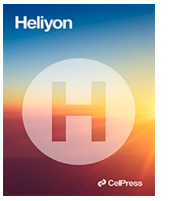
Heliyon 10 (2024) e24715

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Qualification of a flow cytometry-based method for the evaluation of *in vitro* cytotoxicity of GTA002 natural killer cell therapy

Monica Raimo [*,1], Alexandra G. Zavoianu [1], Wilma Meijs, Pascal Scholten, Jan Spanholtz

Glycostem Therapeutics, Kloosterstraat 9, 5349, AB Oss, the Netherlands

A R T I C L E I N F O

Keywords:
ATMP
Flow cytometry
Immunotherapy
Method qualification
Method validation
Quality control
Cell-based potency assay
Cytotoxicity assay
Natural killer cells
GMP manufacturing

A B S T R A C T

Background: Natural Killer (NK) cell-based therapies represent a ground-breaking opportunity for the treatment of solid tumors and hematological malignancies. NK cell manufacturing under good manufacturing practice (GMP) is complex and requires attentive assessment the product's safety and efficacy through quality control (QC). Release testing includes monitoring of *in vitro* cell expansion, differentiation, purity, phenotype, and cytotoxicity. As NK cells are biologically active products, the establishment of potency methods is particularly relevant; surrogate or improper assays can lead to rejection of qualifiable batches or to release of products that falsely meet potency specifications, potentially causing low efficacy during clinical trials. As cell-based therapeutics are highly heterogeneous, no universal guidelines for product characterization are available, and developers must invest significant effort in establishing and validating robust and fit-to-purpose assays. In this study, we describe the qualification procedure of a flow cytometry-based analytical method to assess *in vitro* potency of GTA002 NK cells, to be applied to oNKord®/inaleucel allogeneic off-the-shelf NK cell product from Glycostem Therapeutics, undergoing a Phase I/IIa clinical trial in acute myeloid leukemia (AML) patients (NCT04632316).

Methods: First, we established multi-color flow cytometry panels to quantitatively determine the count of effector (E) GTA002 cells and leukemia target (T) K562 cells alone and in co-culture at different E:T ratios (10:1, 3:1, 1:1). Effector potency was then qualitatively expressed as percentage of cytotoxicity. Next, we defined protocols for method qualification to assess the pivotal features of the assays, including accuracy, precision, linearity, range, specificity, robustness, and carryover; quantitative acceptance criteria were determined for all parameters. Results of the qualification procedure are reported and discussed against pre-defined acceptance criteria.

Results: Overall, our methods show robust performance across all parameters, ensuring QC-compliant assessment of NK cell potency as part of the release test panel for clinical batches. Notably, we identified relevant aspects to address when progressing towards method validation to support pivotal clinical studies.

Conclusions: This article provides a "case-study" of how analytical method development for cell therapeutics is planned and executed from early clinical stages, anticipating the need to establish robust procedures to overcome scientific and regulatory challenges during method validation.

* Corresponding author.
E-mail address: monica@glycostem.com (M. Raimo).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2024.e24715
Received 13 October 2022; Received in revised form 12 January 2024; Accepted 12 January 2024
Available online 17 January 2024

1. Introduction

The establishment of robust analytical methods for advanced therapy medicinal products (ATMPs) is essential to advance pre-clinical products to the clinical stage [1]. Analytical tests are developed to characterize the product through the manufacturing process and at the final stage, ensuring the safety and efficacy of each batch. Test results are compared to pre-established specifications to determine if a certain lot can be released for clinical use.

The translation process of a cell-based therapy to a commercial product follows the same path of small molecules, biopharmaceuticals or medical devices, starting from a proof of concept, providing evidence on safety and efficacy for a particular indication and then moving on to production in a good manufacturing practice (GMP)-compliant facility [2]. However, the intrinsic heterogeneity of patient-derived or allogeneic donor-derived materials poses a challenge to lot-to-lot consistency and to product standardization. Therefore, establishing analytical methods that ensure accurate, precise, and robust results is of paramount importance. Challenges of method development for cell therapy are the need to validate self-developed methods [3], and the lack of appropriate standard materials [4–6]. Guidance is provided by regulatory authorities, namely the European Medicines Agency (EMA), the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH), the United States Food and Drug Administration (FDA), the European Pharmacopeia (Ph. Eur.), the United States Pharmacopeia (USP) or the International Organization for Standardization (ISO). For the development and market authorization of ATMPs in the European Union, the "Guidelines on Good Manufacturing Practice specific to Advanced Therapy Medicinal Products", issued on November 24, 2017, advise that method validation can be applied gradually during clinical development. Safety-related, sterility and microbial assays should be validated before first in-human application. The suitability of analytical methods used to measure critical quality attributes (CQAs) should be established throughout clinical development, but full validation is required only at the stage of marketing authorization application. Analytical methods for batch release and stability testing, including potency assays, are instead expected to be validated prior to pivotal clinical trials [7].

After a method is developed, its qualification ensures that it is appropriate for its intended use, before proceeding with validation. Qualification requires acceptance limits and validation parameters (specificity, linearity, range, precision, accuracy, quantification, and detection limit, as applicable) to be determined [8]. Acceptance ranges are established according to product specifications and to the intended use of the assay. By meeting these criteria, the assay is deemed "fit for purpose" [9,10]. Although guidelines strongly recommend thorough characterization of all aspects of ATMPs, specific recommendations are lacking. Regulatory bodies acknowledge that potency evaluation is determined by the individual characteristics of each product and evaluate the adequacy of potency assays on a case-to-case basis, according to the manufacturing and analytical testing approach chosen [11]. Authorities can be consulted for scientific advice, but procedures must be established by product developers. As potency methods should ideally reflect the quality and the clinical efficacy of the product, definition of the "potency strategy" [11] is of paramount importance. Methods used for potency testing include biological assays, as *in vivo* animal studies, *in vitro* tissue or cell culture systems, and non-biological analytical assays, evaluating immunochemical, biochemical, and/or molecular attributes of the product [12].

This study describes the qualification of the *in vitro* flow cytometry-based cytotoxicity method for the assessment of the potency of GTA002 Natural Killer (NK) cells, to be applied to in-process and final product quality control (QC) during the GMP manufacturing of oNKord®/inaleucel NK cell therapy developed by Glycostem Therapeutics. oNKord® is currently evaluated in a Phase I/IIa clinical trial to treat acute myeloid leukemia (AML) patients (ClinicalTrials.gov Identifier: NCT04632316). The qualification protocol was designed based on the guideline from ICH (specifically Q2 (R1) [8]) for the validation of analytical procedures in regulated settings, including flow cytometry-based assays used for in-process testing, release, and stability assessment of immunotherapy products. The different assay parameters assessed, as well as the results of the qualification protocol, are presented in this article and discussed against pre-defined acceptance criteria. Considerations for future method validation, required before progression to Phase III, are also inferred. The flow cytometry method, the qualification protocol and the results presented in this article were part of the investigational medicinal product dossier (IMPD) application submitted to EMA to apply for clinical trial authorization.

2. Materials and methods

2.1. Counting beads

123count eBeads™ Counting Beads from Invitrogen (Thermo Fisher Scientific, Waltham, MA, USA, cat. no. 01-1234-42) were used as surrogate material for the qualification of the flow cytometry-based methods for the enumeration of effector and target cells. One lot of beads (E127050, with a concentration of 1,009,000 particles/ml) was used over four days of testing.

2.2. Effector cells

Hematopoietic stem cell-derived GTA002 NK cells were manufactured during the development campaign of the NK cell product oNKord®, according to internal proprietary procedures at Glycostem Therapeutics (Oss, NL). Two batches, lot GSV19-006 (development batch) and lot GSV19-010 (engineering batch), were used for the assay development and qualification procedures. After harvest, GTA002 cells were cryopreserved according to Glycostem's standard operating procedures; before use, cells were thawed at 37 °C in a water bath, diluted 1:10 and maintained in cell culture medium, composed of Glycostem basal growth medium (GBGM, FertiPro, Beernem, BE, cat. no. CCT-CLIN-500-T) supplemented with 2 % human serum (HS, Sanquin, Amsterdam, NL, cat. no. B0618). During co-culture with target cells, GTA002 cells were supplemented with cytokines interleukin (IL)-2 at 1000 international units (IU)/

ml (Proleukin®, Clinigen, DE, cat. no. 05060229220264) and IL-15 at 0.02 μg/ml (Sartorius CellGenix, Freiburg im Breisgau, DE, cat. no. 1013–050). A new vial of GTA002 cells was thawed on every test day.

2.3. Target cells

The human chronic myelogenous leukemia cell line K562 was obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA, cat. no. CCL-243). A mycoplasma-free and short tandem repeat (STR)-authenticated master cell bank was established after pre-staining with pacific blue succinimidyl ester (PBSE, Thermo Fisher, cat. no. P-10163) at a concentration of 0.012 mg/ml. Pre-stained K562 cells were aliquoted at 5×10^6 cells/ml and cryopreserved in CryoStor® CS10 medium (BioLife Solutions, Bothell, WA, USA, cat. no. 210102). On every day of testing, one vial was thawed at 37 °C in a water bath, diluted and maintained in Iscove's modified Dulbecco's medium (IMDM) cell culture medium, supplemented with L-glutamine, 25 mM HEPES (Lonza, Basel, CH, cat. no. BE12-722F) and 10 % heat-inactivated fetal bovine serum (FBS) (Gibco, Thermo Fisher, cat. no. 10500–064).

2.4. Sample preparation for flow cytometry

123count eBeads™ Counting Beads were diluted to the appropriate concentrations, expressed as beads/test volume, in 200 μl of phosphate buffer saline (PBS, Lonza, cat. no. 17-516Q).

From GTA002 cell suspensions (with varying concentrations), 50 μl were collected and stained with the appropriate fluorochrome-labelled antibody mix (39 μl) for 15 min at room temperature (RT) in the dark. The final volume was filled to 200 μl with PBS before sample acquisition. The staining mix included the lymphocyte marker CD45 (anti-CD45-Krome Orange, clone J.33, isotype mouse IgG1k, cat. no. B36294) and the 7-amino-actinomycin D (7-AAD) viability dye (cat.no. A07704), both from Beckman Coulter (Marseille, FR). Clones and dyes were chosen based on a panel designed during assay development.

From K562 target cell suspensions (with varying concentrations), 50 μl were collected and stained with 4 μl of 7-AAD for 15 min at RT in the dark. The final volume was filled to 200 μl with PBS before sample acquisition.

For co-culture assays, GTA002 and K562 cells were plated in a flat bottom 96-well plate at three effector (E) to target (T) cells ratios (E:T), 10:1, 3:1 and 1:1, including controls of effector cells and target cells alone. The final co-culture medium contained effector cell culture medium and target cell culture medium at a 1:1 ratio in all wells, for a total of 100 μl per well. Cells were co-cultured overnight (O/N) at 37 °C, in a humidified atmosphere with 5 % CO_2. The next day, samples were moved to a V-bottom 96-well plate, stained with 4 μl of 7-AAD, for 15 min at RT in the dark, then acquired.

2.5. Sample acquisition and data analysis

2.5.1. Equipment and system suitability testing (SST)

All flow cytometry assays were performed using the CytoFLEX S flow cytometer from Beckman Coulter, equipped with a 96-well plate loader. All equipment was validated and released for use by quality assurance (QA). The flow cytometer detectors were configurated as follows: 6 fluorescence channels for the 488 nm Blue laser (50 mW), namely SSC (488/8 nm), FITC (525 nm), PE (585/42 nm), ECD (610/20 nm), PC5.5 (690/50 nm) and PC7 (780/60 nm); 4 channels for the 405 nm Violet laser (80 mW), namely PB450 (450/45 nm), KO525 (525/40 nm), Violet-610 (610/20 nm) and Violet-660 (660/20 nm); and 3 channels for the 638 nm Red laser (50 mW), namely APC (660/20), APC-A700 (712/25 nm) and APC-A750 (780/60 nm). The peristaltic pump-based fluidics system used for sample injection allowed for volumetric cell counting without dilution effect (absolute counting), dependent on the regular calibration of the flow rate. Two flow cytometers were used (instrument A and B). Instrument A was used for the qualification of the enumeration of GTA002 cells and instrument B was used for the qualification of the enumeration of K562 cells and of the potency method. Both instruments were used for the qualification using counting beads (4 runs total, 2 per device). CytoFlex Daily QC fluorospheres (Beckman Coulter, cat. no. B53230) were used for daily verification of the optical alignment and fluidics system of the flow cytometers, as SST. The test was performed after instrument start up and was considered valid for a 24-h period. A maximum of three verification tests per day were allowed; if the check failed all three attempts, any subsequent acquisition would be invalidated.

2.5.2. General acquisition settings

Detector gain settings were set to "Recommended" and threshold to "Automatic" for all tests, to use the instrument's QC settings based on the daily SST. Laminar flow conditions were maintained with the Cytoflex Sheath Fluid (Beckman Coulter, cat. no. B51503) and all samples were acquired from 96-well plates with V-bottom (Corning, Amsterdam, NL, cat. no. 3897). Acquisition of one well was considered equivalent to one test ($n = 1$).

2.5.3. Software

The CytExpert software (Beckman Coulter, version 2.3), with Electronic Record Management installed, was used for instrument operation control and raw data acquisition, visualization, and processing. Standard templates were developed for each method, with saved acquisition settings and fixed gates on dot plots to reduce variability. No changes on settings or gates were performed between assays or runs. Compensation matrices were established during method development and saved as separate files via the CytExpert software on the Cytoflex S instruments. For every method, the compensation was always applied at the start of sample acquisition.

2.5.4. Sample acquisition workflow

For every counting beads sample, a volume of 200 µl PBS was acquired in consecutive technical triplicates, at a flow rate of 60 µl/min; beads were identified as double-positive events in the FITC and PE channels. Compensation was not necessary.

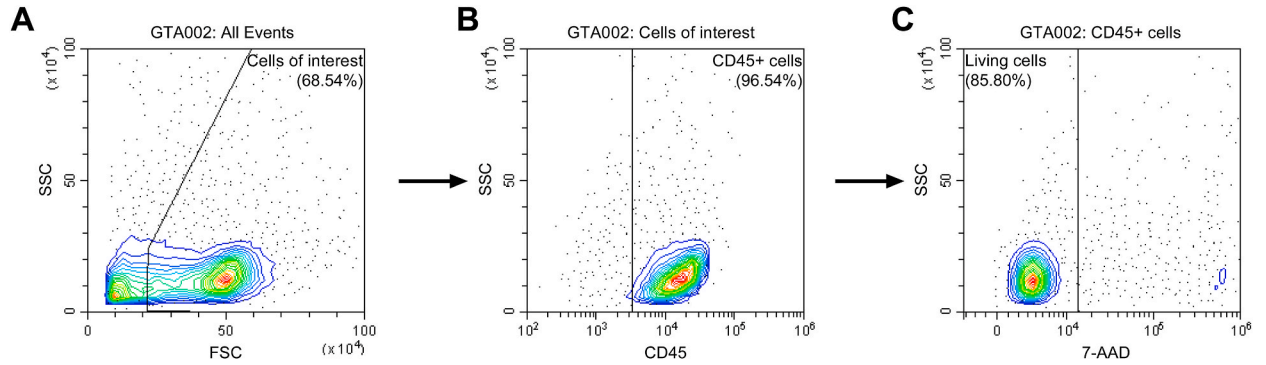


Fig. 1. – Gating strategy for the enumeration of GTA002 effector cells via flow cytometry. **(A)** GTA002 cells, stained and acquired with CytoFLEX S, are identified in the FSC versus SSC plot ("All Events"). The cell population is separated from debris and dying cells in the "Cells of interest" gate. **(B)** From "Cells of interest", surface antigen characterization further identifies CD45-expressing lymphocytes in the CD45 versus SSC plot, "CD45+ cells" gate. **(C)** Viability analysis of "CD45+ cells" in the 7-AAD versus SSC plot separates the living, 7-AAD-negative, lymphocytes in the gate "Living cells". Such sub-population of $CD45^{+}/7\text{-}AAD^{-}$ cells is used to determine the total GTA002 count. The CD45 fluorochrome is detected in the KO525 channel and 7-AAD in the PC5.5 channel. All gates are pre-defined and are unaltered between samples and runs. FSC: forward scatter; SSC: side scatter; CD45: cluster of differentiation 45; 7-AAD: 7-amino-actinomycin D.

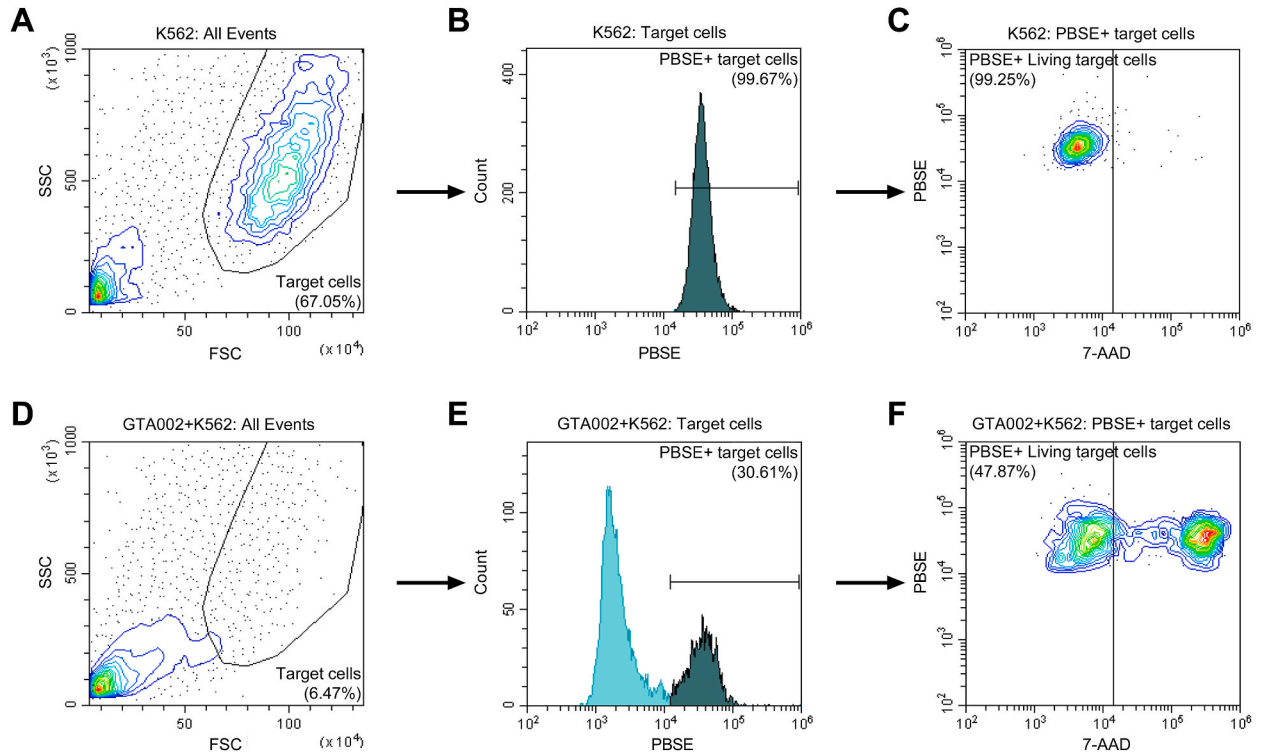


Fig. 2. – Gating strategy for the enumeration of K562 target cells and for the assessment of GTA002 potency via flow cytometry. **(A-C)**: K562 target cells alone; **(D–F)**: Effector GTA002 + target K562 cells, after overnight co-culture at a 10:1 E:T ratio. All gates are pre-defined based on the target-only sample and are unaltered between samples and runs. **(A)** and **(D)** From "All events", the target cell population of interest is separated from effector cells, debris and dying cells in the FSC vs SSC plot, in the "Target cells" gate. **(B)** and **(E)** From "Target cells", PBSE-stained K562 are selected in the PBSE vs Count histogram as "PBSE + target cells". **(C)** and **(F)** Viable target cells are then identified by plotting 7-AAD against PBSE, in the "PBSE + Living target cells" gate. Such sub-population of $PBSE^{+}/7\text{-}AAD^{-}$ cells is used to determine the total K562 count. The PBSE fluorochrome is detected in the PB450 channel and 7-AAD is detected in the PC5.5 channel. FSC: forward scatter; SSC: side scatter; PBSE: pacific blue succinimidyl ester; 7-AAD: 7-amino-actinomycin D. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

For the enumeration of GTA002 cells, 50 μl of 200 μl of stained sample were acquired in consecutive technical triplicates, at a flow rate of 120 μl/min, using the set acquisition template with pre-defined gates (determined during assay development phase) to identify the negative and positive populations for the analyzed antigens. First, the population of interest was identified in the forward scatter (FSC) vs side scatter (SSC) plot (Fig. 1A). From there, CD45-expressing lymphocytes were gated (Fig. 1B). The total living cell count was determined as the absolute number of CD45-expressing, 7-AAD negative ($CD45^+/7\text{-}AAD^-$) cells, identified in the "Living cells" gate (Fig. 1C). The cell concentration, defined as total cells per ml of culture, was obtained by multiplying the average number living cells (from the technical replicates) by the dilution factor (80), as shown in Equation 1 (E1). This method is considered quantitative.

$$(E1)\ Living\ CD45^+cells\ /\ ml = (average\ CD45^+\ /\ 7AAD^-events \times 80).$$

For the enumeration of K562 target cells alone or after co-culture with effector cells, 50 μl of a 200 μl stained sample were acquired in consecutive technical triplicates, at a flow rate of 120 μl/min, using the set acquisition template with pre-defined gates (determined during assay development phase). The negative and positive populations for the antigens of interest were identified for the target-only samples (Fig. 2A–C) or after co-culture (Fig. 2D–F). The total living target cell count was determined as the absolute number of PBSE-positive, 7-AAD-negative cells ($PBSE^+/7\text{-}AAD^-$), identifiable in the "$PBSE^+$ living target cells" gate (Fig. 2C and F) and multiplied by the dilution factor (80) (Equation 2 (E2)):

$$(E2)\ Living\ targetcells\ /\ ml = (average\ PBSE^+\ /\ 7AAD^-\ events \times 80).$$

As the effector cell enumeration, this assay is considered quantitative.

For the co-culture, $PBSE^+/7\text{-}AAD^-$ cells (Fig. 2C and F) were used to calculate the potency of the effector cells for every E:T ratio, expressed as cytotoxicity (%) (Equation 3 (E3)):

$$(E3)\ Cytotoxicity(\%) = 100 - \left(\frac{number\ of\ PBSE^+\ living\ target\ events\ in\ co-culture}{number\ of\ PBSE^+\ living\ target\ events} \times 100\right).$$

The determination of cytotoxicity is considered a qualitative measurement, as such variable is not a measurable quantity of the effector cells, and its value does not strictly depend on the input quantity in the assay, but rather on the analyte's biological properties, which are qualitative [5].

2.5.5. Statistics

Results from all samples and all runs were analyzed using Microsoft Excel; all data were double-checked to confirm validity. Raw values were rounded up if the last digit was 5–9, or down if 0–4. Calculations and statistical analyses were determined for each parameter according to the qualification plan and are described in Table 2. GraphPad Prism (v.9.0.1) was used to generate plots and graphs for data visualization.

Table 1
Overview of the parameters tested during method qualification. List and description of the parameters assessed for the qualification of the quantitative enumeration of GTA002 effector and K562 target cells and the qualitative determination of GTA002 cytotoxicity. Source: ICH Topic Q2 (R1) Validation of Analytical Procedures: Text and Methodology [8].

Parameter	Description
Accuracy	The closeness of agreement between the value which is accepted either as conventional true value or an accepted reference value and the value found.
Precision	The closeness of agreement between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions. It can be considered at three levels: repeatability, intermediate precision, and reproducibility. Repeatability expresses the precision under the same operating conditions over a short interval of time (intra-assay precision). Intermediate precision expresses within-laboratories variations: different days, different analysts, different equipment, etc. (inter-assay precision).
Linearity	The ability of an analytical procedure (within a given range) to obtain test results which are directly proportional to the concentration (amount) of analyte in the sample.
Range	The interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy, and linearity.
Limit of quantification (LoQ)	The lowest (low limit of quantification, LLoQ) or the highest (high limit of quantification, HLoQ) amount of analyte in a sample which can be quantitatively determined with suitable precision and accuracy.
Specificity	The ability to assess unequivocally the analyte in the presence of components which may be expected to be present, typically including impurities, degradants, matrix, etc.
Robustness	A measure of the capacity of an analytical procedure to remain unaffected by small, but deliberate variations in method parameters, providing an indication of its reliability during normal usage.
Sample stability	The capability of a sample to retain the initial properties of a measured constituent for a period of time within specified limits when the sample is stored under defined conditions.
Carryover	The residual level of a substance that is detectable in a sample where that substance is absent (i.e., a blank sample), if it is analyzed immediately after a sample containing that substance.

Table 2
Protocol for the qualification of the flow cytometry-based method for the evaluation of *in vitro* potency of GTA002 NK cells. Tests are presented per parameter and per analyte. The samples used, the predefined acceptance criteria and the formulas used for calculations are described.

Parameter	Analyte	Description	Acceptance criteria	Calculation formulas
Accuracy	Counting beads	123count eBeads™ were diluted to 8 different concentration levels C_1–C_8: 2.5×10^3, 5.0×10^3, 7.5×10^3, 1.0×10^4, 2.5×10^4, 5.0×10^4, 7.5×10^4 and 1.0×10^5 beads per 200 µl test volume, then acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on two flow cytometers (N = 4 runs, N = 12 replicates per level). Accuracy was expressed as Recovery (%) of the theoretical count, calculated from the average of triplicate values measured for each level and each run. The average Recovery (%) was calculated as the mean of the recovery values from the 4 runs, for every sample level.	Recovery (%): 80 %–120 %	(*E4*) $Recovery\ (\%) = \frac{Measured\ bead\ count_{C_{1-8}}}{Theoretical\ bead\ count_{C_{1-8}}} \times 100$, where the measured bead count is the mean of single bead events measured per sample $C_{1\text{-}8}$ (n = 3) and the theoretical bead count is the expected single bead count per sample $C_{1\text{-}8}$.
Precision - Repeatability	Counting beads	123count eBeads™ were diluted to 8 different concentration levels C_1–C_8: 2.5×10^3, 5.0×10^3, 7.5×10^3, 1.0×10^4, 2.5×10^4, 5.0×10^4, 7.5×10^4 and 1.0×10^5 beads per 200 µl test volume, then acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on two flow cytometers (N = 4 runs, N = 12 replicates per level). Repeatability was expressed as the coefficient of variation CV(%), calculated from the triplicate values for each level and each run. Results were calculated per day and per analyst and reported as *min* and *max CV*(%) (n = 3 tests per sample level). The average CV(%) was calculated as the mean of the CV(%) values from the 4 runs, for every sample level.	CV(%) ≤ 25 %	(*E5*) $SD_{C_{1-8}} = \sqrt{\frac{\sum_1^3 (x_{1-3} - \overline{x_{C1-8}})^2}{N-1}}$, where $SD_{C_{1-8}}$ is the standard deviation of $N = 3$ sample measurements , x_{1-3} are the triplicate values per sample level C_{1-8} per run and $\overline{x_{C_{1-8}}}$ is the mean of triplicate measurements for each sample level C_{1-8}; then, (*E6*) $CV_{C_{1-8}}\ (\%) = \frac{SD_{C_{1-8}}}{\overline{x_{C_{1-8}}}} \times 100$, where $CV_{C_{1-8}}(\%)$ is the coefficient of variation for triplicate values of each sample level C_{1-8}.
	GTA002 cells	Undiluted GTA002 cells were stained and acquired in consecutive technical sextuplicate (2 runs) or triplicates (2 runs), on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 18 replicates). Repeatability was expressed as the coefficient of variation CV(%), calculated from the sextuplicate/triplicate values for each run. Results were calculated per day and per analyst and reported as *min* and *max CV*(%) (n = 6 or 3 tests per run).	CV (%) ≤ 30 %	(*E7*) $SD_{C_0} = \sqrt{\frac{\sum_1^6 (x_{1-6} - \overline{x_{C_0}})^2}{N-1}}$, where SD_{C_0} is the standard deviation of $N = 6$ or $N = 3$ sample measurements , x_{1-6} is the result of each sextuplicate value (or 1–3, for triplicates), and $\overline{x_{C_0}}$ is the mean of sextuplicate (or triplicate) values of the undiluted sample C_0; then, (*E8*) $CV_{C_0}\ (\%) = \frac{SD_{C_0}}{\overline{x_{C_0}}} \times 100$, where $CV_{C_0}(\%)$ is the coefficient of variation for sextuplicate or triplicate values of the undiluted sample C_0.
		GTA002 cells were diluted to 8 different concentration levels C_1–C_8: 1.25×10^4, 2.5×10^4, 5.0×10^4, 2.5×10^5, 5.0×10^5, 1.0×10^6, 2.0×10^6, and 4.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). Repeatability was expressed as the coefficient of variation CV(%), calculated from the triplicate values for each level and each run. Results were calculated per day and per analyst only and reported as *min* and *max CV*(%) (n = 3 tests per sample level). The average CV(%) was calculated as the mean of the CV (%) values from the 4 runs, for every sample level.		(*E9*) $SD_{C_{1-8}} = \sqrt{\frac{\sum_1^3 (x_{1-3} - \overline{x_{C1-8}})^2}{N-1}}$, where $SD_{C_{1-8}}$ is the standard deviation of $N = 3$ sample measurements , x_{1-3} are the triplicate values results per sample level C_{1-8} per run and $\overline{x_{C_{1-8}}}$ is the mean of triplicate values for each sample level C_{1-8}; then, (*E10*) $CV_{C_{1-8}}\ (\%) = \frac{SD_{C_{1-8}}}{\overline{x_{C_{1-8}}}} \times 100$, where $CV_{C_{1-8}}(\%)$ is the coefficient of variation for triplicate values of each sample level C_{1-8}.
	K562 cells	K562 cells were diluted to 6 different concentration levels C_1–C_6: 4.0×10^4, 8.0×10^4, 2.0×10^5, 4.0×10^5, 8.0×10^5, and 2.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). Repeatability was expressed as the coefficient of variation CV(%), calculated from the triplicate values for each level and each run. Results were calculated per day and per analyst only and reported as *min* and *max CV*(%) (n = 3 tests per sample level). The average CV(%) was calculated as the mean of the CV (%) values from the 4 runs, for every sample level.	CV (%) ≤ 30 % for C_1–C_2 CV (%) ≤ 20 % for C_3–C_6	(*E11*) $SD_{C_{1-6}} = \sqrt{\frac{\sum_1^3 (x_{1-3} - \overline{x_{C_{1-6}}})^2}{N-1}}$, where $SD_{C_{1-6}}$ is the standard deviation of $N = 3$ sample measurements , x_{1-3} are the triplicate values per sample level C_{1-6} per run and $\overline{x_{C_{1-6}}}$ is the mean of triplicate values for each sample level C_{1-6}; then, (*E12*) $CV_{C_{1-6}}\ (\%) = \frac{SD_{C_{1-6}}}{\overline{x_{C_{1-6}}}} \times 100$, where $CV_{C_{1-6}}(\%)$ is the coefficient of variation for triplicate values of each sample level C_{1-6}.

(*continued on next page*)

Table 2 (*continued*)

Parameter	Analyte	Description	Acceptance criteria	Calculation formulas
	Co-cultured GTA002 and K562 cells	GTA002 and K562 cells were co-cultured overnight at different E:T ratios (10:1, 3:1 and 1:1), stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per ratio). Raw values were used to calculate the % cytotoxicity (Equation 3 (E3)). Repeatability was expressed as the coefficient of variation CV(%), calculated from the % cytotoxicity triplicate values for each E:T ratio and each run. Results were calculated per day and per analyst only and reported as *min* and *max CV*(%) (n = 3 tests per E:T ratio). The average CV(%) was calculated as the mean of the CV(%) values from the 4 runs, for every E:T ratio.	CV (%) ≤ 50 %	$(E13)\ SD_{ratio} = \sqrt{\frac{\sum_1^3 (x_{1-3} - \overline{x_{ratio}})^2}{N-1}}$, where SD_{ratio} is the standard deviation of $N = 3$ sample measurements per E:T ratio , x_{1-3} are the triplicate values per E:T ratio and $\overline{x_{ratio}}$ is the mean cytotoxicity % of triplicate values for each E:T ratio; then, $(E14)\ CV_{ratio}\ (\%) = \frac{SD_{ratio}}{\overline{x_{ratio}}} \times 100$, where $CV_{ratio}(\%)$ is the coefficient of variation for triplicate values of each sample E:T ratio.
Precision - Intermediate precision	Counting beads	123count eBeads™ were diluted to 8 different concentration levels C_1–C_8: 2.5×10^3, 5.0×10^3, 7.5×10^3, 1.0×10^4, 2.5×10^4, 5.0×10^4, 7.5×10^4 and 1.0×10^5 beads per 200 µl test volume, then acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on two flow cytometers (N = 4 runs, N = 12 replicates per level). Intermediate precision was expressed as the overall coefficient of variation Overall CV(%), calculated for each level across all runs, from the mean and the standard deviation of the 4 mean values obtained for each run.	Overall CV (%) ≤ 30 %	$(E15)\ SD_{C_{1-8}} = \sqrt{\frac{\sum_1^4 (x_{1-4} - \overline{x_{C1-8}})^2}{N-1}}$, where $SD_{C_{1-8}}$ is the standard deviation of $N = 4$ mean values from the 4 runs , x_{1-4} is the mean of the triplicate values from each run, and $\overline{x_{C_{1-8}}}$ is the mean of the 4 mean values for each sample level C_{1-8} over the 4 runs; then, $(E16)\ Overall\ CV_{C_{1-8}}\ (\%) = \frac{SD_{C_{1-8}}}{\overline{x_{C_{1-8}}}} \times 100$, where $Overall\ CV_{C_{1-8}}(\%)$ is the overall coefficient of variation for the 4 runs of each sample level C_{1-8}.
	GTA002 cells	Undiluted GTA002 cells were stained and acquired in consecutive technical sextuplicate (2 runs) or triplicates (2 runs), on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 18 replicates). Intermediate precision was expressed as the overall coefficient of variation Overall CV(%), calculated across all runs, from the mean and the standard deviation of the 4 mean values obtained for each run.	Overall CV (%) ≤ 40 %	$(E17)\ SD_{C_0} = \sqrt{\frac{\sum_1^4 (x_{1-4} - \overline{x_{C_0}})^2}{N-1}}$, where SD_{C_0} is the standard deviation of $N = 4$ mean values from the 4 runs, x_{1-4} is the mean of the sextuplicate/ triplicate values form each run, and $\overline{x_{C_0}}$ is the mean of the 4 mean values for the undiluted sample C_0 over the 4 runs; then, $(E18)\ Overall\ CV_{C_0}\ (\%) = \frac{SD_{C_0}}{\overline{x_{C_0}}} \times 100$, where $Overall\ CV_{C_0}(\%)$ is the overall coefficient of variation for the 4 runs of the undiluted sample C_0.
		GTA002 cells were diluted to 8 different concentration levels C_1–C_8: 1.25×10^4, 2.5×10^4, 5.0×10^4, 2.5×10^5, 5.0×10^5, 1.0×10^6, 2.0×10^6, and 4.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). Intermediate precision was expressed as the overall coefficient of variation Overall CV(%), calculated for each level across all runs, from the mean and the standard deviation of the 4 mean values obtained for each run.		$(E19)\ SD_{C_{1-8}} = \sqrt{\frac{\sum_1^4 (x_{1-4} - \overline{x_{C1-8}})^2}{N-1}}$, where $SD_{C_{1-8}}$ is the standard deviation of $N = 4$ mean values from the 4 runs , x_{1-4} is the mean of the triplicate values from each run, and $\overline{x_{C_{1-8}}}$ is the mean of the 4 mean values for each sample level C_{1-8} over the 4 runs; then, $(E20)\ Overall\ CV_{C_{1-8}}\ (\%) = \frac{SD_{C_{1-8}}}{\overline{x_{C_{1-8}}}} \times 100$, where $Overall\ CV_{C_{1-8}}(\%)$ is the overall coefficient of variation for the 4 runs of each sample level C_{1-8}.
	K562 cells	K562 cells were diluted to 6 different concentration levels C_1–C_6: 4.0×10^4, 8.0×10^4, 2.0×10^5, 4.0×10^5, 8.0×10^5, and 2.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). Intermediate precision was expressed as the overall coefficient of variation Overall CV(%), calculated for each level across all runs, from the mean and the standard deviation of the 4 mean values obtained for each run.	Overall CV (%) ≤ 30 % for C_1–C_2 Overall CV (%) ≤ 25 % for C_3–C_6	$(E21)\ SD_{C_{1-6}} = \sqrt{\frac{\sum_1^4 (x_{1-4} - \overline{x_{C_{1-6}}})^2}{N-1}}$, where SD_{1-6} is the standard deviation of $N = 4$ mean values from the 4 runs , x_{1-4} is the mean of the triplicate values from each run, and $\overline{x_{C_{1-6}}}$ is the mean of the 4 mean values for each sample level C_{1-6} over the 4 runs; then, $(E22)\ Overall\ CV_{C_{1-6}}\ (\%) = \frac{SD_{C_{1-6}}}{\overline{x_{C_{1-6}}}} \times 100$ where $Overall\ CV_{C_{1-6}}(\%)$ is the overall coefficient of variation for the 4 runs of each sample level C_{1-6}.
	Co-cultured GTA002 and K562 cells	GTA002 and K562 cells were co-cultured overnight at different E:T ratios (10:1, 3:1 and 1:1), stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per ratio). Raw values were used to calculate the % cytotoxicity (Equation 3 (E3)). Intermediate precision was expressed as the overall coefficient of variation	Overall CV (%) ≤ 50 %	$(E23)\ SD_{ratio} = \sqrt{\frac{\sum_1^4 (x_{1-4} - \overline{x_{ratio}})^2}{N-1}}$, where SD_{ratio} is the standard deviation of $N = 4$ mean % cytotoxicity values from the 4 runs, x_{1-4} is the mean cytotoxicity % from the triplicate values from each run, $\overline{x_{ratio}}$ is the mean of the % cytotoxicity from the 4 mean values for each E:T ratio over the 4 runs;

(*continued on next page*)

Table 2 (*continued*)

Parameter	Analyte	Description	Acceptance criteria	Calculation formulas
		Overall CV(%), calculated from the mean and the standard deviation of the 4 mean values obtained as % cytotoxicity for each ratio.		then, (*E*24) $Overall\ CV_{ratio}\ (\%) = \frac{SD_{ratio}}{\overline{x}_{ratio}} \times 100$, where $Overall\ CV_{ratio}(\%)$ is the overall coefficient of variation for the 4 runs of each E:T ratio.
Linearity	Counting beads	123count eBeads™ were diluted to 8 different concentration levels C_1–C_8: 2.5×10^3, 5.0×10^3, 7.5×10^3, 1.0×10^4, 2.5×10^4, 5.0×10^4, 7.5×10^4 and 1.0×10^5 beads per 200 μl test volume, then acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on two flow cytometers (N = 4 runs, N = 12 replicates per level). For each run, linear regression analysis was performed by plotting the average of the measured count (number of events in the gate of interest) against the theoretical count (beads/test volume), for each level. The slope, y-intercept, coefficient of determination (R^2) and the residual sum of squares (RSS) for both x and y axes were determined.	$R^2 \geq 0.80$	(*E*25)RSS (x axis) $= \sum_{n=1}^{8}(x_{C_{1-8}} - \widehat{x_{C_{1-8}}})^2$; (*E*26) RSS (y axis) $= \sum_{n=1}^{8}(y_{C_{1-8}} - \widehat{y_{C_{1-8}}})^2$, where $x_{C_{1-8}}$ *and* $y_{C_{1-8}}$ are the mean of the triplicate values for each sample level C_{1-8} and $\widehat{x_{C_{1-8}}}$ *and* $\widehat{y_{C_{1-8}}}$ are the values predicted by the linear model; and (*E*27) $R^2 = 1 - \frac{\sum_{n=1}^{8}(y_{C_{1-8}} - \widehat{y_{C_{1-8}}})^2}{\sum_{n=1}^{8}(y_{C_{1-8}} - \overline{y_{C_{all}}})^2}$, where R^2 is the coefficient of determination, $\overline{y_{C_{all}}}$ is the mean of all means of triplicate values for every sample level C_{1-8}.
	GTA002 cells	GTA002 cells were diluted to 8 different concentration levels C_1–C_8: 1.25×10^4, 2.5×10^4, 5.0×10^4, 2.5×10^5, 5.0×10^5, 1.0×10^6, 2.0×10^6, and 4.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). For each run, linear regression analysis was performed by plotting the average of the measured count (number of events in the gate of interest) against the theoretical count (beads/test volume), for each level. The slope, y-intercept, coefficient of determination (R^2) and the residual sum of squares (RSS) for both x and y axes were determined.	$R^2 \geq 0.80$	(*E*28)RSS (x axis) $= \sum_{n=1}^{8}(x_{C_{1-8}} - \widehat{x_{C_{1-8}}})^2$; (*E*29) RSS (y axis) $= \sum_{n=1}^{8}(y_{C_{1-8}} - \widehat{y_{C_{1-8}}})^2$, where $x_{C_{1-8}}$ *and* $y_{C_{1-8}}$ are the mean of the triplicate values for each sample level C_{1-8} and $\widehat{x_{C_{1-8}}}$ *and* $\widehat{y_{C_{1-8}}}$ are the values predicted by the linear model; and (*E*30) $R^2 = 1 - \frac{\sum_{n=1}^{8}(y_{C_{1-8}} - \widehat{y_{C_{1-8}}})^2}{\sum_{n=1}^{8}(y_{C_{1-8}} - \overline{y_{C_{all}}})^2}$, where R^2 is the coefficient of determination, $\overline{y_{C_{all}}}$ is the mean of all means of triplicate values for every sample level C_{1-8}.
	K562 cells	K562 cells were diluted to 6 different concentration levels C_1–C_6: 4.0×10^4, 8.0×10^4, 2.0×10^5, 4.0×10^5, 8.0×10^5, and 2.0×10^6 cells/ml, then stained and acquired in consecutive technical triplicates, on four days, by two analysts (two days per analyst), on one flow cytometer (N = 4 runs, N = 12 replicates per level). For each run, linear regression analysis was performed by plotting the average of the measured count (number of events in the gate of interest) against the theoretical count (beads/test volume), for each level. The slope, y-intercept, coefficient of determination (R^2) and the residual sum of squares (RSS) for both x and y axes were determined.	$R^2 \geq 0.80$	(*E*31)RSS (x axis) $= \sum_{n=1}^{6}(x_{C_{1-6}} - \widehat{x_{C_{1-6}}})^2$; (*E*32) RSS (y axis) $= \sum_{n=1}^{6}(y_{C_{1-6}} - \widehat{y_{C_{1-6}}})^2$, where $x_{C_{1-6}}$ *and* $y_{C_{1-6}}$ are the mean of the triplicate values for each sample level C_{1-6} and $\widehat{x_{C_{1-6}}}$ *and* $\widehat{y_{C_{1-6}}}$ are the values predicted by the linear model; and (*E*33) $R^2 = 1 - \frac{\sum_{n=1}^{6}(y_{C_{1-6}} - \widehat{y_{C_{1-6}}})^2}{\sum_{n=1}^{6}(y_{C_{1-6}} - \overline{y_{C_{all}}})^2}$, where R^2 is the coefficient of determination, $\overline{y_{C_{all}}}$ is the mean of all means of triplicate values for every sample level C_{1-6}.
Range	Counting beads, GTA002 cells and K562 cells	The range was determined from the linearity analysis, as the interval between the lowest and the highest level of analyte to be detected in the linear range that meets accuracy and precision acceptance criteria.	Acceptable level of accuracy and precision within the linear range	Not applicable.
Lower limit of quantification (LLoQ)	Counting beads, GTA002 cells and K562 cells	The LLoQ was determined from the range, as the lowest level of analyte in the linear range meeting the accuracy and precision acceptance criteria.	Lowest acceptable level of accuracy and precision within the linear range	Not applicable.
Higher limit of quantification (HLoQ)	Counting beads, GTA002 cells and K562 cells	The HLoQ was determined from the range, as the highest level of analyte in the linear range meeting the accuracy and precision acceptance criteria.	Highest acceptable level of accuracy and precision within the linear range	Not applicable.
Specificity	Counting beads	Blanks (PBS) were acquired in consecutive technical triplicates before the acquisition of the bead samples in the linearity tests on four days, by two analysts (two days per analyst), on two flow cytometers (N = 4 runs, N = 12 replicates).	≤10 event counts from blank	Not applicable.

(*continued on next page*)

Table 2 (*continued*)

Parameter	Analyte	Description	Acceptance criteria	Calculation formulas
		Specificity was evaluated by considering the highest blank count over the 4 runs as the test result.		
	GTA002 cells	GTA002 cells were stained with the appropriate antibodies ("Stained") or incubated with PBS ("Unstained") and acquired in consecutive technical triplicates, on three days, by two analysts, on one flow cytometer (N = 3 runs, N = 9 replicates per sample). Specificity was evaluated by considering the highest count from the Unstained sample over the 3 runs as the test result.	≤100 events from Unstained	Not applicable.
	K562 cells	Different amounts of K562 cells were spiked with effector GTA002 cells at multiple E:T ratios (N = 4 conditions, 90:1, 45:1, 9:1, 1.8:1), or non-spiked (N = 4 conditions). A "blank" sample of GTA002 cells (E:T 100:0) was used as a control (N = 1 condition). Samples were stained and acquired in consecutive technical triplicates on one day, by one analyst, on one flow cytometer (N = 1 run, N = 3 tests per condition). Specificity was evaluated [1] by considering the highest count of target cells from the blank sample as the test result, and [2] by calculating the % difference (Δ(%)) in K562 count between non-spiked and spiked samples for each E:T ratio.	[1] ≤ 10 events from "blank" E:T 100:0 [2] Δ(%) ≤ 30 % (Non-spiked compared to Spiked, for each E:T ratio)	[1] Not applicable. [2] $(E34)\quad \Delta_{K562_{ratio}}(\%) = = \frac{\left\lvert Non - spiked_{K562_{ratio}} - Spiked_{K562_{ratio}} \right\rvert}{Spiked_{K562_{ratio}}} \times 100,$ where $Non - spiked_{K562}$ and $Spiked_{K562}$ are the mean of the triplicate measurements of K562 cells in samples not spiked or spiked with effector cells, and Δ_{K562} (%) is the percentage difference between K562 count in spiked versus non-spiked samples, for each E:T ratio.
Robustness	GTA002 cells, K562 cells and co-cultured GTA002 and K562 cells	GTA002 or K562 cells were stained for the standard (STD) time, 15 min, or for ± 10 min, i.e., for 5 or 25 min. Samples were acquired in consecutive technical triplicates, on one day, by one analyst on one flow cytometer (N = 1 runs, N = 3 replicates per test). GTA002 and K562 cells were co-cultured overnight at 10:1 E:T ratio, stained for the standard (STD) time, 15 min, or for ± 10 min, i.e., for 5 or 25 min. Samples were acquired in consecutive technical triplicates, on one day, by one analyst on one flow cytometer (N = 1 runs, N = 3 replicates per test). Robustness was expressed as the % of difference in cell count (Δ(%)) between each test and the standard condition.	For GTA002 cells: Δ(%) ≤ 30 % For K562 and co-cultured cells: Δ(%) ≤ 20 % (Test compared to STD, for 5 or 25 min)	$(E35)\ \Delta(\%) = \left\lvert 100 - \frac{\overline{x_{test}}}{\overline{x_{STD}}} \times 100 \right\rvert,$ where $\overline{x_{test}}$ is the mean of the triplicate values for the test condition (5 min or 25 min), and $\overline{x_{STD}}$ is the mean of the triplicate values for the standard condition.
Sample stability	GTA002 cells	After thawing, GTA002 cells were immediately stained (standard condition, STD), or stored for 30 or 60 min at 37 °C before staining. Samples were acquired in consecutive technical triplicates, on one day, by one analyst on one flow cytometer (N = 1 runs, N = 3 replicates per test). Sample stability was expressed as % of difference in cell count (Δ(%)) between each test and the standard condition.	Δ(%) ≤ 25 % (Test compared to STD, for 30 or 60 min)	$(E36)\ \Delta(\%) = \left\lvert 100 - \frac{\overline{x_{test}}}{\overline{x_{STD}}} \times 100 \right\rvert,$ where $\overline{x_{test}}$ is the mean of the triplicate values for the test condition (30 min or 60 min), and $\overline{x_{STD}}$ is the average of the triplicate values for the standard condition.
Carryover	Counting beads, GTA002 cells and K562 cells	Blanks (PBS) were acquired in consecutive technical triplicates immediately after the acquisition of the highest level of analyte in the linearity tests, by two analysts, on four days (two days per analyst), on one or two (for beads) flow cytometer (N = 4 runs, N = 12 tests). Carryover (CO) was assessed as the % of the number of events from the analyte sample measured from the blank sample. Results were calculated per day and per analyst only and reported as *min* and *max CO*(%) (n = 3 tests per run).	CO (%) ≤ 1 %	$(E37)\ CO\ (\%) = \frac{x_{blank}}{\overline{x_{sample}}} \times 100,$ where x_{blank} is the mean of the triplicate values for the blank sample and $\overline{x_{sample}}$ is the mean of triplicate values for the analyte sample (highest level from linearity tests).

2.6. Method qualification

Method validation, necessary to progress to pivotal stages of clinical development, is often preceded by qualification, to prove that the method is fit-for-purpose and to pre-define suitable ranges and acceptance criteria for the observed results. To qualify the suitability of the flow cytometry method to determine the potency of GTA002 cells for oNKord® product batch release, protocols were developed in house, defining the parameters to be assessed, the respective acceptance criteria and the test samples to be used. Tests were established as quantitative or qualitative, according to the aim and to the nature of the sample material. First, different qualification procedures were established for quantitative (enumeration of effector and target cells) and qualitative (evaluation of cytotoxicity) methods. Then, the appropriate test parameters were defined, as summarized in Table 1. Quantitative methods were assessed for accuracy, precision (repeatability and intermediate precision), linearity, range, lower and higher limit of quantification (LLoQ and HLoQ), specificity, robustness, sample stability and carryover (CO). Limit of detection (LoD) was outside the scope of the qualification, as the application of the enumeration is intended for a definite range of values, based on the cell content of the product or of the assay. The qualitative potency method was tested for repeatability and intermediate precision. For the enumeration methods, counting beads, GTA002 and K562 cells (in different preparations) were used; for the potency method, different ratios of E:T co-cultures were used.

Detailed description of samples, tests, acceptance criteria and calculation formulas are given for each method in Table 2. Data are presented and discussed by parameter in the Results section.

2.6.1. Method qualification protocol for the enumeration of GTA002 effector cells (quantitative determination)

The qualification of the enumeration of effector GTA002 cells was performed assessing multiple parameters with 123count eBeads™ Counting Beads diluted to a range of 8 different concentrations (beads/volume), or GTA002 cells undiluted or diluted at 8 concentrations (cells/ml) as test samples. Due to the lack of reference material for flow cytometry assays [4], counting beads with a size of 7 μm, similar to the 6–7 μm reported size of NK cells [13], were used as surrogate materials to determine accuracy, precision, linearity, range, LLoQ and HLoQ, specificity and CO. The synthetic nature of beads does not make them suitable to evaluate robustness and sample stability. GTA002 cells could not be used to determine accuracy, as there is no standard preparation of such cell population with a true count value. Undiluted cell preparations, however, were used for precision, specificity, robustness, and sample stability. Additionally, a diluted range of cell concentrations was used to assess precision, linearity, range, LLoQ and HLoQ and CO. Runs were performed by two operators on different days and on flow cytometer A or B (for beads) or A (for GTA002), as indicated for each run. A summary of the qualification protocol is given in Table 2.

2.6.2. Method qualification protocol for the enumeration of K562 target cells (quantitative determination)

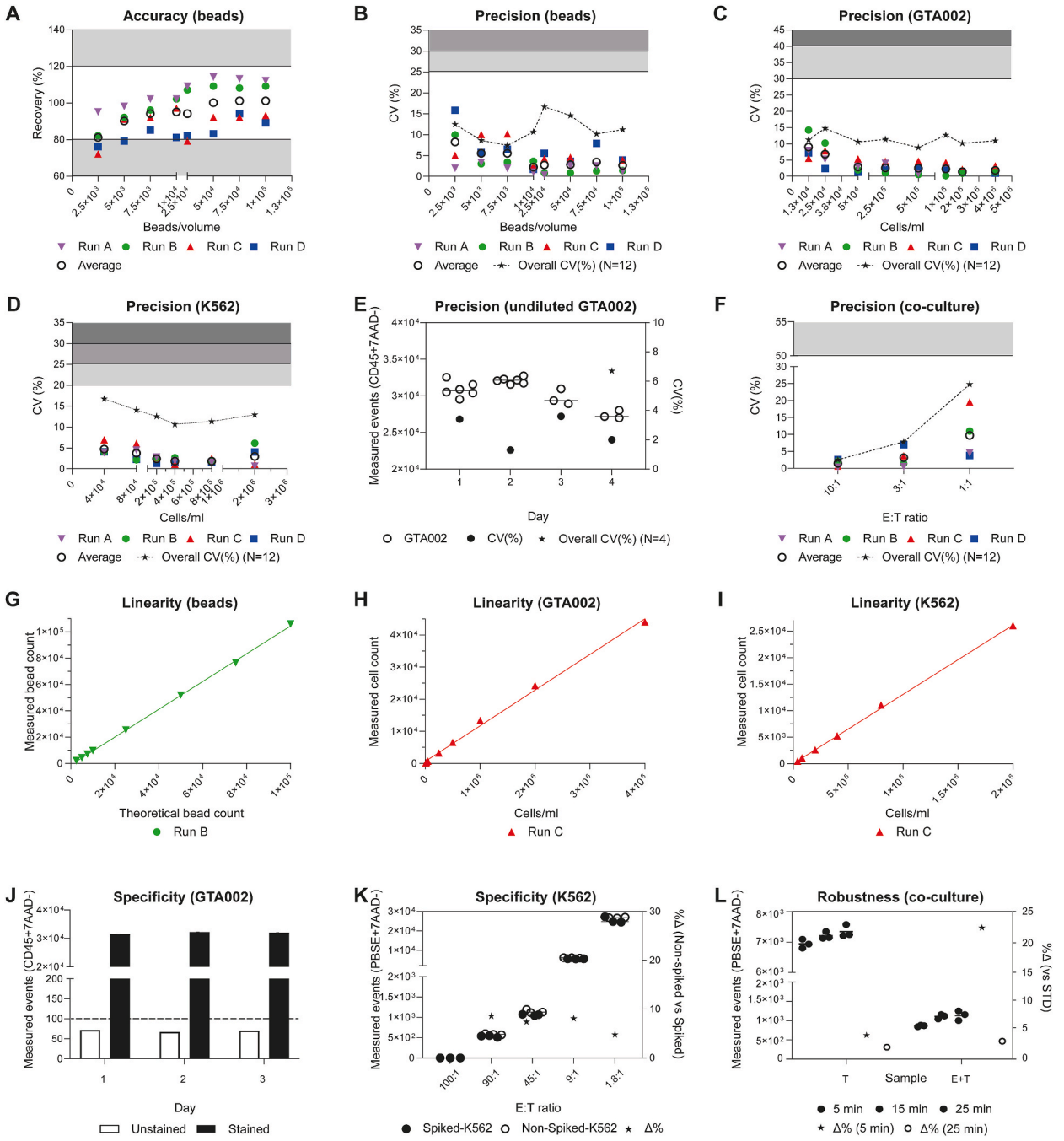
The qualification of the enumeration of target K562 cells was established in a similar manner as for GTA002 cells. The use of count beads surrogate material was described earlier. PBSE-stained K562 cells were diluted to a range of 6 different concentrations (cells/ml) to assess precision, linearity, range, LLoQ and HLoQ and CO. Specificity was tested with target cells alone (non-spiked) or spiked with GTA002 cells at multiple ratios, while robustness was evaluated with target cells alone or at a 10:1 E:T ratio. Runs were performed by two operators, on different days, on flow cytometer B. The detailed procedure is summarized in Table 2.

2.6.3. Method qualification protocol for the determination of potency from co-culture of effector with target cells (qualitative determination)

The evaluation of the cytotoxicity of effector GTA002 cells after co-culture with target cells depends on donor-dependent, variable, and uncontrollable biological activity. During incubation, the ratio of living effector to target cells is continuously changing, until reaching the end-point enumeration of remaining viable target cells. This part of the potency assay is not linear, i.e., it does not show proportionality between number of seeded cells and the final value calculated as cytotoxicity, and the exact interference of dead and dying cells during co-culture on the endpoint result cannot be evaluated. The determination of cytotoxicity relies mainly on the ability of the potency method to correctly enumerate remaining viable target cells, and the only step that could interfere with it would be the transfer of samples from the co-culture plate to the V-bottom staining and acquisition plate after co-culture. Therefore, the qualitative nature of the method is not compatible with the assessment of accuracy, linearity, and other parameters, however it is crucial to assess the precision of the method to determine target cell count after co-culture. The reliability of the method was evaluated with 3 different E:T ratios (10:1, 3:1 and 1:1). Tests were performed by two operators, on different days, on flow cytometer B. Table 2 describes the details of the qualification protocol.

3. Results

The qualification procedure of the flow cytometry-based methods for GTA002 effector and K562 target cell enumeration, and for the evaluation of GTA002 cytotoxicity, is summarized in Table 2. Results are presented per test parameter, in comparison with the predefined acceptance criteria, and visualized in Fig. 3. Cell enumeration results are reported for GTA002, including testing of the counting beads (Fig. 3A, B, C, E, G, H and J), for K562 (Fig. 3D, I, K), and for co-cultured effectors and targets (Fig. 3 F and L).



(caption on next page)

Fig. 3. – Qualification of the enumeration of effector GTA002 and target K562 cells and of the assessment of the potency of GTA002 cells. Qualification was performed by assessing accuracy, precision, linearity, specificity, and robustness with 123count eBeads™ Counting Beads, GTA002 cells and K562 cells. For accuracy (A), precision (B) and linearity (G) with beads, 123count eBeads™ Counting Beads were diluted at 8 different levels C_1–C_8 (2.5×10^3 -1.0×10^5 beads/test volume), then acquired by flow cytometry (n = 3 technical replicates, 4 runs, 4 days, 2 analysts, 2 flow cytometers). For precision (C) and linearity (H) with effector cells, GTA002 cells were diluted at 8 different levels C_1–C_8 (1.25×10^4-4.0×10^6 cells/ml), stained, then acquired by flow cytometry (n = 3 technical replicates, 4 runs, 4 days, 2 analysts, 1 flow cytometer). Cells of interest were defined as $CD45^+$/7-AAD^-. For precision (D) and linearity (I) with target cells, K562 cells were diluted at 6 levels C_1–C_6 (4.0×10^4-2.0×10^6 cells/ml), stained, then acquired by flow cytometry (n = 3 technical replicates, 4 runs, 4 days, 2 analysts, 1 flow cytometer). Cells of interest were defined as $PBSE^+$/7-AAD^-. **(A)** Accuracy was assessed with the 8 levels of beads and expressed as Recovery (%). The light grey background marks areas outside of acceptance criteria. **(B, C, D)** Precision was assessed with the 8 levels of beads, 8 levels of GTA002 cells and 6 levels of K562 cells, as repeatability or as intermediate precision per level. Repeatability was determined as the CV(%) for each run. Intermediate precision was determined as the Overall CV(%) across the 4 runs (black stars). Areas outside of acceptance criteria are marked in grey. **(E)** Precision was determined with undiluted GTA002 cells, stained and acquired (n = 6 or n = 3 technical replicates, 4 runs, 4 days, 2 analysts, 1 flow cytometer). Cells of interest were defined as $CD45^+$/7-AAD^-. **(F)** Precision was analyzed with co-cultured samples at three E:T ratios (10:1, 3:1, 1:1). After overnight incubation, samples were stained and acquired (n = 3 technical replicates, 4 runs, 4 days and 2 analysts, 1 flow cytometer). Target cells were identified as $PBSE^+$/7-AAD^-. The area outside acceptance criteria is marked in light grey. **(G, H, I)** Linearity was determined by linear regression analysis with the 8 levels of counting beads **(G)**, of GTA002 cells **(H)** and the 6 levels of K562 cells **(I)**, using counts derived from the precision analysis. The theoretical count was plotted against the measured count for each level and each analyte, and the linear trendline equation and coefficient of determination R^2 were calculated. Representative runs are shown. **(J)** Specificity of GTA002 enumeration was assessed by acquiring an "Unstained" sample (white bars) and a "Stained" sample (black bars) (n = 3 technical replicates, 3 runs, 3 days, 2 analysts, 1 flow cytometer). Positive events were determined as $CD45^+$/$7AAD^-$. **(K)** Specificity of K562 enumeration was investigated by staining and acquiring K562 target cell samples spiked with different amounts of effector GTA002 cells, in 4 E:T ratios: 90:1, 45:1, 9:1, 1.8:1. A "blank" sample, only containing effector cells (E:T 100:0), was included as a control (n = 3 technical replicates, 1 analyst, 1 run). Target cells were identified as $PBSE^+$/7-AAD^-. **(L)** Robustness of the potency method was assessed as influence of sample staining time. K562 cells alone (T), or after overnight co-culture with GTA002 in a 10:1 ratio (E + T), were stained for 15 min, and with a variation of ±10 min, for 5 or 25 min. Target cells were identified as $PBSE^+$/7-AAD^-. The difference from the 15-min standard condition were expressed as Δ%. 7-AAD: 7-amino-actinomycin D; PBSE: pacific blue succinimidyl ester; CV: coefficient of variation; PBS: phosphate buffer saline. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.1. Accuracy

As described earlier, true accuracy of the quantification of cell populations can only be assessed with the aid of a well characterized and homogenous standard or reference material, which is currently not available for the flow cytometry analysis of cellular products. Therefore, 123count eBeads™ Counting Beads were used as surrogate material to determine the accuracy of the detection system. Counting beads were diluted to 8 different concentration levels, C_1–C_8, i.e., 2.5×10^3, 5.0×10^3, 7.5×10^3, 1.0×10^4, 2.5×10^4, 5.0×10^4, 7.5×10^4 and 1.0×10^5 beads per 200 μl test volume; each dilution was acquired in consecutive technical triplicates on four different days, by two analysts, on flow cytometers A and B. Accuracy was calculated as Recovery (%) of the theoretical number of beads, measured for every run and for every bead level (Table 2, Equation 4 (E4)). The acceptance range was defined as 80–120 % recovery of the theoretical number of beads, for each bead level (Table 2). Fig. 3A shows the results for each run (A-D), as well as the average of 4 runs. The average accuracy over the four runs met the acceptance criteria at all levels C_1–C_8 (81 %–101 %). The recovery range was within the 120 % upper bound for every individual measurement, but was lower than 80 % for level C_1, runs C (72 %) and D (76 %), for C_2 run D (79 %) and for C_5 run C (79 %).

Overall, average bead recovery met the acceptance criteria for accuracy; however, lower recovery was observed with low bead levels for some runs.

3.2. Precision – repeatability and intermediate precision

Precision was assessed at two levels: repeatability and intermediate precision, and calculated as the percentage coefficient of variation, CV(%). Repeatability was determined intra-assay, by calculating the mean CV(%) of the technical replicates for each sample and each run; intermediate precision was determined inter-assay, as the Overall CV(%) calculated from 4 runs of the same analyte over 4 days.

Beads, GTA002 cells and K562 cells were tested with multiple sample preparations. Precision of the enumeration of beads and of cells was analyzed by preparing a range of dilution levels for each analyte. Counting beads were diluted at 8 levels, C_1–C_8 (as described for accuracy assessment), GTA002 cells were diluted at 8 levels, C_1–C_8, i.e., 1.25×10^4, 2.5×10^4, 5.0×10^4, 2.5×10^5, 5.0×10^5, 1.0×10^6, 2.0×10^6, and 4.0×10^6 cells/ml, and K562 cells at 6 levels, C_1–C_6, i.e., 4.0×10^4, 8.0×10^4, 2.0×10^5, 4.0×10^5, 8.0×10^5, and 2.0×10^6 cells/ml (Table 2). Each dilution level was acquired in consecutive technical triplicates on four different days, by two analysts (N = 12 tests).

For beads, acceptance criteria were set at CV(%) ≤25 % for repeatability (Table 2, Equations 5 and 6 (E5 and E6)) and CV(%) ≤30 % for intermediate precision (Table 2, Equations 15 and 16 (E15 and E16)). All data fall well below the acceptance threshold, within a range of 1%–16 % CV(%) for repeatability and 7%–17 % Overall CV(%) for intermediate precision (Fig. 3B).

For GTA002 cells, the intra-assay CV(%) threshold was set at maximum 30 % (Table 2, Equations 9 and 10 (E9 and E10)), while the inter-assay Overall CV(%) was considered acceptable at maximum 40 % (Table 2, Equations 19 and 20 (E19 and E20)). All results were

well-below such limits (Fig. 3C). The highest CV(%) value for repeatability was observed at the lowest cell concentration C_1, (14 %, run B). Increase in cell concentration improved repeatability, as the lowest CV(%) values were observed for C_7 and C_8, (maximum 2 % and 3 %, respectively, both in run C). The variation was higher for intermediate precision, with Overall CV(%) values between 9 and 15 %, but still meeting the predefined acceptance criteria (Fig. 3C).

Analysis of K562 cells showed similar results. For intra-assay precision, acceptance criteria were set at CV(%) $\leq$30 % for lower levels (C_1–C_2) and $\leq$20 % CV for higher levels (C_3–C_6) (Table 2, Equations 11 and 12 (E11 and E12)); for inter-assay precision, they were set as Overall CV(%) $\leq$30 % for lower levels and at $\leq$25 % for higher levels (Table 2, Equations 21 and 22 (E21 and E22)). As shown in Fig. 3D, all tests from all runs were within specification for both intra- and inter-assay precision, with the highest variation recorded for level C_1: 7 % CV(%) (run C) and 17 % Overall CV(%). In general, lower variations were observed for C_3–C_6, with maximum CV(%) 6 % (run B) and Overall CV(%) 13 %, both for C_6. Thus, the precision for target cell enumeration was found acceptable for all levels.

Notably, target cell sample concentrations counted with the method during routine QC testing showed low variation: analysis of C_4 (4.0×10^5 cells/ml), which represents the average K562 density measured before effector cell killing, reached a maximum 3 % CV(%) for intra-assay (run B) and 11 % Overall CV(%) for inter-assay precision, while C_2 (8.0×10^4 cells/ml), representing the average density measured after co-culture with effectors at 10:1 E:T ratio, showed maximum 6 % CV(%) (run C) and 14 % Overall CV(%).

Precision was also evaluated with undiluted GTA002 cells after thawing and reconstitution, using the same cell batches as in the dilution runs. The intra-assay CV(%) threshold was set at maximum 30 % for repeatability (Table 2, Equations 7 and 8 (E7 and E8)), while the inter-assay Overall CV(%) was set at 40 % for intermediate precision (Table 2, Equations 17 and 18 (E17 and E18)). Undiluted GTA002 cells were stained and acquired in consecutive triplicates (2 runs) or in sextuplicate (2 runs), by two analysts on four days on one flow cytometer (Fig. 3E). The highest CV(%) was recorded on day 3 of testing (4 %), meeting the acceptance criteria for repeatability of 30 %, while the overall CV(%) was 7 %, well within the acceptance criteria of 40 %.

Finally, precision of flow cytometry-based method for assessment of effector cell potency was determined using samples of co-cultured effector and target cells, after overnight incubation at E:T ratios of 10:1, 3:1 and 1:1. Samples were acquired in consecutive triplicates on four independent runs, on four days, by two analysts, using one flow cytometer. The coefficient of variation CV(%) was calculated for each E:T ratio and for each run (Table 2, Equations 13 and 14 (E13 and E14)); the Overall CV(%) was determined from all four runs (N = 12 tests) (Table 2, Equations 23 and 24 (E23 and E24)). Acceptance criteria were set at CV(%) $\leq$50 % for repeatability and at Overall CV(%) $\leq$50 % for intermediate precision. Results are shown in Fig. 3F. The highest variability was observed at the 1:1 E:T ratio, with 20 % at intra-assay level (run C) and 25 % CV at inter-assay level. The 3:1 ratio showed 7 % maximum CV (run D) and 8 % inter-assay precision. The highest precision was observed for the 10:1 ratio, with 3 % highest CV(%) reported for both intra (run D) and inter-assay. Notably, in the ongoing Phase I/II trial, the 10:1 E:T potency analysis is one of the release tests for oNKord®.

Overall, all results met the acceptance criteria set for the precision parameter of the method.

3.3. Linearity

Linearity was assessed for the enumeration of counting beads, GTA002 and K562 cells using the data generated for the determination of accuracy and precision. Serial sample dilutions (C_1–C_8 for beads and GTA002 cells, and C_1–C_6 for K562 cells) were acquired in consecutive triplicates, in four runs performed by two analysts on four days (Table 2). For every run, the expected count for each level (expressed as theoretical bead count or as cells/ml) was plotted against the measured count, determined as single beads per test volume for beads, as total living $CD45^+$ cells for GTA002 cells, or as living $PBSE^+$ cells for K562. A linear regression model was applied to determine the goodness of fit (Table 2, Equations 25–27 for beads (E25, E26 and E27), Equations 28–30 for GTA002 (E28, E29 and E30), Equations 31–33 for K562 (E31, E32 and E33)). For all samples and runs, the equation of the trend line, the R^2 coefficient of determination and the residual sum of squares (RSS) for the x and y axes are shown in Table 3. Acceptance criteria were defined as R^2

Table 3
Linearity analysis. Linearity tests were performed using 123count eBeads™ Counting Beads, GTA002 cells and K562 cells. Beads and GTA002 cells were diluted to 8 levels, K562 cells to 6 levels. Samples were acquired in technical triplicates, in 4 runs, on 4 days, by 2 analysts. For each run (A, B, C or D), the linear trendline equation (expressed as $y = ax + b$, where a is the slope and b is the y-intercept), the coefficient of determination R^2, and residual sum of squares (RSS) for both x and y axes are reported. *: 5 test levels instead of 8.

Analyte	Run ID	Slope (a)	y-intercept (b)	R^2	RSS (y axis)	RSS (x axis)	RSS total
Counting beads	Run A	1.13	−690	0.99	2.4×10^{10}	1.9×10^{10}	4.3×10^{10}
	Run B	1.10	−823	0.99	2.2×10^{10}	1.9×10^{10}	4.1×10^{10}
	Run C	0.93	−589	0.99	1.6×10^{10}	1.9×10^{10}	3.5×10^{10}
	Run D	0.91	−854	0.99	1.5×10^{10}	1.9×10^{10}	3.4×10^{10}
GTA002 cells	Run A*	0.014	34	0.99	6.6×10^{6}	4.9×10^{7}	1.2×10^{8}
	Run B	0.009	665	0.99	2.0×10^{9}	3.3×10^{9}	5.3×10^{9}
	Run C	0.011	707	0.99	2.8×10^{9}	3.3×10^{9}	6.1×10^{9}
	Run D	0.011	810	0.99	3.0×10^{9}	3.3×10^{9}	6.3×10^{9}
K562 cells	Run A	0.010	222	0.99	5.2×10^{8}	7.6×10^{8}	1.3×10^{9}
	Run B	0.011	219	0.99	6.5×10^{8}	7.6×10^{8}	1.4×10^{9}
	Run C	0.013	95	0.99	8.4×10^{8}	7.6×10^{8}	1.6×10^{9}
	Run D	0.014	218	0.99	9.1×10^{8}	7.6×10^{8}	1.7×10^{9}

$\geq$ 0.80; for all runs, values were >0.99, well above threshold. Notably, the slope, y-intercept and RSS were similar between all runs of the same analyte, indicating that the linearity trends are comparable between assay repeats. The discrepancy observed in the data from GTA002 run A is due to the acquisition of only 5 sample levels (C_1–C_6), instead of the 8 used for the other runs in the data set; nevertheless, data were within limits. Representative trend curves for beads, GTA002 and K562 cells are shown in Fig. 3 G, H, I, respectively.

Overall, beads, effector and target cell count showed a linear trend across the tested range, indicating proportionality between the amount of analyte measured and its theoretical concentration in the sample.

3.4. Range

The range of analyte concentration that can be reliably measured with the analytical method was determined for counting beads, GTA002 cells and K562 cells from the linearity analysis, by identifying the progressive concentrations that meet the acceptance criteria for accuracy, precision, and linearity. The boundaries of the range, i.e., the highest and lowest reliable concentrations, represent the Higher and the Lower Limit of Quantification, HLoQ and LLoQ (Table 2).

All four runs of bead enumeration showed acceptable accuracy, precision and linearity, for all tested levels. Therefore, LLoQ was determined from the average value of bead count measured over the four runs for C_1, being 2,130 (264) beads/volume (mean (SD)), and HLoQ from the average of four runs for C_8, being 100,767 (11,282) beads/volume.

GTA002 effector cell count was qualified for linearity, accuracy, and precision for the enumeration of $CD45^+$/7-AAD^- cells over the entire range of 8 levels tested. C_1, reported a count of 169 (19) events, while C_8, of 42,039 (4,599). These values, according to Equation 1 (E1), define an acceptable range of 13,540 (1,516) (LLoQ) - 3,363,147 (367,897) (HLoQ) cells/ml for the count of GTA002 cells.

Similarly, all levels of K562 were compliant with precision and linearity acceptance criteria, and the range was therefore defined between C_1 and C_6. The measured events were between 493 (83) and 24,030 (3,099), corresponding to 39,473 (6,611) (LLoQ) - 1,922,420 (247,940) (HLoQ) cells/ml, when Equation 2 (E2) is applied to raw counts. This range, approximated to 0.04–1.9 $\times$ 10^6 cells/ml, extensively covers the 0.08–0.4 $\times$ 10^6 density that is generally observed during target cell sample handling, before and after assessment of cytotoxicity, making the method fit for reporting on the cytotoxicity of GTA002 effector cells for final product release.

Thus, quantitative ranges for the enumeration of effector and target cells were satisfactorily established.

3.5. Specificity

Specificity was assessed in multiple ways, with counting beads, GTA002 and K562 cells (Table 2). With counting beads, bead events were detected from a blank sample containing only the matrix used for resuspension (PBS buffer). Blank samples were acquired in consecutive triplicates before the acquisition of test samples for the four linearity runs on flow cytometers A and B. The maximum acceptable number of events in the bead gate was set to 10. For each replicate and each run, no bead events were recorded (data not shown), and the method was considered specific to the analyte of interest.

Specificity of effector cell count was evaluated by comparing the results from the acquisition of a stained sample to an unstained one, taken from the same GTA002 batch, where the antibody cocktail was replaced by PBS. The unstained sample was acquired before the stained; both were analyzed in consecutive triplicates over 3 days. Acceptance criteria were set as $\leq$100 events recorded from the unstained sample. As shown in Fig. 3J, the unstained sample showed on average 67–72 events per test day in the "Living cells" gate, corresponding to 0.2 % of the events measured in the stained, therefore meeting specificity criteria for GTA002 enumeration. To test the capability of the potency method to specifically detect the cells of interest in presence of other cells in the matrix, K562 target cells were spiked with effector cells at various E:T ratios, then immediately counted with the method. Four theoretical E:T ratios were prepared, 90:1, 45:1, 9:1 and 1.8:1, by keeping the amount of effector cells fixed to 4.5 $\times$ 10^5 cells/well but varying the number of pre-stained target cells between 5.0 $\times$ 10^3-2.5 $\times$ 10^5 cells/well. K562 control samples (non-spiked), with the same four concentrations of target cells but without effector cells, were measured alongside. A "blank" sample, containing only effector cells, was included (E:T 100:0). All samples were acquired in technical triplicates on one day. Acceptance criteria were set as $\leq$10 living target cells measured in the blank effector-only control and less than 30 % difference (Δ% $\leq$30 %) between non-spiked and spiked samples (Table 2, Equation 34 (E34)). A comparable amount of living target cells was quantified from both spiked and non-spiked samples and the specificity test met the predefined criteria for all samples, with the highest difference (Δ%) of 9 % identified at the highest spiking ratio of 90:1, and the lowest, 5 %, for 1.8:1 ratio (Fig. 3K). Additionally, the blank E:T 100:0 control recorded only 2 events, meeting the predefined criteria of less than 10 events.

The evaluation of specificity met the acceptance criteria for all assessments, showing that the method is able to discriminate the analyte of interest and to specifically enumerate target cells in co-culture with effector cells, without being affected by the presence of other components in the matrix.

3.6. Robustness and sample stability

In flow cytometry, robustness is ensured by measures established during method development, such as choosing materials guaranteeing lot-to-lot consistency, operating daily verification of instrument performance, and defining solid acquisition settings. However, variations often occur during sample handling, especially involving time and scheduling. The duration of the staining procedure, or of the sample processing from thawing to acquisition, can affect the outcome of the assay. Understanding the impact of such variables on method performance can help define tolerance windows.

Table 4
Robustness and sample stability analysis. Robustness and sample stability were evaluated as the impact of sample staining and storage times before processing. Robustness: GTA002 (E), K562 (T) and co-cultured (E + T) samples were collected and stained with the appropriate antibodies for 15 min (standard condition), and with a variation of ±10 min, for 5 or 25 min, then acquired by flow cytometry in technical triplicates, in 1 run, on 1 day, by 1 analyst. The average number of events measured for each sample in the appropriate gate, and the difference (Δ(%)) from the standard condition are reported. Sample stability: GTA002 samples were collected and stained with the appropriate antibodies immediately after thawing and reconstitution (standard condition), or 30 min or 60 min later, then acquired by flow cytometry in technical triplicates, in 1 run, on 1 day, by 1 analyst. The average number of events measured for every sample in the "Living cells" gate, and the difference Δ(%) from the standard condition are reported. STD: standard condition.

Analyte	Parameter	Staining time: 5 min	Staining time: 15 min (STD)	Staining time: 25 min
GTA002 cells (E)	Average "Living cells" events	30,452	30,270	29,182
	Δ(%)	1	–	4
K562 cells (T)	Average "Living targets" events	6,906	7,164	7,299
	Δ(%)	4	–	2
GTA002 + K562 cells (E + T)	Average "Living targets" events	865	1,108	1,137
	Δ(%)	22	–	3
Analyte	***Parameter***	***Storage time*: 0 min *(STD)***	***Storage time*: 30 min**	***Storage time*: 60 min**
GTA002 cells (E)	Average "Living cells" events	30,270	29,605	30,534
	Δ(%)	–	2	1

Robustness was assessed by varying the cell sample staining time by ±10 min, compared to the standard 15 min indicated by the method (i.e., 5 or 25 min) with GTA002 cells, K562 cells and co-cultured effectors and targets (Table 2). All tests were performed in technical triplicates on one day. Robustness was calculated as the difference (Δ(%)) between the tests and the standard condition (Table 2, Equation 35 (E35)).

For the count of GTA002 cells (E), acceptance was set as Δ(%) ≤30 %, and was met by both 5 min and 25 min conditions, with a maximum difference of 4 % for 25 min (Table 4). For the count of K562 cells alone (T) and after overnight co-culture with effectors (E + T, 10:1 E:T ratio), acceptance criteria were set as Δ(%) ≤20 %. Results are shown in Tables 4 and in Fig. 3L. For T only samples, the maximum count difference observed was 4 %, for 5 min condition. Highest variation was observed for E + T samples, reaching 22 % for 5 min, and thus exceeding the acceptance criteria of Δ(%) ≤20 %. Notably, increasing the staining time to 25 min only minimally affected the count of the T and E + T samples by 3 %, which is well within the acceptance limits.

Sample stability was investigated with GTA002 cells delaying the staining and acquisition by 30 or 60 min after cell thawing, while cells were kept at 37 °C (Table 2). Acceptance limits (Table 2, Equation 36 (E36)) were defined as less Δ(%) ≤25 % between the delayed and the standard condition. Table 4 shows that the observed difference was maximum 2 %, for 30 min, well below the acceptance threshold, indicating that sample storage up to 1 h does not affect cell enumeration.

Overall, varying sample staining time did not affect effector or target count alone but only impacted co-cultures, for which it should not be decreased. Besides, delaying sample processing only had minimal effect on sample stability. Together, these data define time boundaries for sample acquisition within which the reliability of the analytical system is ensured.

3.7. Carryover

Carryover was assessed as the number of events measured from a blank sample immediately after acquisition of counting beads, GTA002 or K562 cells. PBS blanks were acquired in consecutive triplicates after acquisition of the highest concentration level of beads (1.0×10^5 beads/volume) or of cells (4.0×10^6 and 2.0×10^6 cells/ml for effectors and targets, respectively) during linearity testing, in four runs. Acceptance threshold was set as CO ≤ 1 % of the average number of events measured in the test sample, for each analyte

Table 5
Carryover analysis. Carryover was evaluated by the acquisition of a blank sample (phosphate buffer saline, PBS) immediately after the acquisition of the highest concentration level of 123count eBeads™ Counting Beads, GTA002 cells or K562 cells from linearity runs. Blanks were acquired in technical triplicates, in 4 runs, on 4 days, by 2 analysts.

Analyte	Run ID	CO (%) after highest analyte level	Result (PASS/FAIL)
Counting beads	Run A	<0.01	PASS
	Run B	<0.01	PASS
	Run C	0.02	PASS
	Run D	0.01	PASS
GTA002 cells	Run A	0.02	PASS
	Run B	0.01	PASS
	Run C	0.01	PASS
	Run D	0.01	PASS
K562 cells	Run A	0.01	PASS
	Run B	0.00	PASS
	Run C	0.01	PASS
	Run D	0.01	PASS

and each run (Table 2, Equation 37 (E37))). Data showed a maximum 0.02 % CO for beads and GTA002 cells, of 0.01 % with K562 cells (Table 5). Thus, carryover was shown to be minimal, meeting acceptance criteria for all analytes.

4. Discussion

Successful development of cellular therapies heavily relies on the assessment of key product attributes, influencing batch quality and consistency [14]. Potency assessment of cytotoxic immune cells, such as NK and T cells, is critical to confirm the biological quality of the products and their clinical efficacy. Specific guidelines for potency testing are provided by regulatory authorities [12,15], but as cellular therapies are highly heterogeneous, they allow considerable flexibility in determining the appropriate potency measurement (s) for each product. Therefore, the adequacy of tests is evaluated on a case-by-case basis, and developers must invest time and resources in designing the optimal method for their unique product to ensure clinical advancement towards market authorization. Potency testing likely includes multiple assays representative of the product's mechanism of action for the evaluation of attributes that are most relevant to predict clinical efficacy. Assessment of the fitness and robustness of potency assays should begin in early clinical development stages and be adjusted when progressing to pivotal studies. Additionally, it is important that sponsors, when submitting data for regulatory approval (based on Phase III clinical trial results), show strong reliability of potency estimation based on evidence from the use of several complementary assays to the cell-based cytotoxicity assay, using orthogonal approaches, especially with genetically engineered cell products [16]. Thus, the potency assay should constantly be refined and correlated with other functional characterization assays, as more information about mode of action emerges.

Given the complexity of the task, the performance of the flow cytometry-based analytical methods established for release of the allogeneic clinical product oNKord® was challenged prior to the start of a Phase I/IIa clinical trial in AML. Qualification procedures were defined for the quantitative enumeration of effector and of target cells, and for the qualitative assessment of the cytotoxicity of effector cells. Research & Development-grade GTA002 NK cells, manufactured in compliance with the clinical product, were used as effector cell analyte; the chronic myelogenous leukemia cell line K562 was used as target cell analyte. As the performance of the cytotoxicity assay depends on the efficient enumeration of effector and of target cells, qualification protocols were designed to investigate both the quantitative and qualitative aspects of the methods. Therefore, test parameters were chosen based on the most important features of the assays, and acceptance criteria were defined on the expected values (and their variations) from routine, QC-use of the assay for in-process-control and clinical batch release, which were determined during assay development. Nevertheless, acceptance criteria were maintained quite broad, with the goal of narrowing them further for method validation, based on the increasing knowledge on product development acquired during clinical stages.

Accuracy is one of the main parameters required by authorities to be analyzed and validated [4]. This poses a challenge to cell product developers, due to the need to find appropriate surrogate materials. Our method proposes the use of counting beads as a robust and well-controlled material to assess accuracy, precision and linearity of flow cytometry count methods. Bead results qualified the capability of our method to enumerate analytes accurately and precisely at concentration levels between 2.5×10^3-1.0×10^5 beads/test volume. Results that did not meet the acceptance criteria were reported for some concentrations in the accuracy analysis, especially in the lower range, although average values were acceptable. Notably, these were observed only when instrument B was used. Nevertheless, precision results were highly reproducible between runs, and could allow for the intra- and inter-assay precision threshold to be reduced from 25 % to 20 % and from 30 % to 20 %, respectively, for method validation, and possibly to even lower values by assessing the optical alignment between flow cytometers A and B. Further, bead count should be evaluated over a wider range, to match the routine concentrations of GTA002 and K562 cells, e.g., 1.0×10^2-1.0×10^6 beads/test volume; this would help draw more relevant conclusions on accuracy, which could then be safely extrapolated to cell analytes. Slightly bigger counting beads (17–20 μm diameter) could be taken into consideration to better match the size of K562 cells.

Precision was investigated with undiluted GTA002 samples, diluted GTA002 and K562 cells, and with co-cultured cells. Undiluted GTA002 cells were thawed from cryopreserved batches, which underwent the same harvest and fill-and-finish procedure as oNKord®. Low inter- and intra-assay variability was observed with undiluted GTA002 cells enumerated immediately after thawing, demonstrating the precision of the method when performed on a minimally diluted matrix, in a similar procedure as in-process control QC testing of oNKord® cell cultures. Furthermore, diluted samples were used to define a range of acceptable precision. Results showed satisfactory repeatability and intermediate precision for both effector and target cells. Acceptance criteria can thus be adapted for future validation based on batch-to-batch variability and trending of reference material, which is continuously monitored during manufacturing, to 20 % for intra- and inter-assay precision with diluted GTA002 cells and to 10 % with undiluted samples. For K562, acceptance criteria can be lowered to 10 % and 20 % for inter- and intra-assay evaluation. For co-culture studies, effectors and target cells were combined at multiple E:T ratios (10:1, 3:1 and 1:1); although the lowest ratio, 1:1, showed the highest intra- and inter-assay variability, all were significantly within acceptance limits. Notably, the 10:1 ratio, used for potency assessment and release testing of clinical batches, consistently showed the highest precision. Lowering the threshold to 30 % for validation would suit all co-culture ratios; further lowering to 10 % would be possible for 10:1.

Linearity of the enumeration, assessed with diluted ranges of counting beads, GTA002 and K562 cells, was successfully demonstrated for all analytes. Each run showed a robust linear correlation between the expected and the measured amount of analyte, for the entire range tested. The R^2 threshold of 0.80 could be increased to 0.90 for method validation. Defined on accuracy, precision, and linearity data, suitable ranges are quite wide and are appropriate for the purpose of the method, as the concentration of effector and target cells routinely measured in in-process control and release testing fit well into the qualified limits of quantification. However, the qualified ranges should be reviewed if changes in the product or in the QC workflow occur, so that appropriate re-qualification/re-validation steps can be taken. To widen the spectrum, ranges could be challenged towards lower concentrations during validation.

Specificity was analyzed with GTA002 and K562 cells. For GTA002 cells, detection of the analyte of interest in a blank sample was considered sufficient to qualify enumeration for early clinical stages. Results clearly showed that the matrix used for sample acquisition has no influence on cell count. Notably, K562 living cell enumeration was not affected by the presence of different ratios of effector cells in the matrix, as demonstrated by the minimal differences observed between spiked and non-spiked samples when assessed immediately after. As for GTA002 cells, absence of analyte in the matrix did not report any events. In the future, however, specificity should be addressed using a cell sample that does not express the antigen of interest, or presents it at a lower density (e.g., CD45 antigen), and a sample of dead cells, to further challenge the reliability of the viability analysis.

Studying the robustness of the method, although optional for earlier stages, is useful to address as soon as possible which sources of variation in the analytical system should be evaluated and contained within pre-defined limits. Several measures, such as verification of critical material incoming goods, use of lot-to-lot consistent antibodies, evaluation of flow cytometer performance, and establishment of fixed sample acquisition procedures, have been applied during assay development. Time, however, remains a permanent source of variation. For GTA002 and K562, alone or in co-culture, the impact of the duration of sample staining procedure was assessed. In all cases, the established standard time of 15 min was shortened to 5 min or extended to 25 min. Enumeration of effector or target cells alone was not affected; interestingly, shortening the staining time of co-cultured samples significantly influenced the results, making the 5-min period not acceptable. During validation, robustness should further be assessed from antibody titration analysis.

Sample stability was assessed with GTA002 cells. Comparing immediate analyte processing to 30- or 60- minutes delayed conditions after thawing showed no impact of sample storage time. Up to 1 h delay of sample preparation, which can occur during testing while waiting for equipment or analysts to be available, does not impact the outcome of the analysis. However, more insight about the influence of time on the enumeration of cells and on the evaluation of potency must be obtained during validation, for example by comparing different equipment and operators, by testing more GTA002 batches, by analyzing the effect of incubation on K562 cells after thawing. Defining more challenging time windows of sample storage before and during preparation steps will ensure the generation of appropriate and robust data in the context of real daily lab activities, when analysts can have particularly busy days and samples cannot be processed without delay. Additionally, given the importance of time in potency testing, investigating the effect of the duration of the co-incubation of effectors and targets on the determination of cytotoxicity will further strengthen the method.

Lastly, evaluation of carryover must be carefully considered when using equipment that relies on a fluidics system (pressurized liquid flow through a tubing set) for automatic sampling with a sample probe (as also described in the International Standard ISO15189:2012 Medical laboratories – Requirements for quality and competence [17]). Carryover, evaluated by measuring a blank sample after a sample with high content of test analyte (highest level of counting beads, GTA002 and K562 from linearity runs), was proven to be negligible.

5. Conclusions

Notwithstanding the complex nature of the analyte and of the matrix composition, the variability inherent to the application of the potency method for GTA002 batch release was shown to be acceptable, when compared to the pre-defined acceptance criteria. The results of the qualification demonstrated the capability of the method, inclusive of trained analysts, documentation, equipment, reagents and samples of effector and target cells, to deliver reliable results, and provide useful insight for the definition of acceptance criteria for further stages of the method validation process. Such learnings will be applied during the establishment of the method validation protocols, when progressing to pivotal clinical Phase III, by the end of which full method validation data will have to be provided to regulatory authorities.

Data availability statement

The data associated with this study has not been deposited into a publicly available repository and is confidential.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Monica Raimo: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alexandra G. Zavoianu:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation **Wilma Meijs:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Pascal Scholten:** Writing – review & editing, Validation, Investigation. **Jan Spanholtz:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

Acknowledgements

The authors would like to thank Lars von Oerthel and Jolanda Paardekooper for executing some of the flow cytometry assays.

List of abbreviations

ATMP	Advanced Therapy Medicinal Product
GMP	GoodManufacturing Practice
EMA	European Medicines Agency
ICH	Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
FDA	United States Food and Drug Administration
Ph. Eur	European Pharmacopeia
USP	United States Pharmacopeia
ISO	International Organization for Standardization
CQA	Critical Quality Attribute
NK	Natural Killer
QC	Quality Control
AML	Acute Myeloid Leukemia
IMPD	Investigational Medicinal Product Dossier
GBGM	Glycostem Basal Growth Medium
HS	Human Serum
IL	Interleukin
IU	International Units
ATCC	American Type Culture Collection
STR	Short Tandem Repeat
PBSE	Pacific Blue Succinimidyl Ester
IMDM	Iscove's Modified Dulbecco's Medium
FBS	Fetal Bovine Serum
PBS	Phosphate Buffer Saline
RT	Room Temperature
7-AAD	7-Amino-Actinomycin D
E	Effector
T	Target
O/N	Overnight
SST	System Suitability Testing
QA	Quality Assurance
FSC	Forward Scatter
SSC	Side Scatter
LLoQ	Low Limit of Quantification
HLoQ	High Limit of Quantification
CO	Carry Over
LoD	Limit of Detection
CV	Coefficient of Variation
RSS	Residual Sum of Squares

References

[1] A. Campbell, T. Brieva, L. Raviv, J. Rowley, K. Niss, H. Brandwein, et al., Concise review: process development considerations for cell therapy, Stem. Cells. Transl. Med 4 (10) (2015) 1155–1163, https://doi.org/10.5966/sctm.2014-0294. PubMed PMID: 26315572; PubMed Central PMCID: PMCPMC4572896.

[2] R. Isasi, V. Rahimzadeh, K. Charlebois, Uncertainty and innovation: Understanding the role of cell-based manufacturing facilities in shaping regulatory and commercialization environments, Appl. Transl. Geno 11 (2016) 27–39, https://doi.org/10.1016/j.atg.2016.11.001.

[3] B. Mfarrej, J. Gaude, J. Couquiaud, B. Calmels, C. Chabannon, C. Lemarie, Validation of a flow cytometry-based method to quantify viable lymphocyte subtypes in fresh and cryopreserved hematopoietic cellular products, Cytotherapy 23 (1) (2021) 77–87, https://doi.org/10.1016/j.jcyt.2020.06.005. PubMed PMID: 32718876.

[4] D.M. O'Hara, Y. Xu, Z. Liang, M.P. Reddy, D.Y. Wu, V. Litwin, Recommendations for the validation of flow cytometric testing during drug development: II assays, J. Immunol. Methods 363 (2) (2011) 120–134, https://doi.org/10.1016/j.jim.2010.09.036.

[5] B. Wood, D. Jevremovic, M.C. Béné, M. Yan, P. Jacobs, V. Litwin, Validation of cell-based fluorescence assays: practice guidelines from the ICSH and ICCS - part V - assay performance criteria, Cytometry. B. Clin. Cytom 84 (5) (2013) 315–323, https://doi.org/10.1002/cyto.b.21108. PubMed PMID: 24022854.

[6] C. Lambert, G. Yanikkaya Demirel, T. Keller, F. Preijers, K. Psarra, M. Schiemann, et al., Flow cytometric analyses of lymphocyte markers in immune oncology: a comprehensive guidance for validation practice according to laws and standards, Front. Immunol. 11 (2020), https://doi.org/10.3389/fimmu.2020.02169.
[7] Guidelines on Good, Manufacturing Practice Specific to Advanced Therapy Medicinal Products 4 (2017).
[8] ICH Topic Q 2 (R1), Validation of Analytical Procedures: Text and Methodology Step 5 NOTE for GUIDANCE on VALIDATION of ANALYTICAL PROCEDURES: TEXT and METHODOLOGY, 1995. CPMP/ICH/381/95.
[9] S.P. Boudreau, J.S. Mcelvain, L.D. Martin, T. Dowling, S.M. Fields, Method validation by phase of development: an acceptable analytical practice, Pharmaceut. Technol. 28 (2004) 54–66.
[10] N.M.A.S. Ritter, J. Hennessey, et al., What is test method qualification? Proceedings of the WCBP CMC strategy forum, 24 July 2003, BioProcess. Int., 2 (8) (2004) 32–46. Ritter NM, Advant SJ, Hennessey J, et al (2004) What is test method qualification? Proceedings of the WCBP CMC strategy forum, 24 July 2003. BioProcess Int. 2004(2):32-46.
[11] C.A. Bravery, J. Carmen, T. Fong, W. Oprea, K.H. Hoogendoorn, J. Woda, et al., Potency assay development for cellular therapy products: an ISCT* review of the requirements and experiences in the industry, Cytotherapy 15 (1) (2013) 9–19, https://doi.org/10.1016/j.jcyt.2012.10.008.
[12] Guidance for Industry, Potency Tests for Cellular and Gene Therapy Products, 2011. FDA-2008-D-0520.
[13] A.J. Dickinson, M. Meyer, E.A. Pawlak, S. Gomez, I. Jaspers, N.L. Allbritton, Analysis of sphingosine kinase activity in single natural killer cells from peripheral blood, Integr. Biol. 7 (4) (2015) 392–401, https://doi.org/10.1039/c5ib00007f. PubMed PMID: 25786072; PubMed Central PMCID: PMCPMC4566154.
[14] Guidance for Industry, Guidance for Human Somatic Cell Therapy and Gene Therapy, 1998. FDA-2009-D-0132-0016.
[15] Guideline on Potency Testing of Cell Based Immunotherapy Medicinal Products for the Treatment of Cancer, EMA/CHMP/BWP, 2016.
[16] Considerations for the Development of Chimeric Antigen Receptor (CAR) T Cell Products - Draft Guidance for Industry, 2022. FDA-2021-D-0404.
[17] Standardization IOf, ISO 15189:2012 Medical Laboratories — Requirements for Quality and Competence, 2012.