Greater Than Random HLA-B Homogeneity in HIV-1 Transmission Chains

(Abstract Number: 948)

Huyen Nguyen[1,2], Christian Thorball[3], Jacques Fellay[3], Jürg Böni[2], Sabine Yerly[4], Matthieu Perreau[5], Thomas Klimkait[6], Huldrych F. Günthard[1,2], Roger Kouyos[1,2], for the Swiss HIV Cohort Study (SHCS)[7]

1. Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland. 2. Institute of Medical Virology, Swiss National Center for Retroviruses, University of Zurich, Zurich, Switzerland. 3. Global Health Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 4. Laboratory of Virology, Geneva University Hospital, University of Geneva, Geneva, Switzerland. 5. Division of Immunology and Allergy, University Hospital Lausanne, University of Lausanne, Lausanne, Switzerland. 6. Department of Biomedicine, University of Basel, Basel, Switzerland. 7. The Swiss HIV Cohort Study, Zurich, Switzerland.

Take Home Message

- **We observed greater HLA-B homogeneity within HIV-1 transmission pairs and clusters compared to when the patients were randomly assorted into the same number of pairs/clusters**
- **The mechanism behind this observed distribution is unclear and deserves greater attention, but may involve how HLA-B modulates HIV transmission or superinfection**

Introduction

- HIV's capacity to escape immune recognition by Human Leukocyte Antigen (HLA) on cytotoxic T cells is a core component of HIV pathogenesis
- The role of HLA in viral escape is well understood at the individual level but not at the population level
- How the individual makeup of HLA class I in HIV-infected patients is distributed within transmission chains has not been previously described
- A deeper understanding of how HLA diversity may affect HIV transmission patterns may have large implications, e.g. in vaccine design[1]

Methods

Phylogenetic Tree

- Analyses were performed with data from the Swiss HIV Cohort Study (SHCS), a prospective, multicenter HIV study with high nationwide coverage
- Of 20,089 patients in the SHCS, HIV viral sequences (pol gene) were available for 11,922 patients (earliest sequence for each patient was used), which were then combined with 11,390 Los Alamos background sequences to produce a phylogenetic tree of HIV transmission

HLA Data

- HLA allele information on 4,892 SHCS patients were obtained via HLA imputation techniques[2,3,4] from exome/SNP data
- HLA-A, HLA-B, and HLA-C were examined (MHC class I)
- For all three MHC-1 HLA types, a broader two digit (corresponding to allele group) and a more specific four digit (corresponding to alleles sharing an identical amino acid sequence) level analysis were done
- For HLA-B, an additional supertype analysis was done, which is broader in categorization than two digit[5]

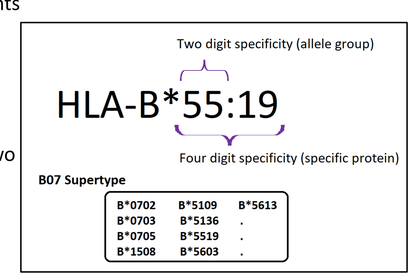


Figure 1. Example of levels of HLA-classification

Pairs Analysis

- Pairs were identified as: 1. being adjacent pairs in the tree that form a monophyletic clade, 2. having a maximum genetic distance between them of 3%, 3. both being SHCS patients, 4. both having data on at least a single HLA allele, 5. not having any individual who contracted HIV via perinatal transmission
- For each pair analysis (for HLA-A, -B, and –C, at 3 levels of allelic specificity), the SHCS pairs were scrambled 10,000 times to obtain a distribution of the average proportion of pairs with an allelic match (Figure 2)
- The true proportion in the unscrambled SHCS pairs was compared to the distribution from the randomizations to obtain a p value to see if HLA homogeneity within pairs is greater than if the individuals were randomly assorted



ACTUAL PAIRS RANDOMIZED PAIRS

Figure 2. Hypothetical example of how actual pairs (shown as rows) in SHCS tree are scrambled, with an HLA match being designated if at least one HLA allele matches between the paired individuals. Each individual is represented by a single-color block with two HLA alleles each. Here the proportion matching is 3/6 in the actual pairs compared to 1/6 in the randomized pairs.

Clusters Analysis

- Clusters were identified as: 1. forming a monophyletic clade within the tree with at least 2 individuals, 2. having a maximum genetic distance between individuals of 3%, 3. comprising of sequences which are over 80% from the SHCS, 4. having data on at least a single HLA allele for at least a third of the cluster individuals (disregarding individuals who were perinatally infected with HIV)
- An individual can only belong to a single cluster
- Individuals in the clusters were scrambled 10,000 times to obtain a distribution of the average proportion of all possible patient-to-patient comparisons within each cluster with at least a single allelic match (Figure 3)
- The true proportion in the original SHCS clusters were compared to the distribution of average proportions from the randomizations, to obtain a p value to see if HLA homogeneity is greater in the clusters than if the individuals were randomly assorted

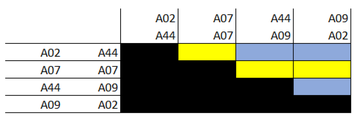


Figure 3. Hypothetical example of how a cluster of 4 individuals (with two HLA alleles each) is calculated for HLA homogeneity: with 4 individuals there are 6 possible comparisons—an HLA allelic match is denoted in three comparisons by blue, non-matches in yellow. As 3/6 comparisons have a match, the calculated proportion of homogeneity for this cluster is 0.5.

Results

General Characteristics of SHCS Patients and Those Analyzed in the Pairs and Clusters

	Swiss HIV Cohort Study (SHCS)	SHCS Patients with Sequences in the Transmission Phylogeny	SHCS Patients in Transmission Pairs Fitting Criteria	SHCS Patients in Transmission Clusters Fitting Criteria
Total Number	20089	11922	672	1319
Median Age (IQR)	32.69 (26.78-41.00)	27.00 (33.04-41.23)	33.60 (28.04-39.68)	33.88 (27.85-41.00)
Male (%)	14548 (72.42)	8554 (71.75)	527 (78.42)	1072 (81.27)
Risk Group				
MSM (%)	7764 (38.65)	4738 (39.74)	334 (49.70)	665 (50.42)
Heterosexual (%)	6624 (32.97)	4245 (35.61)	205 (30.51)	340 (25.78)
Intravenous Drug User (%)	4824 (24.01)	2430 (20.38)	116 (17.26)	276 (20.92)
Other (%)	877 (4.37)	509 (4.27)	17 (2.53)	38 (2.88)
White (%)	13611 (67.75)	9199 (77.16)	640 (95.24)	1276 (96.74)

Table 1. Overall demographic information of patients in the SHCS, as well as those represented in the transmission phylogeny, and those with HLA data in the analyzed pairs and clusters as fitting the criteria (described in Methods section).

Statistical Analyses of HLA Homogeneity in Transmission Clusters

Table 2. Overview of statistical analyses looking at HLA allelic homogeneity (average proportion of all comparisons within clusters where at least one match exists). Different analyses were performed looking at varying HLA types (A, B, and C) and levels of HLA classification specificity (4-digit, 2-digit, and supertype). Significant p values of tests (<0.05) denoted in bold.

HLA-Type	HLA Analysis Level	Number of Clusters	Average Cluster Size	Average Percentage of Individuals in Clusters with HLA Data	Mean Proportion of Intracluster Comparisons with a Match	Average Value in Randomizations	P Value
A	4-digit	469	3.981	65.82	0.379	0.376	0.447
A	2-digit	469	3.981	65.82	0.413	0.411	0.456
B	4-digit	496	3.911	66.47	0.191	0.185	0.329
B	2-digit	496	3.911	66.47	0.291	0.251	**0.009**
B	Supertype	496	3.911	66.47	0.659	0.611	**0.002**
C	4-digit	475	3.979	66.35	0.294	0.293	0.478
C	2-digit	475	3.979	66.35	0.410	0.425	0.810

- 672 SHCS patients both had HLA data and were identified as belonging to monophyletic pairs (301, 325, and 310 pairs for HLA-A, -B, and –C respectively)
- 1319 SHCS patients both had HLA data and belonged to identified clusters (469, 496, and 475 clusters for HLA-A, -B, and –C respectively)
- Transmission pairs were shown to be significantly more homogenous when looking at HLA-B supertype
- Transmission clusters were shown to be significantly more homogenous when looking at HLA-B at the 2-digit and supertype level
- Results are in line with literature suggesting greater importance of HLA-B in HIV immune escape [6]
- Mechanism behind this observed greater homogeneity is unclear and needs further investigation (HLA effect on transmission or superinfection?)

Statistical Analyses of HLA Homogeneity in Transmission Pairs

HLA-Type	HLA Analysis Level	Number of Pairs	Proportion of Pairs Matching	Average Value in Randomizations	P Value
A	4-digit	301	0.405	0.380	0.131
A	2-digit	301	0.442	0.415	0.120
B	4-digit	325	0.200	0.182	0.168
B	2-digit	325	0.265	0.241	0.136
B	Supertype	325	0.655	0.608	**0.014**
C	4-digit	310	0.271	0.287	0.718
C	2-digit	310	0.406	0.420	0.691

Table 3. Overview of statistical analyses looking at HLA allelic homogeneity (proportion of pairs with at least one matching HLA-allele). Different analyses were performed looking at varying HLA types (A, B, and C) and levels of HLA classification specificity (4-digit, 2-digit, and supertype). Significant p values of tests (<0.05) denoted in bold.

Sources

1.M. P. Davenport, L. Loh, J. Petravic, S. J. Kent, Rates of HIV Immune Escape and Reversion: Implications for Vaccination. Trends Microbiol. 16, 561–566 (2008).
2. X.M. Jia, et al., Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. PLoS ONE. 8.6: e64683 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064683 (2013).
3. A. Szolek, et al., OptiType: Precision HLA Typing from Next-Generation Sequencing Data. Bioinformatics. 30, 3310-3316 (2014).
4. A.T. Dilthey, et al., High-Accuracy HLA Type Inference from Whole-Genome Sequencing Data Using Population Reference Graphs. PLoS Comput. Biol. 12.10: e1005151 https://www.ncbi.nlm.nih.gov/pubmed/27792722 (2016).
5. J. Sidney, et al., HLA Class I Supertypes: A Revised and Updated Classification. BMC Immuno. 9, 1 (2008).
6. P. Kiepiela, et al., Dominant Influence of HLA-B in Mediating the Potential Co-evolution of HIV and HLA. Nature. 432(7018): 769-75 (2004).