Supplemental methods

RT-PCR, PCR (outer and inner), sequencing and vector primer sequences

C2-C4 env primer:
Seq2: 5’-TCCTCCATATCTCTCTCCAGGTC-3’ (RT-PCR, outer PCR)
Seq3: 5’-TATGGGATCAAAGCCTAAAGCATG-3’ (outer PCR)
Seq5: 5’-GTCAACTCACTGCTGTAAATGGC-3’ (inner PCR, seq)
Seq6: 5’-ATCTAATTTGTCCACTGATGGGAGG-3’ (inner PCR, seq)

Vector primer:
T7: 5’-TAATACGACTCACTATAGGG-3’
SP6: 5’-GATTTAGTGACACTATAG-3’
Supplemental methods and results

Patient population

The prospective Amsterdam Cohort Studies on HIV Infection and AIDS (ACS) is a well-documented cohort study among HIV-infected and HIV-uninfected individuals who are at risk for acquiring HIV infection. Enrolment of men who have sex with men (MSM) started in October 1984 and up to December 31, 2008, 2,383 MSM have had at least one visit, with 1,588 testing HIV seronegative, 585 testing HIV seropositive and 210 HIV seroconverters. Clinical and epidemiological data is collected, laboratory markers such as CD4+ T cell numbers and plasma viral load (VL) are determined, serum or plasma is stored at -70°C and peripheral blood mononuclear cells are cryopreserved at 3-monthly intervals. The ACS are conducted in accordance with the ethical principles set out in the Declaration of Helsinki and are approved by the Medical Ethics Committee of the Academic Medical Center in Amsterdam. Written informed consent is obtained for every participant.

Molecular cloning and sequencing of HIV-1 env C2–C4

Second-round PCR products were cloned into the pGEM-T Easy Vector system (Promega), transformed into competent DH5a Escherichia coli (Invitrogen), and plated on Luria-Bertani agar with blue-white screening. White colonies were picked at random (4–32 colonies per reaction). Cloned PCR products were amplified (vector primers, T7 [5’-TAATACGACTCACTATAGGG- 3’] and SP6 [5’-GATTTAGGTGACACTATAG-3’]) using the above-described PCR program. After purification of PCR products (ExoSAP-IT; USB), the T7 and SP6 primers were used for sequencing (Big Dye Terminator Cycle Sequencing kit, version 1.1; Applied Biosystems). Sequences were determined using an automated DNA sequencer (Applied Biosystems).
When sequence diversity did not reflect the heteroduplex pattern obtained by heteroduplex mobility assay (HMA), an additional 4 nested PCRs were performed, and products were cloned and plated. Two clones per plate were picked, PCR with the SP6 and T7 primers was performed, and PCR products were sequenced. Clonal sequences were aligned for each patient by means of the ClustalW algorithm, and alignments were manually edited using BioEdit software (Bio-Edit, version 7.0.5.3; Ibis Biosciences). Alignments were visually inspected for the presence of mismatches, insertions, and deletions.

Statistical analysis

For the SC time point HMA pattern, we analyzed time from SC to each end point; for the 1 year post-SC time point HMA pattern or sequence diversity, we analyzed time from 1 year after SC to each end point. In addition, we studied time to AIDS-related death from the moment of AIDS diagnosis onwards. Log Rank P-value was used to determine significant differences in the clinical course of infection between both groups. Univariate and multivariate analyses in a model that included HMA heteroduplexes or sequence diversity (continuous variable), CCR5 Δ32 heterozygous genotype, HLA-B57, positive MT-2 cultures, set point CD4+ T-cells > 500 cells/µl blood and set point VL < 10^{4.5} copies/ml plasma were performed from 2 years after SC to AIDS (SPSS 16.0 software package, SPSS Inc., Chicago, IL, USA).

The association of HMA pattern or sequence diversity with VL or CD4+ T-cell count at set point was tested using Student’s t-test as implemented in GraphPad Prism 5 (GraphPad Software, La Jolla, CA, USA).
**Predictive value of HMA pattern for HIV-1 disease course**

We tested the predictive value of HMA pattern in the group of 89 MSM seroconverters versus several established markers of disease progression, namely CCR5 Δ32 genotype, HLA-type, detection of CXCR4-using virus by MT-2 assay (CXCR4 use), VL and CD4+ T cell count at VL set point in a model analyzing the time from 2 years after SC to AIDS CDC 1993. Univariate analysis showed significant effects only for HMA pattern, CXCR4 use and CD4+ T cell count at VL set point in this group (Table 1). Subsequent multivariate analysis with only these 3 predictors in the model analyzing time to AIDS from 2 years after SC to AIDS CDC 1993 indicated that all three were independently predictive of HIV-1 disease progression (Table 1).

As expected, univariate and multivariate analyses also confirmed the independent predictive value of envelope sequence diversity for HIV-1 disease progression (Table S2).
### Table S1a. Kaplan-Meier survival analysis of association between presence of heteroduplexes in HMA and time to clinical endpoint

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>First time point (SC)(^a)</th>
<th>Second time point (12 months post-SC)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heteroduplex present</td>
<td>Heteroduplex present</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>CD4&lt;400(^c)</td>
<td>37/45</td>
<td>41/44</td>
</tr>
<tr>
<td>AIDS93(^e)</td>
<td>29/45</td>
<td>26/44</td>
</tr>
<tr>
<td>Death</td>
<td>23/45</td>
<td>18/44</td>
</tr>
</tbody>
</table>

HMA, heteroduplex mobility assay; SC, seroconversion; ns, not significant (\(P\) value > 0.10). No associations were found with time to first emergence of CXCR4-using HIV-1 or time between AIDS diagnosis and AIDS-related death (data not shown).

\(^a\) Survival analysis for time from SC to each end point

\(^b\) Survival analysis for time from 1 year post-SC to each end point

\(^c\) CD4+ T cell count less than 400 cells/\(\mu\)l

\(^d\) Number of individuals with events/total number of individuals

\(^e\) AIDS according to the 1993 CDC definition
Table S1b. Kaplan-Meier survival analysis of association between viral envelope sequence diversity and time to clinical endpoint

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Second time point (12 months post-SC)(^a)</th>
<th>Absolute sequence diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n  (^b)</td>
</tr>
<tr>
<td>CD4&lt;400(^c)</td>
<td>47/65</td>
<td>ns</td>
</tr>
<tr>
<td>AIDS93 (^d)</td>
<td>47/70</td>
<td>0.0003</td>
</tr>
<tr>
<td>Death</td>
<td>34/71</td>
<td>0.0029</td>
</tr>
</tbody>
</table>

SC, seroconversion; ns, not significant (\(P\) value > 0.10). No associations were found with time to first emergence of CXCR4-using HIV-1 or time between AIDS diagnosis and AIDS-related death (data not shown).

\(^a\) Survival analysis for time from 1 year post-SC to each endpoint

\(^b\) Number of individuals with events/total number of individuals

\(^c\) CD4+ T cell count less than 400 cells/µl

\(^d\) AIDS according to the 1993 CDC definition
Table S2. Predictive value of viral envelope sequence diversity one year post-SC, CCR5 genotype, HLA type, set point CD4+ T cell count, set point HIV-1 VL and set point MT-2 assay result for progression to AIDS (CDC 1993) in 74 MSM seroconverters.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Comparison</th>
<th>Univariate</th>
<th></th>
<th></th>
<th>Multivariate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>P value</td>
<td>RH</td>
<td>n</td>
<td>P value</td>
<td>RH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(95% CI)</td>
<td></td>
<td></td>
<td>(95% CI)</td>
</tr>
<tr>
<td>Diversity</td>
<td>Continuous</td>
<td>63</td>
<td>0.0006</td>
<td>2.50</td>
<td>63</td>
<td>0.005</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.48-4.22) *</td>
<td></td>
<td></td>
<td>(1.25-3.54) *</td>
</tr>
<tr>
<td>CCR5</td>
<td>WT vs Δ32</td>
<td>63</td>
<td>ns</td>
<td>1.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.46-2.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA</td>
<td>X vs B57</td>
<td>63</td>
<td>ns</td>
<td>4.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.62-32.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP CD4</td>
<td>&lt;500 cells/µl</td>
<td>63</td>
<td>0.009</td>
<td>2.28</td>
<td>63</td>
<td>0.023</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.23-4.20)</td>
<td></td>
<td></td>
<td>(1.11-4.01)</td>
</tr>
<tr>
<td>SP VL</td>
<td>&gt;1 x 10^4.5 cp/ml</td>
<td>63</td>
<td>ns</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.93-3.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP MT-2</td>
<td>POS vs NEG</td>
<td>63</td>
<td>&lt;0.001</td>
<td>9.64</td>
<td>63</td>
<td>0.004</td>
<td>5.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.17-29.3)</td>
<td></td>
<td></td>
<td>(1.74-17.41)</td>
</tr>
</tbody>
</table>

SC, seroconversion; VL, viral load; MSM, men who have sex with men; n, number of individuals included in analysis; RH, relative hazard; CI, confidence interval; Δ32, 32 base pair deletion; WT, wild type; X, all other HLA types; SP, set point; POS, positive MT-2 assay; NEG, negative MT-2 assay.

Univariate and multivariate analyses in a model analyzing the time from 2 years after SC until AIDS (CDC 1993) diagnosis. P values from univariate and multivariate Cox proportional hazard analyses.

*Relative hazard for a 0.01 increase in diversity.
Supplemental figure

Supplemental figure legend

Figure S1. Set point VL, set point CD4+ T cell count, and CD4+ subsets by HMA pattern.

Heteroduplexes in HMA at one year post-SC are correlated with set point CD4+ T cell count (A) but not with set point VL (B). HMA patterns showed a trend for correlation with the percent CD38+ (activated) CD4+ T cells (C) and were inversely correlated with the percent CD45RO+ (memory) CD4+ T cells (D). Comparisons were analyzed for statistical significance using Student’s t-test as implemented in GraphPad Prism 5 software. Mean and SEM (error bars) are represented for the respective groups. P values for significance are shown above each panel.
Figure S1

A

\[ P = 5.36 \times 10^{-1} \]

\begin{align*}
\text{HIV RNA (log}_{10} \text{cp/ml)} & \\
\text{No (n=36)} & \text{Yes (n=49)}
\end{align*}

B

\[ P = 2.25 \times 10^{-2} \]

\begin{align*}
\text{CD4+ T cells (cells/μl)} & \\
\text{No (n=36)} & \text{Yes (n=49)}
\end{align*}

C

\[ P = 5.05 \times 10^{-2} \]

\begin{align*}
\% \text{CD38+ CD4 cells} & \\
\text{No (n=26)} & \text{Yes (n=30)}
\end{align*}

D

\[ P = 4.64 \times 10^{-2} \]

\begin{align*}
\% \text{RO+ CD4 cells} & \\
\text{No (n=26)} & \text{Yes (n=30)}
\end{align*}