Standardisation of EGFR FISH in colorectal cancer: results of an international interlaboratory reproducibility ring study

Andrea Sartore-Bianchi,1 Steffen Fieuws,2 Silvio Veronese,3 Mauro Moroni,1 Nicola Personeni,4 Milo Frattini,5 Valter Torri,6 Federico Cappuzzo,4 Sara Vander Borght,7 Vittoria Martin,5 Margaret Skokan,8 Armando Santoro,4 Marcello Gambacorta,3 Sabine Tejpar,9 Marileila Varella-Garcia,8 Salvatore Siena1

ABSTRACT

Aims Epidermal growth factor receptor (EGFR) gene copy number evaluated by fluorescence in situ hybridisation (FISH) can discriminate among KRAS wild-type patients those with better outcome to EGFR-targeted therapy in metastatic colorectal cancer, further enhancing selection of patients. Nevertheless, enumeration of gene copies is challenging and the lack of analytical standardisation has limited incorporation of the test into the clinical practice. We therefore assessed EGFR FISH interlaboratory consensus among five molecular diagnostic reference centres.

Methods A set of 12 colorectal cancer samples circulated among laboratories, and samples were scored according to commonly agreed guidelines. Reproducibility was quantified using the standard error of measurement (SEM).

Results A SEM of 0.865 and a within-subject coefficient of variation (WSCV) of 28.8% for mean EGFR gene/nuclei and a SEM of 0.235 and a WSCV of 19.4% for the mean EGFR gene/CEP7 ratio were observed. Measurement of the fraction of cells displaying chromosome 7 polysomy showed WSCV of 46.6%, 34.0% and 51.0% for the fraction of cells displaying chromosome 7 polysomy.

Conclusions Molecular diagnosis of EGFR gene copy number by FISH varied largely among pathology centres, with fluctuations covering the whole range of proposed cut-offs of predictive usefulness from literature. Definition of a detailed scoring system and implementation of comprehensive training programmes for laboratories are therefore necessary before including the test into clinical practice.

INTRODUCTION

Targeted therapy of metastatic colorectal cancer (mCRC) with anti-epidermal growth factor receptor (EGFR) monoclonal antibodies (moAbs) cetuximab or panitumumab has been revolutionised by the introduction of genetic profiling of individual tumours. Although initial response rates of about 10% were seen in patients with chemotherapy-refractory mCRC, it was subsequently discovered that higher response rates in the range of 13%–17% were achievable in tumours without mutations in codon 12 or 13 of the KRAS gene, whereas only 0%–1.2% of the KRAS mutant tumours responded to therapy.1 Nevertheless, even in KRAS wild-type CRCs, about 40% of the previously untreated and about 60%–70% of the previously treated do not respond to anti-EGFR treatment and additional detection of NRAS, BRAF and PIK3CA exon 20 mutations and loss of PTEN protein or better discrimination among KRAS mutations by excluding G13D carriers may further enhance selection of patients.

In addition to these negative predictive molecular alterations, the EGFR gene copy number (GCN) stood up as a candidate biomarker for predicting response of CRC to anti-EGFR therapy.7 EGFR GCN could indeed further discriminate among KRAS wild-type patients those better candidates to cetuximab or panitumumab, enhancing patients’ selection by achieving response rates as high as 80%.12–14 This notion has been recently supported also by a study in which a molecularly annotated platform of patient-derived xenografts (‘ xenopatients’) was exploited for identifying novel mechanisms of resistance to cetuximab, confirming that EGFR GCN gain (as assessed by quantitative PCR) tended to segregate responders also in this preclinical context.15 Fluorescence in situ hybridisation (FISH) has been used almost invariably in retrospective clinical studies for assessing EGFR GCN in CRC.11 13 14 However, signal enumeration in solid tumour sections by FISH is challenging to interpret16 17 and guidelines dealing with key technical issues and reading strategies like those available for non-small-cell lung cancer18 are not available for CRC. Thus, the lack of standardisation of analytical methods and scoring systems may partly explain why the EGFR GCN testing as a predictive biomarker has not been incorporated into the clinical practice yet.

We therefore designed this international ring study in order to assess interlaboratory consensus in EGFR copy number enumeration among five highly experienced molecular diagnostic centres with the aim of establishing variability in scoring and identifying issues that may contribute to discordant results.

MATERIAL AND METHODS

Study design

A slide-exchange programme was used to compare EGFR GCN FISH testing results among five pathology reference centres located in Belgium.
Specimen selection and logistics
Tumour samples were selected by the coordinating laboratory from anonymised CRC surgical specimens in such way to represent different FISH patterns. All samples were fixed with 10% neutral buffered formalin (12–48 h) and embedded into paraffin blocks. Tissue sections (4 μm thick) were mounted on positively charged glass slides with similar orientation and 25 slides were prepared from each of the 12 tumour specimens selected. Five slides from each specimen were then sent to each of the other four testing centres in a blinded manner, while five remaining slides from each specimen were retained by the sending laboratory for its own evaluation. Sequential slides were numbered following the scheme: laboratory A, slides 1, 6, 11, 16, 21 (the last one stained with H&E for morphology), laboratory B, slides 2, 7, 12, 17, 22 and so on for the other three laboratories.

Specimen analysis
FISH assays were performed by each testing centre using the EGFR/CEP7 FISH Probe Kit (Abbott Molecular, Des Plaines, Illinois, USA), according to its own operating protocol as previously described. The equipment used in each laboratory was the following: ZEISS Z 1 microscope with high-resolution camera and full Metasystem software (Ospedale Niguarda Ca’ Granda, Milano, Italy), ZEISS Axiosplan 2 microscope with Metasystems CCD camera and ISIS software from Metasystems (University Hospital Gasthuisberg, Leuven, Belgium), Olympus BX 61 microscope with high-resolution camera and Applied Imaging CytoVision Genus software (Istituto Clinico Humanitas, Rozzano, Italy), ZEISS Axioskop 2 plus microscope with high-resolution camera and AxioVision software from ZEISS (Istituto Cantonale di Patologia, Locarno, Switzerland) and ZEISS Axiosmager Z1 microscope with CCD camera and CytoVysion/Genus software from Leica Microsystems (University of Colorado School of Medicine, Aurora, Colorado, USA). Analysis was performed according to guidelines commonly agreed by laboratory directors followed by distribution of written instructions including details on how to (1) assess quality of specimen for analyses, (2) select eight tumour foci per specimen, (3) select nuclei for scoring, (4) count the signals in each nucleus and (5) define gene amplification (online supplementary document 1). Analyses in individual cells were reported in electronic worksheets for each individual specimen and subsequently sent to statisticians for analysis. FISH scores were based on counting of $EGFR$ and $CEP7$ signals measured as: (1) mean $EGFR$ signals per nucleus, (2) mean $EGFR$/mean $CEP7$ ratio, (3) mean of percentage of cells displaying ≤2 $EGFR$ signals, (4) mean of percentage of cells displaying ≥5 $EGFR$ signals and (5) mean of percentage of cells displaying ≥4 $CEP7$ signals. An additional testing was performed by one of participating centres (University Hospital Gasthuisberg, Leuven, Belgium) analysing intralaboratory reproducibility of FISH assay among different slides of the same specimen.

Statistical analysis
The reproducibility among laboratories was quantified using the standard error of measurement (SEM), which is the SD of the values within a specimen. If all laboratories assign the same value to a patient, the SEM equals 0. A SEM equal to 0.5 implies that for a specific patient, 95% of the obtained values (from various laboratories) are expected to fall in a range of ±0.5 in the true value. Furthermore, the differences in values between two laboratories are expected to fall in the range ±1.96× SD, which is known as the within-subject (specimen) coefficient of variation (WSCV). These indices reflect different sources of variability, that is, differences between laboratories (and/or observers), differences between slides within a specimen and measurement error. In the between-slides variability performed on the same specimen, the SEM and WSCV reflect only between-slide variability and measurement error. Mean $EGFR$/nucleus signals have also been categorised according to the following cut-offs: (1) from 0 to ≤2, (2) between 2 and ≤3 and (3) >3. The interlaboratory agreement for this categorisation was assessed using a (weighted) $k$ coefficient for multiple raters.

RESULTS
As depicted in figures 2 and 3, scoring of mean $EGFR$ GCN per nucleus (defining absolute $EGFR$ GCN of a given sample) and $EGFR$/CEP7 ratio (defining $EGFR$ status relative to the number of copies of chromosome 7 centromere and discriminating disomy, aneusomy or amplification) showed a low level of consensus among centres for both parameters. For mean $EGFR$ gene per nucleus, SEM was 0.865, thus indicating, with a mean value of 3.22, a WSCV of 26.8%. This denotes that if these

---

**Figure 1** Workflow of the ring study. The slide-exchange programme scheduled testing rounds in which tumour specimens were selected and sent by centre A (coordinator) to centres B, C, D and E in a blinded manner; after returning to centre A, results were subjected to statistical analysis performed by two independent biostatisticians. FISH, fluorescence in situ hybridisation.
95% of the measurements would be expected to fall in the range between values being 52.5% (\(\frac{-1.96}{3}\)) lower and 52.5% higher than the true value. As an example, for a patient with a true value of 3.22, the 95% range would be between 1.53 and 4.92. For the mean \(\text{EGFR/CEP7}\) ratio, the WSCV was 19.4% (SEM = 0.235). A trend towards having higher variability was noticeable in specimens with higher \(\text{EGFR}\) copy number.

Since it was proposed that \(\text{EGFR}\) GCN could be proficiently scored not only as mean \(\text{EGFR}\) GCN/nucleus but also in terms of fraction of chromosome 7 polysomy within the tumour specimen,\(^{15, 23}\) we additionally elected to test consensus among laboratories according to the \(\text{EGFR/CEP7}\) ratio. Similarly to mean \(\text{EGFR}\) GCN/nucleus, this parameter was associated with high variation, showing WPCV of 46.6%, 34.0% and 51.0% for percentage of cells displaying \(\leq 2\), \(\geq 3\) and \(\geq 4\) \(\text{EGFR}\) signals, respectively (figure 4).

**Figure 2** Reproducibility of scoring of mean epidermal growth factor receptor (\(\text{EGFR}\)) gene per nuclei (defining absolute \(\text{EGFR}\) gene copy number of a given sample) among the five institutions involved in the study. Each line represents the scorings of one laboratory given for each of the 12 specimens. The specimens are ordered according to their mean score. SEM, standard error of measurement (\(=\)within patient SD); WSCV, within-subject coefficient of variation, that is, the SEM expressed relatively to the mean value.

**Figure 3** Reproducibility of scoring of the mean ratio of epidermal growth factor receptor (\(\text{EGFR}\)) gene and CEP7 (defining \(\text{EGFR}\) status relative to the number of chromosome 7 centromeres and discriminating among disomy, aneusomy or amplification) among the five institutions involved in the study. Each line represents the scorings of one laboratory given for each of the 12 specimens. The specimens are ordered according to their mean score. SEM, standard error of measurement (\(=\)within patient SD). To represent the ratios appropriately, the y-axis is on a logarithmic (base 2) scale. WSCV, within-subject coefficient of variation, that is, the SEM expressed relatively to the mean value.

**Figure 4** Scoring of epidermal growth factor receptor (\(\text{EGFR}\)) gene copy number was also performed according to the fraction of chromosome 7 polysomy within tumour specimens, adopting a cut-off of percentage of cells displaying \(\leq 2\) signals (A), \(\geq 3\) signals (B) or \(\geq 4\) signals (C). As in previous graphics, each line represents the scorings of one laboratory given for each of the 12 specimens. The specimens are ordered according to their mean score. SEM, SE of measurement (\(=\)within patient SD). WSCV, within-subject coefficient of variation, that is the SEM expressed relatively to the mean value.
Finally, consensus analysis was supported by a non-parametric approach, by ordering data of mean EGFR gene per nuclei according to the following cut-offs for EGFR GCN per nuclei: (1) from 0 to ≤2, (2) between 2 and ≤3 and (3) >3 (table 1). In accordance with previous results, category ratings analysis indicated a low level of agreement among laboratories (κ=0.202, SE=0.072).

Analysis of FISH concordance within tumour specimens

Given the potential tumour heterogeneity of EGFR GCN within samples, an additional testing was performed analysing intralaboratory reproducibility of FISH assay among different slides of the same specimen. As depicted in figure 5, the WSCV was 6.1% for mean EGFR gene per nuclei and 3.9% for mean of EGFR/CEP7 ratios, hence accounting only for a small fraction of the observed interlaboratory disagreement. Evaluation of fraction of chromosome 7 polysomy within the same tumour specimen resulted in WSCV of 10.5%, 12.4% and 24.0% for the % of cells displaying ≤2, ≤3 and ≥4 EGFR signals, respectively (data not shown).

DISCUSSION

EGFR GCN has been proposed from several studies as a candidate biomarker for predicting response of CRC to anti-EGFR therapy by discriminating among KRAS wild-type patients those better candidates to cetuximab or panitumumab, thus enhancing patients’ selection. Nevertheless, these data come from retrospective analyses of patients’ cohorts and there is not a reference technique for scoring. EGFR gene status deregulation due to true amplification, defined as more than a doubling of the EGFR gene compared with the CEP7 copy number, rarely occurs in CRCs. Therefore, correlation with response has been mainly based on an increase in EGFR gene dosage caused by chromosome 7 polysomy, even though it is unknown whether balanced polysomy could have an equivalent biologic effect as compared with gene amplification in driving cancer progression and thus predicting response to EGFR-targeted agents. Discrepant cut-offs, in the range of 2.5–2.92 copies per cell, were proposed for discriminating responders from non-responders. Results of this slide-exchange ring study show that, even under standardised conditions by means of shared written guidelines and among highly experienced pathology centres, there was a low level of consensus for enumerating EGFR copy number in FISH assays in mCRC and that the observed variability in scoring translates into fluctuations alongside the whole range of marker usefulness. Subcategorisation of disomy (figure 4), which was recently reported to be associated with lack of response to cetuximab, did not improve reproducibility in our study. The study also suggests that discordant data are not due to tumour heterogeneity within samples but that the major factors for the lack of consensus should be technical, like the quality of the slide, the equipment used for the analyses and the personnel difference in interpretation of the guidelines. In this study, guidelines were discussed

Table 1 Distribution of category ratings (top) and relative frequencies (bottom) of mean EGFR gene copy number in participating laboratories according to the following cut-offs: (1) from 0 to ≤2, (2) between 2 and ≤3 and (3) >3 (κ=0.20168, SE=0.072)

<table>
<thead>
<tr>
<th>Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

Frequencies

<table>
<thead>
<tr>
<th>Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

*Rating, categories: (1) 0 to ≤2, (2) 2 to ≤3 and (3) >3.

EGFR, epidermal growth factor receptor.
and agreed upon by the laboratory directors and implemented, but a preliminary exchange of a training set of slides for refining interpretation could have helped in improving reproducibility. We assumed indeed ‘a priori’ that all laboratories involved in the study had the same experience and would have had similar interpretation of the guidelines, but there are many relevant skills necessary to perform enumeration of GCN in solid tumours, including identification of tumour versus non-tumour cells, focus across full depth of the section to account for complete nuclear area and, very important in CRC sections, identification of the correct individual nucleus since the individual cells may be diffusely overlapped. From this study, we learnt that in order to reach reproducible levels for all the variables explored, this strategy was not enough and that enumeration of GCN by FISH requires more intensive training to be reproducible among laboratories. For these reasons, a preliminary test of a subset of tumours according to the drafted guidelines in a process of ‘familiarisation’ would help by confirming that involved laboratories are doing hybridisation at the same quality level and applying same interpretation to the guidelines. The use of familiarisation for testing EGFR by FISH in non-small-cell lung cancer does indeed support this alternative strategy (Marielela Varella-Garcia, personal communication).

Various reports applied alternative assays for determination of EGFR copy number gain in mCRC, such as chromogenic in situ hybridisation,15 26 silver in situ hybridisation (SISH)27 and quantitative real-time PCR.15 28 Chromogenic in situ hybridisation and SISH, if scored manually, will have at least the same limitations of FISH, although enabling simultaneous morphological tissue characterisation which could facilitate interpretation. Quantitative PCR also has limitations, both due to the poor quality of material extracted from formalin-fixed paraffin-embedded sections and the fact that genomic regions used for control may be involved in losses or gains therefore impairing the estimative of the level of gain for the test gene. Furthermore, tumour DNA dilution by healthy cells during DNA extraction can affect results of this assay. Methods to improve GCN enumeration could therefore include automation of scoring by means of software enhancing statistical power that are already available for FISH and SISH. However, automated software analysis does not work well for colorectal histology because of the irregular shape of colon cancer cells, the admixture with surrounding stroma and the need to distinguish malignant cells from normal colonic cells. These systems work better indeed for tumours with large tumour nests and for assays based on a ratio between two probes (such as HER2/CEP17 in breast cancer) than for pure enumeration of GCN since the ratio is not affected by correctly identifying the contours of individual cells. Finally, since there is a poor correlation between EGFR IHC and GCN,12 26 28 further studies addressing whether in situ hybridisation techniques could be effectively guided by preliminary or synchronous IHC—and thus performed in areas with high EGFR IHC staining—should be warranted.

In conclusion, implementation of a comprehensive programme for standardisation of protocol and guidelines for EGFR GCN detection in mCRC is warranted in order to enhance reproducibility and subsequent dissemination into the clinic as a biomarker for predicting outcome to EGFR-targeted therapies. Such programme should be performed by setting up a larger follow-up study to address more in depth each parameter taken into consideration in present study in a larger sample size. A uniform protocol should be used since current results already show that existing literature cut-offs cannot be implemented and the programme should deliver not only written protocols and instructions but also technical training, possibly coupled by a familiarisation preliminary phase. Periodical proficiency testing should also be required from the clinical laboratories offering the test as a predictive assay for selecting patients.

Funding This work was supported by Oncologia Ca’ Granda ONLUS Fondazione and Associazione Italiana per la Ricerca sul Cancro (AIRC): Investigator Grant 4412 and 2010 Special Program Molecular Clinical Oncology 5 · 1000, project 9970.

Competing interests AS-B: following relevant financial activities outside the submitted work: payment for board membership (AMGEN); payment for lectures including service on speakers bureaus (AMGEN, Merck-Serono). FC: following relevant financial activities outside the submitted work: payment for consultancy (Roche). MV-G: following relevant financial activities outside the submitted work: payment to my Institution (University of Colorado) as co-inventor in a patent to use EGFR biomarker for selection of non-small-cell lung cancer patients to EGFR inhibitors therapy. SS: following relevant financial activities outside the submitted work: payment for Board membership (Roche, Merck-Serono, AMGEN). ST: following relevant financial activities outside the submitted work: payment to my Institution for research grants (Merck-Serono). Other co-authors: none declared.


Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES


Quan H, Shih WJ. Assessing reproducibility by the within-subject coefficient of variation with random effects models. Biometrics 1996;52:1195–203.


Figure 3  (A) An example of aspirated material from a cystic brain lesion, clinicoradiologically suspected to be a neoplasm, prepared into a cell block using the method as described. (B) The yield from the cystic material is ample, with excellent cytomorphological preservation, allowing a diagnosis of an inflammatory non-neoplastic lesion.

Figure 4  Scoring of epidermal growth factor receptor (EGFR) gene copy number was also performed according to the fraction of chromosome 7 polysomy within tumour specimens, adopting a cut-off of percentage of cells displaying ≥2 signals (A), ≥3 signals (B) or ≥4 signals (C). As in previous graphics, each line represents the scorings of one laboratory given for each of the 12 specimens. The specimens are ordered according to their mean score. SEM, SE of measurement (=within patient SD). WSCV, within-subject coefficient of variation, that is the SEM expressed relatively to the mean value.