CCQM P123 Pilot Study

Number and geometric property of cells adhered to a solid substrate

Final report

Coordinating laboratories: Istituto Nazionale di Ricerca Metrologica (INRIM); National Measurement Laboratory at LGC (LGC); National Institute of Standard and Technology (NIST)

Contact
Carla Divieto, INRIM, Italy
c.divieto@inrim.it

Authors
Laura Revel[1], Stefano Pavarelli[1], Carla Divieto[1], Francesca Pennecchi[1], Massimo Zucco[1], Stephen L.R. Ellison[2], Maria Paola Sassi[1]

Affiliations
[1]Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Torino (Italy)
[2]National Measurement Laboratory at LGC, Queens Road, Teddington, Middlesex, TW11 0LY (UK)

Participants
INMETRO - Instituto Nacional de Metrologia, Qualidade e Tecnologia
INRIM - Istituto Nazionale di Ricerca Metrologica
KRISS - Korea Research Institute of Standards and Science
LGC - National Measurement Laboratory (NML) at LGC
NIM – National Institute of Metrology P. R. China
NIMT - National Institute of Metrology Thailand
NIST - National Institute of Standards and Technology
PTB - Physikalisch-Technische Bundesanstalt
VNIIM - D.I. Mendeleyev Institute for Metrology, Rosstandart

Summary

The CCQM-P123 study was set up within the Cell Analysis Working Group to demonstrate capability of participants in the quantification of 4 measurands using one reference material: number of adherent cells on a planar surface, cell density, cell confluency and average cell area. These made the study intrinsically ambitious. Nine Institutions participated in the study. Metrology for cells provides confidence in both, scientific and applicative researches. Several scientific fields, such as cytology, biochemistry, molecular biology and molecular genetics have cells as their direct topic of study. The cell is a fundamental unit of highly technological systems, employed, for example, in the development of new chemicals and drugs, in the evaluation of compound toxicity, in novel approaches to health, such as regenerative medicine. High content screening technologies are cellular imaging-based assays. In this framework, the cell density, the cell confluency and the cell size are measurands that can describe a biological effect of a molecule under examination. Reference materials for three levels of cell density and cell confluency were prepared for fluorescence microscopy and circulated among participants. Each reference material was a commercial imaging dish on which human cells were seeded, fixed and stained for nucleus and for whole cell. Each dish had a glass bottom divided in 400 squares; 4 squares for each dish has been selected for the analysis and these 12 squares are the measurement standards of the CCQM-P123 pilot study. In each square a defined area was selected as area of measurement. The use of a calibrated ruler to define the area of measurement assured the traceability to the International System of Units (SI). Results for cell number, cell density, cell confluency and average cell area showed an overall good agreement between the laboratories. In addition, the study allowed several considerations, fundamental in the frame of a cell-based measurement comparison, in terms of: traceability, comparability of results and complexity of the study protocol. The major strength of the pilot study was the possibility to compare four measurement capabilities by using a single reference material. In conclusion, the study has been a preparatory exercise for a key comparison on at least one of the 3 measurements claims (cell density, cell confluency and cell area) related to adhered cells features such as cell proliferation, morphology and size. The results from such a key comparison could be utilized by the NMIs to provide traceability to different methods of bio-analysis.

Contents

CCQM P123 Number and geometric property of cells adhered to a solid substrate ..... p. 1

Coordinating laboratories ..... p. 1

Contact ..... p. 1

Authors ..... p. 1

Affiliations ..... p. 1

Participants ..... p. 1

Summary ..... p. 2

Contents ..... p. 3

List of figures ..... p. 4

List of tables ..... p. 5

1. Introduction and Study Rationale ..... p. 6
2. Measurement claim ..... p. 7
3. Participation in CCQM-P123 ..... p. 7
4. Materials and Methods ..... p. 8

4.1 Reference Material and Circulation ..... p. 8

4.2 Measurement protocol ..... p. 10

4.3 P123 measurement standard – squares to be acquired ..... p. 12

5. CCQM-P123 Study Results ..... p. 13

5.1 Uncertainty analysis ..... p. 14

5.2 Cell number per square ..... p. 15

5.3 Cell density ..... p. 17

5.4 Cell confluency ..... p. 20

5.5 Average cell area ..... p. 23

6. Discussion ..... p. 27

6.1 Pilot study reference value ..... p. 27

7. Conclusion ..... p. 30
8. Acknowledgements ..... p. 30
9. References ..... p. 30

Appendix 1: Reference material preparation, homogeneity and stability ..... p. 32

Appendix 2: Study Protocol ..... p. 41

List of figures

Fig. 1.	Imaging dish (A and B) and example of measurement standard (C) of the CCQM-P123 pilot study.	p. 8
Fig. 2.	Dual Axis Linear Scale Micrometer (Edmund Optics) namely “ruler”.	p. 9
Fig. 3.	Scheme of the standards circulation.	p. 9
Fig. 4.	XY ruler, traceable to the SI.	p. 11
Fig. 5.	Example of acquired images in phase contrast.	p. 11
Fig. 6.	Example of acquired and processed image to visualize only the area of measurement.	p. 12
Fig. 7.	Cell number by target and laboratory.	p. 16
Fig. 8.	Mandel’s h for cell number.	p. 16
Fig. 9.	Cell density by target and laboratory.	p. 18
Fig. 10.	Mandel’s h for cell density.	p. 18
Fig 11.	MSD for cell density.	p. 19
Fig. 12.	Confluency by target and laboratory.	p. 21
Fig. 13.	Mandel’s h for confluency.	p. 21
Fig. 14.	MSD for confluency.	p. 22
Fig. 15.	Cell area by target and laboratory.	p. 24
Fig. 16.	Mandel’s h for average cell area.	p. 24
Fig. 17.	MSD for average cell area.	p. 25
Fig. 18.	Plot of average cell area.	p. 25
Fig. 19.	Box plot of the 4 measurands grouped by levels.	p. 26

List of tables

Table 1.	CCQM-P123 participants and instruments.	p. 7
Table 2.	Scheme of the squares (sq.) of set 1 and set 2 to be acquired.	p. 12
Table 3.	Uncertainty source considered by participants.	p. 14
Table 4.	Participant's results on Cell number.	p. 15
Table 5.	Participant's results on Cell density.	p. 17
Table 6.	Participant's results on Cell confluency.	p. 20
Table 7.	Participant's results on Average cell area.	p. 23
Table 8.	Results of Cox's procedure A and B for cell number.	p. 28
Table 9.	Results of Cox's procedure A and B for cell density.	p. 28
Table 10.	Results of Cox's procedure A and B for confluency.	p. 29
Table 11.	Results of Cox's procedure A and B for average cell area.	p. 29

1. Introduction and Study Rationale

The study was set up under the auspices of the Consultative Committee for Amount of Substance (CCQM) – Cell Analysis Working Group (CAWG) to demonstrate capability of participants in the quantification of cell number per area (cell density), area occupied by cells (cell confluency) and average cell area on a planar surface and to identify sources of measurement uncertainty. Cells are intended to be adherent in monolayer to a solid bidimentional (2D) substrate.

Nine NMIs, from the CAWG, participated in the CCQM-P123 study. CAWG mission is to "*identify, establish and underpin global comparability of cell measurement capabilities through reference measurement systems of the highest possible metrological order with traceability to the SI or to other internationally agreed units. The group benchmarks claimed competences of the National Measurement Institutes for measurement services in the quantification of intact cells and cell properties*" as stated in [1].

Metrology for cells is a horizontal underpinning activity providing confidence in both, scientific and applicative researches [1]. The cell is the basic structural, functional and biological unit of all known living organisms. Cells are the smallest unit of life classified as a living thing, often called the "building blocks of life"[2]. Several scientific fields, such as cytology, biochemistry, molecular biology and molecular genetics have cells as their direct topic of study. Moreover, the cell is also a fundamental unit of highly technological systems, for the development of new chemicals and drugs, the evaluation of the toxicity of active molecules, the building of novel approaches to health, such as regenerative medicine [3, 4]. In these systems, the cells cease being topic of study, and become tools [5]. In this framework, counting cells at different time points allows following the growth of a cell culture (proliferation) and to evaluate the effect of specific substances (drugs, toxics) that may interfere with cell behaviour. The number of cells per area and the fraction of the area occupied by cells, represent measurands that can describe a biological effect of a molecule under examination.

The development of high content screening technologies including automated immunostaining, automated image acquisition and automated image analysis have enabled higher throughput of cellular imaging-based assays [6]. These cell-based assays allow quick evaluation of the effects of thousands of molecules and provide starting points for drug design and for understanding the interaction or role of particular biochemical processes [7].

The pharmaceutical industries have declared their need of metrology in the cellular field, in order to improve reliability and comparability of results, reduce the number of the tests, narrow the time from discovery to the marketing of new drugs and limit costs [8].

The study of cell morphology and the biological study of the form and structure of living cells is useful in determining behaviour and function of cells. Cell morphology is essential in identifying the shape, structure, and size of cells. It generally entails microscopy to identify the morphological features of cell under study. Morphological parameters of a specified cell line, typically dimensions and shapes of cells in defined conditions are essential for successful cell culture experiments and to confirming the healthy status of cells.

Signs of non-healthy conditions of the cells include granularity around the nucleus, detachment of the cells from the substrate, cytoplasmic vacuolation and modification in dimension and/or shape of the cells. These conditions may be caused by a variety of reasons, including contamination of the cell culture, senescence of the cells, presence/accumulation of toxic substances in the cell growth medium. Biological tests made with non-healthy cells produce non-reliable results. One of the majoir need remains the validation of cell counting measurements and the development of new or improved biological and non-biological reference materials [9].

The protocol of the present study applies to cell measurements based on fluorescence microscopy.

2. Measurement claims

The measurement claims for the CCQM-P123 study are:

1. number of cells in monolayer adhesion on a defined area[1], in the range 0 to 5000 cell/mm^2, defined as cell density and calculated as:

$$Cell\ density = \frac{number\ of\ cells}{area} \quad [\text{unit count value}/\text{mm}^2]$$

2. confluency fraction of cells in monolayer adhesion on a defined area[1], in the range 0 % to 100 %, defined as cell confluency and calculated as:

$$Cell\ confluency = \frac{area\ occupied\ by\ cells}{area} \times 100 \quad [\text{mm}^2/\text{mm}^2 \quad \%]$$

3. average cell area of cells in monolayer adhesion on a defined area[1], derived from the previous measurements and calculated as:

$$Average\ Cell\ Area = \frac{area\ occupied\ by\ cells}{number\ of\ cells} \quad [\mu\text{m}^2]$$

[1] the defined area was intended on a planar (2D) surface. The area of measurement defined for the comparison was S = (0.2098 ± 0.0006) mm^2 and its definition was described in the protocol and reported within this document in Appendices 1 and 2.

In Appendix 1 section 5 the selection of measurands ranges has been addressed.

3. Participation in CCQM-P123

The nine organizations participating in the study are reported in Table 1, also listing the microscopes, cameras and imaging software used. One Institution was dropped out of the study because of an issue of sample retrieving through the customs.

Institute	Country	Instrument-Manufacturer
INMETRO - Instituto Nacional de Metrologia, Qualidade e Tecnologia	Brazil	Microscope: Zeiss Axio Observer.D1 Camera: Zeiss Axiocam MRc5 Software: ImageJ
INRIM - National Institute of Metrological Research	Italy	Microscope: Zeiss AxioObserver Z1 Camera: Olympus Monochrome XM10 Software: ImageJ
KRISS - Korea Research Institute of Standards and Science	Rep. of Korea	Microscope: Olympus IX 71 Camera: Olympus DP70 Software: ImageJ
LGC - National Measurement Laboratory (NML) at LGC	UK	Microscope: Zeiss PALM CombiSystem Camera: Zeiss Axiocam MRM Software: ImageJ
NIM – National Institute of Metrology P. R. China	P. R. China	Microscope: Zeiss Observer Z1 Camera: Zeiss AxioCam MRm r3.0 Software: ImageJ
NIMT - National Institute of Metrology Thailand	Thailand	Microscope: Nikon Eclipse TI Camera: NIkon MC 100 Software: Image J
NIST - National Institute of Standards and Technology	USA	Microscope: Zeiss Axiovert 200M Camera: Photometrics CoolSnap HQ Software: ImageJ
PTB - Physikalisch-Technische Bundesanstalt	Germany	Microscope: Zeiss AxioObserver Z1 Camera: Zeiss AxioCam MRm Software: ImageJ
VNIIM - D.I. Mendeleyev Institute for Metrology, Rosstandart	Russia	Microscope: Zeiss Axiovert 200M-DFC420 Camera: Leica 420C Software: manual / ImageJ

Table 1. CCQM-P123 participants and instruments

4. Materials and Methods

4.1 Reference Material preparation and circulation

In this pilot study three *reference materials* (defined according to [10]) were prepared for three levels of cell density and cell confluency “low”, “medium”, “high” and within each reference material, 4 *measurement standards* (defined according to [10]) for cell density and cell confluency were identified. The 4 measurement standards for each of the three reference materials had to be measured.

INRIM, NIST and NML at LGC defined the suitable reference materials characteristics and a suitable measuring method based on fluorescence microscopy, INRIM developed the study protocol, prepared and tested the reference materials for stability and homogeneity, coordinated the reference materials circulation among participants, collected results from the participants, INRIM and LGC analysed the results, INRIM and LGC wrote the final report.

Each **reference material** was a commercial **imaging dish** on which cells (human cell line A549) are seeded, fixed and stained for nucleus and for whole cell (see fig.1). Each dish has a 21 mm diameter glass bottom and an imprinted grid divided in 400 squares. Each square has nominal dimension of 0.5 mm x 0.5 mm and is identified by letter and number; 4 squares for each dish has been selected for the analysis and these **12 squares are the measurement standards of the CCQM-P123 pilot study**.

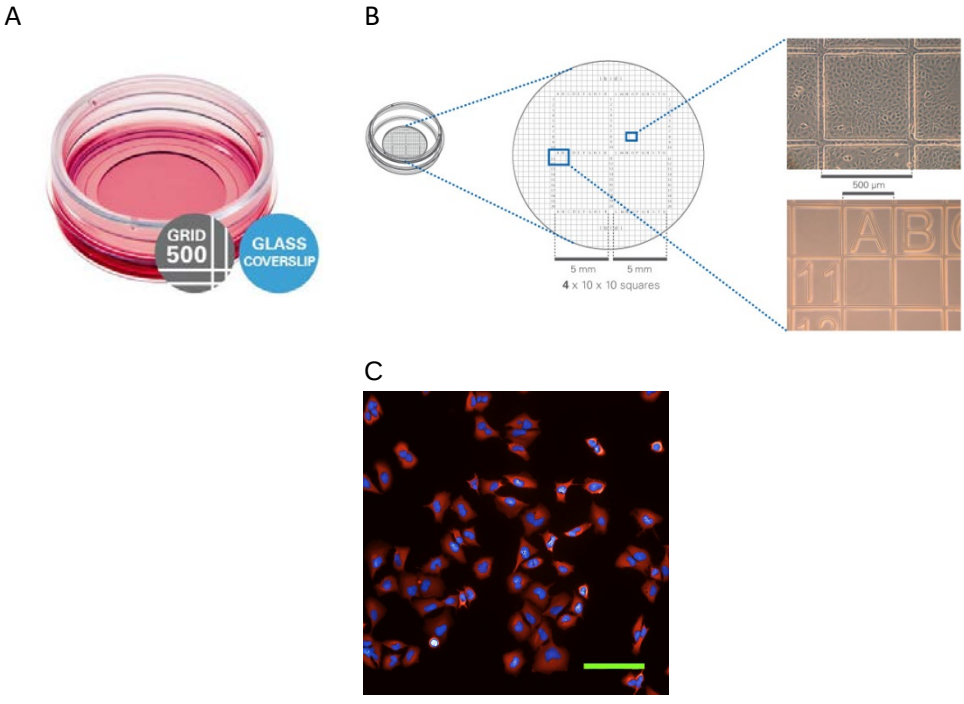


Fig. 1. Imaging dish (A and B) and example of measurement standard (C) of the CCQM-P123 pilot study.
Images in A and B is taken from https://www.ibidi.com/

The A549 cell line (human alveolar basal epithelial cells) was purchased from ATCC (reference number CCL-185); they have been seeded at three cell density nominal values, $1x10^5$, $2x10^5$ and $4x10^5$ cells per dish, in order to cover a wide range of cell density and confluency.

Two dyes have been used to stain the nucleus and the whole cell: DAPI (Invitrogen, cod. D1306) has been selected to stain the cell nucleus and Texas Red® c2 maleimide (Invitrogen, cod. T6008) has been selected to stain intracellular proteins.

The cell seeding was optimized to obtain the best possible homogeneity of cell distribution among the squares. The confluency variability among the squares was 3 % for low confluency, 13 % for medium confluency and 3 % for high confluency.

Additional technical details and protocols used for preparation and stability and homogeneity tests of the reference materials are reported in **Appendix 1**.

To have SI traceable measurements the use of a xy traceable ruler was adopted to define the area of measurement and crop this area. In addition, the use of a ruler allowed to minimize the differences among squares in terms of area of measurement.
At these aims, participants were asked to define the area of measurement within the squares by using the xy calibrated ruler. This was a dual axis linear stage micrometer (Edmund Optics Ltd #58-608) with a xy scale engraved with a grating pitch of 25 µm and was traceable to the INRIM length primary standard.

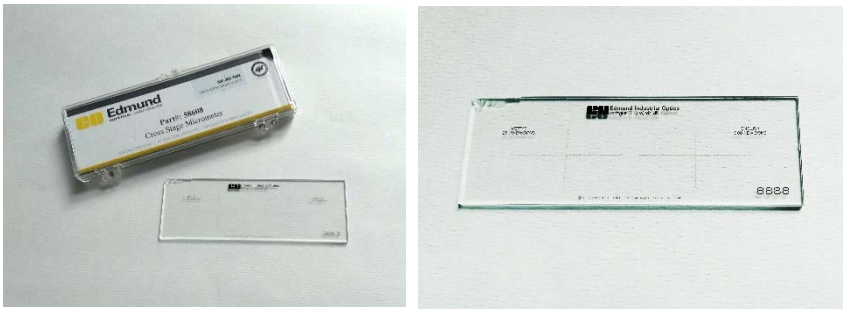


Fig. 2. Dual Axis Linear Scale Micrometer (Edmund Optics) namely "ruler".

The **area of measurement** defined for the comparison was S = (0.2098 ± 0.0006) mm^2

See **Appendix 2** for further details on the definition of the area of measurement.

The comparison was carried out by circulating a set of three reference materials (dishes) among the participating laboratories and by measuring measurement standards (squares) for cell density and cell confluency.

Circulations:

The comparison was organized in two circulations. Each circulation used a specific set of reference materials, respectively "set1" and "set2", as represented in figure 1. Each of the two sets were composed of three reference materials, representing the three levels of cell density and cell confluency.
The 12 measurement standards were measured at the beginning of the circulation by INRIM and sent to the second participant of each loop. Both sets of standards were returned to INRIM and checked at the end of the circulations for fluorescence intensity stability of the dyes.

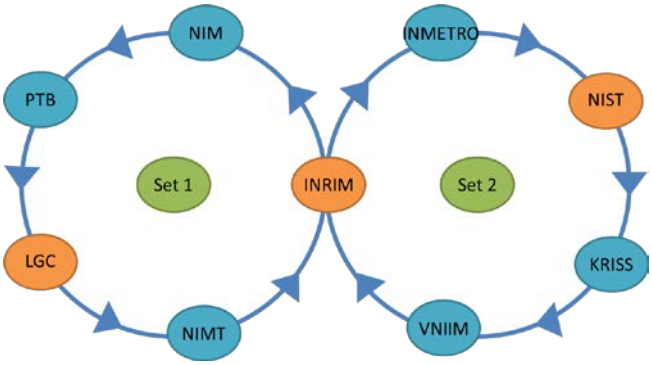


Fig. 3. Scheme of the standards circulation.

Note: set 1 was damaged after five shipments and set 2 was introduced to complete the comparison.
Set 2 high and medium level also were damaged after 4 shipments. VNIIM could only analyse the low level.

Participants received a package contained 3 reference materials, a xy calibrated ruler and a fluorescent glass filter. Each participant received before the comparison, and together with the protocol, a "test dish" for setting the microscope and exercising.

For "set1" the three reference materials were named:
- P123 set 1 LOW CONFLUENCY (P123_s1_L)
- P123 set 1 MEDIUM CONFLUENCY (P123_s1_M)
- P123 set 1 HIGH CONFLUENCY (P123_s1_H)

For "set2" the three reference materials were named:
- P123 set 2 LOW CONFLUENCY (P123_s2_L)
- P123 set 2 MEDIUM CONFLUENCY (P123_s2_M)
- P123 set 2 HIGH CONFLUENCY (P123_s2_H)

4.2 Measurement protocol

The measurement protocol (see **Appendix 2**) required to each laboratory:
- to set the microscope up according to the given instructions,
- to acquire the xy ruler image and design the area of measurement (fig. 3),
- to acquire 3 images setting the appropriate focus (phase contrast, DAPI and Texas Red), of each of the 12 selected measurement standards (4 measurement standards for each of the 3 reference materials) (see example in fig. 4 and 5),
- to process all the images according to the given instructions, in order to cut the area of measurement (see example in fig. 4) on each image (i.e. overlap the three acquired images - phase, DAPI, Texas Red, center the ruler image on the overlapped image, cut the defined area on the four-layers image and save the cut images as non-compressed tiff file),
- to measure the cell number per area for each measurement standard,
- to measure the cell confluency for each measurement standard,
- to measure the average area of the cells, derived from the model: average area of the cells = cell confluency x area of measurement / number of cells,
- to provide an uncertainty budget for the three measurements.

The protocol assumed the use of wide field fluorescence microscopy with the following set up:
- phase contrast or brightfield mode (to acquire both the square edges and the xy ruler to determine the area of measurement);
- filters for DAPI (Ex=358 nm, Em=461 nm) and Texas Red® c2 maleimide (Ex=595 nm, Em=615 nm);
- 10x magnification objective;
- digital camera (monochromatic camera was preferred).

Participants used the "square TEST" to set the exposure times.
The pilot study data were extracted from images manually and/or with algorithms for image analysis software. As example, a procedure based on an open source software was described in the protocol.

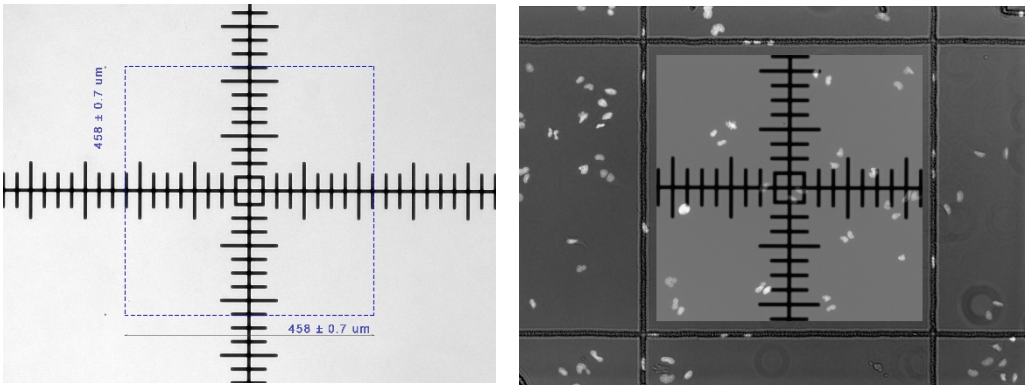


Fig. 4. XY ruler, traceable to the SI. In the image on the left, dashed lines indicate the ruler pitches to be used to draw the area of measurement on the square, as shown in the image on the right.

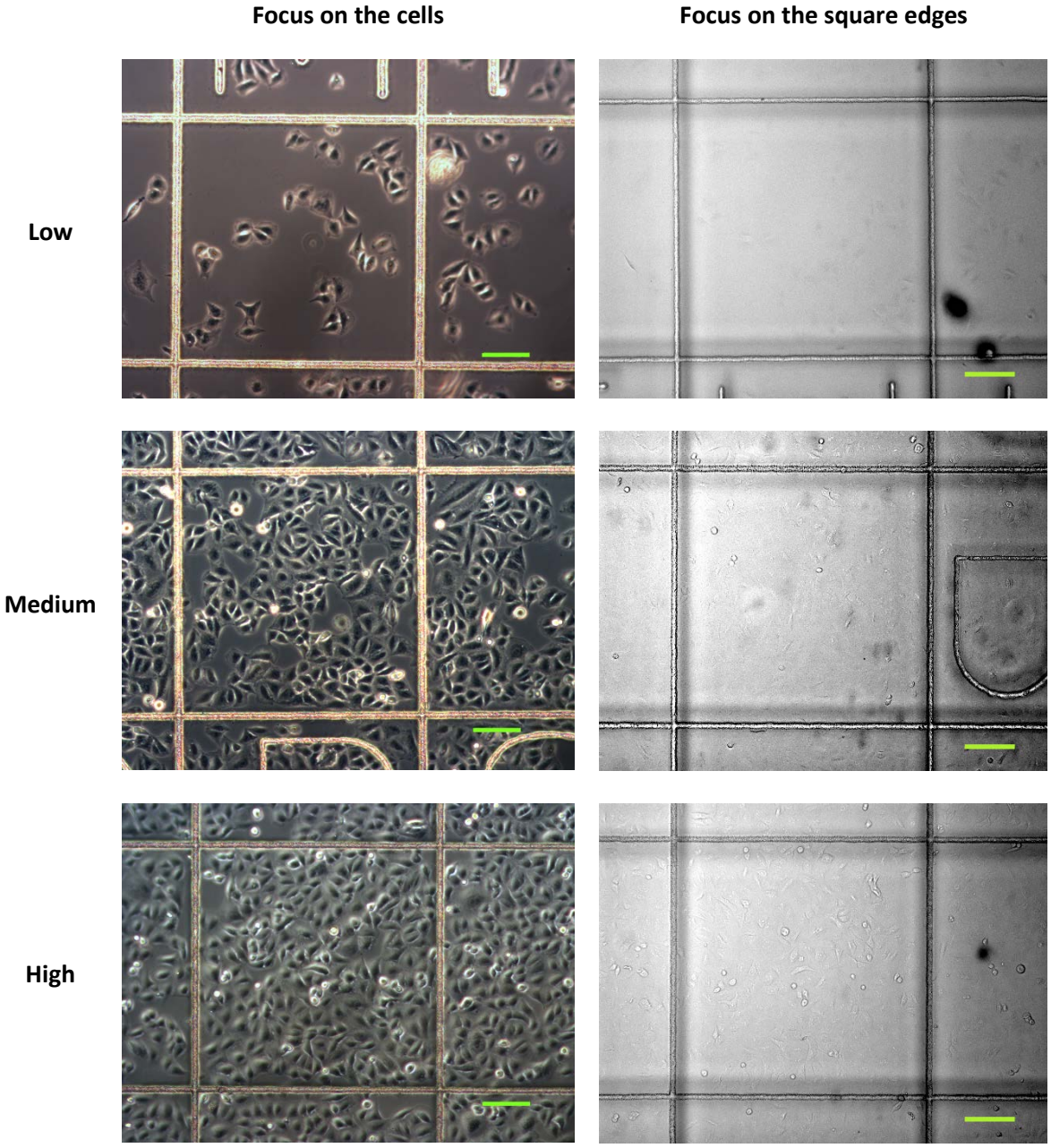


Fig. 5. Example of acquired images in phase contrast. Focus on the cells to select the level (low, medium and high). Focus on the square edges to process the image and cut the area of measurement. Scale barr = 100 μm.

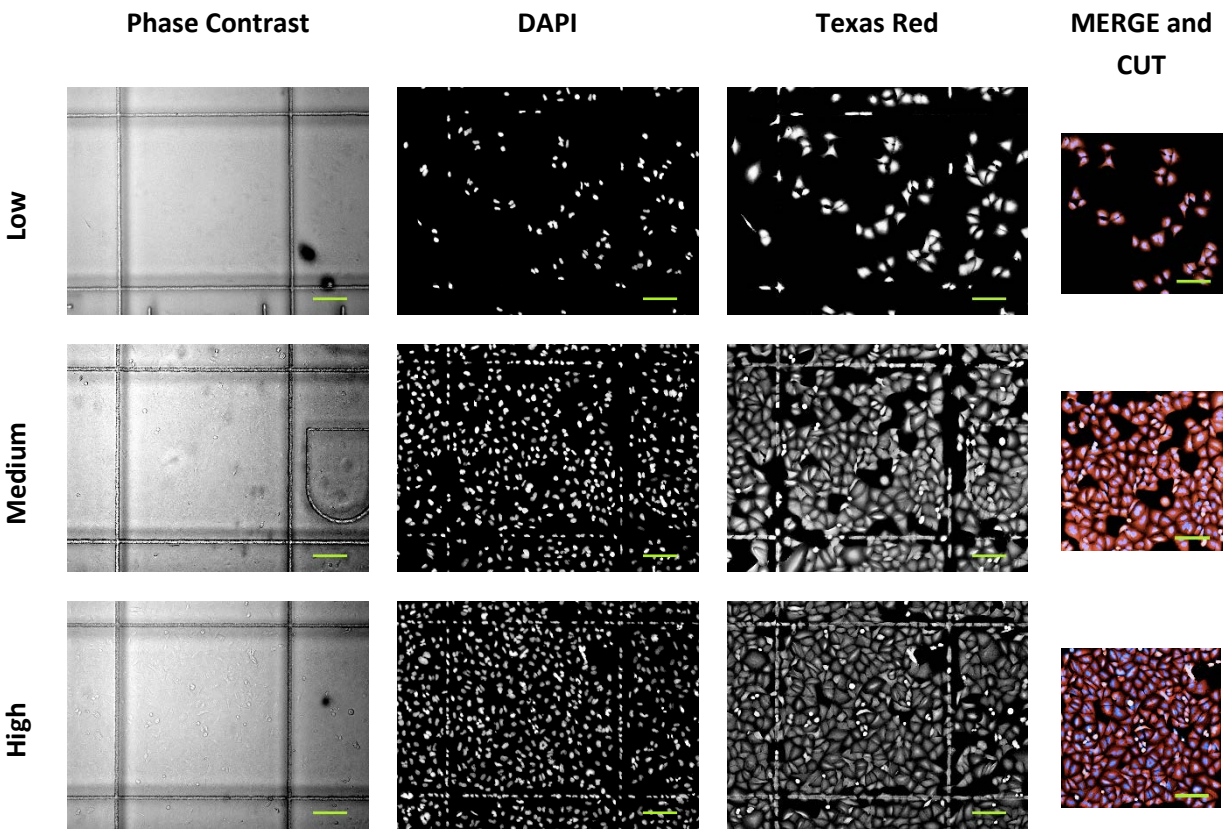


Fig. 6. Example of acquired and processed image to visualize only the area of measurement: acquired images (phase contract, DAPI and Texas Red), overlapped and cut image. Scale barr = 100 μm.

4.3 P123 measurement standards – squares to be acquired

Several dishes were prepared, among whom to select the best possible reference material.
The selection of the P123 reference materials and measurement standards responded to several criteria, described in Appendix 1 – section 5. Briefly, for each reference material four squares were selected as measurement standards, one for each quadrant, for a total of **12 P123 measurement standards**. One square was selected as test square.
Each P123 measurement standards were to be acquired for Texas Red, DAPI and phase contrast.

P123	id.	sq.1	sq.2	sq.3	sq.4	sq. TEST
SET 1	s1_L	A1	M1	K12	U13	A20
	s1_M	I8	R10	K15	U14	A1
	s1_H	D9	R10	F16	O17	L1
SET 2	s2_L	U5	Q11	K15	G10	A10
	s2_M	K10	P8	D15	L19	G2
	s2_H	H2	P3	I16	U20	A10

Table 2. Scheme of the squares (sq.) of set 1 and set 2 to be acquired.

See **Appendix 1** - fig. 8 for a graphical representation of the P123 reference materials and measurement standards selected for set 1 and set 2.

5. CCQM-P123 Study Results

The nine institutes performed the measurements using Inverted Phase Contrast Fluorescence Microscopes, high resolution monochrome or color digital cameras and ImageJ software for image analysis. Microscopes, cameras and software used are listed in Table 1.
The results of the study were expressed as:
- cell number: the number of cells (in unit count value) was given by the nuclei counting in the area of measurement defined by the use of the xy ruler;
- cell density: the number of cells in monolayer adhesion on the defined area (in unit count value/mm^2); it is derived from "cell number", traceable to the SI;
- cell confluency: the percentage of area occupied by the cells on the total area of measurement (in mm^2/mm^2 %); it is obtained by image analysis software;
- average cell area: derived from the previous measurements (in μm^2).

To note that each measurement standard (square) was considered as an independent measurement object. The study generated 24 indipendent measurement standards.

Uncertainty sources considered by partcipants are listed in table 3.
In the figures, the results are coded by institute (letters A-L) and an additional number indicates that two methods for image analysis were used. Results submitted by participants are provided in Tables 4 to 7 below.

Participants were asked to send all the acquired images, to fill and to send back report forms related to:
- method employed for images acquisition (Annex 3),
- cell number per area and uncertainty evaluation (Annex 4),
- confluency and uncertainty evaluation (Annex 5),
- mean area of cells and uncertainty evaluation (Annex 6).

However, 2 Institutions (D, G) sent only data, without information on the methods employed, the measurands estimation and the uncertainty evaluation.
Five Institutions (B, E, F, G, H, I) acquired one wrong square (a different one in each lab and close to the right one) hence the data of the right square are missing; four Institutions (B, D, G, I) did not send the processed images, hence the correctness of the area of measurement could not be assessed; two Institutions (C, E) did not follow properly the instructions for processing the images as illustrated in the study protocol and made an overestimation of the area of measurement; three Institutions (A, F, H) processed correctly the images defining the correct area of measurement and following the instructions on the protocol.
One Institution (I) did not report the values in Annex 4 as "number of cells per area" but as "cells per mm^2", justifying the choice: the measurand "cells per mm^2"could have a wider range of use than "cell number per area", being independent from the cell culture vessel. Hence, results on "cell number per area" reported in Table 4 do not include any data from Institution I.

The participant data are displayed in figures 7 to 17.
Due to the use of two sets of reference materials, all data sets are relatively small.
The data sets were initially examined by means of **dot plots** by target (i.e square) and laboratory (Figure 7, 9, 12, 15). The dot plots were obtained in R version 3.2.1 [11]: they show the measurements performed by the laboratories in each square with the associated expanded uncertainty ($k = 2$). **Plots of Mandel's *h*** and **median scaled differences** (Figure 8, 10, 13, 16 and 11, 14, 17 respectively) were generated using the metRology package for R, version 0.9-17 [12]. Both were calculated by target (square). Mandel's *h* is an indication of

relative deviation from the mean value for a group of data. Each vertical line is the value of Mandel's *h* statistic for one reported value from each laboratory, showing its deviation from the mean of the subset of measurements to which it pertains (i.e., the mean of the measurements performed by all the laboratories on the same square) divided by the standard deviation of this measurement subset. The larger the *h* value, the bigger the deviation, the poorer is the accuracy of that laboratory. Horizontal dashed lines are indicators of significance at a 95 % confidence level, based on an assumed normal distribution. Mandel's *h* showed for each target if the laboratories overestimated or underestimated with respect to the mean of the measurements performed by all the laboratories on the same target.
Median scaled difference (MSD) for a reported value x_i is calculated as Median ($|x_j - x_j|/\sqrt{(u_j^2 + u_j^2)}$) where j runs over all the values reported by the other laboratories on the same target and $u_{i,j}$ denotes the standard uncertainty associated with $x_{i,j}$. MSD increases as a data point deviates from the majority of the others, in relation to the uncertainty of the difference ($x_j - x_j$). The horizontal dashed line is at MSD = 2: it is an approximate 95 % confidence level indicator based on independent normally distributed data with correctly reported uncertainty. Broadly, MSD values larger than two indicate that a result is unusually distant from others, compared to its uncertainty.

5.1 Uncertainty analysis

The combined uncertainty was evaluated maily accordingly to [13, 14] and the expanded uncertainty was calculated considering as coverage factor, in the majority of the cases, k=2. Briefly, for both cell number and cell confluency participants considered repeatability and reproducibility of the measurement as main sources of uncertainty; for cell density also the uncertainty of the area of measurement was considered. Other sources of uncertainty, considered by the majoirity of participants, are reported and described in table 3.
INRIM performed studies on dye stability and image exposure time, reported in Appendix 1.

Component	Source	Description
uncertainty due to the sample	cell proliferation	dividing cells (e.g. two nuclei sharing the cytoplasm)
	omogeneity / distribution	cells agglomerates or overlapped cells
	edge error	cells on the edges of the area of measurament
	dyes stability	decay of fluorescence intensity
	sample quality	dust, scratches
uncertainty due to acquisition	image assembly	automatic image overlapping
	instrument set-up	brightness, contrast, illumination distribution over the image field of view, detection efficiency distribution, focus plane adjustment, image exposure time
uncertainty due to analysis	measurement repeatibility	standard deviation of the means of repeated measurements
	reproducibility regarding operator / automatic machine	mean of the standard deviation of the values of repeated measurements for every operator / automatic machine
	image quality	image and background noise
	threshold setting	manual setting of threshold for image analysis
	algorithms	algorithm selected for image analysis

Table 3. Uncertainty sources considered by partcipants.

5.2 Cell number

Table 4 reports the number of cells counted within the defined area in each square by each lab with absolute and relative expanded uncertainty. The defined area is calculated as described in section 4.1 and in details in Appendix 2. Figure 7 and 8 show the dotplot and the Mandel's *h* indicator, respectively, for cell number.

SET 1 - CELL NUMBER								SET 2 – CELL NUMBER						
Lab Square	A	B	C	D	E			Lab Square	F	G	H	H1	L	
s1.L.A1	28.00 0.05 0.18	28.00 0.05 0.17	28.67 1.29 4.50	26.00 0.00 0.00	30.667 3.609 11.770	cell number *U* *U* %		s2.L.U5	57.0 2.2 3.9	58.0 2.5 4.3	60.0 14.4 24.0	50.0 3.0 6.1	57.7 0.8 1.4	cell number *U* *U* %
s1.L.M1	42.00 2.70 6.40	42.00 0.07 0.17	43.67 0.57 1.32	43.67 0.00 0.00	44.889 0.462 1.030	cell number *U* *U* %		s2.L.Q11	101.0 1.4 1.4	97.7 2.9 2.9			97.0 3.3 3.4	cell number *U* *U* %
s1.L.K12	28.33 2.60 9.10	28.33 4.96 17.52	30.33 0.57 1.90	27.00 0.58 1.32	34.000 0.859 2.530	cell number *U* *U* %		s2.L.K15	76.0 2.2 2.8	80.3 7.6 9.4	80.0 14.5 18.3	67.0 3.5 5.2	77.6 5.8 7.5	cell number *U* *U* %
s1.L.U13	55.00 2.30 4.20	56.33 4.96 8.81	57.67 0.57 1.00	54.33 0.58 1.06	57.778 2.084 3.610	cell number *U* *U* %		s2.L.G10	60.0 0.0 0.0	64.0 2.5 3.9	69.0 20.9 30.4	56.0 3.9 7.0	64.0 3.9 6.1	cell number *U* *U* %
s1.M.I8	153.00 10.00 6.50	155.00 7.42 4.79	164.67 3.78 2.22	149.67 1.15 0.77	174.33 2.43 1.39	cell number *U* *U* %		s2.M.P8	113.0 1.7 1.5	115.0 10.8 9.4			108.9 4.7 4.3	cell number *U* *U* %
s1.M.R10	212.00 17.00 8.00	220.33 13.79 6.26	230.67 1.91 0.83	193.33 0.58 0.30	236.00 10.95 4.64	cell number *U* *U* %		s2.M.K10	206.0 1.0 0.5	210.0 15.1 7.2			191.7 7.9 4.1	cell number *U* *U* %
s1.M.K15	197.00 10.00 5.10	199.67 16.26 8.14	209.67 12.19 5.81	206.33 0.58 0.28	207.22 9.42 4.54	cell number *U* *U* %		s2.M.D15		142.0 0.0 0.0			140.4 2.4 1.7	cell number *U* *U* %
s1.M.U14	180.00 8.00 4.00	182.00 8.57 4.71	181.33 7.47 4.11	152.67 0.58 0.38	188.89 4.25 2.25	cell number *U* *U* %		s2.M.L19	165.0 1.5 0.9	173.0 6.6 3.8			156.6 7.0 4.5	cell number *U* *U* %
s1.H.D9	275.00 23.00 8.40	282.67 25.17 8.90	292.67 17.67 6.03	220.33 6.11 2.77	301.11 9.81 3.26	cell number *U* *U* %		s2.H.H2	381.0 3.4 0.9	394.7 73.7 18.7			374.0 7.0 1.9	cell number *U* *U* %
s1.H.R10	455.00 28.00 6.20	467.00 61.99 13.27	497.33 41.93	253.00 3.46 8.43	507.22 2.79 1.37	cell number *U* *U* %		s2.H.P3	378.0 1.9 0.5	394.7 35.5 9.0			365.0 13.0 3.6	cell number *U* *U* %
s1.H.F16	398.00 17.00 4.30	380.33 16.21 4.26	412.67 21.06 5.10	390.33 2.52 0.64		cell number *U* *U* %		s2.H.I16	365.0 1.7 0.5	376.3 29.6 7.9			360.0 13.0 3.6	cell number *U* *U* %
s1.H.O17	294.00 25.00 8.50		317.33 13.51 4.26	75.67 1.16 1.53	320.78 10.81 3.37	cell number *U* *U* %		s2.H.U20	263.0 2.6 1.0	273.7 10.3 3.8			241.0 11.0 4.6	cell number *U* *U* %

LEGENDA	
Cell number	Cell number [count]
U	Expandend absolute uncertainty [count]
U %	Expandend relative uncertainty [%]

Table 4. Participant's results on Cell number.

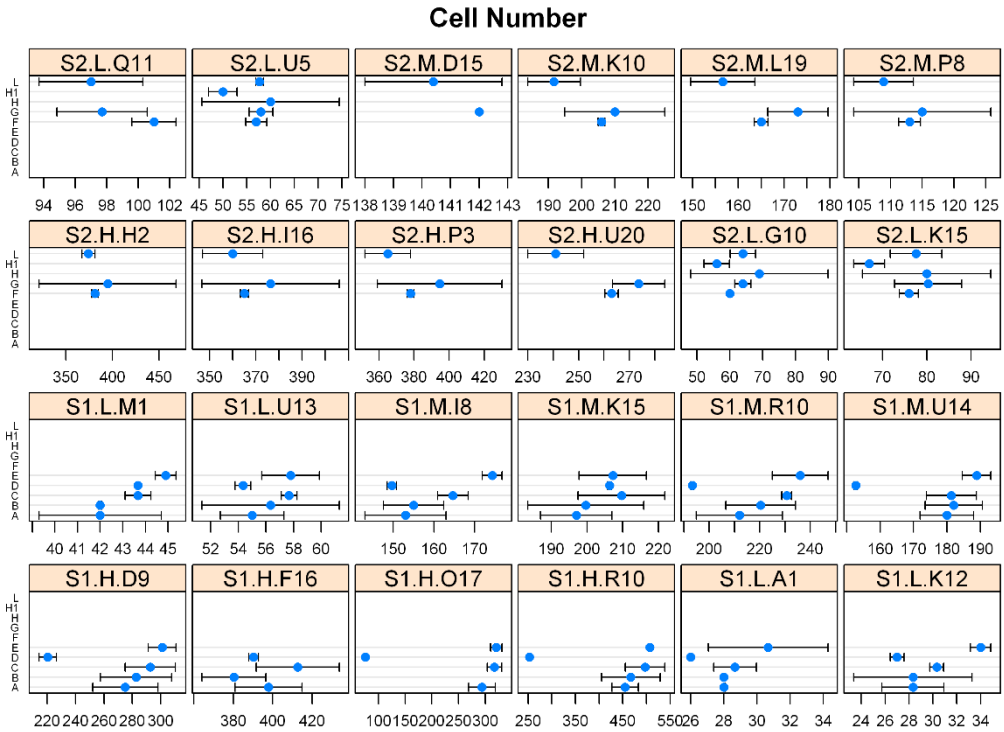


Fig. 7. Cell number by target and laboratory. Error bars show expanded uncertainties calculated with coverage factor $k = 2$.

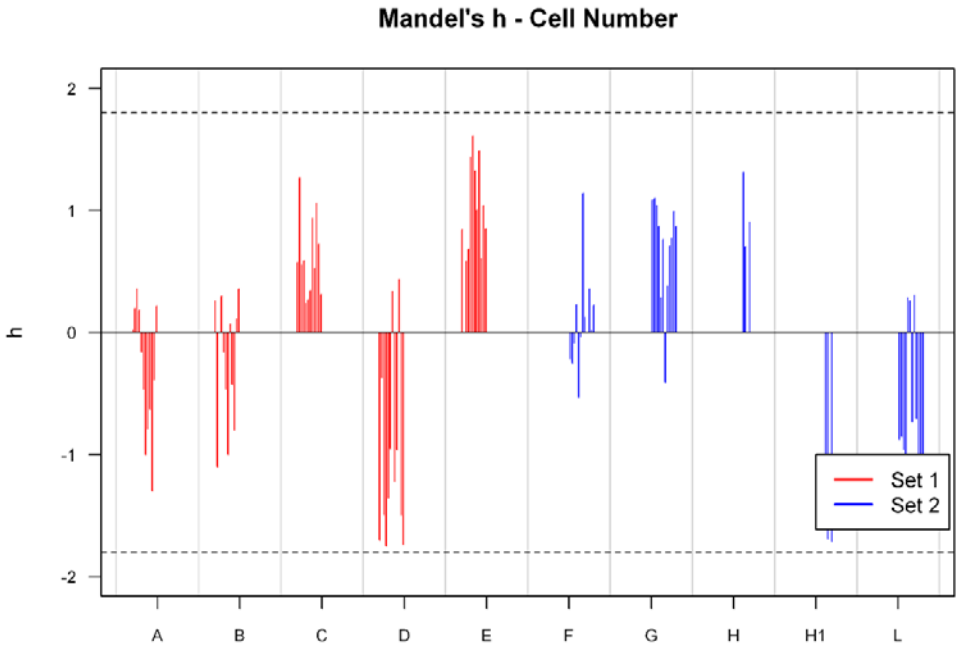


Fig. 8. Mandel's h for cell number

Institutions reported different uncertainty values, either being very conservative (hence stating large uncertainties) or being optimistic to some extent (even zero uncertainties were declared). A general agreement between measurements is encountered, in the majority of the targets, within the declared uncertainties, except for some measurements which falls apart from the others and/or have a very small associated uncertainty. The agreement would be improved if the uncertainties were more similar across Institutions.

Another characteristic is that there is often a systematic behavior among the Institutions, also well visible from Mandel's *h* (figure 8), in which it is evident the tendency of some Institutions to overestimate or

underestimate with respect to the others. There is not, however, any measurement exceeding the 95 % confidence limits.

5.3 Cell density

Table 5 reports the cell density measured within the defined area in each square by each lab. The defined area is calculated as described in section 4.1 and in details in Appendix 2. Figure 9, 10, 11 show the dotplot, the Mandel's *h* and MSD indicators, respectively, for cell density.

SET 1 - CELL DENSITY						
Lab	A	B	C	D	E	
Square						
	133.5	133.5	136.7	123.9	146.2	cells/mm^2
s1.L.A1	6.7	6.7	9.2	6.2	18.7	*U*
	5.0	5.0	6.7	5.0	12.8	*U* %
	200.2	200.2	208.2	208.2	214.0	cells/mm^2
s1.L.M1	16.3	10.0	10.8	10.4	10.9	*U*
	8.1	5.0	5.2	5.0	5.1	*U* %
	135.1	135.1	144.6	128.7	162.1	cells/mm^2
s1.L.K12	14.0	24.6	7.7	6.7	9.1	*U*
	10.4	18.2	5.3	5.2	5.6	*U* %
	262.2	268.5	274.9	259.0	275.4	cells/mm^2
s1.L.U13	17.1	27.2	14.0	13.2	17.0	*U*
	6.5	10.1	5.1	5.1	6.2	*U* %
	729.4	738.9	785.0	713.5	831.1	cells/mm^2
s1.M.I8	59.8	51.2	42.9	36.1	43.1	*U*
	8.2	6.9	5.5	5.1	5.2	*U* %
	1010.7	1050.4	1099.7	921.7	1125.1	cells/mm^2
s1.M.R10	95.3	84.2	55.7	46.2	76.7	*U*
	9.4	8.0	5.1	5.0	6.8	*U* %
	939.2	951.9	999.6	983.6	987.9	cells/mm^2
s1.M.K15	67.1	90.9	76.6	49.3	66.7	*U*
	7.1	9.6	7.7	5.0	6.8	*U* %
	858.1	867.6	864.4	727.8	900.5	cells/mm^2
s1.M.U14	54.9	59.6	56.0	36.5	49.4	*U*
	6.4	6.9	6.5	5.0	5.5	*U* %
	1311.0	1347.6	1395.2	1050.4	1435.5	cells/mm^2
s1.H.D9	128.2	137.6	109.3	60.1	85.7	*U*
	9.8	10.2	7.8	5.7	6.0	*U* %
	2169.1	2226.3	2370.9	1206.1	2418.1	cells/mm^2
s1.H.R10	172.8	315.7	232.4	62.5	121.6	*U*
	8.0	14.2	9.8	5.2	5.0	*U* %
	1897.4	1813.1	1967.3	1860.8		cells/mm^2
s1.H.F16	125.1	119.1	140.5	93.8		*U*
	6.6	6.6	7.1	5.0		*U* %
	1401.6		1512.8	360.7	1529.2	cells/mm^2
s1.H.O17	138.2		99.4	18.9	92.2	*U*
	9.9		6.6	5.2	6.0	*U* %

SET 2 - CELL DENSITY								
Lab	F	G	H	H1	I	I1	L	
Square								
	271.7	276.5	286.0	238.4	252.3	251.5	275.1	cells/mm^2
s2.L.U5	17.2	18.2	70.2	18.8	13.3	40.5	14.0	*U*
	6.3	6.6	24.5	7.9	5.3	16.1	5.1	*U* %
	481.5	465.8			432.7	434.4	462.4	cells/mm^2
s2.L.Q11	25.0	27.0			19.7	34.9	27.0	*U*
	5.2	5.8			4.5	8.0	5.8	*U* %
	362.3	383.0	381.4	319.4	361.8	352.4	369.9	cells/mm^2
s2.L.K15	20.8	40.9	72.4	23.0	21.5	62.8	31.8	*U*
	5.8	10.7	19.0	7.2	6.0	17.8	8.6	*U* %
	286.0	305.1	328.9	267.0	295.8	281.7	305.1	cells/mm^2
s2.L.G10	14.3	19.3	101.2	23.0	3.0	43.6	23.8	*U*
	5.0	6.3	30.8	8.6	1.0	15.5	7.8	*U* %
	538.7	548.2					519.2	cells/mm^2
s2.M.P8	28.2	58.4					33.2	*U*
	5.2	10.7					6.4	*U* %
	982.1	1001.1			937.1	924.7	913.9	cells/mm^2
s2.M.K10	49.3	87.7			29.4	56.6	58.5	*U*
	5.0	8.8			3.1	6.1	6.4	*U* %
		677.0			616.9	612.6	669.3	cells/mm^2
s2.M.D15		33.8			11.0	33.4	36.0	*U*
		5.0			1.8	5.5	5.4	*U* %
	786.6	824.7			742.0	730.2	746.6	cells/mm^2
s2.M.L19	40.0	51.8			20.1	40.1	47.8	*U*
	5.1	6.3			2.7	5.5	6.4	*U* %
	1816.3	1881.5			1719.0	1614.0	1783.0	cells/mm^2
s2.H.H2	92.2	363.5			124.2	272.0	96.0	*U*
	5.1	19.3			7.2	16.9	5.4	*U* %
	1802.0	1881.5			1723.7	1676.1	1740.1	cells/mm^2
s2.H.P3	90.6	193.5			105.4	206.1	111.4	*U*
	5.0	10.3			6.1	12.3	6.4	*U* %
	1740.1	1794.1			1640.1	1570.1	1716.2	cells/mm^2
s2.H.I16	87.4	167.3			95.8	197.4	109.9	*U*
	5.0	9.3			5.8	12.6	6.4	*U* %
	1253.8	1304.7			1197.7	1176.9	1148.9	cells/mm^2
s2.H.U20	63.9	81.8			50.6	128.9	81.2	*U*
	5.1	6.3			4.2	11.0	7.1	*U* %

LEGENDA	
cells/mm^2	Cell density [cells/mm^2]
U	Expandend absolute uncertainty [cells/mm^2]
U %	Expandend relative uncertainty [%]

Table 5. Participant's results for Cell density.

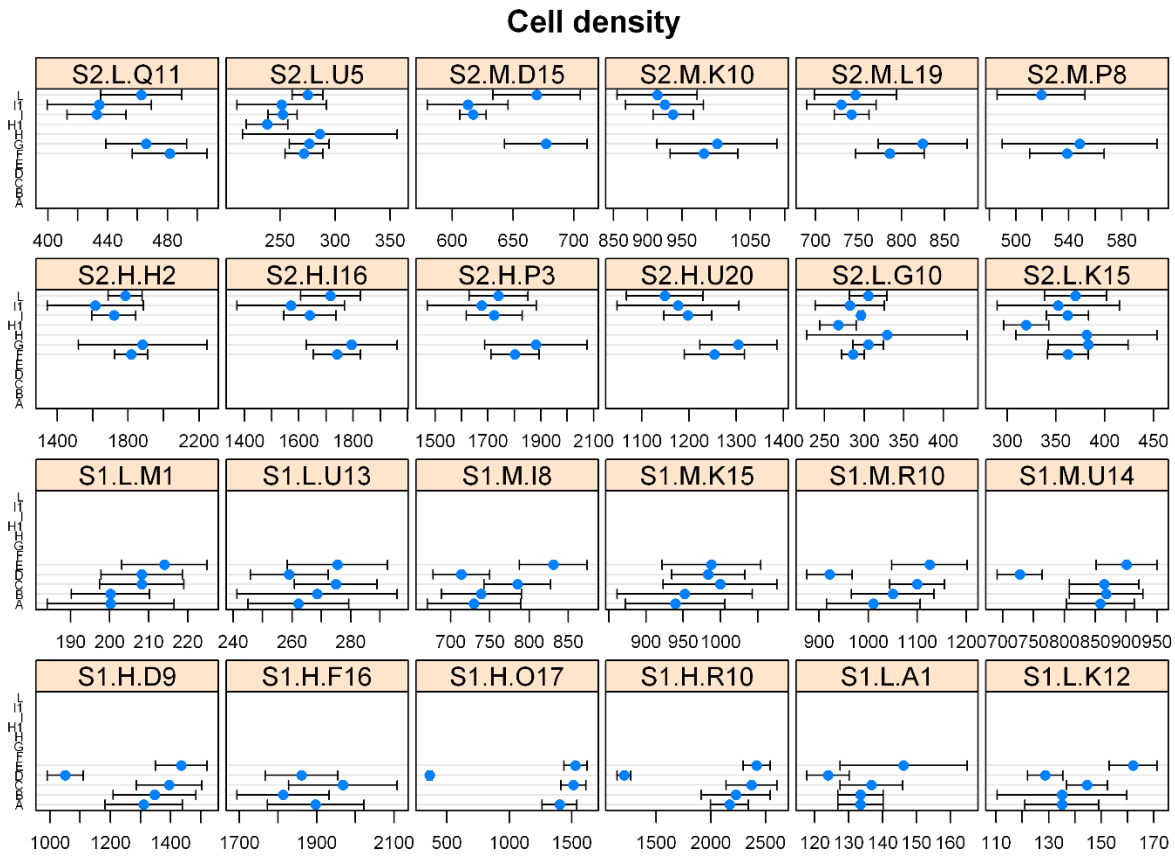


Fig. 9. Cell density by target and laboratory. Error bars show expanded uncertainties calculated with coverage factor $k = 2$

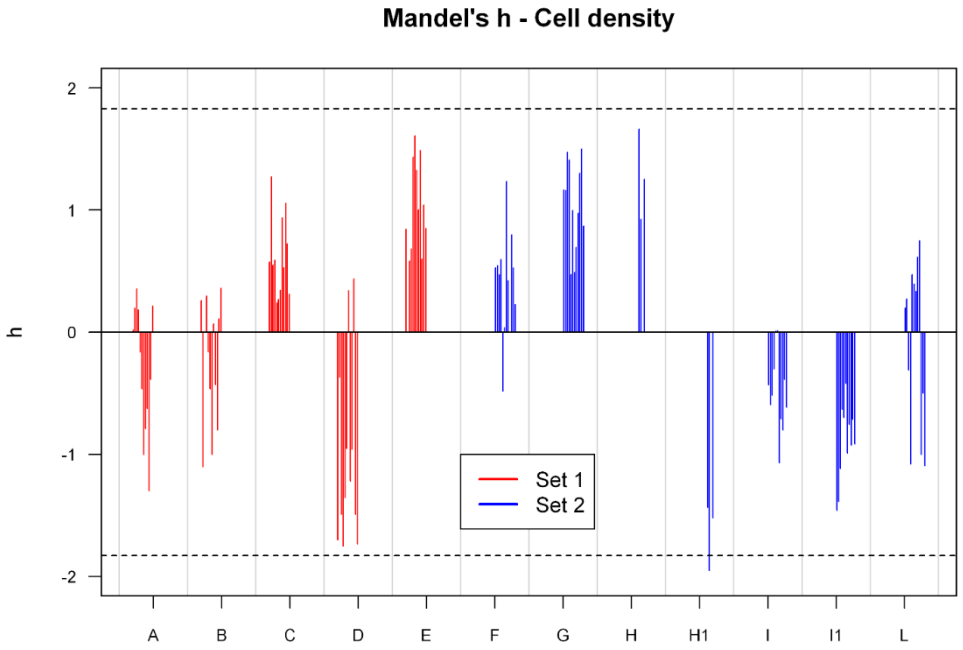


Fig. 10. Mandel's h for cell density

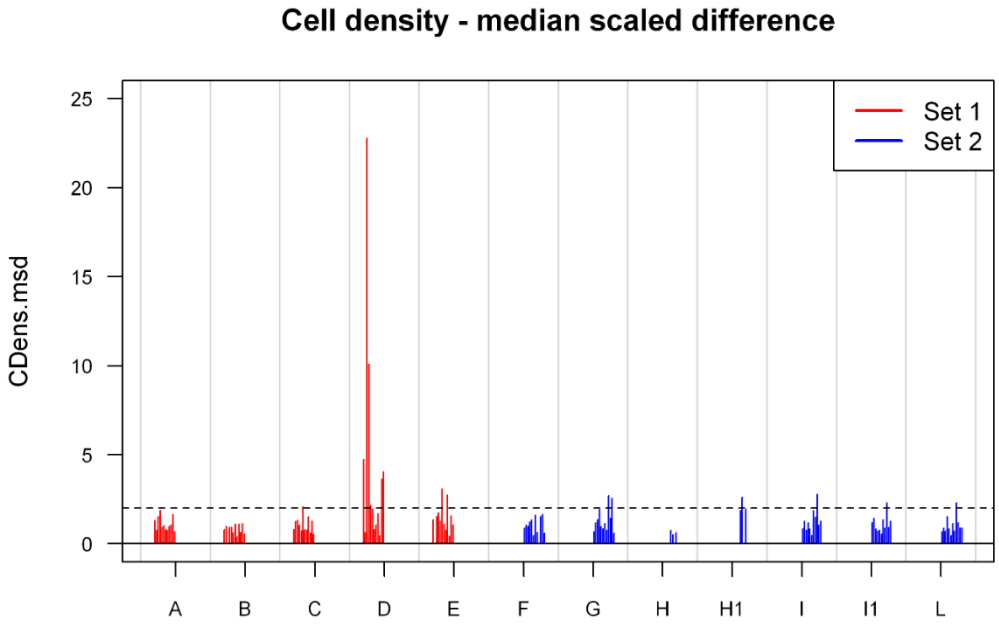


Fig. 11. MSD for cell density

Institutions reported variable uncertainties: some Institutions appeared very conservative, reporting large uncertainty, others appeared very optimistic, reporting narrow uncertainty. There is greater agreement between results with respect to the cell number results.

For set 1, from Mandel's *h* (figure 10), it is evident the tendency of some Institutions to overestimate or underestimate with respect to the others. Only one lab exceeded the 95 % confidence limits. For set 2 data of I and I1 are added: this modify the trend with respect to the cell number. Apart from one lab (L), the presence of systematic deviations (underestimation or overestimation) is evident for all the Institutions. Only for one target of Institution H1 the 95 % confidence limits were excedeed. In figure 11, MSD data showed that, apart from Institution D, almost all of the targets were below the 95% confidence limit.

5.4 Cell Confluency

Table 6 reports the cell confluency measured within the defined area in each square by each lab. The defined area is calculated as described in section 4.1 and in details in Appendix 2. Figure 12, 13, 14 show the dotplot, the Mandel's *h* and MSD indicators, respectively, for cell confluency.

SET 1 - CELL CONFLUENCY								SET 2 - CELL CONFLUENCY						
Lab / Square	A	B	C	D	E			Lab / Square	F	G	H	I	L	
s1.L.A1	14.02	10.34	8.80	8.71	13.00	confluency		s2.L.U5	20.84	17.98	13.53	20.30	22.40	confluency
	2.50	0.32	2.07		2.50	*U*			1.97	8.79	2.31	0.90	2.40	*U*
	17.83	3.04	23.58		19.20	*U* %			9.45	48.91	17.07	4.60	11.00	*U* %
s1.L.M1	17.72	14.79	12.70	11.35	18.30	confluency		s2.L.Q11	34.85	26.37		32.90	34.00	confluency
	2.70	0.50	0.93		3.00	*U*			1.74	11.00		4.30	5.20	*U*
	15.24	3.39	7.35		16.17	*U* %			4.99	41.72		13.10	15.00	*U* %
s1.L.K12	14.36	10.26	9.26	9.69	13.70	confluency		s2.L.K15	26.15	22.98	19.74	24.10	29.40	confluency
	2.10	0.21	0.73		4.00	*U*			2.59	7.88	3.01	1.70	4.00	*U*
	14.62	2.02	8.00		29.34	*U* %			9.91	34.33	15.23	7.00	14.00	*U* %
s1.L.U13	24.94	20.69	16.27	14.48	24.10	confluency		s2.L.G10	20.24	16.79	17.78	18.40	20.30	confluency
	3.10	0.53	1.21		5.20	*U*			1.29	1.00	4.44	2.30	3.10	*U*
	12.43	2.55	7.41		21.70	*U* %			6.37	5.94	25.01	12.30	15.00	*U* %
s1.M.I8	63.39	50.59	43.46	48.69	68.10	confluency		s2.M.P8	48.89	47.11			53.50	confluency
	10.80	6.18	2.54		8.20	*U*			3.23	4.21			4.40	*U*
	17.04	12.22	5.47		11.98	*U* %			6.61	8.93			8.30	*U* %
s1.M.R10	74.75	68.59	45.87	50.27	76.60	confluency		s2.M.K10	64.48	63.79		70.60	75.30	confluency
	9.10	3.25	15.29		7.90	*U*			7.35	18.93		5.90	5.70	*U*
	12.17	4.73	33.35		10.27	*U* %			11.40	29.68		8.40	7.50	*U* %
s1.M.K15	70.17	67.18	52.48	63.19	76.20	confluency		s2.M.D15		44.58		48.10	55.10	confluency
	8.52	5.01	8.55		6.70	*U*				9.13		4.80	5.60	*U*
	12.14	7.46	16.28		8.79	*U* %				20.48		9.90	10.00	*U* %
s1.M.U14	72.27	63.55	53.29	55.04	75.70	confluency		s2.M.L19	57.84	56.38		62.20	64.60	confluency
	8.40	4.50	7.28		5.40	*U*			4.73	8.02		4.80	5.60	*U*
	11.62	7.08	13.66		7.10	*U* %			8.18	14.23		7.70	8.60	*U* %
s1.H.D9	91.87	83.94	67.86	65.05	89.90	confluency		s2.H.H2	97.71	78.72		96.60	97.60	confluency
	12.20	15.58	11.25		7.20	*U*			1.27	35.89		3.30	2.50	*U*
	13.28	18.56	16.58		8.06	*U* %			1.30	45.60		3.40	2.60	*U* %
s1.H.R10	93.77	95.27	74.46	86.18	97.00	confluency		s2.H.P3	96.85	77.50		97.90	97.90	confluency
	8.50	9.27	5.01		1.10	*U*			1.76	41.55		1.70	1.10	*U*
	9.07	9.73	6.73		1.09	*U* %			1.82	53.62		1.80	1.10	*U* %
s1.H.F16	92.25	92.71	70.10	84.59		confluency		s2.H.I16	95.42	70.20		91.30	94.90	confluency
	10.00	12.24	14.63			*U*			3.40	50.22		11.40	3.20	*U*
	10.84	13.21	20.87			*U* %			3.56	71.54		12.50	3.40	*U* %
s1.H.O17	91.49		66.15	88.76	92.90	confluency		s2.H.U20	93.91	75.58		93.60	93.70	confluency
	11.70		9.79		1.90	*U*			0.76	37.15		3.90	1.60	*U*
	12.79		14.80		2.08	*U* %			0.81	49.15		4.10	1.80	*U* %

LEGENDA	
confluency	Cell confluency [area/area, %]
U	Expandend absolute uncertainty [area/area, %]
U %	Expandend relative uncertainty [%]

Table 6. Participant's results on Cell confluency.

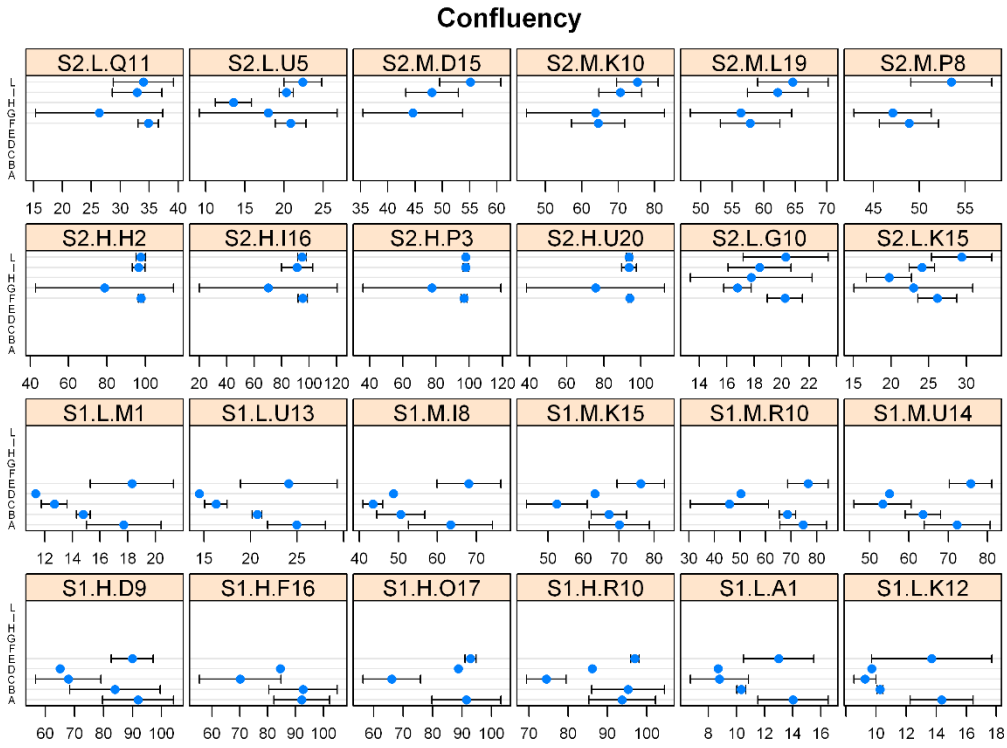


Fig. 12. Confluency by target and laboratory. Error bars show expanded uncertainties calculated with coverage factor $k = 2$.

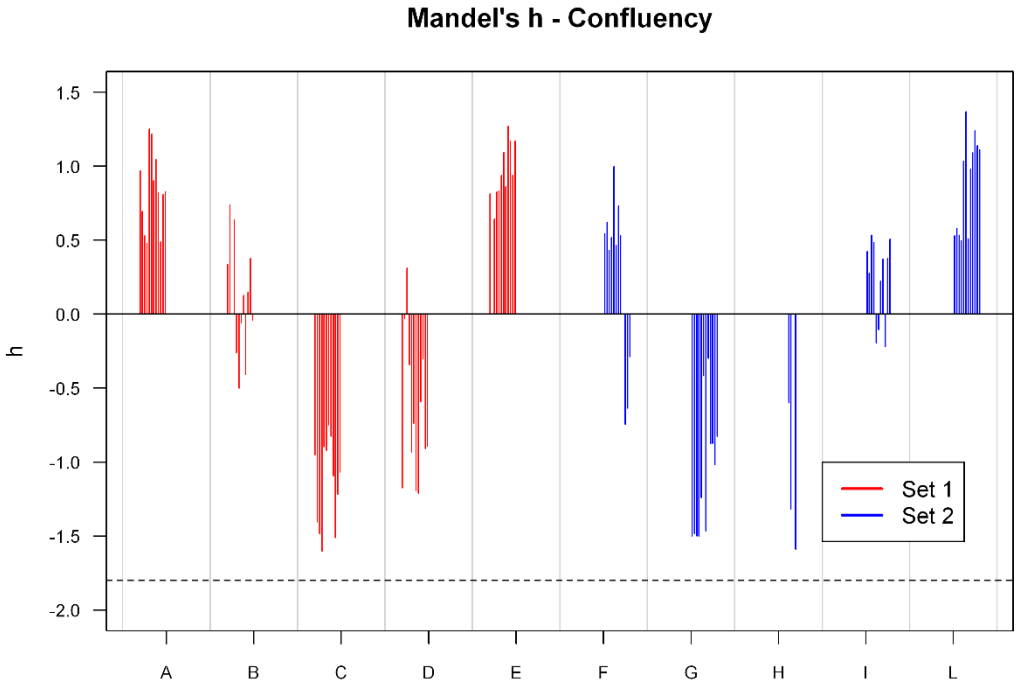


Fig. 13. Mandel's h for confluency

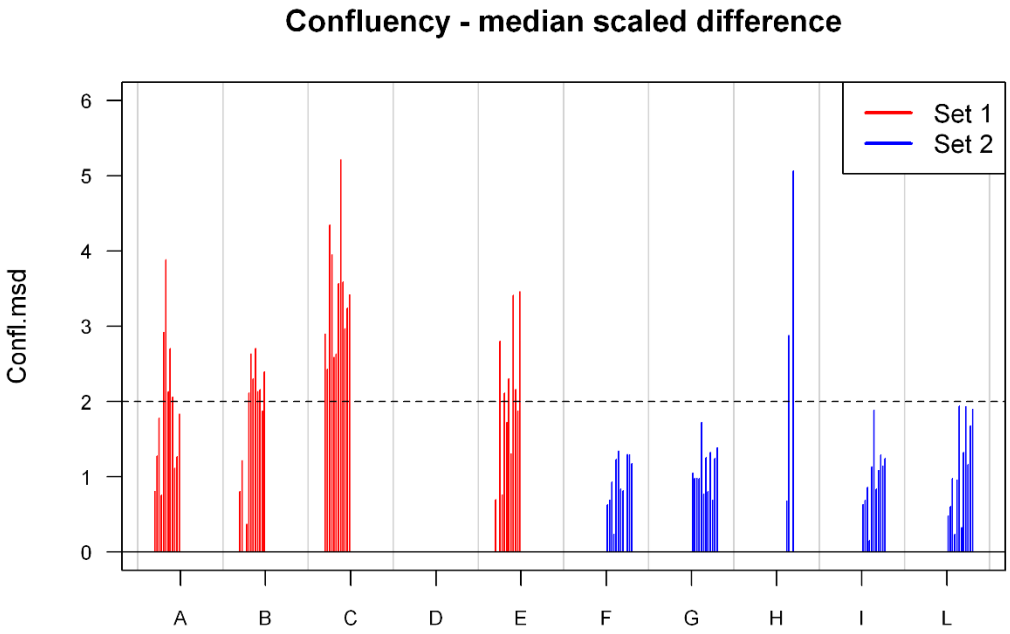


Fig. 14. MSD for confluency

Institutions reported variable uncertainties: some Institutions appeared very conservative, reporting large uncertainty, others appeared very optimistic, reporting narrow uncertainty.

From Mandel's *h* (figure 13), with the exception of B, F, I, it is evident the tendency of the Institutions to overestimate or underestimate with respect to the others. No one exceeded the 95 % confidence limits. In figure 14, MSD data showed that for set 1 the majority of target results exceed the dashed line, indicating that results were unusually distant from others, compared to their uncertainty. In set 2, the results are much better and, only for Institution H, 2 targets out of 3 are unusually distant from the others.

5.5 Average cell area

Table 7 reports the average cell area measured within the defined area in each square by each Insitution. The defined area is calculated as described in section 4.1 and in details in Appendix 2. Figure 15, 16, 17 show the dotplot, the Mandel's *h* and MSD indicators, respectively, for average cell area.

SET 1 - CELL AREA							SET 2 - CELL AREA							
Lab / Square	A	B	C	D	E		Lab / Square	F	G	H	I	I2	L	
s1.L.A1	802.6	907.4	732.6	828.7	1024.4	cell area	s2.L.U5	788.83	726.16	716.06	804.00	806.90	811.45	cell area
	123.10	59.60	165.74		52.93	*U*		109.51	353.39	52.74	54.50	134.60	87.72	*U*
	15.34	6.57	22.62		5.17	*U* %		13.88	48.67	7.36	6.80	16.70	10.81	*U* %
s1.L.M1	762.3	866.5	692.5	640.0	955.8	cell area	s2.L.Q11	731.78	626.31		760.00	757.50	732.65	cell area
	117.17	58.60	64.03		145.25	*U*		49.23	224.95		102.60	113.50	114.97	*U*
	14.60	6.76	9.25		15.20	*U* %		6.73	35.92		13.50	15.00	15.69	*U* %
s1.L.K12	819.9	905.1	736.5	872.3	940.4	cell area	s2.L.K15	700.28	660.16	1094.81	665.00	682.80	791.91	cell area
	124.69	103.40	68.69		266.65	*U*		72.74	172.87	490.52	59.70	130.10	122.64	*U*
	15.54	11.43	9.33		28.36	*U* %		10.39	26.19	44.80	9.00	19.00	15.49	*U* %
s1.L.U13	776.3	894.5	671.0	648.7	989.2	cell area	s2.L.G10	726.45	608.36	1123.47	622.00	652.80	662.99	cell area
	92.13	67.70	74.53		182.71	*U*		30.78	58.53	888.69	53.80	115.50	108.86	*U*
	11.48	7.57	11.11		18.47	*U* %		4.24	9.62	79.10	8.60	17.70	16.42	*U* %
s1.M.I8	741.0	823.9	666.7	802.7	908.5	cell area	s2.M.P8	915.70	950.53				1026.87	cell area
	81.70	120.10	66.17		104.79	*U*		64.17	212.30				95.44	*U*
	10.18	14.58	9.92		11.53	*U* %		7.01	22.33				9.29	*U* %
s1.M.R10	630.1	767.8	466.9	641.4	771.1	cell area	s2.M.K10	648.12	709.31		753.00	763.20	821.03	cell area
	75.95	60.40	141.36		43.25	*U*		89.27	289.81		53.40	67.38	70.91	*U*
	9.46	7.86	30.28		5.61	*U* %		13.77	40.86		7.10	8.80	8.64	*U* %
s1.M.K15	695.2	826.0	592.0	756.5	868.5	cell area	s2.M.D15		729.15		779.00	784.70	820.30	cell area
	47.33	87.20	93.54		112.63	*U*			184.97		77.80	881.10	84.72	*U*
	5.90	10.56	15.80		12.97	*U* %			25.37		10.00	11.20	10.33	*U* %
s1.M.U14	747.9	883.5	697.5	900.3	960.5	cell area	s2.M.L19	745.69	765.66		839.00	852.40	862.24	cell area
	53.35	80.10	110.15		47.02	*U*		69.25	158.34		56.20	70.13	84.28	*U*
	6.65	9.07	15.79		4.89	*U* %		9.29	20.68		6.70	8.20	9.77	*U* %
s1.H.D9	691.1	784.5	544.4	719.8	705.2	cell area	s2.H.H2	540.07	472.62		562.00	598.60	545.46	cell area
	69.57	181.30	110.08		37.90	*U*		7.85	322.34		44.50	102.80	17.32	*U*
	8.67	23.12	20.22		5.38	*U* %		1.45	68.20		7.90	17.20	3.18	*U* %
s1.H.R10	437.5	550.5	355.2	829.4	461.1	cell area	s2.H.P3	535.65	462.67		568.00	583.90	560.63	cell area
	52.26	75.30	67.91		5.12	*U*		11.83	309.56		36.10	72.51	20.98	*U*
	6.51	13.69	19.12		1.11	*U* %		2.21	66.91		6.40	12.40	3.74	*U* %
s1.H.F16	488.6	685.8	402.2	524.9		cell area	s2.H.I16	544.67	438.32		556.00	581.20	551.00	cell area
	40.03	123.80	85.44			*U*		24.60	358.03		75.30	101.80	27.20	*U*
	4.99	18.06	21.24			*U* %		4.52	81.68		13.50	17.50	4.94	*U* %
s1.H.O17	647.1		495.0	2854.5	700.8	cell area	s2.H.U20	749.22	652.67		781.00	795.10	812.66	cell area
	69.70		105.88		29.62	*U*		1.75	373.05		46.00	93.00	39.69	*U*
	8.68		21.39		4.23	*U* %		0.23	57.16		5.90	11.70	4.88	*U* %

LEGENDA	
Cell area	Cell area [μm^2]
U	Expandend absolute uncertainty [μm^2]
U %	Expandend relative uncertainty [%]

Table 7. Participant's results on Average cell area.

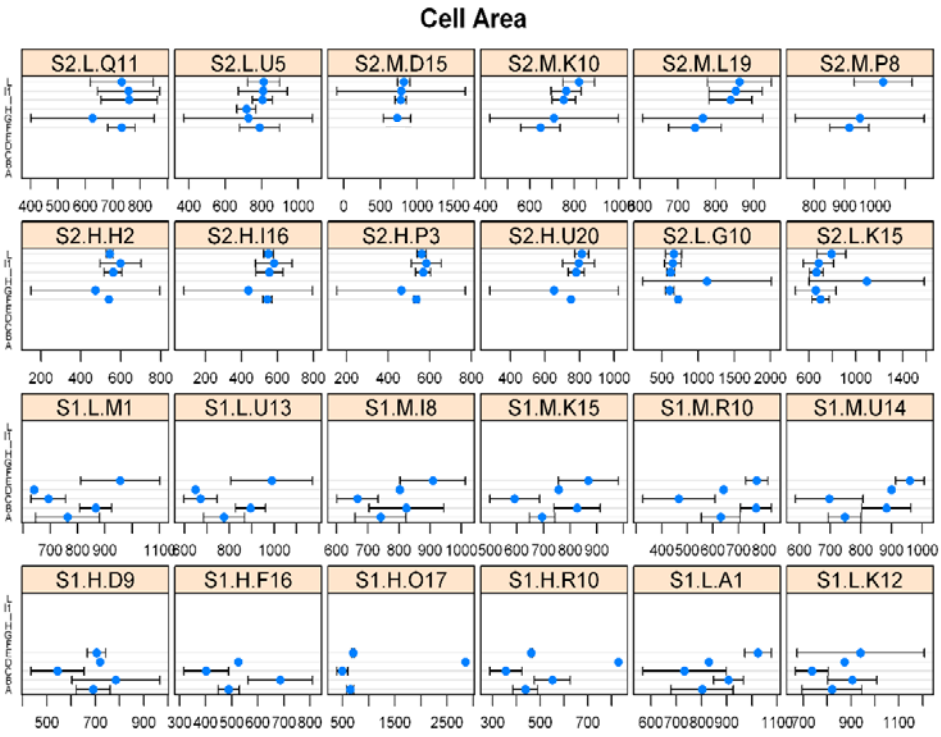


Fig. 15. Cell area by target and laboratory. Error bars show expanded uncertainties calculated with coverage factor $k = 2$.

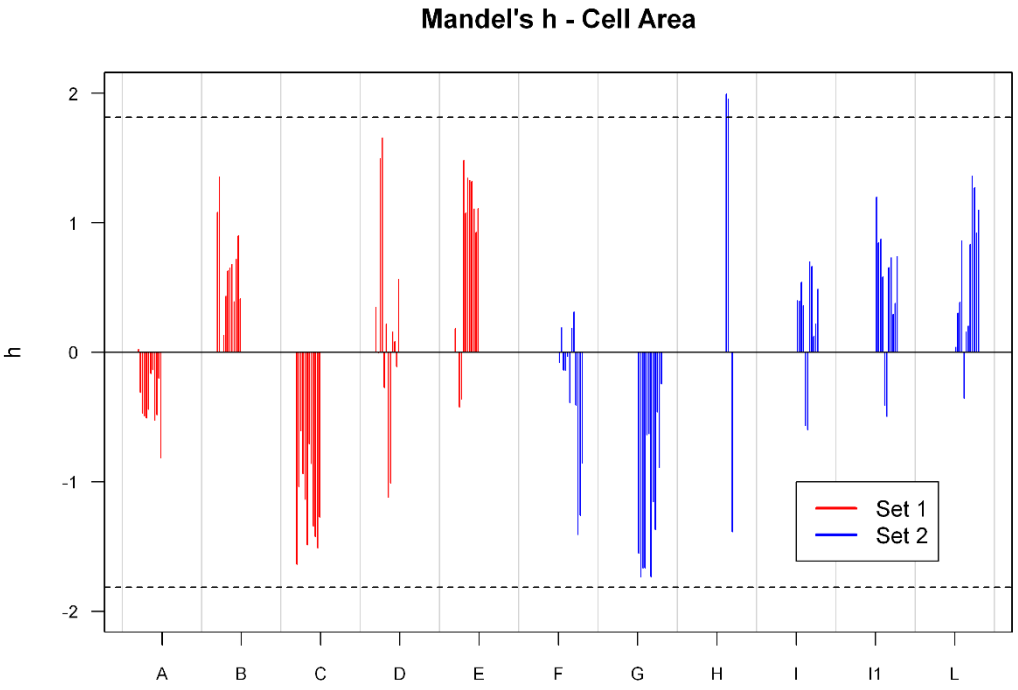


Fig. 16. Mandel's h for average cell area

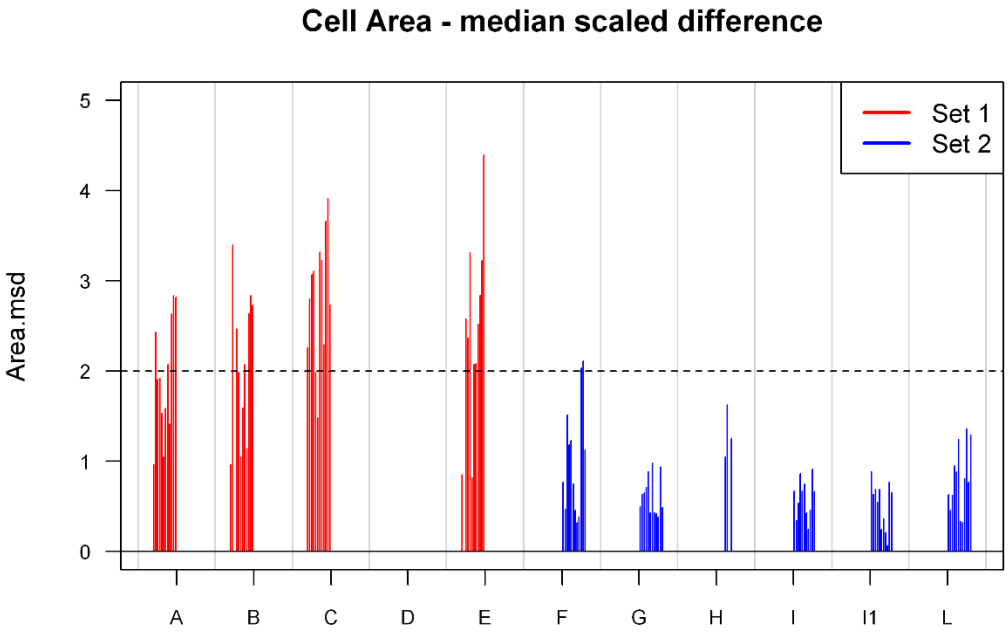


Fig. 17. MSD for average cell area

Institutions reported different uncertainty values, either being very conservative, reporting large uncertainty or being optimistic, reporting narrow uncertainty.
For set 1 the dispersion of data was much higher than for set 2.
From Mandel's *h* (figure 16), 4 Institutions sistematically overestimate or underestimate with respect to the others. For the others, such effect is slightly reduced. Only one Institution (H) for two targets exceeded the 95 % confidence limits. In figure 17, MSD data showed that for set 1 the majority of target results exceed the dashed line, indicating that results were unusually distant from others, compared to their uncertainty. In set 2, the results are much better and, only for Institution F, 2 targets out of 12 exceed the dashed line.

Average cell area was also plotted as mean values of Institutions results, by target (figure 18).

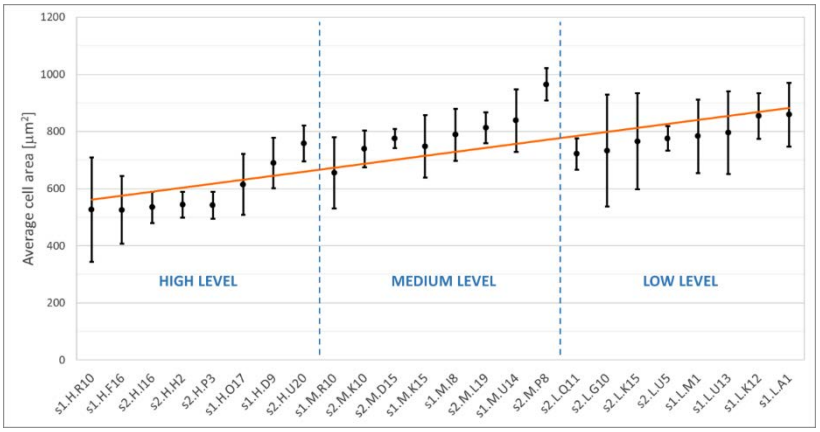


Fig. 18. Plot of average cell area with best-fitting line (orange). Error barrs shows standard deviations.

In the study, each measurement standard (square) was considered as an independent measurement object, as reported and represented in the last figures. However, data were also displayed combining the target by level – low, medium and hight (L, M and H, respectively) to show the general trend for each measurand, irrespectively of the targets. The **boxplots** of the raw data, divided level by level, were plotted in R version 3.4.3 for all the measurands (figure 19).

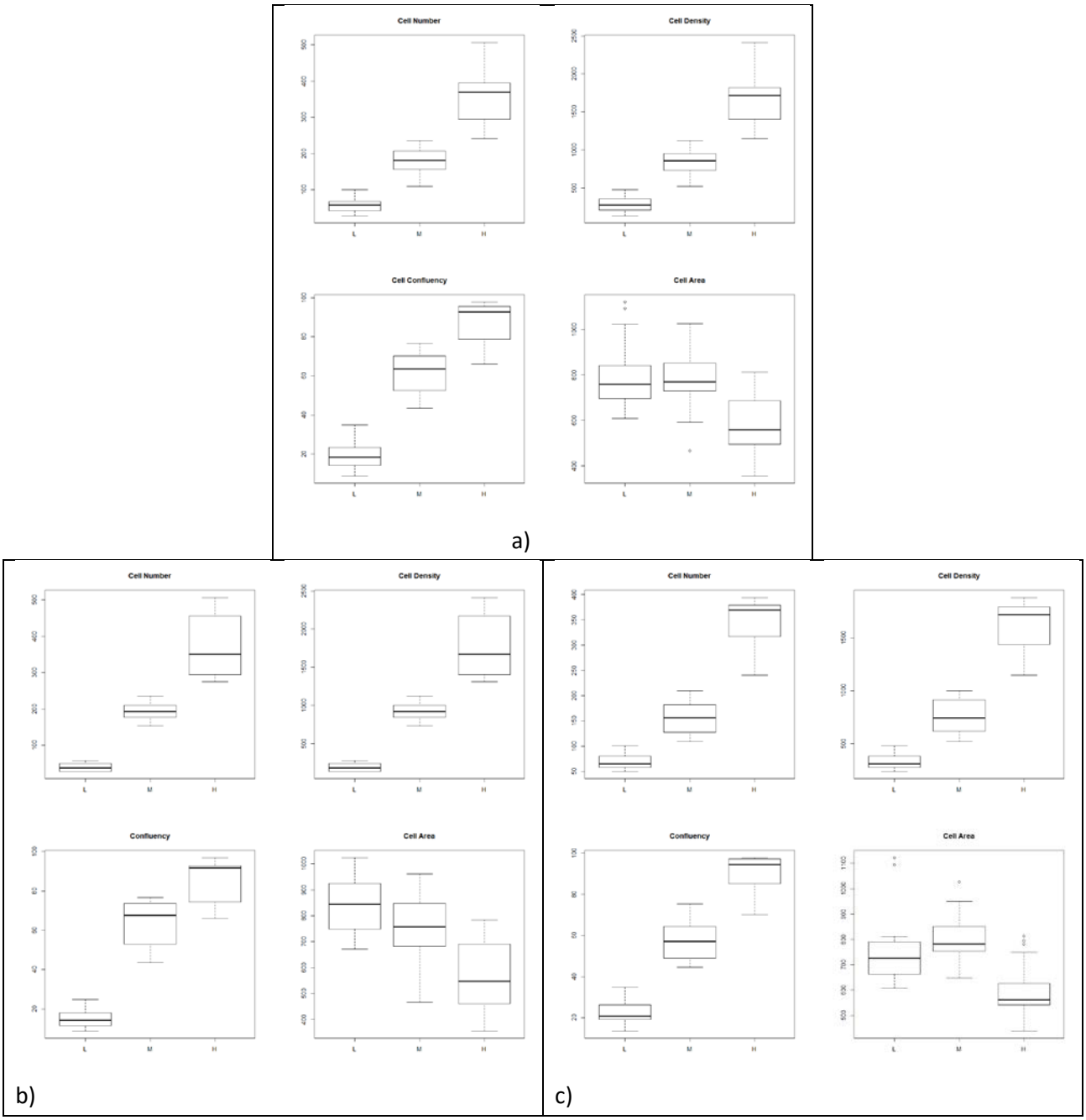


Fig. 19. Box plot of the 4 measurands grouped by levels. a) all the data; b) only set 1 data; c) only set 2 data.

6. Discussion

Due to the use of two sets of reference materials, all data sets are relatively small, making extreme values hard to identify. In addition, reported uncertainties are variable: some laboratories appeared very conservative, reporting large uncertainty, others appeared very optimistic, reporting narrow or even zero uncertainty. One Institution, for some measurands, did not report uncertainty data.

The dot plots show some evidence of systematic deviations by laboratory. This is evident by the plots of Mandel's *h*, calculated by target (Figure 8, 10, 13, 16): the figures show systematic laboratory effects.
The systematic effect could be induced by deviations in the defintion of the area of measurement.
The definition of the area of measurement could be not univocal if referred to the square edges: irregularities of the square edges and corners could lead to under- or overestimation of the area and dispersion among participants. To avoid these problems and to have SI traceable results, the area of measurement was defined within the protocol. The protocol also gave instructions to use a traceable ruler to cut the images and to get the same area of measurement. The correct use of the traceable ruler, provided to each laboratory, should have assured traceable measurements and comparable results among laboratories.
However, not all the laboratories measured on the correct area. Some of them did not use the ruler according to the protocol and considered a different area, sometime the deviation from the define area was very large and results were strongly affected. The expression of results in cell number per mm^2 (i.e. cell density) mitigated this disagreement: the participants' results were normalized on the defined area of measurement, with its associated uncertainty. All participants agreeded with the strategy and checked their data after normalization. The average cell area could have been expected as a constant across the levels (low, medium, high). However, the results of this study showed that the average area of a cell decreased when cell density and cell confluency increased. The measurement uncertainty was larger at lower levels of cell number per area. An explanation for this trend could simply be related to the availability of adhesion surface for each cell which decreases with increasing number of cells per area.

Taken together all these considerations, the definition of a reference value for each measurand could be difficult. However, some investigations were done in this sense.

6.1 Pilot study reference values

Among the variety of techniques and tools for determing the reference value of interlaboratory comparisons, **Procedures A and B** described in the guideline of M.G.Cox [15] have been considered here. They were applied to the laboratory data, target by target. The two procedures are conceived for the evaluation of key comparison data, providing a key comparison reference value (KCRV) and the associated uncertainty, and the degree of equivalence between each laboratory and the KCRV. Procedure A is based on the use of the weighted mean, together with a consistency check assessing its applicability. Whenever such test fails, Procedure B can be applied based on the use of the median, for example, as a more robust estimator. The calculations for the results produced within this pilot study were performed by means of a dedicated code, developed in R version 3.4.3. In the implementation of both procedures, a 95 % coverage interval is calculated for the deviation of each laboratory result for a specific target from the reference value (i.e. either the weighted mean or the median) obtained from all the measurements on that target. An interval encompassing the zero value indicates a compatibility of that result with the reference value: this situation is indicated with "1" in the tables 8, 9, 10 and 11.

Mandel's *h* and MDS results helped in the identification of those data that could be excluded in the Procedures A and B (e.g. data without uncertainty, measurement carried out on the wrong area, much higher than the defined area).

LEGENDA of tables 8-11	
	Missing data (i.e. square not acquired)
1	Cox's procedure passed
	All labs in agreement with the underlying reference value
0	Cox's procedure not passed
	Excluded data according to Mandel's *h* and MSD
A	Application of procedure A based on the weighted mean
B	Application of procedure A and then B based on the median

SET 1 – Cell number						
	A	B	C	D	E	Cox's procedure
s1.L.A1	1	1	1	0	1	B
s1.L.M1	1	0	1	1	0	B
s1.L.K12	1	1	1	1	0	B
s1.L.U13	1	1	1	1	1	B
s1.M.I8	1	1	0	0	0	B
s1.M.R10	1	1	1	0	1	B
s1.M.K15	1	1	1	1	1	B
s1.M.U14	1	1	1	0	0	B
s1.H.D9	1	1	1	0	1	B
s1.H.R10	1	1	1	0	1	B
s1.H.F16	1	1	0	1		B
s1.H.O17	1		1	0	0	B

SET 2 – Cell number						
	F	G	H	H1	L	Cox's procedure
s2.L.U5*	1	1	1	0	1	B
s2.L.Q11*	1	1			1	B
s2.L.K15*	1	1	1	0	1	B
s2.L.G10*	1	1	1	0	1	B
s2.M.P8	1	1			1	A
s2.M.K10	1	1			0	B
s2.M.D15		1			1	A
s2.M.L19	1	0			0	B
s2.H.H2	1	1			1	A
s2.H.P3	1	1			1	A
s2.H.I16	1	1			1	A
s2.H.U20	1	0			0	B

Table 8. Results of Cox's procedure A and B for cell number. *Agreement among all the laboratories was reached excluding H1 data.

For cell number, agreement between all the laboratories was reached in 2 and 9 cases out of 12 for set 1 and 2, respectively. H and H1 data were presented by the same lab, obtained with different methods. The exclusion of H1 was proposed to obtain agreement between all the participants.
For set 2, in particular, Procedure A was satisfactorily applied 5 times, indicating a strong consistency between the data. Note that for set 1, in 6 cases out of 10 the agreement was reached among all laboratories except one (not always the same one).

SET 1 - Cell density						
	A	B	C	D	E	Cox's procedure
s1.L.A1	1	1	1		1	A
s1.L.M1	1	1	1		1	A
s1.L.K12	1	1	1		0	B
s1.L.U13	1	1	1		1	A
s1.M.I8	1	1	1		0	B
s1.M.R10	1	1	1		1	A
s1.M.K15	1	1	1		1	A
s1.M.U14	1	1	1		1	A
s1.H.D9	1	1	1		1	A
s1.H.R10	1	1	1		1	B
s1.H.F16	1	1	1			A
s1.H.O17	1		1		1	A

SET 2 - Cell density								
	F	G	H	H1	I	I1	L	Cox's procedure
s2.L.U5*	1	1	1	0	1	1	1	B
s2.L.Q11*	1	1			1	1	1	B
s2.L.K15*	1	1	1	0	1	1	1	B
s2.L.G10*	1	1	1	0	1	1	1	B
s2.M.P8	1	1					1	A
s2.M.K10	1	1			1	1	1	A
s2.M.D15		0			0	0	0	B
s2.M.L19	1	0			1	1	1	B
s2.H.H2	1	1			1	1	1	A
s2.H.P3	1	1			1	1	1	A
s2.H.I16	1	1			1	1	1	A
s2.H.U20	1	1			1	1	1	B

Table 9. Results of Cox's procedure A and B for cell density. *Agreement among all the laboratories was reached excluding H1 data.

For cell density, Institution D was excluded from the Cox's procedure evaluation because MSD results showed unusual behavior in 6 out of 12 targets (see fig. 11). For set 1, agreement between all the laboratories was reached in 10 cases out of 12; in the remaining 2 cases, the agreement was reached for 3 laboratories out of 4. Procedure A was satisfactorily applied 9 times, indicating a strong consistency between the data.

For set 2, H and H1 data were presented by the same lab, obtained with different methods. The exclusion of H1 was proposed to obtain agreement between all the participants. Agreement between all the laboratories was reached in 10 cases out of 12. In 1 case out of 12 the agreement was reached for 4 laboratories out of 5. Procedure A was satisfactorily applied 5 times.
Note that for set 2, in 1 case (s2.M.D15) there was no agreement at all.

SET 1 - Confluency						
	A	B	C	D	E	Cox's procedure
s1.L.A1	1	0			1	B
s1.L.M1	1	1			1	B
s1.L.K12	1	1			1	B
s1.L.U13	1	1			1	B
s1.M.I8	1	0			1	B
s1.M.R10	1	1			1	B
s1.M.K15	1	1			1	B
s1.M.U14	1	1			1	B
s1.H.D9	1	1			1	A
s1.H.R10	1	1			1	A
s1.H.F16	1	1				A
s1.H.O17	1				1	A

SET 2 - Confluency						
	F	G	H	I	L	Cox's procedure
s2.L.U5	1	1	0	1	1	B
s2.L.Q11	1	1		1	1	A
s2.L.K15	1	1	0	1	0	B
s2.L.G10	1	1	1	1	1	B
s2.M.P8	1	1			1	B
s2.M.K10	1	1		1	1	B
s2.M.D15		1		1	1	B
s2.M.L19	1	1		1	1	A
s2.H.H2	1	1		1	1	A
s2.H.P3	1	1		1	1	A
s2.H.I16	1	1		1	1	A
s2.H.U20	1	1		1	1	A

Table 10. Results of Cox's procedure A and B for confluency

For cell confluency, Institution C and D were excluded from the Cox's procedure evaluation because C carried out the measurements on a wrong area, much higher than the defined one (effect which is highlighted even by MSD data in fig. 14), and D did not declared any uncertainty. These exclusions contributed to reduce the number of laboratories evaluated: in 10 cases only 3 labs were compared while in 2 cases (s1.H.F16 and s1.H.O17) the labs compared were only 2.
For set 1, agreement between all the laboratories was reached in 10 cases out of 12. Procedure A was satisfactorily applied 4 times.
For set 2, agreement between all the laboratories was reached in 10 cases out of 12. In 1 case out of 12 the agreement was reached for 4 laboratories out of 5. Procedure A was satisfactorily applied 6 times.

SET 1 – Average cell area						
	A	B	C	D	E	Cox's procedure
s1.L.A1	1	1			0	B
s1.L.M1	1	1			1	A
s1.L.K12	1	1			1	A
s1.L.U13	1	1			1	B
s1.M.I8	1	1			1	B
s1.M.R10	0	1			1	B
s1.M.K15	0	1			1	B
s1.M.U14	0	1			1	B
s1.H.D9	1	1			1	A
s1.H.R10	1	1			1	A
s1.H.F16	0	0				B
s1.H.O17	1				1	A

SET 2 – Average cell area							
	F	G	H	I	I1	L	Cox's procedure
s2.L.U5	1	1	0	1	1	1	B
s2.L.Q11	1	1		1	1	1	A
s2.L.K15	1	1	1	1	1	1	A
s2.L.G10	0	1	1	1	1	1	B
s2.M.P8	1	1				1	A
s2.M.K10	0	1		1	1	1	B
s2.M.D15		1		1	1	1	A
s2.M.L19	1	1		1	1	1	B
s2.H.H2	1	1		1	1	1	A
s2.H.P3	1	1		1	1	1	B
s2.H.I16	1	1		1	1	1	A
s2.H.U20	1	1		1	1	1	B

Table 11. Results of Cox's procedure A and B for average cell area

For average cell area, Institution C and D were excluded from the Cox's procedure evaluation because C carried out the measurements on the wrong area, much higher than the defined one (effect which is

highlighted even by MSD data in fig. 14), and D did not declared any uncertainty. These exclusions contributed to reduce the number of laboratories evaluated: in 10 cases only 3 labs were compared while in 2 cases (s1.H.F16 and s1.H.O17) the labs compared were only 2.
For set 1, agreement between all the laboratories was reached in 7 cases out of 12. Procedure A was satisfactorily applied 5 times.
For set 2, agreement between all the laboratories was reached in 9 cases out of 12 and in the remaining 3 cases the agreement was reached for 4 laboratories out of 5. This indicated a strong consistency between the data. Procedure A was satisfactorily applied 6 times.

7. Conclusion

The CCQM P123 pilot study has been the first study proposed on adherent cells within the EURAMET frame and had the unique feature to propose 4 measurands in one reference material and 3 measurement claims in one comparison. In addition, the three measurements required several operations and different level of complexity. These made the study intrinsically ambitious. The major strength of the pilot study was the possibility to compare four measurement capabilities by using a single reference material.
The study protocol was particularly detailed and overall complex to be executed and the reference material was not enough resistant to survive several shipments and improper handlings. The first set of reference material was severly dagamed after 5 shipments and relative handlings to perform the measurements. The second set was damaged as well. This study allowed several considerations in terms of: traceability, comparability of results and complexity of the comparison protocol. The traceability was assured by the use of the ruler. However, the instructions on the use of the ruler were not easily followed by all the participants. Comparability of results were compromised by failing to follow the protocol. However, the protocol itself should have given more precise and synthetic instructions.
Taken these considerations into account, in general the results for the four measurands showed a good agreement between the laboratories. In the 75 % of the measurement standards of set 2, the consistency between the laboratories' result and the proposed reference value was reached by all the laboratories for at least three measurands. For set 1, this consistency was obtained in the 50 % of the measurement standards.
In conclusion, the study has been a preparatory exercise for a key comparison on at least one of the 3 measurements claims (cell density, cell confluency and cell area) related to adhered cells features such as cell proliferation, morphology and size. The study has been applied to the A549 cell line, selected for some properties (mainly adhesion and size). However, the measurement claims are not intended to be limited to this cell line, can be applied to any adherent cell type matching the measurands ranges of the measurement claims. In conclusion, the study has been a preparatory exercise for a key comparison on at least one of the 3 measurements claims (cell density, cell confluency and cell area) related to adhered cells features such as cell proliferation, morphology and size. The results from such a key comparison could be utilized by the NMIs to provide traceability to different methods of bio-analysis.

8. Acknowledgements

We are grateful to Jayne Morrow for her assistance during the drafting of the pilot final report as Chair of the CAWG until April 2019.

9. References

[1] Faruqui N, Kummrow A, Fu B et al. Cellular Metrology: Scoping for a Value Proposition in Extra- and Intracellular Measurements *Front Bioeng Biotechnol.* 2020 14;7:456
[2] Alberts B, Cell Movements and the Shaping of the Vertebrate Body. Molecular Biology of the Cell, 4th rev. ed., Garland Science, 2002, Chapter 21.
[3] Mišík M, Nersesyan A, Ropek N et al. Use of human derived liver cells for the detection of genotoxins in comet assays. *Mutat Res.* 2019 Sep; 845:402995.

[4] Sachinidis A, Cardiotoxicity and Heart Failure: Lessons from Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes and Anticancer Drugs. *Cells*. 2020 Apr. 9, 1001.
[5] Guo Y, Chu X, Parrott N J et al. Advancing Predictions of Tissue and Intracellular Drug Concentrations Using In Vitro, Imaging and PBPK Modeling Approaches. *Clin Pharmacol Ther*. 2018 Nov; 104(5): 865–889.
[6] Götte M, Hofmann G, Michou-Gallani A et al. An imaging assay to analyze primary neurons for cellular neurotoxicity. *Neurosci Methods*. 2010 Sep 30; 192(1):7-16.
[7] Bickle M. The beautiful cell: high-content screening in drug discovery. *Anal Bioanal Chem*. 2010 Sep; 398(1):219-26. Review.
[8] Horizon 2020 - Work Programme 2018-2020 Health, demographic change and wellbeing
[9] Lin-Gibson S, Sarkar S, Elliott JT. Summary of the National Institute of Standards and Technology and US Food And Drug Administration cell counting workshop: Sharing practices in cell counting measurements. *Cytotherapy*. 2018 Jun; 20(6):785-795.
[10] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition) JCGM 200:2012 (JCGM 200:2008 with minor corrections)
[11] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/
[12] Stephen L R Ellison. (2014). metRology: Support for metrological applications. R package version 0.9-17. http://CRAN.R-project.org/package=metRology
[13] Guide to the expression of uncertainty in measurement (GUM), BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, 1993 amended 1995
[14] EURACHEM/CITAC Guide: Quantifying Uncertainty in Analytical Measurement Second Edition Editors S L R Ellison (LGC, UK) M Rosslein (EMPA, Switzerland) A Williams (UK), QUAM:2000.P1
[15] Cox M G. The evaluation of key comparison data: An introduction. *Metrologia*. 2002 Dec 1, Vol 39-6.

Appendix 1: Reference material preparation, homogeneity and stability

Authors:
Laura Revel[1], Carla Divieto[1], Stefano Pavarelli [1], Roberto Bellotti [1], Mariapaola Sassi [1]
[1] Istituto Nazionale di Ricerca Metrologica (INRIM)

1. Overview

The document describes the preparation of the reference material of the *CCQM-CAWG Pilot study P(123): Number and geometric property of cells adhered to a solid substrate*. This procedure has been developed and adopted by INRIM to seed and stain the A549 cells on a defined area in order to produce a cell culture dish useful as reference material.

2. Dish preparation

2.1. Selecting the appropriate dish

Four different dishes, produced by IBIDI, were tested to select the best material for this study:

1. Dish Grid-500 with ibiTreat® coating (a physical treatment that optimize the adhesion of most cell types) (cod. 80156),
2. Dish Grid-500 uncoated (cod. 80151) and coated at INRIM with fibronectin,
3. Dish Grid-500 uncoated (cod. 80151) and coated at INRIM with poly(lysine),
4. Dish Grid-500 with glass bottom uncoated (cod. 81168).

A549 cells were seeded at the concentration of $1x10^5$/dish.
After an over-night adhesion, homogeneity, morphology and distribution of cells were evaluated: ibiTreat® coating and in the glass bottom dishes, the cell adhesion and morphology were better than in the others dishes. The dish Grid-500 with glass bottom showed less auto-fluorescence (see 4.3.).

2.2. IBIDI Dish Grid-500

The IBIDI Dish Grid-500 was selected as substrate for cells adhesion in the CCQM Pilot Study P(123).
This commercial cell culture dish has a glass bottom with 400 squares engraved, identified by letters and numbers. Each square has a nominal area of 0.25 mm^2 and an INRIM optical profilometer traceable to the International System of Units (SI) was used to validate this area measurement.

The dish has the following characteristics:

- μ-Dish^{35mm,low}, glass bottom, sterile
- Diameter dish = 35 mm
- Diameter observation area = 21 mm
- Lettered and numbered 4 x 100 squares (A-U; 1-20)
- Bottom made D 263 M Schott glass with a thickness of 170 μm ± 10 μm
- Grid and cells in one focal plane
- Low birefringence and auto-fluorescence
- Lid with lock position, which minimizes evaporation
- No effect on cells growth

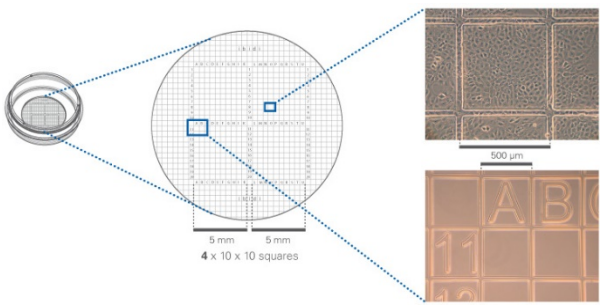


Fig 1. IBIDI dish layout and example images

2.3. Measurements of the dish square area

The area of the square was measured by the Optical Profilometer SensorFar model PLμ 2300. The x–y measurement system of the optical profilometer is traceable to the SI, as consequence the square area measurement provided below is traceable to the SI. The expanded uncertainty of the measurement, due to the standard deviation of the measurements and the instrumental resolution, was 9.5 x 10^{-3} mm^2. The expanded uncertainty of one measurement due to the instrumental resolution is 6 μm^2, that corresponds to a relative expanded uncertainty of the square area of about 0.03‰

The area of 15 squares of 3 different dishes was measured and the mean area of the 45 measurements was:

$$A_{mean} = (0.2172 \pm 0.0095)\ mm^2$$

3. Standard dish preparation

A549 were chosen among 4 cell lines for their adhesion rate and adhesion strength to the dish. The A549 cells are a human alveolar basal epithelial cell line deriving from a human lung carcinoma (characterized by the American Type Culture Collection, ATCC, reference number CCL-185) [1]. A549 cells are used as an *in vitro* model for a type II pulmonary epithelial cell model for drug metabolism. This cell line was initiated in 1972 by D.J. Giard, et al. through explant culture of lung carcinomatous tissue from a 58-year-old Caucasian male [2, 3]. Cells morphology stability and homogeneity have been investigated by flow cytometry.

Three cells density nominal values were chosen to cover a wide range of cell densities and confluency:

- from 40 to 70 cells/square (1x10^5 cells/dish) for low confluency (range: 15% - 30%),
- from 100 to 200 cells/square (2x10^5 cells/dish) for medium confluency (range: 40% - 60%),
- from 300 to 550 cells/square (4x10^5 cells per dish) for high confluency (range: 70% - 90%).

3.1. Cell culture

Reagents

- Roswell Park Memorial Institute (RPMI) cell culture medium (Sigma Aldrich, cod. R8758.)
- L-Glutamine 200 mM (Sigma Aldrich, cod. G7513)
- Kanamycin (Sigma Aldrich, cod. K1876)
- Fetal Bovine Serum (FBS) (Sigma Aldrich, cod. F2442)
- 0.25% trypsin/EDTA (Sigma Aldrich, cod. T4049)

Cells were grown in RPMI medium added with 1% L-Glutamine 200 mM, 1% Kanamycin and 10% FBS in 10 cm diameter Petri dishes at 37°C, 5% CO2 in a humidified incubator. In order to have the lowest number of proliferating cells and cell divisions in process, cells were detached when reached 80% confluency by using 0.25% trypsin/EDTA. After centrifugation, cells were resuspend in RPMI medium containing only 1% of FBS to avoid cell proliferation and counted by using a Neubauer cell counting chamber. The appropriate number of cells was seeded on the dishes that were placed in the incubator at 37°C, 5% CO2 over-night. Cell seeding was assessed by optical microscopy and the cells were labeled with two fluorescent dyes.

3.2. Cell staining

Reagents

- DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride) (Invitrogen, cod. D1306) (Ex=358 nm, Em=461 nm)
- Texas Red® c2 maleimide (Invitrogen cod. T6008) (Ex=595 nm, Em=615 nm)
- Paraformaldehyde 1% solution (Sigma-Aldrich, cod.15817)
- Triton-X100 0.1% solution in PBS 1x (Sigma-Aldrich, cod. T8532)
- MOWIOL (Sigma- Aldrich, cod. 324590)

Two fluorescent dyes were selected, based on two main factors: the dyes stability and the dyes ability to selectively stain the nucleus or the whole cell DAPI to stain nucleus and Texas Red to stain intracellular proteins.
Cells were fixed with 1% paraformaldehyde, incubated for 45 minutes with 0.5 μg/ml of Texas Red in 0.1% Triton-X100 and for 5 minutes with 0.5 μg/ml of DAPI, at room temperature and in the dark. A glass 20 mm diameter coverslip was mounted on the cell growing area of the dish by using 30 μl of warmed mounting medium MOWIOL. The dishes were placed overnight at 4 °C on a planar surface before imaging.

4. Testing the P123 reference materials

4.1. Cell distribution homogeneity

Several methods of seeding cells were tested to obtain the highest possible homogeneity in terms of cell distribution on the dish, both within the same square and among different squares. This is an important aspect of the preparation of dishes because the A549 cells tends to create cell aggregates. The homogeneity was measured through a tool developed for the confluency analysis. Several images was acquired crosswise the entire area of the dish, each of them covering a portion of the dish consisting in 24 squares. The confluency dispersion among the dish portions was 3 % for low confluency, 13 % for medium confluency and 3 % for high confluency. These data were indicative of the homogeneity of the entire dish, did not indicate any additional information on the confluency variability of the single squares to be measured by participant laboratories. The following table and graph report the data.

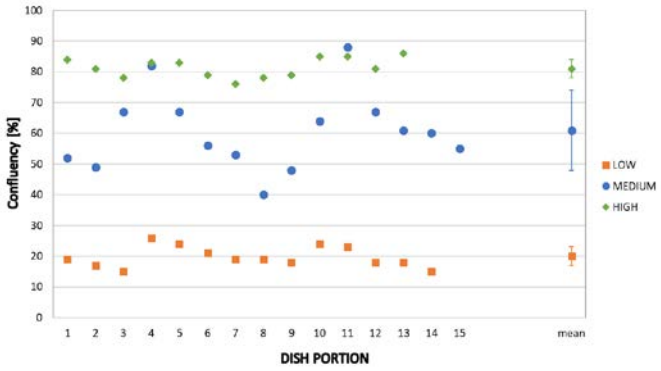


Fig 2. Homogeneity calculated for the three ranges of confluency

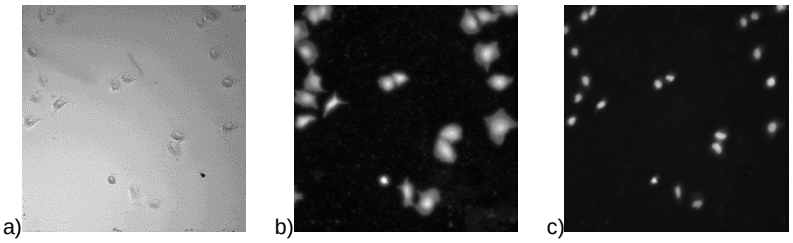
	Confluency [%]		
dish portion	Dish 1x10^5 cells	Dish 2 x10^5 cells	Dish 4 x10^5 cells
1	19	52	84
2	17	49	81
3	15	67	78
4	26	82	83
5	24	67	83
6	21	56	79
7	19	53	76
8	19	40	78
9	18	48	79
10	24	64	85
11	23	88	85
12	18	67	81
13	18	61	86
14	15	60	-
15	-	55	-
min	15	40	76
max	26	88	86
mean	20	61	81
Rel STD DEV	3	13	3

Table 1. Homogeneity calculated for the three ranges of confluency selected

4.2 Decay of fluorescent dyes brightness

To evaluate the brightness decay of fluorescent dyes during time, qualitative and quantitative evaluations were done. For the qualitative evaluation, images of a selected square (fig. 3) of a test dish were acquired and were evaluated soon after the preparation, after 6 and 12 months. The brightness did not show appreciable decay.

For the quantitative analysis, the same square (fig.3) of the test dish was analyzed in terms of mean fluorescence intensity (MFI), cell confluency and cell number, several times from the dish preparation till 7 months. Images of the square were acquired at 4 time points (0, 197, 200 and 215 days) and at two different value of exposure times (150 ms and 250 ms for Texas Red and 1,3 s and 1,7 s for DAPI). The MFI of Texas Red and of DAPI and the mean value of confluency and of cell number were evaluated over time. The 4 time points were selected to simulate the use of the dish as reference material in the long, short and medium term, respectively and at different temporal frequencies.



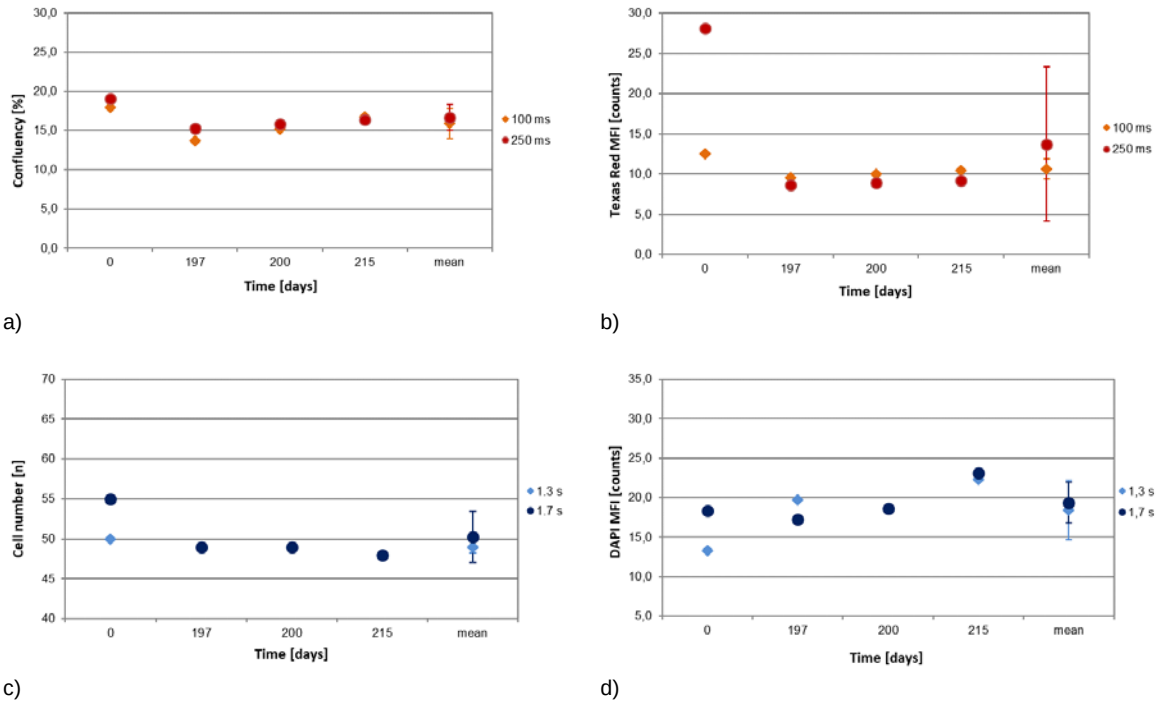
a) b) c)

Fig 3. Image of the square test acquired at time 0 day in phase contrast mode (a), Texas Red (b) and DAPI (c).

In table 2 and fig. 4 results of these evaluations are reported.

		confluency [%]	MFI [counts]
Texas Red	**100 ms**	15.9 (σ = 1.9)	10.7 (σ= 1.3)
	250 ms	16.6 (σ = 1.7)	13.7 (σ = 9.6)
		cell number [n]	**MFI [counts]**
DAPI	**1,3 s**	50 (σ = 3.2)	19.4 (σ = 2.6)
	1,7 s	49 (σ = 0.8)	18.5 (σ = 3.8)

Table 2. Results of fluorescent dyes brightness decay analysis. All the values are mean values over 7 months (σ = standard deviation).



a) b)

c) d)

Fig 4. Results of fluorescent dyes brightness decay analysis.
Confluency (a), MFI of Texas Red (b), cell number (c) and MFI of DAPI (d) are reported as mean values over time (7 months).

Results showed no significant differences in terms of confluency and cell number, despite the dyes have been investigated under stressed conditions. However, it should be noted that the exposure time used for the DAPI acquisition is greater by a order of magnitude compared to the values normally employed (~ 100 ms).
Another test related to the fluorescent dyes was considering the variations of confluency, cell number and mean fluorescence intensity of Texas Red and DAPI to vary the exposure time. The exposure time employed for Texas Red acquisition was 100 ms, 250 ms, 400 ms, 550 ms, 700 ms and 1000 ms; the exposure time employed for DAPI acquisition was 1.3 s, 1.7 s, 2.5 s, 3 s, 4 s and 5 s.

In table 3 and fig. 5 results of these analysis are reported: no significant differences were observed. For Texas Red, an encrease of MFI corresponded to an encrease of confluency; the same relationship was not evident for cell number and MFI of DAPI.

		confluency [%]	**MFI [counts]**
Texas Red	**day 200**	16.9 (σ = 1.6)	12.1 (σ = 3.5)
	day 215	17.8 (σ = 1.5)	13.2 (σ = 4.8)
		cell number [n]	**MFI [counts]**
DAPI	**day 200**	49 (σ = 0.6)	19.2 (σ = 1.0)
	day 215	48 (σ = 0.4)	22.9 (σ = 0.7)

Table 3. Results of fluorescent dyes brightness decay analysis. All the values are mean values at different exposure time

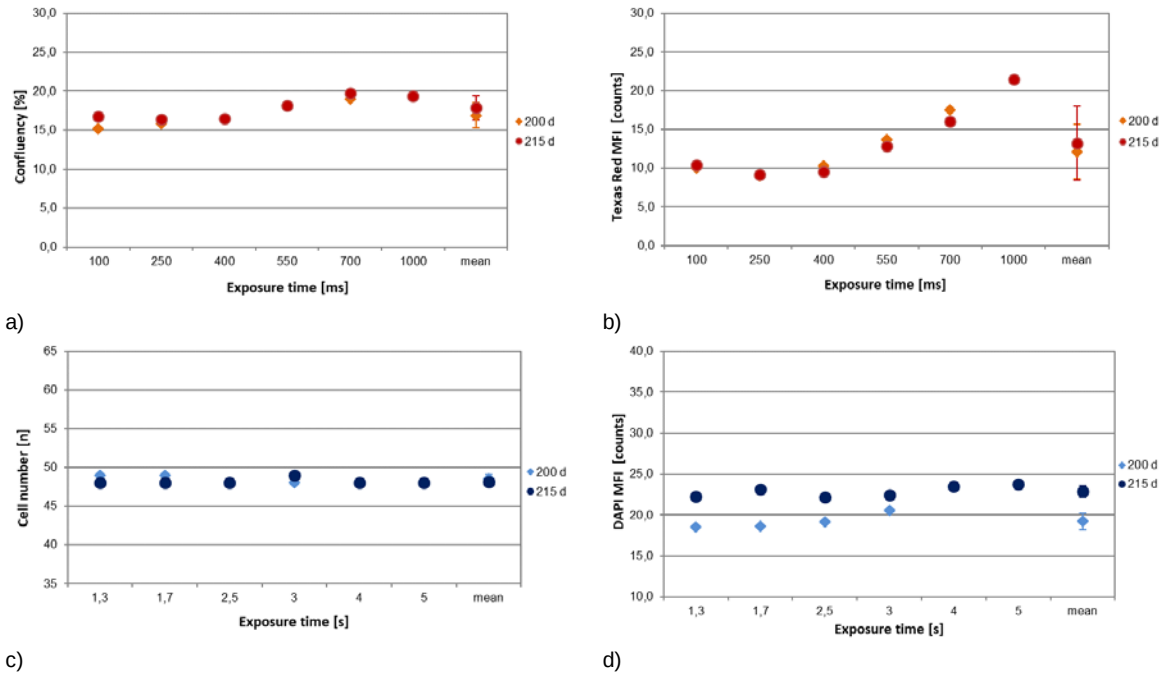


Fig 5. Results of fluorescent dyes brightness decay analysis.
Confluency (a), MFI of Texas Red (b), cell number (c) and MFI of DAPI (d) are reported as mean values at different exposure time.

4.3 Dish auto-fluorescence

To select the best dish between the ibiTreat® coating dish and the glass bottom dish, the auto-fluorescence of dishes was evaluated.

Images of two type of empty dishes (i.e. with no cells) were acquired at the following conditions:

- the square L10, placed in the center of the dish, was imaged,
- fluorescence lamp switched on 1 hour before the acquisition,
- fluorescence filter set to Texas Red wavelenghts,
- focus selected at the center of the dishes,
- exposure time of 500 ms.

The images acquired were cut excluding the squares edges to eliminate differences on the planar illumination, obtaining a square of 300x300 pixels. The images were analyzed with ImageJ software to obtain the histogram that represents the distribution of the fluorescence intensity of the pixel.

Histograms showed the glass bottom dish had a value of mean fluorescence half of the value of ibiTreat® dish.

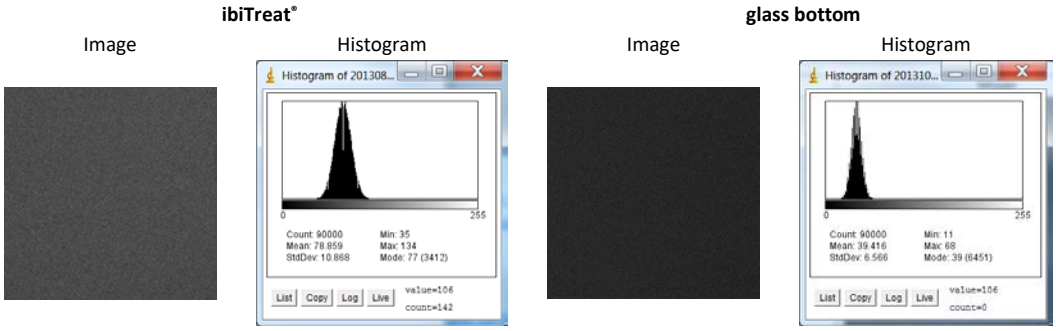


Fig 6. Images of L10 square and histogram of dishes with ibiTreat® coating and with glass bottom.

4.4 Mounting medium auto-fluorescence

The auto-fluorescence of several mounting media was evaluated in order to select the appropriate mounting media.

The mounting media compared were:

- Glycerol (Sigma-Aldrich cod. G2025)
- Mowiol (Sigma-Aldrich cod. P1763)
- Vectashield with DAPI (Vector Laboratories cod. H-1200)

Vectashield and Mowiol are glycerol-based mounting media. Glycerol-based media are aqueous mountant that do not solidify but remain viscous liquids after application, they inhibit photobleaching also after prolonged storage. Mowiol and Vectashield are enriched with anti-fade compounds and Vectashield is also enriched with DAPI stain.

For these test three ibiTreat® dishes without cells were mounted with 30 μl of each type of mounting media, the growth area were covered with glass coverslips of 20 mm diameter and the dishes were put overnight at 4 °C on a planar surface. Then the coverslips were sealed with transparent varnish and images were acquired at the microscope.

The images of dishes with mounting media were acquired at the same conditions, cut and analyzed as described for the auto-fluorescence evaluation (see 4.3).

Histogram showed that the mounting media had an auto-fluorescence intensity double respect to the auto-fluorescence of the empty dish with ibiTreat® coating (fig.6). Mowiol showed an intensity of auto-fluorescence 10% less than the others.

The mounting media MOWIOL was chosen as mounting medium for the P123 reference material because of its enrichment with anti-fade molecules, the absence of DAPI stain and the lowest value of auto-fluorescence.

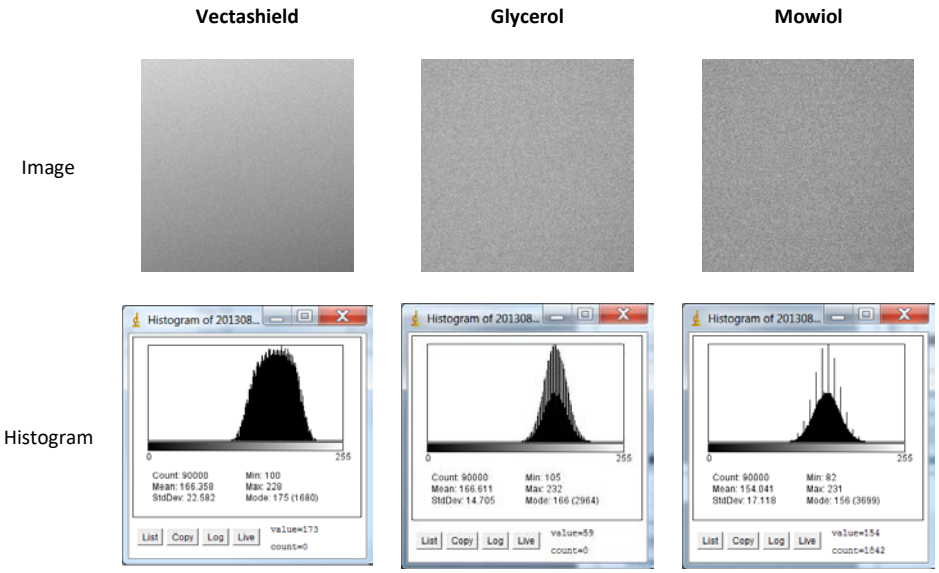


Fig 7. Images of the three dishes with different mounting media and histogram related

4.5 Stability and homogeneity of A549 cells

In order to evaluate the stability and homogeneity of the morphology of A549 cells, the physical parameters "Side scattered light" (SSC) and "Forward-scattered light" (FSC) have been measured by a CyAn ADP Cytofluorimeter (Beckman Coulter).

Aliquots of cells derived from the same batch and sub-cultured in standard conditions were fixed in 70% iced ethanol and stored for at least 24 hours at -20°C before the analysis.

The day of the analysis, the aliquots were removed from the –20°C and placed on ice. After centrifugation at 270 g and removal of the ethanol, the cells were allowed to re-hydrate in PBS 1X for 10 minutes at 4°C.

During that time, the flow cytometer was switched on and the lasers warmed up for at least 30 minutes; routine controls were used to assure the quality controls.

At least $1x10^4$ cells for each sample were acquired by flow cytometry and the physical parameters SSC and FSC were measured. The analysis of the data was performed through the software "Summit" (Beckman Coulter) and the medians of the FSC and the SSC were calculated. The cells kept in standard condition of culture and sub-culturing did not show significant changes in morphology in terms of both SSC, parameter related to cell granularity or internal complexity and FSC, parameter related to cell surface area or size.

5. Selection of the P123 measurement standards

It was decided that the P123 reference material was composed by three dishes at different cell density and confluency levels in order to cover the widest possible ranges of these measurands.

The three levels were selected: low, medium and high. At each level a range of confluency was associated and a corresponding range of cells per square was calculated. As consequence, the number of cells to seed per dish was calculated. Table 4 summarize the level and associated ranges.

Level	Confluency ranges	cells per square	n. of cells per dish (nominal)
LOW (l.c.)	10% - 30%	40 - 70	$1x10^5$
MEDIUM (m.c.)	40% - 60%	100 - 200	$2x10^5$
HIGH (h.c.)	70% - 90%	300 - 550	$4x10^5$

Table 4. Nominal values of confluency, cells per square and cells per dish for each P123 reference material

About 50 dishes were prepared, to select the best possible reference material from.

The selection of the P123 reference material responded to several criteria, described below:

- dyes brightness,
- dish cleaning (absence of PBS crystals, opaque zones, fractures on the bottom of the dishes),
- cells morphology (linearity of the margins and practically absence of branching),
- cells adhesion (very high percentage of well-adherent cells, almost total absence of non-adherent, rounded cells),
- homogeneity of cells distribution, at least in the area containing the grid of the squares.

Within each reference material (dish) selected for circulation, two squares for each quadrant (see fig.1), for a total of 8 squares for each dish, were selected by considering:

- homogeneity in the cells distribution on the square,
- cell morphology and adhesion,
- edges of the square, clearly distinguishable,
- appropriate cells number, with respect to the range in table 4.

Images of the 24 selected squares were acquired for Texas Red, DAPI and phase contrast.

These images were analyzed using ImageJ software and for each quadrant the best square were chosen: **the 12 selected squares have been named as P123 measurement standards**.

Summarizing, for each dish four squares were selected as measurement standards, one for each quadrant, for a total of **12 P123 measurement standards**. One square was selected as test square, positioned as far as possible from the measurement standards to avoid any damage (e.g. photobleaching) for the latters.

6. References

[1] www.lgcstandards-atcc.org/Products/Cells_and_Microorganisms/Cell_Lines/Human/Alphanumeric/CCL-185.aspx
[2] Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, Parks WP, “In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors.” J Natl Cancer Inst., vol. 51(5): 1417-23, November 1973
[3] Lieber M, Smith B, Szakal A, Nelson-Rees W, Todaro G, “A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells.”,Int J Cancer, vol. 17(1): 62-70, January 1976

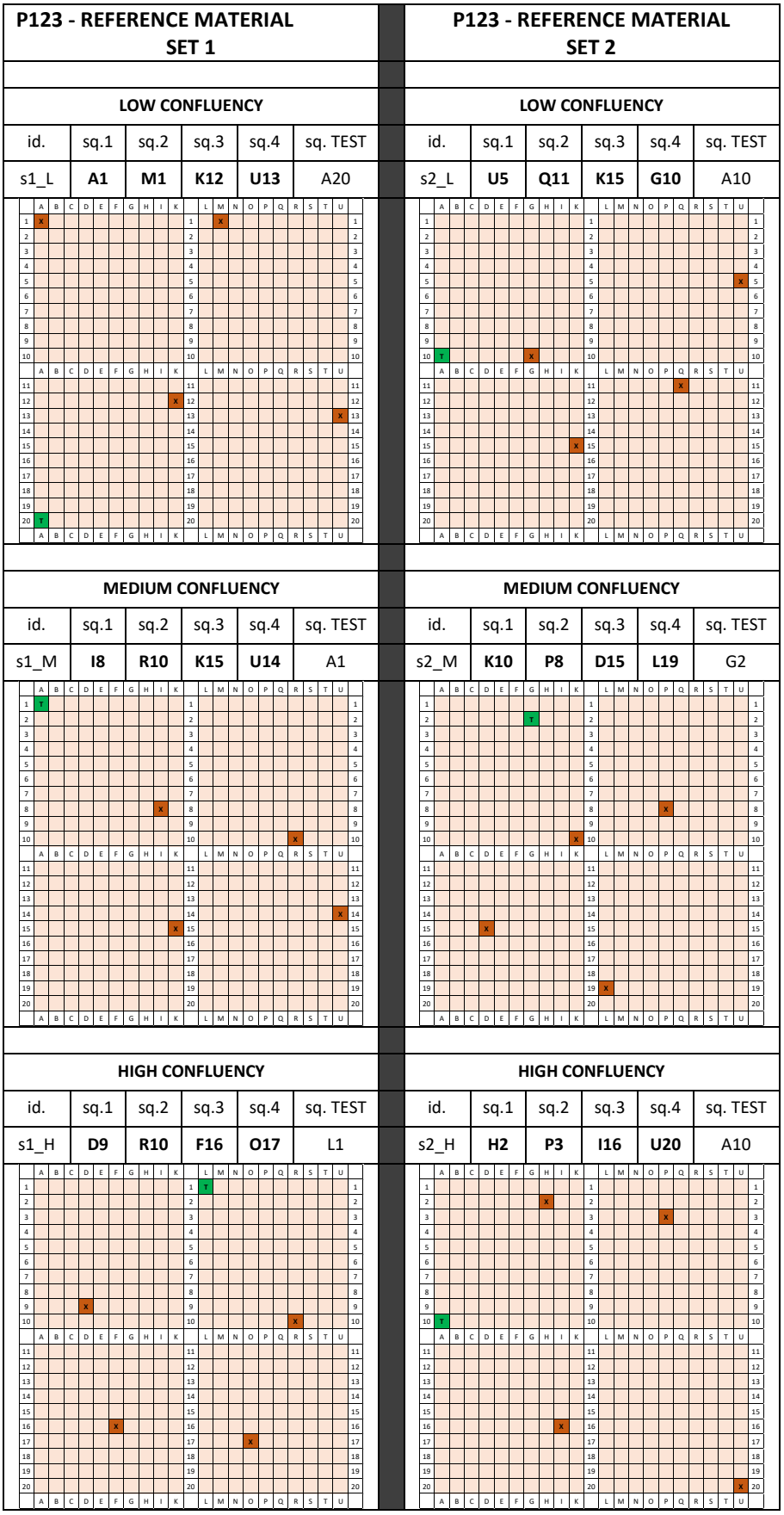


Fig. 8. P123 measurement standard selected for set 1 and set 2.

Appendix 2: study protocol

Appendix 2 summarizes the instructions and the technical protocol provided to the participants.
The main paragraphs of the study protocol are reported below.

Study aim and measurement protocol

This paragraph described the study aims and the measurement protocol that required to each laboratory:

- to acquire the appropriate images of 4 specific and defined measurement standards (squares) for each of the 3 reference materials (dishes), for a total of 12 images;
- to perform the following operations and suggestions:
  - to measure, for each square, the number of cells per area in a defined area (cell density),
  - to measure, for each square, the percentage area covered by the cells over the same total area (confluency fraction or cell confluency),
  - to calculate the average area of the cells, derived from the previous measurements (from the model: average area of the cells = confluency fraction x measured area / number of cells),
  - to provide an uncertainty budget for the three measurements,
  - to use a widefield fluorescence microscopy (suggestion),
  - to collect images in phase contrast or brightfield mode to identify the squares and to acquire both the square edges and the xy ruler to determine the area of measure,
  - to collect images in fluorescence mode for imaging DAPI stained nuclei (Ex=358 nm, Em=461 nm) and Texas Red® c2 maleimide stained cell proteins (Ex=595 nm, Em=615 nm),
  - to collect images with a 10x magnification objective and by a digital camera (a monochromatic camera is preferred) (suggestion),
  - to extract data from the images manually and/or with algorithms for image analysis software (a suggestion, based on an open source image analysis software package, was described in details),
  - to report the microscope setup, images filename, the acquisition and image analysis procedures on the provided spreadsheet named P123_datareporting_spreadsheet.xls. This spreadsheet will be used to link the image data to the analysis results.

Image acquisition

This paragraph described how to fulfill the following steps for image acquisition:

- **Microscope setting**;
- **Flat-field evaluation** (to evaluate variations in the lamp illumination, sensitivity of the detector, distortions in the optical path);
- **Squares recognition** by the use of phase contrast/brightfield mode. **SAFETY NOTE:** in order to avoid dyes quenching, the time of total exposure to lamp, with maximum possible attenuation, has **not to exceed 5 seconds**. **This rule is valid for each square acquisition**. A cleaning procedure of the dish, if necessary, was reported;
- **Exposure times setting,** to have the image histogram showing no more than 1% saturated pixels;
- **Collecting Images:** center the XY position on the phase contrast image, acquire the three (phase contrast, DAPI, Texas Red) images of the square and save them as non-compressed tiff file;
- **Pixel dimensions measurement**: each laboratory must use the provided **xy calibrated ruler** to check the measure of the image pixel

Image Analysis

This paragraph reported the instructions to analyze images manually and/or with algorithms for image analysis software.

The use of an open source image analysis software package called ImageJ, available from the NIH website (http://rsb.info.nih.gov/ij/), was reported as an example.

Area of measurement definition

This subparagraph was intended to give instructions on how to define the same area of measurement, in terms of region of the square and dimensions, between laboratories and for each image. The image of the provided XY calibrated ruler allowed to define the area of measurement and to calculate the pixel dimension. The nominal pitch of the X and Y scales was 25 µm. The length of the segment, nominal -225 µm ; +225 µm, was calibrated, both on the horizontal and on the vertical scale. The measured values x = (458 ± 0.7) µm and y = (458 ± 0.7) µm **correspond to the external edge of the scale lines** as shown in Figure 1.

The area of measurement defined for the comparison is *S* = (458 x 458) µm²

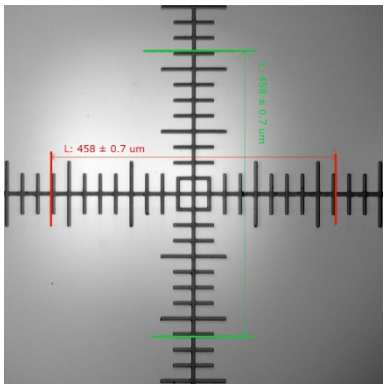


Figure 1. XY calibrated ruler image.

To have this area on the measurement standard images, instructions were done:

1. for each square, overlap the three acquired images (phase, Texas Red, DAPI)
2. center the ruler image on the overlapped square image
3. cut the defined area on the four-layers image
4. save the cut images as non-compressed tiff file

Uncertainty evaluation

This paragraph required that all participating laboratories:

- state the method and the mathematical model they apply for each measurements and their uncertainty budget established according to [1, 2]
- submit the mathematical model as an equation and define each of the variables included in it
- fill an uncertainty budget table according to the submitted model

Results report

This paragraph gave instructions to report the results (images and data) by filling the forms in annexes 3 - 6.

Annexes to the protocol

The protocol had 12 annexes, listed below:

- Annex 1: Reception of Standards
- Annex 2: Shipment of Standards
- Annex 3: Results report form – method employed for images acquisition
- Annex 4: Results report form – cell number calculation and uncertainty estimation
- Annex 5: Results report form - confluency and uncertainty estimation
- Annex 6: Results report form – mean area of cells and uncertainty estimation
- Annex 7: Standard preparation protocol
- Annex 8: Rules for handling the dishes
- Annex 9: Rules for cleaning the dish surface
- Annex 10: Example of uncertainty budget for the measurement
- Annex 11: P123 standard – square to be acquired
- Annex 12: Data reporting spreadsheet

References

[1] Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, 1995, Geneva, Switzerland.

[2] EURACHEM/CITAC Guide: Quantifying Uncertainty in Analytical Measurement Second Edition Editors S L R Ellison (LGC, UK) M Rosslein (EMPA, Switzerland) A Williams (UK), QUAM:2000.P1