

TITLE PAGE

Leticia Fernández, M. José Bleda, M. José Gómara and Isabel Haro*

DESIGN AND APPLICATION OF GB VIRUS C (GBV-C) PEPTIDE MICROARRAYS FOR DIAGNOSIS OF GBV-C / HIV-1 CO-INFECTION

Unit of Synthesis and Biomedical Applications of Peptides. IQAC-CSIC. Jordi Girona, 18-26 08034 Barcelona, Spain

*Corresponding author:

isabel.haro@iqac.csic.es

Telephone: +34 934006109

Fax: +34 932045904

Abstract

The main objectives of the design of GBV-C peptide microarrays are the miniaturisation of antigen-antibody interaction assays, the simultaneous analysis of several peptide sequences and the reduction in the volume of serum required from patients, since this always represents a limiting factor in studies to develop new systems for diagnosing human diseases. We herein report the design of a microarray immunoassay based on synthetic peptides derived from the GBV-C E2 protein to evaluate their diagnostic value in detecting anti-E2 antibodies in HIV-1 patients. To this end, peptide microarrays were initially prepared to identify the most relevant epitopes in the GBV-C E2 protein. Thus, 124 peptides composed of 18 amino acids covering the whole E2-protein sequence, with 15 residue overlaps, were spotted in triplicate onto γ -aminopropyl silane-functionalised adsorbent binding slides. The procedure to select the E2-protein epitopes was carried out using serum samples from HIV-1 infected patients. The samples had previously been tested for the presence or absence of GBV-C anti-E2 antibodies by means of the Abbott test. Thus, 11 specific epitopes in the GBV-C E2 protein were identified. Subsequently, peptide antigen microarrays were constructed using the E2 epitopes identified to detect GBV-C anti-E2 antibodies in the serum of HIV-1 infected patients with no known GBV-C co-infection. The 11 peptides selected identified anti-E2 GBV-C antibodies among HIV-1 infected patients and a reactivity of 47% was established. The potential antigenic peptides selected could be considered a useful tool for designing a new diagnostic system based on peptide microarrays to determine anti-GBV-C E2 antibodies in the serum of HIV-1 infected patients.

Keywords: GBV-C/HIV-1 co-infection, Diagnosis, GBV-C peptides, Microarrays.

Introduction

The GB virus C (GBV-C) (formerly known as hepatitis G virus) is a virus in the *Flaviviridae* family made up of a single chain of RNA. There is clear evidence that GBV-C is transmitted by sexual and percutaneous routes and is frequently found in populations at risk for blood-borne or sexually transmitted viruses. Thus, GBV-C is more frequently detected in groups at higher risk for hepatitis infection by similar routes of transmission and in patients treated with multiple haemodialysis procedures and a high number of transfused blood product units [1]. GBV-C prevalence among patients with HIV mono-infection varies from 14% to 43%, depending on the population studied [2].

However, no impact on health [3, 4] was identified until Prof. Tillmann's research team demonstrated that GBV-C viraemia is associated with a significant beneficial effect on the survival of HIV-infected patients [5]. These results were subsequently confirmed by this group and other research groups [6, 7]. Although the results were not always clearly significant [8], a meta-analysis underlined GBV-C's association with a beneficial effect of the course of the disease [9, 10]. Although there are currently no commercial systems for detecting specific markers of GBV-C infection, active GBV-C infection has been detected by RT-PCR [11], while past infection has been detected by the presence of anti-E2 antibodies [12] using an ELISA assay involving the E2 recombinant protein. The development of an anti-E2 detection method has led to a complete definition of the prevalence of GBV-C since E2 antibodies are several times more frequently detectable than RNA in blood donors [13]. Although the appearance of anti-E2 antibodies is considered to be an indication of past infection, and detection of these antibodies and the active presence of GBV-C [14] are unlikely to be concomitant, recent literature provides descriptions of results that feature duality in terms of the positive presence of anti-E2 antibodies and the presence of viral RNA [15].

On the other hand, it has been proposed that E2 antibodies are not a reliable marker of past GBV-C infection in populations with impaired immune function [16]. In particular, HCV co-infected patients receiving interferon therapy are able to eliminate the virus without the development of anti-E2 antibodies. Thus, the elimination of the infection without antibodies against the E2 protein has raised questions about the usefulness of E2 antibodies as a marker of past GBV-C infection among HIV and HIV-HCV co-infected patients.

The abovementioned results point to the need for further studies on the development of the antibodies in order to develop an understanding of the effect of exposure to GBV-C on human health in the context of other viral infections.

Our group studied synthetic peptides derived from GBV-C for the development of new systems for diagnosing infections caused by the virus. We investigated the capacity of the synthetic peptides to recognise anti-GBV-C antibodies in HIV-1 and HCV/HIV-1 co-infected patients in order to secure a better understanding of the effect of exposure to GBV-C on the progression of illness caused by HIV-1 infection, as well as its putative role as a prognostic marker in the context of other viral infections. Furthermore, we recently described specific domains of the E2 envelope protein of GBV-C that interfered with the HIV-1 fusion peptide vesicle interaction, notably reduced cellular membrane fusion and interfered with HIV-1 infectivity in a dose-dependent manner [17].

In recent years, microarrays have become invaluable research tools for life scientists. The main interest in array technology lies in the capacity to analyse a large number of molecules in one single experiment. More specifically, the protein or peptide microarrays allow for the analysis of large-scale, high-sensitivity protein-protein interactions, making them a very attractive technique in proteomic studies in search of biomarkers, in the screening of new drugs and in

disease diagnosis trials [18, 19]. In fact, the use of microarrays has emerged as a very promising diagnostic tool in medicine. In particular, peptide microarrays have been used in both autoimmune and allergic human diseases. Therefore, immunoenzymatic peptide assays have been designed for the diagnosis of the human herpes and corona viruses [20] and to identify different serum profiles in rheumatoid arthritis [21].

We herein report the design of a microarray immunoassay based on synthetic peptides derived from the GBV-C E2 protein to evaluate their diagnostic value in detecting anti-E2 antibodies in HIV-1 patients. To this end, peptide microarrays were used to map the binding epitopes of the anti-E2 IgG antibodies produced in GBV-C infection. Eleven specific IgG surface-binding epitopes on the E2 GBV-C protein were identified. Peptide antigen microarrays were then constructed using the identified E2-binding epitopes for serum detection of IgG in GBV-C infection. Thus, serodiagnostic tests were performed with the designed GBV-C E2 peptide microarrays to establish exposure to GBV-C in HIV-1 co-infected patients.

Material and methods

Peptide synthesis

The synthesis of the peptides was performed using an Fmoc-based solid-phase procedure in a semi-automatic multiple peptide synthesiser, as previously described [17]. The 124 peptides consisting of 18 amino acids in length were overlapped by 15 residues, covering the sequence of the best-preserved primary structure of the GBV-C E2 protein.

Two irrelevant peptides were used as negative controls.

Human serum samples

Three different serum panels were analysed:

The first panel consisted of 76 sera from HIV-infected patients who attended an outpatient clinic at Medizinische Hochschule in Hannover (Germany). Thirty-eight sera tested positive for GBV-C anti-E2 antibodies (using the Abbott test) and negative for GBV-C RNA (HIV-1+/E2 GBV-C+), and 38 tested negative for GBV-C anti-E2 antibodies and negative for GBV-C RNA (HIV-1+/E2 GBV-C-).

The second panel consisted of 38 control sera from volunteer blood donors at Hospital Clínic Barcelona (negative controls).

The third panel consisted of 60 sera from HIV-1 infected patients at the Hospital Universitari de Bellvitge in Barcelona (serum testing panel).

This study was approved by the hospitals' ethics committees.

Optimisation of microarrays

In order to identify the best conditions for performing the assays, optimisation experiments were conducted. For these purposes, two different surfaces for immobilisation were used: γ -aminopropyl silane-functionalised adsorbent binding slides (UltraGAPS, Gamma Amino Propyl Silane, Corning) and 3-glycidoxypropyltrimethoxysilane-functionalised covalent slides (Epoxide, Corning). Different concentrations of peptides and serum samples were also tested.

The best results were obtained with the UltraGAPS slides. Minimisation of the non-specific interactions was obtained by working at a peptide concentration of 1 $\mu\text{g}/\mu\text{l}$ and a serum sample dilution of 1:200.

Preparation and processing of microarrays

Peptides were diluted in 10% DMSO/PBS to 1 $\mu\text{g}/\mu\text{l}$ and transferred into 384-well plates. A robotic microarrayer from ArrayIt Corporation was used to spot the peptides onto the amino-silane surface on the UltraGAPS slides. The spotted microarrays were stored overnight at 4°C, kept free from moisture and blocked the next day with PBS containing 0.005% of Tween 20 and 5% of bovine serum albumin (BSA) for 1 h at room temperature.

Each slide was divided into 14 wells with silicon seals that were attached to a solid support, creating a sealed incubation chamber inside.

Each well of the array was incubated overnight with 100 μl of serum (diluted 1:200 with PBS containing 0.005% Tween and 5% BSA). After being rinsed and washed six times for 2 min in PBS with 0.05% Tween, arrays were incubated with 100 μl of a 1:1000 dilution of a donkey anti-human IgG Dylight 649-conjugated secondary antibody (Jackson ImmunoResearch, West Grove, PA, USA) in PBS (0.005% Tween + 5% BSA) for 1.5 h at room temperature. The arrays were then rinsed and washed again six times for 2 min. After removing the silicon seals, the arrays were washed twice for 10 min in PBS-T and twice for 15 seconds in distilled water. The arrays were spun dry and scanned using a GenePix 4000B Scanner.

Statistical analysis

Fluorescence intensities (medians after subtraction of the local background) were calculated from the scanned array images with GenePix Pro 6.1.0.4 (Molecular Devices, Sunnyvale, CA, USA). Fluorescence data distribution was skewed and log transformation did not normalise the data. As described by others authors, some normalisation methods may even be more prejudicial than beneficial when trying to make data from different peptide microarrays more comparable [22]. Therefore, to compare non-normal data, non-parametric techniques, such as the Wilcoxon rank-sum test [23], were used. To take account of the large number of pair comparisons between the panels of patients, a multiple-test procedure method featuring multiple comparisons, known as the false discovery rate, was applied [24-27].

In order to quantify the accuracy of a diagnostic test and estimate its specificity (the fraction of non-co-infected patients correctly classified by the diagnostic test) and sensitivity (fraction of co-infected patients correctly classified), a non-parametric estimation of the receiver operating characteristic (ROC) curve was used [28]. The area under the ROC curve (AUC) indicates the global performance of a diagnostic test and was estimated using the trapezoidal rule [29, 30].

The greater the AUC, the better the global performance of the diagnostic test. All statistical analyses were performed using the statistical software STATA 12.0 [31].

Results and discussion

The most commonly used method to diagnose viral infectious diseases is based on the detection of antigen-specific antibodies from serum samples by conventional ELISA assays. Although this format offers sensitivity, specificity and automation, a drawback is the lack of sample multiplexing. Transferring these immunological assays from microtiter plates to microarray formats makes it possible to miniaturise the antigen-antibody interaction assays, analyse several peptide sequences simultaneously and reduce the volume of serum required from patients, since this always represents a limiting factor in studies to develop new systems for diagnosing human diseases. Figure 1 shows the scheme of a peptide-based microarray for detecting specific antibodies.

Furthermore, high-density peptide microarrays have been described as powerful tools for measuring the reactivity of antibodies against thousands of peptides simultaneously [22, 32]. Peptide microarrays can be applied in the context of infectious diseases to determine specific antibody recognition sequences. In this work, peptide microarrays were prepared to identify and define relevant epitopes in E2 proteins from GBV-C. 124 peptides of 18 amino acids covering the whole E2 protein sequence, with 15 residue overlaps, were spotted in triplicate on planar slides. After testing different surfaces for the immobilisation of the 124 synthesised peptides [17] during screening of the primary structure of the E2 envelope protein of GBV-C, the solid medium that made it possible to minimise the non-specific interactions with the serums was that of γ -aminopropyl silane-functionalised adsorbent binding (UltraGAPS, Gamma Amino Propyl Silane).

The procedure to select the relevant E2-protein epitopes was carried out using serum samples from HIV-1 infected patients. The samples had previously been tested for the presence or absence of GBV-C anti-E2 antibodies by means of the discontinued Abbott test **that have as antigenic substrate the E2 recombinant protein. Within each slide we have incorporated positive [33] and negative controls to address the specificity of the recognition signal we found.** By using a human/monkey anti-IgG as a secondary antibody conjugated to Dylight 649 fluorescent dye, the required volume of human serum was reduced considerably compared to the amount used in ELISA assays [33]. The selection procedure of E2 antigenic peptides was carried out in three steps, as described below.

Peptide selection procedure

Initial peptide selection: GBV-C E2 protein screening

Eight HIV-1+/E2 GBV-C+ sera, eight HIV-1+/E2 GBV-C- sera and four negative control sera were randomly selected in order to identify a subset of the 124 initial peptides indicating potentially antigenic zones. The fluorescence of each peptide (in triplicate) was obtained for each serum sample **(Figures 1 and 2A).**

Using the results of the eight HIV-1+/E2 GBV-C+ sera, three zones with very high fluorescence values were identified: zone I (peptide 1 to peptide 9), zone II (peptide 10 to peptide 27) and zone III (peptide 45 to peptide 55), covering 38 peptides (31% of the initial 124 peptides) (Figure 2B). Focusing on these three zones, we compared the distribution of the fluorescence values of the HIV-1+/E2 GBV-C+, HIV-1+/E2 GBV-C- and negative control sera between the 38 peptides identified and the 86 that remained. The non-parametric Wilcoxon rank-sum test was used. The selected peptides showed statistically significant higher fluorescence values than the non-selected peptides.

Likewise, in the 38 peptides selected, the fluorescence values were significantly higher in the HIV-1+/E2 GBV-C+ sera than in either the HIV-1+/E2 GBV-C- or negative control sera. These results indicate that the 38 identified peptides could differentiate the serum panels analysed. This set of peptides showed a higher reactivity with serum from the HIV-1/GBV-C co-infected patients compared to the reactivity obtained with serum from the HIV-1 patients without GBV-C E2 antibodies and serum from the healthy volunteers.

Second selection: comparison of the HIV-1+/E2 GBV-C and negative control panels

Taking into account the first selection of peptides with the limited number of characterised sera, an antigenic study was carried out using the whole first panel of 76 sera from HIV patients and the second serum panel from the volunteer blood donors.

All the serum samples in each panel were analysed using the 38 peptides previously identified plus two irrelevant peptides that were spotted in triplicate on the UltraGAPS slides, as previously described, and the median fluorescence value was obtained. Descriptive results (Figure 3, Table S1 Supplementary Material) showed the variation in fluorescence for the three groups of serum samples within a peptide and the differences between peptides within a group of sera.

For some peptides, the fluorescence values of the HIV-1+/E2 GBV-C- sera were higher than the values of the negative controls and, at the same time, lower than the values of the HIV-1+/E2 GBV-C+ patients (e.g. P3, P15). Nevertheless, in other peptides this was not observed (e.g. P1, P27). These results indicate that some peptides seem to specifically recognise antibodies in the HIV-1+/E2 GBV-C+ sera previously tested with the Abbott test.

To select a new set of peptides, we conducted two comparisons of the fluorescence data distribution between serum panels.

1) HIV-1+/E2 GBV-C- versus negative control groups and 2) HIV-1+/E2 GBV-C+ versus HIV-1+/E2 GBV-C- groups. The distribution of the fluorescence values was compared for each peptide using the non-parametric Wilcoxon rank-sum test. Given the large number of comparisons, a multiple-test procedure method featuring multiple comparisons, known as the false discovery rate, was applied in order to reduce the chance of establishing differences among groups where there were none. As a result of the statistical analysis, the peptides whose fluorescence distribution in the serum panel from the HIV-1 patients was significantly higher than the fluorescence distribution in the negative control panel (the volunteer blood donors) were selected. Furthermore, those peptides whose fluorescence data distribution of the serum from the HIV-1 patients with GBV-C anti-E2 antibodies was higher and significantly different from those of the serum from the HIV-1 patients without GBV-C anti-E2 antibodies were selected. In addition to the above criteria, the

fluorescence median of the HIV-1+/E2 GBV-C E2+ group should be higher than 1000 (twice the background noise). P11 was considered a special case and was not selected regarding the 75th percentile values of fluorescence in HIV-1+/E2GBV-C- and HIV-1+/E2GBV-C+ panels. P19 and P54 were also considered special cases and were selected regarding the 75th percentile values of fluorescence (Table S2).

Overall, 17 peptides from the initial 38 sequences corresponding to the three putative antigenic regions of GBV-C E2 protein were selected. Thus, 44.7% of the peptides in the initial step (17/38) were selected: 33% of the peptides in zone I (P3, P4 and P6); 61% of the peptides in zone II (P10, P12 to P20, and P24); and 27% of the peptides in zone III (P51, P52 and P54).

Third selection: diagnostic capacity

The objective of diagnostic tests is to determine the presence or absence of a pathological event (illness, infection, etc.). Receiver operating characteristic (ROC) curves were used to calculate the sensitivity and specificity values of the test. The area under the curve (AUC) was also calculated as a measure of the discriminatory power of the diagnostic test. For a fixed specificity, established at 94.74%, the cut-off point of each peptide was determined as the value of maximum sensitivity.

Using the HIV-1+/E2 GBV-C+ and negative control serum samples, 17 ROC curves were estimated using a non-parametric method, one per peptide. The negative controls were selected as representative of a population with a very low risk of contact with GBV-C. Results are shown in Table 1. ROC curve analysis is shown in Figure S1 and fluorescence variation of the HIV-1+/E2 GBV-C+ group (as positive control) and negative control serum samples, together with the cut-off value of each peptide in Figure S2 of the section Supporting Information.

Regarding these results, peptides with a sensitivity higher than 79% and a percentage of serums correctly classified greater than or equal to 85% were selected. Eleven of the 17 peptides (64.7%) were then selected: zone I (P3, P4); zone II (P12, P13, P15, P16, P17, P18, P19, and P24); and zone III (P54). The AUC of five of these peptides was greater than 0.98, which indicates excellent discriminatory power; five peptides had an AUC of between 0.90 and 0.98, which indicates very good discriminatory power, and the AUC of peptide P54 was 0.87.

After all of these selection steps, 11 peptides (9% of the initial 124 peptides) of the GBV-C E2 protein were selected as potential antigenic peptides for the detection of anti-GBV-C antibodies.

Diagnostic value of antibodies against selected peptides in HIV-1 infected patients

In an attempt to apply GBV-C E2 peptides for the diagnosis of GBV-C in HIV-1 patients, we investigated the capacity of the 11 peptides selected to interact with anti-E2 antibodies in HIV-1 infected patients with no known GBV-C/HIV-1 co-infection. At this point we also analysed the selected 11 peptide sequences by means of the Basic Local Alignment Search Tool (BLAST) (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). The Blast programme finds regions of local similarity between sequences comparing nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

After analysing all selected peptide domains we proved that they just belong to different strains of GB virus C and they do not match to any other known virus. Thus, the peptides-sera reactivity observed could specifically be related to antibodies anti-E2.

Microarray assay was performed as described using UltraGAPS slides printed with the 11 peptides previously selected and the 60 serum samples from the third panel (serum testing panel). The median fluorescence value was obtained in triplicate. Descriptive results (Figure 4) showed the variation of fluorescence within a peptide for the testing panel, together with the HIV-1+/E2 GBV-C+ group (as a positive control) and negative control serum samples. The cut-off value for each peptide is shown.

In order to determine whether these 60 serum samples could be considered anti-E2 GBV-C antibody positive, serum patterns were created. The pattern of each serum is a description of the peptides selected in which the median fluorescence value of the serum was greater than or equal to the corresponding cut-off ($P_{\#CO}$) (Table 2). Thus, taking into account the number of $P_{\#CO}$ in each pattern, we defined a serum as anti-E2 GBV-C antibody positive when the pattern included at least six peptides. Moreover, the three antigenic zones previously identified had to appear within these patterns in their entirety, and we therefore ensured that at least one peptide from each zone was also present in a pattern for a serum to be considered anti-E2 GBV-C antibody positive (Table 3). These patterns demonstrate that patient serum shows reactivity against different peptides, which on their own are less specific markers for GBV-C infection than when combined. The creation of these patterns is important for the improvement of diagnostic accuracy with a combined set of peptides. Thus, peptide microarray technology can help establish reliable diagnostic biomarkers by employing a combination of antigenic peptides.

Twenty-eight of the 60 sera from the testing panel were classified as anti-E2 GBV-C antibody positive. This corresponded to a reactivity of 47%. This value is in agreement with the reported prevalence of GBV-C viraemia, which ranges from 14% to 43% among HIV-1 infected people [2]. Bearing in mind these results, the potential antigenic peptides selected could be considered a useful tool for designing a new diagnostic system based on peptide microarrays to determine anti-E2 GBV-C antibodies in the serum of HIV-1 infected patients.

Conclusions

This work demonstrates that peptide microarrays are instrumental in identifying relevant antigenic peptide sets in the GBV-C E2 protein since they make it possible to measure anti-E2 antibody responses to multiple peptide sequences simultaneously.

Our results show the usefulness of synthetic peptides as potential antigens for the development of a new GBV-C antibody peptide-based microarray for diagnosis in HIV-1 infected patients. The 11 peptides selected in this GBV-C E2 protein antigenic study identified anti-E2 GBV-C antibodies among HIV-1 infected patients, establishing a reactivity of 47%. Furthermore, these results reinforce the need for a combination of potentially antigenic epitopes of different GBV-C E2 domains to achieve a more accurate and precise diagnostic system.

Acknowledgements: The authors greatly acknowledge to Professors Tillmann (Duke Clinical Research Institute, USA), Ercilla (Hospital Clinic, Barcelona, Spain) and Casanova (Hospital de Bellvitge, Barcelona, Spain) for the donation of the serum samples. We also acknowledge to Drs. Lidia Sevilla and Ignacio Pons (Scientific Parc of Barcelona, Spain)

for their helpful advice during the preparation and processing of microarrays. This work was funded by grants CTQ2009-13969-C02-01 and CTQ2012-37589-C02-01 from the Spanish Ministerio de Ciencia e Innovación. LF is a recipient of a FPI grant from the Spanish Ministerio de Ciencia e Innovación.

Table 1. ROC Curves results by antigenic zones

	Peptide	AUC	95% CI	Sensitivity	Percentage of serums correctly classified	Cut-off
Zone I	P3	0.992	(0.98 ; 1.00)	97%	96%	≥ 882
	P4	0.942	(0.87 ; 1.00)	97%	96%	≥ 1426
	P6	0.787	(0.68 ; 0.90)	5%	50%	≥ 6315
Zone II	P10	0.751	(0.64 ; 0.86)	45%	70%	≥ 2484
	P12	0.989	(0.97 ; 1.00)	97%	96%	≥ 2448
	P13	0.939	(0.87 ; 1.00)	87%	91%	≥ 384
	P14	0.829	(0.73 ; 0.93)	24%	59%	≥ 6932
	P15	1.000	(1.00 ; 1.00)	100%	97%	≥ 1251
	P16	0.994	(0.98 ; 1.00)	97%	96%	≥ 1595
	P17	0.996	(0.99 ; 1.00)	97%	95%	≥ 1248
	P18	0.978	(0.95 ; 1.00)	87%	91%	≥ 911
	P19	0.910	(0.84 ; 0.99)	79%	87%	≥ 427
	P20	0.775	(0.66 ; 0.89)	55%	75%	≥ 1223
	P24	0.975	(0.95 ; 1.00)	87%	91%	≥ 368
Zone III	P51	0.704	(0.59 ; 0.82)	24%	59%	≥ 3524
	P52	0.838	(0.75 ; 0.93)	50%	72%	≥ 3104
	P54	0.869	(0.77 ; 0.97)	84%	90%	≥ 78

Table 2. Classification of testing serum samples by the selected peptides

Number of P##CO	Number of positive sera with peptides in the three antigenic zones	Number of positive sera with peptides in any antigenic zone	Total
1	0	1	1
2	0	4	4
3	1	5	6
4	0	2	2
5	0	2	2
6	0	4	4
7	5	5	10
8	3	2	5
9	4	5	9
10	9	1	10
11	7	0	7
Total	29	31	60

Table 3. Testing serum's patterns listed by number of P##_{CO}Sera with 7 P_{CO}

	Sera	Pattern									
		Zone I		Zone II						Zone III	
1.	31		P4 _{CO}	P12 _{CO}		P16 _{CO}		P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
2.	33		P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}		P19 _{CO}	P24 _{CO}	P54 _{CO}
3.	46		P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}		P18 _{CO}	P24 _{CO}	P54 _{CO}
4.	51	P3 _{CO}	P4 _{CO}			P15 _{CO}	P16 _{CO}		P18 _{CO}	P24 _{CO}	P54 _{CO}
5.	55		P4 _{CO}			P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}

Sera with 8 P_{CO}

	Sera	Pattern										
		Zone I		Zone II						Zone III		
1.	29		P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}		P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
2.	36	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}		P19 _{CO}	P24 _{CO}	P54 _{CO}	
3.	54		P4 _{CO}		P13 _{CO}	P15 _{CO}	P16 _{CO}		P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}

Sera with 9 P_{CO}

	Sera	Pattern										
		Zone I		Zone II						Zone III		
1.	14	P3 _{CO}		P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
2.	21		P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
3.	35		P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
4.	49	P3 _{CO}			P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}

Sera with 10 P_{CO}

	Sera	Pattern										
		Zone I		Zone II						Zone III		
1.	1	P3 _{CO}		P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
2.	3	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
3.	7	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
4.	16		P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
5.	26	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
6.	27	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
7.	28	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
8.	34	P3 _{CO}	P4 _{CO}	P12 _{CO}		P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
9.	38	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}		P24 _{CO}	P54 _{CO}

Sera with 11 P_{CO}

	Sera	Pattern										
		Zone I		Zone II						Zone III		
1.	5	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
2.	6	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
3.	30	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
4.	39	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
5.	41	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
6.	48	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}
7.	53	P3 _{CO}	P4 _{CO}	P12 _{CO}	P13 _{CO}	P15 _{CO}	P16 _{CO}	P17 _{CO}	P18 _{CO}	P19 _{CO}	P24 _{CO}	P54 _{CO}

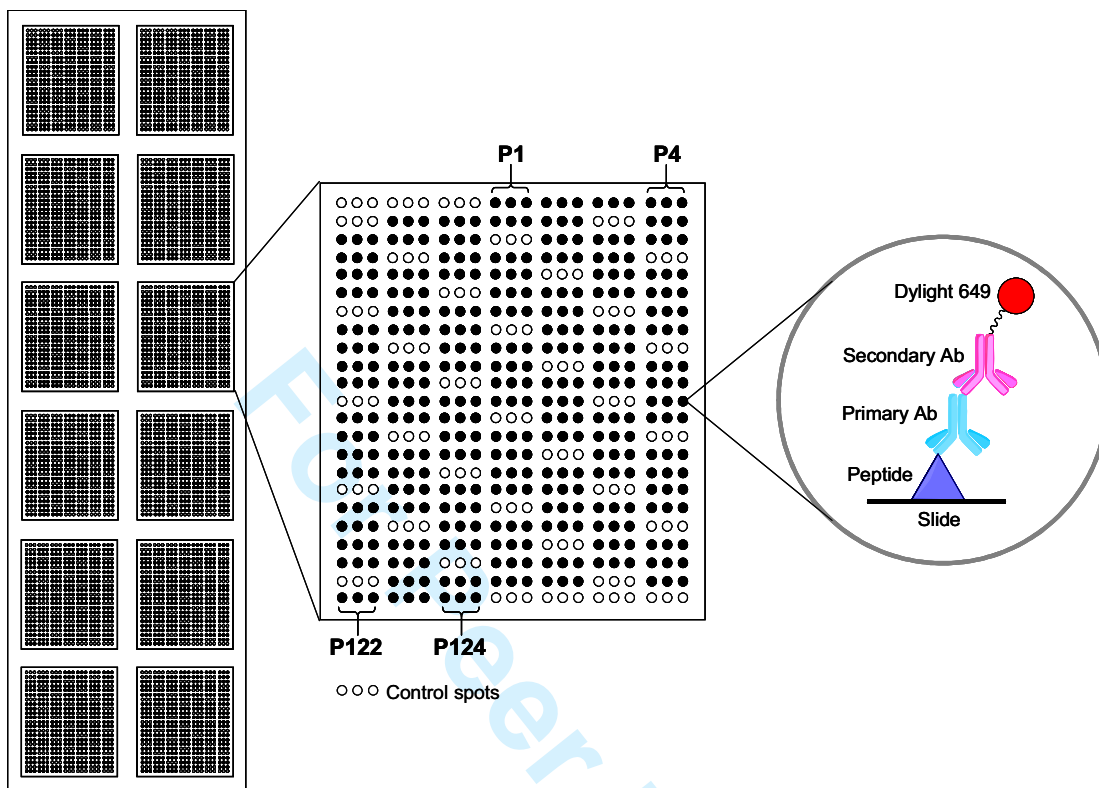
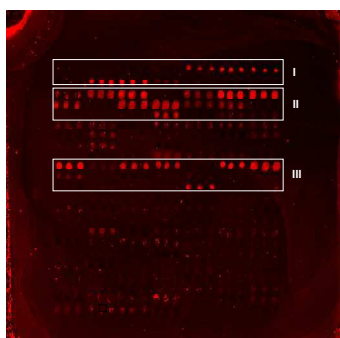


Figure 1

A)



B)

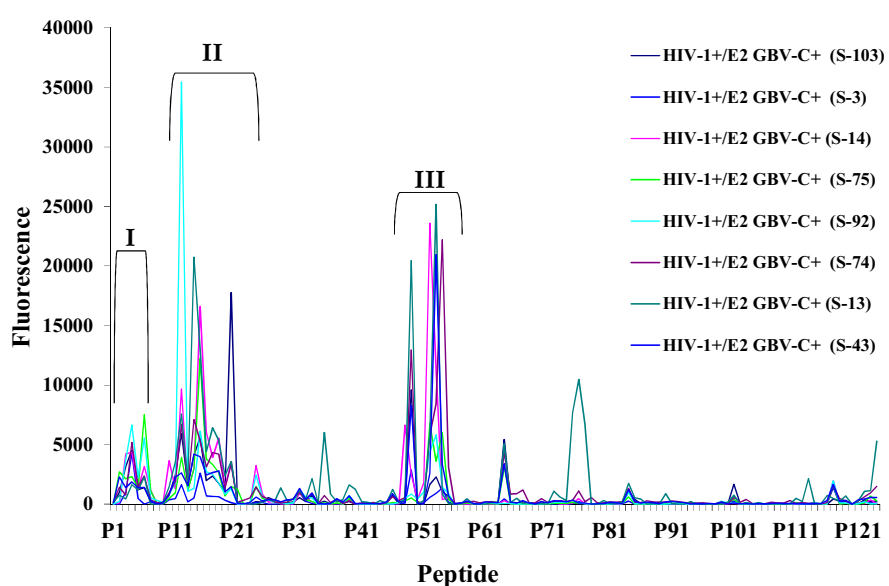
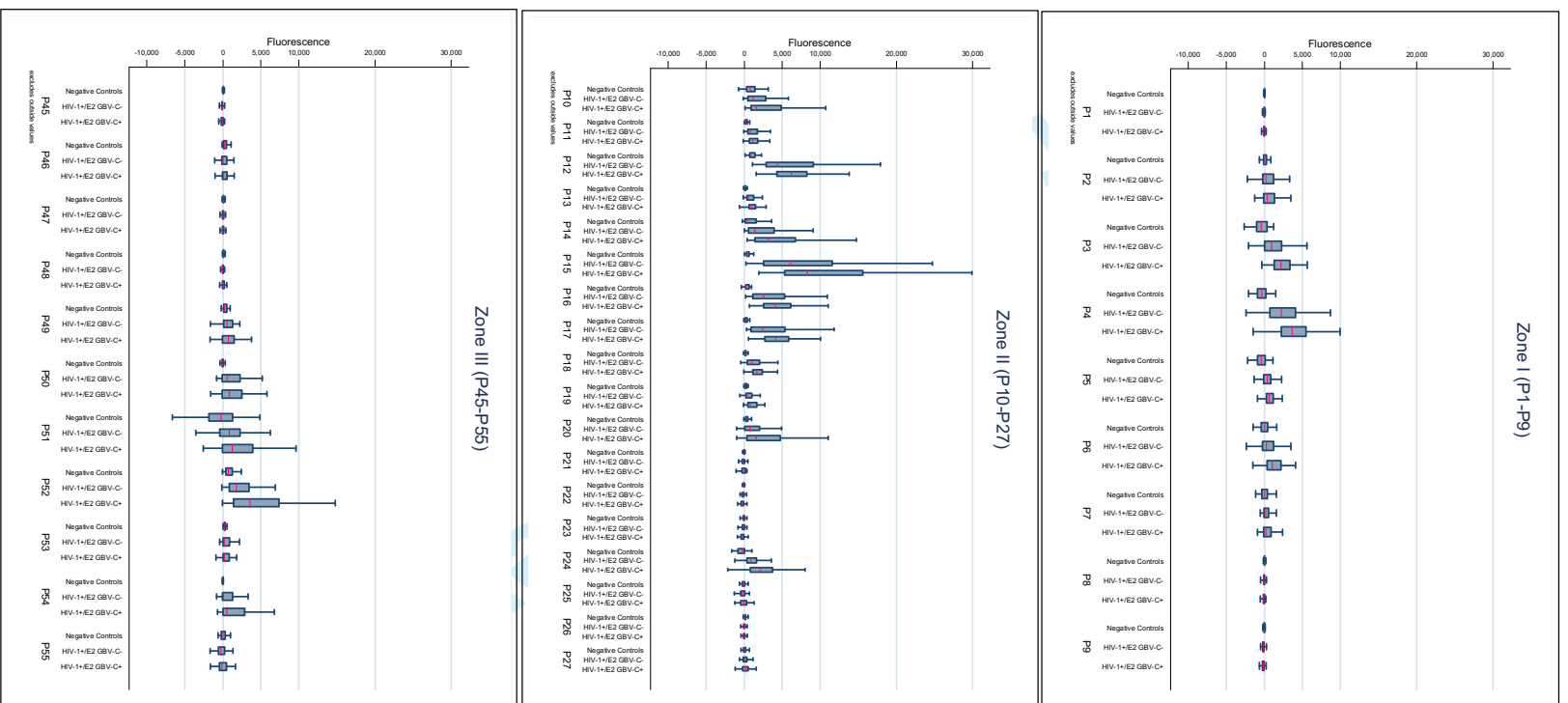


Figure 2

Figure 3



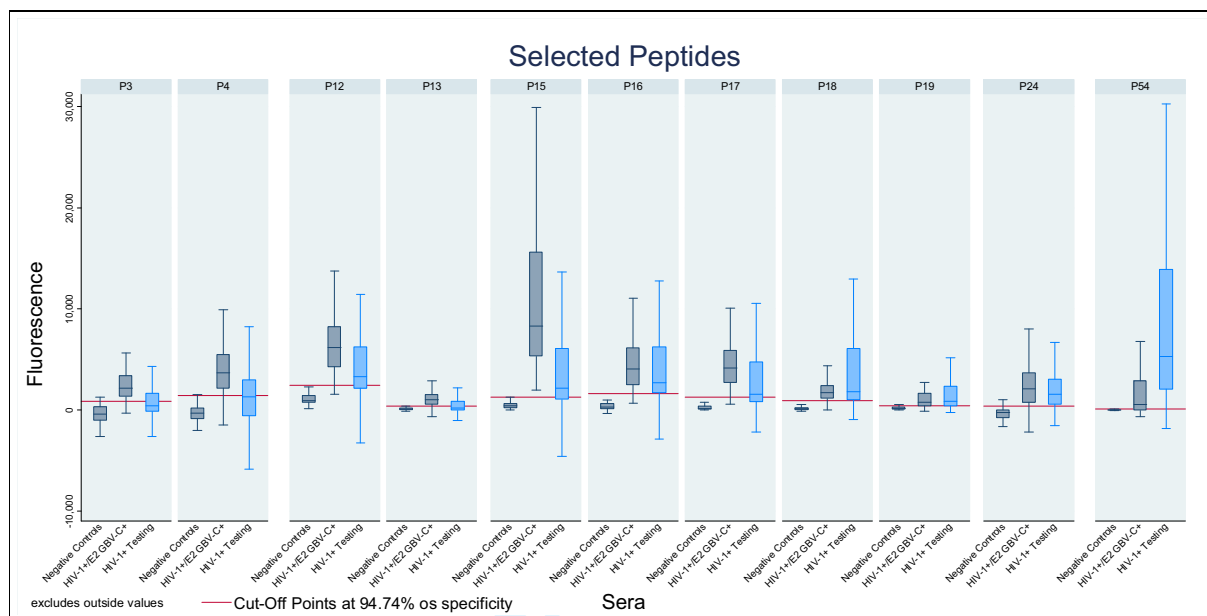


Figure 4

FIGURE CAPTIONS

Figure 1. Scheme of a peptide-based microarray.

Figure 2. A) Scanned image of a slide. B) Fluorescence values of 124 GBV-C E2 peptides for 8 HIV-1+/E2 GBV-C+ sera.

Figure 3. Distribution of fluorescence values for the three groups of serum samples by antigenic zones.

Figure 4. Distribution of fluorescence values for the testing panel together with HIV-1+/E2 GBV-C+ group (as positive control) and negative controls serum samples by peptide. Horizontal red line indicates the ROC cut-off value of each peptide.

Supplementary material

Table S1. Description of fluorescence values of microarray assays for the three groups of sera by antigenic zone. Median and Interquartile Range (IR).

	Peptide	Sera					
		Negative controls n=38		HIV-1+/E2 GBV-C- n=38		HIV-1+/E2 GBV-C+ n=38	
		Median	IR	Median	IR	Median	IR
Zone I	P1	1	39	-37	98	-70	151
	P2	15	445	185	1480	359	1459
	P3	-427	1336	959	2298	2160	2099
	P4	-317	1093	2249	3355	3661	3353
	P5	-425	1041	400	1023	656	919
	P6	12	855	285	1562	1060	1857
	P7	37	755	203	657	187	1019
	P8	-11	90	-60	199	-84	241
	P9	-4	80	-113	231	-154	265
Zone II	P10	773	1153	1156	2420	1571	3982
	P11	250	246	783	1349	1130	1099
	P12	912	650	4383	6182	6184	3975
	P13	81	167	547	930	1014	942
	P14	391	1463	1427	3419	3232	5373
	P15	413	417	6088	9051	8277	10297
	P16	263	458	2557	4147	4078	3638
	P17	159	316	2453	4475	4156	3130
	P18	133	194	1011	1648	1687	1294
	P19	159	184	423	781	800	1232
	P20	296	330	742	1964	1573	4453
	P21	-3	88	-130	328	-76	507
	P22	-49	110	-150	230	-177	342
	P23	-92	221	-134	316	-170	356
	P24	-255	795	878	1313	2087	2922
	P25	-86	366	-278	548	-18	754
	P26	86	210	-11	243	-17	221
P27	13	327	55	547	208	753	
Zone III	P45	47	68	-88	172	-94	243
	P46	203	436	43	744	155	669
	P47	67	119	8	196	20	192
	P48	71	107	-18	148	39	248
	P49	246	422	551	1188	738	1606
	P50	-59	199	593	2412	887	2619
	P51	-248	3187	826	2712	1216	4028
	P52	763	859	1799	2573	3550	6006
	P53	191	186	183	912	165	871
	P54	-25	46	96	1369	536	2945
Controls	C1	59	77	-18	253	-3	330
	C2	27	186	3	141	5	192

Table S2. Second Selection: fluorescence comparison between pairs of panels. Results of Wilcoxon Rank-Sum Test (p-values). Multiple comparison adjustment using *False Discovery Rate* (FDR) correction with Simes's method. A comparison is considered statistically significant if the corresponding p-value is lower or equal the specific $\alpha_{\text{corrected}}$ for this comparison. Selected peptides (shadowed) were those peptides with greater fluorescence values in HIV-1+/E2GBV-C- panel than in Negative Controls panel AND greater values in HIV-1+/E2GBV-C+ panel than in HIV-1+/E2GBV-C- panel AND with a median of fluorescence greater than 1,000 in HIV-1+/E2GBV-C+ panel.

Peptide		Negative controls vs HIV-1+/E2GBV-C-		HIV-1+/E2GBV-C- vs HIV-1+/E2GBV-C+	
		p-value	$\alpha_{\text{corrected}}$	p-value	$\alpha_{\text{corrected}}$
Zone I	P1	---		---	
	P2	0,0239*	0,0350	0,4907	0,0457
	P3	0,0000*	0,0007	0,0000*	0,0129
	P4	0,0000*	0,0014	0,0000*	0,0136
	P5**	0,0000*	0,0021	0,0003*	0,0236
	P6	0,0175*	0,0343	0,0000*	0,0143
	P7	0,0008*	0,0257	0,9316	0,0500
	P8	0,0008*	0,0264	0,1433	0,0379
	P9	---		0,1125	0,0364
Zone II	P10	0,0007*	0,0250	0,0001*	0,0221
	P11 ^o	0,0000*	0,0029	0,0128*	0,0336
	P12	0,0000*	0,0036	0,0009*	0,0271
	P13	0,0000*	0,0043	0,0000*	0,0150
	P14	0,0000*	0,0050	0,0000*	0,0157
	P15	0,0000*	0,0057	0,0000*	0,0164
	P16	0,0000*	0,0064	0,0000*	0,0171
	P17	0,0000*	0,0071	0,0000*	0,0179
	P18	0,0000*	0,0079	0,0000*	0,0186
	P19 ^{oo}	0,0000*	0,0086	0,0000*	0,0193
	P20	0,0268*	0,0357	0,0000*	0,0200
	P21	0,0000*	0,0093	0,3264	0,0421
	P22	0,0000*	0,0100	0,1823	0,0407
	P23	0,0064*	0,0321	0,2023	0,0414
P24	0,0000*	0,0107	0,0000*	0,0207	
P25**	0,0027*	0,0286	0,0039*	0,0293	
P26	---		0,6538	0,0464	
P27	0,3617	0,0429	0,1809	0,0400	
Zone III	P45	---		0,0900	0,0493
	P46	---		0,3993	0,0436
	P47	---		0,4277	0,0450
	P48	---		0,0054*	0,0314
	P49	0,0005*	0,0243	0,7403	0,0471
	P50	0,0000*	0,0114	0,9233	0,0486
	P51	0,0002*	0,0229	0,0044*	0,0307
	P52	0,0000*	0,0121	0,0000*	0,0214
	P53	0,1500	0,0386	0,7815	0,0479
	P54 ^{oo}	0,0016*	0,0279	0,0042*	0,0300
P55	---		0,1570	0,0393	

--- Inconsistent comparison: inconsistent fluorescence values between panels (greater values of fluorescence in Negative control panel than in HIV-1+/E2GBV-C- AND/OR greater values of fluorescence in HIV-1+/E2GBV-C- panel than in HIV-1+/E2GBV-C+).

1
2
3 * Comparison statistically significant ($p\text{-value} \leq \alpha_{\text{corrected}}$).

4 ** Median of fluorescence less than 1,000 in HIV-1+/E2GBV-C+ panel.

5 ° P11: special case of a non selected peptide because even though meeting all the criteria, the 75th percentiles of
6 fluorescence in HIV-1+/E2GBV-C- and HIV-1+/E2GBV-C+ panels were almost identical (1,749 and 1764,5).
7

8 °° P19 and P54: special cases of selected peptides even though not meeting one criteria. Median of fluorescence less
9 than 1,000 but with 75th percentile values high (1,661.5 and 2,902 respectively).
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Figure S1. ROC curves.

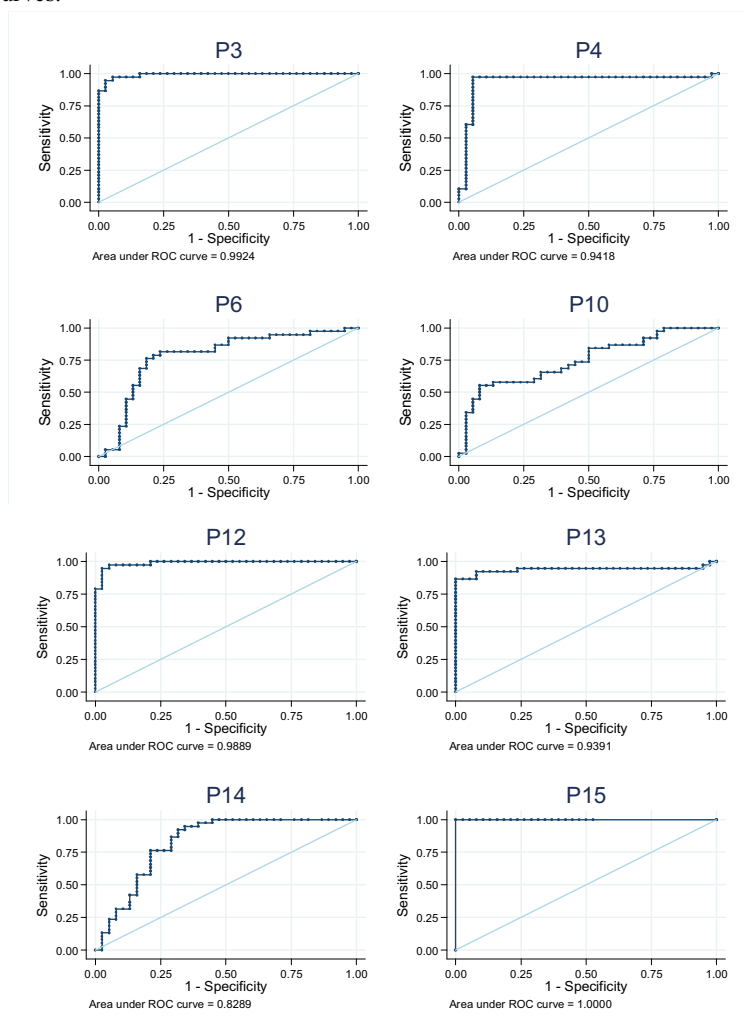


Figure S1 (continuation). ROC curves.

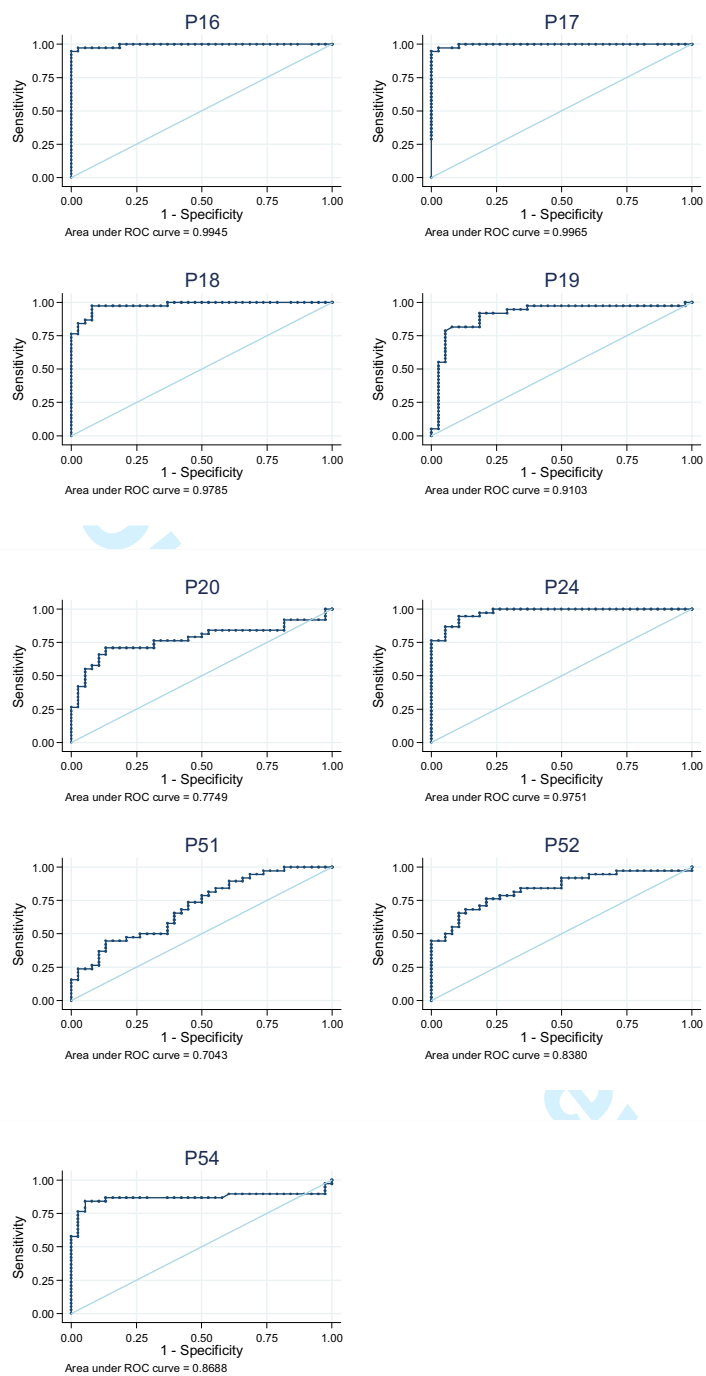
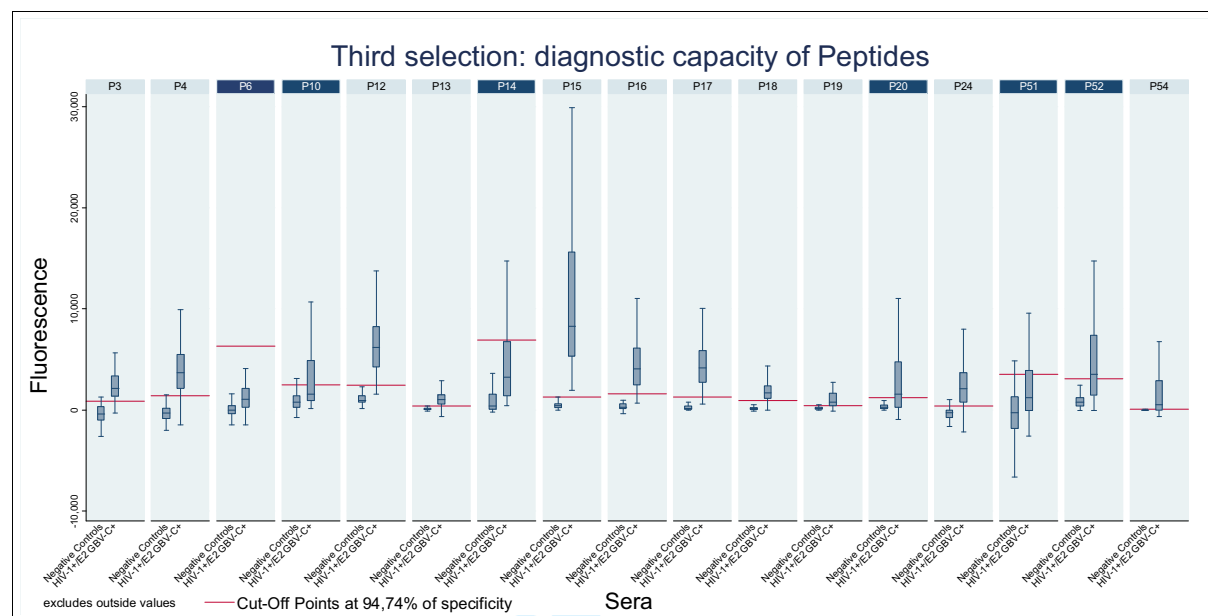


Figure S2. Distribution of fluorescence values for HIV-1+/E2 GBV-C+ group (as positive control) and negative controls serum samples by peptide. The cut-off value of each peptide is the horizontal red line.



Bibliografia

- [1] G.L. Baggio-Zappia, C.F.H. Granato, *Clinical Chemistry and Laboratory Medicine*, 47 (2009) 12-19.
- [2] S. George, D. Varmaz, J.T. Stapleton, D. Varmaz, J. Stapleton, *The Journal of infectious diseases*, 193 (2006) 451-454.
- [3] M.J. Mphahlele, G.K.K. Lau, W.F. Carman, *Liver*, 18 (1998) 143-155.
- [4] D. Theodore, S.M. Lemon, *Hepatology*, 25 (1997) 1285-1286.
- [5] S. Heringlake, J. Ockenga, H.L. Tillmann, C. Trautwein, D. Meissner, M. Stoll, J. Hunt, C. Jou, N. Solomon, R.E. Schmidt, M.P. Manns, *Journal of Infectious Diseases*, 177 (1998) 1723-1726.
- [6] H.L. Tillmann, H. Heiken, A. Knapik-Botor, S. Heringlake, J. Ockenga, J.C. Wilber, B. Goergen, J. Detmer, M. McMorro, M. Stoll, R.E. Schmidt, M.P. Manns, *New England Journal of Medicine*, 345 (2001) 715-724.
- [7] J.H. Xiang, S. Wunschmann, D.J. Diekema, D. Klinzman, K.D. Patrick, S.L. George, J.T. Stapleton, *New England Journal of Medicine*, 345 (2001) 707-714.
- [8] W. Zhang, K. Chaloner, H.L. Tillmann, C.F. Williams, J.T. Stapleton, *Hiv Medicine*, 7 (2006) 173-180.
- [9] C.F. Williams, D. Klinzman, T.E. Yamashita, J.H. Xiang, P.M. Polgreen, *The New England journal of medicine*, 350 (2004) 981-990.
- [10] A.K. Van der Bij, N. Kloosterboer, M. Prins, B. Boeser-Nunnink, R.B. Geskus, J.M.A. Lange, R.A. Coutinho, H. Schuitemaker, *Journal of Infectious Diseases*, 191 (2005) 678-685.
- [11] U. Künkel, M. Höhne, T. Berg, U. Hopf, A.S. Kekulé, G. Frösner, G. Pauli, E. Schreier, *Journal of Hepatology*, 28 (1998) 978-984.
- [12] M. Tacke, K. Kiyosawa, K. Stark, V. Schlueter, B. OfenlochHaehnle, *Lancet*, 349 (1997) 318-320.
- [13] V. Reshetnyak, T. Karlovich, L. Ilchenko, *World Journal of Gastroenterology*, 14 (2008) 4725-4734.
- [14] D.L. Thomas, D. Vlahov, H.J. Alter, J.C. Hunt, R. Marshall, J. Astemborski, K.E. Nelson, *Journal of Infectious Diseases*, 177 (1998) 539-542.
- [15] P. Toniutto, C. Fabris, F. Barbone, D. Liani, S.G. Tisminetzky, T. Galai, G. Barillari, F. Biffoni, V. Gasparini, M. Pirisi, *Clin. Diagn. Lab. Immunol.*, 6 (1999) 573-576.
- [16] M.D. Berzsenyi, S.K. Roberts, *Journal of Infectious Diseases*, 194 (2006) 407-409.
- [17] E. Herrera, S. Tenckhoff, M. Gomara, R. Galatola, M. Bleda, C. Gil, *Journal of medicinal chemistry*, 53 (2010) 6054-6063.
- [18] M. Uttamchandani, S.Q. Yao, *Curr. Pharm. Design*, 14 (2008) 2428.
- [19] Y.M. Foong, *Curr. Opinion Chem. Biol.*, 16 (2012) 234-242.
- [20] H. Andresen, C. Grötzinger, K. Zarse, O.J. Kreuzer, E. Ehrentreich-Förster, F.F. Bier, *Proteomics*, 6 (2006) 1376-1384.
- [21] W. Hueber, B.A. Kidd, B.H. Tomooka, B.J. Lee, B. Bruce, B. Kidd, B. Tomooka, B. Lee, B. Bruce, J. Fries, G. Snderstrup, P. Monach, J. Drijfhout, W. van Venrooij, P. Utz, M. Genovese, W. Robinson, *Arthritis & rheumatism*, 52 (2005) 2645-2655.
- [22] Hecker et al., *Autoimmunity Reviews*, 11 (2012) 180-190.
- [23] F. Wilcoxon, *Biometrics Bulletin*, 1 (1945) 80-83.
- [24] W.H. Kruskal, *Journal of the American Statistical Association*, 52 (1957) 356-360.
- [25] Y. Benjamini, Y. Hochberg, *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1995) 289-300.
- [26] R.J. Simes, *Biometrika*, 73 (1986) 751-754.
- [27] Y. Benjamini, D. Yekutieli, *The Annals of Statistics*, 29 (2001) 1165-1188.
- [28] E.R. DeLong, D.M. DeLong, D.I. Clarkepearson, *Biometrics*, 44 (1988) 837-845.
- [29] J.A. Hanley, B.J. McNeil, *Radiology*, 143 (1982) 29-36.