Copy Number Variation within Human β-Defensin Gene Cluster Influences Progression to AIDS in the Multicenter AIDS Cohort Study

Rajeev K. Mehlotra1, Jean-Eudes Dazard2, Bangan John1,3, Peter A. Zimmerman1,3, Aaron Weinberg4, and Richard J. Jurevic4*

1Center for Global Health and Diseases, Case Western Reserve University School of Medicine, Cleveland, OH, USA
2Center for Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, OH, USA
3Department of Biology, College of Arts and Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA
4Department of Biological Sciences, Case Western Reserve University School of Dental Medicine, Cleveland, OH, USA

Abstract

Study background: DEFB4/103A encoding β-defensin 2 and 3, respectively, inhibit CXCR4-tropic (X4) viruses in vitro. We determined whether DEFB4/103A Copy Number Variation (CNV) influences time-to-X4 and time-to-AIDS outcomes.

Methods: We utilized samples from a previously published Multicenter AIDS Cohort Study (MACS), which provides longitudinal account of viral tropism in relation to the full spectrum of rates of disease progression. Using traditional models for time-to-event analysis, we investigated association between DEFB4/103A CNV and the two outcomes, and interaction between DEFB4/103A CNV and disease progression groups, Fast and Slow.

Results: Time-to-X4 and time-to-AIDS were weakly correlated. There was a stronger relationship between these two outcomes for the fast progressors. DEFB4/103A CNV was associated with time-to-AIDS, but not time-to-X4. The association between higher DEFB4/103A CNV and time-to-AIDS was more pronounced for the slow progressors.

Conclusion: DEFB4/103A CNV was associated with time-to-AIDS in a disease progression group-specific manner in the MACS cohort. Our findings may contribute to enhancing current understanding of how genetic predisposition influences AIDS progression.

Keywords: AIDS progression; β-Defensin; DEFB4; DEFB103A; MACS; X4 emergence

Introduction

The antimicrobial peptides human β-defensin 2 (hBD-2) and -3, encoded by DEFB4 and DEFB103A, respectively, are induced during infections and inflammation, and can reach biologically relevant concentrations in the systemic circulation [1-4]. To date, all evidence suggests that the β-defensin gene cluster on chromosome 8p23.1, carrying DEFB4 and DEFB103A, varies en bloc, with copy numbers ranging from 2 to 12 per diploid genome (PDG) [5,6]. DEFB4 Copy Number Variation (CNV) correlates with its transcript levels [6], and with hBD-2 serum concentrations [2]. Recently, we demonstrated that: (1) CCR5-tropic (R5) and CXCR4-tropic (X4) strains of HIV-1 induce DEFB4 and DEFB103A transcripts in normal human oral epithelial cells; (2) hBD-2 and -3 inhibit HIV-1 infection by both strains, with greater activity against X4 viruses; and (3) this inhibition is due to direct binding with virions, and through down-modulation of the CXCR4 co-receptor [7]. A later study confirmed the anti-HIV activity of these hBDs [8]. We subsequently discovered that hBD-3 competes with stromal-derived factor 1 (SDF-1), the natural ligand for CXCR4, and acts as an antagonist of CXCR4 on T cells by promoting its internalization without subsequent activation of the cell [9].

Chemokine receptor usage (viral tropism), a complex phenomenon where many aspects are still unresolved, plays a major role in HIV-1 disease pathogenesis. It is generally accepted that R5 viruses predominate at early stages of infection, whereas X4 or R5/X4 (dual- or mixed-tropic) viruses tend to emerge at later stages and their emergence appears to be associated with accelerated progression to disease [10-12]. Viral subtype, drug resistance, and impact of highly active antiretroviral therapy (HAART) may be important factors in influencing viral tropism [13-17]. However, the role of host genetic factors, to our knowledge, is not known. Based on our previous findings [7,9], we hypothesized that there may be an association between DEFB4/103A CNV and time to emergence of X4 viruses (time-to-X4), and thus between DEFB4/103A CNV and time to AIDS. To test our hypothesis, we utilized samples from the previously published Multicenter AIDS Cohort Study (MACS), which provides a longitudinal account of viral tropism in relation to rates of HIV-1 disease progression [18].

Materials and Methods

Study population

De-identified packed white blood cell pellets (n = 50) were obtained from the MACS. These samples were part of the study, conducted on HIV-1 seroconverters (n = 67) for whom emergence of X4 viruses was monitored and used to examine whether X4 emergence before onset of AIDS was related to the rate of AIDS progression [18]. The baseline characteristics of this study population have been previously described [18]. The self-identified racial/ethnic distribution of the 50 subjects was: white, non-Hispanic (White, n = 43), white, Hispanic (Hispanic, n = 4), and black, non-Hispanic (Black, n = 3), which reflects the fact that the MACS cohort is predominantly white (Caucasian). For these...
subjects, the information regarding their dates of seroconversion, emergence of X4 viruses, defined as the detection of R5/X4 viruses, and AIDS diagnosis were made available for further analysis (Table 1 and Supplemental Figure 1).

Disease progression groups

Previously, the rate of AIDS progression was defined in terms of the interval (τ) between seroconversion and onset of an AIDS-defining event (CD4+ T cell count of <200 cells/µL on two consecutive visits and/or an AIDS-defining illness) [18]. AIDS progression was considered Rapid (τ ≤ 5 years), Moderate (5 < τ < 9 years), Slow (9 < τ < 11 years), and Very Slow (τ > 11 years) [18]. To increase statistical power, the disease progression groups were merged by similarities of CD4+ and CD3+ T cell count profiles and HIV tropism patterns (Figure 1 of [18]). We thus identified a “Fast” group (τ ≤ 9 years, n = 36) after merging the Rapid and Moderate groups, and a “Slow” group (τ > 9 years, n = 14) after merging the Slow and Very Slow groups.

Determination of CNV

Genomic DNAs were extracted from 200 µL of each of the pellets using the QiAamp DNA Blood Mini Kit (Qiagen). For the determination of DEFB4 and DEFB103A copy numbers, the real-time quantitative PCR assay was used as described [19]. Reference genes TBP (TATA-Box Binding Protein, GenBank accession #AL031259) and DEFB1 (hBD-1, GenBank accession #NT_023736), specific primer sets producing only one specific product of ~150 bp at 54°C annealing temperature, reaction mix, and conditions were used as described [19], and the Bio-Rad CFX96™ system (Bio-Rad Laboratories) was used for the PCR analysis. Each sample was run in triplicate. Data were analyzed by the comparative Ct method, and the copy numbers were calculated as described [19]. Recently, we assessed the reliability and validity of this assay using 151 multi-population samples, which included 46 well-characterized samples from five diverse populations from the Coriell Cell Repositories [20].

Statistical analysis

All 50 subjects were assumed independent from each other. Missing DEFB4/103A copy numbers (n = 5) were determined by simple-imputation results from the expectation-maximization algorithm, assuming missing data at random [21]. The missing CNVs were imputed by using other genotype data (CCR5 -2459G>A, CCR5 32-bp deletion, CCR2 190G>A and CXCL12 801G>A [unpublished]), as well as disease progression groups and race. Initial exploratory data analyses to compare trends between the two time-to-event outcomes, and to investigate a trend in a given time-to-event as a function of CNV’s were carried out with a standard non-parametric trend test requiring minimal assumptions (1-sided Cox-Stuart test) [22]. For measuring the monotonic association between the two right-censored time to events, a non-parametric partial correlation coefficient (partial Kendall’s τ) was used [23].

We used the right-censoring modeling framework of risk or survival analysis to model either time-to-X4 or time-to-AIDS outcome, for which the event was either X4 emergence or AIDS diagnosis, each of which was observed or not during each patient’s follow-up time (Table 1, Supplemental Figure 1). We assumed non-informative (or random) censoring and non-competing risks for the two outcomes (e.g., Supplemental Figure 2). We used the non-parametric Kaplan-Meier estimator [24] to model the relationship of the Disease-Free (DF) or X4-Free (XF) probability as a function of the right-censored time-to-event outcome of interest. We report from Kaplan-Meier models the time-to-event at median DF or XF probability and log-rank test results adjusting for race/ethnicity. We also used Cox proportional hazards modeling (Cox model) [25] adjusting for race/ethnicity.

For all statistical analyses, differences were considered significant at the a = 0.05 level. All statistical modeling, computations, and plotting were performed in the R language and environment for statistical computing (http://www.r-project.org/). We used our own code in addition to the following R packages: “survival” for Kaplan-Meier and Cox regression modeling and “NADA” for correlation analysis in the presence of censoring (http://cran.r-project.org/).

Results

Distribution of CNV

The DEFB4/103A copy numbers ranged from 2 to 5 (median = 3). The distribution of copy numbers was: 2 (n = 13), 3 (n = 11), 4 (n = 17) and 5 (n = 4). The copy numbers could not be determined for five samples due to low genomic DNA concentrations (<50 ng). After imputation, the distribution of copy numbers was: 2 (n = 14), 3 (n = 13), 4 (n = 19), and 5 (n = 4). To increase statistical power, we aggregated DEFB4/103A CNV into either two categories: a “minimal” (CNV = 2, n = 14) and a “higher” (CNV > 2, n = 36), or three categories: below the median (CNV < 3 (CNV = 2), n = 14), median (CNV = 3, n = 13), and above the median (CNV > 3, n = 23).

Relationship between time-to-event outcomes

Although a positive slope was observed between time-to-X4 and time-to-AIDS (Cox-Stuart P = 0.007), these two outcomes were weakly correlated (τ = 0.284). When we performed this analysis using the Fast and Slow progression groups, we observed a stronger relationship.

Table 1: Racial/ethnic distribution and clinical information of study subjects.

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>X4 emergence</th>
<th>AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n TTX</td>
<td>n TTA</td>
</tr>
<tr>
<td>White/non-Hispanic</td>
<td>43 29 14</td>
<td>0.99 0.93 0.99 14.33</td>
</tr>
<tr>
<td>White/Hispanic</td>
<td>4 2 2</td>
<td>0.95 1.45 4 0.95 8.18</td>
</tr>
<tr>
<td>Black/ non-Hispanic</td>
<td>3 1 2</td>
<td>3.68 2 1 3.68 6.08</td>
</tr>
<tr>
<td>Total</td>
<td>50 32 18</td>
<td>0.95 9.93 49 1 0.95 14.33</td>
</tr>
</tbody>
</table>

* = event present, - = event absent
TTX = Time-to-X4
TTA = Time-to-AIDS
*range, in subjects with events

Figure 1: Kaplan-Meier plots of aggregated DEFB4/103A copy number effect on time-to-AIDS (A) and time-to-X4 (B).
between the two outcomes in the Fast group (Supplemental Figure 2A) than in the Slow group (Supplemental Figure 2B).

Consistent with this observation, Kaplan-Meier and Cox model analyses showed a much stronger significant difference between the Fast and Slow groups in time-to-AIDS (Fast, 6.46 years; Slow, 10.14 years; log-rank \( P = 9.91 \times 10^{-12} \); Cox \( P < 1 \times 10^{-12} \); Supplemental Figure 2C), compared with time-to-X4 (Fast, 4.71 years; Slow, 8.00 years; log-rank \( P = 0.124 \); Cox \( P = 0.046 \); Supplemental Figure 2D). This was also evident among the four disease progression groups (see Methods) (time-to-AIDS (log-rank \( P < 1 \times 10^{-12} \)); Cox \( P = 2.83 \times 10^{-8} \), time-to-X4 (log-rank \( P = 0.004 \); Cox \( P = 0.031 \)).

**Time-to-event analysis of DEFB4/103A CNV**

A trend analysis revealed a stronger relationship between DEFB4/103A CNV and time-to-AIDS (Cox-Stuart \( P = 0.084 \); Supplemental Figure 3A) than between DEFB4/103A CNV and time-to-X4 (Cox-Stuart \( P = 0.412 \); Supplemental Figure 3B).

In line with the above trend analysis, both Kaplan-Meier and Cox model analyses detected a trend towards significant difference between DEFB4/103A CNV categories in time-to-AIDS (CNV = 2, 7.55 years; CNV > 2, 8.13 years; log-rank \( P = 6.2 \times 10^{-4} \); Cox \( P = 0.113 \); Figure 1A), but not in time-to-X4 (CNV = 2, 6.96 years; CNV > 2, 5.21 years; log-rank \( P = 0.83 \); Cox \( P = 0.947 \); Figure 1B). We observed no difference in the results obtained by using either two or three categories of DEFB4/103A CNV (data not shown).

**Interaction effect of DEFB4/103ACNV with AIDS progression groups**

Whereas a trend towards significant difference in time-to-AIDS was observed between DEFB4/103A CNV categories (Figure 1A), a stronger significant difference was detected by both Kaplan-Meier and Cox model analyses, after fixing the AIDS progression groups. In the Slow group, time-to-AIDS appeared to be significantly slower when the DEFB4/103A CNV > 2 (CNV = 2, 9.54 years; CNV > 2, 10.92 years; log-rank \( P = 4.03 \times 10^{-4} \); Cox \( P = 0.005 \); Figure 2A), in contrast with the Fast group, where no significant difference in time-to-AIDS was observed between the CNV categories (CNV = 2, 7.22 years; CNV > 2, 6.09 years; log-rank \( P = 0.127 \); Cox \( P = 0.888 \); Figure 2B). We observed no difference in the results obtained by using either two or three categories of DEFB4/103A CNV (data not shown). Because a significant difference in time-to-AIDS between the CNV categories was observed in the slow group, but not in the Fast group, it strongly implies an interaction effect between DEFB4/103A CNV and disease progression.

In an alternative approach, we compared time-to-AIDS between AIDS progression groups, after fixing the DEFB4/103A CNV categories. We observed a larger difference between the Fast and Slow groups when the DEFB4/103A CNV > 2 (log-rank \( P = 6.41 \times 10^{-8} \); compared with the DEFB4/103A CNV = 2 (log-rank \( P = 4.03 \times 10^{-4} \);). Collectively, these results suggest that the protective effect of higher DEFB4/103A copy number (>2) strongly depends on whether the patients are slow or fast progressors.

**Discussion**

By utilizing samples previously analyzed for viral tropism in relation to rates of HIV-1 disease progression [18], we observed a weak correlation between time-to-X4 and time-to-AIDS. However, there was a stronger relationship between these two outcomes in the Fast progression group than in the Slow progression group, which is consistent with previous findings [10-12,14,18]. Although the exact biological mechanisms that regulate viral tropism in vivo remain elusive, it has been clear that emergence of the X4 phenotype heralds a rapid clinical deterioration, suggesting that co-receptor usage strongly influences disease progression [10-12]. A literature analysis of 498 patients, who were in late-stage disease, found high frequencies (60-77%) of CXCR4 tropism for five major HIV-1 subtypes [14]. Similarly, in the sample set used here, the frequency of X4 viruses and the likelihood of X4 virus emergence were significantly higher among patients with more rapid disease progression than among those with slower disease progression [18]. In the present study, our direct analysis to measure a relationship between emergence of X4 viruses and progression to AIDS suggests that these two outcomes are distinct, and that their relationship strongly depends on whether the patients are fast or slow progressors. Time-to-AIDS showed a much larger difference between fast and slow progressors than did time-to-X4, further supporting that these two outcomes reflect different levels of biological complexity [12,18].

In line with our first observation, we observed that higher DEFB4/103A copy number was associated with a delayed onset of AIDS, but not with delayed X4 virus emergence. Furthermore, the effect of DEFB4/103A CNV on AIDS progression was more pronounced when we considered the interaction between DEFB4/103A CNV categories and AIDS progression groups: copy number >2 PDG was significantly associated with slower time-to-AIDS in the Slow group, but not in the Fast group. A considerable advancement has been made in understanding the ways that slow progressors can control HIV-1 infection and AIDS progression naturally. It is understood that the slow progression phenotype is an outcome of interplay among viral, host genetic, and immunologic components [26]. Recent studies have shown that higher copy numbers of certain host genes, such as chemokine (CCL3L1, and CCL4L1, ligands for CCR5) [27-30] and immune receptor family (KIR) [31], exert a protective effect against HIV infection and AIDS progression. In addition, CNV of β-defensin genes may also play an important role in susceptibility to HIV infection [32]. In Brazilian HIV-positive children, median copy number of DEFBI04, which encodes hBD-4 and is located in the same chromosomal region as DEFB4 and DEFBI03A, was lower compared with HIV-exposed uninfected children and healthy controls, suggesting that DEFBI04 may be involved in protection from vertical transmission of HIV [33]. On the other hand, in Ethiopian and Tanzanian patients, higher β-defensin CNV was associated with increased HIV viral load prior to HAART, and poor immune reconstitution following initiation of HAART, suggesting that higher β-defensin copy number may be a risk factor [32]. Here we show that higher DEFB4/103A copy number is an additional genetic factor associated with slower progression to AIDS.

---

**Figure 2.** Kaplan-Meier plots of aggregated DEFB4/103A copy number effect, after fixing the AIDS progression groups Slow (A) and Fast (B).
since the MACS cohort is predominantly Caucasian, our finding, together with the results of recent studies [32,33], indicates that the association between β-defensin CNV and HIV infection/disease progression may be population-specific.

Higher β-defensin CNV correlates with higher peptide concentration in serum [2]. The anti-HIV activity of β-defensins [7,8], together with down-modulation of CXCR4 [7], and antagonism and internalization of CXCR4 by hBD-3 [9] may be responsible for the observed association between DEF4/103A CNV and AIDS progression. Testing this plausible mechanistic model requires further studies.

We acknowledge that our study has some limitations. We combined the three racial/ethnic groups, due to very low numbers of Hispanic and Black subjects, and thus were not able to perform a comparative analysis of DEF4/103A CNV vis-à-vis AIDS progression in various racial/ethnic groups. The observed association between DEF4/103A CNV and AIDS progression is based on either two or three categories of DEF4/103A CNV. A more in depth analysis using distinct copy numbers, and their diplootypes [34], may provide a better resolution of DEF4/103A CNV vis-à-vis AIDS progression. Addressing these limitations and reproducing our findings will need longitudinal analysis of viral tropism in larger HIV/AIDS cohorts. For such a study, the cost, in addition to accessibility and performance, of assay for the determination of viral tropism may pose a challenge [35-37].

In conclusion, we observed that higher DEF4/103A CNV was associated with a significant delay in time-to-AIDS in individuals with slow disease progression phenotype. Further exploration to unravel the possible mechanism(s) underlying this association may enhance our understanding of genetic predisposition to AIDS progression.

Acknowledgments

This study was supported by a grant from the National Institute of Dental and Craniofacial Research at the National Institutes of Health (1P01DE019759, A.W.; Project 4, R.J.J.). J.-E.D. was supported by this grant as a Statistician (Core B).

We are thankful to Ms. Janet Schollenberger, Senior Project Coordinator, CAMACS, for providing us the samples as well as data related to the samples to perform the present study. We thank Dr. Vinay Cheruvu for his initial efforts on this project, and Dr. John Tilton, Dave McNamara, Dr. Carolyn Myers, and Tenisha Phipps for their critical comments on the manuscript.

References


