Regulatory Polymorphisms in the Cyclophilin A Gene, PPIA, Accelerate Progression to AIDS

Ping An
National Cancer Institute at Frederick

Li Hua Wang
National Cancer Institute at Frederick

Holli Hutcheson-Dilks
National Cancer Institute at Frederick

George Nelson
National Cancer Institute at Frederick

Sharyne Donfield
Rho, Inc.

See next page for additional authors

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

Part of the Genetics and Genomics Commons, Immunology and Infectious Disease Commons, and the Medicine and Health Sciences Commons

NSUWorks Citation
An, Ping; Li Hua Wang; Holli Hutcheson-Dilks; George Nelson; Sharyne Donfield; James J. Goedert; Charles Rinaldo; Susan Buchbinder; Gregory D. Kirk; Stephen J. O'Brien; and Cheryl Winkler. 2007. "Regulatory Polymorphisms in the Cyclophilin A Gene, PPIA, Accelerate Progression to AIDS." PLoS ONE 3, (6 e88): 849-857. https://nsuworks.nova.edu/cnso_bio_facarticles/780

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Authors

This article is available at NSUWorks: https://nsuworks.nova.edu/cnso_bio_facarticles/780
Regulatory Polymorphisms in the Cyclophilin A Gene, *PPIA*, Accelerate Progression to AIDS

Ping An1, Li Hua Wang2, Holli Hutcheson-Dilks3, George Nelson1, Sharyne Donfield4, James J. Goedert5, Charles R. Rinaldo6, Susan Buchbinder7, Gregory D. Kirk8, Stephen J. O’Brien9, Cheryl A. Winkler1*

Introduction

As an obligate intracellular parasite, HIV type 1 (HIV-1) utilizes host cell factors for its replication. Human cyclophilin A (CypA), also known as peptidyl prolyl isomerase A (*PPIA*), is a ubiquitous cytoplasmic protein (by convention, we refer to the protein as CypA and the gene as *PPIA*). CypA has long been known for its incorporation into HIV-1 virions and its important role in facilitating HIV-1 replication in host cells [1,2]. CypA is a member of the cyclophilin family, members of which all possess peptidyl-prolyl *cis/trans* isomerase activity. Peptidyl prolyl *cis/trans* isomases (*PPIAses*) catalyze the *cis/trans* isomerization of prolyl peptide bonds and are believed to be involved in protein folding [3]. The incorporation of CypA into the HIV-1 virion capsid is mediated through the direct binding between prolyl peptide bond located in a proline-rich loop of the fourth and fifth helices of the HIV-1 capsid and the active sites of CypA [1,5]. Disruption of CypA incorporation, either by HIV-1 Gag mutations or by cyclosporine A, an immunosuppressive drug that prevents HIV-1 Gag binding to CypA, leads to an attenuation of HIV-1 infectivity [2,6]. Braaten and Luban found that HIV-1 replication was decreased in CypA-null human CD4+ T cells, in which the gene encoding CypA (*PPIA*) was deleted through homologous recombination [7]. CypA is therefore an important host factor that regulates HIV-1 replication.

Recently, the role of CypA in HIV-1 has gained even greater attention with the discovery of a fusion protein of CypA and TRIM5α, a host restriction factor against HIV-1 [8], which confers HIV-1 resistance in owl monkey [9–11]. Both TRIM5α and CypA recognize and act on the capsid of HIV-1, but apparently confer opposite effects. TRIM5α restricts HIV-1 by promoting premature disassembly of HIV-1 capsid [12], while CypA increases viral infectivity by facilitating proper uncoating. Although the interaction between CypA and TRIM5α is still unclear, it appears that the modulation of HIV-1 infectivity by CypA is independent of TRIM5α [11,13–15]. It has been postulated that binding of CypA to capsid protects HIV-1 from an unknown restriction factor in humans [15].

The study of the influence of human gene variation on susceptibility to HIV-1 infection and disease progression is an approach that may reveal the in vivo host factor HIV-1 interactions and their epidemiologic importance at the population level. With this approach we have identified several AIDS-modifying variants in genes TRIM5 [16],


Received March 9, 2007; Accepted May 11, 2007; Published June 22, 2007

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Abbreviations: AA, African American; CI, confidence interval; CypA, cyclophilin A; EA, European American; HIV, HIV type 1; HRIU, high-risk exposed uninfected; LD, linkage disequilibrium; *PPIA*, peptidylprolyl isomerase A; RH, relative hazard; SC, seroconverter; SN, seronegative; SNP, single nucleotide polymorphism; SP, serore prevalent*

* To whom correspondence should be addressed. E-mail: winkler@mail.nclcfcr.gov
Author Summary

Individual risk of acquiring HIV type 1 (HIV-1) infection and developing AIDS is not equal; some people are more prone to HIV/AIDS than others. Susceptibility to HIV-1/AIDS is likely determined by a combination of environmental, viral, and host genetic factors. Genetic variations in host cellular factors involved in HIV-1 cell entry, replication, and host defense have been found to affect susceptibility to HIV-1/AIDS. In this report, we focused on the gene PPIA that encodes cyclophilin A, a human cellular protein that is incorporated into the HIV-1 virion and promotes viral replication. We studied genetic variation in the PPIA gene in persons with different susceptibility levels to HIV-1 infection or different rates of disease progression. We found that individuals who processed two functional variants in the promoter region of PPIA had higher risk of CD4+ T-cell loss or progression to AIDS-defining diseases. We also observed that an additional variant occurred more frequently in HIV-1-infected individuals compared to HIV-1-exposed, but uninfected, individuals. These results suggest that genetic variation in PPIA may influence host susceptibility to HIV-1 infection or disease progression and targeting PPIA might provide therapeutic benefit.

APOBEC3G [17], and CUL5 [18] that encode human innate HIV-1 restriction factors or related proteins. In a recent study performed in the Swiss HIV Cohort Study, a variant of PPIA was implicated as a potential factor affecting HIV-1 disease progression [19]. As CypA is an essential human protein for completion of the HIV-1 life cycle, we have assessed the influence of genetic variation in the PPIA gene on HIV-1 infection and disease progression in five United States–based HIV-1 natural history cohorts.

Results

Description of PPIA Variations

The PPIA gene is approximately 7 kb in length, consisting of five exons (Figure 1A). Using PPIA-specific primers we resequenced virtually the entire PPIA gene to screen for single nucleotide polymorphisms (SNPs) in 92 European Americans (EA) and 92 African Americans (AA). Seven polymorphisms were discovered: three in the putative promoter region, one in the 5' UTR, one in the 3' UTR, one in intron 4, and one in the second Alu repeat region upstream of the putative promoter. No SNPs were found in the coding regions. These SNPs and four additional SNPs available from dbSNP database located in regions not detected by our sequencing were selected for genotyping in the AIDS cohorts (Figure 1 and Table 1). SNP1 genotypes deviated from the frequencies expected under the Hardy-Weinberg equilibrium (p < 0.001), probably due to its location in the Alu repetitive element; thus, SNP1 was excluded in the analysis. SNP2 was also excluded from analysis due to its rarity. The genotypic frequencies of all other SNPs conform to Hardy-Weinberg expectations.

Allele frequency distributions of PPIA SNPs differed in EA and AA with Fst values ranging from 0.03 to 0.40 (average 0.22). The difference is particularly pronounced for SNPs 7, 8,
Haplotypes of PPIA SNPs

The extent of linkage disequilibrium (LD) among eight SNPs was assessed by calculating all pairwise D’ values, separately for AA and EA (Figure 2A and 2B). Strong LD was observed among almost all SNPs in AA (D’ range 0.87–1.0) and even stronger LD in EA (D’ range 0.97–1.0). SNP3 and SNP4 were in perfect LD in both populations (D’ = 1). As each of these two SNPs carries the same information content, SNP4 was selected as a proxy for SNP3 in the analysis. The eight SNPs in both population groups formed a single LD block as defined by the solid spine of LD method ([20]). This provided a justification to use all eight SNPs spanning the entire region as one block for subsequent haplotype-based association analyses.

Among AA, SNP6 presented intermediate level of LD (D’ ~0.60) with the SNPs downstream (SNPs 8–11); SNP6–SNP8 recombinant G-C haplotype had a frequency of 5%. Among EA, SNP5 had weak LD with SNP6–11 (D’ range 0.47–0.71); the recombinant haplotype SNP5–SNP6 (G-G) had a frequency of 2.5%. This suggests an occurrence of recombination in the region between SNP5 and SNP8. The presence of multiple Alu elements in PPIA may have introduced the recombination event by promoting genomic instability ([21]).

Effect of PPIA SNPs and Haplotypes on Susceptibility to HIV-1 Infection

We compared the PPIA SNP allele and haplotype frequency distributions among three groups with increasing resistance to HIV-1 infection: seroconverters (SCs), seronegatives (SNs), and HREU, respectively (Mantel-Haenszel trend test, p = 0.21). The reduced carriage of SNP5G in HREU was significant when SCs were compared to HREU (odds ratio = 1.78, p = 0.02), suggesting that the SNP5G allele carriers may have increased susceptibility to HIV-1 infection (Table 2). No distortion of frequency distribution between risk groups were observed for all other SNPs in AA or EA (unpublished data).

Effect of SNP4 on AIDS Progression

Among AA SCs, the minor allele of SNP4 (the G allele) was associated with accelerated loss of CD4+ T cells (relative

Table 2. Effects of PPIA SNP4 and SNP5 on HIV-1 Infection Risk in EA

<table>
<thead>
<tr>
<th>SNP</th>
<th>Risk Group</th>
<th>n</th>
<th>Genotype</th>
<th>Number of Carriers (Frequency)</th>
<th>OR (95% CI)a</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP4</td>
<td>SC</td>
<td>659</td>
<td>CC</td>
<td>574 (0.871)</td>
<td>2 (0.003)</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>83 (0.126)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>2 (0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC + GG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>447</td>
<td>CC</td>
<td>392 (0.877)</td>
<td>2 (0.005)</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>53 (0.119)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>1 (0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>155</td>
<td>CC</td>
<td>130 (0.839)</td>
<td>1.05 (0.73–1.52)</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CG</td>
<td>23 (0.155)</td>
<td>0.77 (0.47–1.25)</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>1 (0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP5</td>
<td>SC</td>
<td>652</td>
<td>AA</td>
<td>485 (0.744)</td>
<td>12 (0.002)</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>155 (0.238)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>15 (0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SN</td>
<td>603</td>
<td>AA</td>
<td>458 (0.76)</td>
<td>1.09 (0.84–1.41)</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>130 (0.216)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HREU</td>
<td>153</td>
<td>AA</td>
<td>128 (0.837)</td>
<td>1.76 (1.11–2.80)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>21 (0.137)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td>4 (0.003)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aOR represents the odd ratio of being an HIV-1-SC compared to being either HIV-1-SN or HREU.

n, number of individuals.

doi:10.1371/journal.ppat.0030088.t002
hazard (RH) = 3.7, 95% confidence interval [CI] = 1.59–8.63) in a Cox proportional hazards model (Table 3). After stratifying by cohort in the Cox regression analysis, this association became slightly stronger (RH = 4.08, 95% CI = 1.71–9.70). Kaplan-Meier survival analysis presented a clear separation of curves stratified for the SNP4 C/C and C/G genotypes (G/G absent) on progression to CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b},}\textsubscript{\textgreek{c}} 200 in AA (p = 0.002, log-rank test, Figure 3). Notably, all SNP4 carriers progressed to CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b},}\textsubscript{\textgreek{c}} 200 within 6 y. The effect on AIDS-1987 was in the same direction but short of significance (Table 3).

To affirm the observed genetic influence of SNP4, we performed a categorical analysis of HIV-1-infected people that included seroprevalents (SPs) in the long-term survivors (Figure 4). SPs who had remained AIDS-free for greater than 7.5 y for AA or 10 y for EA after study enrollment (the median time to AIDS for each) were included in this analysis. The frequency distributions of SNP4 in six discrete time intervals progressing to disease outcome after HIV-1 infection were tested for statistical trend using a Mantel-Haenszel test. As shown in Figure 4, the SNP4 G allele occurred more frequently among fast progressors in both AA and EA. With an increased sample size of 389 patients in AA, a significant trend toward rapid CD4 loss is still preserved (p = 0.04, Figure 4A). Among 970 EA patients, the trend toward rapid progression to AIDS was more pronounced (p = 0.004) (Figure 4B). Thus, the results from SC plus SP groups corroborate the findings from the SC-alone survival analyses, indicating that the accelerating effect of SNP4 is consistent in both AA and EA populations.

### Table 3. Effects of PPIA SNPs on AIDS Progression in AA and EA

<table>
<thead>
<tr>
<th>SNP Number</th>
<th>Endpoint</th>
<th>AA (n = 284–290)</th>
<th></th>
<th></th>
<th>EA (n = 643–650)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RH</td>
<td>95% CI</td>
<td>p</td>
<td>RH</td>
<td>95% CI</td>
<td>p</td>
</tr>
<tr>
<td>SNP4(SNP3)</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>3.7</td>
<td>1.59–8.63</td>
<td>0.003</td>
<td>1.19</td>
<td>0.88–1.62</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>2.03</td>
<td>0.63–6.53</td>
<td>0.24</td>
<td>1.34</td>
<td>0.97–1.85</td>
</tr>
<tr>
<td>SNP5</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>0.67</td>
<td>0.44–1.01</td>
<td>0.06</td>
<td>0.95</td>
<td>0.74–1.22</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>0.86</td>
<td>0.50–1.48</td>
<td>0.58</td>
<td>1.26</td>
<td>0.98–1.64</td>
</tr>
<tr>
<td>SNP6</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>1.16</td>
<td>0.78–1.74</td>
<td>0.46</td>
<td>1.16</td>
<td>0.92–1.45</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>1.20</td>
<td>0.71–2.04</td>
<td>0.49</td>
<td>1.05</td>
<td>0.82–1.34</td>
</tr>
<tr>
<td>SNP7\textsuperscript{b}</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>1.04</td>
<td>0.60–1.82</td>
<td>0.88</td>
<td>1.08</td>
<td>0.85–1.37</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>0.82</td>
<td>0.40–1.67</td>
<td>0.58</td>
<td>1.32</td>
<td>1.03–1.70</td>
</tr>
<tr>
<td>SNP8</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>0.72</td>
<td>0.39–1.34</td>
<td>0.30</td>
<td>0.95</td>
<td>0.74–1.22</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>0.64</td>
<td>0.29–1.43</td>
<td>0.28</td>
<td>1.26</td>
<td>0.98–1.64</td>
</tr>
<tr>
<td>SNP11\textsuperscript{c}</td>
<td>CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200</td>
<td>0.88</td>
<td>0.49–1.60</td>
<td>0.68</td>
<td>1.08</td>
<td>0.85–1.37</td>
</tr>
<tr>
<td></td>
<td>AIDS-87</td>
<td>0.84</td>
<td>0.38–1.87</td>
<td>0.68</td>
<td>1.32</td>
<td>1.03–1.70</td>
</tr>
</tbody>
</table>

The Cox proportional hazards model was employed to test the impact of SNPs on the rates of progression to CD4\textsubscript{\textgreek{a},}\textsubscript{\textgreek{b}} 200 or to AIDS clinical diseases (AIDS-87); p-values were from a Wald test.

aSNP3 and SNP4 are in perfect LD.
bSNP7 is representative of SNP8, SNP9, and SNP11 in EA.
cSNP9 is representative of SNP11 in AA.

n, number of individuals.

doi:10.1371/journal.ppat.0030088.t003

**PPIA SNPs and AIDS Progression**

In a Swiss Caucasian HIV-1 cohort, SNP5G, referred to as 1650A/G [19], was reported to be associated with a rapid CD4 cell depletion and with a trend toward higher in vitro HIV-1 replication [19]. In our analysis of EA SCs, SNP5G showed a nonsignificant trend toward rapid progression to AIDS-1987 (RH = 1.26, p = 0.08), but had no effect on CD4 cell depletion, while an opposite trend was observed in AA (Table 3). Among 970 EA patients, the trend toward rapid progression to AIDS was more pronounced (p = 0.004) (Figure 4B). Thus, the results from SC plus SP groups corroborate the findings from the SC-alone survival analyses, indicating that the accelerating effect of SNP4 is consistent in both AA and EA populations.

**Effects of PPIA Haplotypes on AIDS Progression**

In the haplotype analysis, we first used the Cox regression model to test the global null hypothesis of no association between haplotype frequency and AIDS progression, by comparing a model with all haplotypes to the covariants-
HIV-1 infection but not progression [22]. In the Cox regression model analysis, we observed no obvious interaction for infection or progression (p > 0.15, unpublished data).

**Interactions between TRIM5 and PPIA Variants**

Because TRIM5α interferes with and CypA facilitates post-entry uncoating of the HIV-1 capsid, it is possible that there may be genetic interactions between two genes. Potential interactions were tested only between variants in two genes with functional plausibility or high strength of association. Specifically, PPIA SNP4 was tested for its interaction with TRIM5-rs16934386 and TRIM5-R136Q, which was previously reported using the same patient population to be associated with HIV-1 infection but not progression [22]. In the Cox regression model analysis, we observed no obvious interaction for infection or progression (p > 0.15, unpublished data).

This suggests that CypA and TRIM5α likely act independently, consistent with recent in vitro observations [11,13–15,29].

**Electrophoretic Mobility-Shift Assay**

As SNP3, 4, and 5 reside in the regulatory region of PPIA, we performed gel shift assays to assess whether these SNPs differentially bind to transcriptional factors (Figure 5). In this experiment, ~25-bp DNA probes containing either of the SNP alleles were incubated with the nuclear extracts from Th1 cytokine-stimulated human T lymphocytes. Both SNPs showed differential binding affinity to nuclear protein(s) with the same mobility (Figure 5A, top band). Averaged from two independent experiments, a 4-fold decrease or 2.8-fold increase in the band density was observed for the minor allele of SNP3 (G) and SNP4 (G), respectively. The cold unlabeled oligonucleotides with minor alleles were unable to compete with wild-type oligonucleotides, suggesting the bindings are specific. These results indicate that these two SNPs within the promoter alter the binding affinity with certain transcriptional protein(s). On the contrary, no obvious differential binding was observed for SNP5 (Figure 5A).

In silico analysis of the sequence at these three SNP sites by using TESS software also predicted differential binding for SNP3 and 4, but not for SNP5. SNP3 and 4 were each located in one of the consensus-binding motifs of transcription factor SP1, as shown in Figure 5B. The local sequence with SNP3C was identical to the consensus sequence of SP1 site, GGAGGC, whereas SNP3C>G change perturbs this motif. In contrast, the sequence with SNP4C increased the mismatch to one, predicting a stronger binding. There is, therefore, a plausible biochemical basis for the differential binding affinity observed in the gel shift experiment. These results also suggest that SNP3 and SNP4 affect transcription, likely through their interactions with SP1. The sequence of the region of human PPIA promoter resides in by SNP3 and SNP4 was compared to the corresponding sequence from other species (Figure 5C). SNP3 and SNP4 are both located in the highly conserved regions. The C allele of SNP3 is present in silico analysis of the sequence at these three SNP sites by using TESS software also predicted differential binding for SNP3 and 4, but not for SNP5. SNP3 and 4 were each located in one of the consensus-binding motifs of transcription factor SP1, as shown in Figure 5B. The local sequence with SNP3C was identical to the consensus sequence of SP1 site, GGAGGC, whereas SNP3C>G change perturbs this motif. In contrast, the sequence with SNP4C increased the mismatch to one, predicting a stronger binding. There is, therefore, a plausible biochemical basis for the differential binding affinity observed in the gel shift experiment. These results also suggest that SNP3 and SNP4 affect transcription, likely through their interactions with SP1. The sequence of the region of human PPIA promoter resides in by SNP3 and SNP4 was compared to the corresponding sequence from other species (Figure 5C). SNP3 and SNP4 are both located in the highly conserved regions. The C allele of SNP3 is present
in both primates and rodents. The C allele of SNP4 occurs only in primates and the minor allele G occurs in rodents, suggesting a possible association with speciation. These interspecies data support the hypothesis that SNP3 and SNP4 changes may have functional consequences.

Discussion

The role of CypA in promoting HIV-1 replication has been well established through extensive in vitro experiments [1,2,6,7]. In this study, we have undertaken a systematic investigation of the association between genetic variations in the \textit{PPIA} gene and susceptibility to HIV-1 infection and disease progression in HIV-1 natural history cohorts. We found that two promoter variants in perfect LD, SNP3 and SNP4, in \textit{PPIA}, were associated with rapid HIV-1 disease progression. This effect was revealed for CD4 cell depletion or for AIDS progression, reflecting early or late stage of disease progression, in AA and EA, respectively. The racial difference in the strength or timing of the effects may be due to interactions with other sequence variations that are distributed differently in two populations or to immunologic and viral variables. We further found that both SNPs had differential binding efficiency to nuclear proteins in the gel shift experiment. These findings suggest that the functional promoter SNP3/SNP4 influences HIV-1 disease progression.

\textit{PPIA} in human encodes a 165-amino acid protein that is highly conserved across species, with 100% and 96% sequence identity to rhesus monkey and mouse, respectively. The nucleotide sequence of \textit{PPIA} is 98% identical between human and chimpanzee. No variation in the coding sequence of human \textit{PPIA} was found in this study or in the dbSNP database. This suggests that differential genetic impact of \textit{PPIA} SNPs and haplotypes on infection and progression are likely due to regulatory region variation affecting protein expression. The regulation and promoter function of human \textit{PPIA} is still undefined. \textit{PPIA} is highly expressed in almost all cell types and is considered a housekeeping gene. Sequence analysis revealed that \textit{PPIA} promoter contains a TATA box and two SP1 binding sites and is highly GC-rich (70%) with overrepresented CpG dinucleotides [24]. In this study, we identified two SNPs, each located in two additional SP1 binding elements. SP1 is generally believed to function as a transactivating factor for many housekeeping genes [25]. These two SNP changes, SNP3C and SNP4G, always occurring together through LD, had opposite effects on the nuclear protein binding. Assuming the predicted stronger SP1 binding effect of SNP4 is dominant over the weaker SP1 binding of SNP3, a higher \textit{PPIA} gene expression would be predicted. This would be consistent with more rapid disease course as observed in this study.

Based on our analyses of eight SNPs covering the entire gene region of \textit{PPIA} and their haplotypes, we found that the \textit{PPIA} disease-modifying effects were most likely afforded by promoter SNP3 or SNP4 and the associations with the downstream SNPs 7–11 were largely due to tracking of these
two SNPs through LD. In a previous report, Bleiber et al. studied the ex vivo HIV-1 infectivity in CD4+ T cells from healthy individuals carrying the SNP4 and SNP5 alleles (named as PPIA-1604 and PPIA-1650, respectively, in the Bleiber’s report), as well as their impact on the CD4 gradient (from CD4+ T-cell counts 500/μm^2 to 200/μm^2) among SPs enrolled in the Swiss HIV cohort, largely comprised of Caucasians [19]. In the ex vivo experiment, both SNPs showed nonsignificant tendencies toward greater HIV-1 replication, with the effects of SNP5 being more pronounced. In the population study, both SNPs were associated with faster CD4+ T-cell depletion in a model also containing two other candidate genes (model 2); with the SNP5 of PPIA effect again being more pronounced. Thus, the detrimental effect of SNP4 observed from these two studies is largely consistent. In contrast, no evidence supporting a role of SNP5 in AIDS progression was found, although the SNP5 G allele was carried significantly less often in HREU than in SC but was similarly distributed in SC and SN. Thus, the role of CypA polymorphism in HIV-1 infection remains inconclusive and warrants further investigation. Taken together, the association and functional results from these two studies point to a role of PPIA genetic variation in HIV-1/AIDS. In future studies, it may be informative to determine the mechanism of association between PPIA variants and host susceptibility to HIV-1 infection or disease progression in view of the interaction between CypA and the capsid domain of the Gag polyprotein.

Population-based genetic association studies provide a powerful approach to uncovering host genes that confer disease susceptibility or resistance. This approach has led to identification of various genetic factors that affect HIV-1/AIDS, and has provided unique insights into the host HIV-1 interaction [26,27]. The successful identification of true genetic associations requires a large sample size (to minimize the type II error), replication in independent studies (to ward off type I error), and plausible functional evidence (to infer causal relationship). In this study, the plausible genetic-modifying role of PPIA is supported by replication in two independent ethnic groups and the demonstration that SNPs 3 and 4 are each associated with altered binding affinities to transcription factors. Moreover, the previous independent study also provides supportive association and functional evidence [19].

Our comprehensive analysis of nearly all variant alleles and their haplotypes in the PPIA gene is, in essence, equivalent to a gene-based approach [28]. A major problem facing genetic association studies is the difficulty in replication, particularly for single SNP associations. One reason for failure to replicate previous studies is due to potential differences in allele frequencies and LD structure across populations. Gene-based replication has recently been advocated as a gold standard for replication studies [28]. The entire gene with prior SNP association is considered the functional unit and is examined for association with effectively all genetic variation in the gene. Nonreplication due to population differences may be minimized as local allele frequencies and LD structure from study populations are considered with this approach [28]. Thus, our comprehensive survey of PPIA genetic variation and LD structure may facilitate future comparisons of replication studies using a gene-based approach.

In summary, through genetic epidemiological and functional approaches, we have identified two promoter SNPs in PPIA as potential genetic modifiers of HIV-1 disease progression. Our findings corroborate the notion that genetic variation of PPIA influences AIDS pathogenesis and provide in vivo evidence that CypA is a critical host protein in interaction with HIV-1. Manipulation of PPIA may be considered as a plausible option for anti-HIV-1 therapeutic development as previously explored [29].

Materials and Methods

Study participants. Study participants were enrolled in five United States–based natural history HIV/AIDS cohorts. AIDS Link to the Intravenous Experience (ALIVE) is a community-based cohort of intravenous injection drug users in Baltimore enrolled in 1988–1989 [30], consisting of 92% AA; Multicenter AIDS Cohort Study (MACS) is a longitudinal prospective cohort of men who have sex with men (MSM) from four United States cities: Chicago, Baltimore, Pittsburgh, and Los Angeles, enrolled in 1984–1985 [31], consisting of 83% EA and 10% AA; the San Francisco City Clinic Study (SFCC) is a cohort of MSM originally enrolled in a hepatitis B study in 1978–1980 [32], consisting of 96% EA; Hemophilia Growth and Development Study (HGDS) is a multicenter prospective study that enrolled children with hemophilia who were exposed to HIV-1 through blood products between 1982 and 1983 [33], consisting of 72% EA and 11% AA; the Multicenter Hemophilia Cohort Study (MHCS) is a retrospective study that enrolled persons with hemophilia [34], consisting of 90% EA and 6% AA. The participant group is comprised of HIV-1 SCs (infected after study enrollment), SPs (infected at study enrollment), SNs, and HREU. The MACS, MHCS, SFCC, and ALIVE cohorts consist of both SCs and SPs among HIV-1-infected individuals. Due to the potential frailty bias (missing the most rapid progressors to AIDS and death) among SPs, only SCs from these four cohorts were used in the survival analysis. SPs were also included for allele frequency estimation, haplotype inference, and disease categorical analysis. The number of participants studied in each risk or disease category: SC = 604 EA, 295 AA; SN = 604 EA, 350 AA; HREU = 153 EA, 81 AA; and SP = 1199 EA, 773 AA. Of 290 AA SCs, 237, 42, 5, and 5 were from ALIVE, MACS, MHCS, and HGDS, respectively.

The date of seroconversion after study enrollment was estimated as the midpoint between the last seronegative and first seropositive HIV-1 antibody test; only individuals with less than 2 y elapsed time between the two tests were included in the seroconverter progression analysis. The censoring date was the earliest of the date of the last recorded visit, or December 31, 1995 for the MACS, MHCS, HGDS, and SFCC or July 31, 1997 for the ALIVE cohort to avoid potential confounding by highly effective anti-retroviral therapy (HAART). The censoring date was extended in the ALIVE cohort because of delayed uptake of HAART in this group [30,35].

HIV-1-uninfected individuals were classified into two categories based on individual’s documented exposure levels to HIV-1. HREU individuals were those 80 AA and 145 EA with documented HIV-1 exposure through sharing of injection equipment [36], who had anal receptive sex with multiple partners [37], or transfusions with Factor VIII clotting factor prior to 1984, when heat treatment was initiated [38]. SN individuals (n = 420 and 571, respectively, for AA and EA) are those enrolled in the cohorts who remained HIV-seronegative despite ongoing or prior risk activity. The study protocols were approved by the Institutional Review Boards of participating institutions and informed consent was obtained from all study participants.

Identification of SNPs. Nucleotide polymorphisms were discovered in a panel of 92 EA and 92 AA, representing the extremes of the distribution for rapid and slow progression to AIDS and HREU. A nonsistopic RNA cleavage assay following PCR was employed to screen for polymorphisms [39]. PPIA has a high degree of homology to other processed pseudogenes that varies from 75% to 95% in the exon regions. PPIA also contains six copies of Alu repeats, each with a length of approximately 300 bp, located in the immediate upstream of the putative promoter region and introns [24]. Sequence comparison and BLAST search were performed to select PPIA-specific PCR primers. Overlapping primers covered nearly the entire PPIA gene region including the putative promoter region, 5’ and 3’ UTRs, all five exons, as well as the Alu repeat regions. A part of intron 1 was not covered due to the high GC content. Primer sequences are presented in Table S1 and are numbered according to the GenBank
DNA sequence X52851. Additional intronic SNPs were selected from NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP) and HapMap databases (http://www.hapmap.org), by considering location, spacing, and allele frequency at least 10%. Haplotype tagging (ht)SNPs were given preference in the SNP selection. The ancestral allele state of the SNPs was determined using chimpanzee sequence data.

**Genotyping of SNPs.** Genotyping was performed using PCR-restriction fragment length polymorphism (PCR-RFLP) assay or TaqMan assays. PCR-RFLP was carried out with 35 cycles of denaturing at 94°C for 30 s, annealing at 60°C for 30 s, and extending at 72°C for 45 s. The PCR product was digested with respective restriction enzymes (New England Biolabs, http://www.neb.com) overnight and then separated on 3% agarose gels. TaqMan assays were obtained from the Assay-by-Demand service of Applied Biosystems (http://www.appliedbiosystems.com). Genotyping primers and conditions were presented in Table S2. Eight water controls were included for each PCR to monitor the potential PCR contamination, and 10% of SC and HREU samples were genotyped twice. The genotypes obtained were free of water contamination or of inconsistencies between duplicates.

**Statistical analysis.** To assess the difference in allele frequency distribution in two populations, we performed a contingency χ² test for each marker to test the null hypothesis that the allele frequencies are the same in the two populations. FS values were estimated by Wier and Cookerham's method [40]. 

Pairwise D' was calculated using the absolute value of D'. Absolute values of D' range from 0 for independence to 1 for complete LD between the pairs of loci. LD plots were generated utilizing Haploviev (http://www.broad.mit.edu/mpg/haploviev) [20]. A triangular matrix of D' value was used to demonstrate LD patterns within AA and EA. Haplotype blocks were estimated using the solid spine of LD method [41]. Haplotype blocks were defined with a default threshold of 0.8 based on confidence intervals of D' [41], or the solid spine of LD method, which creates blocks of SNPs that have contiguous pairwise D' values of greater than 0.8. With the latter method, the first and last SNPs in a block are in strong LD with all intermediate SNPs, but the intermediate SNPs are not necessarily in LD with each other [20]. Eight SNP haplotype frequencies were inferred separately for each population by means of an expectation-maximization algorithm [42].

Association analyses were conducted using the statistical packages SAS (version 9.0, SAS Institute, http://www.sas.com) and GA2 (version 2.2; http://ga2www.uwaterloo.ca). We used chi square or Fisher's exact test to compare genotypic frequencies between HIV-1 HREU and HIV-1 SC participants, regardless of the exposure route, persons in the HREU or SC groups were at risk to HIV-1 infection based on their inclusion in a HIV-1 risk group; therefore, we combined participants across cohorts to achieve a reasonable statistical power.

**Kaplan-Meier survival statistics and the Cox proportional hazards model (Cox model) were used to assess the effects of SNPs and haplotypes on the rate of progression to AIDS. Two separate endpoints reflecting advancing AIDS pathogenesis were considered for Scs: (1) HIV-1 infection plus a decline of CD4 T-cell counts <200 cells/μm³ (CD4 <200); (2) the 1987 Center for Disease Control definition of AIDS (AIDS-87): HIV-1 infection plus AIDS-defining illness [43].** The significance of genotypic associations and RH was determined by unadjusted and adjusted Cox model regression analyses for each SNP, we compared the minor allele genotypes to the most common genotype as a reference group. All p-values were two-tailed. Genetic factors previously shown to affect progression to AIDS were pre-determined to be included as confounding covariates in the Cox model analysis: CCR5Δ32, CCR2-64I, CCR5-P1, HLA-B*27, HLA-B*57, HLA-A*31, HLA-A*26, HLA-B*3502, HLA-B*3503, HLA-B*3504, and HLA-B*501, and HLA Class I homozygosity for AA (reviewed in [26,27]); HLA-B*57 and HLA Class I homozygosity for AA. CCR2-64I, HLA-B*27, and HLA-B*55Px were not considered as covariates in AA due to no or weak effects in the AA participants, and CCR5Δ32 was considered due to its rarity in AA. Analyses were stratified by sex and age at seroconversion: 0–20, >20–40, and >40 y [27]. Further stratification by cohort was also performed for the exploratory analyses. In the stratified Cox regression model, the overall log-likelihood of hazards obtained is the sum over strata of the stratum-specific hazards, as estimated by the method of partial likelihood. As the same criteria for determining strata (seroconversion date) and endpoints and the similar sampling strategy and follow-up settings were used across cohorts, we combined SCs from all cohorts for the survival analysis to increase the power. Although these cohorts differ in routes of HIV-1 transmission, no appreciable effect of mode of infection on AIDS progression has been found through re-analysis of more than ten thousands of SCs from 38 studies around the world (including three used in this study) [44].

To test the association of PP1A haplotypes and HIV-1 disease progression, we first performed a global test in the Cox regression model for each of two disease outcomes separately for AA and EA. The global null hypothesis is that the odds ratios of all haplotypes are equal between cases and controls. Likelihood ratio tests were used to compare a full model with all haplotypes and a base model with only covariates. When the significance of the global test exceeded a relaxed nominal model, p < 0.10, the associations of individual haplotypes were further tested.

**Electrophoretic mobility-shift assay.** Cell culture and an electrophoretic mobility-shift assay were performed as described [46]. Freshly explanted human T lymphocytes were obtained from normal donors or HIV-1-infected, and activated for 24 h with 1 μg/ml PMA in RPMI 1640 medium containing 10% FCS (Sigma, http://www.sigmaaldrich.com), 2 μM L-glutamine, and penicillin-streptomycin (50 IU/ml and 50 mg/ml, respectively). T lymphocytes were made quiescent by washing and incubating for 24 h in RPMI 1640 medium containing 1% FCS before exposure to cytokines. Cells were then stimulated with 100 nmol/ml IL-2 (PeproTech, http://www.peprotechec.com) or 100 nmol/ml human rIL-2 (Hoffmann-La Roche, http://www.roche.com) at 37°C for 10 min. Cell pellets were frozen at −70°C. The probe sequences were 5'-ttgccccggggCcgattgctg-3' for SNP3G and ttcggggggggCgaaacctgtttg-3' for SNP3G, 5'-gctgccgacCagcttcatgctt-3' for SNP4G, and 5'-gctccgggagCgacgcctg-3' for SNP4G allele; 5'-agggaaacccgCgcttcatgct-3' for SNP5A and 5'-agggaaacccgCgcttcatgct-3' for SNP5G (SNP site is capped). In cold oligonucleotide competition assay, 100-fold excess of cold unlabeled probe was added as a competitor. The band density was measured by using the software ImageJ [http://rsb.info.nih.gov/ij/index.html]. The sequence in and around the SNP sites in the promoter region were searched for the presence of transcription factor binding sites using the program TESS: Transcription Element Search System (http://www.cbil.upenn.edu/tess). Sequence alignment was performed using program ClustalW (http://www.ebi.ac.uk/clustalw).

**Supporting Information**

**Table S1. Primers Used in the Resequencing of the PP1A Gene**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP3G</td>
<td>For SNP3C and ttgcgggcgggCgacagcctgtt-3' (SNP3G)</td>
<td>F: 5'-ttgccccggggCcgattgctg-3', R: 5'-aggaaaaccgtgtActattagc-3'</td>
</tr>
<tr>
<td>SNP4G</td>
<td>For SNP4C and ttgcgggggggCgaaacctgtttg-3' (SNP4G)</td>
<td>F: 5'-gctgccgacCagcttcatgctt-3', R: 5'-gctccgggagCgacgcctg-3'</td>
</tr>
<tr>
<td>SNP5A</td>
<td>For SNP5A and 5'-agggaaacccgCgcttcatgct-3' (SNP5G)</td>
<td>F: 5'-agggaaacccgCgcttcatgct-3', R: 5'-agggaaacccgCgcttcatgct-3'</td>
</tr>
</tbody>
</table>

**Table S2. Genotyping Methods for PP1A SNPs**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP3G</td>
<td>For SNP3C and ttgcgggcgggCgacagcctgtt-3' (SNP3G)</td>
<td>F: 5'-ttgccccggggCcgattgctg-3', R: 5'-aggaaaaccgtgtActattagc-3'</td>
</tr>
<tr>
<td>SNP4G</td>
<td>For SNP4C and ttgcgggggggCgaaacctgtttg-3' (SNP4G)</td>
<td>F: 5'-gctgccgacCagcttcatgctt-3', R: 5'-gctccgggagCgacgcctg-3'</td>
</tr>
<tr>
<td>SNP5A</td>
<td>For SNP5A and 5'-agggaaacccgCgcttcatgct-3' (SNP5G)</td>
<td>F: 5'-agggaaacccgCgcttcatgct-3', R: 5'-agggaaacccgCgcttcatgct-3'</td>
</tr>
</tbody>
</table>

**Accession Numbers**

The Entrez Gene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene) accession numbers for the genes mentioned in this paper are PP1A (5478) and TRIM5 (85363). The GeneBank (http://www.ncbi.nlm.nih.gov) accession numbers for the human and chimpanzee PP1A genomic sequence are X52851 and NW_001237949, respectively. The Transfac binding site (http://www.cbi.upenn.edu/cgi-bin/tess) accession number is R04195 and R03868 for the SP1 factor.

**Acknowledgments**

We thank Yuchun Zhou and Beth Binns-Roemer for excellent technical assistance.

**Author contributions.** PA, SJO, and CA conceived and designed the experiments and wrote the paper. PA, LHW, and HHD performed the experiments and wrote the paper.
the experiments. PA and CAW analyzed the data. GN, SD, JG, CRR, SB, GDK, and SJO contributed reagents/materials/analysis tools.

Funding. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. This research was supported in part by the Intramural Research Program of the National Institute of Mental Health, National Cancer Institute, Center for Cancer Research, and National Human Genome Research Institute.

References


The authors have declared that no competing interests exist.

Competing interests. The authors have declared that no competing interests exist.

The National Institute on Drug Abuse, National Institutes of Health, under grant DA-04334 provided funding for the ALIVE cohort. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.