



Contents lists available at ScienceDirect

## Journal of Immunological Methods

journal homepage: [www.elsevier.com/locate/jim](http://www.elsevier.com/locate/jim)

## Review

## Development and implementation of a proficiency testing program for Luminex bead-based cytokine assays

Heather E. Lynch<sup>a</sup>, Ana M. Sanchez<sup>a</sup>, M. Patricia D'Souza<sup>b</sup>, Wes Rountree<sup>a</sup>, Thomas N. Denny<sup>a</sup>, Michael Kalos<sup>c,d</sup>, Gregory D. Sempowski<sup>a,\*</sup><sup>a</sup> Duke Human Vaccine Institute, Department of Medicine, Duke University Medical Center, Durham, NC 27710, United States<sup>b</sup> Division of AIDS, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, United States<sup>c</sup> Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, United States<sup>d</sup> Cancer Immunotherapy Consortium of the Cancer Research Institute, United States

## ARTICLE INFO

## Article history:

Received 3 October 2013

Received in revised form 10 April 2014

Accepted 23 April 2014

Available online xxxx

## Keywords:

Luminex  
Cytokines  
Proficiency  
GCLP  
Multiplex

## ABSTRACT

Luminex bead array assays are widely used for rapid biomarker quantification due to the ability to measure up to 100 unique analytes in a single well of a 96-well plate. There has been, however, no comprehensive analysis of variables impacting assay performance, nor development of a standardized proficiency testing program for laboratories performing these assays. To meet this need, the NIH/NIAID and the Cancer Immunotherapy Consortium of the Cancer Research Institute collaborated to develop and implement a Luminex assay proficiency testing program as part of the NIH/NIAID-sponsored External Quality Assurance Program Oversight Laboratory (EQAPOL) at Duke University. The program currently monitors 25 domestic and international sites with two external proficiency panels per year. Each panel includes a de-identified commercial Luminex assay kit with standards to quantify human IFN $\gamma$ , TNF $\alpha$ , IL-6, IL-10 and IL-2, and a series of recombinant cytokine-spiked human serum samples. All aspects of panel development, testing and shipping are performed under GCLP by EQAPOL support teams. Following development testing, a comprehensive site proficiency scoring system comprised of timeliness, protocol adherence, accuracy and precision was implemented. The overall mean proficiency score across three rounds of testing has remained stable (EP3: 76%, EP4: 75%, EP5: 77%); however, a more detailed analysis of site reported results indicates a significant improvement of intra- (within) and inter- (between) site variation, suggesting that training and remediation for poor performing sites may be having a positive impact on proficiency. Through continued proficiency testing, identification of variables affecting Luminex assay outcomes will strengthen efforts to bring standardization to the field.

© 2014 Elsevier B.V. All rights reserved.

## Contents

1.	Introduction . . . . .	0
2.	Program development . . . . .	0
2.1.	EQA platform and panel design . . . . .	0
2.2.	Reagents and assay kits . . . . .	0
2.3.	Support for proficiency testing . . . . .	0
2.4.	Participating site requirements. . . . .	0
2.5.	Biannual proficiency panel and send-out . . . . .	0

\* Corresponding author at: Human Vaccine Institute, Duke University Medical Center, 909 S. La Salle St. GHRB Room 1005, Durham, NC 27710, United States.  
E-mail address: [gregory.sempowski@duke.edu](mailto:gregory.sempowski@duke.edu) (G.D. Sempowski).

2.6.	Data collection and analysis . . . . .	0
2.7.	Proficiency evaluation (Scoring) . . . . .	0
2.7.1.	Timeliness (10 points) . . . . .	0
2.7.2.	Protocol adherence (10 points, split) . . . . .	0
2.7.3.	Accuracy to the consensus (40 points) . . . . .	0
2.7.4.	Precision (40 points) . . . . .	0
2.7.5.	Central analysis . . . . .	0
2.7.6.	Reporting results to participating laboratories . . . . .	0
2.7.7.	Site performance remediation . . . . .	0
3.	Results . . . . .	0
3.1.	EP1 . . . . .	0
3.2.	EP2 . . . . .	0
3.3.	EP3, EP4, and EP5 . . . . .	0
4.	Overall summary . . . . .	0
	Acknowledgments . . . . .	0
	References . . . . .	0

## 1. Introduction

Enzyme-linked immuno-assays have revolutionized our ability to rapidly quantify critical biological mediators such as cytokines, chemokines, hormones, and signal transduction proteins. Traditionally, these assays are performed in 96-well plates using an enzyme-linked monoclonal antibody to cleave a substrate and produce a color change. Although these assays are quite powerful, they have been limited to one dimension – detection of a single analyte per well (50–100  $\mu$ L sample). With the creation of 100 color-coded fluorescent bead sets by Luminex, each of which can be conjugated with a unique specific reactant (antibody, substrate, etc.), biomarker analysis can now be done with a multiplex approach.

Custom-made or commercially available Luminex bead array assays can be partnered with Luminex bead readers, data analysis software, and instrument validation/calibration kits to generate comprehensive immune monitoring and discovery technology. The power of this system is that up to 100 unique analytes can be quantified in a single well of a 96-well plate (25–50  $\mu$ L of sample). The automated dual laser or CCD camera flow-based array reader identifies each analyte based on a unique fluorescent bead signature and quantifies fluorescence intensity of the reporter associated with that bead set. This multiplex approach to biomarker analysis has become widely used over the last decade for low- to high-throughput analysis of small volume samples. This assay platform has a broad range of applications, with commercial and lab-developed Luminex-based assays (e.g., mouse, rat, human and non-human primate cytokine panels) being used in basic science discovery, pre-clinical/translational and clinical trial immune monitoring laboratories.

Despite the high-level of integration of Luminex bead-based multiplex assays in pre-clinical and clinical research applications over the past 10 years, there has been no comprehensive analysis of the variables that impact assay performance, nor a national or international proficiency testing program for laboratories performing Luminex-based assays. Efforts to identify inter-laboratory variables which impact assay performance have been a long-standing effort within the Cancer Immunotherapy Consortium (CIC) and NIAID/NIH immune monitoring programs, and have led to implementation of assay harmonization, and optimization across a number of immune assay platforms (van der Burg et

al., 2011) as well as development of proficiency testing programs (Jaimes et al., 2011). Luminex assays in particular are complex and subject to variability due to such factors as assay manufacturer and lot number, kit components, antibody pairs/clones, analyte standards, assay execution, instrumentation and data analysis/interpretation (Khan et al., 2004; Djoba Siawaya et al., 2008; Nechansky et al., 2008; Butterfield et al., 2011; Scott et al., 2011). To address these issues, a collaborative effort between the NIH/NIAID and the CIC of the Cancer Research Institute (CRI) was initiated to specifically assess outcome variation of Luminex bead-based cytokine assays performed in immune monitoring laboratories around the world and to develop an external quality assurance (EQA) program to monitor on-going laboratory performance.

The two-pronged mission of this joint NIAID/CIC Luminex proficiency initiative is to 1) identify variables that significantly impact outcomes of Luminex-based assays; and 2) establish a routine proficiency testing program for this assay platform to facilitate assay quality improvement and harmonization among participating laboratory sites. A central oversight laboratory is critical for achieving the later point. The focus of this manuscript is on the development and implementation of an international EQA program for Luminex bead-based cytokine assays.

## 2. Program development

As a first-step in developing a Luminex EQA program, the NIH/NIAID-sponsored External Quality Assurance Program Oversight Laboratory (EQAPOL) at Duke University developed and implemented a data gathering survey in December 2010. The survey was developed based on draft questions from the Steering Committee and run electronically through a web survey tool. Survey participants were national and international laboratories identified by NIAID and CIC representatives as performing Luminex assays for single or multi-center NIAID- or CIC-supported trials. Topics covered in the survey included the following: instrumentation (plate washers/bead readers), software (acquisition/analysis), desired assay format (polystyrene/magnetic), commercial kit vendor experience, analyte experience, and sample type/matrix experience. Survey results were then used by EQAPOL and the Steering Committee to inform EQA program development.

Participant contact information and survey responses per site were used to populate the EQAPOL Luminex EQA program database.

The initial mandate from the Luminex Steering Committee and NIAID/CIC for the EQAPOL Luminex program was to develop and implement an international Luminex EQA program that would determine laboratory accuracy and precision when given a standard protocol, reagents, and samples. The overall statistical approach to this consisted of the development of a model-based method to assess laboratory performance (i.e. score). A more comprehensive presentation of the statistical methods and models used to determine site reported accuracy to a consensus mean and precision are described in Rountree et al. (2014) in this issue. The results of these statistical analyses, coupled with questionnaire responses, are the basis of proficiency evaluation.

### 2.1. EQA platform and panel design

It was determined that Luminex EQA testing would occur at least twice a year and that a commercial vendor would be solicited to produce a custom 5-plex kit to detect human cytokines: IFN- $\gamma$ , TNF- $\alpha$ , IL-6, IL-10 and IL-2. These five analytes were selected based on survey results that indicated that they were the five most highly quantified serum analytes by participating sites. The kit would use the polystyrene Luminex bead platform, incorporate a 5-analyte pooled lyophilized standard, and be run in a single 96-well filter plate. The choice of polystyrene bead-based kits was due to the fact that the majority of sites did not have the capability for, or experience with, running magnetic bead-based kits at the time of the survey. It is important to note that commercial kits would be de-identified to avoid vendor bias and to avoid the perception that this program was a vendor comparison study or would lead to vendor endorsement by the NIAID/CIC. In the development phase of the EQA, all sites would be required to use the provided kit and follow a common protocol. The decision to require all sites to use the same kit and protocol during initial rounds of testing, allowed for standardization of all site assay platforms and facilitated identification of site-specific variables. Future EQA phases will allow sites to run kits/assays of their choice as a way to gauge routine assay performance in a laboratory.

Each external proficiency panel (EP), would include test samples (human serum, pre-diluted recombinant cytokine standards or human PBMC culture supernatant) prepared by a central laboratory compliant with Good Clinical Laboratory Practices (GCLP). In addition to the assay kit, participating sites would receive three coded replicates of each test sample and be asked to run each of these in triplicate in the assay plate, thereby running nine replicates of each test sample.

Lastly, sites would be assigned a due date by which to provide their final determined pg/mL for each analyte in each sample (Site Reported) and their raw mean fluorescence intensity (MFI) readings for centralized data analysis (EOL-analysis; EQAPOL Oversight Laboratory). In addition, sites would be asked to complete a short questionnaire to provide supporting information regarding instrument and assay performance, problems encountered, and data analysis approaches.

### 2.2. Reagents and assay kits

EQAPOL and the Steering Committee identified a vendor to produce bulk Luminex kits to quantify the five identified human analytes. The Steering Committee decided on the vendor for the initial phases of EQAPOL Luminex EQA based on input from participating sites regarding vendor preferences, followed by evaluation of quality control practices and documents for the top two site-identified vendors.

The EQAPOL Repository team procures all reagents/kits and records all relevant information (quantity, location, lot number, expiration date, QA documents etc.) into the repository module of the EQAPOL web-based application. The assay kit components are de-identified and re-labeled with unique EQAPOL labels. In collaboration with the Luminex EOL, the EQAPOL Repository team also prepares large quantities of test samples for Luminex EQA EPs. Human AB serum test samples are artificially created with known levels of the five target human cytokines using third-party recombinant cytokines. Spike concentrations are purposely selected to span the assay's detectable range of analytes (pg/mL). Mitogen-stimulated human PBMC culture supernatants were also generated as test samples, but phased out after EP1 (see below). Similarly, aliquots of pre-diluted recombinant cytokine standards were prepared as test samples, but phased out after EP2 (see below).

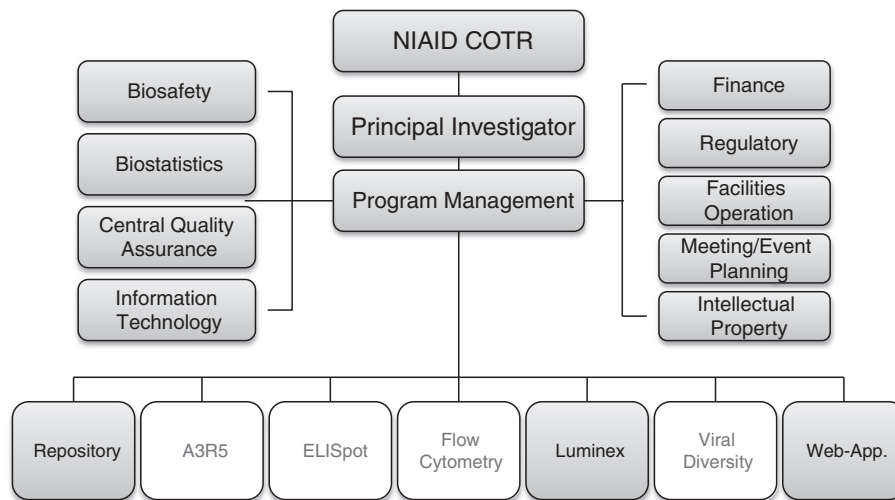
Sample inventory, unique ID, lot number, QA documents, etc. are loaded into the inventory module of the EQAPOL web-based application. Sample stability at  $-80^{\circ}\text{C}$  has been monitored in selected samples over time by the Luminex EOL.

### 2.3. Support for proficiency testing

Continued evolution of the Luminex proficiency tests is informed by input and guidance from the NIAID/CIC Luminex Steering Committee and the overall EQAPOL Scientific Advisory Board. Each round of testing is supported by an EQAPOL infrastructure that includes teams for the following: overall administration and compliance, program management/scheduling, biosafety, biostatistics, quality assurance, and information technology (Fig. 1). Samples for each proficiency panel are produced, tested, and stored under GCLP conditions in the EQAPOL Repository (T. Denny) and the EQAPOL Luminex Oversight Laboratory (EOL; G. D. Sempowski). The expert guidance of the EQAPOL-Central Quality Assurance Unit (CQAU) (see Todd et al., 2013) was critical for establishment of GCLP for all Luminex EQA activities/processes and continues to monitor compliance.

### 2.4. Participating site requirements

Requirements for site participation in the EQAPOL Luminex EQA are as follows: 1) have access to Luminex instrumentation, have prior experience with platform; 2) agree to run a universal kit within two to four weeks of receiving materials; 3) submit all requested data/questionnaires to EQAPOL by EP due date; and 4) be recommended or approved by NIAID or CIC representatives on the Luminex Steering Committee. There is no cost for participation in the current NIAID-sponsored EQAPOL Luminex EQA program. All national and international laboratories that participated in the initial survey study were invited to enroll in the Luminex



**Fig. 1.** EQAPOL administrative structure and support teams. Highlighted along the bottom are the key groups involved in executing the international Luminex EQA.

EQA program. At present the program is monitoring 25 sites that represent the North America, Europe, Africa, and Asia (Fig. 2). Participating sites represent a spectrum from basic research laboratories to immune monitoring core to biotech/pharmaceutical companies.

### 2.5. Biannual proficiency panel and send-out

In collaboration with EQAPOL program management, each EP concept is formally developed into a comprehensive study plan. Study plans are reviewed and approved by EQAPOL leadership and the EQAPOL CQAU, and contain an overview of EP organization/management, list of participating sites, proficiency panel design (kit and samples), assay details, data submission process, confidentiality protections, statistical analysis plan, and result reporting process.

Attachments to the study plan are a list of selected samples from the repository, a randomized plate layout, a confidential data key for the web-application to link coded assay results to sample identifiers in the database, a detailed kit-specific assay protocol, data submission templates, and EP orientation/training materials. EPs are designed so that there is always at least a subset of samples that are common to the previous EP to allow for tracking of EP-to-EP performance. All EP materials are pre-tested in the Luminex EOL on at least one Luminex reader to confirm protocol validity and assay kit performance, and to ensure that test samples perform within tolerance ranges with respect to observed pg/mL of each target analyte. It is important to note that EQAPOL overall has taken the position that the central laboratories are not the gold-standard reference for each testing panel (see Rountree et al., 2014 “Statistical Methods for the Assessment of EQAPOL Proficiency Testing: ELISpot, Luminex, and Flow Cytometry”). For each EP, pre-test assay results from the Luminex EOL are uploaded along with all participating site data to establish a consensus observed mean pg/mL for each analyte in each test sample.

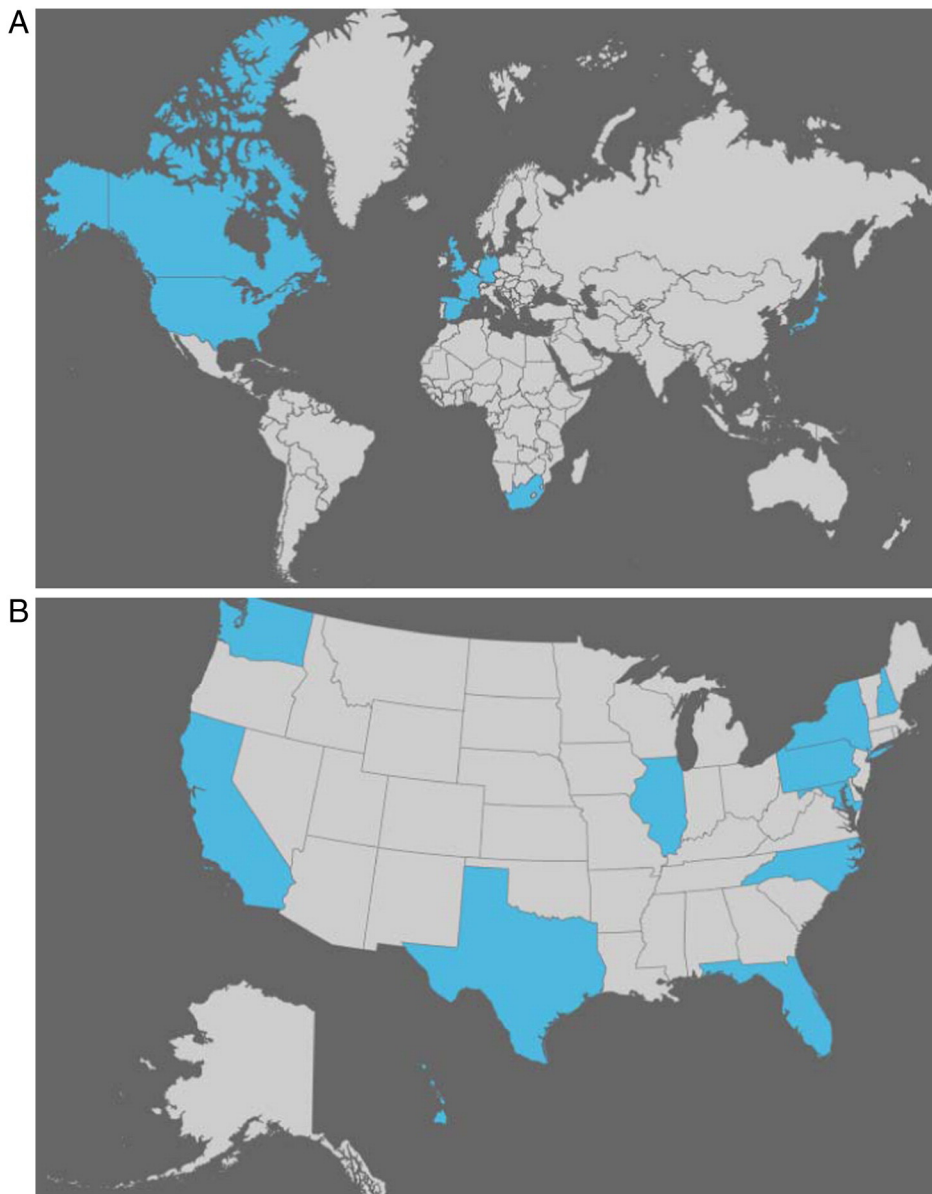
EQAPOL program management sets the EP send-out schedule, alerts all participating sites, and schedules EP-specific orientation and training teleconferences. Each EP send-out for the Luminex EQA consists of a wet ice (assay kit, protocol and plate) and a dry ice (test samples) shipments. Both shipments contain detailed shipping manifests and commercial invoices.

EP execution has been substantially enhanced by the development of the EQAPOL web application. For each EQAPOL EQA program the application maintains a comprehensive database of participating site contact and shipping information, sample, reagent and assay kit inventory, details of all completed, active and pending EPs, shipping records, results, proficiency scores, and remediation notes. For each EP send-out, the EQAPOL web application electronically sequesters EP kit reagents and test samples from inventory and alerts the EQAPOL Repository team to prepare shipments. In addition, the system notifies participating sites of their pending EP shipment. EP-specific instructions, documents, protocols, templates etc. are uploaded and made available via the application for easy retrieval by participating sites.

All wet ice shipments are monitored with an electronic temperature data logger that is sent back to EQAPOL upon package receipt. International dry ice shipments are replenished, if needed, by a specialized courier (i.e., World Courier). All sites are provided electronic copies of their manifests and FedEx/Courier tracking numbers to facilitate package tracking. Upon package receipt, participating sites log in to the EQAPOL web application and indicate receipt of their shipments. Based on the date of receipt, the site's EP due date is assigned according to the timeline specifications in the EP study plan.

### 2.6. Data collection and analysis

The data submission process has evolved with development of the EQAPOL web application and central database. In its simplest form all participating sites are required to



**Fig. 2.** Location of sites participating in EQAPOL Luminex EQA. (A) International. (B) Domestic, United States of America.

provide three essential data sets: 1) a table of final observed pg/mL for each of the five selected analytes in each of the blinded samples; 2) a table of raw MFI values for each analyte in each plate well used in the given EP; and 3) response to a short online questionnaire regarding instrument set-up/QC, assay performance/curve limits, problems encountered, and data analysis approaches. According to Luminex Steering Committee suggestion, the EQAPOL EP protocol does not specify the acquisition or analysis software to be used, nor does it specify how best to establish curve fits or address outliers. These are assumed to be critical site-specific variables that may impact assay outcomes.

In EPs 1 and 2, data sets were emailed back to the EQAPOL Program Management team for integration into the database by the EQAPOL Biostatistics group. With later enhancements

to the secure EQAPOL web application, data from EPs 3, 4, and 5 were submitted using the web application. The submitted files are validated for data format and then imported into the database. The EP-specific questionnaire is also integrated into the results/data submission module of the web application. The power of this tool is that site result files and questionnaire answers are readily available to authorized members of the EQAPOL teams for analysis and scoring.

### 2.7. Proficiency evaluation (Scoring)

Four main criteria are being used to evaluate assay proficiency in the EQAPOL Luminex EQA program: timeliness, protocol adherence, accuracy to the consensus, and precision. These criteria are presented in Table 1 and further descriptions

can be found below. Criteria are weighted such that there is a maximum of 100 points to be earned in a given EP, the majority (80 points) of which are allotted to accuracy and precision. Performance ranges of “excellent, good, fair and poor” are detailed in Table 1. These standard ranges have been adopted by all EQAPOL EQA programs. The *Points Earned Score* and the *Performance Ranges* are to aid sites, EQAPOL, and sponsors in identifying areas of strength and weakness in site proficiency with respect to Luminex assay performance. Mock scoring was done for EP2, and actual proficiency scoring has been done since EP3.

#### 2.7.1. Timeliness (10 points)

Sites are expected to complete an EP panel and provide valid data via our online web application by a specified due date, as detailed above. Invalid data (i.e., data in the wrong format) can significantly degrade analysis of both site-reported and centrally analyzed results. Failure to upload valid data and survey responses by the due date will result in loss of all proficiency points for timeliness.

#### 2.7.2. Protocol adherence (10 points, split)

The ability of a laboratory to accurately set up their instrumentation, follow a specified protocol, and provide data in the correct format is essential to the success of this program and to proficiency of laboratories working in a multi-center consortium. Five Protocol Adherence points are awarded (all or none) for adherence of instrument set-up, plate layout, and assay procedures based on data file review and EP questionnaire responses. A specific plate layout is required for each EP panel so that all incoming data can be centrally analyzed, and all sites must follow a single standard protocol to ensure consistency among sites. Failure to comply with these expectations results in a deduction from an overall proficiency rating. Even minor protocol changes can result in point deductions.

Five additional points are awarded for standard curves that fall within the acceptable fit range with 1 point available for each of the five analyte curves. Site-reported MFI data and expected standard concentrations are used to generate EOL-derived standard curves for each analyte. A four-parameter logistic (4PL) function is fit in SAS 9.2 using Proc NLIN. Fit probability is calculated using the weighted sum of squared errors (SSE). The SSE follows a chi-square distribution with two degrees of freedom (number of curve points minus the number of parameters). A p-value is obtained to assess fit.

#### 2.7.3. Accuracy to the consensus (40 points)

Accuracy to the EP consensus mean analyte concentration is assessed for site-reported pg/mL data to determine whether a site can accurately quantify the concentration of an analyte in a sample. Site-reported data is natural log-transformed for analysis, and a mixed effects model is used to estimate whether a site-reported concentration value is significantly different at the alpha 0.05 level from the other participating sites using a Bonferroni correction (Benjamini, 1995). The number of evaluations (i.e., number of unique test samples multiplied by 5 analytes) of accuracy varies per EP, and therefore, the point value fluctuates to maintain the total 40 point weighting for accuracy.

#### 2.7.4. Precision (40 points)

As previously noted, each blinded sample has nine replicates (three replicate wells per sample, three blinded replicates for each sample). As an estimate of precision, the upper bound of the 1 SD range around the mean for each sample and analyte is used to evaluate sites. Observations per site, per sample, per analyte that are above the established 1 SD range will result in point deductions. The number of evaluations (i.e., number of unique test samples multiplied by 5 analytes) of precision varies per EP, and therefore, the point value fluctuates to maintain the total 40 point weighting for precision.

#### 2.7.5. Central analysis

To identify site-specific data analysis approaches that may be impacting accuracy or precision, all site-provided raw MFI data are centrally analyzed (EOL) using a standardized 4PL curve fit in SAS 9.2 with Proc NLIN. The resultant EOL-generated observed pg/mL per sample, per analyte is used to calculate an EOL-accuracy and an EOL-precision (as above). These data are then used to determine an overall EOL-Score. This EOL-score is not used for proficiency rating. Rather, it is used internally by the EQAPOL Luminex team to assist with site remediation.

#### 2.7.6. Reporting results to participating laboratories

All site proficiency scores, data analysis charts/tables and site-specific comments are uploaded into the EQAPOL web application/database. Comprehensive reports are automatically generated by the application as a PDF document incorporating site-specific scores and comments. Participating laboratories are notified when reports are available online.

**Table 1**  
Luminex EQA grading criteria and point distribution.

Criteria	Target	Max. points
Timeliness	On time upload of valid EP data and survey	10
Protocol adherence	Followed prescribed instrument setup/QC, plate layout, assay procedure, and analysis instructions	5
	Acceptable fit probability of site-reported standard curve(s)	5
Accuracy	Site reported results are not significantly different than consensus mean	40
Precision	Well to well variability (9 estimates per sample; % CV, site reported results for EP5)	40
Overall performance		
Excellent		91–100
Good		75–90
Fair		66–74
Poor		0–65

### 2.7.7. Site performance remediation

Sites that receive a rating of “fair” or “poor” are requested to have a one-hour remediation teleconference with the EQAPOL Luminex team to review performance and identify potential factors negatively affecting proficiency. Site submitted data, central EOL analysis, and post-run questionnaire responses are used by EQAPOL Luminex staff to investigate possible causes of low performance. Specifically, a comparison of submitted data and centrally-analyzed data can indicate a failure to incorporate the required sample dilution factor into analysis software. Decoding of the blinded site-submitted data can indicate errors in the sample order, either due to inaccurate plate loading or mishandling of data when preparing data templates for upload. Questionnaire responses provide information about instrument settings, technician experience, and technical difficulties faced while running the assay, all of which can impact performance. Any common problems identified during site remediation are emphasized in subsequent EP protocols and training materials. Overall, the most common barriers to success have been dilution factor omission, plate leaking/clogging, and failure to submit data on time.

## 3. Results

### 3.1. EP1

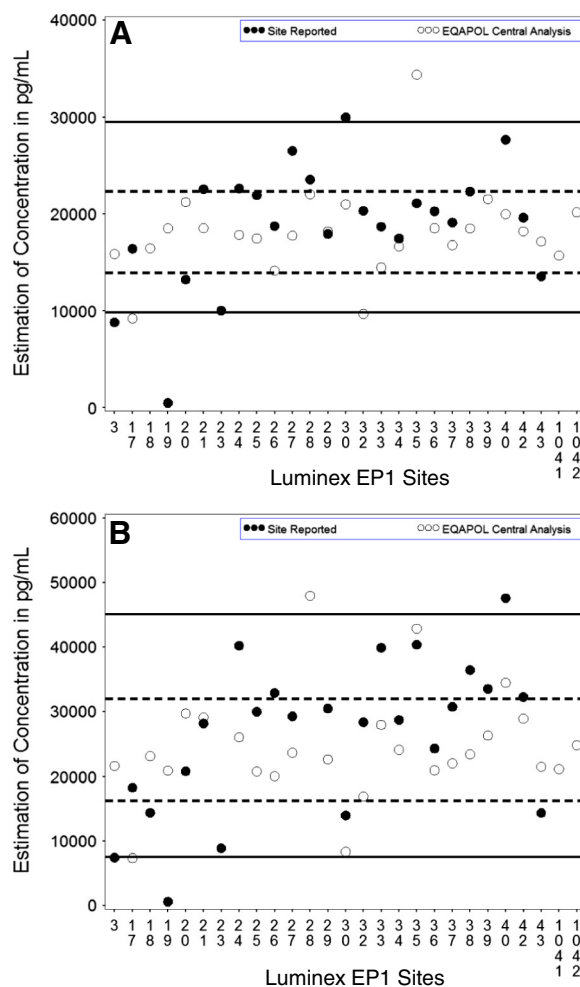
Twenty-five laboratories participated in EP1. All program participants received a de-identified custom 5-plex Luminex-bead assay kit and two test samples containing known amounts of the five target cytokines (IFN $\gamma$ , TNF $\alpha$ , IL-6, IL-10 and IL-2). One test sample was culture supernatant from PMA/ionomycin-stimulated human PBMCs, and the other was human AB serum spiked with known concentrations of recombinant cytokines. These samples were created to contain high cytokine concentrations so that sample dilution would be required for analyte quantification using the Luminex assay. Sites also received pre-diluted assay standards provided in two matrices: serum and culture medium. The goals for this EP were straightforward: 1) work through the logistics of the send-out and ensure that all sites could efficiently receive intact, useable assay reagents and samples; 2) assess site capability in setting up an assay with pre-diluted standards; 3) assess site capability in generating a dilution series of test samples; and 4) determine distribution of reported pg/mL for the five analytes in the samples.

Some shipping delays due to international custom regulations were encountered in EP1, but these issues were easily resolved for future EPs. The Steering Committee was concerned that if a site was unable to perform a recombinant standard dilution series, then the entire assay would fail. Therefore, it was decided that in EP1 pre-diluted frozen aliquots of standards would be provided to the sites ready to use. This allowed all sites to have essentially identical standard curves for their analytes. To monitor the ability of the sites to generate a dilution series, they were asked to serially dilute the provided test samples. Curve fit analysis of these data revealed that all sites were able to generate a dilution series with acceptable curve fit probabilities (data not shown), so future EPs employed a lyophilized standard

that would require reconstitution and serial dilution at each site (industry standard).

Analysis of the site reported pg/mL for both test samples (spiked-serum and culture supernatant) revealed a high level of variability in observed pg/mL (Fig. 3). This was striking as all sites received the same kit, standards, samples and protocol. Furthermore, there was a distinct lack of overlap between site reported pg/mL and EQAPOL centrally calculated pg/mL. This later point suggested a lack of consistency among sites on how data analysis is performed to determine observed pg/mL.

As a follow-up to EP reports, teleconferences were held with the majority of sites to get feedback on the EP design and process. There was an overwhelming desire for the program to use only serum as the matrix/sample type, eliminate sample dilution, provide lyophilized standards, and provide more training on the assay protocol and data transfer process. A common reported problem with the polystyrene bead assay was plate leaking and/or clogging.



**Fig. 3.** EP1 reported concentration of IFN $\gamma$  in serum (A) and activated PBMC culture supernatant (B). Solid lines indicate the upper and lower bounds for site reported data. Dashed lines indicate upper and lower bounds for data from central analysis.

**Table 2**

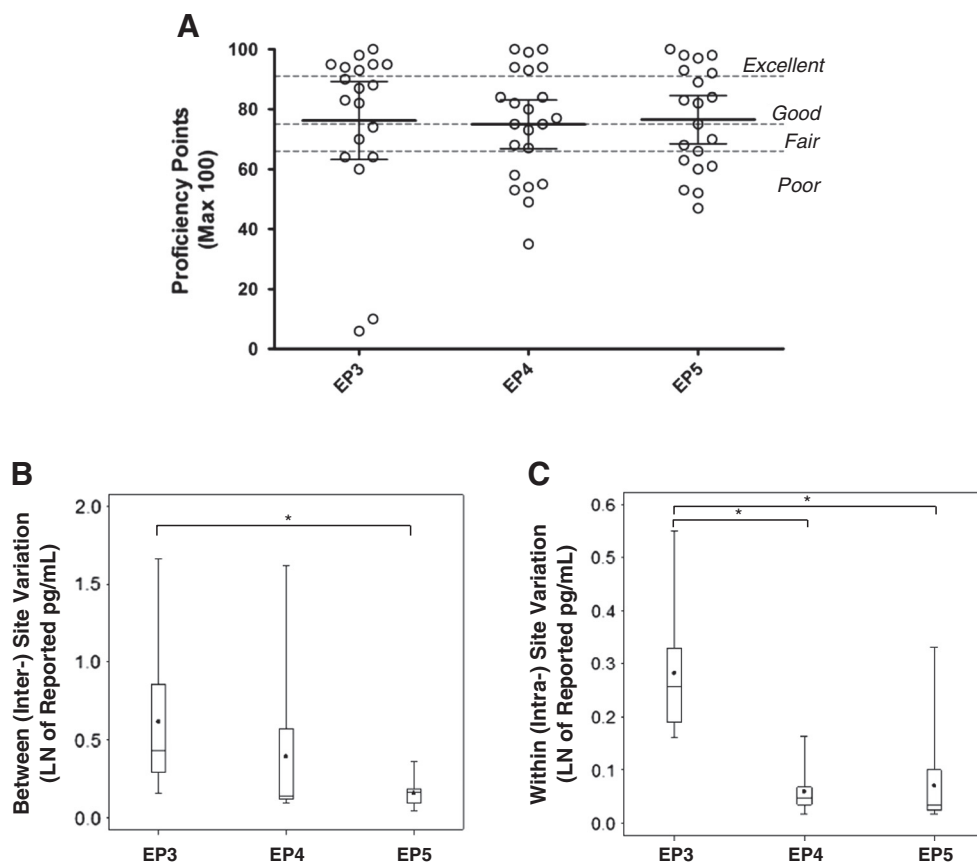
EP2 summary of performance (draft criteria/targets).

Criteria	Target	Results (average for 24 sites)
Timeliness	On time upload of valid EP data	21/24 sites on time (88%)
	On time submission of EP survey	20/24 sites on time (83%)
Protocol adherence	Followed prescribed instrument setup/QC, plate layout, assay procedure, and analysis instructions	15/24 sites complied (63%)
	Acceptable fit probability of all five site-reported standard curve(s)	19/24 sites (79%)
Accuracy	Average site reported results within 95% CI boundaries of EP consensus	45/55 evaluations (82%)
Precision	Average acceptable well to well variability (% CV, 3–9 EOL-determined estimates per sample)	11/15 evaluations (74%)

### 3.2. EP2

EQAPOL Luminex EP2 focused on establishing a proficiency grading system to be used for future EP rounds. Luminex EQA program participants received a de-identified custom 5-plex Luminex-bead assay kit that included all necessary reagents to run the assay, lyophilized standards and 19 blinded test samples containing the five target cytokines at varying concentrations. The NIAID/CIC Luminex Steering committee and EQAPOL developed four main criteria to assess Luminex assay proficiency, as shown in Table 1. These

draft criteria and results from the 24 EP2 participating sites are detailed in Table 2. These data provided descriptive analyses and a general impression of how sites compared to one another, and established a baseline data set for development of a point scheme by the EQAPOL statistical group (Table 1). As described above, a weighted point system was developed for an overall assessment of site proficiency in performing Luminex bead-based cytokine assays. This schema was subsequently proposed to and ratified by the NIAID/CIC Luminex Steering Committee and the EQAPOL Scientific Advisory Board for use in future EPs, as shown in Table 1.



**Fig. 4.** Site performance across three scored EPs. (A) Individual site results. Performance ranges: Excellent (91–100), good (75–90), fair (66–74), and poor (0–65). (B) Between (inter-) site variation. (C) Within (intra-) site variation. Box and whisker plots represent the 25th and 75th percentiles, the line represents the median and the end of the whiskers is drawn to the most extreme values. The dot represents the mean. Note: Two sites were excluded from this analysis, as they received zeros for failure to complete a shipped EP (LN = natural log-transformed).



### 3.3. EP3, EP4, and EP5

EQAPOL Luminex EPs 3, 4, and 5 were run approximately six months apart and incorporated the comprehensive scoring schema for quantitative assessment of assay proficiency. Participating sites received proficiency testing panel kits and reagents as described above (Section 2). Although there were four samples that were the same in EPs 3, 4, and 5, new samples were included in EPs 4 and 5 that were not in EP3. Proficiency points earned for all participating sites in these three EPs are shown in Fig. 4A. The overall mean proficiency score has remained stable in the “good” range from EP3 to EP5. A more detailed analysis of the site reported results indicated that there has been a significant improvement in within (intra-) and between (inter-) site variation (i.e., reduction in variability) (Fig. 4B and C).

Linear mixed effects models (Fitzmaurice et al., 2004) were used to assess the within-site and between-site variabilities over EPs 3, 4 and 5. There were 20 models run, one for each sample by analyte combination. These models included a random and repeated statement in Proc Mixed, which allows for direct estimation of between-site and within-site variances. These two variance components were appended into two separate data sets with 60 observations each (three EPs by 20 models). These variance estimates were then compared among the three EPs using a nonparametric test via Proc Rank and Proc GLM (Conover, 1980) in SAS 9.2. It was found that between-site variance for EP3 is significantly larger than that for EP5, but not significantly larger than that for EP4 (Fig. 4B). It was also determined that within-site variance for EP3 is significantly larger than that of EP4 and EP5 (Fig. 4C). There were no statistically significant differences in between-site or within-site variance comparing EP4 and EP5.

A summary of site participation, scores, and remediation requirements is shown in Table 3. Although lack of timeliness and/or protocol adherence sometimes negatively impact proficiency, the scoring rubric was designed so that a site will not fall below the “good” proficiency range if they have no problems with accuracy or precision. Therefore, areas for site remediation always include improvement of accuracy to the consensus and/or precision.

Sites performing poorly with respect to accuracy are typically struggling with lyophilized standard reconstitution and a general lack of familiarity with their post-run data analysis software, specifically the incorporation of sample dilution factors, bead region identifiers and expected concentrations of standards. Dilution factor problems are readily identified by the approach taken of comparing site-reported scores with EOL-derived scores. These errors are easily corrected and, in most cases, are not recurring problems.

The more challenging area for remediation is assay precision. The predominant site-suggested source of poor precision is leakage and clogging of the assay filter-plate (required for polystyrene bead platform). The EQAPOL EQA program is currently pilot-testing magnetic Luminex bead-based assay kits to determine if kit platform/vacuum washing is indeed a significant source of reduced precision (manuscript in preparation). Site technician turnover and lack of familiarity with the EQA assay have also been identified as correlates with poor precision. These later issues are addressed on a site-by-site basis with targeted training.

### 4. Overall summary

The NIH/NIAID-supported EQAPOL at Duke University has developed and launched the first international EQA/proficiency testing program for Luminex bead-based human cytokine assays. This is the result of a collaborative effort between the EQAPOL, NIH/NIAID and CIC and currently monitors 25 US domestic and international sites with two scored EP panels per year.

Comprehensive scoring based on timeliness, protocol adherence, accuracy, and precision is performed with mixed effects model-based statistics and allows for identification of sites in need of remediation. Approximately 41% of sites across the initial three rounds of scored testing have required remediation by the central laboratory. Overall mean proficiency scores across the three initial rounds have remained relatively stable at ~75/100 points. A more detailed analysis of site reported results, however, indicated a significant improvement in within (intra-) and between (inter-) site variation, suggesting that remediation for poor performing sites and program training may be having a positive impact on assay proficiency or that sites are improving in assay proficiency through assay familiarity and protocol repetition. Poor and fair performing sites typically see point reductions across all four score categories, but the majority of reductions are seen in the area of precision.

Timeliness and protocol adherence are readily corrected with training and continue to be a focus in all EP orientation sessions. Accuracy to the consensus pg/mL for each analyte is typically improved with stricter protocol adherence, equipment

**Table 3**  
EQAPOL Luminex EQA site remediation.

Site #	EP3	EP4	EP5
003	99	81	100
017	<b>61</b>		
019	83	<b>49</b>	<b>61</b>
020	<b>72</b>	94	<b>53</b>
021	<b>22</b>	<b>53</b>	DNF
022		84	<b>52</b>
023	92	<b>55</b>	DNF
024	87	<b>35</b>	89
025		99	
026	89	80	84
027	100	84	92
028	98	100	<b>60</b>
029	<b>59</b>	<b>58</b>	98
030		<b>73</b>	
033	95	93	98
034	<b>71</b>	75	<b>66</b>
035	98		
036	92	<b>68</b>	<b>63</b>
037		<b>54</b>	82
038	95	77	100
039	95	94	97
040	<b>73</b>	<b>75</b>	<b>47</b>
042	94	82	83
043	<b>61</b>	84	93
045		100	<b>68</b>
046		<b>67</b>	<b>75</b>
047			<b>70</b>
Average	82	76	78

Blank = site that did not participate.

DNF = site that participated but was not able to complete the EP.

Bold = site that required remediation.

calibration, and focused attention on site-performed data analysis. Assay precision is a more complicated parameter to remediate. The two sources of poor precision identified in this program are the filter plate/vacuum manifold and the laboratory technician performing the assay. The polystyrene bead platform has chronic plate leaking and clogging issues that appear to be impacting precision. A migration of the Luminex EQA to magnetic kits in the near future will alleviate this confounding issue due to elimination of filter plate use. Overall attention to proper and consistent laboratory techniques at each site will undoubtedly improve assay precision.

In conclusion, the overall goal of this program is to develop and use proficiency testing to identify variables affecting Luminex bead-based assay outcomes in the international community of immune monitoring laboratories. To our knowledge, this is the first proficiency testing program to be developed for the Luminex platform. As we look to future development for the program, our ultimate goal would be to have sites perform their own in-house assay as part of the proficiency testing scheme. In order to achieve this goal, we will need to identify ways to assess performance across different manufactures and kit lots. By continuing to build upon this program as we move towards this goal, we hope to play a critical role in improving harmonization for this platform.

#### Acknowledgments

Current and past members of the NIAID/CIC Luminex Steering Committee include: Dr. Patricia D'Souza, representing NIAID; Drs. Michael Kalos and Michael Pride, representing CIC; Dr. Lisa Butterfield, University of Pittsburg and Dr. Gregory Sempowski, Duke University. We are grateful for the guidance and leadership of Dr. Jim Lane (NIAID). The authors are grateful to Jeff Lovingood (EQAPOL Program Management), Ambrosia Garcia, Linda Walker, Sara Brown, Holly Alley and Jennifer Baker (EQAPOL Repository), Dr. Marcella Sarzotti-Kelsoe and Chris Todd (EQAPOL CQAU), Dr. Nathan Vandergrift and John Bainbridge (EQAPOL Biostatistics) and Paul Morrow (EQAPOL Luminex). Finally, the authors thank the anonymous CIC and NIAID sites for their participation, patience with the development of this program, and their thoughtful suggestions along the way.

This project was funded by the Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract No. HHSN272201000045C entitled "External Quality Assurance Program Oversight Laboratory (EQAPOL)."

#### References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289.
- Butterfield, L.H., Potter, D.M., Kirkwood, J.M., 2011. Multiplex serum biomarker assessments: technical and biostatistical issues. *J. Transl. Med.* 9, 173.
- Conover, J.W., 1980. *Practical Nonparametric Statistics*, 2nd ed. John Wiley & Sons, New York, NY.
- Djoba Siawaya, J.F., Roberts, T., Babb, C., Black, G., Golakai, H.J., Stanley, K., Bapela, N.B., Hoal, E., Parida, S., van Helden, P., Walzl, G., 2008. An evaluation of commercial fluorescent bead-based Luminex cytokine assays. *PLoS One* 3, e2535.
- Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2004. *Applied Longitudinal Analysis*. Wiley-Interscience, Hoboken, NJ.
- Jaimes, M.C., Maecker, H.T., Yan, M., Maino, V.C., Hanley, M.B., Greer, A., Darden, J.M., D'Souza, M.P., 2011. Quality assurance of intracellular cytokine staining assays: analysis of multiple rounds of proficiency testing. *J. Immunol. Methods* 363, 143.
- Khan, S.S., Smith, M.S., Reda, D., Suffredini, A.F., McCoy Jr., J.P., 2004. Multiplex bead array assays for detection of soluble cytokines: comparisons of sensitivity and quantitative values among kits from multiple manufacturers. *Cytometry B Clin. Cytom.* 61, 35.
- Nechansky, A., Grunt, S., Roitt, I.M., Kircheis, R., 2008. Comparison of the calibration standards of three commercially available multiplex kits for human cytokine measurement to WHO standards reveals striking differences. *Biomark. Insights* 3, 227.
- Rountree, W., Vandergrift, N., Bainbridge, J., Sanchez, A.M., Denny, T.N., Jan 20 2014. Statistical methods for the assessment of EQAPOL proficiency testing: ELISpot, Luminex, and flow cytometry. *J. Immunol. Methods*. <http://dx.doi.org/10.1016/j.jim.2014.01.007>, pii: S0022-1759(14)00017-9.
- Scott, M.E., Wilson, S.S., Cosentino, L.A., Richardson, B.A., Moscicki, A.B., Hillier, S.L., Herold, B.C., 2011. Interlaboratory reproducibility of female genital tract cytokine measurements by Luminex: implications for microbicide safety studies. *Cytokine* 56, 430.
- Todd, C.A., Sanchez, A.M., Garcia, A., Denny, T.N., Sarzotti-Kelsoe, M., Oct 9 2013. Implementation of good clinical laboratory practice (GCLP) guidelines within the External Quality Assurance Program Oversight Laboratory (EQAPOL). *J. Immunol. Methods*. <http://dx.doi.org/10.1016/j.jim.2013.09.012>, pii: S0022-1759(13)00268-8.
- van der Burg, S.H., Kalos, M., Gouttefangeas, C., Janetzki, S., Ottensmeier, C., Welters, M.J., Romero, P., Britten, C.M., Hoos, A., 2011. Harmonization of immune biomarker assays for clinical studies. *Sci. Transl. Med.* 3, 108 (ps44).