# AN INTERNATIONAL INTER-LABORATORY DIGITAL PCR STUDY DEMONSTRATES HIGH REPRODUCIBILITY FOR THE MEASUREMENT OF A RARE SEQUENCE VARIANT

Alexandra S. Whale[1], Alison S. Devonshire[1], George Karlin-Neumann[2], Jack Regan[2], Leanne Javier[2], Simon Cowen[3], Ana Fernandez-Gonzalez[1], Gerwyn M. Jones[1], Nicholas Redshaw[1], Julia Beck[4], Andreas W. Berger[5], Valérie Combaret[6], Nina Dahl Kjersgaard[7] Lisa Davis[8], Frederic Fina[9], Tim Forshew[10], Rikke Fredslund Andersen[11], Silvia Galbiati[12], Álvaro González Hernández[13], Charles A. Haynes[14], Filip Janku[15], Roger Lacave[16], Justin Lee[17], Vilas Mistry[18], Alexandra Pender[19], Anne Pradines[20], Charlotte Proudhon[21], Lao Saal[22], Elliot Stieglitz[23], Bryan Ulrich[24], Carole A. Foy[1], Helen Parkes[1], Svilen Tzonev[2] and Jim F. Huggett[1,25]*

[1] Molecular and Cell Biology Team, LGC, Queens Road, Teddington, Middlesex TW11 0LY, UK

[2] Digital Biology Center, Bio-Rad Laboratories, 5731 West Las Positas, Pleasanton, California 94588, USA

[3] Statistics Team, LGC, Queens Road, Teddington, Middlesex TW11 0LY, UK

[4] Chronix Biomedical, Goetheallee 8, 37073 Goettingen, Germany

[5] Department of Internal Medicine I, Ulm University, Albert-Einstein-Allee 23, Ulm, Germany

[6] Laboratoire de Recherche Translationnelle, Centre Léon-Bérard, Lyon, France

[7] Sjællands Universitetshospital, Klinisk Biokemisk Afdeling, Molekylærenhed, Sygehusvej 10, 4000 Roskilde, Denmark

[8] Genoptix Inc., 1811 Aston Avenue, Carlsbad, California 92008, USA

[9] Service de Transfert d'Oncologie Biologique, Laboratoire de Biologie Médicale, Faculte de médecine Nord, Boulevard Pierre Dramard, 13916 cedex 20 Marseille, France

[10] UCL Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London WC1E 6DD, UK

[11] Department of Clinical Immunology and Biochemistry, Vejle Hospital, Kabbeltoft 25, 7100 Vejle, Denmark

[12] Division of Genetics and Cell Biology, IRCCS San Raffaele Scientific Institute, via Olgettina 60, 20132 Milano, Italy

[13] Department of Pediatrics and CIMA LAB Diagnostics, Clínica Universidad de Navarra, Avenida Pío XII 36, 31008 Pamplona, Spain

[14] Michael Smith Laboratories, 301 Michael Smith Building, 2185 East Mall, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[15] The University of Texas MD Anderson Cancer Center; Department of Investigational Cancer Therapeutics (Phase I Clinical Trials Program), 1515 Holcombe Boulevard 0455, Houston, Texas 77030, USA

[16] Solid Tumours Genomics Unit, Tenon Hospital, APHP and Université Pierre et Marie Curie, 4 rue de la Chine, 75970 Paris, France

[17] Johns Hopkins, 1650 Orleans St, Baltimore, Maryland 21287, USA

[18] Department of Cancer Studies, University of Leicester, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester LE2 7LX, UK

[19] Lung Cancer Group, Division of Molecular Pathology, The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP, UK

[20] Institut Claudius Regaud – IUCTO, Laboratoire de Biologie Medicale Oncologique, 1 avenue Irène Joliot-Curie, 31059 Toulouse cedex 9, France

[21] Institut Curie, PSL Research University, SiRIC, Laboratory of Circulating Tumor Biomarkers, 26 rue d'Ulm, 75005 Paris, France

[22] Department of Oncology and Pathology, Lund University, Scheelevägen 2, MV 404-B2, SE-22381, Lund, Sweden

[23] Department of Pediatrics, Benioff Children's Hospital, UCSF Helen Diller Family Comprehensive Cancer Center, 1450 Third Street, San Francisco, California 94158, USA

[24] Dana Farber Cancer Institute, Belfer Center for Applied Cancer Science and Department of Medical Oncology, Boston, Massachusetts 02115, USA

[25] School of Biosciences & Medicine, Faculty of Health & Medical Science, University of Surrey, Guildford, Surrey GU2 7XH, UK

* Corresponding author: Jim Huggett, telephone: +44 (0)20 8943 7655, email: jim.huggett@lgcgroup.com.

## ABSTRACT

This study tested the claim that digital PCR (dPCR) can offer highly reproducible quantitative measurements in disparate labs. Twenty-one laboratories measured four blinded samples containing different quantities of a *KRAS* fragment encoding G12D, an important genetic marker for guiding therapy of certain cancers. This marker is challenging to quantify reproducibly using qPCR or NGS due to the presence of competing wild type sequences and the need for calibration. Using dPCR, eighteen laboratories were able to quantify the G12D marker within 12% of each other in all samples. Three laboratories appeared to measure consistently outlying results; however, proper application of a follow-up analysis recommendation rectified their data. Our findings show that dPCR has demonstrable reproducibility across a large number of laboratories without calibration and could enable the reproducible application of molecular stratification to guide therapy, and potentially for molecular diagnostics.

## KEYWORDS

digital PCR, ddPCR, reproducibility, *KRAS*, single nucleotide variant, rare sequence variant

## SIGNIFICANCE STATEMENT

The poor reproducibility of molecular diagnostic methods limits their application in part due to the challenges associated with calibration of what are relative measurement approaches. In this study we investigate the performance of one of the only absolute measurement methods available today, digital PCR (dPCR), and demonstrated that when compared across twenty-one laboratories, dPCR has unprecedented reproducibility. These

results were achieved when measuring a challenging single nucleotide variant and without calibration to any reference samples. This opens the possibility for dPCR to offer a method to transform reproducibility in the molecular diagnostic field, both by direct use as well as in support of other currently used clinical methods.

## INTRODUCTION

Quantification using real-time quantitative PCR (qPCR) was first described over two decades ago (1, 2). Since then, despite its application to a wide range of preclinical research areas and its capacity for performing precise nucleic acid quantification, there are only a small number of examples where it has been successfully translated into the clinic for quantification, primarily in the fields of clinical virology (3) and management of patients with CML (4). Obstacles to its widespread clinical adoption have been the challenges of standardisation and defining the technical reproducibility of the method (5, 6). In qPCR, reproducibility is often poor due to the fact that while it can be very precise, it can be biased (6); it is not unusual for viral titres to vary by several orders of magnitude between different laboratories in the absence of a calibrator material (7).

Digital PCR (dPCR), where the presence or absence of a target molecule is detected in a binary and absolute fashion, is an alternative method that has the potential to be considerably more reproducible than qPCR. It also offers a number of advantages that include high precision (8-13) and improved sensitivity and specificity (14-18), all without reliance on a calibration curve (19, 20). This high precision has been employed to value-assign certified reference materials that include a plasmid for *BCR-ABL1* monitoring in chronic myeloid leukaemia (21) and genomic DNA for *HER2* amplification detection in

breast cancer (22). Prototype reference materials of plasmid, genomic DNA and whole bacteria in synthetic sputum for Tuberculosis have also been evaluated using dPCR (12, 23). However, the method has not yet been demonstrated to surmount the challenges of end-user inter-laboratory reproducibility that will be necessary if it is to be used for routine clinical use.

The central aim of this study was to evaluate the reproducibility of droplet-based dPCR using an international end-user inter-laboratory comparison. The chosen target was a *KRAS* single nucleotide variant (SNV), one of the most challenging types of cancer biomarkers, which is being pursued vigorously in liquid biopsy translational studies that predict non-response to specific therapies in colorectal carcinomas (24, 25) and non-small cell lung cancers (26).

## RESULTS

**Design of the study.** Twenty-one participant laboratories were enrolled in the study to measure blinded samples that contained varying fractional abundances of a 186 bp plasmid fragment containing the *KRAS* G12D variant and/or the wild-type (*wt*) sequence (**Supplementary Document 1** and **Fig. S1**). Each participant laboratory was provided with three units of each of the four blinded test samples. These test samples were at different nominal G12D fractional abundances: 12%, 0.9%, 0.17% and 0% and randomly assigned a sample letter (A-D) (**Table 1**). Additionally two control samples were provided as three units of the negative control (0% G12D; which was the same as sample B and denoted NEG) and a single unit of the positive control (12% G12D; which was the same as sample D and denoted POS). The %G12D of the POS control was not revealed to the participants and so

could not be used as a calibrator for the blinded samples. Reagents and QX100/200™ consumables from common manufacturing lots (**Supplementary Document 2**) were provided to the participants along with a full protocol (**Supplementary Document 3**), that included the reaction preparation procedure and the plate layout, to minimise these as possible sources of variation. All packages were distributed by LGC who coordinated the study.

Participants used their own QX100/200™ Droplet Digital PCR System (Bio-Rad) with manual droplet generation and were requested to analyse their data using their routine procedure. Each laboratory submitted their mean value for the *KRAS* G12D and *wt* copy number concentrations and %G12D with the associated 95% confidence intervals for each sample, control and no template controls (NTCs) in a pre-designed spreadsheet format (summarised in **Table S1**). Details of their analysis method (**Table S2**) and screen shots of the analysis (not shown) were also requested.

**Inter-laboratory comparison.** For the quantification of the *KRAS* G12D and *wt* copy number concentrations and %G12D, dPCR measurements from the twenty-one laboratories were highly accurate over the three orders of magnitude (~0.42 to ~2200 copies/µL) reflected in the blinded test samples (**Fig. 1**) and two control samples (**Fig. S2**). There were two groups of results. The majority of the laboratories (18/21) measured both the copy number concentrations and fractional abundance of all samples within 20% of each other. The remaining three laboratories consistently under-quantified the G12D copy number concentration and associated fractional abundance in all samples, though they did report consensus values for the *wt* target concentration. There was no significant

146 difference in the copy number concentrations or fractional abundance of sample B and NEG
147 ($p > 0.12$) or sample D and POS ($p > 0.82$). The median participant values for all six samples
148 were within 6% of the mean values obtained in the homogeneity study performed by the
149 co-ordinating laboratory (LGC) that manufactured and characterised the samples (**Table**
150 **1**). The exception was for the G12D copy number concentration and fractional abundance
151 of sample B and NEG, that contained no *KRAS* G12D target, that varied by 16-28%.

152 Of the first group of results, fourteen laboratories had no outlier values in any of the test or
153 control samples (**Fig. 1** and **S2**). The remaining four laboratories (3, 7, 10 and 16) had
154 outlier values in one or more of the measurements for one or more samples which were
155 associated with moderately higher or lower *wt* or G12D copy number concentration or
156 fractional abundance (>1.5-fold different from the median value). The group that
157 consistently under-quantified the G12D molecules (2, 17 and 21) were all identified as
158 outliers and submitted values that were between 2- and 9-fold lower than the median for
159 the G12D (but not *wt*) copy number concentration and fractional abundance.

160 **Factors affecting reproducibility.** Following visual inspection of the two-dimensional
161 (2D) scatter plots produced by the QuantaSoft ddPCR analysis software, it was
162 hypothesised that the cause of the G12D under-quantification in laboratories 2, 17 and 21
163 was misclassification of the droplets (**Fig. S3**). This hypothesis was tested by preparation of
164 guidelines for droplet classification that were circulated to all participants
165 (**Supplementary Document 4**). At this stage all participants remained blind to the copy
166 number concentrations. The seven laboratories with one or more outlier were invited

167 optionally to reanalyse their original data following the guidelines. In parallel, the
168 coordinating laboratory reanalysed the seven laboratory data sets following the guidelines.

169 Four laboratories resubmitted their values (2, 10, 17 and 21) (**Table S3**). In contrast to
170 their originally submitted values, for two of the laboratories (2 and 21), no significant
171 difference was observed between their resubmitted values and the median values from the
172 original inter-laboratory comparison ($p$ >0.35) suggesting that the guidelines rectified the
173 droplet misclassification (**Fig. 2** and **S4**). However, laboratory 17 resubmitted values that
174 were ~1.2-fold higher than their original submission, but still significantly lower (>2-fold)
175 than the median values from the original inter-laboratory analysis ($p$ ≤0.012) (**Fig. 2** and
176 **S4**). Reanalysis of this data set by the coordinating laboratory identified no significant
177 difference between their values and the median values from the original inter-laboratory
178 analysis ($p$ >0.14) (**Fig. 2** and **S4**) suggesting that droplet misclassification, and not the
179 actual data generated, was still responsible for the under-quantification of this laboratory.
180 Laboratory 10, which was an outlier in only one value in one test sample (the G12D
181 fractional abundance in sample D), resubmitted values that were not significantly different
182 from those obtained by the coordinating laboratory (**Fig. 2** and **S4**) indicating that droplet
183 misclassification was not responsible for this single outlier result. The other three
184 laboratories (3, 7 and 16) communicated that they were satisfied that their original
185 analysis aligned with the guidelines; there was no significant difference between their
186 original values and those obtained by the coordinating laboratory ($p$ >0.34).

187 Between-laboratory precision was calculated for all samples and controls using the four
188 resubmitted laboratory values along with the original values from the other seventeen

laboratories. For nominal copy number concentrations >0.1 copies/µL, the precision was high across all the samples and targets; CVs were between 7.39% and 11.39% (**Table 2**). Calculation of the theoretical CV (tCV) for each value for each sample showed that the precision based on Poisson alone was generally only slightly smaller than the between-laboratory precision for the G12D and *wt* copy number concentrations in the samples that contained G12D molecules (**Table 2**) indicating a minimum of experimental error. For the corresponding G12D fractional abundances, the precision was high for all samples that contained G12D fragments (CV <4.8%).

For the *wt* only samples, that contained ~40,000 *wt* copies/reaction and no G12D fragments (Samples B and NEG), the identification of positive partitions for the G12D assay represented the false positive rate (FPR) of the assay. A median FPR of 0.02% G12D was observed in these samples that corresponded to a median of ~6 G12D-positive partitions/reaction (minimum of 0 and maximum of 14 across all 21 participants NEG reactions; n=198) (**Table S4**). Measurements of the *KRAS* G12D molecule in these samples had reduced precision (CV >24%) as would be expected for such low measured values. For the no template controls (NTCs), a median and mode of 0 G12D-positive partitions/reaction was observed (maximum of 3 across all 21 participants, n=186) (**Table S5**) indicating low, if any, contamination in the dPCR reactions.

The *KRAS wt* copy number concentrations were approximately equal for samples A, B and C (**Table 1**). To generate a metric to measure the impact of various deviations from the study protocol, the ratio of the *KRAS wt* copy number concentration of samples C and A was used. From the results of the questionnaire relating to the study protocol (**Table S2**) no

211 significant difference ($p$ >0.14) was observed related to the QX system or version of the
212 QuantaSoft analysis software used, number of times the samples were thawed prior to
213 analysis or any other deviations from the protocol such as the supplier of pipettes, plate
214 sealer or PCR cycler (**Fig. S5**).

215 Characterisation of each study material established that the between-unit variation did not
216 exceed the within-unit variation ($p$ >0.24 for all study materials) (**Fig. S6**) and so estimated
217 standard uncertainties due to possible inhomogeneity were calculated using the between-
218 unit variation. The relative standard uncertainties were <6% for samples containing G12D
219 molecules (**Table 1**) and so it was concluded that the unit homogeneity of all the study
220 materials was acceptable.

221 Stability studies identified no significant time or temperature effect over the short term for
222 the G12D and *wt* copy number concentrations with the exception of sample D/POS that had
223 a 1.1-fold decrease in both G12D and *wt* copy number concentrations (but no effect on the
224 fractional abundance, data not shown) when stored on dry ice for 7 days (**Fig. S7**). For the
225 long term stability study, small differences in the G12D and *wt* copy number concentrations
226 were observed over the six month duration of the stability study (**Fig. S8**). For both
227 stability studies the differences were minor (<12%), non-directional (with the exception of
228 sample D/POS), and all participants performed their experiments within one month of
229 receiving the samples, it was concluded that the stability of the materials was satisfactory.

230 **DISCUSSION**

231 The aim of this study was to investigate whether dPCR could offer reproducible
232 quantification between laboratories in the absence of calibration using a droplet-based
233 technology. We used an international inter-laboratory study focusing on the quantification
234 of a range of SNV concentrations, of varying fractional abundances, using droplet dPCR.
235 Participant laboratories were recruited from a range of institutes including hospitals,
236 universities and research industries. To our knowledge, this is the first study of its kind and
237 could form the basis of a framework for other studies investigating the reproducibility and
238 robustness of a given dPCR method. For this purpose, we used a clinically relevant
239 prognostic model of a *KRAS* point mutation that needs to be determined prior to selection
240 of treatment of specific cancers (27, 28).

241 We chose a range of G12D to *wt* fractional abundances, several of which are challenging to
242 quantify with currently used analytical methods (29, 30). Furthermore, as the interest in
243 liquid biopsies continues to rise, we directed the selection of the concentration and size of
244 the molecules quantified in this study to match this template type (31, 32). Comparison of
245 the independent measurements from twenty-one laboratories demonstrated the high level
246 of reproducibility of dPCR, defined as a CV of <12%. Further investigation identified that
247 most of this variation could be attributed to deviations in tube-to-tube unit homogeneity
248 and stability of the test materials as well as the Poisson error, thereby rendering this
249 reproducibility all the more impressive.

250 This study clearly demonstrated that fractional abundancies down to ~0.2% can be
251 reproducibly quantified using dPCR. As the observed FPR was ~0.02%, it suggests that a
252 lower limit of detection of the G12D molecule, such as 0.1%, may be achievable with dPCR.

253 However, the observed ~0.2% fractional abundance was enabled, in part, by the addition of
254 ~40,000 *KRAS* copies per reaction; this level of sensitivity may not always be achievable
255 using *in vivo* extracts such as cfDNA from plasma, that typically yields ~500 copies/mL
256 plasma (33). In addition to cancer models, other rare SNV models would also be applicable
257 to this framework (14, 15) as well as simpler targets that do not require specific
258 measurement of a given variant in the presence of the *wt* allele, such as quantification of
259 specific pathogens or gene fusions (34, 35).

260 Two distinct aspects of this technique demonstrate that dPCR is reproducible for absolute
261 quantification measurements. Firstly, all participant laboratories were blind to the target
262 quantities of each test sample, yet twenty laboratories were able to correctly quantify all
263 four test samples. Secondly, all the measurements were performed without any form of
264 calibration to a material of known quantity; the materials were simply quantified and
265 results reported directly.

266 In order to determine sources of discrepancy where they arose, we investigated a number
267 of parameters that could influence reproducibility and robustness. Deviations from the
268 study protocol, such as droplet handling using different pipettes and tips to those
269 recommended did not have significant effects on the quantification. Furthermore,
270 participants used different software versions to collect their data. Versions 1.3 and below
271 use a 0.91 nL droplet volume that has been identified as a source of bias in other studies
272 (36), whilst the smaller volume of 0.85 nL is used in versions 1.6 and above. While the use
273 of different partition volumes can impact on accuracy, in this study, the two partition

274 volumes used did not appear to impact on the between-laboratory reproducibility (**Fig.**
275 **S5**), suggesting that technical error masked this potential source of bias.

276 All participants used the same assays, reagents and consumables in this study; another
277 study where these were varied demonstrated good reproducibility (12). Additionally, this
278 study did not evaluate pre-analytical steps such as the extraction, however, with careful
279 optimisation, quantification can be reproducible even when the extraction step is
280 incorporated (23).

281 Incorrect classification of the positive and negative droplets was a main cause of under-
282 quantification by some laboratories. Partition misclassification has been highlighted as a
283 source of bias in dPCR quantification (37) and can be complicated by a number of factors
284 that include template type, source and matrix in addition to assay and reagent choice and
285 thermocycling conditions (12). When measuring SNV quantities further factors that can
286 impact on correct partition classification include assay specificity, partition specific
287 competition (PSC) and assay FPR. Assay specificity and PSC can make it more difficult to
288 classify the four cluster types (38) while the FPR, if not determined correctly can impact on
289 the sensitivity of the assay and lead to an underestimation.

290 The presence of G12D false-positive droplets in the *wt*-only (NEG) control in two of the
291 participant laboratories directed the analysts to be stringent with their droplet
292 classification and opt for increased specificity over sensitivity; such a method has been
293 adopted in another study (39). While this reduced the FPR (increasing specificity), it was at
294 the expense of sensitivity and resulted in a 2 to 3-fold negative bias due to the high total
295 DNA input per reaction. Therefore, for the most sensitive measurements, characterisation

296 of the FPR is needed. However, in order for the clinical utility of this method to be realised,
297 rigorous criteria for calling positive results, along with detailed investigation of the limits
298 of detection and quantification need to be developed and evaluated as illustrated with
299 *EGFR* mutations in lung cancer (40, 41).

300 **CONCLUSIONS**

301 This study has confirmed that dPCR can perform highly reproducible absolute
302 quantification of an SNV between laboratories differing in their reported value by <12%.
303 This was achieved without calibration, which is not possible using conventional qPCR when
304 measuring samples that contain challenging fractional abundances of <1%. Clinical
305 quantification of SNVs has been proposed for a range of applications such as monitoring
306 donor organ rejection (17) or in the treatment of cancer patients (16) and will likely be
307 needed for the development of precision/stratified medicine to be fully realised. These
308 findings suggest dPCR will have an important role in enabling such measurements. In the
309 short term, this study will contribute to a growing body of evidence that demonstrate that
310 dPCR can offer a valuable method to ensure reproducible quantification of nucleic acids.

311 **MATERIALS AND METHODS**

312 The coordinating laboratory for this study was LGC (UK). All experiments pertaining to the
313 production and characterisation of the study materials were performed at LGC. All kits and
314 instruments were used following the manufacturer's instructions.

315 **Production of the study materials.** A portion of the human *KRAS* gene encoding exon 2
316 and its flanking introns was downloaded from the NCBI database (NCBI accession:

NG_007524.1, bases 9788 to 11411). The location of the drug resistant mutation (c. 35G>A) that encodes the amino acid substitution G12D was identified. The inserts were synthesised using overlapping synthetic oligonucleotides and then blunt-end cloned into the *Eco*RV restriction site of the pUC57 plasmid (Eurofins) to generate two constructs: one containing the wild type (*wt*) sequence and one containing the G12D mutation. Sequence validation was performed by Sanger sequencing on both strands (Eurofins) (**Fig. S9A** and **S9B**).

Ten double restriction digest reactions were set up for each *KRAS* construct each containing approximately 10 µg of DNA with 25 units each of *Afl*III and *Nsi*I (New England Biolabs) in a 50 µL reaction volume. The reactions were incubated at 37 °C for three hours and subjected to electrophoresis through a 3% agarose gel. Each 186 bp fragment was identified using the 1kb+ DNA ladder (Invitrogen), gel purified using the Gel Extraction kit (Qiagen) and eluted in 30 µL of EB buffer; the 10 replicate reactions were pooled to give a total eluent volume of 300 µL. The two 186 bp fragments (KRAS *wt* and G12D)were assessed for size and purity using the 2100 Bioanalyzer with the DNA 7500 series II assay (Agilent) (**Fig. S9C** and **S9D**) and the concentration was estimated using the Qubit® 2.0 fluorometer with the dsDNA HS Assay Kit (Invitrogen) to give the 'nominal' stated values for each test sample and control. The copy number concentration was estimated based on the molecular weight of the 186 bp fragment (approximately 116 kDa) and the Avogadro constant using a standard method (42); 1 ng corresponds to approximately 2.01 x $10^8$ copies.

The four study materials were manufactured in 50 mL tubes by gravimetrically diluting the *KRAS wt* and G12D purified 186 bp fragments in non-human carrier to their nominal concentrations (**Table 1**). Carrier was commercially available sonicated salmon sperm genomic DNA (Ambion) at a final concentration of 25 ng/µL in the study materials. The study materials were mixed by rotation at 50 rpm at 4 °C for 30 minutes. One test material was prepared from each study material; these were randomly labelled as samples A, B, C or D to reduce analyst bias. In addition, two control materials were prepared from two of the study materials; the negative control (NEG) was the same as Sample B and the positive control (POS) was the same as Sample D. Following mixing, the four test and two control samples were aliquoted (>100 µL) in pre-labelled 0.5 mL tubes (P/N E1405-2120; StarLabs) to generate between 75 and 300 units and stored at -20 °C. To reduce contamination risks, the test materials were manufactured in order of increasing G12D concentration.

**Characterisation of study materials.** Unit homogeneity was evaluated across twelve randomly selected units for each study material that were maintained at -20 °C. The short term stability (STS) study was set up isochronously at a range of temperatures to monitor the stability of the study materials during shipment to the participants and the accelerated stability at high temperature. Three units of each of the four study materials were incubated on dry ice (shipment temperature), at 4 °C and 28 °C (accelerated stability temperature) for 2 and 7 days and compared to the baseline temperature (three units maintained at -20 °C). Following all temperature and time point conditions, the study materials were incubated at -20 °C overnight before dPCR analysis. The long term stability (LTS) study was performed to establish the stability of the study materials over the full

course of the study. Three units of each of the four study materials were maintained on dry ice for 5 days to mimic shipping conditions, before being transferred to -20 °C before storage for 1 or 7 months prior to dPCR analysis; parallel experiments were compared with samples maintained at -20 °C for the duration of the LTS. Details of the dPCR analysis performed are in the relevant section below.

**Droplet Digital PCR.** All dPCR experiments were implemented in accordance with the dMIQE guidelines (**Table S6**) (20) using the QX100™ or QX200™ Droplet Digital PCR System (Bio-Rad). Briefly, 22 µL reactions were established containing ddPCR Supermix for Probes with no dUTP, PrimePCR™ ddPCR™ Mutation Assays for KRAS p.G12D and KRAS WT (assay details in the relevant section below), 8.8 µL of study material and made up to volume with nuclease-free water. No template controls (NTCs) were performed using nuclease-free water in place of template in every experiment. Droplets generation was performed manually (*not* with the AutoDG™ system) from 20 µL of the reaction following manufactures guidelines. Thermocycling conditions were 95 °C for 10 minutes, 40 cycles of 94 °C for 30 seconds and 55 °C for 1 minute, followed by 98 °C for 10 minutes and a 4 °C hold. The ramp rate for each step was set to 2 °C/second. Droplet reading was performed following manufactures guidelines. Details of the protocol are given in **Supplementary Document 3**.

For the study material characterisation, experiments were performed using the QX200™ droplet digital PCR system with a randomised plate layout. Every study unit was analysed with triplicate wells. Aliquots of the carrier the study materials were made up with were also evaluated for the presence of *KRAS* sequences (carrier only controls); no positive

383 droplets for the *KRAS* G12D were observed. If multiple 96-well plates were necessary,
384 replicate units were distributed between the different plates with the triplicate dPCR wells
385 for each unit located on the same plate. The commercially available PrimePCR™ ddPCR™
386 Mutation Assays used were KRAS p.G12D, Human (P/N 10031246, Assay ID
387 dHsaCP2000001) and KRAS WT for p.G12D, Human (P/N 10031249, Assay ID
388 dHsaCP2000002) that generate a 90 nt amplicon. All PrimePCR assays are designed and
389 validated by the manufacturer on the QX100/200 droplet digital PCR system. The
390 QuantaSoft software version 1.7.4.0917 was used to collect the data in the 'ABS' mode. The
391 status of each well was checked and wells containing <10,000 droplets or a baseline drop
392 in the double-negative droplets compared with the NTC wells were omitted from the data
393 set. Thresholds were set using the 'auto analyze' and 'combined wells' setting in the 2D
394 amplitude mode to define the positive and negative droplets. The data was exported as a
395 .csv file for further analysis.

396 Prior to the inter-laboratory comparison, a pilot study was performed to test the
397 international shipment method of the study materials, reagents and consumables, and the
398 clarity of the protocol. Three Bio-Rad laboratories (USA, France and Germany) participated
399 in the pilot study to confirm that the shipment conditions were acceptable. Following the
400 data analysis, the preparation protocol for the reagents and the dPCR protocol were
401 amended. The final documents and details are included in the supplementary information
402 as cited below.

403 Participant laboratories were provided with the protocol by the coordinating laboratory
404 (**Supplementary Document 3**). The commercially available PrimePCR™ ddPCR™ Mutation

Assays used were KRAS p.G12D, Human (P/N 10031246, Assay ID **dHsaCP2500596**) and KRAS WT for p.G12D, Human (P/N 10031249, Assay ID dHsaCP2500597) that generates a smaller amplicon of 57 nt for each assay compared with the assay used for material characterisation. Each unit was analysed in triplicate wells alongside eight no template controls (NTCs) to give a total of 56 reaction wells in a single experiment; a plate layout was provided in the protocol. The droplets were read in the 'ABS' mode using a QuantaSoft template file provided by the coordinating laboratory. Participants could use either the QX100™ or QX200™ Droplet Digital™ system. Each laboratory was asked to complete a questionnaire with details of their ddPCR instrumentation, software version and data analysis methods (**Supplementary Document 3**). They were requested to report the concentration (copies/µL in the reaction) and percentage allelic frequency of the *KRAS* G12D in each of the four test samples and two controls with the associated 95% confidence intervals as well as the result of their NTCs. In addition to the questionnaire, laboratories were asked to submit their raw data (Quantasoft .csv and .*qlp* files) and screen shots of the 2D scatter plots of the control samples (NEG, POS and NTC).

**Data analysis**. For the characterisation of the test materials, exported .csv files were imported into MS Excel 2010 in the first instance with further analysis performed in the R statistical programming environment version 3.1.1 (http://www.r-project.org/) and Prism 6 (GraphPad). A tab deliminated file was generated in MS Excel 2010 containing the well number, sample name, unit number, replicate well number, software generated copies/µL in the reaction for the *KRAS* G12D and *wt* molecule.

For assessment of unit homogeneity, ANOVA and mixed effects models with maximum likelihood estimation described previously (43) were used to assess plate and unit variation, and to estimate the between-unit standard deviation. Analysis of STS and LTS study was performed by fitting a linear model to each test material with time and temperature as the covariates to determine if the copy number concentration varied with storage time; the time variate was treated as a continuous variable. Statistical significance was identified using the Bonfernoni correction for false discovery rate (44) so that significance was identified when $p$ <0.004 for the STS and $p$ <0.01 for the LTS.

The reported values and confidence intervals from each end-user laboratory were transcribed into Prism 6 for analysis and generation of graphs. The data was tested for normality using the Shapiro-Wilk test with non-normality assigned to data when $p$ <0.05. Robust statistics were used for the non-normally distributed data to estimate the consensus values and the between lab dispersion using the median and median absolute deviation (MAD), respectively. The MAD was calculated as follows using equation [1]:

$$MAD = median_i \left(\left|X_i - median_j(X_j)\right|\right) \quad [1]$$

Where $median_i$ is the median of the reported values from the laboratories, and $median_j$ is the median of the absolute difference between the reported values and median. The MAD was then used to estimate the $MAD_E$ using equation [2]:

$$MAD_E = MAD \times 1.4826 \quad [2]$$

Where the 1.4826 was selected as the consistency constant; for details on the calculation of this value see Staudt *et al.*, 1990 (45). Submitted values outside the upper and lower 99%

447 confidence intervals (calculated using 2.58 $MAD_E$ distributions away from the median)
448 were identified as outliers; this level of confidence was chosen as there were more than 20
449 laboratories enrolled in the study. A robust estimate of reproducibility that down weighs
450 the influence of outliers was calculated using equation [3]:

451 $$\%CV = \frac{MAD_E}{median} \quad [3]$$

452 Paired t-tests were used to compare two groups of data; for example, to test for differences
453 between sample B and the negative control, or between the participant resubmitted data
454 and the coordinating laboratory analysis.

455 **ACKNOWLEDGEMENTS**

456 The work described in this paper was funded by the UK government Department for
457 Business, Energy & Industrial Strategy (BEIS) and the European Metrology Research
458 Programme (EMRP) joint research project (SIB54) Bio-SITrace
459 (http://biositrace.lgcgroup.com) which is jointly funded by the EMRP participating
460 countries within EURAMET and the European Union. The authors would like to thank
461 David Littlemore (LGC) for his central support in the distribution of the materials needed
462 for this study and Pia Scheu (Bio-Rad, Munich, Germany) and Ronald Lebofsky (Bio-Rad,
463 Marnes, France) for their assistance in running pilot experiments to test shipping of
464 samples and kits, and the clarity of the protocol prior to the end-user inter-laboratory
465 study. The authors would also like to thank Cloud Paweletz (Dana Farber Cancer Institute)
466 for critical review of the manuscript prior to submission.

The following authors provided technical and analytical support to the participant laboratories: Isabelle Iacono and Stéphanie Brejon (Centre Léon Berard), Ekkehard Schuetz (Chronix Biomedical), Ana Patiño-Garcia (Clínica Universidad de Navarra), Antoine Carlioz (Assistance Publique Hôpitaux de Marseille), Reinhold Pollner (Genoptix), Virginie Poulot (Tenon Hospital), Anne Casanova and Gilles Favre (Institut Claudius Regaud), Charles Decraene (Institut Curie), Ben Ho Park (Johns Hopkins), Anthony M. George (Lund University), Peter Kaltoft Böhm Nielsen (Sjællands Universitetshospital), Vito Lampasona (Division of Genetics and Cell Biology, and Division of Genetics and Cell Biology & Diabetes Research Institute, San Raffaele Scientific Institute), Julian Downward (Institute of Cancer Research), Daniel Schwerdel (Ulm University), Alice Gutteridge (University College London), Curtis B. Hughesman (University of British Columbia), Daniel Fernandez Garcia and Jacqueline A. Shaw (University of Leicester) and Helen Huang (MD Anderson Cancer Center). The Dana Farber Cancer Institute would also like to acknowledge the Expect Miracles Foundation.

## AUTHOR CONTRIBUTIONS

A.S.W., A.S.D., G.K-N., J.R., S.C., S.T and J.F.H. designed the study, and analysed and interpreted the data. A.S.W., A.F-G., G.M.J. and N.R. produced and characterised the material. L.J. prepared and performed the quality control on the reagents and assays. G.K-N, J.F.H. and S.T. conceived the study. C.A.F., H.P. and J.F.H. obtained funding for the study. J.B., A.W.B., V.C., N.D.K., L.D., F.F., T.F., R.F.A., S.G., A.G.H., C.A.H., F.J., R.L., J.L., V.M., A.P., A.P., C.P., L.S., E.S. and B.U performed the experiments for the inter-lab study. The manuscript was

488 written by A.S.W., G.K-N., and J.F.H. All authors reviewed and provided editorial comments
489 on the manuscript.

490 **DISCLOSURES & CONFLICT OF INTERESTS**

491 G.K.N., J.R., L.J. and S.T. are employees of Bio-Rad Laboratories who make and supply
492 droplet digital PCR instruments and reagents. dPCR reagents, assays and consumables for
493 the inter-laboratory study were supplied by Bio-Rad Laboratories to the twenty-one end-
494 user laboratories. The production and characterisation of the study materials was
495 performed at LGC using separately purchased reagents, assays and consumables. B.U. is an
496 employee of The Dana Farber Cancer Institute that has a pending patent related to non-
497 invasively monitoring of genomic alterations in cancer for the reproducible measurement
498 of rare variants using droplet digital PCR.

499 **REFERENCES**

500 1. Higuchi R, Fockler C, Dollinger G, & Watson R (1993) Kinetic PCR analysis: real-time
501 monitoring of DNA amplification reactions. *Bio/technology (Nature Publishing*
502 *Company)* 11(9):1026-1030.
503 2. Wittwer CT, Herrmann MG, Moss AA, & Rasmussen RP (1997) Continuous
504 fluorescence monitoring of rapid cycle DNA amplification. *BioTechniques* 22(1):130-
505 131, 134-138.
506 3. Madej RM, *et al.* (2010) International Standards and Reference Materials for
507 Quantitative Molecular Infectious Disease Testing. *J Mol Diagn* 12(2):133-143.

4. Hughes T, *et al.* (2006) 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* 108(1):28-37.

5. JCGM (2008) Evaluation of measurement data – Guide to the expression of uncertainty in measurement (ISO/IEC Guide 98-3). *http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf*.

6. Bustin SA, *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55(4):611-622.

7. Fryer JF, *et al.* (2008) Development of working reference materials for clinical virology. *J Clin Virol* 43(4):367-371.

8. White RA, 3rd, Quake SR, & Curr K (2012) Digital PCR provides absolute quantitation of viral load for an occult RNA virus. *J Virol Methods* 179(1):45-50.

9. Whale AS, *et al.* (2012) Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res.* 40(11):e82.

10. Hayden RT, *et al.* (2013) Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. *J Clin Microbiol* 51(2):540-546.

11. Pinheiro LB, *et al.* (2012) Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem* 84(2):1003-1011.

12. Devonshire AS, *et al.* (2015) Highly reproducible absolute quantification of Mycobacterium tuberculosis complex by digital PCR. *Analytical chemistry* 87(7):3706-3713.

13. Hindson CM, *et al.* (2013) Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nature methods* 10(10):1003-1005.

14. Kelley K, Cosman A, Belgrader P, Chapman B, & Sullivan DC (2013) Detection of methicillin-resistant Staphylococcus aureus by a duplex droplet digital PCR assay. *J Clin Microbiol* 51(7):2033-2039.

15. Lench N, *et al.* (2013) The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made. *Prenatal diagnosis* 33(6):555-562.

16. Taly V, *et al.* (2013) Multiplex Picodroplet Digital PCR to Detect KRAS Mutations in Circulating DNA from the Plasma of Colorectal Cancer Patients. *Clin Chem* 59(12):1722-1731.

17. Beck J, *et al.* (2013) Digital Droplet PCR for Rapid Quantification of Donor DNA in the Circulation of Transplant Recipients as a Potential Universal Biomarker of Graft Injury. *Clin Chem* 59(12):1732-1741.

18. Hindson BJ, *et al.* (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* 83(22):8604-8610.

19. White RA, 3rd, Blainey PC, Fan HC, & Quake SR (2009) Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC genomics* 10:116.

20. Huggett JF, *et al.* (2013) The Digital MIQE Guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. *Clin Chem* 59(6):892-902.

21. White H, *et al.* (2015) A certified plasmid reference material for the standardisation of BCR-ABL1 mRNA quantification by real-time quantitative PCR. *Leukemia* 29(2):369-376.

553 22. Lih CJ, *et al.* (2016) Certified DNA Reference Materials to Compare HER2 Gene
554 Amplification Measurements Using Next-Generation Sequencing Methods. *J Mol Diagn*
555 online advanced publication.
556 23. Devonshire AS, *et al.* (2016) The use of digital PCR to improve the application of
557 quantitative molecular diagnostic methods for tuberculosis. *BMC Infectious Diseases*
558 16:366.
559 24. Lievre A, *et al.* (2008) KRAS mutations as an independent prognostic factor in patients
560 with advanced colorectal cancer treated with cetuximab. *Journal of clinical oncology :*
561 *official journal of the American Society of Clinical Oncology* 26(3):374-379.
562 25. Lievre A, *et al.* (2006) KRAS mutation status is predictive of response to cetuximab
563 therapy in colorectal cancer. *Cancer research* 66(8):3992-3995.
564 26. Jackman DM, *et al.* (2009) Impact of epidermal growth factor receptor and KRAS
565 mutations on clinical outcomes in previously untreated non-small cell lung cancer
566 patients: results of an online tumor registry of clinical trials. *Clinical cancer research :*
567 *an official journal of the American Association for Cancer Research* 15(16):5267-5273.
568 27. van Krieken JH, *et al.* (2008) KRAS mutation testing for predicting response to anti-
569 EGFR therapy for colorectal carcinoma: proposal for an European quality assurance
570 program. *Virchows Archiv : an international journal of pathology* 453(5):417-431.
571 28. Hocking CM & Price TJ (2014) Panitumumab in the management of patients with
572 KRAS wild-type metastatic colorectal cancer. *Therapeutic Advances in*
573 *Gastroenterology* 7(1):20-37.
574 29. Huggett JF, Cowen S, & Foy CA (2015) Considerations for Digital PCR as an Accurate
575 Molecular Diagnostic Tool. *Clin Chem* 61(1):79-88.

576 30. Roma C, *et al.* (2013) Detection of EGFR mutations by TaqMan mutation detection
577 assays powered by competitive allele-specific TaqMan PCR technology. *BioMed*
578 *research international* 2013:385087.
579 31. Giacona MB, *et al.* (1998) Cell-free DNA in human blood plasma: length measurements
580 in patients with pancreatic cancer and healthy controls. *Pancreas* 17(1):89-97.
581 32. Wang BG, *et al.* (2003) Increased plasma DNA integrity in cancer patients. *Cancer*
582 *research* 63(14):3966-3968.
583 33. Devonshire AS, *et al.* (2014) Towards standardisation of cell-free DNA measurement
584 in plasma: controls for extraction efficiency, fragment size bias and quantification.
585 *Anal Bioanal Chem* 406(26):6499-6512.
586 34. Heredia NJ, *et al.* (2013) Droplet Digital PCR quantitation of HER2 expression in FFPE
587 breast cancer samples. *Methods* 59(1):S20-23.
588 35. Alikian M, *et al.* (2016) Next-Generation Sequencing-Assisted DNA-Based Digital PCR
589 for a Personalized Approach to the Detection and Quantification of Residual Disease
590 in Chronic Myeloid Leukemia Patients. *J Mol Diagn* 18(2):176-189.
591 36. Corbisier P, *et al.* (2015) DNA copy number concentration measured by digital and
592 droplet digital quantitative PCR using certified reference materials. *Anal Bioanal*
593 *Chem* 407(7):1831-1840.
594 37. Lievens A, Jacchia S, Kagkli D, Savini C, & Querci M (2016) Measuring Digital PCR
595 Quality: Performance Parameters and Their Optimization. *PLoS One* 11(5):e0153317.
596 38. Whale AS, Huggett JF, & Tzonev S (2016) Fundamentals of multiplexing with digital
597 PCR. *Biomolecular Detection and Quantification* online advanced publication.

598 39. Zheng D, *et al.* (2016) Plasma EGFR T790M ctDNA status is associated with clinical
599 outcome in advanced NSCLC patients with acquired EGFR-TKI resistance. *Scientific*
600 *reports* 6:20913.
601 40. Oxnard GR, *et al.* (2014) Noninvasive detection of response and resistance in EGFR-
602 mutant lung cancer using quantitative next-generation genotyping of cell-free plasma
603 DNA. *Clinical cancer research : an official journal of the American Association for*
604 *Cancer Research* 20(6):1698-1705.
605 41. Sacher AG, Paweletz C, Dahlberg SE, & et al. (2016) Prospective validation of rapid
606 plasma genotyping for the detection of egfr and kras mutations in advanced lung
607 cancer. *JAMA oncology* 2(8):1014-1022.
608 42. Dhanasekaran S, Doherty TM, & Kenneth J (2010) Comparison of different standards
609 for real-time PCR-based absolute quantification. *Journal of immunological methods*
610 354(1-2):34-39.
611 43. Devonshire AS, *et al.* (2016) An international comparability study on quantification of
612 mRNA gene expression ratios: CCQM-P103.1. *Biomolecular Detection and*
613 *Quantification* 8:15-28.
614 44. Dunn OJ (1961) Multiple Comparisons Among Means. *Journal of the American*
615 *Statistical Association* 56(293):52-64.
616 45. Staudte RG & Sheather SJ ( 1990) Robust Estimation and Testing. *Robust Estimation*
617 *and Testing*, (John Wiley and Sons, New York).

618 **FIGURE LEGENDS**

**Figure 1. Inter-laboratory comparison.** The originally submitted values from all twenty-one laboratories are shown for Samples A-D in descending KRAS G12D copy number concentration from left to right. For each graph the anonymised laboratory number is shown in the x-axis with the (A) KRAS G12D or (C) wt copy number concentration (copies/µL in the reaction) shown on the y-axis as indicated. For the fractional abundance graphs (B), the %G12D is shown on the y-axis. For each participant, the submitted mean value is plotted as a short black horizontal line together with the 95% confidence interval based on triplicate measurements of three units (n=9). The red horizontal dashed line represents the median value across labs and the red horizontal dotted lines represent the upper and lower $MAD_E$ intervals with 99% confidence. For Sample B, the lower confidence interval is not shown as it is approximately zero. Asterisks (*) just above the x-axis indicate laboratories that reported either a zero value or values below the range of the y-axis. All graphs show two orders of magnitude, shown on the $\log_{10}$ scale, though the range varies according to the sample.

**Figure 2. Quantification of *KRAS* G12D following threshold setting guidelines.** For each sample the comparison between the copy number concentrations originally submitted (black bar) with the resubmitted values (orange bar) and the coordinator analysis (teal bar) is shown. The red horizontal dashed line represents the median from the *original* data set with the associated red horizontal dotted lines representing the upper and lower $MAD_E$ intervals with 99% confidence. The scales are the same as those in Figure 1. Four of the laboratories reanalysed their data (2, 10, 17 and 21) and the remaining three laboratories were satisfied with their original submission (3, 7 and 16) and so do not contain data points for resubmission. The coordinating laboratory analysed the data sets from the seven

642 laboratories. The dashed ovals highlight laboratory 17 that were still under-quantifying the
643 G12D concentration in the four samples compared with the median; while the analysis by
644 the coordinating laboratory did not.

## Table 1. Values for the *KRAS* target materials obtained during characterisation and inter-laboratory comparison.

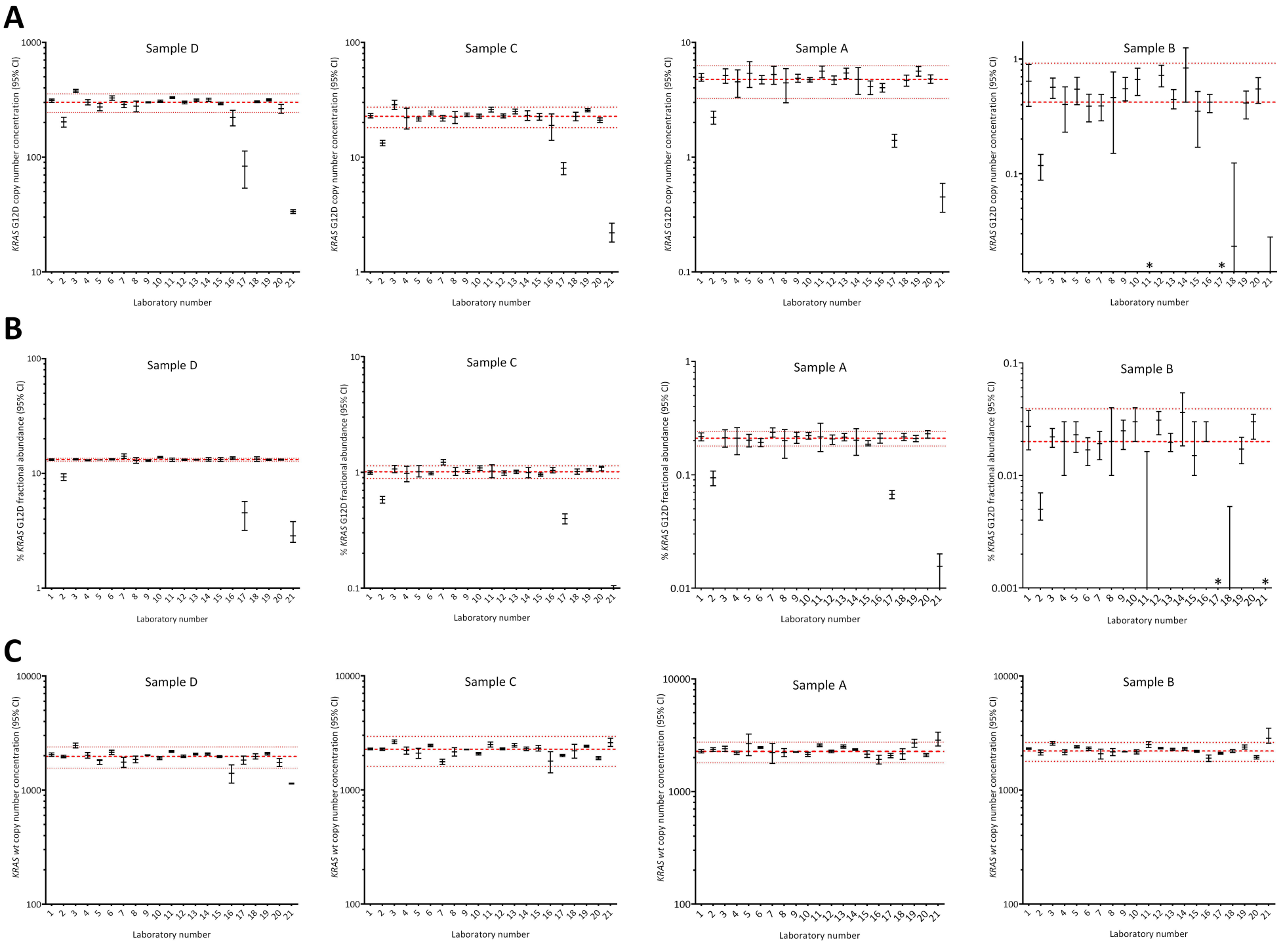
| Material | | Characterisation | | | Inter-laboratory participant results[#] | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Name | Target | Nominal* | Mean* | $U_{rel}$ | Median* | % difference[$] | $MAD_E$ | Lower 99% CI | Upper 99% CI |
| | [WT] | 2000 | 2400 | 1.0% | 2267.00 | 6% | 182.85 | 1795.24 | 2738.76 |
| A | [G12D] | 3 | 4.9 | 5.9% | 4.74 | 3% | 0.59 | 3.23 | 6.26 |
| | % G12D | 0.17 | 0.20 | 6.0% | 0.21 | 5% | 0.01 | 0.18 | 0.24 |
| | [WT] | 2000 | 2300 | 1.7% | 2197.00 | 4% | 162.31 | 1781.50 | 2612.50 |
| B | [G12D] | 0 | 0.58 | 16% | 0.42 | 28% | 0.19 | -0.08 | 0.92 |
| | % G12D | 0.00 | 0.025 | 16% | 0.02 | 20% | 0.01 | 0.00 | 0.04 |
| | [WT] | 2000 | 2300 | 1.4% | 2273.78 | 1% | 257.71 | 1608.88 | 2938.68 |
| C | [G12D] | 18 | 22 | 4.5% | 22.67 | 3% | 1.78 | 18.08 | 27.26 |
| | % G12D | 0.90 | 0.96 | 3.5% | 1.01 | 6% | 0.05 | 0.89 | 1.14 |
| | [WT] | 1760 | 2100 | 1.5% | 1975.00 | 6% | 161.98 | 1557.10 | 2392.90 |
| D | [G12D] | 240 | 310 | 1.6% | 300.80 | 3% | 21.55 | 245.19 | 356.41 |
| | % G12D | 12.00 | 13.00 | 0.92% | 13.16 | 1% | 0.15 | 12.79 | 13.54 |
| | [WT] | 2000 | 2300 | 1.7% | 2264.67 | 2% | 217.51 | 1707.85 | 2821.48 |
| NEG | [G12D] | 0 | 0.58 | 16% | 0.49 | 16% | 0.14 | 0.12 | 0.85 |
| | % G12D | 0.00 | 0.025 | 16% | 0.02 | 19% | 0.01 | -0.01 | 0.05 |
| | [WT] | 1760 | 2100 | 1.5% | 1990.00 | 5% | 139.90 | 1629.07 | 2350.93 |
| POS | [G12D] | 240 | 310 | 1.6% | 294.67 | 5% | 23.23 | 234.72 | 354.61 |
| | % G12D | 12.00 | 13.00 | 0.92% | 13.10 | 1% | 0.32 | 12.27 | 13.93 |

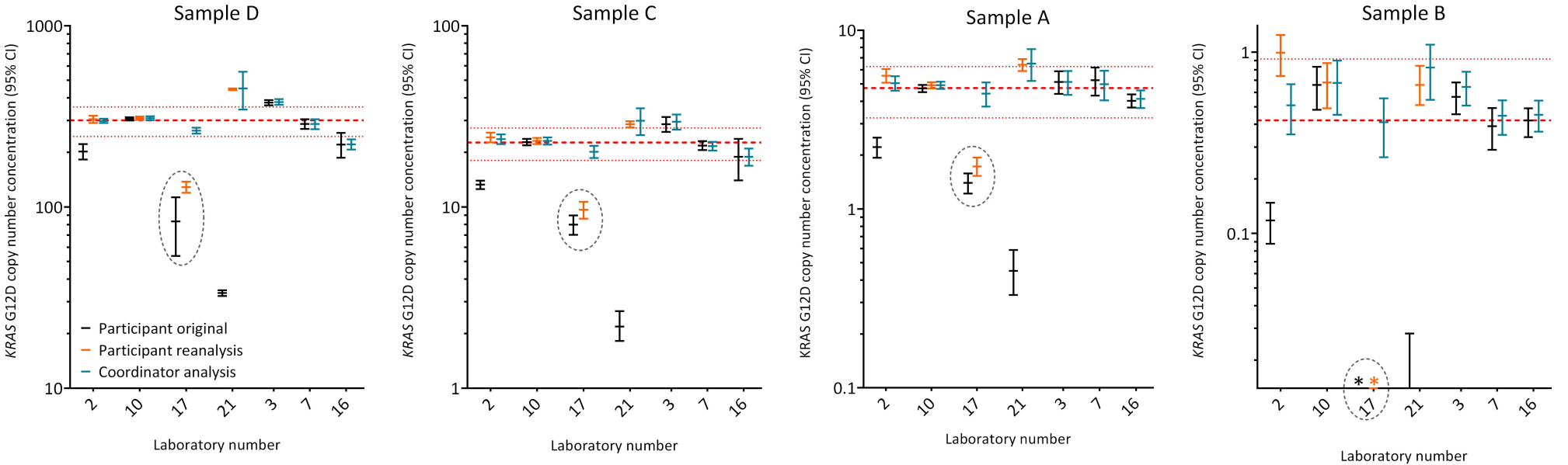
In order to remain ambiguous to the participants, the sample name (A-D) was randomly assigned to each of the four materials so that there was no correlation between sample name and G12D fractional abundance. The two control materials, NEG and POS, were generated from the sample materials as Sample D and Sample B, respectively and so have identical characterisation values. *[G12D] and [WT] are given as copies/µL in the reaction based on following the study protocol where 8.8 µL of each study material is added to 22 µL reaction during set up; this is the metric participants were asked to report. The nominal concentrations are the approximate values the materials were prepared at. For characterisation of the materials, the mean concentration and the relative expanded uncertainty ($U_{rel}$), that is the equivalent of the 95% confidence interval reported as a percentage of the mean, were calculated from the homogeneity study with triplicate measurements (within-unit) from six units (between-units). All values are given to 2 significant figures. [#]For the inter-laboratory participant results, the values are calculated from the twenty-one original submitted values. [$]The percentage difference between the mean concentration obtained from the homogeneity study with the participant median value. The $MAD_E$ is the median absolute deviation.

## Table 2. Precision of the inter-laboratory comparison following resubmission.

| Material Name | Target | Recalculated Median* | % difference$ | CV between labs | tCV |
|---|---|---|---|---|---|
| A | [WT] | 2267.00 | 6% | 8.07% | 1% |
| | [G12D] | 4.87 | 1% | 11.39% | 10% |
| | % G12D | 0.21 | 5% | 4.22% | |
| B | [WT] | 2197.00 | 4% | 7.39% | 1% |
| | [G12D] | 0.46 | 21% | 34.50% | 35% |
| | % G12D | 0.02 | 20% | 37.08% | |
| C | [WT] | 2274.00 | 1% | 11.35% | 1% |
| | [G12D] | 22.98 | 4% | 7.87% | 5% |
| | % G12D | 1.02 | 6% | 4.74% | |
| D | [WT] | 1982.00 | 6% | 7.65% | 1% |
| | [G12D] | 303.30 | 2% | 5.88% | 1% |
| | % G12D | 13.20 | 2% | 1.12% | |
| NEG | [WT] | 2264.67 | 2% | 9.60% | 1% |
| | [G12D] | 0.54 | 7% | 24.72% | 32% |
| | % G12D | 0.02 | 20% | 31.26% | |
| POS | [WT] | 1990.00 | 5% | 6.33% | 1% |
| | [G12D] | 298.00 | 4% | 7.30% | 1% |
| | % G12D | 13.24 | 2% | 2.69% | |

*[G12D] and [WT] are given as copies/µL in the reaction as described in Table 1. The values presented in this table are calculated from the four resubmitted values and remaining seventeen original values. $The percentage difference between the mean concentration obtained from the homogeneity study (Table 1) with the recalculated median. The tCV is calculated based on the equation described in Devonshire et al. 2015.


A
Sample D
Sample C
Sample A
Sample B
KRAS G12D copy number concentration (95% CI)
Laboratory number
B
Sample D
Sample C
Sample A
Sample B
% KRAS G12D fractional abundance (95% CI)
Laboratory number
C
Sample D
Sample C
Sample A
Sample B
KRAS wt copy number concentration (95% CI)
Laboratory number


Sample D
Sample C
Sample A
Sample B
KRAS G12D copy number concentration (95% CI)
Laboratory number
Participant original
Participant reanalysis
Coordinator analysis
2
10
17
21
3
7
16