

Keywords: non-small-cell lung carcinoma; *EGFR* gene mutations; quality assessment

Assessing standardization of molecular testing for non-small-cell lung cancer: results of a worldwide external quality assessment (EQA) scheme for *EGFR* mutation testing

S Patton^{*1}, N Normanno², F Blackhall³, S Murray⁴, K M Kerr⁵, M Dietel⁶, M Filipits⁷, S Benlloch⁸, S Popat⁹, R Stahel¹⁰ and E Thunnissen¹¹

¹EMQN, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester M13 9WL, UK; ²Cell Biology and Biotherapy Unit, Istituto Nazionale per lo Studio e la Cura dei Tumori 'Fondazione Giovanni Pascale'—IRCCS, 80131 Naples, Italy; ³Christie Hospital, Manchester M20 4BX, UK; ⁴Biomarker Solutions Ltd, London EC1V 2NX, UK; ⁵Department of Pathology, Aberdeen Royal Infirmary, Aberdeen AB25 2ZN, UK; ⁶Charité, Humboldt-Universität zu Berlin, Berlin 10117, Germany; ⁷Medical University of Vienna, 1010 Vienna, Austria; ⁸Pangaea Biotech, USP Dexeus University Institute, Barcelona 08028, Spain; ⁹Royal Marsden Hospital, London SW3 6JJ, UK; ¹⁰University Hospital Zürich, CH-8091 Zürich, Switzerland and ¹¹Department of Pathology, VU University Medical Center, Amsterdam 1081 HZ, The Netherlands

Background: The external quality assurance (EQA) process aims at establishing laboratory performance levels. Leading European groups in the fields of EQA, Pathology, and Medical and Thoracic Oncology collaborated in a pilot EQA scheme for somatic epidermal growth factor receptor (*EGFR*) gene mutational analysis in non-small-cell lung cancer (NSCLC).

Methods: EQA samples generated from cell lines mimicking clinical samples were provided to participating laboratories, each with a mock clinical case. Participating laboratories performed the analysis using their usual method(s). Anonymous results were assessed and made available to all participants. Two subsequent EQA rounds followed the pilot scheme.

Results: One hundred and seventeen labs from 30 countries registered and 91 returned results. Sanger sequencing and a commercial kit were the main methodologies used. The standard of genotyping was suboptimal, with a significant number of genotyping errors made. Only 72 out of 91 (72%) participants passed the EQA. False-negative and -positive results were the main sources of error. The quality of reports submitted was acceptable; most were clear, concise and easy to read. However, some participants reported the genotyping result in the absence of any interpretation and many obscured the interpretation required for clinical care.

Conclusions: Even in clinical laboratories, the technical performance of genotyping in *EGFR* mutation testing for NSCLC can be improved, evident from a high level of diagnostic errors. Robust EQA can contribute to global optimisation of *EGFR* testing for NSCLC patients.

Assessment of epidermal growth factor receptor (*EGFR*) mutations has become mandatory to choose the most active first-line treatment for patients with advanced non-small-cell lung cancer

(NSCLC). Indeed, randomized phase III clinical trials have demonstrated that first-line administration of an *EGFR* TKI results in a prolonged progression-free survival as compared with

*Correspondence: Dr S Patton; E-mail: simon.patton@cmft.nhs.uk

Received 22 January 2014; revised 19 May 2014; accepted 20 May 2014; published online 1 July 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14



chemotherapy in patients carrying *EGFR* mutations (Mok *et al*, 2009; Maemondo *et al*, 2010; Mitsudomi *et al*, 2010; Fukuoka *et al*, 2011; Zhou *et al*, 2011; Rosell *et al*, 2012). These studies have also confirmed that *EGFR* mutations are a reliable marker that predicts sensitivity to *EGFR* TKIs (Mok *et al*, 2009).

Activating mutations occur in exons 18 through 21 of the TK domain of the *EGFR* gene, and either point mutations or in-frame small deletions or insertions (Sharma *et al*, 2007; De Luca and Normanno, 2010). Although more than 250 mutations of the *EGFR* gene have been described to date, two mutations, a single point mutation in exon 21, the L858R, and a series of small in-frame deletions in exon 19, account for ~90% of all *EGFR* mutations (Sharma *et al*, 2007; Linardou *et al*, 2008). *EGFR* mutations are strongly associated with defined clinical and pathological features: they are far more frequent in female patients as compared with male; in adenocarcinoma as compared with other histological types; in non-smokers as compared with current smokers or former smokers; and in East-Asian NSCLC patients as compared with Non-East-Asian patients (Normanno *et al*, 2006).

External quality assessment (EQA) is a system of objectively checking laboratory results by an independent external agency (van Krieken *et al*, 2013). The main objective of an EQA programme is to establish inter-laboratory comparability. In this respect, the EQA process can identify latent systematic errors in methodology that may not be revealed by a laboratory's own internal QA processes. Representatives from ETOP, ESMO, ESP, EMQN and other leading European groups met in July 2010 to discuss a pan-European approach to EQA for *EGFR* mutation testing in NSCLC. In this paper, we present the results of this pilot EQA scheme for *EGFR* testing that was completed in 2013.

MATERIALS AND METHODS

Organisation of the scheme. A meeting was organised in July 2010 by ETOP and EMQN to bring together a group of professionals representing EMQN, ESP, ETOP, ESMO and other leading European groups involved in NSCLC testing (see Supplementary Information). From this group, a steering group of five individuals was formed who planned, designed and assessed the results of the pilot EQA scheme. The scheme was coordinated and administered by the EMQN and three rounds were organised within a period of 18 months. The workflow of the scheme process is shown in Figure 1.

Validation of samples. The primary aim of this scheme was to develop a flexible, scalable EQA scheme designed to assess issues related to techniques and minimum detection limits used in standard laboratory practice, focusing exclusively on the analytical (that is, sample processing, genotyping) and reporting phases (interpretation of the results in relation to the clinical context). To enable this and to avoid the significant challenges of sample heterogeneity in real tissue samples, 20 artificial materials were used composed of formalin-fixed paraffin-embedded (FFPE) cell line samples. These EQA materials were designed to mimic real tissue samples as closely as possible and contained homogenous mixtures of mutant vs wild-type cell lines at a range of different allelic ratios. The paraffin blocks were cut and 10 μ m sections placed in eppendorf tube at the Pathology department of the VU University Medical Centre in Amsterdam, The Netherlands, by Dr Erik Thunnissen. H&E (4 μ m) sections were used to estimate the number of tumour cells. In each EQA sample section, at least 200 nuclei were present (usually >300), roughly mimicking the amount of cells from a small NSCLC biopsy.

For each EQA sample, one 10- μ m-thick section was sent by EMQN to each of the three validating laboratories for mutational analysis in a blinded fashion. Different sections from the block

were analysed for *EGFR* mutation status to ensure that the mutation was homogeneously represented within each block. The validating laboratories independently analysed the samples by using three different approaches: direct sequencing of the PCR product for exons 18–21 mutations; fragment analysis for exon 19 deletions and an allelic discrimination-based real-time PCR assay for the L858R mutation in exon 21; and the Therascreen *EGFR* RGQ kit (Qiagen, Hilden, Germany), reporting the results directly to the EMQN.

The allelic ratios of mutations in each sample used in rounds 2 and 3 were accurately quantified by a commercial sponsor (Horizon Diagnostics, Cambridge, UK) using droplet digital PCR (ddPCR) on a BioRad QX100 (Hercules, CA, USA) platform. Genomic DNA (gDNA) was extracted from FFPE sections on the Promega (Madison, WI, USA) Maxwell System using the Maxwell 16 FFPE Plus LEV DNA purification kit, according to the manufacturer's protocol. Quantification was performed using a Promega QuantiFluor dsDNA assay kit, according to the manufacturer's protocol. ddPCR was performed using Taqman custom SNP 40 \times primer/probe assays (Life Technologies, Carlsbad, CA, USA) to assess the frequency of each mutation with the exception of the p.(E746_A750) assay, which was designed in-house. DNA (40 ng) was added to each ddPCR reaction. Reactions were performed in quadruplicate and droplets were generated using a Droplet Generator according to the manufacturer's instructions. PCR was performed on a standard thermocycler using previously optimised, assay-specific cycling conditions. Droplets were analysed using a QX100 Droplet Reader as described in the manufacturer's instructions. Data from at least 45 000 useable droplets were collected for each sample. Formalin-fixed paraffin-embedded reference standards (Horizon Diagnostics) were included as assay controls.

Registration of participant laboratories and shipment of samples. Laboratories that performed *EGFR* mutational analysis were invited to participate in the EQA via an open call from the EMQN in conjunction with the ESP, ETOP and ESMO. Participating laboratories registered via the EMQN website (European Molecular Genetics Quality Network (EMQN), 2014), and were requested to perform DNA extraction and analysis using their routine method. In each round, 10 samples (one 10- μ m-thick section for each) with accompanying mock clinical referral information were sent to participating laboratory. Each laboratory was identified only by a unique EMQN ID code to avoid exchange of information between participants and minimise bias in the results' interpretation process. The laboratories were given 8 weeks to complete their analyses and to submit the results of genotyping to the EMQN website. The centres were requested to provide information on the technique used for mutational analysis and metrics relating to their experience of performing *EGFR* mutational analyses.

Evaluation of results. The scheme included three rounds: the first was restricted to a maximum of 30 labs to establish proof of principle and validate the materials. A subsequent second round of the scheme was organised with no restriction on participation. Laboratories that failed the second round were provided with another set of samples in a restricted third round. The steering group evaluated the results according to a pre-defined scoring system. The scoring system assigned two points to correct genotype and zero points to false-positive or -negative results (Table 1). Errors in mutation nomenclature that might lead to misinterpretation of the results (for example, stating 'deletion' without specifying the exon in which the deletion occurs) were assigned 1.50 points. This deduction was applied only once for each center, generally to the first sample for which the error was found. One point was awarded for cases in which the genotype was mispositioned or miscalled: this error sometimes occurs with exon 19 deletions, for which it might be difficult to define the precise base or amino acid

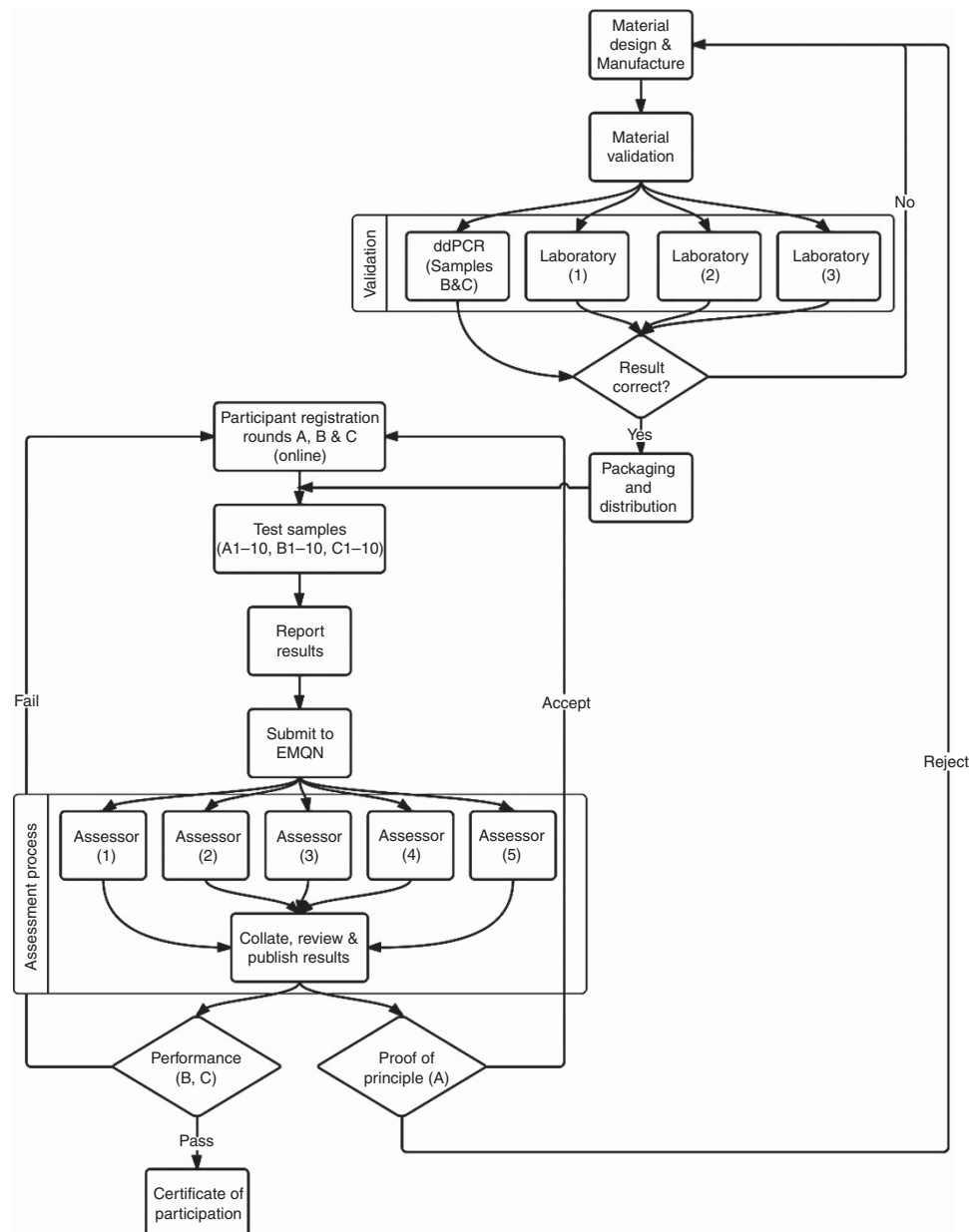


Figure 1. Workflow of the EQA scheme process.

in which the deletion starts or ends. If a test failed giving no result on the sample (analytical failure), then the lab received 1.00 point for that sample. The threshold to pass the EQA was set at a total score for the 10 samples of ≥ 18 out of 20 (Thunnissen *et al.*, 2011) – laboratories with a genotyping score < 18 were classified as poor performers (applied to rounds 2 and 3 only). Performance in the assessment of clinical interpretation and reporting did not contribute to poor performance.

RESULTS

Selection of the samples for the EQA. The first step of the EQA scheme was the selection and the validation of the samples. Twenty materials were manufactured by Dr Thunnissen by mixing four lung cancer cell lines (A549, *EGFR* wild type), H1650 (*EGFR*, p.(E746_A750del), H1975 (*EGFR*, p.(T790M), p.(L858R)) and SW48 (p.G719S). Cell lines with mutations were serially diluted into A549 or SIHA cells at different ratios relevant to establishing

the analytical sensitivity of the tests used by labs. Each material was validated in three different reference laboratories using different techniques to confirm the genotype and the results showed that the mutations were detectable at all the designated ratios, dependent on the technology used (Table 2). A good yield of gDNA was obtained from all the samples. In addition, there was complete concordance on the *EGFR* mutational status of the selected specimens and therefore all were selected for use in the quality assessment scheme with samples A1–A10 used for the pilot, and B1–10 and C1–C10 in subsequent rounds 2 and 3.

To accurately establish quantitative measurements of the allelic frequencies of the *EGFR* mutations, all 10 EQA samples (Table 2; samples B/C1–B/C10) used in rounds 2 and 3 were analysed on a ddPCR platform (BioRad QX100). Three of the samples had allelic frequencies higher than expected (C3, C8 and C9), two were lower (C2 and C10) and in one (C5) it was not possible to establish the true value due to insufficient availability of sample material (Table 3).

First round proof of principle pilot scheme. Twenty-nine laboratories registered from 13 countries, and 25 participated in

Table 1. Scoring system

Criteria	Marks	Scheme round ^a (samples)
Genotyping		
Correct genotype	2.00	A to C (all samples)
Incorrect genotype (false positive or false negative)	-2.00	A to C (all samples)
Genotype mispositioned or miscalled (e.g., incorrect base/amino acid detected)	-1.00	A to C (all samples)
Error in the nomenclature that might lead to misinterpretation of the results	-0.50 ^b	A to C (all samples)
Biological/clinical interpretation		
The mutation is predicted to be a sensitising mutation to <i>EGFR</i> TK inhibitors (TKI).	Comment only	B (1, 4)
The mutation is predicted to confer resistance to <i>EGFR</i> TK inhibitors (TKI).	Comment only	B (9)
General		
Reference sequence and version indices not used, incorrect or inconsistent.	-0.50 ^b	B (1, 4, 9) only
No description of assay used and its limitations	-0.50 ^b	B (1, 4, 9) only
Analytical failure giving no test result	-1.00	B (1, 4, 9) only
No statement present on estimated percentage of tumour cells within the sample	Comment only	B (1, 4, 9) only
Abbreviation: TK = tyrosine kinase.		
^a Pilot round (A), second round (B) and third round (C).		
^b Deduction applied only once.		

Table 2. Validation of EQA materials

Sample no.	PCR/sequencing	Therascreen	Fragment analysis/real-time PCR	Round no./sample no. ^a
1	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); p.(L858R)	p.(T790M); p.(L858R)	A1
2	c.2235_2249del, p.(E746_A750del)	p.(E746_A750del)	p.(E746_A750del)	A2
3	Wild type	Wild type	Wild type	A3
4	Wild type	Wild type	Wild type	A4
5	c.2235_2249del, p.(E746_A750del)	p.(E746_A750del)	p.(E746_A750del)	A5
6	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); p.(L858R)	p.(T790M); p.(L858R)	A6
7	c.2235_2249del, p.(E746_A750del)	p.(E746_A750del)	p.(E746_A750del)	A7
8	Wild type	Wild type	Wild type	A8
9	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); p.(L858R)	p.(T790M); p.(L858R)	A9
10	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); p.(L858R)	p.(T790M); p.(L858R)	A10
11	c.2235_2249del, p.(E746_A750del)	p.(E746_A750del)	p.(E746_A750del)	B1 and C8
12	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); c.2573T>G, p.(L858R)	B2 and C3
13	c.2235_2249del, p.(E746_A750del)	p.(E746_A750del)	p.(E746_A750del)	B3 and C9
14	Wild type	Wild type	Wild type	B4 and C1
15	Wild type	Wild type	Wild type	B5 and C4
16	Wild type	Wild type	Wild type	B6 and C6
17	Wild type	Wild type	Wild type	B7 and C7
18	c.2155G>A, p.(G719S)	p.(G719S)	p.(G719S)	B8 and C10
19	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	p.(T790M); p.(L858R)	p.(T790M); p.(L858R)	B9 and C5
20	c.2155G>A, p.(G719S)	p.(G719S)	p.(G719S)	B10 and C2
Abbreviation: EQA = external quality assessment.				
^a EQA scheme rounds – first (A), second (B) and third (C).				

the pilot EQA scheme (4 labs withdrew due to customs sample importation problems), which was run in fourth quarter of 2011. A set of 10 samples were sent to the laboratories (Table 2; samples A1–A10). All the participating laboratories submitted results within the 8-week time frame. The main methodology used by the participants was PCR/sequencing ($n = 10$ laboratories; 34%) and real-time PCR ($n = 10$; 34%) (Figure 2).

Two analytical errors (false-negative results) were observed. A further five laboratories made process errors (sample swaps) that resulted in an additional 24 genotype errors. In all cases, the genotypes were correct, but reported for the wrong sample. Therefore, 92% of the false-negative results were concentrated in

five laboratories. No false-positive results were reported. The materials performed well and there were no analytical test failures.

As this was designed to ascertain proof of principle, we did not apply a measure of successful laboratory performance. The pilot established that the scheme design and methods used were acceptable for use in a larger scheme.

Second round. One hundred and seventeen laboratories from 30 countries registered, and 101 participated in the second round (due to customs issues we were not able to get samples to 16 labs), run in the second quarter of 2012. Ninety-one laboratories submitted results within the 8-week time frame – the remaining 10 labs gave

Table 3. Results from quantitative material validation

Sample no.	<i>EGFR</i> mutation result ^a	Estimated allelic frequency (%)	Confirmed allelic frequency (%) ^b			
			p.(G719S)	p.(E746_A750del)	p.(T790M)	p.(L858R)
B4, C1	Wild type	0	0	0	0	0
B10, C2	c.2155G>A, p.(G719S)	25	19.5	0	0	0
B2, C3	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	10: 10	0	0	19.3	19.1
B5, C4	Wild type	0	0	0	0	0
B9, C5	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	50: 50	0	0	c	c
B6, C6	Wild type	0	0	0	0	0
B7, C7	Wild type	0	0	0	0	0
B1, C8	c.2235_2249del15, p.(E746_A750del)	25	0	41.2	0	0
B3, C9	c.2235_2249del15, p.(E746_A750del)	5	0	15.3	0	0
B8, C10	c.2155G>A, p.(G719S)	10	8.3	0	0	0

^aExons 18–21.
^bConfirmed by droplet digital PCR.
^cNot measured due to insufficient material.

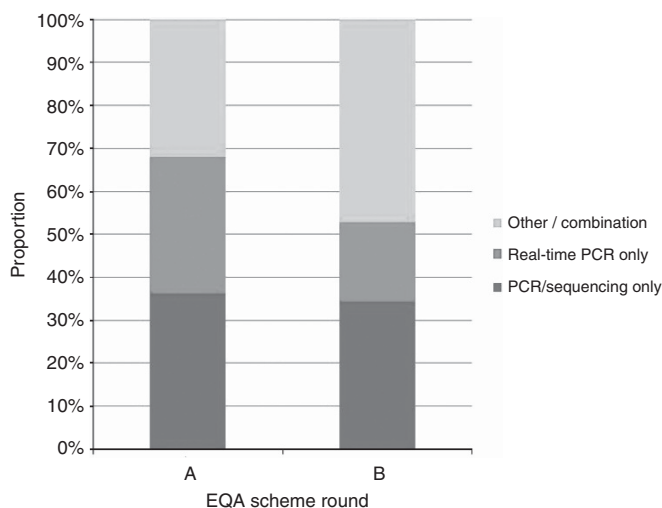


Figure 2. Methodologies used in rounds 1 and 2.

no reason why they did not submit results. A different set of samples from those used in the pilot first round were sent to the laboratories with the emphasis being on the inclusion of mutations at allelic frequencies that would challenge the analytical sensitivity of all the commonly used technologies (Table 2; samples B1–B10). A code number different from the one assigned in the first round was given to the samples. In addition to the genotype results, all participating laboratories were also required to submit for assessment copies of their clinical reports for three samples (B1, B4 and B9).

The main methodology used by the participants was PCR/sequencing ($n = 35$ laboratories; 39%) and real-time PCR ($n = 17$; 18.6%; Figure 2). It was common for labs to use a combination of different methodologies in their testing process (Table 4).

A variety of different errors were detected by the second scheme round, including 74 (8.1%) genotype errors (false-positive ($n = 13$; 1.5%), false-negative ($n = 61$; 82.4%) and a combination of false-negative and -positive results ($n = 1$; 1.4%)), as well as analytical test failures ($n = 31$; 3.4%), mispositioning of the genotype ($n = 7$; 0.8%) and significant errors in the mutation nomenclature ($n = 36$; 3.9%). Two samples (B2 and B8) gave a disproportionately high error rate compared with the other samples used in this round (Table 5), with 94.1% of errors for B2 made by labs using

PCR/sequencing vs 40.7% of errors for sample B8 made by labs using a version of the Therascreen *EGFR* kit (Qiagen). Laboratories did not lose marks if the declared limitations of their assay meant that they would not detect a particular mutation, at the given frequency used in the EQA materials. Eighteen laboratories (19.8%) from 13 countries with a total score below 18 did not pass the second round and were thus classified as poor performers – 72.2% of these labs used PCR/Sequencing as their main diagnostic test for *EGFR* mutation status. The interpretation of the test result relative to the clinical referral was reviewed with laboratories receiving comments on their performance, but no marks assigned.

Overall, 46 (50.5%) of the laboratories had a score ≥ 18 in the second round and passed the EQA. All laboratories received a certificate of participation that displayed their performance in the scheme.

Third round. The 18 laboratories that did not pass the second round were given the opportunity to participate in a third round. One laboratory was unable to participate due to problems with customs sample import permissions. The same set of samples used in the second round was sent to the laboratories in the third quarter of 2012 (Table 2; samples C1–C10). To obscure the sample identity from the labs, a different code number from the one assigned in the second round was given to the samples. Only the genotyping result was assessed for each of the 17 laboratories that returned results.

A total of 18 (10.6%) genotyping errors (reported false-positive ($n = 1$; 5.5%) and false-negative results ($n = 13$; 94.4%)) were made, as well as analytical test failures ($n = 3$; 1.8%). Eight laboratories (47.1%) missed the same mutation in identical samples in rounds 2 and 3 (for example, B3/C9, $n = 1$; B9/C5, $n = 2$; B8/C10, $n = 5$), indicating a failure of their assay for a particular mutation. Three laboratories (17.6%) had a pattern of errors, indicating a more general assay validation problem. Four laboratories (23.5%) did not pass the third round and scored < 18 .

Overall, all 17 laboratories (100%) improved their performance compared with the second round, with 6 (35.3%) labs getting the correct result for all 10 samples.

DISCUSSION

The clinical significance of somatic aberrations in oncology has seen rapid progression in the last couple of years with the

Table 4. *EGFR* testing methodological combinations used by labs in round 2

Methodological combinations	Count
Pyrosequencing + fragment length analysis	3
Pyrosequencing + high-resolution melting	1
Pyrosequencing + high-resolution melting + fragment length analysis + SNaPshot kit	1
Pyrosequencing + NextGen sequencing	1
Pyrosequencing + Therascreen kit	1
Sequencing + AmoyDx kit	1
Sequencing + denaturing capillary electrophoresis	1
Sequencing + fragment length analysis	4
Sequencing + fragment length analysis + high-resolution melt analysis + restriction fragment length polymorphism	1
Sequencing + fragment length analysis + high-resolution melt analysis + SNaPshot	1
Sequencing + fragment length analysis + restriction fragment length polymorphism	1
Sequencing + fragment length analysis + Taqman	1
Sequencing + high-resolution melting	4
Sequencing + MassArray analysis	1
Sequencing + pyrosequencing	1
Sequencing + pyrosequencing + high-resolution melting	2
Sequencing + restriction fragment length polymorphism	1
Sequencing + single-strand conformational analysis	1
Sequencing + Taqman + PNA clamp	1
Sequencing + Therascreen kit	5
Sequencing + Therascreen kit + CAST PCR	1
SNaPshot + single-strand conformational analysis	1
Therascreen kit + fragment length analysis + SNaPshot kit	1

correlation of treatment-related outcomes to gene alterations (Normanno *et al*, 2013). The rapid development and approval for use of therapeutic drugs based on mutational tests has represented a significant innovation for medical oncology, but also a major challenge for oncologists, pathologists and clinical scientists. Companion diagnostics and related guidelines have traditionally lagged behind such clinical indications for a variety of reasons. Although EQA schemes are running for the *KRAS* (sporadic colorectal cancer) and *BRAF* (malignant melanoma) genes, such schemes for *EGFR* are further complicated by the diverse spectrum of mutations, tumour heterogeneity and issues pertaining to the availability of appropriate biological material for use as EQA samples. For this reason, ESMO, ETOP, ESP and other stakeholders collaborated with the EMQN to develop an integrated approach to offer EQA for *EGFR* mutation testing in NSCLC to laboratories around the world. The purpose of the scheme is the provision of accurate and reliable *EGFR* testing for patients by assuring parity of test outcomes among participating laboratories from around the world.

Different types of samples can be used for this type of EQA (van Krieken *et al*, 2013). In the majority of existing schemes, FFPE patient biopsy samples have been used (Bellon *et al*, 2011; Deans *et al*, 2011; Thunnissen *et al*, 2011; Normanno *et al*, 2013). Although this type of sample enables the complete analytical pathway to be assessed ensuring a closer relationship between the EQA and routine clinical activity, it is limited by the amount of human tissue that is available and issues related to transport of samples across national borders. For this scheme, we developed a different approach and used artificial materials composed of FFPE cell lines to allow us to provide exactly the same sample to all the participating laboratories, even when these were numerous. In addition, it was possible to generate homogenous samples with variable content of mutant alleles by mixing wild-type and mutant cells and to use them to uncover hidden weaknesses in test

performance (sensitivity, specificity). This EQA scheme assessed both the genotyping, and the interpretation of the clinical significance of the results. Other EQA schemes that employed tumour samples have also addressed the pre-analytical phase, with laboratories being required to assess the percentage of neoplastic cells and to perform dissection if needed. However, scoring of this phase is not easy, as no consensus on the estimation of tumour cell content has been reached and a huge variability has been reported in previous EQA schemes (Thunnissen *et al*, 2011; van Krieken *et al*, 2013). Our approach is similar to that described by Normanno *et al* (2013) as it reduces the inter-laboratory variability related to the type of technique used for the dissection and allows a comparative evaluation of the sensitivity and specificity of the methods used for the mutational analysis. We used a total of 10 samples for each round of the scheme as has been suggested as the adequate number of cases for proficiency testing in this area of pathology (Thunnissen *et al*, 2011).

Overall, the results of the EQA suggest that there is more work to be done if laboratories are to ensure that the quality of their *EGFR* mutation testing meets an acceptable standard, as only 72 out of 91 (72%) of the laboratories passed the EQA and is in line with the findings of other similar EQA schemes (Deans *et al*, 2013). The threshold that was used (≥ 18 points) is comparable with that of the majority of other EQA schemes in this field (van Krieken *et al*, 2013). We weighted the marking of the errors' type differently depending on its implications for the patient. Errors resulting in no diagnosis (analytical failures) or no change in diagnosis (different mutation) were marked more leniently (deduction of 1.00 mark) when compared with those that resulted in the wrong diagnosis for the patient (false-negative or false-positive results; deduction of 2.00 marks). Given the potentially significant clinical consequences for the patient of any error, we propose to change the threshold criteria to > 18 points (Normanno *et al*, 2013) in future schemes to ensure that any significant errors are picked up in the performance data.

The materials we used in the scheme allowed us to control for some of the causes of variation seen in other EQA schemes, such as influence of fixative on DNA quality (Bellon *et al*, 2011). However, our results show that not all laboratories are able to produce *EGFR* mutation test results to a high standard. At least 1 in 10 samples is genotyped incorrectly in $> 19\%$ of laboratories (round 2), although there is evidence to show that this does improve with continued participation (round 3). In the first round proof of principle EQA, we identified five laboratories that made systematic errors in their pre or post-analytical processes. The way the sample blocks were labelled may have contributed to these errors, but it is impossible to identify the true root cause from our data. However, the outcome of these errors was that the correct results were reported for the wrong patients, indicating a failure of their quality management system.

False-negative and false-positive results were the main sources of genotyping error in the scheme. Both these results are extremely harmful for NSCLC patients. In fact, a negative result will lead to treatment of an *EGFR*-mutant patient with chemotherapy as first line of therapy, which is less effective as compared with *EGFR* TKIs in this subgroup of patients (Mok *et al*, 2009). On the other hand, false-positive findings will lead to treatment of *EGFR* wild-type patients with *EGFR* TKIs that have been shown to be detrimental as first-line treatment in this subgroup of patients, who benefit more from first-line chemotherapy (Mok *et al*, 2009; Fukuoka *et al*, 2011).

False-negative results accounted for 85% of all the genotype errors made in the scheme and might be due to the low sensitivity of the method used for mutational analysis. PCR/sequencing was the most common method used in the scheme for scanning to detect point mutations. The major disadvantage of sequencing is that it is not very sensitive (Angulo *et al*, 2010), especially in samples with low tumour cell content. Real-time allele-specific tests, such as Qiagen's Therascreen *EGFR* kits, are much more

Table 5. Genotyping errors

Sample no.	<i>EGFR</i> mutation result ^a	Samples tested	False positive			False negative			Total errors	Error rate (%)	Round
			A ^a	B ^a	C ^a	A ^a	B ^a	C ^a			
1	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	23	0			4			4	17.4	A1
2	c.2235_2249del, p.(E746_A750del)	24	0			0			0	0.0	A2
3	Wild type	24	5			0			5	20.8	A3
4	Wild type	23	5			0			5	21.7	A4
5	c.2235_2249del, p.(E746_A750del)	24	0			5			5	20.8	A5
6	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	24	0			5			5	20.8	A6
7	c.2235_2249del, p.(E746_A750del)	24	0			1			1	4.2	A7
8	Wild type	23	0			0			0	0.0	A8
9	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	23	0			0			0	0.0	A9
10	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	24	0			1			1	4.2	A10
11	c.2235_2249del, p.(E746_A750del)	107		1	0		1	0	2	1.9	B1 and C8
12	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	107		2	0		16	1	19	17.8	B2 and C3
13	c.2235_2249del, p.(E746_A750del)	100		2	0		3	4	9	9.0	B3 and C9
14	Wild type	104		0	1		0	0	1	1.0	B4 and C1
15	Wild type	103		2	0		0	0	2	1.9	B5 and C4
16	Wild type	104		2	0		0	0	2	1.9	B6 and C6
17	Wild type	103		2	0		0	0	2	1.9	B7 and C7
18	c.2155G>A, p.(G719S)	104		2	0		25	10	37	35.6	B8 and C10
19	c.2369C>T, p.(T790M); c.2573T>G, p.(L858R)	106		0	0		7	2	9	8.5	B9 and C5
20	c.2155G>A, p.(G719S)	104		0	0		9	0	9	8.7	B10 and C2

^aExternal quality assessment scheme rounds – first (A), second (B) and third (C).

sensitive and specific, but only test for a subset of common mutations (Lopez-Rios *et al*, 2013).

However, it is difficult for us to draw strong conclusions from this scheme about the errors made using the different technologies due to the lack of detailed data provided by the labs on exactly which methods were used for each sample, except for samples B2/C3 and B8/C10 that gave a disproportionately high error rate compared with the other samples used. We hypothesised that our estimates of mutational allelic frequency in the samples used in rounds 2 and 3 were inaccurate possibly due to *EGFR* gene copy number variation. We therefore undertook further quantitative validation on these samples using ddPCR to establish the true allelic frequency (Hindson *et al*, 2011). This innovative approach enabled us to establish that in three of the samples the true value was higher than expected, and for the other two samples lower than expected. However, crucially for two samples (B2/C3 and B8/C10), the value established by ddPCR is very close to the expected minimum level of methodological sensitivity (for example, 15% for Sanger sequencing, and 5.43% for the p.(G719S) mutation as defined in version 1 of the Qiagen Therascreen kit packaging insert). We speculate that latent problems with the pre-analytical processes used by these labs (for example, poor recovery of DNA, inaccurate DNA quantification) resulted in false-negative results due to suboptimal analytical conditions. These include insufficient method validation, misinterpretation of raw data, lack of awareness of assay limitations, sample contamination and poor in-house assay design. Nevertheless, all of the mutations were identified by the validating laboratories using a range of different methodologies used by the participant laboratories. These findings confirm that every laboratory should be undertaking an appropriate test validation or verification to define the limits of detection and measurement uncertainty of the techniques they are using. This is a requirement for all labs that are accredited to the ISO 15189:2012 standard (International Organisation for Standardization (ISO), 2010).

A fundamental aspect of all diagnostic testing is the accurate reporting of the results. The quality of reports submitted was

acceptable with a large proportion being comprehensive, stand-alone documents containing most of the basic core elements. However, the report is meaningless if the referring clinician cannot easily extract the relevant information. Therefore, it is essential that that the report is clear, concise and easy to read. Many of the reports obscured the take home message and there was often a lack of clarity and balance between the test information and the clinical context. Standardisation of the reporting and naming of mutations is also important and we assessed labs against the nomenclature guidelines from the Human Genome Variation Society (HGVS) (Human Genome Variation Society (HGVS), 2014). For example, we considered that it was not acceptable to report the amino acid change only, as redundancy in the genetic code means that different changes at the nucleotide level can result in the same change at the amino acid level.

In conclusion, the results of this EQA scheme suggest that the technical quality of *EGFR* mutational analysis could be improved as evidenced from a high level of diagnostic errors. Overall, the standard of reporting was acceptable. These findings also underline the importance of EQA as a mechanism to reveal errors in methodology and to ensure an adequate quality of molecular testing. Regular participation in EQA should be seen as a routine part of the diagnostic testing process for all labs helping to improve and standardise their processes. We have established a model for a robust and scalable EQA that can contribute to global optimisation and improvements in the overall quality of *EGFR* testing for patients with NSCLC.

ACKNOWLEDGEMENTS

We thank the laboratories for their participation in this study. We also thank AstraZeneca for the provision of an unrestricted grant to allow the development to the scheme; and Horizon Diagnostics for their support in quantitative validation of the materials.

Dr Normanno is supported by a grant from Associazione Italiana per la Ricerca sul Cancro (AIRC).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Angulo B, Garcia-Garcia E, Martinez R, Suarez-Gauthier A, Conde E, Hidalgo M, Lopez-Rios F (2010) A commercial real-time PCR kit provides greater sensitivity than direct sequencing to detect *KRAS* mutations: a morphology-based approach in colorectal carcinoma. *J Mol Diagn* **12**: 292–299.
- Bellon E, Ligtenberg MJ, Tejpar S, Cox K, de Hertogh G, de Stricker K, Edsjo A, Gorgoulis V, Hofler G, Jung A, Kotsinas A, Laurent-Puig P, Lopez-Rios F, Hansen TP, Rouleau E, Vandenberghe P, van Krieken JJ, Dequeker E (2011) External quality assessment for *KRAS* testing is needed: setup of a European program and report of the first joined regional quality assessment rounds. *Oncologist* **16**: 467–478.
- Deans ZC, Bilbe N, O'Sullivan B, Lazarou LP, de Castro DG, Parry S, Dodson A, Taniere P, Clark C, Butler R (2013) Improvement in the quality of molecular analysis of *EGFR* in non-small-cell lung cancer detected by three rounds of external quality assessment. *J Clin Pathol* **66**: 319–325.
- Deans ZC, Tull J, Beighton G, Abbs S, Robinson DO, Butler R (2011) Molecular genetics external quality assessment pilot scheme for *KRAS* analysis in metastatic colorectal cancer. *Genet Test Mol Biomarkers* **15**: 777–783.
- De Luca A, Normanno N (2010) Predictive biomarkers to tyrosine kinase inhibitors for the epidermal growth factor receptor in non-small-cell lung cancer. *Curr Drug Targets* **11**: 851–864.
- European Molecular Genetics Quality Network (EMQN) (2014) EMQN home page. Available at <http://www.emqn.org> (accessed 18 January 2014).
- Fukuoka M, Wu YL, Thongprasert S, Sunpaweravong P, Leong SS, Sriuranpong V, Chao TY, Nakagawa K, Chu DT, Saijo N, Duffield EL, Rukazekov Y, Speake G, Jiang H, Armour AA, To KF, Yang JC, Mok TS (2011) Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *J Clin Oncol* **29**: 2866–2874.
- Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**: 8604–8610.
- Human Genome Variation Society (HGVS) (2014) Nomenclature for the description of sequence variants. Available at <http://www.hgvs.org/mutnomen/> (accessed 18 January 2014).
- International Organisation for Standardization (ISO) (2012) ISO 15189:2012 Medical laboratories—requirements for quality and competence. Available at http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=56115 (accessed 18 January 2014).
- Linaridou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S (2008) Assessment of somatic *k-RAS* mutations as a mechanism associated with resistance to *EGFR*-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *Lancet Oncol* **9**: 962–972.
- Lopez-Rios F, Angulo B, Gomez B, Mair D, Martinez R, Conde E, Shieh F, Tsai J, Vaks J, Current R, Lawrence HJ, Gonzalez de Castro D (2013) Comparison of molecular testing methods for the detection of *EGFR* mutations in formalin-fixed paraffin-embedded tissue specimens of non-small cell lung cancer. *J Clin Pathol* **66**: 381–385.
- Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, Gemma A, Harada M, Yoshizawa H, Kinoshita I, Fujita Y, Okinaga S, Hirano H, Yoshimori K, Harada T, Ogura T, Ando M, Miyazawa H, Tanaka T, Saijo Y, Hagiwara K, Morita S, Nukiwa T (2010) Gefitinib or chemotherapy for non-small-cell lung cancer with mutated *EGFR*. *N Engl J Med* **362**: 2380–2388.
- Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, Seto T, Satouchi M, Tada H, Hirashima T, Asami K, Katakami N, Takada M, Yoshioka H, Shibata K, Kudoh S, Shimizu E, Saito H, Toyooka S, Nakagawa K, Fukuoka M (2010) Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol* **11**: 121–128.
- Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA, Fukuoka M (2009) Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**: 947–957.
- Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, Carotenuto A, De Feo G, Caponigro F, Salomon DS (2006) Epidermal growth factor receptor (*EGFR*) signaling in cancer. *Gene* **366**: 2–16.
- Normanno N, Pinto C, Taddei G, Gambacorta M, Castiglione F, Barberis M, Clemente C, Marchetti A (2013) Results of the First Italian External Quality Assurance Scheme for somatic *EGFR* mutation testing in non-small-cell lung cancer. *J Thorac Oncol* **8**: 773–778.
- Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Munoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombardieri P, Bernabe R, Bearz A, Arta A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M, Paz-Ares L (2012) Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced *EGFR* mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* **13**: 239–246.
- Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**: 169–181.
- Thunnissen E, Bovee JV, Bruinsma H, van den Brule AJ, Dinjens W, Heideman DA, Meulemans E, Nederlof P, van Noesel C, Prinsen CF, Scheidel K, van de Ven PM, de Weger R, Schuurin E, Ligtenberg M (2011) *EGFR* and *KRAS* quality assurance schemes in pathology: generating normative data for molecular predictive marker analysis in targeted therapy. *J Clin Pathol* **64**: 884–892.
- van Krieken JH, Normanno N, Blackhall F, Boone E, Botti G, Carneiro F, Celik I, Ciardiello F, Cree IA, Deans ZC, Edsjo A, Groenen PJ, Kamarainen O, Kreipe HH, Ligtenberg MJ, Marchetti A, Murray S, Opdam FJ, Patterson SD, Patton S, Pinto C, Rouleau E, Schuurin E, Sterck S, Taron M, Tejpar S, Timens W, Thunnissen E, van de Ven PM, van Krieken JH, Siebers AG, Normanno N. Quality Assurance for Molecular Pathology group (2013) European consensus conference for external quality assessment in molecular pathology. *Ann Oncol* **24**: 1958–1963.
- Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S, Lu S, Zhang L, Hu C, Hu C, Luo Y, Chen L, Ye M, Huang J, Zhi X, Zhang Y, Xiu Q, Ma J, Zhang L, You C (2011) Erlotinib versus chemotherapy as first-line treatment for patients with advanced *EGFR* mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* **12**: 735–742.



This work is licensed under the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)