Validation of peptide epitope microarray experiments and extraction of quality data

Tatjana Nahtman a, Alexander Jernberg b, Shahnaz Mahdavifar b, Johannes Zerweck c, Mike Schutkowski c, Markus Maeurer b,⁎, Marie Reilly a,⁎

a Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden
b Microbiology and Tumor Cell Biology Center, Karolinska Institutet, Sweden
c JPT, Berlin, Germany

Received 19 February 2007; received in revised form 12 July 2007; accepted 16 July 2007
Available online 13 August 2007

Abstract

Introduction: Within the last decade, the development of antigen microarray slides has enabled the simultaneous measurement of serum reactivity to hundreds of peptides in a single biological sample. Despite this considerable scientific progress, many issues remain regarding the quality, analysis and interpretation of the data these slides produce. There is currently no accepted approach to guide data analysis, and researchers use a wide variety of statistical methods and software tools. We designed and implemented a laboratory experiment to assess the reliability and range of measurement of peptide microarray data, and present graphical and statistical procedures for pre-processing so that quality data can be extracted for addressing biological hypotheses.

Methods: Synthetic peptides spanning the proteins Ag85A, Ag85B, CFP10, MPT51/MPB51, TB10.4 and ESAT-6 were chosen as a paradigm to screen for serum reactivity to Mycobacteria tuberculosis (MTB). We explored various quantitative and graphical methods for presenting the responses from a slide. We replicated assays of samples from five TB-positive individuals to examine reproducibility, and used linear mixed models to investigate the various sources of variability, and to assess the range of measurement. We use our methods to extract data from the five TB-positive individuals and five healthy controls, and analyse the “normalized” responses using the freely available SAM package.

Results: The ratio of foreground to background signal (on a log scale) provides an appropriate response index. A two-dimensional graphical display clearly illustrates the responses from the control and peptide features on a slide. Mixed model analysis of the replicated slides found a high reproducibility of the assay between operators, days and experiments. The range of measurement was also satisfactory. Our analysis of the normalized responses from the five TB-positive patients and five healthy controls suggested that 10 of the 363 peptides assessed had significantly higher responses in the TB-positive group.

Conclusions: Carefully designed laboratory experiments and rigorous statistical analysis can enable the removal of technical artefacts to produce quality peptide array data for addressing biological hypotheses. These instruments, which enable valid

Abbreviations: TB, tuberculosis; MTB, mycobacteria tuberculosis, M.tuberculosis (M.tub); MOTT, mycobacteria other than tuberculosis (MOTT); ESAT-6, early secreted antigenic target 6-kDa protein; FCS, fetal calf serum; FDR, false discovery rate; CFP10, culture filtrate protein 10; Ag85A, antigen 85A, a member of the Ag85 protein complex (Ag85A,-B, -C); MPT51, mycobacterial protein secreted by M.bovis; MPB51, BCG/mycobacterial protein secreted by M.tuberculosis; GAL, Genepix Array List; GPR, GenePix Result; LSMEANS, least squares means; BLUP, best linear unbiased predictor.

⁎ Corresponding authors. Reilly is to be contacted at Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Nobels väg 12A, SE 171 77, Stockholm. Tel.: +46 8 5248 3982; fax: +46 8 31 49 75. Maeurer, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Nobels väg 18, SE 171 82, Stockholm. Tel.: +46 8 457 26 50; fax: +46 8 33 74 60.
E-mail addresses: Markus.Maeurer@ki.se (M. Maeurer), marie.reilly@ki.se (M. Reilly).

0022-1759/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jim.2007.07.015
comparisons across slides and/or batches of slides, will escort future comparative analyses targeting high content serum reactivity profiling against a broad array of B-cell epitopes.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Peptide microarray; Statistical analysis; Normalization; Mixed models

1. Introduction

In the 1990s, endeavours to gauge specific signatures of immune responses progressed from testing immune reactivity to a few antigens to measuring multiple and more complex immune responses, e.g. in Elispot assays. Recent developments have made it possible to measure immune responses in even greater detail. Glass slides, called antigen arrays, are prepared by fixing hundreds or even thousands of proteins or peptides on a solid surface. After incubation with patient serum, a digital scanner provides raw data that reflects the antibody immune response to each individual antigen in the form of fluorescence intensities (see Fig. 1). Nowadays, such arrays are developed and tested for their potential in the detection of immune responses directed against infectious agents and in the monitoring of immune responses in vaccine trials (Neuman de Vegvar et al., 2003).

The development of a peptide microarray assay requires extensive validation related to the comprehensive assessment of antigen/peptide reactivity, and establishing robust assay performance. However, the rapid development of microarray technology has not been accompanied by a simultaneous development of appropriate statistical methods to guide data validation, extraction, analysis, and linkage of results with biological processes in the immune system. There is currently no accepted approach to guide data analysis in complex immunological systems, and researchers use a wide variety of statistical methods and software tools.

Fig. 1. Schematic representation of the capture of antibody responses from peptide microarray slides.
Antigen arrays have many similarities with their genomic microarray precursors, so that some of the available methodologies and software tools developed for genomics data may be useful. However, antigen peptide arrays differ in several important ways from genomic microarrays, including their design and the nature and interpretation of the generated signals, all of which impact on extraction of meaningful data for analysis. For example, the normalization of genomic microarray data routinely assumes that only a few genes or a limited set of genes are involved in the process being studied, and the analytical approaches are tailored to identifying these special few. In contrast, the availability of peptide antigen microarrays makes it possible to appreciate global patterns of immune reactivity in the immune response, and there is no a priori reason to assume that the response will be focused or directed against just a few epitopes. Another important feature of antigen arrays is the use of a comprehensive set of negative and positive control spots. Negative controls provide a benchmark against which the background of all signals generated by peptide locations can be assessed, while positive controls enable the assessment of the quality of the foreground signal, and also provide a means of making adjustments for variation within and between slides. Antigen array data should thus be handled with an understanding of the technology and measurement process, and not just subjected to “data mining” using any available bioinformatics tools.

Technological developments drive the field of immunomics rapidly, and analysis of resulting antigen microarray data is proceeding in the absence of adequate development of accepted methods of data quality assurance. For example, the visual inspection and removal of outlying observations from a peptide microarray slide is not simple, although it is not uncommon to see reports of such an exercise. For some researchers, another step in data cleaning is the removal of any peptides with foreground less than background (i.e., a negative response) although this may disturb the distribution of the remaining responses and the behaviour of algorithms used to analyse them. To “normalize” each slide in order to obtain “comparable values”, some researchers simply rescale the measured responses to a constant level (Neuman de Vegvar et al., 2003) or subtract the median of all antigens on the same array (Quintana et al., 2004).

In this paper we design and implement a validation study to assess the reproducibility and range of measurement of peptide responses in antigen microarrays. The various steps in data screening can be guided by intuitive statistical graphics and models that provide the analyst with clear and quantifiable decisions. We provide methods to quantify and remove biases due to between- and/or within-slide effects, background intensities, and other sources of systematic error. Our mixed model approach uses the information in the negative control, positive control, and peptide responses, where the replicated blocks on each slide provide important information for the removal of variation that may be due to the measurement process (Kerr et al., 2000; Wolffinger et al., 2001), and the remaining variability can be reasonably attributed to measurement noise. Using these techniques, which are available in standard statistical packages and well known to trained statisticians, enables the extraction of quality normalized peptide array data for use in the investigation of various biological hypotheses.

In the following section, we describe the methods used for preparing the slides, and outline the approach to the analysis of the responses, starting with the identification of a suitable transformation and mode of display. In the Results section, we describe the precision and range of measurement of the assay, and present graphical displays and results of statistical modelling of the transformed responses from negative controls, positive controls and peptides, finishing with normalized peptide responses for each of the individuals included in the validation study. Finally, we provide an illustrative example where the normalized peptide responses of the TB+ patients included in the validation study are compared with the normalized responses of a group of healthy controls, using a standard software package that is available for free download. The Discussion section presents some implications of our work and its relevance for scientific investigation using peptide microarray data. We used TB-associated antigens as a paradigm, since recent data indicate that TB-specific antibody responses may be helpful in differentiating between latent and active TB (Banerjee 2004).
These antigens may also be helpful in monitoring future vaccine trials targeting molecularly defined TB antigens.

2. Materials and methods

2.1. Peptides

Subtractive hybridization yielded the identification of the RD1 region in mycobacterial species, which is present in M. tuberculosis, in pathogenic M. bovis strains, but lacking in all BCG (bacille Calmette-Guérin) strains and in almost all MOTTs (Behr et al., 1999). Two antigens encoded by RD1 are the early secreted antigenic target 6-kDa protein, ESAT-6 (Arend et al., 2000) and the culture filtrate protein 10, CFP10 (Berthet et al., 1998). ESAT-6 belongs to a family of 22 low molecular mass proteins, most of them strongly immunogenic. One of the subfamilies is TB10.4 (Skjot et al., 2002). The antigen 85A (Ag85A), a member of the Ag85 protein complex (Ag85A,-B and-C) with mycolyltransferase activity, represents one of the most potent antigens and a lead vaccine candidate (Huygen et al., 1996). The antigen MPT51/MPB51 (mycobacterial protein secreted by M. bovis, BCG/mycobacterial protein secreted by M. tuberculosis) binds fibronectin and elicits strong immune responses (Ohara et al., 1995). For the work presented here, we used “first-generation” peptide microarray slides containing 15 mer peptides, with 5 aa overlaps, which cover the entire proteins listed in Table 2.

2.2. Preparation of micro-arrays

Amino-oxy-acetylated peptides were synthesized on cellulose membranes in a parallel manner using SPOT synthesis technology (Frank, 1992; Wenschuh et al., 2000). Following side chain deprotection the solid phase bound peptides were transferred into 96-well microtiter filtration plates (Millipore, Bedford, USA) and treated with 200 uL of aqueous triethylamine (0.5% by vol.) in order to cleave the peptides from the cellulose support. Peptide-containing triethylamine solution was filtered off and solvent was removed by evaporation under reduced pressure. Resulting peptide derivatives (50 nmol) were re-dissolved in 25 microliter of printing solution (70% DMSO, 25% 0.2 M sodium acetate pH 4.5, 5% glycerol; by vol.) and transferred into 384-well microtiter plates. Two droplets of 0.5 nL peptide solution (1 mM) were deposited per spot on epoxy functionalized glass slides (Corning Epoxy # 40042) using the non-contact printer Nanolotter of GESIM (Groβerkmannsdorf, Germany) equipped with a piezoelectric NanoTip (GESIM, Groβerkmannsdorf, Germany).

Antibody controls, human IgG, human IgM, human IgA and human IgE (catalogue no. 7050-x) were obtained from Sigma Munich Germany, from Diatec Oslo Norway and printed from 100 mM PBS buffer pH 8.0 at a concentration of 500 ug/mL. The peptide and control spots were printed in three identical replications (blocks) on the slide (see Fig. 1). Printed peptide microarrays were kept at room temperature for 5 hours, washed with deionised water, quenched for 1 hour with 0.1 mg/mL BSA in 75 mM SSC buffer pH 7.0 containing 0.1% SDS and 750 mM NaCl at 42 °C, washed extensively with 1.5 mM SSC buffer pH 7.0, followed by washings with water, and dried using a chip centufuge. Resulting peptide microarrays were stored at 4 °C and maintained reactivity for more than 18 months.

2.3. Samples

Plasma from 5 patients with a clinical history of pulmonary TB were collected: all patients underwent partial pulmonectomy in order to remove scar tissue associated with TB. Material from the lesions of one patient tested positive for viable bacilli, but cultures were negative for the other four patients. Sputum samples from all five patients tested positive for the presence of acid-fast bacilli, and no patient showed serological evidence for HIV or Hepatitis B infection. Concentration of IgG, IgA and IgM in serum from these patients was measured by nephelometry. Samples were aliquoted and kept at −70 °C. For comparison, plasma from five individuals with no evidence of active TB were obtained from the Aeras Foundation, Washington, DC, USA. Based on sputum culture, X-ray analysis, as well as immunological testing (Quantiferon), these individuals had no evidence of active or latent TB. Ethical approval was obtained from the Stockholm South ethical committee (Dnr 238/02) for use of the specimens from Huddinge hospital, and from St Louis University, USA (number 12968), for use of the samples provided by the Aeras Foundation. Informed consent was obtained from the patients included in the study.

2.4. Sample processing

An adhesive Gene-Frame (Abgene, UK) incubation chamber was attached to the peptide microarray glass slide (JPT, Berlin, Germany). Plasma was diluted 1:100 in washing solution containing filtered PBS, 3% fetal calf serum (FCS, Lot nr 45K3397, Sigma, Munich Germany) and 0.5% Tween to a final volume of 300 uL.
which was pipetted at the end of the adhesive frame, and covered with a cover slip (VWR, International). This dilution buffer tested superior, defined by less background as compared to other diluents/blocking solutions (data not shown).

The plasma-solution was spread evenly across the entire microarray without air bubbles. The slide was incubated for 16 h at 4 °C in a humid chamber. After this first incubation step, the cover slip and gene-frame incubation chamber were removed and the slide was placed in a box with a slide holder (VWR, International, USA) and was washed 5 times for 5 min each (two times with washing solution, followed by two washes with sterile water and finally one wash with filtered Milli Q water). The washing procedure was performed on a shaker. A hydrophobic pen (DakoCytomation, Denmark) was used to define the incubation area. Without allowing the slide to dry, 300 uL of secondary antibody, Cy5-labeled goat anti-human IgG (Abcam, UK) diluted 1:500 in washing solution was pipetted onto the slide, which was covered with a cover slip and incubated for one hour at room temperature in a humid chamber, followed by the same five washing steps as above. The slide was spun dry for 10 seconds using a slide spinner (Euro Tech, UK). Incubation time, temperature and dilution for plasma as well as for the secondary antibody reagent was determined in preliminary experiments. Negative controls included scanning of the slide without any reagent (i.e. blank), and with only the reagent buffer plus the secondary antibody used for detection.

2.5. Data acquisition

Peptide microarray slides were scanned at a wavelength of 635 nm using a GenePix 4000B microarray scanner (Axon Instruments) at a PMT level of 100% and a resolution of 10 µm. Images were saved electronically in TIFF and JPG formats. Image analysis was performed utilizing the circular feature alignment of the GenePix Pro 5.1 software (Axon Instruments) and GenePix Array List (GAL) files supplied by JPT (see Fig. 1). GenePix Result (GPR) files were saved on CDs, and the median information in a slide. We analyzed the data from the range of measurement study to test for the effect of concentration on the chosen measure of response. We proceeded to inspect the negative control, positive control, and peptide responses separately in order to reveal any systematic errors that should be corrected. For each of these features we first used graphical analysis to identify outlying values that should be further investigated for manufacturing/technical errors. These were excluded from further analysis but flagged in the original data so that they could be noted when interpreting responses in the neighbouring locations on the slide. Finally we estimated the contribution of analyst, day and experiment to the variation of the results, enabling an assessment of assay precision. Each of these analytical steps is explained in detail in the following sections.

2.6. Statistical analysis

Using graphical tools we first examined the distribution of fluorescent intensities on various scales to find an appropriate transformation of the response to individual peptides that had a symmetric distribution, thus facilitating statistical analysis. We explored various two-dimensional representations of the data in order to find an efficient means of conveying the information in a slide. We analyzed the data from the range of measurement study to test for the effect of concentration on the chosen measure of response. We proceeded to inspect the negative control, positive control, and peptide responses separately in order to reveal any systematic errors that should be corrected. For each of these features we first used graphical analysis to identify outlying values that should be further investigated for manufacturing/technical errors. These were excluded from further analysis but flagged in the original data so that they could be noted when interpreting responses in the neighbouring locations on the slide. Finally we estimated the contribution of analyst, day and experiment to the variation of the results, enabling an assessment of assay precision. Each of these analytical steps is explained in detail in the following sections.

2.6.1. Range of measurements

Range of measurement was assessed using twenty slides prepared from four successive dilutions of each of five specimens with known IgG concentration. To assess whether peptide response is affected by dilution, we fit a linear model of the following form:

\[
Y_{ijk} = \mu + P_h + I_i + C_j + \varepsilon_{ijk}
\]

where \(Y_{ijk}\) is the transformed response of the \(h\)th peptide for the \(i\)th individual recorded at the \(j\)th concentration, and our objective is to test for a significant effect of concentration \(C\). In this model, \(P_h\) is the fixed effect of peptide \(h=1,...,378\); \(I_i\) is a random effect representing the individual, \(I_i \sim N(0, \sigma^2_I)\), \(i=1,..., 5\); \(C_j\) is the fixed effect of the \(j\)th concentration, \(j=1,2,3,4\), and \(\varepsilon_{ijk} \sim N(0, \sigma^2_{\varepsilon})\) is random error, where \(k\) denotes the replication due to block, \(k=1,2,3\). One could additionally check for intra-slide (block) effects.
2.6.2. Modelling of responses

The information in the replicated slides and the replicated blocks on each slide enables us to estimate the effects of operator (analyst), day and the order within day. The approach we pursue is to formulate a linear mixed model (McCulloch and Searle, 2001) that fits the patterns of variability in the data. One can include interaction between factors in the model to check if the levels of one factor behave differently at the various levels of other factors. After model fitting, residual analysis helps to reveal outliers that might indicate “overspill” from neighbouring spots in the case of negative controls, or spurious responses in the case of peptides. The model should be re-fit without these values if necessary. Using the linear mixed model we estimate and remove systematic errors to yield “normalized” data that provides a valid representation of the peptide response profile of an individual.

2.6.2.1. Negative controls. We fit the following linear mixed model to the transformed responses from the negative controls:

\[ Y_{ijklmn} = \mu + I_i + A_j + D_k + R_l + B_m + \epsilon_{ijklmn} \]  

where \( Y_{ijklmn} \) is the transformed response of the \( i \)th individual recorded by the \( j \)th analyst on the \( k \)th day in the \( l \)th experiment in the \( m \)th block of the slide; \( I_i \) is a random effect representing the individual, \( I_i \sim N(0, \sigma_I^2) \), \( i = 1, \ldots, 5 \); \( A_j \) is a random effect representing the effect of the \( j \)th analyst, \( A_j \sim N(0, \sigma_A^2) \), \( j = 1,2 \); \( D_k \) is a random day effect, \( D_k \sim N(0, \sigma_D^2) \), \( k = 1,2 \); \( R_l \) is a random effect representing the effect of the \( l \)th replicate of the experiment on the same day, \( R_l \sim N(0, \sigma_R^2) \), \( l = 1,2 \); \( B_m \) is a fixed effect representing block, \( m = 1,2,3 \), and \( \epsilon_{ijklmn} \sim N(0, \sigma_\epsilon^2) \), is random error, \( n = 1, \ldots, n_{NC} \), where \( n_{NC} \) denotes the number of negative controls in each block of a slide.

2.6.2.2. Positive controls. In contrast to the NC responses, which are assumed to be multiple observations of the same underlying process, each type of positive control represents a different process, and thus we fit the following model:

\[ Y_{hijklmn} = \mu + PC_h + I_i + A_j + D_k + R_l + B_m + \epsilon_{ijklmn} \]  

where the additional term \( PC_h \) is the fixed effect of the \( h \)th positive control (IgA, IgE, IgG and IgM), \( h = 1, \ldots, 4 \), and \( \epsilon_{ijklmn} \sim N(0, \sigma_\epsilon^2) \), is random error, \( n = 1, \ldots, n_{PC} \), with \( n_{PC} \) denoting the number of replicates of a positive control within each block, which may vary with the type of positive control.

2.6.2.3. Peptides. To study variation in peptide responses we use the following linear mixed model:

\[ Y_{hijklmn} = \mu + P_h + I_i + (PI)_h + A_j + D_k + R_l + B_m + (PB)_h + \epsilon_{ijklmn} \]  

where \( Y_{hijklmn} \) is the transformed response as before. The \( P_h \) term is a fixed effect for the \( h \)th peptide, \( h = 1, \ldots, n_P \) (the number of distinct peptides studied), and \( \epsilon_{ijklmn} \sim N(0, \sigma_\epsilon^2) \).

Fig. 2. Index versus log2 for background response of the raw data from a single TB+ patient.

Fig. 3. Average index across all peptides for five TB+ patients at successive dilutions 1:10, 1:100, 1:500, 1:1000. The five patient samples were first adjusted to obtain identical IgG concentrations (7.4g/l).

Fig. 4. Mean index of peptide responses across all dilutions for five TB+ patients.
(0, \sigma^2_\varepsilon), is random error, n = 1, \ldots, n_P, with n_P denoting the number of replicates of peptides within each block, which in our case is just one. Peptides are excluded from analysis if they exhibit a high response on the slide with only buffer and secondary antibody. The interaction effects (PB)_{hm} are used to identify extreme peptide responses that do not appear systematically in each block, as these are likely to be spurious responses that should be removed from analysis. The interactions (PI)_{hi} represent individual-specific responses for each peptide. The steps outlined above leave us with individual peptide responses for each individual adjusted for the variation, if any, due to slide, analyst, day and block. The fixed effects of peptides are estimated by least square means, and the SOLUTIONS option produced a file containing the best linear unbiased estimators (BLUPs).

2.8. Illustrative example

In order to illustrate how data extraction and analysis would proceed in practice, we compared the peptide responses of the five TB+ patients to the responses of five healthy controls. Since in practical applications, one will typically not have experiments replicated by analyst, day, and order within day, we extracted normalized responses for the five TB+ patients from a single experiment using the tools described in detail above. The normalized responses, together with the normalized responses from the healthy controls were used as input data for the freely-available SAM analysis.

2.6.3. Precision study

Using the information in the slides replicated by different analysts at different times, on different days, we express the magnitude of the variation due to each of these factors as a variance component obtained from the mixed models described above. The effect of each factor on the precision of the assay can be assessed by comparing its variance component to the variation due to random error (i.e. experimental noise).

2.7. Statistical software

All data preparation and graphical analysis was performed using R version 2.3.1. Statistical models were fitted using PROC MIXED in the SAS system for Windows (Version 9.1, SAS Institute Inc., Cary, NC, USA), where the LSMEANS subcommand yielded the
package of G. Chu and R. Tibshirani (http://www-stat.stanford.edu/sim~tibs/SAM) to determine whether there was evidence of significant effects of any peptides while controlling the false discovery rate. Although the software was developed for significance analysis of genomic microarray data, similar principles apply, and one can provide the identities of peptides as the "genes" in the input to the package.

3. Results

After inspection of various transformations of the data, we chose log2 (foreground/background) as an appropriate measure of response. This log-ratio, which will heretofore be referred to as the index, overcame the extreme positive skew in the raw data while still providing good discrimination between negative controls, positive controls and peptides. In Fig. 2 we plot the index versus the log2-transformed background values for the raw data from one of the slides. This plot offers a clear representation of the behaviour of the negative controls, positive controls and peptides. For any level of background, we can see that the index values for positive controls and peptides had similar range of values, a feature that is important for the validity of our statistical models. In addition, the index had a distribution characteristic of white noise for negative controls. As expected, the negative controls are tightly clustered around the zero-index line, the positive controls have index values above one, and the peptides have a wide range of responses (from −1 to 6). An index of at least 2 indicates that a foreground reading is at least twice as high as background on the log2 scale, i.e., four times higher. Note that because of the log-transform, the index is negative for any peptides with a foreground signal less than background. While the small group of peptides with high indices may be of particular interest, their clinical significance can only be assessed by comparative studies after quality data extraction.

3.1. Range of measurements

The effect of concentration on peptide response is illustrated in Fig. 3. We found no significant effect of dilution on peptide responses (p-value=0.21) when we analysed four patients (excluding patient 1). With
patient 1 included there is a significant effect of concentration.

3.2. Analysis of negative controls

Fig. 4 illustrates the behaviour of the negative controls from the analysis of the TB+ patients in a single experiment. The histograms reveal a few negative control spots (0.5% of all negative controls) with outlying response values.

To check for sources of systematic errors, we fitted model (2) to the negative control indices. Residual analysis revealed outliers, that were removed, their location flagged, and the model was re-fit. We found a significant difference in results between the two analysts (p-value=0.04) who performed the experiments, although the magnitude of this difference was very small (an average difference in the index of 0.003). There was no significant block effect (p-value=0.28) in the negative control responses, and little or no evidence of inconsistency in responses across days (p-value=0.57) or within day (p-value=0.17).

3.3. Analysis of positive controls

The distribution of the positive control responses measured for five TB+ individuals in a single experiment is displayed in the boxplot in Fig. 5, which also includes the blank slide and the slide containing only buffer and secondary antibody. As expected, the blank slide exhibits just low-level experimental noise. The ‘buffer’ slide which contains all dilution/detection reagents (including the fluorochrome labelled anti-human IgG antibody) but no plasma, showed a strong positive control reaction, indicating a ‘plasma-effect’, i.e. that the composition of individual plasma impacts on the recognition of the positive controls (i.e. immunoglobulins) by the secondary antibody. We can see that for all slides the positive controls are generally responding well, although there is evidence of differences between slides. A similar pattern was observed for each of the four types of positive controls (IgG, IgM, IgA, IgE) when studied separately, and for each of the three blocks (data not shown).

To assess between day and within day variation in positive control responses we fitted model (3). We found a significant positive control feature effect (p-value<0.0001), block effect (p-value=0.02, with the difference in index means between any two blocks ≤0.07), a significant analyst effect (p-value<0.0001; difference between index means=0.35), but no significant effect for replication within day (p-value=0.64) nor between day (p-value=0.20). The variation between analysts, days, and replications of the experiment within day is displayed graphically in Fig. 6a and b. In addition we found evidence that positive control responses are modified by patient (p-value<0.0001, see Fig. 6c).

3.4. Analysis of peptides

We first plot the raw values of indices for all peptides on a boxplot to study their behaviour (Fig. 7). On the slide containing only buffer and secondary antibody, we observed high responses for a small number of peptides. To allow for random error, we considered as valid all peptides whose responses on this slide were no further from the median than the minimum negative response. The remaining 15 peptides with more extreme responses on this control slide were excluded from analysis.

We fit model (3) in order to quantify and remove the different sources of variability. We found a strong

![Fig. 7. Boxplot of the average response (on the three blocks) for each peptide on the slides from TB+ patients in a single experiment. The plot also includes a blank slide and a slide containing only buffer and secondary antibody.](image)
overall peptide effect ($p$-value < 0.0001), no strong individual effect ($p$-value > 0.05), but a strong individual-specific peptide effect ($p$-value < 0.0001). There was no significant block effect ($p$-value = 0.23), and no inter-day ($p$-value = 0.24) nor intra-day ($p$-value = 0.66) effects. There was no significant difference in peptide indices between analysts (i.e., no analyst effect, $p$-value = 0.54).

3.5. Precision study

The mixed models described above provide us with variance components representing the variation due to different factors. The estimates of these variance components are presented in Table 1 together with the estimate of the variation due to random experimental noise. All variance components are of small magnitude compared to random variation indicating good experimental precision.

3.6. Illustrative example

For the five TB+ individuals and five healthy controls, the input to the SAM package consisted of the 363 peptide responses with all systematic effects removed as explained in Section 2.6.2.3. Thus the data consisted of the overall peptide effects and the individual-specific peptide effects, representing the response profile of the group and the departure of individuals from this profile for specific peptides. We set the threshold for the fold change value at 2.5, and selected the delta parameter so that false discovery rates (FDR) were at most 5%. We found 11 peptides with significantly higher responses in the TB+ group. One of these peptides had three neighbouring negative control locations with outlying responses, and we present the results for the other ten peptides in Table 2 and Fig. 8. In an analysis of the raw data without the systematic effects removed, we identified 13 peptides: 9 of those on Table 1 (all except LPDWLAANRGLAPGG), and four new peptides, one of which was located close to negative controls which were removed from the analysis.

4. Discussion

In this paper, the validation of first generation peptide microarray chips experiments is performed with the help of statistical methodology that can be used to assess quality, minimize random noise and remove systematic measurement errors. Using graphical tools, we identified a scale for measuring response that had a convenient range, a meaningful distribution for negative controls (i.e., experimental noise), a reasonably constant spread of values for all background levels, and good discrimination between negative controls, positive controls, and peptides. The response measure, $\log_2(\text{foreground/background})$, which we refer to as the index, measures the relative strength of the foreground signal compared to background, using a log scale to represent simple doubling effects. Thus a spot whose foreground is double the background intensity has an

<table>
<thead>
<tr>
<th>Protein accession number</th>
<th>Position</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag85A POA4V2</td>
<td>112–126</td>
<td>VMPVGGQSSFYSWDYW</td>
</tr>
<tr>
<td></td>
<td>154–168</td>
<td>ANRHVKPTGSAVGL</td>
</tr>
<tr>
<td></td>
<td>259–273</td>
<td>PSDLGGNNLPAPKLFLA</td>
</tr>
<tr>
<td></td>
<td>289–303</td>
<td>AGGHGNGYFDPDSG</td>
</tr>
<tr>
<td>Ag85B AAF13448</td>
<td>241–253</td>
<td>PNGTHSWEYWGQA</td>
</tr>
<tr>
<td>MPT51/MPB51 POA4V6</td>
<td>97–111</td>
<td>PAGGAYSMTYNEWQD</td>
</tr>
<tr>
<td></td>
<td>124–138</td>
<td>LPDWLAANRGLAPGG</td>
</tr>
<tr>
<td>TB10.4 POA568</td>
<td>25–39</td>
<td>LQSLGAELIAVEQAAL</td>
</tr>
<tr>
<td>ESAT-6 Q54OD8</td>
<td>53–66</td>
<td>QQKWDATATELNN</td>
</tr>
<tr>
<td>CFP10 POA566</td>
<td>22–34</td>
<td>LKTQIDQVESTAG</td>
</tr>
</tbody>
</table>

Amino acids in bold designate differences in the corresponding peptide stretch in Ag85A versus Ag85B and vice versa.
indexes of 1, while an index of 4 represents a foreground intensity 16 times higher than background. Our index differs from the “reactivity” measure of Quintana et al. (2004), which is the log-transformed difference between foreground and background for all spots where this difference is positive. Our method allows us to study peptides whose foreground is less than background, which may yield insight into natural variation in peptide responses. We present a graphical display of the index that captures the information in the data from a slide and presents it in a succinct form with unusual responses clearly visible. This display has three dimensions, the background, the index, and the feature (negative control, positive control, or peptide) and is a convenient tool for visualising a slide in a quantitative way.

To establish whether the range of measurement was satisfactory, we used a linear model to estimate the effects of plasma concentration on the peptide-specific response adjusted for differences between individuals. This analysis demonstrates that a 1:100 dilution of plasma is acceptable in TB+ individuals with active pulmonary TB defined by acid-fast positive sputum samples. A significant effect of plasma dilution (note that all specimens were standardized for the IgG content) was observed in material obtained from a single patient, who showed viable TB bacilli.

Our statistical tool, the Linear Mixed Model, enables us to model the contributions of various sources to the variability between assays done on the same patient. By modelling the effect of analyst, day, and replication within day as random effects, we not only treat our analysts, days and replicates as a sample from a larger set of possibilities, but we also obtain a variance component that allow us to assess the variation due to each of these factors relative to random experimental noise (i.e. unavoidable variation). This provides a clear quantification of the precision of the assay, which is a crucial pre-requisite for its use in addressing clinical hypotheses.

From our mixed models of negative control, positive control and peptide responses, we found that none of these three features had a significant inter-day or intra-day effect, which means that the assay is repeatable. We detected no statistically significant differences in peptide responses measured by the two analysts, although the positive controls did have a significant difference, and the negative controls also had a statistically significant difference but of very small magnitude. The lack of significant difference for peptide responses may be a result of reduced statistical power due to their variability. In contrast to the negative controls and positive controls, the response of a patient to individual peptides will be subject to natural (biological) variability in the patient specimen applied to each slide. It is reassuring to note that the magnitude of the variation between replications is small relative to experimental noise. We also found significant block effects for positive controls, but these were of small magnitude.

The key estimates from our analysis are the estimates of variation due to analyst, day and replication. By demonstrating that these are not statistically significant or of negligible magnitude, we validate the comparability of data gathered at different times on different days by different analysts, which is how data is typically obtained in real clinical studies or in a clinical ‘routine’ testing setting. Thus our focus in the analysis of effects of analyst, day and replication, is in establishing their ignorability. In contrast, a significant block effect is of no concern, since due to the replicate blocks this can be estimated and removed in the data pre-processing step.

In the analysis of peptides we used individual by peptide interaction terms to represent how the response profile of an individual differs from the average profile of their group. These terms are obtained from a linear mixed model (Model 4) that adjusts for any systematic effects of analyst, day, replication and block. Thus the peptide profiles of individuals are “normalized” to remove these effects and an individual’s overall response ($I_i$ in the model). This normalisation has been achieved without the need to assume balanced or few peptide responses, in contrast to many of the normalisation methods used for genomic microarrays (Yang et al., 2002, Irizarry et al., 2003). Interactions between peptide and block also enable us to detect peptide responses that are likely spurious, as a valid response would be expected to be similar in every block (i.e. no interaction). Not surprisingly we found evidence of significant differences between the TB+ individuals for their average response to the different peptides presented on the slide, and a significant individual-specific peptide effect that reflects a profile that is characteristic of the individual. Such profiles, extracted from TB+ and healthy individuals, were used as input data to a freely downloadable package for comparative analysis. This analysis identified 11 peptides that were significantly different in their responses in the two groups. For some of these, there was a shift in the response for all individuals in one group compared to the other, while others showed dramatic differences in just some individuals. We checked whether any of these were in locations where neighbouring negative controls had extreme (outlier) responses and found that one of the identified peptides had three neighbouring locations whose negative controls were outliers on slides from TB+ individuals. Since this
peptide’s response cannot be interpreted with any confidence, we presented the results for the remaining 10. We found that analysis of the data without removal of the systematic effects would have yielded a further four peptides, including one located close to negative control locations whose responses were suspect. This suggests that our pre-processing helps to reduce the false discovery rate (FDR) of peptides due to systematic artefacts in the data.

It is not surprising that we identified a very limited set of peptides in the small study conducted here. Larger patient cohorts and more comprehensive peptide antigen arrays are needed to visualize shared and ‘private’ B-cell responses in TB+ patients. The humoral epitope focus on the antigens Ag85A and -B is not surprising, as both antigens have been reported to stimulate B- and T-cell responses in humans (Launois et al., 1994, Mustafa et al., 2000). Of note, the peptide epitope VMPVGGQSSFYSDDWY which we identified in the Ag85A antigen serves as a strong T-helper epitope defined by interferon gamma production in CD4+ T-cell lines from patients with TB (Mustafa et al., 2000). The TB10.4 (aa 25-39) peptide also provides a T-helper epitope in patients with TB associated with distinct HLA-DR haplotypes (Skjot et al., 2002). It is interesting that only single epitopes from Ag85B, ESAT-6, TB10.4 or CFP10 were found to be recognized by the humoral immune system in TB+ patients, while four peptides from Ag85A were recognized. This may reflect the person-to-person heterogeneity of antigen recognition associated with M.tuberculosis infection (Lyashchenko et al., 1998) or stronger immunogenicity associated with Ag85A.

Most studies in TB-vaccine trials involve several blood draws at different time points: specimens will be frozen and tested for antigen-specific reactivity in a single experiment. Some study protocols include the harvesting of serum, while other protocols target the preservation of peripheral blood mononuclear cells and subsequent collection of plasma. We obtained similar data from plasma and serum (data not shown).

The inclusion of a blank slide and a slide with only buffer and secondary antibody is a critical part of the design of a validation experiment. Unexpected high responses from the blank slide can identify problems in manufacture, where peptides interact with the glass causing crystallization (Wenschuh et al., 2000). The slide with only buffer and secondary antibody can highlight where peptides are interacting with the secondary antibody in the absence of patient plasma or serum. Clearly these locations cannot provide valid data on the responses of the individuals under study. In our example, we identified such locations involving 15 distinct peptides, which were excluded from the comparative analysis of TB patients and controls. It is noteworthy that including these in the analysis using SAM, some can appear to be significant (data not shown), so that failure to recognize them would have compromised the validity of the analysis. Our analysis of negative control responses can identify locations on a slide which unusual behaviour, so that peptides in that neighbourhood can be discounted from any comparative analysis and/or from interpretation.

A strength of our study is the designed experiment, where the various factors (analyst, day, and experiment) are replicated. This enables us to quantify the magnitude of systematic biases, and to extract data that has been adjusted for such effects. If clinical experiments are to be conducted with only a single slide per patient, then it is important to first conduct a validation study to determine the precision of such measurements. Even where the validation study indicates that there is high reproducibility, failure to replicate the assay in a real application limits the information in the data extracted: for example, the individual effect and slide effect are completely confounded so that one can only extract an overall peptide effect for a group of individuals and an individual-specific peptide effect, as we have done in our illustrative example. However, it may also be of interest to estimate an individual effect, i.e. whether the entire spectrum of peptide responses are elevated for an individual, and in this case one would need replicated slides for the individual.

In conclusion, the approach presented here provides a paradigm for validation and extraction of data not only from first-generation slides but also from high-density chips currently under development where the responses of thousands peptides are measured. Such a standardised approach is crucial to enable valid comparisons across slides, batches of slides, and research studies.

Acknowledgements

We are grateful to Jan Andersson and Susanna Grundstrom, Department of Infectious Disease, Karolinska University Hospital, Stockholm, Sweden for providing patient samples, Frank Weichold, Aeras Foundation, Washington, DC, USA for providing samples from TB-negative individuals, and P. Andersen, SSI, Denmark for providing recombinant ESAT-6 antigen.

This work is supported by the Stockholm City Council (KI-Tuberculosis Center) to improve diagnostics in TB.
References


