Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method

Lina Cekaite[1], Ola Haug[2], Ola Myklebost[3], Magne Aldrin[2], Bjørn Østenstad[4], Marit Holden[2], Arnoldo Frigessi[5], Eivind Hovig[1], Mouldy Sioud[6]

[1]Department of Tumor Biology, [6]Department of Immunology, Molecular Medicine Group Institute for Cancer Research, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway

[2]Norwegian Computing Center, P.O. Box 114 Blindern, 0314 Oslo, Norway

[3]Department of Molecular Biosciences, [5]Section of Medical Statistics University of Oslo, Norway

[4]Department of Oncology, Ullevål University Hospital, Oslo, Norway.

Correspondence: Lina Cekaite, Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway

E-mail: linac@radium.uio.no

Fax: +47 22 52 24 21

Abbreviations: BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay; GAPS, gamma aminopropyl silane; PBS, phosphate-buffered saline; PMT, photomultiplier tube.

Key words: phage display • targeting peptides • protein microarray • regression

Abstract

We describe a novel approach for high-throughput analysis of the immune response in cancer patients using phage-based microarray technology. The recombinant phages used for fabricating phage arrays were initially selected via the use of random peptide phage libraries and breast cancer patient serum antibodies. The peptides displayed by the phages retained their ability to be recognized by serum antibodies after immobilisation. The recombinant phage microarrays were screened against either breast cancer or healthy donor serum antibodies. A model-based statistical method is proposed to estimate significant differences in serum antibody reactivity between patients and normals. A significant tumour effect was found with most of the selected phage-displayed peptides, suggesting that recombinant phage microarrays can serve as a tool in monitoring humoral responses towards phage-displayed peptides.

1 Introduction

Protein biomarkers in the serum hold great promise for early detection, classification and prognosis of many diseases such as cancer. The priming of the humoral arm of the immune system by tumour cells and/or other cancer-causing agents is expected to be printed in patient sera as a specific antibody response. Therefore, probing of such humoral immune responses could provide an effective strategy for cancer screening and therapy.

With the use of molecular approaches such as serological analysis of recombinant cDNA expression libraries of human tumours with autologous serum, it has been shown that the humoral response in cancer patients can be used to isolate novel tumour antigens [1]. We previously demonstrated that the display of random peptides or proteins on the surface of filamentous phages is well suited for the enrichment of serum antibody binding ligands [2, 3]. Notably, the antibody binding ligands do not necessarily resemble the natural antigens, but they can mimic their binding properties. Thus, it is possible to probe specific humoral immune responses in patients without preconceptions about the parental antigens. The application of such technology in cancer has led to the identification of a series of peptides and proteins that are recognized more frequently by patient serum antibodies when compared to antibodies from controls [4-6]. However, immune responses to single proteins or peptides is not expected in the majority of the cancer patients. It is therefore necessary to measure the antibody level for multiple antigens. The peptides presented on the phage surface could greatly enhance the possibilities for diagnosis, classification and prognosis of the disease [7, 8].

Historically, proteins or phage library screens for immunoreactivity have been carried out by the evaluation of individual proteins or recombinant phages by ELISA, Western blots, phages immobilized on a membrane by a plaque lift procedure, or by construction of macroarrays on filters [9]. These methods have been effective for a variety of applications, but they have several limitations. First, they require large sample volumes for both printing of antigens and for patient samples; second, these assays are time-consuming and laborious. Therefore, a complementary technology is necessary to permit the analysis of multiple serological markers. In the present study we analysed the reactivity pattern of patient and normal serum antibodies against immunoselected phage-displayed peptides in a high-throughput format. The findings demonstrate that peptides expressed by the phages retained their ability to be recognized by serum antibodies after immobilisation. The immunoreactivity clusters of patients and healthy donors were distinguished.

2 Materials and Methods

2.1 Sera

Sera from annotated patients with breast cancer of various stages were obtained from Ullevål University Hospital. Normal sera were obtained from blood donors (Ullevål University Hospital). All sera were aliquoted and stored at –70 ^{o}C until use.

2.2 Construction of a phage array

Immunoselected phages were used to assemble a phage array, and the selection protocol is described in ref [10]. The phage-displayed peptides were immediately frozen and stored at –70 ^{o}C until use. Phage clone stocks were thawed on ice, diluted in PBS and spotted on gamma aminopropyl silane slides (Ultra GAPS, Corning, Life Sciences, Corning, NY) at 10^{10-11} TU/ml concentration in a volume of 1 nl, using a robotic Micro Grid II robotic printer (Bio Robotics, Cambridge, UK) with a spot to spot distance of 350 μm. The humidity (45 - 55%) and temperature (15°C) was maintained during all of the print runs. Two identical arrays were printed on each slide, and the arrays were assembled from four sub-arrays, where each phage clone was replicated seven times. Bovine serum albumin (BSA) and empty phage (wild type) were also included in one of the sub-arrays as negative controls for the specificity of immunoreactivity. To control the experimental variation in hybridizations, 100 ng/ml of purified human Ig G (positive control) was included in seven replicates in all four sub-arrays. The printed slides were either used immediately or stored at +7°C for short time periods until used.

2.3 Labelling of secondary antibodies

Polyclonal rabbit anti-human Ig G antibodies, specific for gamma chains (DAKO, Denmark) and goat anti-mouse Ig G antibodies, specific for Fab fragments (Sigma-Aldrich, MI), were labelled with Cy-3 dye (Amersham Pharmacia Biotech, UK) as recommended by the manufacturer. Unconjugated dye was removed by PD10 (Amersham Pharmacia Biotech, UK) gel filtration columns (10-ml bed volume), which were pre-equilibrated in PBS prior to filtration. The efficiency of labelling was evaluated by measurement of the final protein/dye ratio.

2.4 Hybridisation and data acquisition

Hybridisations were performed with a hybridisation station (Perkin Elmer Life Sciences, Cambridge, UK). All incubation steps were performed at 37°C for 1 h with agitation. Slides were blocked in 3% fat-free milk powder in PBS/0.05% Tween-20, which was spun down for 10 min at 13 K before use. Thereafter, slides were incubated with serum samples diluted 1/1000 in PBS/0.05% Tween-20, followed by addition of Cy-3 labelled rabbit anti-human Ig G antibodies adjusted 1μg/ml (1/1000 dilution) in PBS/0.05% Tween-20. In experiments carried out to test the variation of concentrations of presented antigenic epitopes of the distinct phage clones on the spotted phage array, slides were incubated with mouse anti-serum raised against wild phage capsid proteins at 1/1000, then hybridised with Cy-3 labelled goat anti-mouse Ig G antibodies at 0.5 μg/ml (1/1000 dilution) in PBS/0.05% Tween-20, and passed through 0.22 μm filters (Millipore, Bedford, MA). All washes between different incubation steps were performed automatically at 37 °C, flow rate 10 s, hold 20 s for

five cycles in PBS/0.05% Tween-20, which had been passed through a 0.22 μm filter (Millipore, Bedford, MA) at 37°C. After drying by centrifugation, the slides were scanned using a ScanArray 4000XL Microarray scanner (Packard Biosciences, Biochip Technologies, LLC, Meriden, CT). The same laser power setting was used for all slides, and PMT values were calibrated according to the signal on Ig G spots within a narrow PMT interval. GenePix Pro 4.1 software (Axon Instruments, Inc., Union City, CA) was used to extract the scan data. The local background was subtracted from signal at each spot, and spots that had obvious defects were removed from further analysis.

A clustering of 14 replicates measurements for each phage clone hybridised both with breast cancer patients and healthy donors were constructed by Pro J-Express software (Molmine, Bergen, Norway).

Sequences alignment for the phage inserts was built using the Neighbour Joining method (AlignX, Vector NTI suite software, InforMax, Oxford, UK)

2.5 Detection of selected phage clones by ELISA

Serum antibodies reacting with antigenic epitopes displayed by phages were detected by ELISA. Briefly, 96-well microtiter plates were coated overnight with the phage particles (10^9 TU) in 100 µl PBS. After blocking with 0.5 % BSA in PBS for 2 h at room temperature and subsequent washing, the plates were incubated with serum samples (1/100 dilution in PBS / 0.2 % Tween-20) for 1 h at 37 °C. After additional washing, plates were incubated with an anti-human Ig G alkaline phosphatase

conjugate (1/5000) for 1 h at 37°C. The immune complexes were detected by adding p-nitrophenyl phosphate as a substrate with subsequent reading at 405 nm.

3 Results and Discussion

3.1 Array design and fabrication

The aim of the present study was to develop a method that allows efficient immobilisation and recognition of the phage-displayed peptides by patient serum antibodies in a high throughput format compatible with low sample volumes. Successful nucleic acid immobilization has been achieved using several different strategies for covalent attachment of modified synthetic nucleic acid oligomers to reactive surfaces. Most of these strategies use well-developed reactions between functional groups deliberately placed on the oligonucleotide terminus via synthetic methods and with complementary chemistry on surfaces. Surface functional groups can be coated, photoimmobilized, or chemically grafted to many support chemistries using well-established coating methods (e.g., silanization, functional polymer coating). Direct covalent approaches to protein arrays represent rational extensions to the same strategy, since several amino acids (e.g., lysine, glutamate, aspartate, histidine, tyrosine, cysteine, or any N- or C-terminus acids) provide a suitable functional capacity for covalent immobilization and modification chemistries [11-14].

Within the last few years, methods based on microarray technology have been adapted to the analysis of proteins, and novel applications have emerged [15-19]. Robotic microarray spotters that can group a large panel of phage-expressed peptides or proteins onto identical chips would facilitate the analysis of multiple antibody

specificities in individuals serum samples. We here report the fabrication of microarrays consisting of selected B cell peptide epitopes displayed by phages that are immobilised on surfaces coated with –gamma aminopropyl silane (GAPS) where immobilization resulted in a stable display of native, biologically active peptides. Phage microarrays were obtained by printing selected phage clone preparations containing distinct antigenic epitopes. It is of importance to maximize the signal-to-noise ratio in order to preserve potentially wide dynamic signal ranges. This may be achieved by attaching the fluorophore to a pure population of secondary antibodies. Also, it is necessary to label each serum sample individually in order to obtain serum specific signals. We therefore selected an immunospecific sandwich design, employing immobilized phage particles, followed by a blocking step (adsorption of skim milk proteins) to limit non-specific binding, and hybridization in a microarray format (Figure 1).

To fabricate phage microarrays, we employed a high-precision contact-printing robot to deliver nanoliter volumes of phage samples to the slides, yielding spots of about 200 to 250 µm in diameter. During printing, humidity was maintained between 45 and 55% in order to retain proteins hydrated throughout this and subsequent steps, to prevent denaturation, or folding of phage particles. As a result, each of the microspots in the array displayed a relatively constant amount of both phage capsid proteins, and the clone specific unique peptides/epitopes within the phage genome. The hybridisations with mouse anti-serum raised against wild phage capsid proteins clearly established the presence of immobilised phage particles on the phage

microarrays at expected locations. This demonstrates that proteins forming phage capsid are competent to react with GAPS-coated surfaces and immobilise whole phage particle on the glass slide.

Figure 2 demonstrates a characteristic representation of microarray hybridisations. The images show that the signals are clone dependent, with minor variation among repeated spots. The signal intensity is expected to be dependent on: First, the printed concentration of phage particles within a spot (an experimental variable), and second, the quantity of specific antibodies reacting to an individual antigenic epitope represented by phage (a biological variable). The data showed that although all phage clones had a similar transduction unit amount (estimated as TU/ml) prior to spotting, signals obtained from the distinct clones in some cases differed more than one log. The most likely explanation is that the plaque detection method cannot predict the exact phage concentration. The fact that the spot itself includes both phage original proteins and an inserted antigenic epitope, permits testing of the experimental variable by implementing an additional control hybridisation. Therefore, one slide from the same batch of printed slides was hybridised with mouse anti-serum raised against empty phage (wild type) capsid proteins (Figure 2A). The effect of the phage control array is presented in more detail in section 3.3.

To determine if the immunoreactivity obtained on phage microarrays was specific for the displayed peptides, we included a number of controls. BSA was chosen as a non-specific control protein (negative control-I) because non-BSA specific antibodies are

unlikely to hybridize to BSA spots, and that the presence of BSA specific antibodies is unlikely. No significant reactivity was observed with BSA, indicating that the immunorecactivity with the phage-displayed peptide was specific. However, two samples showed reactivity against BSA. This reactivity was considered non-specific, and therefore both samples were excluded from further analysis.

Besides recombinant peptide, all spots also include proteins of the phage capsid. Therefore in each array, empty phage (wild type) was printed in seven replicates in one of the sub-arrays (negative control-II). In most cases no reactivity was seen with patient sera in these control spots, and when low signals were detected, an increase of all signal intensities on the array was seen. Notably, each of the phage-displayed peptides displayed a different degree of reactivity with individual patient sera. For each sample the effect of non-specific binding to the empty phage was taken into account by the statistical model, see section 3.3.

To exclude the experimental variation in hybridizations, purified human Ig G antibodies that did not hybridise with sample antibodies, but reacted with fluorescently labelled anti-human secondary antibodies, were spotted on arrays. Strong signals were always obtained with the Ig G spots on slides hybridised with patients and healthy donors, see Figures 2 B and C, respectively. These spots were used for calibration of the PMT settings.

As shown in Figure 2 (and 3), the signal intensity throughout the spot varied slightly from spot to spot, depending on immunoreactivity on the epitope displayed on distinct phage clones. Typically, the epitopes with lower reactivity levels demonstrated less uniform spot morphology. Five of the printed phage clones (P6-P10) exhibited a bad spot morphology, possibly due to the precipitation and/or denaturation of phage particles in the spotting buffer. These phages were excluded from further analysis.

Typical displays of characteristic control spots and two recombinant phages hybridised with sera from a healthy donor and a patient and mouse anti-serum raised against empty phage capsid proteins are presented in Figure 3. The control spots demonstrated specific hybridisations; furthermore the phage-displayed peptides displayed a different degree of reactivity with sera of patients and healthy donors. Nevertheless, the signals here correlated with the signal intensity of the spots on the phage control array.

3.2 Expression of peptides on the surface of the immobilised phages are accessible to serum antibody binding

Having demonstrated that phage particles can be printed and detected on the arrays, we investigated the immunoreactivity of the displayed peptides. The screening of phage-displayed random peptide libraries on breast cancer patients identified a series of peptides, some of which bind selectively to serum antibodies from patients as compared to controls [10] (Sioud, M and Hansen M, unpublished results). Display of

random peptides and proteins on the surface of phages results in the ligands being amenable for affinity selection [20, 21]. With the use of the affinity selection protocol described in references [10, 22], the most immunoreactive phage clones were chosen for the phage array. In the adaptation of the phage display technology to a microarray format, we selected relatively short inserts (7-14 amino acids) that were not expected to have significant secondary structures. Aligned sequences for the inserts are given in Figure 4A. A phylogenetic tree of the calculated distance values is shown in parenthesis following the phage name displayed on the tree (Figure 4B). Three major clusters were distinguished by alignment of the sequences for consensus motifs. Analysis of amino acid residue occurrence patterns shows an overall depression of aspartic, glutamic acids, glycine and methionine residues and an overall abundance of proline, histidine, tryptophan and tyrosine residues. The enrichment of some amino acids in the selected sequences is more likely reflects the specificity of the selection, and the nature of the B cell response in the patients used for the selection of the phages as previously discussed [3, 23, 24].

3.3 Expression pattern of the repertoire of breast cancer antibodies to phage displayed peptides

The experimental data consisted of scanned phage peptide arrays representing 19 individuals (9 breast cancer patients (Table 1) and 10 healthy donors) additional to a phage control array. From the 36 phage clones originally spotted, P6-P10 were removed from further analyses due to bad spot morphology, leaving 31 distinct phage clones in each array. In arrays hybridised with human serum, additional data from 4

positive controls (I) and one negative control (II) were used in the analyses. The positive controls represented each of four sub-arrays, whereas the negative control represented an array in full. Two arrays were prepared for each individual and for the phage control, so that 14 repeated spot signals were available from either of the phage clones or the positive and negative controls (7 repetitions in each array).

Each single spot in the arrays had associated signal foreground and background values. Mean intensities minus background were used for analysis. In order to investigate the performance of recombinant phage microarrays to perform discrimination, a hierarchical clustering of the raw data is presented in Figure 5.

As a measure of similarity between clusters, complete linkage using the Canberra metric function [25] was chosen. In the matrix to be clustered there is one column for each of the 19 individuals, and there is one row for each of the 14 replicate measurements for each phage clone. A striking property of the clustering measurements in Figure 5 is the presence of large contiguous patches of colour representing groups of tested samples that share similar expression patterns over multiple hybridisations. The same batch of the printed slides were used for all hybridisations. Thus individual samples could be compared. The two major cluster groups are distinguished on the tested samples. A disease-related cluster, were mainly patients grouped, and a normal-like cluster, were healthy donors and two patients with low response grouped, were the most prominent. The accuracy of distinguishing between patients and normals depends on the specificity of phage presented peptides. In this microarray setup, we used phages biopanned against the pooled Ig G fraction derived from breast cancer patients with stage IV cancer [10]. The tested patient

samples on microarrays included patients of stages I-IV, prior and after treatment (Table 1). The stage and response to treatment may have an impact on the results. However, a larger number of tested samples for the selected antigenic epitopes is necessary for performance evaluation for disease stage and treatment.

The replicates of phages were similar. However, as we expected the variation in clone concentration to have clear impact on grouping of distinct recombinant phages, the following model was applied on the data:

Each single spot in the arrays has associated signal foreground, S^{fg}, and background, S^{bg}, values in the range $0 - (2^{16}-1)$. Further spot analyses described in this section were based on signal log-differences. Subtracting the background from the observed signal accounts for possible non-specific binding to the background. The logarithmic transformation stabilises the variance and is justified by the multiplicative nature of the various effects. To reduce the influence from spots with foreground intensity of the same order as background intensity, a lower limit equal to 10 is assumed for the signal difference, so that signal log-differences are actually defined as $\log\left(\max\left(S^{fg} - S^{bg}, 10\right)\right)$.

The statistical model was applied to reveal whether there is significant evidence that serum antibodies from breast cancer patients hybridise differently to selected phage clone epitopes as compared to antibodies from healthy control individuals (normals). As a consequence of the way the phage clones were selected, one would expect phage clone signals to be stronger for patients than for normals. This is visually confirmed from the plot shown in Figure 6. For each phage clone, signal log-difference values

from the 14 repetitions for each individual have been averaged and stacked in separate columns for normals and patients. For all phage clones, the centre of mass for responses of patients to phage-displayed peptides are above the ones for normals.

A visual comparison involving also the ELISA data set mentioned in section 2.5 indicates that our approach has a similar capacity as ELISA to detect expression levels and on the basis of these to distinguish different groups of sera. Figure 7 shows as a representative example a plot of ELISA data scattered against corresponding microarray values of epitope (P13) displayed on phage. However, since five of the control individuals were not recorded in the ELISA data set, observations should be interpreted with some care.

Furthermore, we expect signals from phage clone spots on arrays hybridised with human sera (normals and patients) to be positively correlated to signals from corresponding spots in the phage control array hybridised with mouse anti-serum raised against wild type phage capsid proteins. This is clearly indicated from the leftmost scatter plot in Figure 8, where phage clone signal log-differences are plotted against phage control signal log-differences from corresponding spots. There is unambiguous evidence of a positive relationship between the two variables.

From section 3.1 we know that strong negative control-II spots imply increased phage clone spot signals. This is in accordance with what is observed from the middle scatter plot of Figure 8. Here, phage clone signal log-differences seem to increase for increasing log-difference values of negative control-II spots within the same array.

From the properties of the material spotted in the positive controls (purified human Ig G), one would expect their signal intensities to be positively correlated to corresponding repetitions of phage clone spots from the relevant sub-array. However, the PMT tuning carried out in order to bring the average of the positive controls for 7 repetitions (1-7 or 8-14) to a similar level for all individuals, to some extent accounts for variations in efficiency of the hybridisations of secondary antibodies among the individuals (slides). Therefore, systematic coupling between the phage clones and the positive controls vanishes, as confirmed from the scatter plot given in the rightmost position of Figure 8. Here, control signal values are compressed, whereas the phage clone signals are still stretched out, cf. the x-range of the scatter plots. Some of the positive control spot signals are close to saturation, and this also introduces non-linearities compatible with the observations.

Possible distinctions in the hybridisation of antibodies from patient and normal human sera to selected phage clone epitopes can be revealed through the fitting of a statistical regression model. Such a model is able to separate systematic variations in observed phage clone signal, due to experimental settings, from biological variation, which is of actual interest. This is achieved with the help of explanatory variables such as controls, phage clone identification and patient/normal belongings. Define first the indices

i index counting individuals, $i = 1, 2, …, 19$

p index counting phage clone epitopes, $p = 1, 2, \ldots, 31$

r index counting repetitions, $r = 1, 2, \ldots, 14$

Indexed by these quantities, the following response and explanatory variables are derived:

y_{ipr} phage clone signal log-differences as measured from the r'th repetition of epitope p when the slide is hybridised with serum from individual i

C_{pr}^{phage} phage control signal log-differences as measured from the r'th repetition of epitope p when the slide is hybridised with anti-phage serum from mouse

C_{ipr}^{pos} positive control signal log-differences as measured from the r'th repetition of epitope p when the slide is hybridised with serum from individual i (for fixed i and r, C_{ipr}^{pos} attains the same value for all p's within each of the 8 possible sub-arrays)

C_{ir}^{neg} negative control signal log-differences as measured from the r'th repetition when the slide is hybridised with serum from individual i

I_{i}^{tumour} tumour indicator variable for individual i: I_{i}^{tumour} is 1 for tumour patients and 0 for normals

Next, we assume the following additive structure for the expectation of the signal:

$$\mathrm{E}\left(y_{ipr}\right) = \gamma_{p} + \beta^{phage} C_{pr}^{phage} + \beta^{pos} C_{ipr}^{pos} + \beta^{neg} C_{ir}^{neg} + \delta_{p} I_{i}^{tumour}$$

The constant term γ_p is interpretable as a base level contrast reflecting the response from human normal serum antibodies hybridising to epitope p. The middle three terms on the right hand side (those involving controls) are adjustments to this base level. Finally, $\delta_p I_i^{tumour}$ accounts for tumour-specific deviations in the response. Taking exponentials turns the model into factorial form, which emphasises relative effects of the individual terms. Of main interest is the relative additional effect, ($\exp(\delta_p)$), caused by hybridisation of antibody from tumour patients to epitope p.

Clearly, our responses (epitope signal log-differences) cannot be considered independent. In particular, signals from the different repetitions of a specific phage clone will be linked for each individual. This fact makes the theory of ordinary regression and Analysis of Variance inadequate for our problem. Instead, parameters were estimated by the method of Generalized Estimating Equations (GEE) [26, 27] using the S-PLUS add-on software package Oswald [28]. GEE is an extension of ordinary regression, developed specifically to accommodate data that are correlated within clusters (here: individuals). See the appendix for further details.

Fitting our model to the data produced estimates of the coefficients in the model along with their standard deviations. Values for the control variables are given in Table 2.

For the phage control and the negative control-II, the estimated coefficients are clearly positive, in accordance with the plots in Figure 8. Moreover, the phage control

is highly significant, which stresses the importance of taking this term into the model. The positive controls, however, turn out to be non-significant. This fully supports the scatter plot in Figure 8, and indicates that positive controls could be left out from our model with no loss of explanatory properties (as long as PMT tuning is employed). But, since the PMT correction is not quite precise (only based on the first 7 repetitions and averages are brought to a 'similar level'), we decided to keep the positive control as an explanatory variable.

Our main interest is in the recombinant phage clone/tumour cross-terms. All cross-terms except two were estimated to be significant (P4 and P33 were slightly non-significant at the 95% level), and also, the corresponding coefficients were all positive. That is, after correcting for effects implied by the phage controls and the negative (II) and positive (I) controls, antibodies present in tumour serum still have an additional effect on the selected phage clone epitopes beyond the general effect induced by sera from both normals and tumour patients. This is no surprise since all phage clones were selected based on their expected response to serum antibodies from the breast cancer patients. The results are illustrated in Figure 9, and details are given in Table 3 in the appendix. Within the plots, the cross-term of recombinant phage clones that are significantly positive are identified from their confidence intervals spanning positive values only.

4 Concluding remarks

In the present study, we have demonstrated that serum antibodies from breast cancer patients specifically recognize phage-displayed peptides spotted on microarrays. A large panel of recombinant phages showed significant differential reactivity to patients as compared to normal serum antibodies.

We conclude that the proposed approach represents a valid strategy to high-throughput profiling serum antibody specificities, using one-channel signal data. Also, the approach described here for fabricating phage arrays is highly practical; the binding assays on these arrays demonstrated high specificity and no apparent artefacts due to phage particle immobilization. In contrast to other molecular techniques, the use of peptides or proteins expressed on the surface of the phages required no purification steps of the expressed ligands. Notably, the effect of using phage particles as a carrier for contact printing and surface support may, however, have a significant effect on the quality of the results as compared to the direct immobilisation of short peptides on derived glass chips. Thus, future experiments addressing the comparison of free peptides and phage-displayed peptides printing would be valuable in assessing the relative performance of the alternate strategies.

In our opinion, the technology presented here has a potential for the measurement of serum antibody levels against antigenic peptides or proteins in various types of cancer, as well as for the identification and monitoring of biomarkers for disease progression. Moreover, the analysis of multiple antibodies simultaneously may significantly improve the profiling of tissue specific disease signatures, and a probing of humoral immune response

.

5 References

[1]   U. Sahin, O. Tureci, H. Schmitt, B. Cochlovius, T. Johannes, R. Schmits, F. Stenner, G. Luo, I. Schobert, M. Pfreundschuh, *PNAS* 1995, *92*, 11810.

[2]   A. Dybwad, O. Forre, J. Kjeldsen-Kragh, J. B. Natvig, M. Sioud, *Eur J Immunol* 1993, *23*, 3189.

[3]   M. Sioud, O. Forre, A. Dybwad, *Clinical Immunology and Immunopathology* 1996, *79*, 105.

[4]   M. Sioud, M. Hansen, A. Dybwad, *Int.J Mol Med.* 2000, *6*, 123.

[5]   M. Sioud, M. H. Hansen, *Eur J Immunol* 2001, *31*, 716.

[6]   J. C. Miller, H. Zhou, J. Kwekel, R. Cavallo, J. Burke, E. B. Butler, B. S. Teh, B. B. Haab, *Proteomics.* 2003, *3*, 56.

[7]   R. Ueda, H. Shiku, M. Pfreundschuh, T. Takahashi, L. T. Li, W. F. Whitmore, H. F. Oettgen, L. J. Old, *J Exp Med* 1979, *150*, 564.

[8]   A. Folgori, R. Tafi, A. Meola, F. Felici, G. Galfre, R. Cortese, P. Monaci, A. Nicosia, *EMBO J.* 1994, *13*, 2236.

[9]   H. R. Hoogenboom, A. P. de Bruine, S. E. Hufton, R. M. Hoet, J.-W. Arends, R. C. Roovers, *Immunotechnology* 1998, *4*, 1.

[10]   M. H. Hansen, B. Ostenstad, M. Sioud, *Mol Med* 2001, *7*, 230.

[11]   A. S. Hoffman, *J Biomater Sci Polym Ed* 1999, *10*, 1011.

[12]   G. MacBeath, S. L. Schreiber, *Science* 2000, *289*, 1760.

[13]   J. R. Falsey, M. Renil, S. Park, S. Li, K. S. Lam, *Bioconjug Chem* 2001, *12*, 346.

[14]   B. T. Houseman, M. Mrksich, *Trends Biotechnol* 2002, *20*, 279.

[15]   P. Angenendt, J. Glokler, D. Murphy, H. Lehrach, D. J. Cahill, *Analytical Biochemistry* 2002, *309*, 253.

[16]   W. Kusnezow, J. D. Hoheisel, *J Mol Recognit* 2003, *16*, 165.

[17]   W. Kusnezow, J. Hoheisel, *Biotechniques* 2002, *Suppl*, 14.

[18]   Y. Kawahashi, N. Doi, H. Takashima, C. Tsuda, Y. Oishi, R. Oyama, M. Yonezawa, E. Miyamoto-Sato, H. Yanagawa, *Proteomics* 2003, *3*, 1236.

[19]   H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, M. Snyder, *Nat Genet* 2000, *26*, 283.

[20]   A. D. Griffiths, M. Malmqvist, J. D. Marks, J. M. Bye, M. J. Embleton, J. McCafferty, M. Baier, K. P. Holliger, B. D. Gorick, N. C. Hughes-Jones, et al., *Embo J* 1993, *12*, 725.

[21]   G. P. Smith, *Science* 1985, *228*, 1315.

[22]   M. H. Hansen, A. Dybwad, O. Forre, M. Sioud, *Clin Exp Rheumatol* 2000, *18*, 465.

[23]   M. Sioud, A. Dybwad, L. Jespersen, S. Suleyman, J. B. Natvig, O. Forre, *Clin Exp Immunol* 1994, *98*, 520.

[24]   D. J. Rodi, A. S. Soares, L. Makowski, *J Mol Biol* 2002, *322*, 1039.

[25]   J. Yang, R. Parekh, V. Honavar, *Intelligent Data Analysis* 1999, *3*, 55.

[26]   S. L. Zeger, K. Y. Liang, *Biometrics* 1986, *42*, 121.

[27]   P. Diggle, Liang, K. and Zeger, S., *The Analysis of Longitudinal Data*, 1994.

[28]   D. Smith, Robertson, W. and Diggle P., *Object-Oriented Software for the Analysis of Longitudinal Data in S*, Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, UK, 1996.

Table 1. The patient information

PATIENT	STAGE OF BREAST CANCER	TREATMENT
Patient1	IV	Chemotherapy
Patient2	IV	No therapy
Patient3	IV	No therapy
Patient4	II	Chemotherapy
Patient5	IV	Chemotherapy
Patient6	III	Chemotherapy
Patient7	I	Radiation therapy
Patient8	II	Chemotherapy
Patient9	II	Chemotherapy &Radiation therapy

Table 2. Control explanatory variables: Estimated coefficients and their standard deviations.

Coefficient	Estimate	St.Dev.
β^{phage}	0.14	0.02
β^{pos}	0.06	0.18
β^{neg}	0.21	0.09
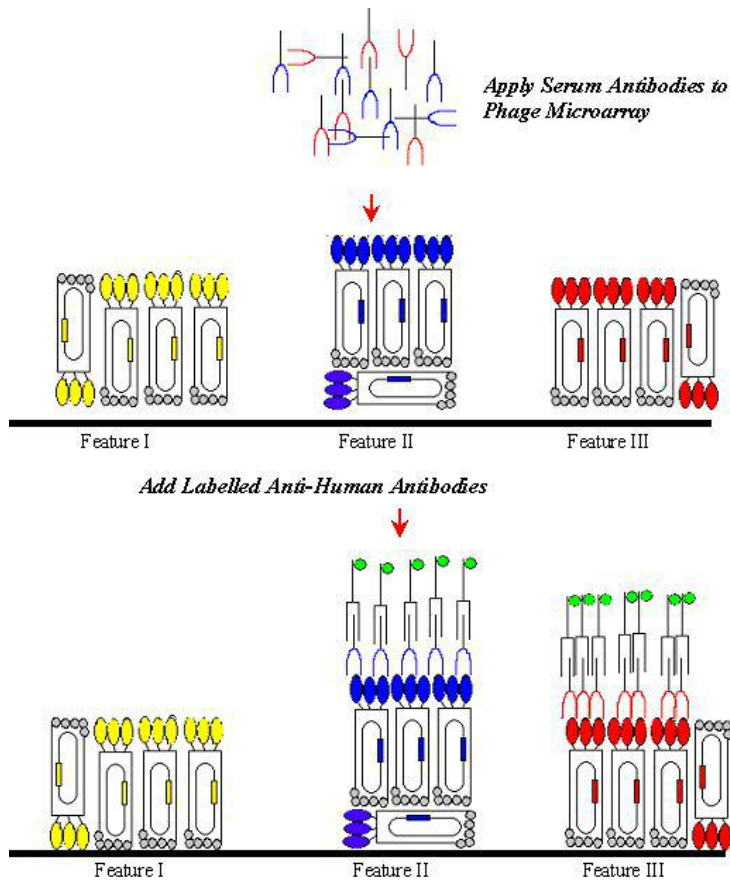
Figure 1

Probing immune responses with antigen arrays. Patient serum, containing multiple antibodies, is exposed to an antigenic epitope presented by phage (red, blue and yellow) so that antibodies specific for the different antigens are captured at specific locations on the array. After washing away unbound antibodies, labelled (green) anti-human antibodies are added, and bound to the human antibodies on the array, thus indirectly revealing the antibody response to the various antigens in the patient serum.

A. A control array   B. Breast cancer array   C. Healthy donor (normal)
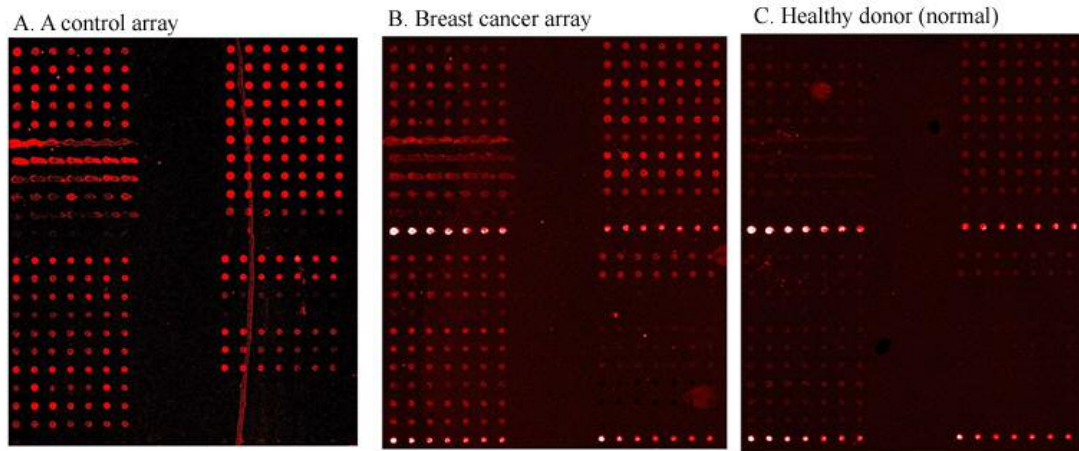
Figure 2

Representation of characteristic microarray images with the spatial arrangement of 36 distinct phage clones, negative and positive controls spotted in seven replicates on each array. The phage arrays were incubated with mouse serum raised against wild/empty phage (A), or with sera from breast cancer patients (B) or healthy donors (C). Then each array was hybridised with secondary Cy-3 labelled antibodies. Phage clones P6-P10 exhibited bad spot morphology, and were excluded from further analysis.

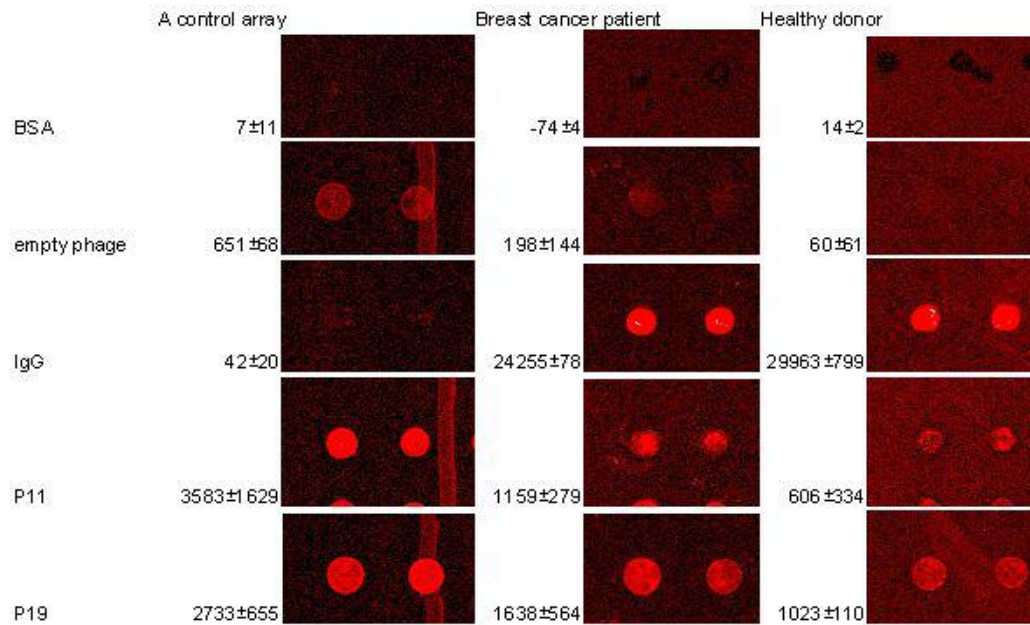|	A control array	Breast cancer patient	Healthy donor
BSA	7±11	-74 ±4	14±2
empty phage	651±68	198±144	60±61
IgG	42±20	24255±78	29963 ±799
P11	3583±1629	1159±279	606 ±334
P19	2733±655	1638±564	1023±110

Figure 3

Representation of typical negative (BSA and empty phage) and positive (Ig G) controls and spots of two randomly selected phage clones (P11 and P19) on arrays hybridised with mouse serum raised against wild (empty) phage and human sera from a breast cancer patient and a healthy donor, respectively. Numbers signify mean values of fluorescence intensity of the two spots; followed by background subtraction. Standard deviations are given.

A

```
          (1) 1          10            23
 P1   (1) --------LTPYHPP--------
 P6   (1) --------APPSPPP--------
 P32  (1) -----SERRIPSNPDTRFAD---
 P3   (1) ---------WPLKPKL-------
 P5   (1) ---------FPMKPVL-------
 P12  (1) ---------IPHKPWW-------
 P34  (1) --SQRIPARIHHFPTSI------
 P13  (1) -------NPIEDQV---------
 P17  (1) -------ALINWQV---------
 P7   (1) ------HPILWPP---------
 P15  (1) -------FMAKDRP---------
 P25  (1) -------GMLKDRP---------
 P22  (1) --------FQKDRPG--------
 P33  (1) NHEPSVTQVILDRPY--------
 P10  (1) --------NVNTPQTVKHPG---
 P36  (1) --------TVETPQTAKTAYYAT
 P27  (1) --------YVNVNRW--------
 P2   (1) -------GSVTQRL---------
 P9   (1) -------GAVAQRL---------
 P21  (1) --------AVPYQGY--------
 P18  (1) ------RPHVPSH---------
 P26  (1) --------HVPSFSP--------
 P24  (1) --------HVPAFVN--------
 P29  (1) --------HVPAFRT--------
 P11  (1) --------SKPFATQ--------
 P23  (1) ---------LPRMLNV-------
 P19  (1) --------SSIQSLP--------
 P31  (1) ------DSSLTSLRNSTR-----
 P30  (1) ---WAPPLFRSSLFY--------
 P4   (1) ---------YSSSLYY-------
 P35  (1) ---SDEQSYSTTRSYYGQ-----
 P8   (1) ---------YSTTLSY-------
 P14  (1) --------HVGIPPS--------
 P28  (1) --------EGIKPTN-------
 P16  (1) --------GLHPGPA--------
 P20  (1) -------SQLSTKP--------
```

B

Phylogenetic tree with the following leaves:
P1 (0.3057), P6 (0.2657), P32 (0.3157), P3 (0.2143), P5 (0.2143), P12 (0.2319), P34 (0.3396), P13 (0.2894), P17 (0.2820), P7 (0.3137), P15 (0.1553), P25 (0.1304), P22 (0.1742), P33 (0.2694), P10 (0.2461), P36 (0.2539), P27 (0.4056), P2 (0.1365), P9 (0.1492), P21 (0.3464), P18 (0.0840), P26 (0.1160), P24 (0.1488), P29 (0.1369), P11 (0.4300), P23 (0.4033), P19 (0.1385), P31 (0.1472), P30 (0.2217), P4 (0.2069), P35 (0.0676), P8 (0.0752), P14 (0.2403), P28 (0.2597), P16 (0.3567), P20 (0.3100)
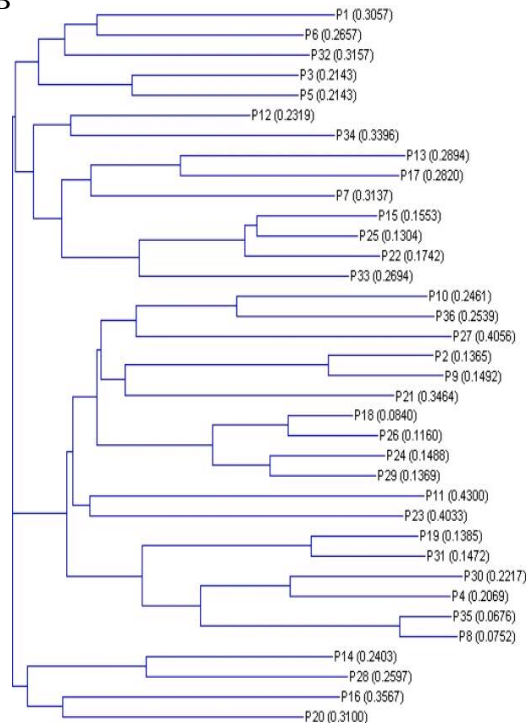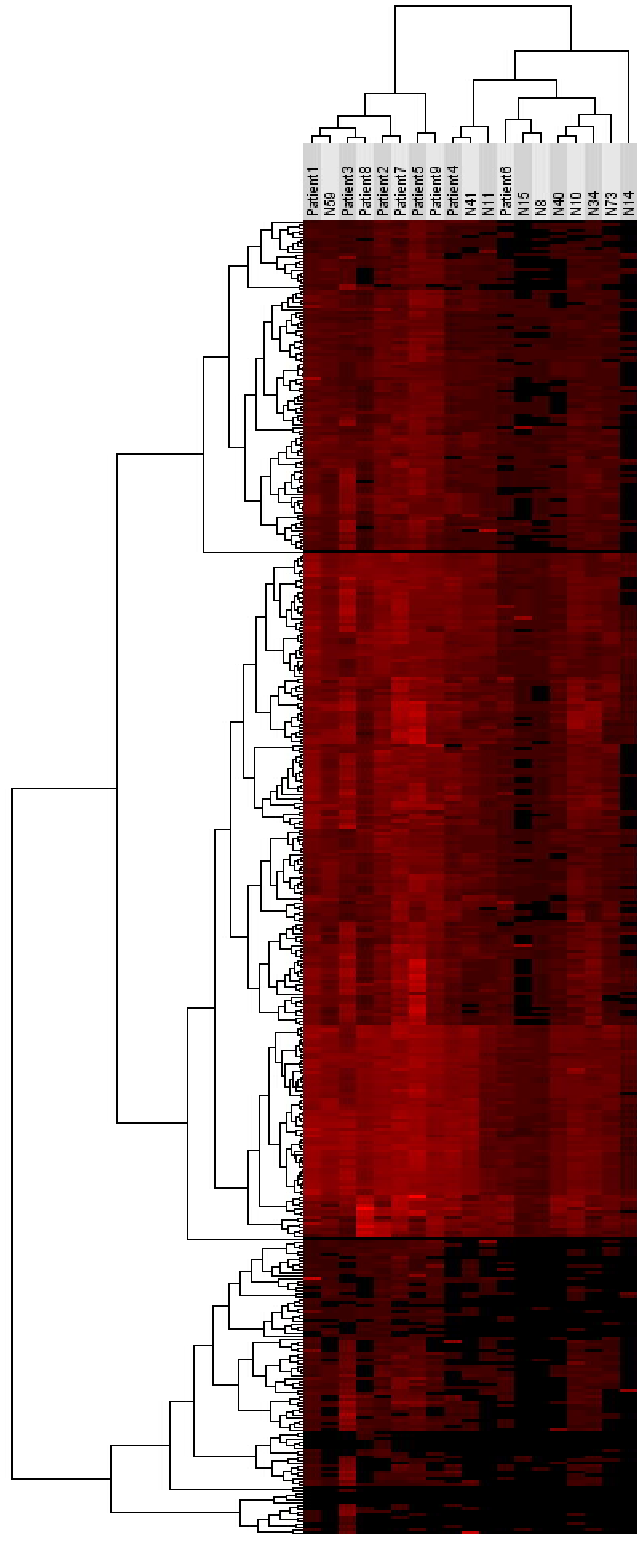
Figure 4

The presentations of aligned sequences for the inserts of recombinant phages are given (A). A phylogenetic tree (Neighbour Joining method) displays three major clusters of the recombinant phages (B). The calculated distance values are shown in parenthesis following the phage name displayed on the tree.

Cluster method: Complete Linkage
Distance metric: Canberra

-20,167.000   0.0   20,167.000

Figure 5

In the matrix to be clustered there is one column for each of the 19 individuals, and there is one row for each of the 14 replicate measurements for each phage clone. As a measure of similarity between clusters, complete linkage using the Canberra metric function was chosen. A striking property of the clustered measurements is the presence of large contiguous patches of colour representing groups of tested samples that share similar expression patterns over multiple hybridisations. The two major groups of samples are depicted above the cluster graph.
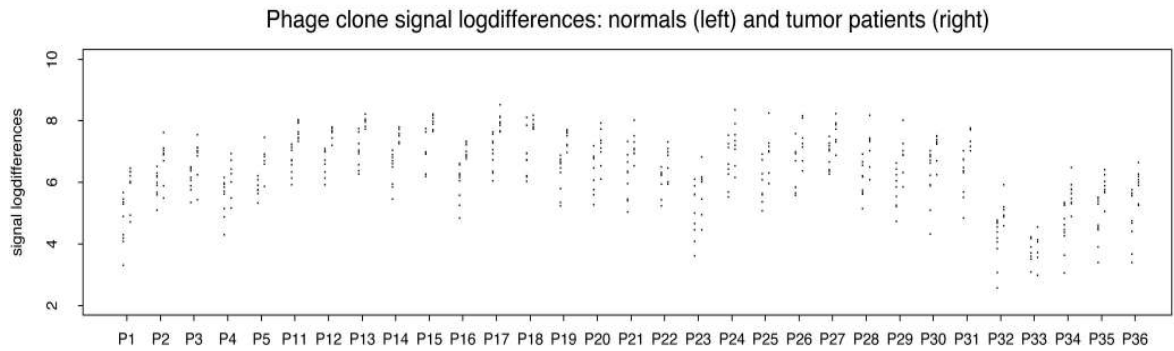
Figure 6

Signal log-difference values for each phage clone split into separate columns for normals (left) and breast cancer patients (right). Each single dot represents the average of signal log-difference values available for each individual from repeated measurements of the phage clone in question (maximum 14 repetitions).
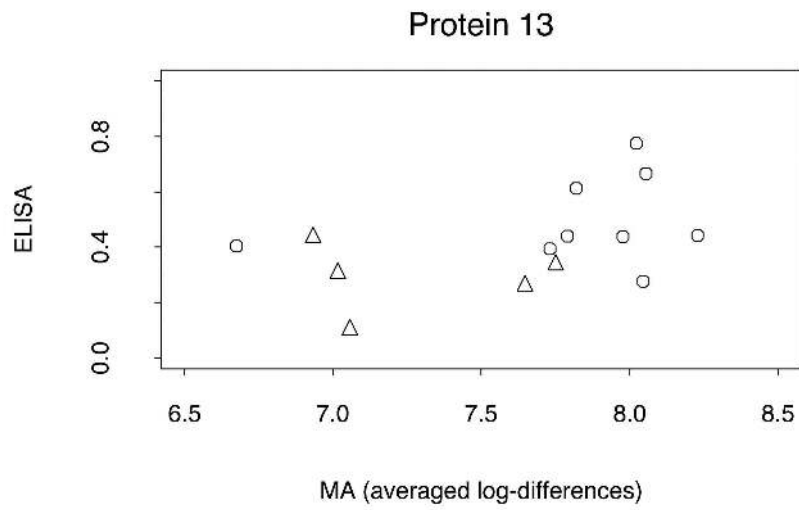
Figure 7

Scatter diagram of recorded ELISA and microarray (MA) data for a representative epitope (P13). The breast cancer patients are symbolised by circles and normals by triangles.
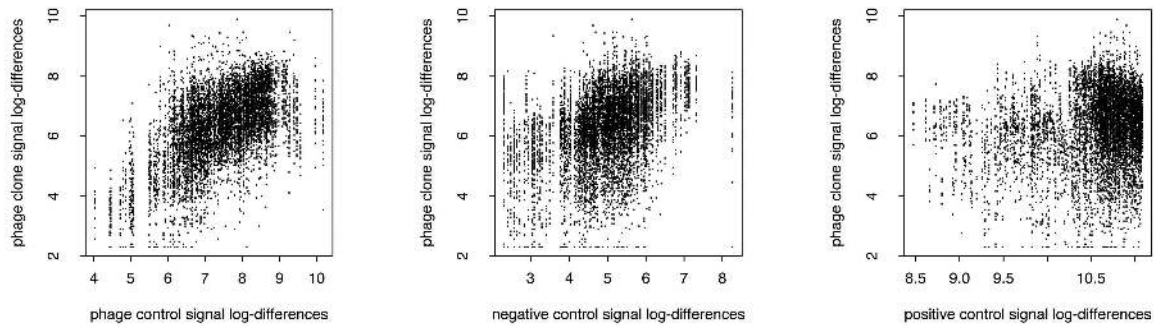
Figure 8

Scatter diagrams of phage clone signal log-differences versus (A): phage control signal log-differences, (B): negative control-II signal log-differences and (C): positive control-I signal log-differences. The effect of the lower cut-off limit is slightly visible in some of the plots through dots piling up at the limit value (=log(10)).
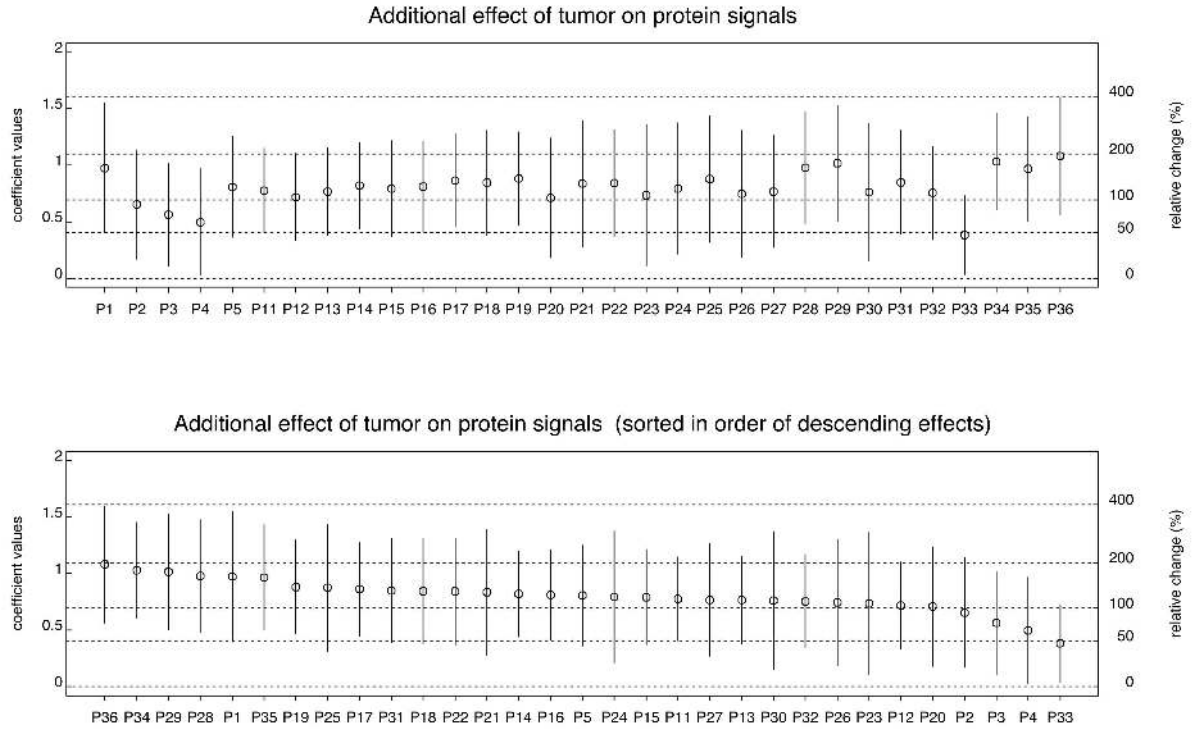
Figure 9

Phage clone additional effects of having tumour: The plots display point estimates and 95% confidence intervals for the coefficients of each of the phage clone/tumour cross-terms in the model. Along the x-axis, phage clones are sorted by clone numbers in the upper panel and by estimated effects in the lower panel. The left y-axis scale gives coefficient values (δ_p 's), whereas the right-hand side specifies the relative change in signal log-differences caused by serum coming from the breast cancer patients versus normals.

Appendix

In the Oswald software package, the user has to specify a so-called working correlation matrix, which states the structure of the correlations between observations within the same individuals. For the present model, we have used an "exchangeable correlation matrix", which means that all observations within the same individual have equal correlation. Then the estimation is carried out under the assumption that (the structure of) the working correlation matrix is true, and standard errors of the estimates are calculated.

The theory of GEE says that the estimates are asymptotically normal distributed with the estimated standard errors. Further, the estimates are consistent (*i.e.* converges to the true values when the number of observations becomes large), *even if the statistical distribution or the working correlation matrix is incorrectly specified*. This important result does not imply that the choices of distribution and working correlation matrix are of no consequence. The closer they are to reality, the more precise the estimates will be (for a given number of observations).

Table 3. Coefficients of base level antibody effects (γ_p's) and phage clone/tumour cross-term effects (δ_p's): Parameter estimates and their standard deviations.

Coefficient	Estimate	St.Dev.	Coefficient	Estimate	St.Dev.
γ_1	2.22	1.60	δ_1	0.90	0.29

γ_2	3.20	1.55	δ_2	0.58	0.24
γ_3	3.30	1.54	δ_3	0.48	0.23
γ_4	2.85	1.57	δ_4	0.42	0.23
γ_5	2.97	1.55	δ_5	0.73	0.22
γ_{11}	3.89	1.46	δ_{11}	0.68	0.17
γ_{12}	3.84	1.46	δ_{12}	0.62	0.17
γ_{13}	4.16	1.45	δ_{13}	0.67	0.18
γ_{14}	3.68	1.47	δ_{14}	0.72	0.18
γ_{15}	4.10	1.46	δ_{15}	0.69	0.20
γ_{16}	3.30	1.48	δ_{16}	0.72	0.19
γ_{17}	4.10	1.46	δ_{17}	0.77	0.20
γ_{18}	4.03	1.46	δ_{18}	0.75	0.22
γ_{19}	3.55	1.47	δ_{19}	0.78	0.20
γ_{20}	3.56	1.52	δ_{20}	0.62	0.25
γ_{21}	3.55	1.53	δ_{21}	0.74	0.26
γ_{22}	3.15	1.53	δ_{22}	0.75	0.22
γ_{23}	2.47	1.56	δ_{23}	0.65	0.31
γ_{24}	3.81	1.54	δ_{24}	0.71	0.28
γ_{25}	3.36	1.53	δ_{25}	0.79	0.27
γ_{26}	3.72	1.52	δ_{26}	0.66	0.27

γ_{27}	4.21	1.52	δ_{27}	0.68	0.24
γ_{28}	3.34	1.52	δ_{28}	0.89	0.24
γ_{29}	3.14	1.53	δ_{29}	0.93	0.25
γ_{30}	3.35	1.51	δ_{30}	0.71	0.29
γ_{31}	3.53	1.50	δ_{31}	0.80	0.22
γ_{32}	1.65	1.54	δ_{32}	0.71	0.21
γ_{33}	1.27	1.53	δ_{33}	0.34	0.19
γ_{34}	1.93	1.52	δ_{34}	0.98	0.21
γ_{35}	2.18	1.52	δ_{35}	0.92	0.22
γ_{36}	2.25	1.52	δ_{36}	1.04	0.25