EUTOS for CML

European Treatment and Outcome Study

Working definitions of Molecular Response in CML

Nick Cross, Martin Müller, Fabrizio Pane, Andreas Hochhaus
EUTOS Molecular Monitoring Steering Group

Version November 2013

1. Background

1.1 It is widely appreciated that second-generation tyrosine kinase inhibitors (TKIs) result in faster and deeper responses compared to imatinib for newly diagnosed chronic phase CML patients[i,ii]. Although the direct clinical benefit of achieving deep levels of molecular response (MR) is a matter of debate, there is considerable interest in the possibility that CML might be a curable disease for some patients. This possibility is being explored in carefully controlled studies that are evaluating cessation of TKI therapy in patients who achieve durable MR. These recent shifts in the CML treatment paradigm highlight the need for robust, standardised and workable definitions of deep MR. Specifically, it is critical that the measurement of MR is standardised in a manner to withstand both intra- and inter-laboratory variability as well as new methodological developments. Proposals for broad standardised definitions of MR at different levels on the International Scale (IS) have been published recently[iii] and endorsed in the recently updated ELN recommendations for treatment of CML patients[iv]. These publications, however, do not provide the detail to enable testing laboratories to convert their local results to IS results in a standardised fashion.

1.2 ENEST1st is a Phase IIIb, multicentre, open-label study of nilotinib in adult patients with newly diagnosed CML in chronic phase. The primary endpoint of ENEST1st trial is $MR^{4.0}$ at 18 months and molecular monitoring for the study has been undertaken by 14 laboratories across Europe, all of which use *ABL1* as a control gene (CG). This group has been piloting laboratory definitions of MR which were distributed widely to testing laboratories in 2011. This document provides an update on these definitions for general implementation locally. It is recognised that these definitions are evolving and feedback is welcome to improve them.

[i] Saglio et al., N Engl J Med. 2010;362(24):2251-9.
[ii] Kantarjian et al., N Engl J Med. 2010;362(24):2260-70
[iii] Cross et al., Leukemia. 2012;26(10):2172-5.
[iv] Baccarani et al., Blood. 2013;122(6):872-84

1.3 Version 1 of the definitions was issued in June 2011. The principal changes in Version 2 (August 2013) were (i) inclusion of *GUSB* as well as *ABL1* transcript numbers (paragraph 1.4), (ii) use of the IRMM ERM-AD623 BCR-ABL traceable calibration plasmid (paragraph 1.7), (iii) the recommendation that each individual amplification reaction should have at least 10,000 *ABL1* transcripts rather than the sum of the replicates being at least 10,000 (paragraph 3.1). In addition the term complete molecular response (CMR) is no longer used. Version 3 (November 2013) updates and clarifies the use of the IRMM ERM-AD623 BCR-ABL plasmid (paragraph 1.7).

1.4 The current definitions are:

MR^{4} = either (i) detectable disease ≤0.01% *BCR-ABL1*^{IS} or (ii) undetectable disease in cDNA with ≥10,000 *ABL1* or ≥24,000 *GUSB* transcripts

MR$^{4.5}$ = either (i) detectable disease ≤0.0032% *BCR-ABL1*^{IS} or (ii) undetectable disease with in cDNA with ≥32,000 *ABL1* or ≥77,000 *GUSB* transcripts

MR^{5} = either (i) detectable disease ≤0.001% *BCR-ABL1*^{IS} or (ii) undetectable disease with in cDNA with ≥100,000 *ABL1* or ≥240,000 *GUSB* transcripts[v]

1.5 Recent performance evaluation studies have indicated that testing laboratories should be able to achieve sufficient sensitivity to enable detection of MR$^{4.5}$ in most samples, but currently many laboratories fail to achieve this. It is important therefore that laboratories interested in measuring deep MR optimise their protocols to enable routine detection of maximal numbers of control gene targets (in the same volume of cDNA used to detect *BCR-ABL1*).

1.6 Until recently no standardized, traceable plasmid calibrator has been available; instead copy number estimates were made from optical density measurements and different calibrator dilutions were compared to each other. Studies have indicated that the use of different calibrators is a source of substantial variation in results between centres[vi] and that comparability of results can be improved by using a common calibrator. This is particularly important when estimating absolute CG numbers, which is essential when *BCR-ABL1* is undetectable. A standardized, traceable calibrator is now available from the Institute of Reference Materials and Measurement (IRMM), Belgium: catalogue number ERM-AD623; https://web.jrc.ec.europa.eu/rmcatalogue/searchrmcatalogue.do) which should either be used directly by testing laboratories in their routine analysis or indirectly to calibrate their own plasmid dilutions.

[v] The correspondence between *ABL1* and *GUSB* was calculated using data from 1495 cases kindly provided by the Vejle and Lund laboratories.
[vi] Müller et al., Leukemia. 2009;23(11):1957-6

1.7 In the performance evaluation study undertaken by EUTOS in 2011/2 it emerged that the values obtained using the IRMM ERM-AD623 BCR-ABL plasmid were roughly 2 fold lower that those obtained using many established calibrators. The reason for this is that the the copy numbers were assigned by digital PCR as numbers of double stranded plasmid molecules. cDNA is single stranded and therefore the assigned numbers need to be doubled to calibrate qRT-PCR assays.

2. Laboratory procedures

2.1 For routine monitoring, laboratories should use their established protocols and conversion to the International Scale for samples in which *BCR-ABL1* is detectable. Samples with detectable *BCR-ABL1* should be scored as MMR (also referred to as MR^3) if ≤0.1%, MR^4 if ≤0.01%, $MR^{4.5}$ if ≤0.0032% etc.

2.2 There is considerable variation in the way labs define 'undetectable' or 'low level positive' *BCR-ABL1*. The consensus opinion is that the Europe Against Cancer definition of undetectable should be used:

- The cut-off for positivity should correspond to a Ct of intercept + 1 (which should generally lead to cut-offs of 41 – 42 Ct). In other words, samples with a Ct higher than intercept + 1 are considered as undetectable.
- The no template control wells and reagent blanks should ideally not cross the threshold at any point but should certainly be at least 2 Cts above the intercept Ct for that run. If this is not the case then the run must be considered as failed.

2.3 Some centres routinely score samples that have a Ct higher than intercept + 1 as 'low level positive' if the Ct is clearly lower than the negative controls. In addition, many centres score positive samples with a Ct higher than that of the lowest plasmid standard as 'low level positive', '<10 *BCR-ABL1*', '<4 *BCR-ABL1*', positive outside the quantifiable range (POQR), etc. This presents a significant problem for scoring low levels of disease and therefore we suggest that:

- All samples considered to be low level positive must be assigned a specific number of *BCR-ABL1* transcripts (which in some cases may be <1 if some replicates are negative). [vii]
- If replicate analyses are routinely performed then the number of *BCR-ABL1* transcripts should be the mean value and the final result expressed on the IS, i.e.

[vii] It is recognized that this is controversial since real time PCR analysis cannot readily distinguish between 1, 2, 3 etc numbers of targets and that such small differences may determine, for example, whether a sample is scored as MR^4 or $MR^{4.5}$. Consequently this may be modified in future iterations of the definitions, e.g. in the light of clinical findings or new technologies such as digital PCR.

[(mean *BCR-ABL1*)/(mean *ABL1*)] x conversion factor x100. If any replicate test on the same sample is positive for *BCR-ABL1* then the overall result should be considered as positive.

2.4 If *BCR-ABL1* is undetectable then laboratories need to ensure that their estimates of CG numbers are standardized and that in their hands *BCR-ABL1* and the CG are amplified with comparable efficiencies. Recommendations for acceptability of run parameters have been published elsewhere [viii,ix].

3. Scoring of MR when *BCR-ABL1* is undetectable

3.1 Many laboratories routinely perform replicate analyses for both *BCR-ABL1* and *ABL1* in order to increase the accuracy of their results, however replicates can also be used to increase sensitivity when *BCR-ABL1* is consistently undetectable (**note: this is only valid when the replicate analyses are performed on exactly the same cDNA sample**). Provided that *BCR-ABL1* is undetectable for all replicates then it is legitimate to give a final result as [undetectable *BCR-ABL1*]/[sum of estimated *ABL1* in all the replicates]. Samples for which each individual replicate has a volume of cDNA estimated to have <10,000 *ABL1* or 24,000 *GUSB* trancripts should be considered as unevaluable.

3.2 For replicates that consistently test negative for *BCR-ABL1* and each replicate is seeded with cDNA estimated to have ≥10,000 *ABL1* or ≥24,000 *GUSB* transcripts should be scored as:

- MR^4 if the sum of *ABL1* is <32,000 or the sum of *GUSB* is <77,000
- $MR^{4.5}$ if the sum of *ABL1* is 32,000-99,999 or the sum of *GUSB* is ≥77,000-239,999
- MR^5 if the sum of *ABL1* is ≥100,000 or the sum of *GUSB* is ≥240,000

If any of the replicates are positive for *BCR-ABL1* then an estimate of the average number of *BCR-ABL1* targets should be made and reported on the IS as for any positive sample.

3.3 Examples:

- Example 1:
  - *BCR-ABL1* replicate 1: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 2: undetectable in 5µl cDNA
  - *ABL1* replicate 1: 16,500 copies in 5µl cDNA
  - *ABL1* replicate 2: 18,000 copies in 5µl cDNA

Result = undetectable *BCR-ABL1* in 34,500 *ABL1* = $MR^{4.5}$

[viii] van Dongen et al., Leukemia. 1999;13(12):1901-28
[ix] Foroni et al., Br J Haematol. 2011;153(2):179-90

- Example 2:
  - *BCR-ABL1* single analysis: undetectable in 5µl cDNA
  - *ABL1* single analysis: 45,000 copies in 5µl cDNA

  Result = undetectable *BCR-ABL1* in 45,000 *ABL1* = MR$^{4.5}$

- Example 3:
  - *BCR-ABL1* replicate 1: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 2: undetectable in 5µl cDNA
  - *ABL1* replicate 1: 7,000 copies in 5µl cDNA
  - *ABL1* replicate 2: 8,000 copies in 5µl cDNA

  Result = unevaluable for MR as *ABL1* <10,000 in each replicate

- Example 4:
  - *BCR-ABL1* replicate 1: undetectable in 2µl cDNA
  - *BCR-ABL1* replicate 2: undetectable in 2µl cDNA
  - *BCR-ABL1* replicate 3: undetectable in 2µl cDNA
  - *ABL1* replicate 1: 16,500 copies in 2µl cDNA
  - *ABL1* replicate 2: 18,000 copies in 2µl cDNA

  Result = undetectable *BCR-ABL1* in (3x mean *ABL1*= 51,750) = MR$^{4.5}$

- Example 5:
  - *BCR-ABL1* replicate 1: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 2: undetectable in 5µl cDNA
  - *ABL1* single replicate: 11,000 copies in 5µl cDNA

  Result = undetectable *BCR-ABL1* in 22,000 *ABL1* = MR^{4}

- Example 6 (Lab conversion factor = 0.8):
  - *BCR-ABL1* replicate 1: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 2: detectable in 5µl cDNA, estimated 2 copies
  - *ABL1* replicate 1: 18,000 copies in 5µl cDNA
  - *ABL1* replicate 2: 16,500 copies in 5µl cDNA

  Result = (mean *BCR-ABL1* = 1)/(mean *ABL1* = 17,250) x0.8 x100 = 0.0046% = MR^{4}

- Example 7 (Lab conversion factor = 0.8):
  - *BCR-ABL1* replicate 1: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 2: undetectable in 5µl cDNA
  - *BCR-ABL1* replicate 3: detectable in 5µl cDNA, estimated 4 copies
  - *ABL1* replicate 1: 14,000 copies in 5µl cDNA
  - *ABL1* replicate 2: 15,000 copies in 5µl cDNA

Result = (mean *BCR-ABL1* = 1.33)/(mean *ABL1* = 14,500) x0.8 x100 = 0.0074% = MR[4]

4. General considerations

For all qRT-PCR runs labs should routinely record:

- Equation of standard curve
- r^2 of standard curve
- Ct values for standard curve and samples
- Threshold setting
- Baseline setting
- If possible A260/A280 and A230

Ct values of the plasmid standards on every run should also be recorded. These should remain constant over time providing the same threshold is used. This enables the laboratory to check for run-to-run variability and drift over time, and gives confidence when comparing results over long periods. It is useful to look at the mean Ct values for the previous 4-6 month period and compare to previous periods to look for evidence of drift.

The calibrators pME-2[x] or the IRMM plasmid[xi] are recommended to normalise copy numbers across different targets. The Ct for equivalent data points for different transcripts (i.e. *BCR-ABL1* versus *ABL1* versus *GUSB*) should be the same (<1Ct) if reaction efficiencies are similar. If this is not the case, at least one of the reactions needs to be optimised.

[x] Müller et al., Leukemia. 2008 Jan;22(1):96-102
[xi] https://web.jrc.ec.europa.eu/rmcatalogue/searchrmcatalogue.do; catalogue number ERM-AD623