14] and were used here to establish that the multiplex assay can reproducibly detect *BCR–ABL1* across a broad linear dynamic range. However, it is important to note that the assay detects total *ABL1* at the junction between exons 10 and 11 (Figure 1). Therefore, high *BCR–ABL1* levels in undiluted RNA can artificially affect the percent ratio. Below 10% dilution, the contribution in total *ABL1* from the t(9;22)-positive cell lines is negligible relative to the *ABL1* contributed by HL-60 and should not impact the measured percent ratios. This effect was evidenced with the BV173 cell line that lacks expression of normal *ABL1* and instead expresses an *ABL1–BCR* fusion transcript commonly found in t(9;22)-positive chronic myeloid leukemia patients.[15,16] As expected, undiluted BV173 RNA reproducibly generated percent ratios higher than with the two other cell lines (around 100%, that is, total *ABL1* = *BCR–ABL1*) and no significant differences between the cell lines were observed after 10-fold serial dilutions (Figure 2c and Supplementary Figure 1).

Multiplexing quantitative assays can also provide specific advantages. For example, multiplex assays have the potential to reduce the burden of validation and operator training for independent assays, to increase the number of samples tested per run, to improve the laboratory throughput and, overall, to streamline the laboratory workflow. As the identity of *BCR–ABL1* fusion transcript is in general determined at initial diagnosis[3] by independent qualitative PCR methods, the BCR/ABL1 Quant assay does not discriminate between e1a2, b2a2 and b3a2 fusion transcripts and a single quantitative *BCR–ABL1* result is generated. However, the assay can distinguish the three *BCR–ABL1* fusion transcripts with the optional CE reflex-testing step. The multiplex assay may also provide additional benefits relative to previously described quantitative LDTs compatible with CE.[17,18] For example, all reagents are manufactured under current good manufacturing practises and are optimized to reduce the number of pipetting steps and to streamline the assay workflow (see Materials and methods). The assay also includes AmpliTaq Gold polymerase and Armored RNA Quant calibrators for the establishment of four-point multiplex standard curves for each target detected by the assay (Figure 2a). Armored RNAs are stable, nuclease resistant and precisely quantified synthetic RNAs already widely used as controls and standards in clinical molecular infectious disease testing.[19] Unlike plasmid standards, these RNA molecules enable assessment of both the RT and PCR steps and have shown promising results (JH and EL, unpublished data) for the development of secondary reference materials anchored to the primary reference standard for quantitation of *BCR–ABL1*.[11]

Cell line dilutions were also used to assess the assay analytical precision. Duplicate testing across the assay linear dynamic range resulted in 95% limits of agreement of plus or minus 2.4-fold, indicating that singleton test results are not very different from the mean of duplicate testing (Figure 2d). As the clinically relevant changes in *BCR–ABL1* expression levels are at minimum 2- to 10-fold and must be confirmed by two subsequent measurements within a period of a few months before any change in therapeutic strategy,[1,2,20] these results suggest that singleton testing should not affect the performance of the assay for quantitative *BCR–ABL1* reporting. This was further confirmed by evaluating precision on multiple days. The standard curves were highly reproducible and most of the daily means were within twofold of each other (Table 1). The within-device variability, also called total precision (S_T), combines the
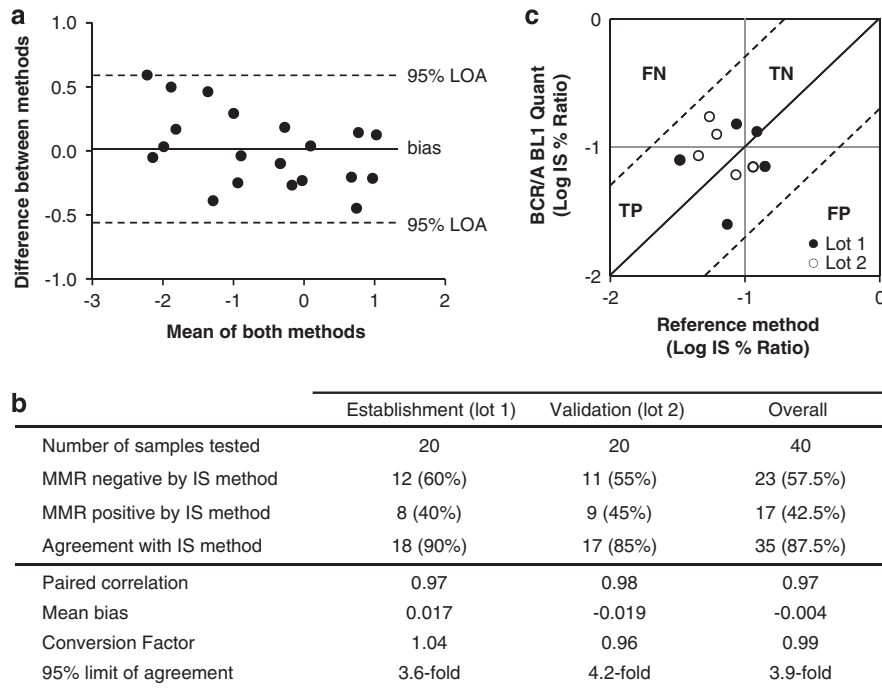
Figure 5 Comparison with the international scale. (**a**) Bland and Altman analysis for the 20 samples tested with BCR/ABL1 Quant lot 1 at Asuragen Inc. and by the reference method at the IS Reference Laboratory. The graph shows the difference between methods (log IS percent ratio−log BCR/ABL1 Quant percent ratio) plotted against the average of both log percent ratio values. The mean difference between methods (bias, solid line) and the 95% limit of agreement between methods (95% LOA, dash lines) are shown. (**b**) Summary of results and assay performance during establishment and validation of the CF and for the combined data (overall). (**c**) Correlation plot between the IS percent ratios obtained with the BCR/ABL1 Quant assay and the reference method for samples within threefold to the MMR point (0.1% IS) according to the reference method. The equality line (solid line) and the fivefold limit of agreement between methods (dash lines) are shown. FN, false MMR negative; FP, false MMR positive; TN, true MMR negative; TP, true MMR positive.

within-run (S_r) and between-day/run (S_{dd}) components of precision,[12] resulting in percent coefficients of variation higher than with other commonly used methods such as the standard deviation of all observed data or the standard deviation of the daily means (used to calculate the percent coefficients of variation in Table 1). As expected, the contribution of the within-run variability to the total precision at very low *BCR–ABL1* input was higher than the contribution of the between-day/run variability (Table 2). This high variability near the LOQ is consistent with the current limitations of quantitative real-time technologies and is, in general, controlled in a clinical setting by defining reportable ranges and by not reporting quantitative values below the validated assay LOQ (for example, 'positive but below LOQ').

By testing in singleton 115 archival total RNA samples from monitored chronic myeloid leukemia patients, we further showed that the BCR/ABL1 Quant assay is compatible with representative clinical specimens, can detect the three different *BCR–ABL1* fusion transcripts in those specimens and has a high qualitative and quantitative agreement with independent singleplex LDTs routinely used in a clinical setting (Figure 3). As the *BCR–ABL1* and *ABL1* transcripts have been shown to have comparable mean stability,[13] a low *ABL1* copy number can indicate poor RNA quality and/or poor efficiency during the extraction, RT or PCR steps. All of the residual RNA samples evaluated had an *ABL1* copy number compatible with quantitative reporting, that is, $>10^4$ copies of *ABL1* per reaction and >50 copies of *BCR–ABL1* for the few samples with $<10^4$ copies of *ABL1* per reaction. Although the quantitative correlation with each LDT was high (0.97), there was a significant difference in relative precision. The calculated mean bias

between methods or the mean difference between the log percent ratio obtained with BCR/ABL1 Quant and each LDT was 0.21 and 1.07 for method 1 and method 2, respectively. In agreement with previously reported data,[5] this observation suggests that individual field methods can generate widely different percent ratios and further emphasizes the need for reporting quantitative *BCR–ABL1* measurements on a unique standardized scale such as the IS.

As neither LDT was IS harmonized, we sought to directly compare the assay with the IS using two distinct approaches. Analysis of four-level reference materials confirmed that multiplexing did not affect the sensitivity and linearity of the assay (Figure 4). The BCR/ABL1 Quant percent ratios were all within twofold of the reference IS percent ratios. However, it cannot be concluded from this single experiment that there is a systematic bias between the multiplex assay and a specific IS reference method. The *BCR–ABL1* to *ABL1* IS percent ratios assigned to these reference materials were obtained by averaging results from six independent laboratories, and as much as fourfold differences were observed between the different methods.[11] By directly comparing the assay to a single IS reference method, we found that there was a minimal bias between methods with two lots of reagents (Figure 5). It should be emphasized that these results do not imply that a CF of 1 can be systematically used to convert percent ratio obtained with the BCR/ABL1 Quant assay to IS percent ratio. Local laboratories reporting on the IS must establish their own CF to factor in potential differences in preanalytical steps and revalidate their CF every 2 years or each time the procedure is changed.[7] In addition, the comparison with the Reference Laboratory was performed using dilutions of cell line RNA, whereas the recommended process is to perform

the comparison using patient RNA.[5] Therefore, the performance reported here may not exactly mimic those of patient samples. However, these results demonstrate that there is not a systematic bias relative to the IS and validate that an analytical CF can be maintained from lot-to-lot through rigorous manufacturing procedures and quality controls.

The primary end point and key marker of molecular response for imatinib and various second-generation tyrosine kinase inhibitors is based on a single-point classification, MMR, originally defined as a 3-log reduction in b2a2/b3a2 fusion transcripts from a standardized baseline value and equivalent to 0.1% IS.[2–4] Previous work showed that quantitative assays must maintain after IS conversion a mean bias relative to the reference method between 0.8 and 1.2 (accuracy) and a reproducibility assessed by the 95% limit of agreement within plus or minus fivefold (precision) to reach optimal MMR classification accuracy.[5] MMR concordance rates of 91% are considered optimal given the technical limitations of current technologies (95% limit of agreement within plus or minus 2.5-fold for each method).[5] In our experiments, we observed a mean bias after conversion of 0.96, with 95% limits of agreement less than fivefold and an overall concordance rate of 74%. It is important to note that this last performance metric is dependent not only on the analytical bias of each method but also on the distribution of the tested samples around the MMR point. In our sample set, the distribution was relatively well balanced, with 57.5% (23/40) of the samples above 0.1% IS (MMR negative) and 42.5% (17/40) at or below 0.1% IS (MMR positive) according to the reference method (Figure 5b). However, 20% (8/40) of the samples had percent ratios very close to the MMR cutoff point (within twofold or 0.05–0.2% IS ratio) and 37.5% (15/40) of the samples were within fivefold of MMR (0.02–0.5% IS ratio). All samples misclassified had a BCR/ABL1 Quant IS percent ratio within the 95% limit of agreement between methods, 80% of the samples had a difference of less than 2-fold and one sample had an IS percent ratio 3.2-fold higher than the reference method (Figure 5c). This variation was indistinguishable from the inherent within-assay variability of both methods, inferring that the observed concordance rate was likely optimal for this sample set with the current quantitative RT-PCR technologies.

In summary, our study established that the BCR/ABL1 Quant research assay has the performance required for the sensitive and multiplex detection of e1a2, b2a2, b3a2 and *ABL1* and for reporting quantitative measurement of *BCR–ABL1* expression levels on the IS. The assay, designed and manufactured under current good manufacturing practises, is currently available as a CE-marked test in Europe and as a Research Use Only assay in the rest of the world. A multisite clinical validation study and regulatory approval of a device based on a similar technology would likely facilitate harmonization of *BCR–ABL1* quantitative measurement, improve clinical laboratories' efficiency and work-flow, and increase adherence to the current recommendations for reporting on the IS.

Conflict of interest

Drs Brown, Laosinchai-Wolf, Hedges and Labourier are employees of Asuragen Inc. The other authors declare no conflict of interest.

Disclaimer

Certain commercial equipment and reagents are identified to specify the experimental procedure. This does not constitute an endorsement by the Adelaide IS Reference Laboratory nor does it imply that the reagents or equipment are the best available for the purpose.

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**: 6041–6051.

2 Baccarani M, Saglio G, Goldman J, Hochhaus A, Simonsson B, Appelbaum F *et al.* Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood* 2006; **108**: 1809–1820.

3 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.

4 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**: 1423–1432.

5 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**: 3330–3338.

6 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**: 970–973.

7 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**: 1957–1963.

8 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**: 2474–2486.

9 Verma D, Kantarjian HM, Jones D, Luthra R, Borthakur G, Verstovsek S *et al.* Chronic myeloid leukemia (CML) with P190 BCR-ABL: analysis of characteristics, outcomes, and prognostic significance. *Blood* 2009; **114**: 2232–2235.

10 Fielding AK. Current treatment of Philadelphia chromosome-positive acute lymphoblastic leukemia. *Haematologica* 2010; **95**: 8–12.

11 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–e117.

12 CLSI. Evaluation of precision performance of quantitative measurement methods: approved guidelines, 2nd edn. *NCCLS document EP5-A2*.

13 van der Velden VH, Boeckx N, Gonzalez M, Malec M, Barbany G, Lion T *et al.* Differential stability of control gene and fusion gene transcripts over time may hamper accurate quantification of minimal residual disease–a study within the Europe Against Cancer Program. *Leukemia* 2004; **18**: 884–886.

14 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**: 2318–2357.

15 Melo JV, Gordon DE, Cross NC, Goldman JM. The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. *Blood* 1993; **81**: 158–165.

16 Westbrook CA, Rubin CM, Carrino JJ, Le Beau MM, Bernards A, Rowley JD. Long-range mapping of the Philadelphia chromosome by pulsed-field gel electrophoresis. *Blood* 1988; **71**: 697–702.

17 Cortes J, Talpaz M, O'Brien S, Jones D, Luthra R, Shan J *et al.* Molecular responses in patients with chronic myelogenous leukemia in chronic phase treated with imatinib mesylate. *Clin Cancer Res* 2005; **11**: 3425–3432.

18 Luthra R, Sanchez-Vega B, Medeiros LJ. TaqMan RT-PCR assay coupled with capillary electrophoresis for quantification and identification of bcr-abl transcript type. *Mod Pathol* 2004; **17**: 96–103.

19 Pasloske BL, Walkerpeach CR, Obermoeller RD, Winkler M, DuBois DB. Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *J Clin Microbiol* 1998; **36**: 3590–3594.

20 Press RD, Willis SG, Laudadio J, Mauro MJ, Deininger MW. Determining the rise in BCR-ABL RNA that optimally predicts a kinase domain mutation in patients with chronic myeloid leukemia on imatinib. *Blood* 2009; **114**: 2598–2605.

Supplementary Information accompanies the paper on Blood Cancer Journal website (http://www.nature.com/bcj)