



OPEN

Systematic reference sample generation for multiplexed serological assays

SUBJECT AREAS:
BIOCHEMICAL ASSAYS
DIAGNOSTIC MARKERS
ASSAY SYSTEMS
TUBERCULOSISHannes Planatscher¹, Stefanie Rimmele¹, Gerd Michel^{2*}, Oliver Pötz¹, Thomas Joos¹
& Nicole Schneiderhan-Marra¹¹NMI Natural and Medical Sciences Institute at the University of Tuebingen, Markwiesenstr. 55, 72770 Reutlingen, Germany, ²FIND, Foundation for Innovative New Diagnostics, Avenue de Budé 16, 1202 Geneva, Switzerland.Received
26 June 2013Accepted
1 November 2013Published
19 November 2013Correspondence and
requests for materials
should be addressed to
N.S.-M.
(schneiderhan@nmi.
de)* Current address:
Vela Research
Singapore Pte Ltd#02-
01 The Kendall 50
Science Park Road,
Singapore Science
Park II, Singapore
117406

Quality controls of serological assays have to contain defined amounts of human antibodies specific for the targeted antigen. A prevailing issue for array-based antigen assays is that dozens of antigens are targeted within the same assay. Commonly different patient sera are combined and optimal pools are empirically identified. Here, we report a mathematical approach how an optimal sample pool composition can be systematically calculated and accurately compiled. The approach was used to compose suitable quality controls for a 71 plex Tuberculosis antigen bead array using a limited number of positive human sera.

Miniaturized multiplexed serological assays have been applied to characterize the serological response against a variety of pathogens including HBV, HCV, *Helicobacter pylori*, *Mycobacterium tuberculosis* (Mtb) and influenza^{1–5}. In addition, (whole) proteome arrays enabled the definition of reactive antigens sets for a variety of pathogens like Mtb⁶, *Plasmodium falciparum* (3D7 strain)⁷, HPV⁸, *Burkholderia pseudomallei*⁹, and *Coxiella burnetii*^{10–12}. Arrays containing thousands of recombinant human proteins were used for the discovery of antibodies directed against self-antigens, which could be potential biomarkers for diagnostic purposes^{13–15}. In current clinical practice, serological assays are well established within the field of autoimmune diseases^{16–18}. All of these serological assays require quality controlled sample testing procedures.

Prior to implementation into diagnostics appropriate assay validation has to be achieved. FDA guidelines for the development of immunoassays state that “sufficient QC samples should be used to ensure control of the assay”¹⁹. As a consequence, such quality controlled samples should be available for assay validation, as well as for large scale screening and diagnostic purposes²⁰. Quality control samples are necessary within every assay to ensure that the assay performs within specifications and should be reviewed before interpretation of the results of individual serum samples. The purpose of a quality control is to report that all experimental steps were executed correctly in an assay experiment and to be able to compare data over a longer period of time.

Reference samples for sandwich immunoassays targeting serum proteins can be easily generated by spiking the target analytes into a plasma or serum matrix. However, any serological assay is based on the presence of human antibodies specific for the selected antigens. For singleplex assays it is usually sufficient to select a serum with strong reactivity towards its respective antigen. However, identifying single sera with appropriate reactivity against a multitude of different antigens, such as required for antigen arrays, has been very difficult if not impossible in many cases. Moreover, a single serum as a quality control, covering all targeted antigens would only be available in limited amounts and may thus confine test development, validation and clinical evaluation. A common, but surprisingly little documented approach is to create pools from multiple sera in order to warrant reactivity towards all target antigens and to generate a sufficiently large quality control stock^{21,22}.

Here we present a mathematical approach towards a sample pooling strategy where a composition of such a pool was calculated from an available data set with the aim that this pooled sample shows a positive response for each analyte. The threshold for a positive signal is defined by a multiple of the negative control population. In our case we chose four times the negative control population. If a signal in the pool exceeds this threshold the analyte is covered.

Results

The serological response of 142 serum samples obtained from patients with active tuberculosis (TB) was analyzed using a bead array consisting of 71 TB proteins (Supplement A). The serological response of these sera was heterogeneous, ranging from 2–69 TB-associated proteins per serum sample (Figure 1). Out of our 142 sera we found no serum reactive to all 71 antigens under investigation (Supplement B). Our mathematical approach

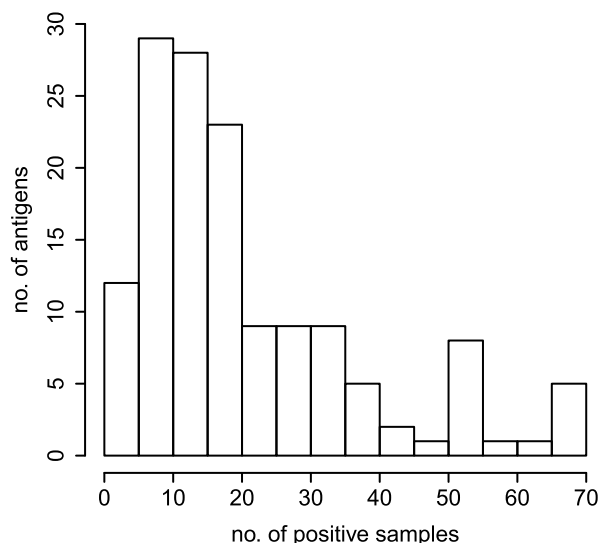


Figure 1 | Distribution of samples that give positive signals. The plot shows the number of TB-associated proteins reacting with a given number of serum samples. No single serum revealed a serological reactivity to all Mtb antigens. Only a small number of samples revealed a serological reactivity for a broad range of TB antigens.

identified sets of positive sera, which could be pooled to generate a quality control serum to react with all 71 TB proteins. This strategy allowed us to create defined reference samples, revealing a simultaneous serological response against all TB antigens employed in the assay.

Appropriate data sets of the serological response pattern against the targeted antigens for a set of available samples provided the basis for our calculation. A mathematical model was developed to predict reactivity characteristics of a given sample pool. We hypothesized that those values could be estimated from the quantitative serological response measured for the individual samples. Our first assumption was that if two samples are combined, assay signals would add up (see

Figure 2). The second assumption was that on average assays show dilution linearity with a slope of 1.0.

A linear integer program was constructed from the model, screening data and the threshold vector. The objective function was to maximize the number of serum reactivities, given a fixed number of serum samples from which the pool should be generated. The relative dilution of the sample pool was kept identical to the dilution of the individual serum. Results of the optimization approach revealed suggestions for the generation of optimal sample pools, differing in composition and size (Supplement C).

For the verification of our theoretical results, the following experiments were performed. In a first experiment the assumption of additivity of the individual signal values of the pooled samples was tested. A suspension bead array displaying the different tuberculosis antigens were incubated with human serum samples. Bound human IgG antibodies were detected with an R-PE-labeled anti-human IgG. The read-out was performed on a fluorescence-based bead array reader (Luminex FlexMAP3D). Sample pools were created by subsequently pooling samples in the scheme S_1 , $S_1 + S_2$, $S_1 + S_2 + S_3$, ... $S_1, S_1 + S_2$, $S_1 + S_2 + S_3$, ... up to a pool consisting of six samples. As shown in Figure 3A–E a strong correlation ($R > 0.98$) was observed between the values predicted from single sample screenings and the signal generated by the sample pool. The slope of the linear regression was 1.04 for the least complex pool, and decreased to 0.7 for the pool containing up to six samples. This data supports our hypotheses about signal additivity. The observation that the signals generated by the pools for a given antigen get stronger when the number of sera in the pool increased is notable (Figure 3). A larger number of different paratopes for the same antigen originating from different individual sera could also explain this observation.

In a subsequent experiment our algorithm was applied to find optimal pools for the total panel of 71 TB antigens. Here a data set derived from 142 previously tested serum samples was used as input. The allowed range for dilution of a single sample within the pool was 1 : 200 to 1 : 2000. Interestingly, we found that we had to consider that the signal intensities of each individual serum added to the pool are “diluted” with the other sera during the pooling process (Figure 3). The final serum dilution of the pool was set to 1 : 200, according to the standard dilution of our serological TB assay. An artificial cutoff

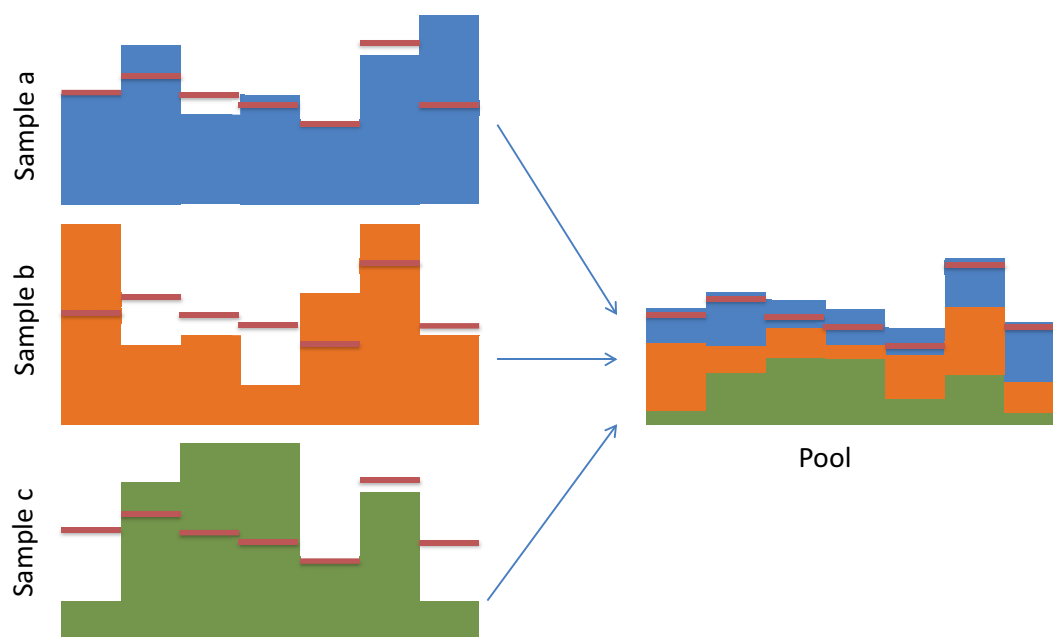


Figure 2 | Schematic representation of a sample pooling. The red lines represent an antigen-specific threshold. Samples (A–C) show individual responses and none exceed all thresholds. By combination a new pooled sample, with a positive response to all antigens, is created.

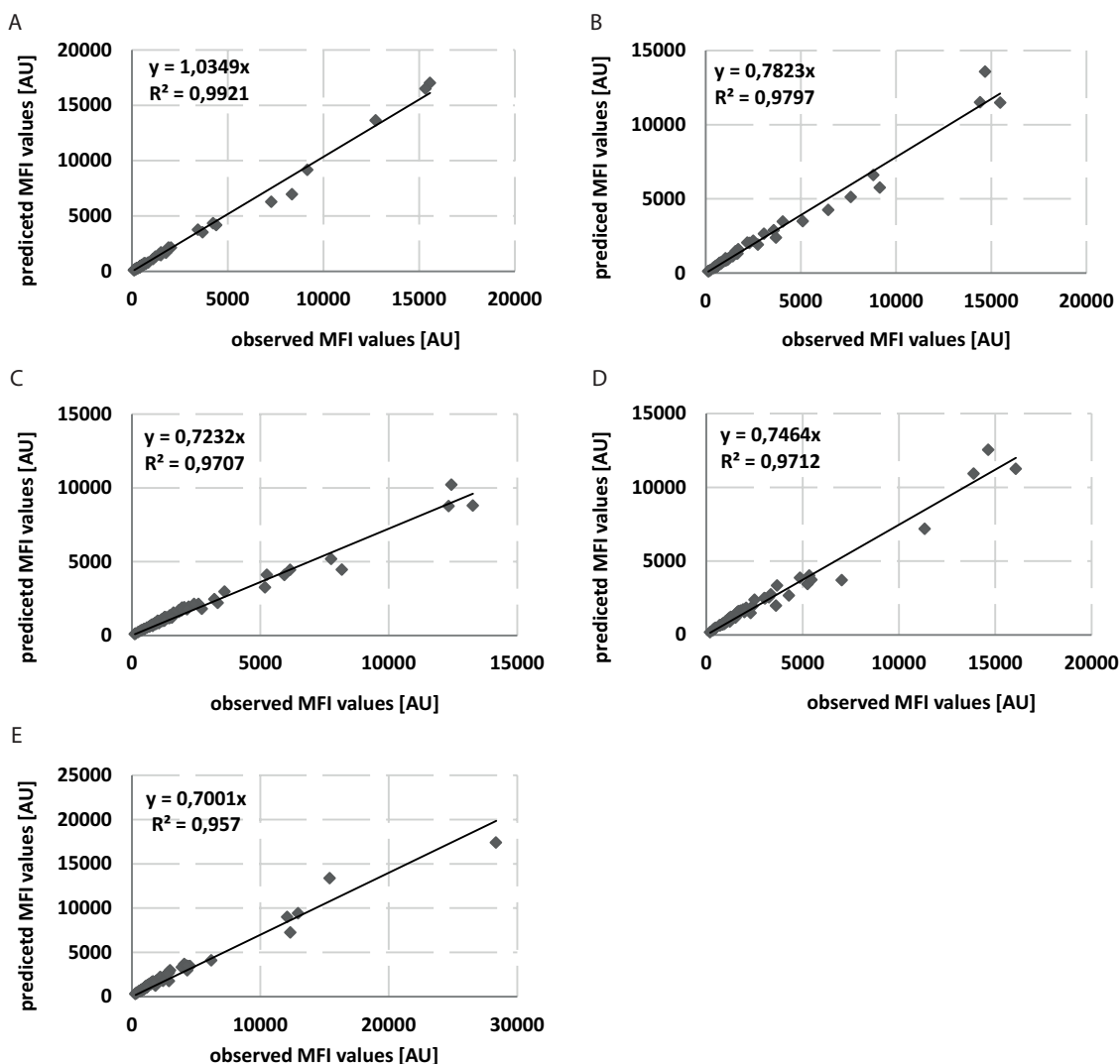


Figure 3 | Correlation between predicted and measured pool data. The graphs show the strong correlation between predicted and measured results, although the linear slope is 0.7 for the most complex pool (E). (A) Correlation of prediction and measurement for pool S1 + S2 (B) pool S1 + S2 + S3 (C) pool S1 + S2 + S3 + S4 (D) pool S1 + S2 + S3 + S4 + S5 (E) pool S1 + S2 + S3 + S4 + S5 + S6.

for each TB antigen was calculated from the quadruple of the values measured in the negative control sample. The algorithm suggested ten solutions consisting of up to ten parts of up to four different samples.

Thus, as expected, it was not possible to cover all analytes by a single sample or by pooling two individual samples. Our algorithm suggested pools consisting of at least three samples to reveal a serological response to all TB antigens. The signals of the pool for all analytes were higher than the defined threshold. The measured values correlate with the predicted values with a correlation coefficient of 0.98 (see Figure 4 A). The correlation between the pool and the three single samples is comparatively low as shown in Figures 4 B–D. This shows that no sample stands out in the pool and that the signal pattern is the result of the composition of all three samples.

Discussion

We have created a technical quality control for multiplexed antigen assays to make sure that all antigens used in the assay has not lost its antigenicity and that all technical steps are executed correctly.

With this mathematical model, we can create quality control samples for roughly 60000 samples from only 1.5 mL of three pooled serum samples. We also created a second pool consisting of four samples (10 parts; 5 parts sample 1, 3 parts sample 2, 1 part sample

3 and 4) with a correlation coefficient of 0.9 between predicted and observed MFI values (data not shown). Once the first pool is running out, one can easily create a second pool consisting of different samples. Our results demonstrate that our mathematical model for sample pools makes adequate predictions. We demonstrated that quality controls for multiplex antigen assays can be created by the systematic selection and pooling of samples. Our systematic approach is scalable and can be easily adapted to other assays platforms. We believe that our method provides an important tool for diagnostic assay development and test evaluation.

A software tool for the generation of reference samples is available at <http://webservices.nmi.de/samplepooler>.

Methods

Coupling of antigens to magnetic carboxylated beads. An automated bead handler (KingFisher 96, Thermo Scientific, Schwerte, Germany) was used to couple the antigens to magnetic carboxylated beads (MagPlex Microspheres, Luminex-Corp., Austin, TX). The Mtb proteins were covalently coupled to the beads using EDC/sulfo-NHS chemistry. The bead stock was vortexed and sonicated thoroughly for at least 10 s. Three hundred μL beads from each bead stock (1.25×10^7 beads/mL) were transferred to respective wells. Beads were washed with 250 μL activation buffer (100 mM Na_2HPO_4 + 0.005% (v/v) Triton X-100). The carboxyl groups on the magnetic beads were activated with 120 μL activation buffer + 15 μL EDC (50 mg/mL) + 15 μL sulfo-NHS (50 mg/mL in water-free DMSO) for 20 min at room temperature with agitation. Activated beads were washed two times with 250 μL

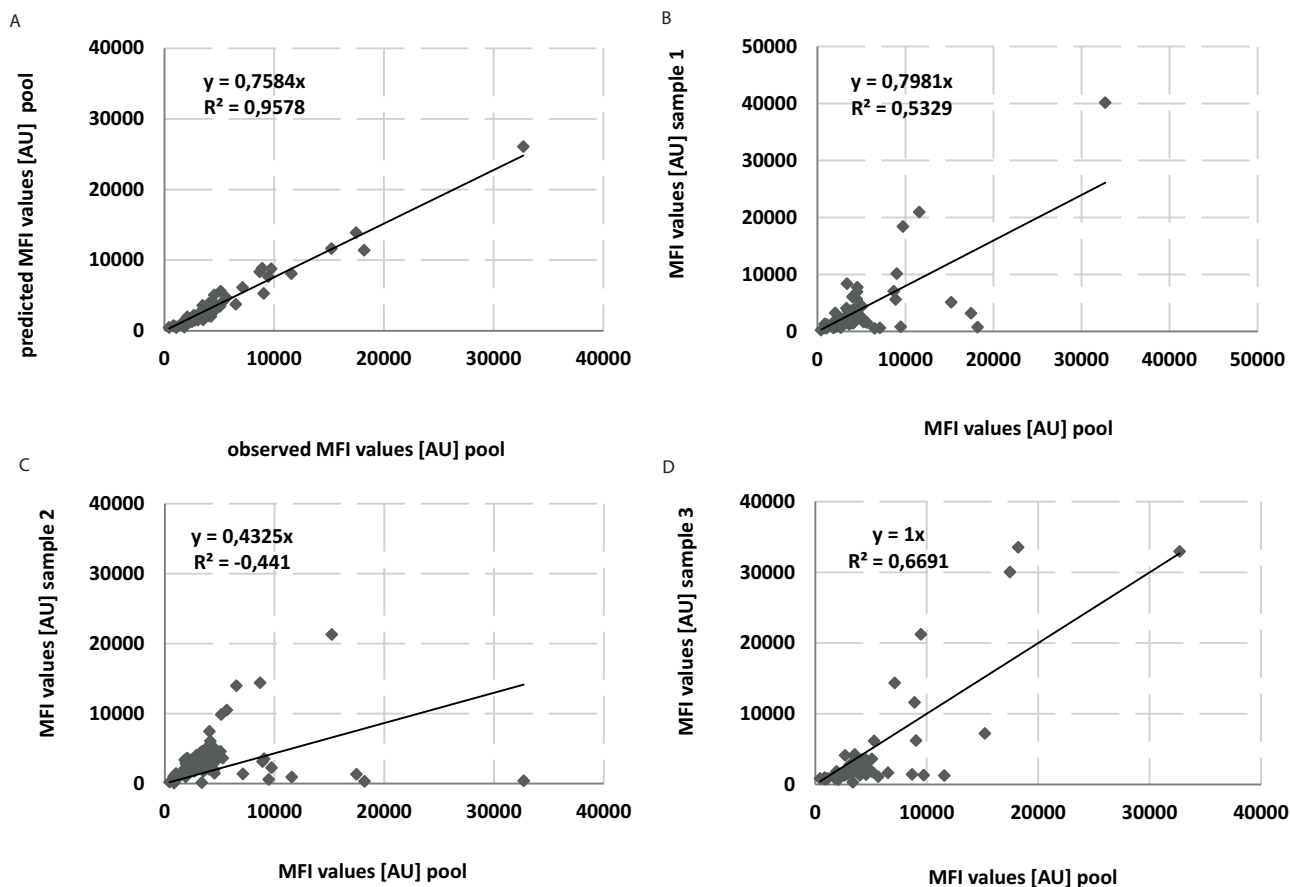


Figure 4 | Correlation of the calculated sample pool MFI values and the measured pool MFI values. While (A) the correlation between the sample pool and the predicted values (a weighted sum of single sample values) are high, the single samples (B–D) show weaker correlation with the pool. This data shows that the unique coverage characteristic of the serum pool is due to the combination of the three samples.

coupling buffer (50 mM MES + 0.005% (v/v) Triton X-100). Antigens diluted to a concentration of 100 µg/mL in coupling buffer were incubated with the activated beads and agitated for 2 h at RT. Coupled beads were washed with 250 µL wash buffer and resuspended in 200 µL block store buffer (PBS + 1% (w/v) BSA) containing 0.05% NaN₃ and stored at 4°C until further use.

Coupling control. The coupling efficiency of the His-tagged antigens was controlled using an anti-His antibody. Bound anti-His antibody was visualized using a secondary R-PE conjugated anti-species antibody.

The mouse-anti-His antibody (Qiagen, Hilden, Germany) was diluted in block store buffer (10 µg/mL, 1 µg/mL, 0.1 µg/mL, 0.01 µg/mL). For each antibody concentration 30 µL of the prepared bead suspension was distributed on a 96 half-well plate (1 µL of each individual bead type). A plate magnet was placed under the plate and the supernatant was removed by quickly inverting the plate. Beads were resuspended in 50 µL of diluted antibody and incubated on a shaker for 45 min at RT in the dark. Beads were washed twice with 100 µL wash buffer with the plate magnet as described before. 50 µL of a PE-conjugated goat-anti-mouse antibody (5 µg/mL in block store buffer) was added and incubated on a shaker for 30 min at RT in the dark. Beads were washed twice with 100 µL wash buffer and were resuspended in 100 µL block store buffer. Measurements were performed using a Luminex FlexMAP3D instrument with Luminex xPONENT software (settings: sample size: 80 µL, time out: 60, bead count: 100 per bead sort). Binding events were displayed as median fluorescence intensities.

After a successful coupling control the antigen-coupled beads were pooled to generate a master mix.

Bead-based serological assay. A serological bead-based assay was performed to detect human IgG antibodies directed against individual *Mycobacterium tuberculosis* proteins.

Serum samples were stored on ice. Serum dilution and incubation were performed at room temperature.

Samples were diluted in assay buffer (PBS + Low Cross Buffer (Candor, Wangen, Germany) + 0.5% BSA (w/v) supplemented with 10% *E. coli* lysate. After dilution, samples were incubated for 20 min on a shaker.

Incubation protocol. The assay was performed in a semi-automated fashion using a bead handling system (KingFisher 96, Thermo Scientific, Schwerte, Germany).

A master bead mix containing antigen-coated beads was prepared in assay buffer without *E. coli* lysate and distributed on a 96 well PCR plate. The beads were transferred from the bead source plate to 50 µL of the diluted human serum samples and were incubated for 2 h at room temperature. Unbound antibodies were removed by washing the beads twice with 100 µL PBS + 0.05% Tween20. To visualize bound human antibodies 50 µL of an R-PE labeled goat-anti-human IgG antibody (5 µg/mL) beads were incubated for 1 h at room temperature. After washing twice with 100 µL PBS + 0.05% Tween20 the beads were resuspended in 100 µL assay buffer without *E. coli* lysate. Measurements were performed using a Luminex FlexMAP3D instrument with Luminex xPONENT software (settings: sample size: 80 µL, time out: 60, bead count: 100 per bead sort). Binding events were displayed as median fluorescence intensities.

Algorithm. The screening of a set of samples S_1, S_2, \dots, S_n generated a dataset M and $m_j(S_i)$ designates the MFI-signal for target j in sample S_i .

An important premise is that the assays have a predominately linear characteristic in the range of interest. If a pooled sample P was created from the sample S_k and S_l the MFI would approximately add up

$$m_j(P) \approx m_j(S_k) + m_j(S_l). \quad (1)$$

If a sample is diluted using factor $\alpha \leq 1$ the MFI will exhibit a linear change:

$$\alpha m_j(P) \approx m_j(\alpha - \text{diluted } P). \quad (2)$$

We have created pools by subsequently adding samples ($S_1, S_1 + S_2, S_1 + S_2 + S_3, \dots$).

The threshold for a positive signal is usually defined by a multiple of the mean intensities measured in a negative control population. The threshold for target j is designated t_j . Furthermore the decision variable x_i describes whether the sample S_i is included in the pool or not.

In our approach the number of samples is fixed to a maximum, and the number of antigens quality controlled by the resulting pool is the target to be optimized. By allowing integer values for x_i , the constraint

$$\sum_{i=0}^n x_i = X_{\max} \quad (3)$$



fixes the number of parts, a pool consist of, to X_{max} . E.g. a pool with $X_{max} = 5$ could consist of 3 parts S_3 and 2 parts S_7 .

The set 71 of decision variables a_1, a_2, \dots, a_m , indicate whether an analyte should be covered by the pool or not. By defining the coverage as

$$\sum_{i=0}^n x_i m_j(S_i) \geq a_j t_j \quad 1 \leq j \leq m \quad (4)$$

it is ensured, that only if $a_j = 1$, the sum of MFIs has to exceed the threshold t_j . The term to maximize is the number of analytes which can be quality controlled using the pooled sample

$$\max \sum_{j=0}^m a_j \quad (5)$$

Another constraint is that the resulting pool should have the same matrix dilution as used in the normal sample preparation. If the input samples have been measured in a 1:n dilution resulting in values $m_j(S_i)$, the values need to be scaled accordingly to 1:n X_{max} . The upper bound for X_{max} is defined by the limit of dilutional linearity. E.g. if the limit is 1 : 2000 and the original dilution was 1 : 200 the maximum for X_{max} would be 10.

Pseudocode.

INPUT: $M, X_{maxupperbound}$

WHILE $X_{max} \leq X_{maxupperbound}$

BEGIN

$$\hat{m}_j(S_i) \leftarrow \frac{m_j(S_i)}{X_{max}}$$

solve ILP:

$$\max \sum_{j=0}^m a_j$$

subject to

$$\sum_{i=0}^n x_i \hat{m}_j(S_i) \geq a_j t_j \quad 1 \leq j \leq m$$

$$\sum_{i=0}^n x_i = X_{max}$$

$$a_j \in \{0, 1\} \quad \forall j \in \{1, \dots, m\}$$

$$x_i \in \mathbb{N}_0$$

$$X_{max} \leftarrow X_{max} + 1;$$

END

Generation of sample pool. For pooling the samples indicated by the algorithm 5 μL of each of the three samples were diluted 1 : 200 in assay buffer + 10% *E. coli* lysate, aliquotted at 60 μL and stored at -80°C .

- Opalka, D. *et al.* Simultaneous quantitation of antibodies to neutralizing epitopes on virus-like particles for human papillomavirus types 6, 11, 16, and 18 by a multiplexed luminex assay. *Clin Diagn Lab Immunol* **10**, 108–115 (2003).
- Waterboer, T. *et al.* Multiplex Human Papillomavirus Serology Based on In Situ-Purified Glutathione S-Transferase Fusion Proteins. *Clin Chem* (2005).
- Waterboer, T. *et al.* Dried blood spot samples for seroepidemiology of infections with human papillomaviruses, *Helicobacter pylori*, Hepatitis C Virus, and JC Virus. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **21**, 287–293 (2012).
- Tong, M. *et al.* A multiplexed and miniaturized serological tuberculosis assay identifies antigens that discriminate maximally between TB and non-TB sera. *J Immunol Methods* **301**, 154–163 (2005).
- Boni, M. F. *et al.* Population-Level Antibody Estimates to Novel Influenza A/H7N9. *J Infect Dis* **208**, 554–558 (2013).
- Kunnath-Velayudhan, S. *et al.* Dynamic antibody responses to the Mycobacterium tuberculosis proteome. *Proc Natl Acad Sci U S A* **107**, 14703–14708.
- Doolan, D. L. *et al.* Profiling humoral immune responses to *P. falciparum* infection with protein microarrays. *Proteomics* **8**, 4680–4694 (2008).
- Luevano, M. *et al.* High-throughput profiling of the humoral immune responses against thirteen human papillomavirus types by proteome microarrays. *Virology* **405**, 31–40 (2010).

- Felgner, P. L. *et al.* A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens. *Proc Natl Acad Sci U S A* **106**, 13499–13504 (2009).
- Beare, P. A. *et al.* Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a Coxiella burnetii protein microarray. *Clin Vaccine Immunol* **15**, 1771–1779 (2008).
- Vigil, A. *et al.* Genome-wide profiling of humoral immune response to Coxiella burnetii infection by protein microarray. *Proteomics* **10**, 2259–2269.
- Vigil, A. *et al.* Profiling the humoral immune response of acute and chronic Q fever by protein microarray. *Molecular & cellular proteomics: MCP* **10**, M110 006304 (2011).
- Gnjatic, S. *et al.* Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational protein arrays. *J Immunol Methods* **341**, 50–58 (2009).
- Hudson, M. E., Pozdnyakova, I., Haines, K., Mor, G. & Snyder, M. Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. *Proc Natl Acad Sci U S A* **104**, 17494–17499 (2007).
- Vizoso Pinto, M. G. *et al.* A systematic approach for the identification of novel, serologically reactive recombinant Varicella-Zoster Virus (VZV) antigens. *Virology journal* **7**, 165 (2010).
- Auger, I. *et al.* New autoantigens in rheumatoid arthritis (RA): screening 8268 protein arrays with sera from patients with RA. *Annals of the rheumatic diseases* **68**, 591–594 (2009).
- Robinson, W. H. *et al.* Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature medicine* **8**, 295–301 (2002).
- Kattah, M. G., Alemi, G. R., Thibault, D. L., Balboni, I. & Utz, P. J. A new two-color Fab labeling method for autoantigen protein microarrays. *Nat Methods* **3**, 745–751 (2006).
- FDA Guidance for Industry Bioanalytical Method Validation. *US Food and Drug Administration* (2001).
- Cummings, J., Ward, T. H., Greystoke, A., Ranson, M. & Dive, C. Biomarker method validation in anticancer drug development. *British journal of pharmacology* **153**, 646–656 (2008).
- Wong, S. J. *et al.* Detection of human anti-flavivirus antibodies with a west nile virus recombinant antigen microsphere immunoassay. *J Clin Microbiol* **42**, 65–72 (2004).
- Cooley, G. *et al.* High throughput selection of effective serodiagnosics for Trypanosoma cruzi infection. *PLoS neglected tropical diseases* **2**, e316 (2008).

Acknowledgments

Support for this Project was provided through funding from The Foundation for Innovative New Diagnostics. The views expressed by the authors do not necessarily reflect the views of the funding agency.

This project was a collaborative effort of the Natural and Medical Sciences Institute at the University of Tuebingen, Germany and The Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland.

Author contributions

H.P. and S.R. contributed equally to this work. H.P. conceived algorithm and mathematical models, implemented the software, and wrote the paper with input from other authors. S.R. carried out the experiments and analyzed the data. G.M. initiated the project while working at FIND, provided the antigens and selected the patient sera. T.J., O.P. and N.S.M. were involved in planning of experiments, in discussions and supervision of the project. The authors are grateful to Eloise Valli and Catharina Böhme, FIND, for collecting the sera for this study.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Planatscher, H. *et al.* Systematic reference sample generation for multiplexed serological assays. *Sci. Rep.* **3**, 3259; DOI:10.1038/srep03259 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>