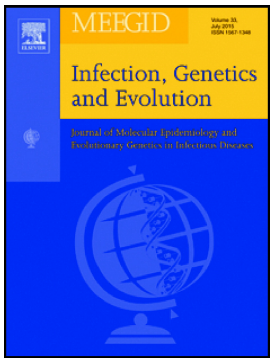
Journal Pre-proof



Bioinformatic analysis of post-transmission viral readaptation in Argentine patients with acute HIV-1 infection

G. Damilano, O. Sued, S. Satorres, M.J. Ruiz, Y. Ghiglione, F. Guzman, G. Turk, F. Quiroga, P. Cahn, H. Salomón, D. Dilernia

PII: S1567-1348(20)30039-3

DOI: https://doi.org/10.1016/j.meegid.2020.104207

Reference: MEEGID 104207

To appear in: *Infection, Genetics and Evolution*

Received date: 27 August 2019

Revised date: 18 January 2020

Accepted date: 22 January 2020

Please cite this article as: G. Damilano, O. Sued, S. Satorres, et al., Bioinformatic analysis of post-transmission viral readaptation in Argentine patients with acute HIV-1 infection, *Infection, Genetics and Evolution*(2020), https://doi.org/10.1016/j.meegid.2020.104207

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Journal Pre-proof

Affiliations, Conflict of Interest, and Authorship Conformation Form

Paper title: **Bioinformatic analysis of post-transmission viral readaptation in Argentine patients with acute HIV-1 infection**

G. Damilano[1], O. Sued[2], S. Satorres[3], M.J. Ruiz[4], Y. Ghiglione[1], F. Guzman[5], G. Turk[1], F. Quiroga[1], P. Cahn[6], H. Salomón[1], D. Dilernia[7].

1. CONICET-Universidad de Buenos Aires, Instituto de Investigaciones Biomédicas en Retrovirus y Sida (INBIRS), Buenos Aires, Argentina.
2. Fundación Huésped-Buenos Aires, Argentina.
3. Facultad de Química, Bioquímica y Farmacia. Universidad Nacional de San luis (UNSL), Argentina.
4. Currently, Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Quebec, Canada.
5. Núcleo de Biotecnología Curauma, Pontificia Universidad Católica de Valparaíso, Chile.
6. Hospital General de Agudos "Dr. JA Fernández" Buenos Aires, Argentina.
7. Emory Vaccine Center, Emory University, Atlanta, GA.

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Journal Pre-proof

Author Contribution section

G. Damilano, D. Dilernia: Concept, design, methodological development, data analysis and writing; O. Sued, P. Cahn, H. Salomón: Resource management, patient recruitment / monitoring, data analysis and review; M.J. Ruiz, Y. Ghiglione, F. Guzman, S. Satorres, G. Turk, F. Quiroga: Data analysis, review and edition.

Journal Pre-proof

Bioinformatic analysis of post-transmission viral readaptation in Argentine patients with acute HIV-1 infection

G. Damilano, O. Sued, S. Satorres, M.J. Ruiz, Y. Ghiglione, F. Guzman, G. Turk, F. Quiroga, P. Cahn, H. Salomón, D. Dilernia.

Abstract

During the acute phase of HIV-1 infection, a strong readaptation occurs in the viral population. Our objective was to analyze the post-transmission mutations associated with escape to the cytotoxic immune response and its relationship with the progression of the infection. In this study, a total of 17 patients were enrolled during acute/early primary HIV infection and 8 subjects that were the HIV positive partner resulting in 8 transmission pairs. Genotyping of the genetic polymorphisms of HLA class I A and B was performed using PCR-SSOP. Viral RNA extraction was from plasma. 570 single Gag-gene amplifications were obtained by limiting-dilution RT-PCR. Epitope prediction was performed with NetMHC CBS prediction server for the 19 HLA-A and -B alleles. Cytotoxic response prediction was performed by using the IEDB Analysis Resource. From our results, we deduce that the transmitted CTL / gag escape frequency in the founder virus was at least double compared to the post-transmission events. Additionally, by means of an algorithm that combines these frequencies, we observed that the founder viruses better adapted to the HLA A / B alleles of the recipient could contribute to a greater progression of the infection. Our results suggest that there is a large adaptation of HIV-1 to the HLA A / B alleles prevalent in our population. However, despite this adaptive advantage, the virus needs to make "readjustments" through new escape and compensatory mutations. Interestingly, according to our results, this readaptation could have a role in the progression of the infection.

1. Introduction

During acute HIV-1 infection, viral production is higher resulting in extremely high viral load (VL) during the first 4 to 6 weeks (Arnott et al. 2010; Girerd-Genessay et al. 2016). The lack of an effective specific immune response in this phase of infection contributes indirectly to an increase in viral fitness (Tomaras et al. 2008; Kramer et al. 2010; Cohen et al. 2011; Borrow 2011; Radebe et al. 2015; Radebe et al. 2011). However, in the presence of a strong specific immune response, above 30 to 60 days of infection, the virus must necessarily sacrifice replicative capacity in order to escape (Price et al. 1999; Song et al. 2012; Kloverpris, Leslie, and Goulder 2015; Sunshine et al. 2015). This coincides with a drop in viral load which establishes a viral set point strongly linked to the effectiveness of the immune response and clinical progression (Selhorst et al. 2017; Goodreau et al. 2018). While several investigations have described post-transmission viral dynamics (Brockman et al. 2010; Frange et al. 2013; Kloverpris, Leslie, and Goulder 2015; Yu et al. 2018), the adaptive role of escape mutations transmitted and/or acquired post-infection is not fully elucidated. Using computational methods, our first objective was to classify the different escape polymorphisms to the cytotoxic immune response in the viral gag protein in transmission events. The criterion for the selection of the gag protein in this study lies in that the immune response against this protein has been strongly associated with the control of viral replication by various studies(Walker and McMichael 2012; Turk et al. 2013; Radebe et al. 2015) Next, with these data, our second objective was to evaluate the dynamics of these mutations over time after infection, and mainly, to analyze how these escape variables observed in the founder (transmitted) virus can influence viral readaptation and the course of infection.

2. Methods

Study subjects: A total of 17 patients were enrolled during acute/early primary HIV infection (average 43.7 days post-infection, SD: 24,3) and 8 subjects that were the HIV positive partners (PP) resulting in 8 transmission pairs, **Table 1**. Two to three blood samples were taken during the first year of infection of the acute HIV individuals (AHI) and one blood sample from their PP (**Figure 1a**). Design: longitudinal prospective cohort study. AHI subjects were enrolled in the Grupo Argentino de Seroconversión (Socias et al. 2011) following an inclusion criteria: detection of HIV RNA or p24 antigen with a simultaneous negative or indeterminate screening (ELISA or particle agglutination test) or incomplete pattern of western blot assay. Clinical stage of primary infection subjects was determined according to the system of Fiebig et al. (Fiebig et al. 2003; Fiebig EW 2003). Infection days was estimated from clinical history and correlated with symptoms and reported exposures. The criteria for PP were defined as known HIV positive subjects with more than 100 days of infection and detectable viral load (VL> 50 copies of HIV RNA / ml plasma) and being reported by his/her partner as the source of infection (this was confirmed by phylogenetic analysis). Both AHI and PP were HAART *naïve* at the time of sampling.

Samples: Blood samples were collected from study participants at enrollment and following points and centrifuged to separate plasma, which was stored at -80°C until use. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque density gradient centrifugation (Amersham, Sweden) and cryopreserved. Plasma VL (branched-DNA, Versant HIV-1 RNA 3.0 assay; Siemens Healthcare) and CD4+ T-cell count (flow cytometry double platform, BD FACSCanto; BD Biosciences) were determined in all samples.

HLA typing: Genotyping of the genetic polymorphisms of HLA class I A and B was performed using PCR-SSOP (polymerase chain reaction-sequence specific oligonucleotide probe, Fujirebio INNO LiPA) (Oh SH 1993; Middleton D 1995), and was reported in medium and high resolution. For this technique, an amplification step was performed by the polymerase chain reaction (PCR) from PBMC DNA with biotinylated primers designed to encompass the polymorphic zone. The biotinylated amplified product was chemically denatured (strand separation) and hybridized to nitrocellulose strips containing specific immobilized oligonucleotide probes, complementary to the polymorphisms known in the alleles of our population (between 30 and 40 probes for HLA-I A and B). Each strip has two internal control bands (hybridization and conjugation controls). Once the process was finished, the interpretation of results for determining the possible combinations of alleles was done through tables (manual) and Liras® HLA software v6.00 (Fujirebio-Europe Company).

DNA/Amino acids sequence analysis: Viral RNA extraction was done from plasma and DNA from PBMC using commercial QIAamp viral extraction kit and QIAamp DNA mini kit (QIAGEN, Hilden, Germany), respectively. 567 single Gag-gene amplifications were obtained by limiting-dilution RT-PCR from plasma samples and 3 gag sequences from proviral DNA, obtaining an average of 12 sequences per patient. For first-strand DNA synthesis 1rdgag-Reverse primer, SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific) was used whit 10µl of RNA sample in a final volume of reaction of 20µl. Then, the cDNA was diluted successively until obtaining the limit dilution. 1:5 to 1:100 dilutions were required to obtain single-gag sequence amplifications. Amplification was conducted under the following conditions: for reverse transcription 5 min at 65°C (1º

RdgagRev/RNA), 50 min at 42°C, 15 min at 70°C (1º RdgagRev/RNA/SSIII), one cycle. The conditions of the nested PCR were: Platinum Taq polymerase HiFi (Thermo Fisher Scientific) was used under appropriate reaction conditions: 3 min at 95°C one cycle; 15 sec at 95°C, 15 sec at 56°C, 1 min 40 sec at 72°C for 5 cycles;15 sec at 90°C, 15 sec at 56°C, 1 min 40 sec at 72°C for 30 cycles and one cycle at 72°C for 10 min. Primers: 1ºrd gagFw 5'-CTAGCAGTGGCGCCCGAACAGG-3', 1ºrd gagRev 5'-CAGTCTTTCATTTGGTGTCCTTC -3', 2ºrd gagFw5'-TCTCTCGACGCAGGACTCG-3', 2ºrd gagRev 5'-TTTCCACATTTCCAACAGCCC-3'.

Sanger sequencing and NGS SMRT technology (Pacific Biosciences) were implemented to sequence the amplicons, as described previously (Dilernia D 2015). The chromatograms obtained from the sequencing of the viral quasispecies were analyzed with Sequencher 4.10 (www.genecodes.com). All HIV-1 gag aminoacid sequences obtained in this study were submitted to GenBank under accession numbers: (MK479303 - MK479872).

Phylogenetic analysis: Following alignment, manual adjustments were made using BioEdit software, taking into consideration protein-coding sequences. HIV-1 genotypes were determined through SimPlot v3.5.1 (Lole et al. 1999). For phylogenetic analysis, MEGA 10.0.4 was used (Kumar et al. 2018), with which the phylogenetic reconstruction was performed for the acute and chronic viral quasispecies amino acid sequences as a whole. Maximum-likelihood trees were generated using a GTR Gamma model, **Figure 1b**. Bootstrap trees were produced using the rapid bootstrapping algorithm and 1000 bootstrap replicates. Clusters with bootstrap values higher than 0.85 (85%) were defined as a phylogenetic cluster. For the modeling of trees, FigTree v1.4.3 was used. Phylogenetic distances were calculated from the nucleotide sequences by constructing a distance matrix using the Poisson model, without considering deletions and insertions (Tajima F. 1984; Zuckerkandl E. 1965; Tamura K. 2013). The founder viruses were determined from the lower phylogenetic distance (nucleotide substitutions per site) compared to the PP or Population consensus sequence (shown below). Subtype reference sequences A, B, C, D and F were used to root (Sequence data were obtained from: www.hiv.lanl.gov/components/sequence/HIV/search).

Immune prediction analysis: Epitope prediction was performed with NetMHC CBS prediction server (http://www.cbs.dtu.dk/services/NetMHC/) for the 19 HLA-A and -B alleles most prevalent in our population. Peptides with a length from 8 to 14 mer and an affinity cut-off of 400nM were considered according to the method described by *Nielsen and Andreatta et. al*. (Nielsen et al. 2003; Andreatta and Nielsen 2016) Cytotoxic response prediction was performed by using the IEDB Analysis Resource (http://tools.iedb.org). Based on the reference (Calis et al. 2013), a difference greater than 0.11 in the affinity score for the TCR molecule was considered significant and associated with escape.

Transmitted escape mutations to recipient alleles were determined by a comparison between founder virus in the AHI with circulating consensus in our population (Ancestral population consensus) obtained during the period 1987 to 2006 by our working group (Dilernia et al. 2008). Ten consensus sequences were constructed with Consensus Maker from Los Alamos HIV Database tools. Those polymorphisms associated with HLAI / TCR escape present in the founder virus and not present in the population gag consensus were considered as escape mutations prior to the transmission event (transmitted

mutations). All epitope/mutation associations were validated as a significant non-random relationship from the Poisson equation with $p < 0.01$. In addition, the greater proportion of known epitopes within the predicted epitope with escape mutation (71.5%, *Chi*-square p<0.01) provides additional support for this relationship.

Strategy for viral readaptation analysis: we included the escape categories evaluated in each recipient that we consider of greatest impact on the progression of the infection. The resulting equation, which describes the level of viral readaptation to the new host from the results obtained, is:

$$Rh(g) = \frac{\Sigma fr\ TE(hlaR)}{\Sigma fr\ EE(hlaR)\ x\ Er(d)}\ x\ c$$

Where:

$\boldsymbol{Rh}$: viral readaptation score in gag protein (g) towards the HLA-I (h) alleles of the recipient. $\boldsymbol{fr\ TE}$: relative frequency of transmitted escape mutations for recipient alleles (*hlaR*). $\boldsymbol{frEE}$: relative frequency of post-transmission escape to recipient alleles. $\boldsymbol{Er}$: rate of HLA-I / TCR post-transmission escape mutations according to days of infection monitored ($\boldsymbol{d}$). $\boldsymbol{c}$: coefficient used to obtain values from 0 to 1. **Note:** DR was not considered in the equation because it was not possible to recruit PP for all transmission events.

Ethics Statement: The study was reviewed and approved by two institutional review boards (IRB): Comité de Ética Humana, Facultad de Medicina, Universidad de Buenos Aires, and Comité de Bioética, Fundación Huésped (Buenos Aires, Argentina). All participants provided written informed consent accepting to participate in this study before any procedure. This research has been carried out in accordance with the guidelines of Good Clinical Practices and the Declaration of the Helsinki World Medical Association.

3. Results

After the process of optimization of the viral sequences, a phylogenetic analysis was carried out in which the transmissions were checked from the PP enrolled and a general comparative vision of diversity was obtained in the sequenced viral populations (**Figure 1b**). The nucleotide sequences were subjected to variability analysis (Entropy) and amino acid sequences epitope prediction was performed for HLA-I (locus A / B) and affinity prediction by the TCR molecule, both associated with a specific cytotoxic immune response. The data obtained from the computational prediction of affinity to HLA-I / TCR allowed defining immuno-dominant segments along with the gag protein. In addition, the escape to the cytotoxic immune response against the possible epitopes contained in these segments was evaluated, and, by combining these variables through an algorithm, we evaluated their possible impact on the progression of the infection.

3.1. The founder virus contains a high frequency of escape mutations

The classification resulting from the prediction analysis based on the presence of escape mutations was that 28% (178) did not present mutations (conserved epitopes), 59% (373) presented escape associated mutations, defined as a significant decrease in the prediction of affinity peptide-HLA-I / TCR, and 13% (82) presented other mutations (non-escape mutations) (see methods and **Figure 2a**). In the group with escape-associated mutations, the resulting distribution of these mutations were: i) based on the AHI HLA-I alleles,

64.2% (240) were transmitted HLA-I/TCR mutations (TE), and 30.6% (114) were HLA-I / TCR escape mutations that arose during post-infection time (EE), and 5.1%(19) were reversals mutations to AHI alleles (RC), **Figure 2b**; ii) based on the PP HLA-I alleles, 89% (116) of escape mutations were transmitted and remained without revert in the recipient even in the absence of the allele, and 11% (15) was reversed in the recipient in the absence of allele (DR), **Figure 2c**. The frequency of escape mutations associated with HLA-I was significantly higher (64.9%) than those associated with an escape to the TCR molecule (35.1%), p <0,01. The distribution of escape mutations along the gag protein is presented in **Figure 2c**.

In addition, we found that the proportion of known epitopes (*www.hiv.lanl.gov/content/immunology*) was higher in epitopes associated with escape (267/373, 71.5%) compared to those not associated with escape (77/229, 33.6%), Chi-square p <0.01, **Table 2**.

3.2. The ability of viral readaptation to the new host could be associated with the progression of infection

Assuming that the virus is capable of accumulating adaptive mutations, (particularly escape mutations) through successive transmissions, we evaluated the association between different escape categories with monitoring laboratory parameters (viral load and CD4+ T lymphocyte count).

The resulting correlation between the values of *Rh* (see methods) obtained for each receptor with the corresponding values of post-acute monitoring of VL and TCD4+ lymphocyte count (viral set point reference) is shown in **Table 3** and **Figure 3 a** and **b**. In the case of recipients with more than two monitoring points (R1, R4, F1, F3, F5, and F7), the average values during the beginning of the chronic stage (2nd and 3rd sample) were taken as reference. From these results, we deduced that a higher level of escape (TE) contributes slightly to a lower number of epitopes associated with an immune response which leads to a reduction in the rate of escape mutations (*Er*) during the passage towards the chronic phase. These results suggest that a higher level of pre-adaptation in the founder virus could favor, but not ensure, its re-adaptation (*Rh*) to new immunological HLA-I context, with a lower cost in viral fitness. This translates in higher VL values and lowers TCD4 + counts, which correlates with greater progression of infection (Selhorst et al. 2017; Goodreau et al. 2018), (Spearman, *p<0,01*). Additionally, the reliability of the viral readaptation algorithm was verified. This was done considering only the epitopes predicted and described in the literature associated with immune response in the gag protein (known epitopes, see supplementary material). The results showed that the correlation with the monitoring parameters (VL and T CD4 +) is maintained (Spearman, *p<0.01*), which suggests that both the prediction parameters and the equation used have a reasonable immunological correlation with that described in the literature (**Figure 3c**).

3.3. The frequency of escape could be influenced by the frequency of the allele in the population

Through a comparative analysis between TE / EE frequencies, we conclude that there is a significant positive correlation between both variables. Interestingly, this tendency is accentuated especially for the most prevalent alleles (frequency≥13%) in our population. Although the same was observed for both escape categories, our results show a more significant correlation for TE (*p=0,0013*). However, we also observed that the frequencies of EE showed a positive correlation with allelic frequencies although significantly less marked (*p=0,0305)*. Non-parametric Spearman tests were used (**Figure 4a)**. Interestingly, although higher levels of Rh score could be expected for patients with more frequent alleles, this was not the case (**Figure 4b**), which highlights the impact of EE polymorphisms on rehabilitation.

3.4. Gag segments associated with escape presented more significant entropy variation

Once we identified the escape mutations during the transmission event (EE), we performed a variability analysis through the measurement of entropy in nucleotide sequences of gag clones as a function of time (days post-infection). For this, Shannon entropy tool and HIV_SNAP (www.hiv.lanl.gov) were used for determining the increase in variability and the nonsynonymous and synonymous nucleotide substitution rates (dN / dS) respectively, in regions associated with post transmission escape in comparison with the rest of the gag sequence. As expected, the dN / dS ratio between both regions was not significant (p> 0.1) with an average of dN / dS = 3.66 (78.5% non-synonymous substitutions). A uniform distribution of synonymous mutations was found throughout the gag gene, which supports the association between evaluated amino acid mutations and escapes to immune response.

For the analysis of variability, the level of entropy was measured within segments of 42 nucleotides (corresponding to the peptide of greater length evaluated in the prediction of epitopes, 14 mer) upstream and downstream of each post transmission mutation. Thus, comparing the different sampling intervals in the patients analyzed, we were able to compare the increase in variability between associated regions and those not associated with immune escape. Our results show that those segments associated with escape (epitopes) present a more rapid increase in post-infection variability (**Figure 5**).

Moreover, considering only 1 epitope/anchor point / HLA-I allele we found that the sub protein p24 showed the lowest frequency of epitopes associated with escape (average 21%) in comparison with Nucleocapside (NC, average 62.7%) and Matrix (MA average 39.2%), t-test p<0.01. Our results coincide with previous studies in which a positive association between variability and presence of epitopes with immunological relevance was established (Bansal et al. 2005; Ferrari et al. 2011).

4. Discussion.

Our main objective was to identify, through computational analysis, the peptide sequences most associated with cytotoxic immune response along the gag protein.

In the first place, we observed that the majority (64.2%) of epitopes contained escape mutations to HLA-I and TCR that were already present at the founder virus during transmission (TE) and did not arise afterward as a result of the escape to HLA-I recipient's alleles, which coincides with previous studies(Kloverpris, Leslie, and Goulder 2015; Katoh et al. 2016). On the other hand, we saw in transmission pairs, that a minority (11%) of the escape mutations that were present before the transmission reverted in the recipient (DR), in the absence of the HLA-I allele from which they escaped. It could be expected in this case that the virus reverses that mutation, assuming that said mutation reduces its efficiency of replication when it is no longer exposed to the HLA-I allele from which it escaped since it is known that the virus, with certain mutations, sacrifices greater replication capacity in exchange of escaping from the immune response (Troyer et al. 2009; Du et al. 2017). However, it is deduced from our results that many of these mutations do not have a negative impact on viral replicative capacity since we did do not observe a high reversion rate. These non-reverting mutations, in the case of persisting, will help in evading the immune response in a new host that presents the alleles and indicates that the escape mutations have the capacity to establish themselves in the viral population, if necessary. It is interesting to note that, 12/19 (63.2%) of the reversals to AHI allele (RC) were manifested with an escape to TCR, which could be evaluated as compensatory mutations (Gijsbers et al. 2013; Nagaraja et al. 2016), such analysis is beyond the scope of this research.

On the other hand, we saw that the proportion of epitopes associated with escape mutations generated in the event of transmission (EE) in the presence of HLA-I alleles of the recipient, was relatively low (30.6%, 20,9% HLA escape polymorphism and 9,7% TCR escape polymorphism). This low proportion (compared to TE) of HLA-I/TCR associated mutations generated in the event is probably because the virus already contained escape mutations that contributed to partial escape but being of great importance during the first days of infection. The post-transmission escape frequency obtained was slightly higher than previous studies(Brumme et al. 2008; Monaco et al. 2016), and it suggests that in at least 1 in 5 epitopes present in the founding virus, in which an escape prior to transmission was not established, an HLA escape was selected and almost 1 in 10 epitopes an escape mutation to TCR was selected. The reason and the role of the fact that the highest escape frequency is observed towards the HLA molecule are not entirely clear(Bronke et al. 2013).

Following the line of analysis, we combined the frequencies of the different escape categories evaluated by means of an equation that summarizes the level of viral readaptation (*Rh*) for each transmission event.

In recent years, a general consensus has been established about how the polymorphisms associated with escape (mainly HLA) can accumulate over the years and establish in the viral population(Moore et al. 2002; Dilernia et al. 2008; Kawashima et al. 2009; Cotton et al. 2014), and above all, how this variable can negatively influence the course of the infection(McLaren et al. 2015; Sunshine et al. 2015; Monaco et al. 2016). Using the Rh

equation, despite the limited number of patients, we were able to estimate a trend of greater or lesser progression of the infection. Surprisingly, the results of the Rh score show us that the increase in viral pre-adaptation (ET) does not necessarily correlate with greater progression in AHI. Taking into account that the virus tends to accumulate mutations with the lowest cost of replicative capacity(Kloverpris, Leslie, and Goulder 2015) (represented by TE in our equation), the measurement of EE becomes relevant, because, if there is no specific immune response, most of these would not have been selected, which indicates a more significant statistical association. We conclude from our data that, the mere accumulation of escape does not ensure viral readaptation, and on the other hand, the measurement of EE during the first months of infection is a representative marker in the assessment of progression.

We also saw that the frequency of TE and EE correlates positively with the frequency of HLA A / B alleles in our population. Although this correlation is significant in both cases, our results suggest a greater association of TE. We believe that this difference is due to the fact that EE are more prone to reversal or compensatory mutations because they still do not represent established mutations in the viral population, compared to TE, which largely represents the level of viral adaptation, in addition, it showed a relatively low reversal rate (DR, 11%). Future analysis will be necessary to challenge that hypothesis. In relation to the Rh equation, the slight but significant increase in the frequency of EE for more prevalent alleles resulted in slightly lower values of Rh score in patients with more frequent alleles as seen in Figure 4b. We interpret this result could be due to an additive effect in the reduction of the replicative capacity caused by TE and EE. While the data is limited, this result is consistent with studies describing the loss of protective impact of the prevalent HLA(Cotton et al. 2014; Kloverpris, Leslie, and Goulder 2015).

Finally, we saw how EE is associated with segments with a greater increase in variability as a function of time, which suggests that there is a selection of certain mutations in immunodominant segments with the ability to reduce subtly and gradually the HLA-I or TCR affinity, and finally facilitates the selection of the escape mutation itself. However, more studies are required to verify this assumption. Additionally, we saw that both the escape frequencies and the associated epitopes were directly proportional to the intrinsic variability of the protein subunits, with the capsid protein (CA, p24) being the most conserved and the one with the lowest frequency of escape (Chakraborty, Rahman, and Chakravorty 2014).

It should be noted that not only the recognition and presentation of the epitope by HLA-I is enough to generate an effective immune response since other factors are involved as well (Genesca 2011; Walker and McMichael 2012; Turk et al. 2013). We must highlight the fact that we did not analyze *in vitro* immune response, and although the prediction algorithms used have been tested and validated by numerous scientific studies(Nielsen et al. 2003; Andreatta and Nielsen 2016), it will be necessary to design *in vitro* immune response assays for studying the most representative escape mutations observed.

However, a high percentage of the predicted epitopes associated with escape were previously reported (Table 2).

5. Conclusion and perspectives

Although the use of predictors to study cytotoxic immune response can generate an excessive number of peptide sequences without immunological importance *in vivo*, with the use of appropriate algorithms it is possible to reveal a large part of the epitopes with possible immunological relevance. The bioinformatic study of regions with reduced evasion capacity in viral proteins could contribute to design an effective vaccine including immunological targets that are more "stable" over time.

6. Author Contribution

G. Damilano, D. Dilernia: Concept, design, methodological development, data analysis and writing; O. Sued, P. Cahn, H. Salomón: Resource management, patient recruitment/monitoring, data analysis and review; M.J. Ruiz, Y. Ghiglione, F. Guzman, S. Satorres, G. Turk, F. Quiroga: Data analysis, review and edition.

7. Acknowledgments

We thank all the patients from the Grupo Argentino de Seroconversión for their participation in this research and to the physicians from Fundación Huésped for their clinical contribution. Also, thanks to Paul Kerston for assistance during the edition of the manuscript. We thank all INBIRS (Universidad de Buenos Aires) colleagues for their interest and direct or indirect collaboration in this research.

This research was funded by grants from the Agencia Nacional de Promoción Científica y Tecnológica (PICT 2008-00771 and PICT 2011-0271) to H. Salomon.

We thank the Universidad Nacional de San Luis (UNSL-FQByF) for their support and collaboration.

We have no conflict of interest to testify.

8. References:

Andreatta, M., and M. Nielsen. 2016. 'Gapped sequence alignment using artificial neural networks: application to the MHC class I system', *Bioinformatics*, 32: 511-7.

Arnott, A., D. Jardine, K. Wilson, P. R. Gorry, K. Merlin, P. Grey, M. G. Law, E. M. Dax, A. D. Kelleher, D. E. Smith, and D. A. McPhee. 2010. 'High viral fitness during acute HIV-1 infection', *PLoS One*, 5.

Bansal, A., E. Gough, S. Sabbaj, D. Ritter, K. Yusim, G. Sfakianos, G. Aldrovandi, R. A. Kaslow, C. M. Wilson, M. J. Mulligan, J. M. Kilby, and P. A. Goepfert. 2005. 'CD8 T-cell responses in early HIV-1 infection are skewed towards high entropy peptides', *Aids*, 19: 241-50.

Borrow, P. 2011. 'Innate immunity in acute HIV-1 infection', *Curr Opin HIV AIDS*, 6: 353-63.

Brockman, M. A., Z. L. Brumme, C. J. Brumme, T. Miura, J. Sela, P. C. Rosato, C. M. Kadie, J. M. Carlson, T. J. Markle, H. Streeck, A. D. Kelleher, M. Markowitz, H. Jessen, E. Rosenberg,

M. Altfeld, P. R. Harrigan, D. Heckerman, B. D. Walker, and T. M. Allen. 2010. 'Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection', *J Virol*, 84: 11937-49.

Bronke, C., C. A. Almeida, E. McKinnon, S. G. Roberts, N. M. Keane, A. Chopra, J. M. Carlson, D. Heckerman, S. Mallal, and M. John. 2013. 'HIV escape mutations occur preferentially at HLA-binding sites of CD8 T-cell epitopes', *Aids*, 27: 899-905.

Brumme, Z. L., C. J. Brumme, J. Carlson, H. Streeck, M. John, Q. Eichbaum, B. L. Block, B. Baker, C. Kadie, M. Markowitz, H. Jessen, A. D. Kelleher, E. Rosenberg, J. Kaldor, Y. Yuki, M. Carrington, T. M. Allen, S. Mallal, M. Altfeld, D. Heckerman, and B. D. Walker. 2008. 'Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection', *J Virol*, 82: 9216-27.

Calis, J. J., M. Maybeno, J. A. Greenbaum, D. Weiskopf, A. D. De Silva, A. Sette, C. Kesmir, and B. Peters. 2013. 'Properties of MHC class I presented peptides that enhance immunogenicity', *PLoS Comput Biol*, 9: e1003266.

Chakraborty, S., T. Rahman, and R. Chakravorty. 2014. 'Characterization of the Protective HIV-1 CTL Epitopes and the Corresponding HLA Class I Alleles: A Step towards Designing CTL Based HIV-1 Vaccine', *Adv Virol*: 321974.

Cohen, M. S., G. M. Shaw, A. J. McMichael, and B. F. Haynes. 2011. 'Acute HIV-1 Infection', *N Engl J Med*, 364: 1943-54.

Cotton, L. A., X. T. Kuang, A. Q. Le, J. M. Carlson, B. Chan, D. R. Chopera, C. J. Brumme, T. J. Markle, E. Martin, A. Shahid, G. Anmole, P. Mwimanzi, P. Nassab, K. A. Penney, M. A. Rahman, M. J. Milloy, M. T. Schechter, M. Markowitz, M. Carrington, B. D. Walker, T. Wagner, S. Buchbinder, J. Fuchs, B. Kaplin, K. H. Mayer, P. R. Harrigan, M. A. Brockman, A. F. Poon, and Z. L. Brumme. 2014. 'Genotypic and functional impact of HIV-1 adaptation to its host population during the North American epidemic', *PLoS Genet*, 10: e1004295.

Dilernia D, Chien JT, Monaco DC, Brown MP, Ende Z, Deymier MJ, Yue L, Paxinos EE, Allen S, Tirado-Ramos A, Hunter E. 2015. 'Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing.', *Nucleic Acids Res. Nucleic Acids Res. 2015 Nov 16; 43(20): e129.* .

Dilernia, D. A., L. Jones, S. Rodriguez, G. Turk, A. E. Rubio, S. Pampuro, M. Gomez-Carrillo, C. T. Bautista, G. Derqui, J. Benetucci, M. B. Lasala, L. Lourtau, M. H. Losso, H. Perez, P. Cahn, and H. Salomon. 2008. 'HLA-driven convergence of HIV-1 viral subtypes B and F toward the adaptation to immune responses in human populations', *PLoS One*, 3: e3429.

Du, Y., T. H. Zhang, L. Dai, X. Zheng, A. M. Gorin, J. Oishi, T. T. Wu, J. M. Yoshizawa, X. Li, O. O. Yang, O. Martinez-Maza, R. Detels, and R. Sun. 2017. 'Effects of Mutations on Replicative Fitness and Major Histocompatibility Complex Class I Binding Affinity Are Among the Determinants Underlying Cytotoxic-T-Lymphocyte Escape of HIV-1 Gag Epitopes', *MBio*, 8.

Ferrari, G., B. Korber, N. Goonetilleke, M. K. Liu, E. L. Turnbull, J. F. Salazar-Gonzalez, N. Hawkins, S. Self, S. Watson, M. R. Betts, C. Gay, K. McGhee, P. Pellegrino, I. Williams, G. D. Tomaras, B. F. Haynes, C. M. Gray, P. Borrow, M. Roederer, A. J. McMichael, and K. J. Weinhold. 2011. 'Relationship between functional profile of HIV-1 specific CD8 T cells and epitope variability with the selection of escape mutants in acute HIV-1 infection', *PLoS Pathog*, 7: e1001273.

Fiebig, E. W., D. J. Wright, B. D. Rawal, P. E. Garrett, R. T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S. H. Kleinman, and M. P. Busch. 2003. 'Dynamics of

HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection', *Aids*, 17: 1871-9.

Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH, Busch MP. 2003. 'Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection'.

Frange, P., L. Meyer, M. Jung, C. Goujard, D. Zucman, S. Abel, P. Hochedez, M. Gousset, O. Gascuel, C. Rouzioux, and M. L. Chaix. 2013. 'Sexually-transmitted/founder HIV-1 cannot be directly predicted from plasma or PBMC-derived viral quasispecies in the transmitting partner', *PLoS One*, 8: e69144.

Genesca, M. 2011. 'Characterization of an effective CTL response against HIV and SIV infections', *J Biomed Biotechnol*: 103924.

Gijsbers, E. F., K. A. Feenstra, A. C. van Nuenen, M. Navis, J. Heringa, H. Schuitemaker, and N. A. Kootstra. 2013. 'HIV-1 replication fitness of HLA-B*57/58:01 CTL escape variants is restored by the accumulation of compensatory mutations in gag', *PLoS One*, 8: e81235.

Girerd-Genessay, I., D. Baratin, T. Ferry, C. Chidiac, V. Ronin, and F. Vanhems. 2016. 'Higher HIV RNA Viral Load in Recent Patients with Symptomatic Acute HIV Infection in Lyon University Hospitals', *PLoS One*, 11: e0146978.

Goodreau, S. M., S. E. Stansfield, J. T. Murphy, K. C. Peebles, G. S. Gottlieb, N. F. Abernethy, J. T. Herbeck, and J. E. Mittler. 2018. 'Relational concurrency, stages of infection, and the evolution of HIV set point viral load', *Virus Evol*, 4: vey032.

Katoh, J., A. Kawana-Tachikawa, A. Shimizu, D. Zhu, C. Han, H. Nakamura, M. Koga, T. Kikuchi, E. Adachi, T. Koibuchi, G. F. Gao, Z. L. Brumme, and A. Iwamoto. 2016. 'Rapid HIV-1 Disease Progression in Individuals Infected with a Virus Adapted to Its Host Population', *PLoS One*, 11: e0150397.

Kawashima, Y., K. Pfafferott, J. Frater, P. Matthews, R. Payne, M. Addo, H. Gatanaga, M. Fujiwara, A. Hachiya, H. Koizumi, N. Kuse, S. Oka, A. Duda, A. Prendergast, H. Crawford, A. Leslie, Z. Brumme, C. Brumme, T. Allen, C. Brander, R. Kaslow, J. Tang, E. Hunter, S. Allen, J. Mulenga, S. Branch, T. Roach, M. John, S. Mallal, A. Ogwu, R. Shapiro, J. G. Prado, S. Fidler, J. Weber, O. G. Pybus, P. Klenerman, T. Ndung'u, R. Phillips, D. Heckerman, P. R. Harrigan, B. D. Walker, M. Takiguchi, and P. Goulder. 2009. 'Adaptation of HIV-1 to human leukocyte antigen class I', *Nature*, 458: 641-5.

Kloverpris, H. N., A. Leslie, and P. Goulder. 2015. 'Role of HLA Adaptation in HIV Evolution', *Front Immunol*, 6: 665.

Kramer, H. B., K. J. Lavender, L. Qin, A. R. Stacey, M. K. Liu, K. di Gleria, A. Simmons, N. Gasper-Smith, B. F. Haynes, A. J. McMichael, P. Borrow, and B. M. Kessler. 2010. 'Elevation of intact and proteolytic fragments of acute phase proteins constitutes the earliest systemic antiviral response in HIV-1 infection', *PLoS Pathog*, 6: e1000893.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Mol Biol Evol*, 35: 1547-49.

Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray. 1999. 'Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination', *J Virol*, 73: 152-60.

McLaren, P. J., C. Coulonges, I. Bartha, T. L. Lenz, A. J. Deutsch, A. Bashirova, S. Buchbinder, M. N. Carrington, A. Cossarizza, J. Dalmau, A. De Luca, J. J. Goedert, D. Gurdasani, D. W. Haas, J. T. Herbeck, E. O. Johnson, G. D. Kirk, O. Lambotte, M. Luo, S. Mallal, D. van Manen, J. Martinez-Picado, L. Meyer, J. M. Miro, J. I. Mullins, N. Obel, G. Poli, M. S.

Sandhu, H. Schuitemaker, P. R. Shea, I. Theodorou, B. D. Walker, A. C. Weintrob, C. A. Winkler, S. M. Wolinsky, S. Raychaudhuri, D. B. Goldstein, A. Telenti, P. I. de Bakker, J. F. Zagury, and J. Fellay. 2015. 'Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load', *Proc Natl Acad Sci U S A*, 112: 14658-63.

Middleton D, Williams F, Cullen C, Mallon E. 1995. 'Modification of an HLA-B PCR-SSOP typing system leading to improved allele determination', *Tissue Antigens 45:232*.

Monaco, D. C., D. A. Dilernia, A. Fiore-Gartland, T. Yu, J. L. Prince, K. K. Dennis, K. Qin, M. Schaefer, D. T. Claiborne, W. Kilembe, J. Tang, M. A. Price, P. Farmer, J. Gilmour, A. Bansal, S. Allen, P. Goepfert, and E. Hunter. 2016. 'Balance between transmitted HLA preadapted and nonassociated polymorphisms is a major determinant of HIV-1 disease progression', *J Exp Med*, 213: 2049-63.

Moore, C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. A. Mallal. 2002. 'Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level', *Science*, 296: 1439-43.

Nagaraja, P., H. K. Alexander, S. Bonhoeffer, and N. M. Dixit. 2016. 'Influence of recombination on acquisition and reversion of immune escape and compensatory mutations in HIV-1', *Epidemics*, 14: 11-25.

Nielsen, M., C. Lundegaard, P. Worning, S. L. Lauemoller, K. Lamberth, S. Buus, S. Brunak, and O. Lund. 2003. 'Reliable prediction of T-cell epitopes using neural networks with novel sequence representations', *Protein Sci*, 12: 1007-17.

Oh SH, Fleischhauer K, Yang SY. 1993. 'Isoelectric focusing subtypes of HLA-A can be defined by oligonucleotide typing.', *Tissue Antigens 41:135*.

Price, D. A., A. O'Callaghan C, J. A. Whelan, P. J. Easterbrook, and R. E. Phillips. 1999. 'Cytotoxic T lymphocytes and viral evolution in primary HIV-1 infection', *Clin Sci (Lond)*, 97: 707-18.

Radebe, M., K. Gounder, M. Mokgoro, Z. M. Ndhlovu, Z. Mncube, L. Mkhize, M. van der Stok, M. Jaggernath, B. D. Walker, and T. Ndung'u. 2015. 'Broad and persistent Gag-specific CD8+ T-cell responses are associated with viral control but rarely drive viral escape during primary HIV-1 infection', *Aids*, 29: 23-33.

Radebe, M., K. Nair, F. Chonco, K. Bishop, J. K. Wright, M. van der Stok, I. V. Bassett, Z. Mncube, M. Altfeld, B. D. Walker, and T. Ndung'u. 2011. 'Limited immunogenicity of HIV CD8+ T-cell epitopes in acute Clade C virus infection', *J Infect Dis*, 204: 768-76.

Selhorst, P., C. Combrinck, N. Ndabambi, S. D. Ismail, M. R. Abrahams, M. Lacerda, N. Samsunder, N. Garrett, Q. Abdool Karim, S. S. Abdool Karim, and C. Williamson. 2017. 'Replication Capacity of Viruses from Acute Infection Drives HIV-1 Disease Progression', *J Virol*, 91.

Socias, M. E., O. Sued, N. Laufer, M. E. Lazaro, H. Mingrone, D. Pryluka, C. Remondegui, M. I. Figueroa, C. Cesar, A. Gun, G. Turk, M. B. Bouzas, R. Kavasery, A. Krolewiecki, H. Perez, H. Salomon, and P. Cahn. 2011. 'Acute retroviral syndrome and high baseline viral load are predictors of rapid HIV progression among untreated Argentinean seroconverters', *J Int AIDS Soc*, 14: 40.

Song, H., J. W. Pavlicek, F. Cai, T. Bhattacharya, H. Li, S. S. Iyer, K. J. Bar, J. M. Decker, N. Goonetilleke, M. K. Liu, A. Berg, B. Hora, M. S. Drinker, J. Eudailey, J. Pickeral, M. A. Moody, G. Ferrari, A. McMichael, A. S. Perelson, G. M. Shaw, B. H. Hahn, B. F. Haynes, and F. Gao. 2012. 'Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome', *Retrovirology*, 9: 89.

Sunshine, J. E., B. B. Larsen, B. Maust, E. Casey, W. Deng, L. Chen, D. H. Westfall, M. Kim, H. Zhao, S. Ghorai, E. Lanxon-Cookson, M. Rolland, A. C. Collier, J. Maenza, J. I. Mullins, and

N. Frahm. 2015. 'Fitness-Balanced Escape Determines Resolution of Dynamic Founder Virus Escape Processes in HIV-1 Infection', *J Virol*, 89: 10303-18.

Tajima F., Nei M. 1984. 'Estimation of evolutionary distance between nucleotide sequences', *Molecular Biology and Evolution 1:269-285.*

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. 2013. 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0', *Molecular Biology and Evolution30: 2725-2729.*

Tomaras, G. D., N. L. Yates, P. Liu, L. Qin, G. G. Fouda, L. L. Chavez, A. C. Decamp, R. J. Parks, V. C. Ashley, J. T. Lucas, M. Cohen, J. Eron, C. B. Hicks, H. X. Liao, S. G. Self, G. Landucci, D. N. Forthal, K. J. Weinhold, B. F. Keele, B. H. Hahn, M. L. Greenberg, L. Morris, S. S. Karim, W. A. Blattner, D. C. Montefiori, G. M. Shaw, A. S. Perelson, and B. F. Haynes. 2008. 'Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia', *J Virol*, 82: 12449-63.

Troyer, R. M., J. McNevin, Y. Liu, S. C. Zhang, R. W. Krizan, A. Abraha, M. M. Tebit, H. Zhao, S. Avila, M. A. Lobritz, M. J. McElrath, S. Le Gall, J. I. Mullins, and E. J. Arts. 2009. 'Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response', *PLoS Pathog*, 5: e1000365.

Turk, G., Y. Ghiglione, J. Falivene, M. E. Socias, N. Laufer, R. S. Coloccini, A. M. Rodriguez, M. J. Ruiz, M. A. Pando, L. D. Giavedoni, P. Cahn, O. Sued, H. Salomon, and M. M. Gherardi. 2013. 'Early Gag immunodominance of the HIV-specific T-cell response during acute/early infection is associated with higher CD8+ T-cell antiviral activity and correlates with preservation of the CD4+ T cell compartment', *J Virol*, 87: 7445-62.

Walker, B., and A. McMichael. 2012. 'The T-cell response to HIV', *Cold Spring Harb Perspect Med*, 2.

Yu, F., Y. Wen, J. Wang, Y. Gong, K. Feng, R. Ye, Y. Jiang, Q. Zhao, P. Pan, H. Wu, S. Duan, B. Su, and M. Qiu. 2018. 'The Transmission and Evolution of HIV-1 Quasispecies within One Couple: a Follow-up Study Based on Next-Generation Sequencing', *Sci Rep*, 8: 1404.

Zuckerkandl E, Pauling L. 1965. 'Evolutionary divergence and convergence in proteins', *Academic Press, New York, 1965. Pages 97–166.* .

Journal Pre-proof

Case	Sample ID	Viral Load (copias/ml)	T CD4 (células/ul)	Days of infection	HLA A	HLA B	Viral subtype	*gag* clones	Year of enrollment
TC1	D1s1	19250	813	>100	A02; A68	B15CXH; B14	B	15	2009
	R1s1	1259	410	26	A68; A31	B44CXC; B51	B	10	
	R1s2	316	710	65			B	10	
	R1s3	112	743	230			B	10	
TC2	D2s1	25599	818	>100	A31; X	B1508(75); B35	B	10	2009
	R2s1	10706	1021	81	A01; A02	B07; B08	B	10	
	R2s2	3981	862	145			B	10	
TC3	D3s1	10673	1111	>100	A33; A26	B38BHS ;B44CX'HZ	F	10	2009
	R3s1	60309	490	86	A02; A03	B44CX[illegible] ;B3[illegible]	F	10	
	R3s2	34566	570	346			F	10	
TC4	D4s1	383841	266	>100	A02 A;68	B7C XPJ; B15CX[illegible]	B	10	2009
	R4s1	17420	580	63	A02; A68	B15, B40	B	14	
	R4s2	17779	468	150			B	10	
	R4s3	11489	592	330			B	10	
TC5	D5s1	25899	352	>100	A02; A24	B15YCK ;B1517	B	14	2010
	R5s1	68465	525	41	A02, A[illegible]2	B40CXJJ(61); B57	BF	10	
	R5s2	17194	419	355			BF	10	
TC6	D6s1	18398	528	>100	A1[illegible] A3001…	B18; B35	BF	3 (*)	2010
	R6s1	4220	634	93	A01; A03	B37; B35	BF	10	
	R6s2	2512	743	290			BF	10	
TC7	D7s1	1358	1166	>[illegible]00	A01; A11	B57; B35	BF	10	2010
	R7s1	79433	952	59	A01; A26	B1501; B1517(63)	BF	10	
	R7s2	38150	541	35[illegible]			BF	8	
TC8	D8s1	264322	311	>1[illegible]	A02; A31	B07; B57	BF	27	2011
	R8s1	74624	489	27	A01; A24	B40; B51	BF	10	
	R8s2	58709	351	318			BF	16	
F1	S1	363614	335	43	A26:NPDV; A33:FMNW	B14:JRTM; B44:NPFA	BF	10	2012
	S1	163597	4[illegible]3	70			BF	10	
	S1	81286	5[illegible]7	160			BF	10	
F2	S2	112615	254	30	A02; A02	B40; B44	B	10	2012
	S2	35048	322	105			B	15	
F3	S3	>500000	118	24	A02; A68	B14; B51	BF	11	2013
	S3	115526	256	72			BF	10	
	S3	18462	575	180			BF	10	
F4	S4	24376	629	37	A11; A24	B08; B35	BF	10	2013
	S4	18991	567	120			BF	10	
F5	S5	102297	526	25	A02; A02	B1501; B51	BF	16	2013
	S5	153012	352	60			BF	10	
	S5	88501	628	133			BF	10	
F6	S6	260028	426	25	A03:DFPJ; A24:HBMP	B07:GZTE; B35:AMYE	BF	11	2014
	S6	51245	536	71			BF	17	
F7	S7	32918	713	30	A29; A29	B40; B44	BF	16	2014
	S7	25824	410	120			BF	22	
	S7	44407	641	190			BF	17	
F8	S8	>500000	361	25	A31:KEYY; A68:BPXK	B35:HDBU; B40:04	BF	10	2015
	S8	392928	386	87			BF	12	

Journal Pre-proof

F9	S9	117489	521	24	A03:DFPJ; A23:MZUY	B35:AMYE; B44:GYTP	B	10	2015
	S9	12439	635	90			B	26	

Table 1. Results of monitoring, HLA A / B typing and genotyping of viral sequences in study subjects. References: TC: transmission couple (donor ID underlined), F: patient without donor enrolled. BF: recombinant CRF12_BF. (*) Amplified from genomic DNA.

Journal Pre-proof

Table 2. Prediction results of epitopes and escape mutations for the HLA-I alleles studied. The percentage of known epitopes is also observed for each allele (significant difference was considered with a $p < 0.05$, Chi-square test, 95% CI). Frequency median (IQR) of TE: 0.67 (0.5-0.8) and EE: 0.44 (0.15-0.59). *Predicted *epitopes (NetMHC)*, **N of TE or EE/N of available epitopes per allele, ****Los Alamos HIV CTL-epitope Database*.

HLA allele	HLA frequency (Population)	n AHÍ allele	Founder Virus epitopes*	TE		EE		Frequency of TE **	Frequency of EE **	Known epitopes (with escape)*** (%)	Known epitopes (without escape) (%)	*p (Chi-square)*
				HLA	TCR	HLA	TCR					
A23	**4%**	1	15	3	0	1	0	*0,20*	*0,07*	25,0	16,7	*0,626*
B57	**5%**	1	23	4	3	1	0	*0,30*	*0,04*	66,7	40,0	*0,177*
A33	**6%**	1	26	6	2	2	0	*0,31*	*0,08*	30,0	18,2	*0,430*
B1517	**7%**	1	29	3	2	4	3	*0,17*	*0,24*	50,0	45,5	*0,723*
B14	**7%**	1	3	1	0	0	0	*0,33*	*0,00*	100,0	33,3	*0,186*
A26	**8%**	2	25	2	1	2	1	*0,12*	*0,12*	50,0	15,4	***<0,05***
A29	**9%**	2	5	2	0	1	0	*0,40*	*0,20*	100,0	50,0	*0,200*
B40	**10%**	5	38	8	5	5	2	*0,34*	*0,18*	71,4	31,3	***<0,01***
A11	**11%**	1	33	4	2	2	0	*0,18*	*0,06*	100,0	63,6	***<0,01***
A31	**12%**	2	35	9	4	6	3	*0,37*	*0,26*	63,6	21,4	***<0,01***
B08	**13%**	2	14	1	2	3	2	*0,21*	*0,36*	100,0	40,0	***<0,01***
B1501	**14%**	3	43	7	6	5	4	*0,30*	*0,21*	55,6	28,6	*0,076*
B07	**14%**	2	28	5	2	2	0	*0,25*	*0,07*	85,7	60,0	*0,086*
B51	**16%**	3	6	2	1	1	0	*0,50*	*0,17*	50,0	0,0	*0,303*
A03	**17%**	4	57	16	11	13	7	*0,47*	*0,35*	82,1	16,7	***<0,01***
B35	**17%**	6	44	17	12	3	1	*0,66*	*0,09*	83,3	36,4	***<0,01***
B44	**19%**	5	26	8	3	6	2	*0,35*	*0,31*	75,0	37,5	***<0,05***
A01	**21%**	6	9	4	2	2	0	*0,67*	*0,22*	93,3	40,0	*0,052*
A68	**24%**	3	59	11	8	4	2	*0,32*	*0,10*	73,3	18,2	***<0,01***
A24	**30%**	3	28	9	5	5	3	*0,50*	*0,29*	77,8	45,5	***<0,05***
A02	**47%**	11	87	32	17	10	6	*0,56*	*0,18*	87,7	52,2	***<0,01***
Total			**633**	**152**	**88**	**78**	**36**			**71,50%**	**33,60%**	***<0,01***

Journal Pre-proof

Case	*Monitoring period (days)*	*Log10 VL (copies/ml)**	*CD4+ T linphocyte (cell/ul) **	*TE*	*EE HLA*	*EE TCR*	*Reversion HLA (DR)*	***Total MAE (**)***	***Total MNAE***	***Total mutations***	***Escape rate (d^{-1})***	***Rh* Score**
TC1	230	2,33	727	16	7	8	1	16	3	19	*0,0696*	***0,05***
TC2	145	3,60	863	9	10	4	1	16	9	25	*0,1103*	***0,018***
TC3	346	4,54	570	14	12	2	3	10	15	25	*0,025*	***0,193***
TC4	330	4,17	530	18	9	7	2	19	30	52	*0,0576*	***0,057***
TC5	355	4,24	419	12	5	6	5	12	31	44	*0,0296*	***0,146***
TC6	290	3,40	743	9	9	4	2	13	8	21	*0,0448*	***0,053***
TC7	350	4,58	541	7	8	8	2	4	18	22	*0,0114*	***0,528***
TC8	318	4,77	351	11	8	5	4	7	6	13	*0,0184*	***0,294***
F1	160	5,09	480	12	2	1	2	3	1	4	*0,0188*	***0,736***
F2	105	4,54	678	5	1	3	4	4	2	6	*0,0381*	***0,113***
F3	180	4,83	416	13	6	3	1	8	0	8	*0,0278*	***0,323***
F4	120	4,28	598	8	3	0	3	3	0	3	*0,025*	***0,368***
F5	133	5,10	490	19	3	0	0	3	0	3	*0,0226*	***0,968***
F6	71	4,71	536	6	3	0	2	3	0	3	*0,0423*	***0,163***
F7	190	4,55	525	16	3	1	3	7	16	23	*0,0368*	***0,214***
F8	87	5,59	386	11	2	0	0	1	1	2	*0,023*	***0,825***
F9	90	4,09	635	13	6	0	1	6	8	14	*0,0667*	***0,112***

Table 3. Escape result associated with recipient alleles. ***Viral set point values. **References: TE: Transmitted escape; **EE:** escape during transmission event; **MAE:** Mutations associated with escape; **MNAE:** Mutations not associated with escape. ******Mutations associated with escape were supported with $p<0.01$ of statistical significance (Poisson distribution).

Journal Pre-proof

Conflict of Interest and Authorship Conformation Form

Paper title: **Bioinformatic analysis of post-transmission viral readaptation in Argentine patients with acute HIV-1 infection**

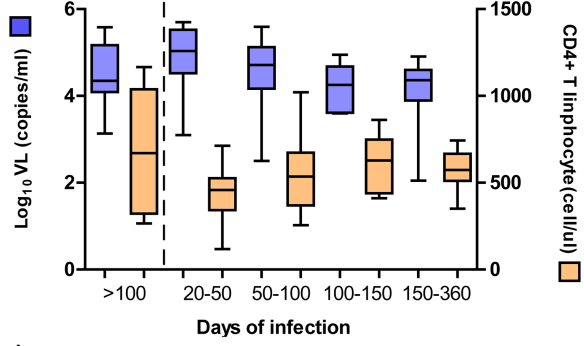
- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Author's name	Affiliation
Gabriel Damilano	Universidad de Buenos Aires, Argentina.
Omar Sued	Fundación Huésped-Buenos Aires, Argentina.
Sara Satorres	Universidad Nacional de San luis (UNSL), Argentina.
Maria Julia Ruiz	Hospitalier de l'Université de Montréal
Yanina Ghiglione	Universidad de Buenos Aires, Argentina.
Fanny Guzman	Universidad Católica de Valparaíso, Chile.
Gabriela Turk	Universidad de Buenos Aires, Argentina.
Florencia Quiroga	Universidad de Buenos Aires, Argentina.
Pedro Cahn	Hospital "Dr. JA Fernández", Buenos Aires, Argentina.
Horacio Salomón	Universidad de Buenos Aires, Argentina.
Dario Dilernia	Emory Vaccine Center, Emory University, Atlanta, GA.

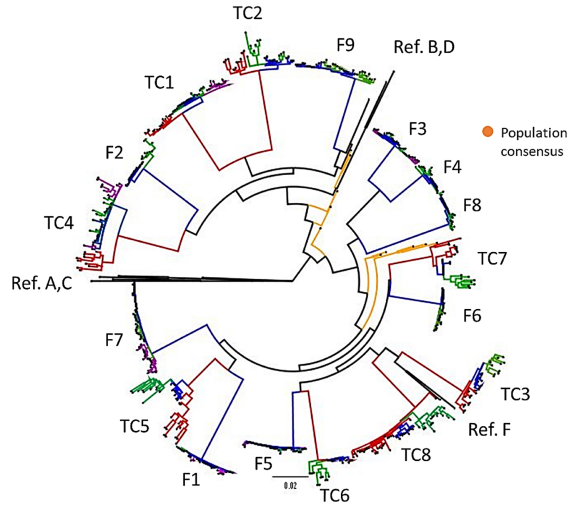
Journal Pre-proof

- The highest proportion of HIV-1 viral escape was transmitted
- Protein p24 showed the lowest proportion of epitopes associated with escape
- The frequency of CTL escape is influenced by the frequency of the allele
- The viral readaptation to new host could influence the progression of infection
- Gag segments associated with escape presented more significant entropy variation

Journal Pre-proof

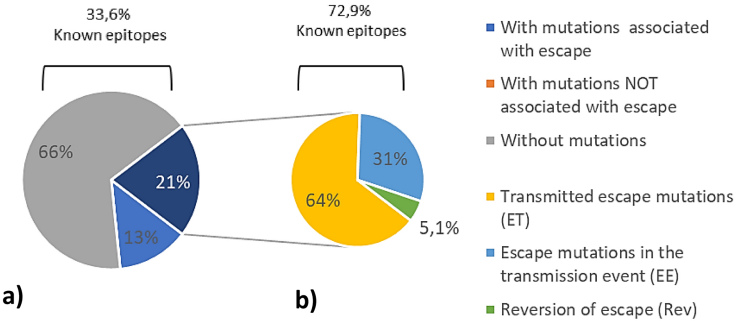

Donor's
(n=8)
Recipient's
(n=17)
Log10 VL (copies/ml)
CD4+ T linphocyte(cell/ul)
6
4
2
0
1500
1000
500
0
>100
20-50
50-100
100-150
150-360
Days of infection

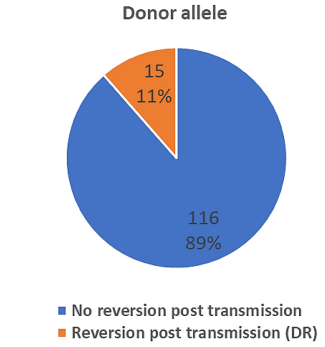
a)


TC2
F9
Ref. B,D
TC1
F3
Population consensus
F2
F4
F8
TC4
TC7
Ref. A,C
F6
F7
TC3
TC5
Ref. F
TC8
F5
F1
TC6
0.02

b)

Figure 1


33,6%
Known epitopes
72,9%
Known epitopes
66%
21%
13%
31%
64%
5,1%
With mutations associated with escape
With mutations NOT associated with escape
Without mutations
Transmitted escape mutations (ET)
Escape mutations in the transmission event (EE)
Reversion of escape (Rev)
a)
b)


Donor allele
15
11%
116
89%
No reversion post transmission
Reversion post transmission (DR)
c)

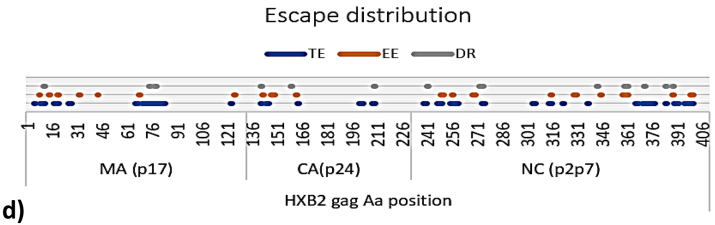

Escape distribution
TE
EE
DR
1
16
31
46
61
76
91
106
121
136
151
166
181
196
211
226
241
256
271
286
301
316
331
346
361
376
391
406
MA (p17)
CA(p24)
NC (p2p7)
HXB2 gag Aa position
d)

Figure 2

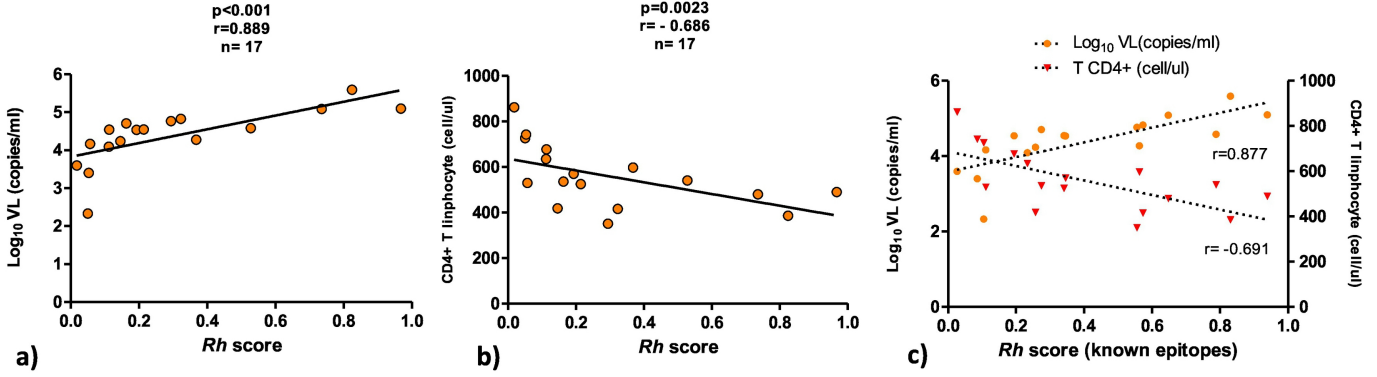
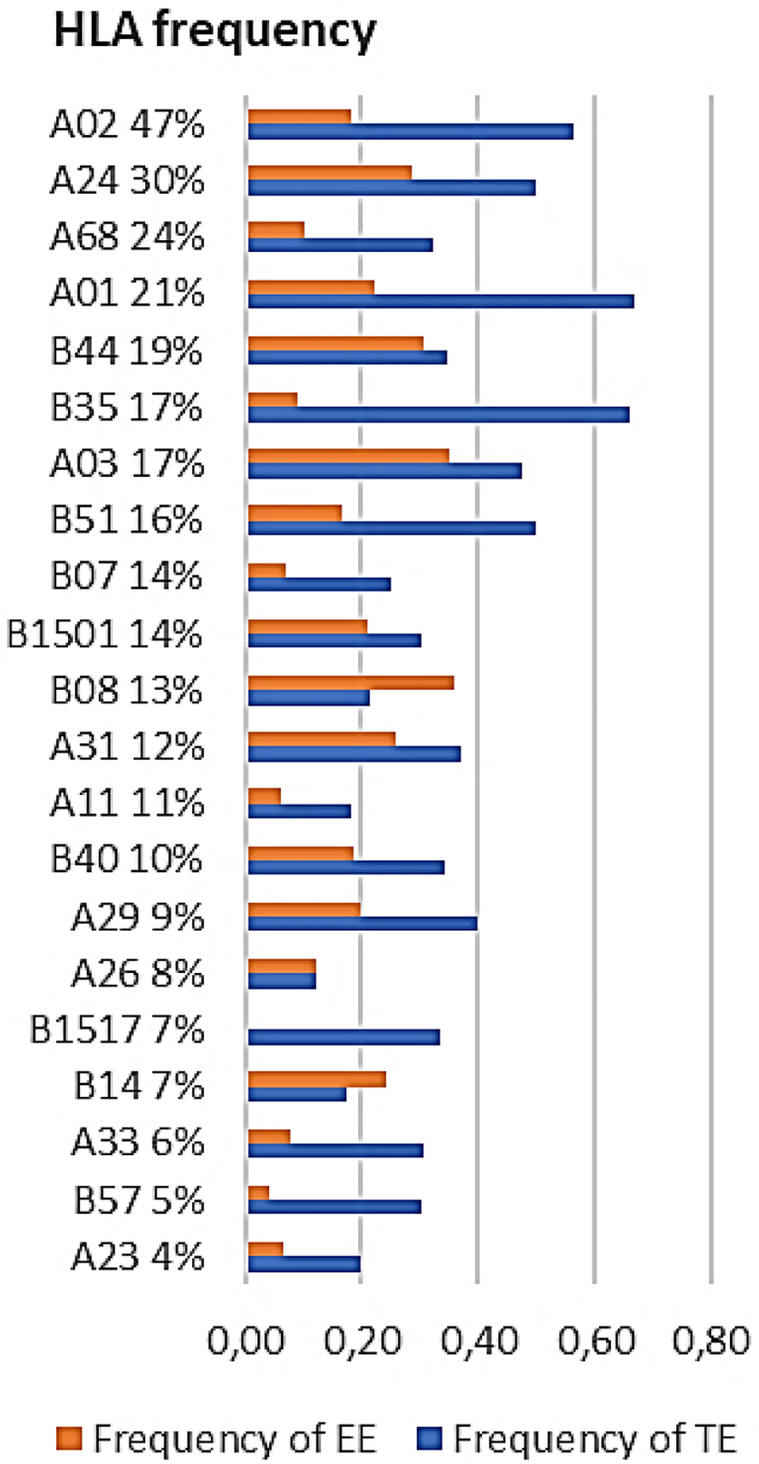
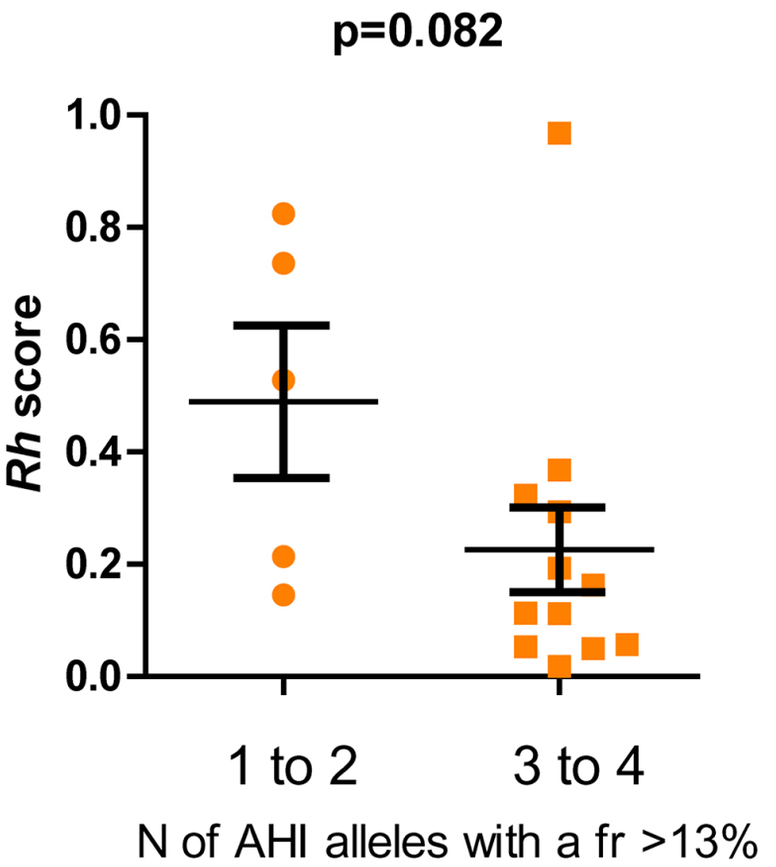

p<0.001
r=0.889
n= 17
Log10 VL (copies/ml)
Rh score
a)
p=0.0023
r= - 0.686
n= 17
CD4+ T linphocyte (cell/ul)
Rh score
b)
Log10 VL(copies/ml)
T CD4+ (cell/ul)
r=0.877
r= -0.691
Log10 VL (copies/ml)
CD4+ T linphocyte (cell/ul)
Rh score (known epitopes)
c)

Figure 3


HLA frequency
A02 47%
A24 30%
A68 24%
A01 21%
B44 19%
B35 17%
A03 17%
B51 16%
B07 14%
B1501 14%
B08 13%
A31 12%
A11 11%
B40 10%
A29 9%
A26 8%
B1517 7%
B14 7%
A33 6%
B57 5%
A23 4%
0,00 0,20 0,40 0,60 0,80
Frequency of EE
Frequency of TE

a)


p=0.082
Rh score
1.0
0.8
0.6
0.4
0.2
0.0
1 to 2
3 to 4
N of AHI alleles with a fr >13%

b)

Figure 4

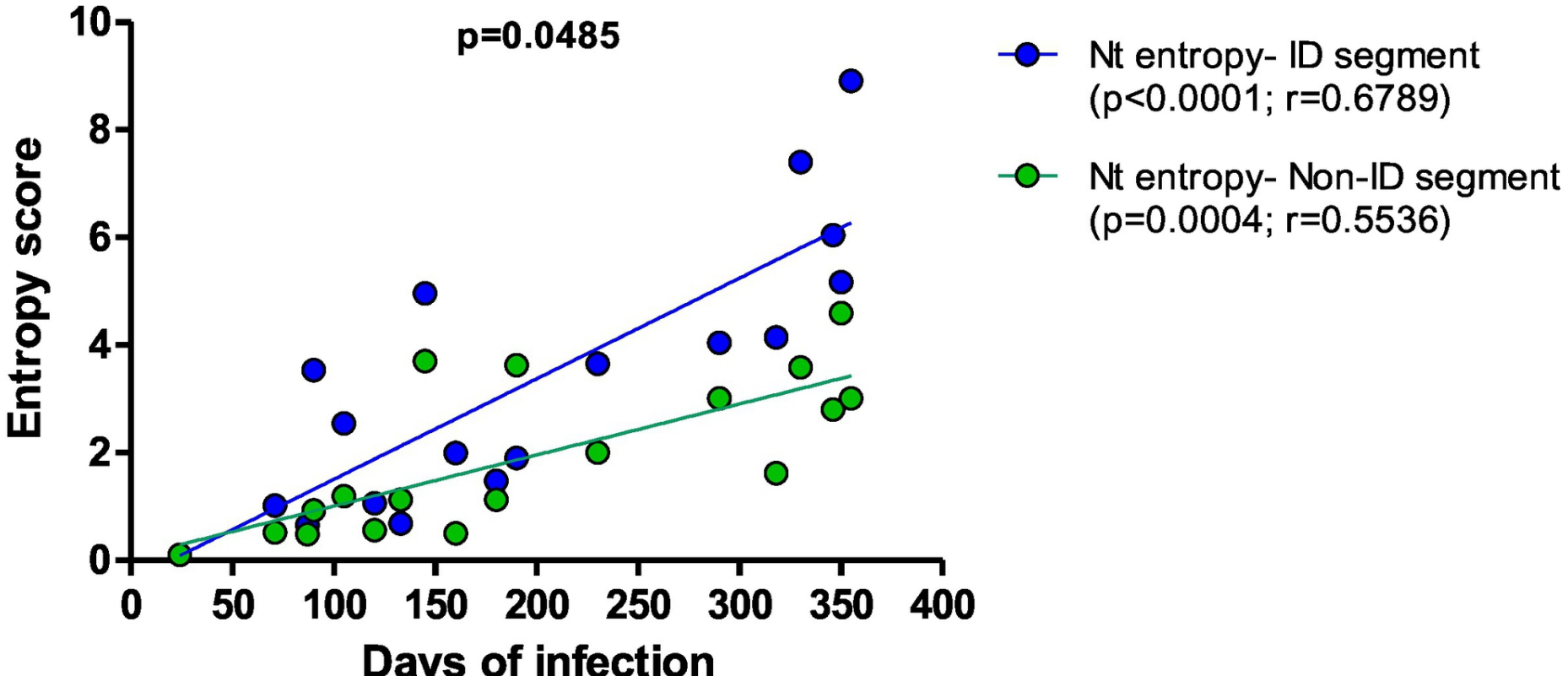

p=0.0485
Entropy score
Days of infection
0
2
4
6
8
10
0
50
100
150
200
250
300
350
400
Nt entropy- ID segment
(p<0.0001; r=0.6789)
Nt entropy- Non-ID segment
(p=0.0004; r=0.5536)

Figure 5