β-defensin Genomic Copy Number Is Associated With HIV Load and Immune Reconstitution in Sub-Saharan Africans

Robert J. Hardwick,1 Wondwossen Amogne,2,3 Sabina Mugusi,4,5 Getnet Yimer,6,7 Eliford Ngaimisi,6,8 Abiy Habtewold,6,7 Omary Minzi,8 Eyasu Makonnen,7 Mohammed Janabi,9 Lee R. Machado,1 Maria Viskaduraki,10 Ferdinand Mugusi,9 Getachew Aderaye,2 Lars Lindquist,3 Edward J. Hollox,1,* and Eleni Aklillu6,†

1Department of Genetics, University of Leicester, United Kingdom; 2Internal Medicine, Addis Ababa University, Ethiopia; 3Institution of Medicine, Unit of Infectious Diseases, Karolinska Institutet, Karolinska University Hospital, Huddinge, Sweden; 4Department of Internal Medicine, Muhimbili National Hospital, Dar es Salaam, Tanzania; 5Venhälsan, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden; 6Department of Clinical Pharmacology, Addis Ababa University, Ethiopia; 6Department of Pharmacology, School of Pharmacy, and 9Department of Internal Medicine, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; and 10College of Medicine, Biological Sciences and Psychology, University of Leicester, United Kingdom

AIDS, caused by the retrovirus human immunodeficiency virus (HIV), is the leading cause of death of economically active people (age, 15–59 years) in sub-Saharan Africa. The host genetic variability of immune response to HIV and immune reconstitution following initiation of highly active antiretroviral therapy (HAART) is poorly understood. Here we focused on copy number variation of the β-defensin genes, which have been shown to have anti-HIV activity, and are important chemoattractants for Th17 lymphocytes via the chemokine receptor CCR6. We determined β-defensin gene copy number for 1002 Ethiopian and Tanzanian patients. We show that higher β-defensin copy number variation is associated with increased HIV load prior to HAART (P = .005) and poor immune reconstitution following initiation of HAART (P = .003). We suggest a model where variable amounts of β-defensin expression by mucosal cells, due to gene copy number variation, alters the efficacy of recruitment of Th17 lymphocytes to the site of infection, altering the dynamics of infection.

African countries have the highest disease burden of human immunodeficiency virus (HIV), with 9.2% prevalence in Addis Ababa in Ethiopia and over 10% in Dar-es-Salaam in Tanzania, yet almost all genetic studies have focused on cohorts from Western countries [1]. The genetic architecture of HIV susceptibility in Africans is likely to be different to Europeans, yet genome-wide association studies of host susceptibility to HIV have not yielded any significant results [2]. These studies miss regions that show copy number variation, particularly structurally complex regions that are not correlated with alleles at flanking single-nucleotide polymorphism (SNP) markers [3].

Copy number variation is defined as the variation in copy number of a given DNA sequence in a diploid genome. Copy number variation is common in the genome, affects gene expression, and involves immune response genes [4, 5], suggesting that it may affect susceptibility of the host to infectious disease. Copy number variation of the killer cell immunoglobulin receptor genes has been shown to affect host control of HIV infection, as determined by the viral load (VL) at setpoint [6]. The chemokine gene CCL3L1, which encodes a ligand for the chemokine receptor CCR5 used as a coreceptor by R5 strains of HIV, also shows copy number variation [7]. It has been shown to affect HIV acquisition, progression to AIDS, and immune reconstitution following highly active antiretroviral therapy (HAART) [8, 9]. However, attempts at
replicating this finding have not always been successful [10, 11],
and using quantitative polymerase chain reaction (PCR) to
determine CCL3L1 copy number may generate false-positive
associations [12]. It is therefore important that the most accu-
rate technologies are used to determine the number of gene
copies in disease association studies. In this study we use the
paralogue ratio test, a robust method to measure copy number
variation that has been validated extensively [13, 14].

β-defensins are a family of multifunctional peptides, with
roles in inflammation and reproduction as well as possessing
direct antimicrobial activity [15]. In humans, 7 β-defensin
genes show extensive copy number variation as a block, with a
modal copy number of 4 per diploid genome [16]. This varia-
tion is reflected in levels of β-defensin in serum, at least for
human β-defensin 2 (hbd2), encoded by the DEFB4 gene [17].
Of the 7 β-defensin genes, 2 (DEFB4, and DEFB103 encoding
hbd2 and hbd3 proteins, respectively) have been shown to
encode peptides that have anti-HIV activity in vitro and che-
motactic activity [18, 19]. High β-defensin copy number is as-
associated with increased susceptibility to psoriasis and high
levels of hbd2 are associated with psoriasis severity [17, 20].

Immune reconstitution following HAART initiation, as
measured by CD4+ T-lymphocyte count, varies between indi-
viduals, and the host factors contributing to this are poorly
deﬁned [21]. Given the important role of β-defensins in the
innate immune response, and evidence of their interactions
with HIV, the β-defensin copy number variable locus is a
strong candidate locus for inﬂuencing the immune response
to HIV. We conducted a cohort study investigating how β-
defensin copy number affects HIV VL immediately prior to
administration of HAART and immune reconstitution follow-
ing initiation of HAART. This is, to our knowledge, the ﬁrst
study to test association of copy number variation with
disease susceptibility in 2 large African patient populations, as
well as the ﬁrst large study of β-defensin copy number varia-
tion in HIV patients.

METHODS

Ethics and Patient Recruitment
The study protocol was approved by the Institutional Review
Board at the Faculty of Medicine, Addis Ababa University and
Ethiopian Science and Technology Ministry, the regional
ethical review board in Stockholm at the Karolinska Institutet,
and the ethical review committee of Muhimbili University of
Health and Allied Sciences. Written informed consent was ob-
tained from each subject before the start of this study.

Study Design and Participants
This study was part of a European and Developing Countries
Clinical Trial Partnership (EDCTP) funded, multicenter
project on optimization of HIV/tuberculosis cotreatment in

Africa. The study was initiated at the same time and run in
parallel at tuberculosis and HIV clinics in Addis Ababa, Ethi-
pia, as well as in Dar es Salaam, Tanzania. A total of 649
newly diagnosed ART-naive Ethiopian patients living in Addis
Ababa were recruited and enrolled prospectively. The Tanzani-
an HIV-only and HIV/tuberculosis coinfected study partici-
pants (n = 353) are described elsewhere [22]. Study partici-
pants were prospectively recruited and followed up to 1
year to monitor clinical, virologic, and immunological out-
comes of HAART.

Treatment
HIV-only patients received efavirenz-based HAART (600
mg/day) containing stavudine/lamivudine/efavirenz or zidovu-
dine/lamivudine/efavirenz or tenofovir/lamivudine/efavirenz
according to the National AIDS Control Program treatment
guidelines. Patients with HIV and tuberculosis coinfection re-
ceived rifampicin-based short course treatment under directly
observed therapy (Initiation phase: rifampicin, isoniazid, pyra-
znamid, and ethambutol for 2 months followed by continuation
phase with isoniazid and rifampicin for 4 months). Efavirenz-
based HAART was initiated on the fourth week of rifampicin-
based antituberculosis therapy and continued throughout.
Compliance was deﬁned as self-reported use of medication.

Clinical and Laboratory Analysis
After informed consent and appropriate pretest counseling, the
demographic characteristics were recorded. A detailed history of
present and past illnesses was taken along with a general physi-
examination. Patient visits for clinical and laboratory exami-
nation were scheduled every week for the ﬁrst month, every 2
weeks during the second month, and thereafter every 4 weeks
until 1 year. The laboratory investigations were performed ac-
cording to the same clinical schedule in both treatment groups.
HIV load and CD4 counts were determined before the initia-
tion of HAART and during HAART on the 12th, 24th, 36th,
and 48th weeks. CD4 T-lymphocyte counts were determined
using standard ﬂow cytometry on whole blood. HIV-1 RNA
plasma load was quantiﬁed by real-time PCR using HIV-1
assay (Abbott RealTime) or COBAS AMPLICOR HIV-1
MONITOR Test, v. 1.5 (Roche Diagnostics) according to the
manufacturer’s guidelines. The standard protocol (limit of de-
tection, <400 copies/mL) was used to analyze plasma samples
collected at baseline and the ultrasensitive protocol (limit of de-
tection, <40 copies/mL) was used for plasma samples collected
during HAART in accordance with the manufacturer’s instruc-
tions and in duplicates to ensure reproducibility.

Extraction of DNA and Analysis of β-Defensin Copy Number
and CCR5 Δ32 Polymorphism
Venous blood samples were collected from study participants.
Genomic DNA was isolated from peripheral blood leukocytes
using QIAamp DNA Maxi Kit (Qiagen). The β-defensin copy number was determined, blind to clinical information, using a triplex paraglogue ratio test described previously [13, 14]. Every experiment included the same 6 positive controls of known copy number, the same samples as used in previous studies [14, 23], which were used to normalize the experimental results and minimize interexperimental technical variation in typing copy number. Integer copy number was called by combining information from all 3 assays and estimating the most likely copy number given the data. In addition, noninteger copy number values were calculated by averaging the estimates from the 3 assays. DNA samples with integer copy number calls that were not statistically supported \( (P > .05) \) were excluded from the analysis. The 32 bp deletion of the CCR5 gene was genotyped by standard PCR using the primers CCR5-delta32-F \( (5'-cagatctaaaaaaaaggtcttct-3') \) and CCR5-delta32-R \( (5'-gattcccgagtagcagatgac-3') \) followed by agarose gel electrophoresis, resulting in an expected size of 119 bp for the deleted allele and 151 bp for the insertion allele.

Statistical Analyses

To rule out population structure within the Ethiopian and Tanzanian cohorts as a confounding factor we stratified our data by site of sample collection and compared the copy number distributions. In Ethiopians 3 sites (BHC, LHC, and MHC) were discounted given their very low sample sizes of 5 or less, leaving 5 sites (AHC, BLH, BT, KHC, and KIR, \( n = 524 \)) for analysis. In Tanzanians all 3 sites were analyzed (IDC, MNH, and MWN, \( n = 314 \), after removal of 2 samples with copy number data but no site data). In all sites, the modal copy number was 4, and there were no significant differences in copy number distributions between sites of sample collection (Ethiopia \( P = .075 \), Tanzania \( P = .375 \), 1-way analysis of variance). We constructed a generalized linear model, using PASW Statistics 18.0, where the dependent variable (VL) was modeled as a gamma distribution, reflecting the skew in distribution toward lower values, and allowed better fitting of the model to the data than a normal distribution, while still permitting the identification of any linear associations with VL (unlike log transforming VL data). In this model, we assigned population and disease status as fixed factors, CD4⁺ count as a scalar covariate and integer copy number as an ordinal covariate, and calculated using Type III sum of squares.

For the copy number variation association test using a Gaussian mixture model, we used CNVtools, implemented in the statistical language R [24]. Its strength is that it incorporates error directly into the association test, rather than assigning copy numbers, which are then assumed to be error-free, and then performing the statistical test. We split each population into 2 groups, one group with viral loads above the median and the other with viral loads below the median. We removed 2 outlying samples (copy number of 1 and 9) that appeared to interfere with fitting the model. We then fitted the raw copy number data (Ethiopian \( n = 397 \)) to a Gaussian mixture model, correcting for batch effect in measurement error between the 2 groups, with variance constant across copy numbers, and 7 components (reflecting the best fit using the Bayesian information criterion). We identified the best fit of the model under the 2 hypotheses when comparing the high VL and low VL groups (H₀ no difference in copy number distribution, and H₁ difference in copy number distribution) and compared the log likelihoods. An allelic disease model was used, that is, the odds of being in one group compared to the other are not constrained by a linear trend. A model where the odds were constrained to be linear with respect to copy number was tested and the null hypothesis could not be rejected at \( \alpha = 0.05 \).

We identified 2 overlapping classes of HAART nonresponders, based on observation of an outlier group in the VL follow-up data (VL > 1000) and on the definition of virological failure given by the US Department of Health and Human Services panel on Antiretroviral Guidelines for Adults and Adolescents (VL > 200). Patients who fell below the threshold on any of their follow-up measurements were deemed to have responded to HAART for the purposes of this analysis, even if the VL rose to above the threshold on a subsequent visit, perhaps as a result of noncompliance. We decided to take an empirical nonparametric approach, using a Monte Carlo randomization test, to test for differences in copy number distribution. There was no evidence of sampling bias: We found no difference in copy number distribution between the 290 patients who had VL follow-up data and the 301 patients who did not, nor any copy number difference between those individuals who attended 1, 2, or 3 follow-up clinics.

For CD4⁺ timepoint data, after transforming the x axis (timepoint in weeks) by log2, we performed regression against CD4 count for both low copy number (below median, \( \leq 3 \)) and higher copy number \( \geq 4 \) groups. Because each \( P \) value represents an independent study set, they could be combined using the method of Fisher by summing the -2ln(likelihood) of each study and comparing to a \( \chi^2 \) distribution with 2n degrees of freedom (where \( n = \) number of independent studies) [25, 26]. We also constructed a generalized linear mixed model, using STATA, where the dependent variable (CD4⁺ count) was modeled as a normal distribution. In this model, we assigned population and disease status as fixed factors, initial CD4 count and log2 (time since HAART initiation) as scalar covariates and either integer copy number as an ordinal covariate, or copy number class as a fixed factor. The model was calculated using Type III sum of squares, with a variance correction to allow for multiple CD4⁺ timepoint readings from a single patient.

RESULTS

We measured β-defensin copy number using a triplex paraglogue ratio test on 649 Ethiopian and 353 Tanzanian HIV-only and
HIV/tuberculosis coinfected patients and inferred integer copy number using a maximum-likelihood approach [21]. Copy number distribution did not differ between HIV-only and HIV/tuberculosis cohorts and was consistent with previous studies (Table 1) [23]. VL is widely regarded as a prognostic marker of HIV infection, severity, and infectiousness. Analysis of 564 patients whose CD4 count was <200 cells/mm$^3$ and whose VL before initiation of treatment was measured (Table 2), showed that there was a significant relationship with VL and increasing β-defensin copy number ($P = .0003$, $F$-test, Figure 1A). Binning of raw gene copy number estimates into integers, and then performing association studies, presumes that the initial data binning is error-free. To show that the association of β-defensin copy number with VL was robust to any errors in copy number calling, we used the CNVTools software [24], which fits a Gaussian mixture model to the raw, unbinned, copy number data. We analyzed VL data from 397 Ethiopian patients, with error explicitly taken into account, and confirmed an association of β-defensin copy number with VL (high VL vs low VL, $P = .002$, likelihood ratio test, Figure 1B and 1C).

Population of origin, coinfection with tuberculosis, and CD4$^+$ count could all possibly confound the association between β-defensin copy number and VL. To test for this possibility, we constructed a generalized linear model with population of origin and presence or absence of tuberculosis as fixed factors. In the resulting analysis, population of origin and CD4$^+$ count significantly influenced viral load ($P = .014$ and $P = .007$, respectively), but presence of tuberculosis did not. Taking into account these other factors, higher β-defensin copy number was associated with higher VL ($P = .005$). We genotyped all samples for the Δ32 mutation of CCR5, which generates a CCR5 null allele, and is known to affect VL and HIV progression [27]. No samples carried the null allele, consistent with previous studies showing a very low frequency outside Eurasia [28], and confirms that CCR5 Δ32 is not a confounding factor in the analysis (data not shown).

### Table 1. Diploid β-Defensin Copy Number Frequencies

<table>
<thead>
<tr>
<th>Copy no.</th>
<th>Ethiopian HIV only</th>
<th>Ethiopian HIV + TB</th>
<th>Tanzanian HIV only</th>
<th>Tanzanian HIV + TB</th>
<th>Zambian$^a$</th>
<th>Yoruban$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
<td>0.13</td>
<td>0.13</td>
<td>0.18</td>
</tr>
<tr>
<td>4</td>
<td>0.47</td>
<td>0.44</td>
<td>0.35</td>
<td>0.36</td>
<td>0.37</td>
<td>0.39</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
<td>0.24</td>
<td>0.28</td>
<td>0.21</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>0.10</td>
<td>0.09</td>
<td>0.12</td>
<td>0.17</td>
<td>0.21</td>
<td>0.13</td>
</tr>
<tr>
<td>7</td>
<td>0.02</td>
<td>0.03</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>N</td>
<td>262</td>
<td>354</td>
<td>192</td>
<td>133</td>
<td>112</td>
<td>79</td>
</tr>
</tbody>
</table>

Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis.

$^a$ Data from reference [21].

### Table 2. Summary of Study Cohorts

<table>
<thead>
<tr>
<th>Pheno-</th>
<th>Ethiopian HIV only</th>
<th>Ethiopian HIV + TB</th>
<th>Tanzanian HIV only</th>
<th>Tanzanian HIV + TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Total</td>
<td>Copy no. not called</td>
<td>CD4$^+$ baseline and timepoint data and CD4$^+ \leq 200^{a}$</td>
<td>Both VL and CD4$^+$ baseline data and CD4$^+ \leq 200^{b}$</td>
</tr>
<tr>
<td>No.</td>
<td>284</td>
<td>22</td>
<td>177</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>11</td>
<td>227</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td>14</td>
<td>157</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>147</td>
<td>14</td>
<td>56</td>
<td>40</td>
</tr>
</tbody>
</table>

Mean CD4$^+$ at baseline, cells/mm$^3$: 106.4 83.28 100.2 94.97

Mean VL at baseline, HIV copies/mL: 243 700 252 400 535 600 755 500

Mean copy no.: 4.308 4.289 4.781 4.847

Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis; VL, viral load.

$^a$ These data were used for the CD4$^+$ timepoint analyses.

$^b$ These data were used for the association tests with pre-HAART (highly active antiretroviral therapy) viral load, and for the calculation of mean CD4$^+$, VL, and copy no. shown in this table.
To test whether $\beta$-defensin copy number variation affected immune reconstitution, we identified nonresponders to HAART on the basis of VL up to 48 weeks (Figure 2A), using thresholds of 1000 copies HIV/mL and 200 copies HIV/mL, the latter being the definition of virological failure given by the US Department of Health and Human Services (DHHS) Panel on Antiretroviral Guidelines. Using both thresholds, higher copy number was significantly associated with nonresponse (or slower response) to HAART ($P = .004$ and $P = .015$, permutation test, Figure 2B). We then used CD4$^+$ count

Figure 1. Association of $\beta$-defensin copy number variation with human immunodeficiency virus (HIV) load. A, HIV load vs integer $\beta$-defensin copy number, for the 564 patients, with the regression line, significance level ($F$-test), and 95% confidence intervals on the regression are shown. Note the 2 different scales shown on the y-axis. B and C, Histogram of raw $\beta$-defensin copy number data from Ethiopian patients with low HIV load (B) and high HIV load (C) showing clustering around integer copy number values. A Gaussian mixture model has been fitted to the data using the CNVtools software, modeling the copy numbers as a series of Gaussian distributions, one distribution for each copy number (line curves show posterior probability distributions of different copy numbers). Comparing the fit of the Gaussian distributions under the null hypothesis (no difference in copy number distribution) and the alternative hypothesis (a difference in copy number distribution) allows statistical comparison of degree of difference between the copy number distributions.
Figure 2. Association of β-defensin copy number variation with response to highly active antiretroviral therapy (HAART). A, Summary plot of viral load distributions of different responders to HAART, with nonresponders partitioned by viral load (VL) > 200 and by VL > 1000. B, Cumulative distribution of β-defensin copy number, divided by VL response to HAART. C–F, Recovery of CD4+ counts after HAART initiation in different cohorts, stratified by copy no. Means ± standard error of the mean are shown.
follow-up data to confirm the effect of β-defensin copy number variation on immune reconstitution following HAART initiation. Given the previous results, we would expect that lower copy number would be associated with improved immune reconstitution, and high copy number associated with impaired immune reconstitution. This was observed for 3 of 4 cohorts, with 2 cohorts showing a statistically significant difference (Figure 2). Analysis across all 4 cohorts shows a highly statistically significant difference ($P = 5.6 \times 10^{-3}$).

Given the difference in response to HAART between different cohorts, as shown in Figure 2, we decided to analyze the data further to investigate the nature of this variation. In particular, we wanted to determine whether variation in the initial CD4$^+$ count, which is known to affect response to HAART [29], was causing the observed difference between cohorts or, indeed, the observed effect of copy number on immune reconstitution. To test this, we constructed a generalized linear model to measure the dependency of CD4$^+$ count on population of origin, presence or absence of tuberculosis, and copy number class (3 copies or fewer and 4 copies or more) as fixed factors, with log(time since HAART initiation) and initial CD4$^+$ count as continuous variables. Unsurprisingly, both log2 (time since HAART initiation) and initial CD4$^+$ count were very strongly associated with CD4$^+$ count in this model (both $P < .001$). Copy number class was also strongly associated with CD4$^+$ count, when these 2 factors were taken into account ($P = .003$), confirming our observation from Figure 2. Perhaps surprisingly, CD4$^+$ count was not associated with the presence of tuberculosis coinfection but was associated with the population of origin ($P = .007$), reflecting a similar pattern of association found with pre-HAART baseline VL. The association of population of origin with both the baseline VL prior to HAART initiation and response to HAART may reflect both genetic and environmental differences between the 2 populations. This emphasizes the importance of taking population of origin/ethnicity into account in similar studies.

**DISCUSSION**

We have determined β-defensin genomic copy number using accurate methods on 1002 HIV African patients and compared with extensive clinical data taken immediately prior to HAART initiation, and with data taken during the initial stages of HAART treatment. Taken together, our data suggest that higher β-defensin genomic copy number is associated with increased VL and impaired immune reconstitution following HAART initiation. Considering the published evidence suggesting that β-defensins bind to HIV and inhibit HIV replication [16], this appears counterintuitive. However, these assays were performed in vitro, often in serum-free media, using unphysiologically high concentrations of β-defensin, so the relevance to the antiviral activity of β-defensins in vivo is unclear. Our alternative model emphasizes the chemotaxant nature of the β-defensins. Because the copy number variation that we measured includes 7 β-defensin genes, any one of these genes, or a combination of 2 or more, may be responsible for the effects that we report. However, hbd2, encoded by the gene DEFB4, is known to be a strong chemotaxant for CD4$^+$ T-lymphocytes at physiological concentrations, acting through CCR6 (a receptor that defines a population of Th17 lineage cells found at mucosal sites) [19, 30]. Thus, hbd2 mediates the arrest of Th17 cells on the vascular endothelium and subsequent infiltration of the mucosa [31]. Th17 cells are preferentially infected by HIV-1 due to high coexpression of the HIV coreceptor CCR5 [32]. The copy number variation is likely to affect levels of mucosal β-defensin expression, thereby affecting the pool of Th17 cells available for HIV infection. Alternatively, β-defensins may act indirectly to affect immune reconstitution, by recruiting other cells, such as dendritic cells, to the mucosa, which then alter the cytokine milieu [33]. Both these hypotheses, derived from our genetic data and previously published work, should be tested directly in functional experiments.

Our model is consistent with the observed association of high β-defensin copy number with increased risk and severity of psoriasis [20], where Th17 cell infiltration represents a critical early stage of disease [34]. More generally, this model also suggests that β-defensin copy number variation may be a risk factor in other infectious and inflammatory diseases where Th17 responses are critical. Identification of variable immune reconstitution stratified by β-defensin copy number variation suggests targeting of the β-defensin-CCR6 axis to control viremia and aid immune reconstitution.

The data we present also support an important role for copy number variation in infectious disease. Follow-up studies on other HIV patient cohorts, from Africa and elsewhere, are a priority, both for β-defensin copy number variation and for other copy number variation regions that contain genes that are involved in the immune system. New approaches for accurately genotyping copy number variation genome-wide in combination with very large African HIV cohorts will be needed to fully explore the role of host genetics in HIV pathogenesis and immune reconstitution following HAART.

**Notes**

**Acknowledgments.** Thanks to Jasper Estranero and Jenny Bowdrey for technical support, and the patients for participation in this study.

**Author contributions.** The project was designed by E. J. H and E. A. Trial management, clinical measurement, DNA collection and extraction, and clinical data recording was performed by W. A., S. M., G. Y., A. H., E. N., O. M., E. M., M. J., F. M., G. A., L. L. and E. A. Copy number typing was performed by R. J. H., with analysis performed by R. J. H., E. J. H., L. R. M. and M. V. The article was written by E. J. H with input from all authors.
**Financial support.** This work was supported by the United Kingdom Medical Research Council New Investigator award (grant no. GO801123) to E. J. H.; European and Developing Countries Clinical Trials Partnership (grant numbers CT.2005.32030.001, CG_TA.05.40204.005); and the Swedish International Development Cooperation Agency/Department for Research Cooperation (grant nos. HIV-2006-031, SWE 2007–270).

**Potential Conflicts of interest.** All authors: no reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**