Background

Acquired Immunodeficiency Syndrome (AIDS) caused by the human immunodeficiency virus (HIV) is one of the major infectious diseases with 34 million people affected worldwide [1]. The introduction of combination antiretroviral treatment converted HIV/AIDS from a deadly disease into a chronic infection [2]. However, short- and long-term side-effects often challenge life-long antiviral treatment, prompting the need for new therapeutic strategies. This search has mainly been focussed on targeting viral enzymes as reverse transcriptase, protease and more recently integrase [3]. Different steps of the HIV-1 cycle are tightly dependent on cellular factors and accordingly host factors have become potential targets for HIV drug development. As an example, a first drug targeting the CCR5 co-receptor was approved for treatment of HIV-1 infection [4]. Further unravelling the host – virus interaction not only leads to better understanding of viral dynamics and disease progression, but also

Abstract

Background: As Lens epithelium-derived growth factor (LEDGF/p75) is an important co-factor involved in HIV-1 integration, the LEDGF/p75-IN interaction is a promising target for the new class of allosteric HIV integrase inhibitors (LEDGINs). Few data are available on the genetic variability of LEDGF/p75 and the influence on HIV disease in vivo. This study evaluated the relation between LEDGF/p75 genetic variation, mRNA expression and HIV-1 disease progression in order to guide future clinical use of LEDGINs.

Methods: Samples were derived from a therapy-naive cohort at Ghent University Hospital and a Spanish long-term-non-progression cohort. High-resolution melting curve analysis and Sanger sequencing were used to identify all single nucleotide polymorphisms (SNPs) in the coding region, flanking intronic regions and full 3’UTR of LEDGF/p75. In addition, two intronic tagSNPs were screened based on previous indication of influencing HIV disease. LEDGF/p75 mRNA was quantified in patient peripheral blood mononuclear cells (PBMC) using RT-qPCR.

Results: 325 samples were investigated from patients of Caucasian (n = 291) and African (n = 34) origin, including Elite progressor cohort. High-resolution melting curve analysis and Sanger sequencing were used to identify all single nucleotide polymorphisms (SNPs) in the coding region, flanking intronic regions and full 3’UTR of LEDGF/p75. In addition, two intronic tagSNPs were screened based on previous indication of influencing HIV disease. LEDGF/p75 mRNA was quantified in patient peripheral blood mononuclear cells (PBMC) using RT-qPCR.

Conclusions: LEDGF/p75 is a highly conserved protein. Two non-coding polymorphisms were identified indicating a correlation with disease outcome, but further research is needed to clarify phenotypic impact. The conserved coding region and the observed variation in LEDGF/p75 expression are important characteristics for clinical use of LEDGINs.


Editor: Mathias Lichterfeld, Massachusetts General Hospital, United States of America

Received August 13, 2012; Accepted October 16, 2012; Published November 30, 2012

Copyright: © 2012 Messiaen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was partly supported by the Flemish Agency for Innovation by Science and Technology (CellCoVir - IWT file nr 60813). Linos Vandekerckhove is supported by the National Fund for Scientific Research – Belgium as Principal Investigator. The Spanish RIS cohort and HIV BioBank are integrated in the Spanish AIDS Research Network supported by Instituto de Salud Carlos III (Grant RD06/0006/0035) and Fundación para la Investigación y Prevención del SIDA en España (FIPSE). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: All authors have declared that no competing interests exist. Prof Dr Bruno Verhasselt serves as a PLOS ONE Academic editor. This does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: Linos.Vandekerckhove@ugent.be

Background

1 HIV Translational Research Unit, Department of General Internal Medicine and Infectious Diseases, Ghent University Hospital, Ghent, Belgium, 2 Unidad de Inmunopatología del SIDA, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain, 3 Center for Medical Genetics, Ghent University, Ghent, Belgium, 4 Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium, 5 Hospital General d’Elx, Alicante, Spain, 6 Centro Sanitario Sandoval, Madrid, Spain
paves the way to validate new targets in the treatment of HIV-1 [5]. An increasing number of co-factors is being studied in the course of new strategies against HIV-1 infection [6].

An important and limiting step during cell infection is the integration of proviral DNA into host cell DNA, catalyzed by the viral enzyme integrase (IN). This requires the interaction with Lens epithelium-derived growth factor p75 (LEDGF/p75). By interacting with the catalytic core domain of IN, LEDGF/p75 functions as a chromatin tethering factor for the pre-integration complex, and targets HIV-1 integration towards actively transcribed genomic regions [7,8,9,10,11,12,13,14]. LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF) related protein family (HRP-family) known as transcriptional co-activators for heat-shock and stress-related genes [15,16,17,18,19,20]. It plays a role as anti-apoptotic protein in diverse oncologic settings [21,22]. The essential role in HIV-1 replication has been elucidated by mutagenesis, RNA interference, transdominant expression of protein domains and knock-out experiments [23,24,25,26,27]. Recently, allosteric inhibitors of the LEDGF/p75 - IN interaction (LEDGIN) were developed as a new class of antiretroviral treatment [28]. In vitro data show that LEDGINs also block the interaction between IN and HRP2, a second host protein of the HRP-family that contains a structurally similar integrase binding domain (IBD) and can substitute for HIV integration in cell lines with LEDGF/p75 knock-down [29].

Several Genome-wide association studies explored the influence of human genetic variation on HIV-1 replication and disease progression [30,31,32,33,34]. Genetic variants in genes associated with HLA-B*57:01 and the HLA-C gene and disease progression [30,31,32,33,34]. Genetic variants in influence of human genetic variation on HIV-1 replication [28].

African cohorts. They showed that the minor allele of one tagSNP was associated with higher CD4 T-cell count, lower viral load below 10,000 copies/ml in the absence of therapy. The slope of CD4 decline and the average plasma HIV viral load were determined for all patients based on at least four CD4+ T-cell counts and plasma HIV RNA measurements on samples collected with minimum three months' time interval during the therapy-naïve follow-up period. Classification of patients according to disease progression was based on broadly applied clinical definitions [39,40,41,42,43]: LTNP elite controllers (LTNP-EC) are long-term non-progressors (10 year follow-up, CD4 >500 cells/µl) maintaining an undetectable viral load without antiretroviral therapy (n = 48), LTNP viremic controllers (LTNP-VC) are long-term non-progressors harboring less than 2000 HIV RNA copies/ml without therapy in 75% of the measurements (n = 63), viremic non-controllers (LTNP-NC) are long-term non-progressors with a viral load between 2000 and 10,000 copies/ml (n = 66). Rapid progressors (RP) are non-LTNP patients with CD4 decline of more than 100 cells/µl per year (n = 35) and normal progressors (NP) are non-LTNP with a CD4 decline less than 100 cells/µl per year (n = 113). Ethical approval was obtained from Ethics Committee of Ghent University Hospital (Reg nr B67020071646) and Instituto de Salud Carlos III (Ref CEI PI 33_2010-v3). All participants provided written informed consent.

PSIP1 genotyping

After gDNA extraction from whole blood (Blood mini kit, Qiagen), 23 fragments were amplified by PCR, spanning the complete coding region of PSIP1, 3’UTR and on average 25 basepairs of the flanking intronic regions (see Figure 1). Two intronic tagSNPs described by Madlala et al. were analyzed as well [37]. Primers were designed with Lightscan software (Idaho Technologies) and listed in Table S1. Single nucleotide polymorphisms were screened by high resolution melting curve (HRM) analysis (Lightscan, Idaho Technologies) with high sensitivity detection and auto-grouping [44,45]. Bi-directional Sanger sequencing with Big Dye Terminator (ABI 3730 xl Sequencer, Applied Biosystems) was performed on all samples with aberrant melting curves for identification of SNPs. This methodology was validated by direct comparison of HRM data with Sanger sequencing in a large group of samples (n = 50). When necessary, detection sensitivity levels were adjusted to assure 100% sensitivity for all variants. All SNPs in the coding region and infrequent SNPs in the non-coding region were re-tested using separate DNA samples to independently verify the results. The impact of SNPs on protein structure was assessed with SIFT (Sorting Intolerant from Tolerant) and Polyphen score (Polymorphism Phenotyping) for
variants in the coding region and MaxEnt scan and NNNSPlce for intronic variants [46,47,48,49]. In silico evaluation of the amino acid code of PSIP1 gene product from different primates (Chimpanzee, Gorilla, Gibbon, Bushbaby) was performed based on the consensus sequences found in Ensembl database [50].

mRNA gene expression analysis

RNA was isolated using Trizol LS reagent (Invitrogen, Carlsbad, USA) from freshly extracted 5·10^6 peripheral blood mononuclear cells (PBMCs) of patients from the Ghent cohort containing genetic variants that indicated clustering in disease progression groups. Patients not harboring PSIP1 genetic variants were used as controls. A strict protocol in accordance with MIQE guidelines was followed [51]. After Dnase treatment (Dnase I, Ambion), the RNA integrity was assessed using automated electrophoresis (Experion, Bio-Rad). Reverse transcription was performed on 400 ng RNA with the iScript complementary DNA (cDNA) synthesis kit and random hexamere primers (Bio-Rad). LEDGF/p75 specific mRNA (i.e. not including LEDGF/p52) and HRP2 mRNA expression was quantified by qPCR using dual-labelled hydrolysis probes and LightCycler® 480 Probes Mastermix on a LC480 platform (Roche Diagnostics). After qPCR, the labelled hydrolysis probes and LightCycler

The same analysis was performed to distinguish clustering of alleles in disease progression subgroups. Bonferroni correction for multiple sampling was applied. After normality testing with Shapiro-Wilk, Analysis of variance (ANOVA) was corrected to determine differences in gene expression between different disease progression groups. Of the genetic variants in the coding region and in the non-coding region with significant clustering in one or more disease progression groups, the influence on gene expression (LEDGF/p75 and HRP2 mRNA) and disease progression (CD4 slope and average viral load) was analyzed for significance by Mann-Whitney U and Kruskal-Wallis test (SPSS 19 software). Pearson correlation and paired T-tests were used to assess correlation between gene expression data and the parameters of disease outcome.

Results

The study comprised 325 chronically infected HIV-1 patients from a diverse ethnic background with an average follow-up time of 15 years (Table 1). Patients were grouped into five distinct disease progression categories as outlined in Methods. Thirty-four percent of study subjects were either LTNP viremic or elite controllers. CD4 decline or average plasma viral load was independent from HIV subtype. The sensitivity of HRM curve analysis for SNP detection was 100% for all amplicons, the specificity ranged from 82% to 97%. In total 23 individual SNPs analysis for SNP detection was 100% for all amplicons, the specificity ranged from 82% to 97%. In total 23 individual SNPs

<table>
<thead>
<tr>
<th>Table 1. Patient characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age at diagnosis, median years (IQR)</td>
</tr>
<tr>
<td>Follow-up time, median years (IQR)</td>
</tr>
<tr>
<td>Sex, N (%)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Ethnicity, N (%)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Disease progression (N, %)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Current ART, N (%)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Overview of the patient characteristics, divided for Ghent and RIS cohort and per ethnicity (Africans, Caucasians). The number of patients per subcategory is presented for sex, ethnicity, disease progression groups and current antiretroviral treatment.

IQR = interquartile range; N = number; LTNP = long-term non-progressor; ART = antiretroviral treatment.

doi:10.1371/journal.pone.0050204.t001

Statistical analysis

Linkage disequilibrium between different variants was assessed with principal component analysis. Observed minor allele frequencies of individual SNPs were compared to expected frequencies per ethnicity with Fisher exact test (SPSS 19 software).
they did not meet the Hardy-Weinberg law. The observed MAFs of individual SNPs were compared with expected MAFs on population level per ethnicity. In order to detect clustering in disease progression groups, the MAFs in subgroups were calculated and compared. We did not detect linkage disequilibrium between different SNPs with higher MAF.

Genetic variants of PSIP1 in the coding region

In the full coding sequence of PSIP1, five SNPs were detected, of which two were not altering the encoded amino-acid (silent mutations) and three were non-synonymous SNPs (Table 2). All these variants have been annotated before and had low abundance [54]. Observed minor allele frequencies were compared to expected frequencies per ethnicity, in comparison with data from HapMap and dbSNP database [55,56]. Fisher exact tests were performed to detect significant lower or higher MAF compared to a randomly selected patient population (Table 2). When comparing the MAFs between the different disease progression groups, no clustering of these variants in one or more subcategories could be detected (data not shown). Kruskal-Wallis test (wild-type versus minor-allele) failed to reveal an association between genetic variants in the coding region and CD4 slope or average viral load. LEDGF/p75 mRNA expression levels did not differ for those patients harboring these SNPs (Figure 2).

Of the two synonymous variants, rs2821529 (S116S) was more abundant and showed comparable MAF to the expected frequencies, both in Africans and Caucasians. No link with disease progression (CD4 decline, average viral load) or LEDGF/p75

Table 2. Observed genetic variants in the PSIP1 coding region.

<table>
<thead>
<tr>
<th>dbSNP rs number</th>
<th>HGVS name</th>
<th>SNP location</th>
<th>African MAF exp</th>
<th>African MAF obs</th>
<th>Caucasian MAF exp</th>
<th>Caucasian MAF obs</th>
<th>Amino acid change</th>
<th>SIFT score</th>
<th>POLYPHEN score</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2821529</td>
<td>c.348T&gt;C</td>
<td>Exon2</td>
<td>0.169</td>
<td>0.107</td>
<td>0.042</td>
<td>0.039</td>
<td>S116S</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs139433616</td>
<td>c.402T&gt;C</td>
<td>Exon5</td>
<td>NA</td>
<td>0.000</td>
<td>NA</td>
<td>0.002</td>
<td>T134T</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs188943134</td>
<td>c.743C&gt;T</td>
<td>Exon8</td>
<td>NA</td>
<td>0.000</td>
<td>NA</td>
<td>0.002</td>
<td>P248L</td>
<td>0.18</td>
<td>0.045</td>
</tr>
<tr>
<td>rs61744944</td>
<td>c.1415A&gt;T</td>
<td>Exon13</td>
<td>0.034</td>
<td>0.059</td>
<td>NA</td>
<td>0.002</td>
<td>Q472L</td>
<td>0.01</td>
<td>0.007</td>
</tr>
<tr>
<td>rs35678110</td>
<td>c.1423C&gt;G</td>
<td>Exon14</td>
<td>0.025</td>
<td>0.000</td>
<td>NA</td>
<td>0.0041</td>
<td>L478V</td>
<td>0.21</td>
<td>0.049</td>
</tr>
</tbody>
</table>

1 signifies variants not in accordance with Hardy Weinberg law.

Overview of the observed SNPs in the PSIP1 coding region for LEDGF/p75. Both expected and observed minor allele frequencies (MAF) are shown per ethnicity. The SIFT score ranges from 0 to 1. The amino acid substitution is predicted as damaging if the score is ≤ 0.05, and tolerated if the score is > 0.05. The POLYPHEN score ranges from 0.00 to ≥ 2.00. The amino acid substitution is predicted possibly damaging if the score is ≥ 1.50 and probably damaging if the score is ≥ 2.00.

HGVS = Human Genomic Variation Society; SNP = single nucleotide polymorphism; MAF = minor allele frequency; NA = not assessed; rs = referenced SNP id number; – means no impact due to silent mutation.

doi:10.1371/journal.pone.0050204.t002
mRNA expression could be established (Figure 2). Only one Caucasian patient harbored SNP rs139433616 (T134T), from whom no mRNA for further analysis could be obtained.

Caucasian. SIFT and PolyPhen in silico missense variant and rs188943134 coding for P248L were both infrequent in Caucasian. SIFT and Polyphen scores indicated good functional tolerance however excluded from further analysis due to Hardy-Weinberg disequilibrium. SIFT and Polyphen scores indicated good functional tolerance in silico.

**Table 3. Observed genetic variants in the PSIP1 non-coding region.**

<table>
<thead>
<tr>
<th>dbSNP rs number</th>
<th>HGVS name</th>
<th>SNP location</th>
<th>African MAF exp</th>
<th>African MAF obs</th>
<th>Caucasian MAF exp</th>
<th>Caucasian MAF obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs536106972</td>
<td>c.1409-G&gt;C</td>
<td>Intron1</td>
<td>0.000</td>
<td>NA</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>rs2277191</td>
<td>-142+226C&gt;T</td>
<td>Intron1</td>
<td>0.034</td>
<td>0.103</td>
<td>NA</td>
<td>0.011</td>
</tr>
<tr>
<td>rs2795128</td>
<td>c.73-59G&gt;A</td>
<td>Intron2</td>
<td>0.124</td>
<td>0.109</td>
<td>0.044</td>
<td>0.066</td>
</tr>
<tr>
<td>rs12339417</td>
<td>c.149+806G&gt;A</td>
<td>Intron3</td>
<td>0.686</td>
<td>0.638</td>
<td>0.119</td>
<td>0.148</td>
</tr>
<tr>
<td>rs2737828</td>
<td>c.289-29T&gt;A</td>
<td>Intron4</td>
<td>0.000</td>
<td>0.15</td>
<td>0.028</td>
<td>0.028</td>
</tr>
<tr>
<td>rs16933270</td>
<td>c.1421-48T&gt;C</td>
<td>Intron14</td>
<td>0.146</td>
<td>0.103</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>rs188404574</td>
<td>c.30T&gt;G</td>
<td>3’UTR</td>
<td>0.000</td>
<td>NA</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>rs144536781</td>
<td>c.410G&gt;T</td>
<td>3’UTR</td>
<td>0.016</td>
<td>NA</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>rs114211035</td>
<td>c.464A&gt;G</td>
<td>3’UTR</td>
<td>0.085</td>
<td>0.000</td>
<td>NA</td>
<td>0.002</td>
</tr>
<tr>
<td>rs78732111</td>
<td>c.592G&gt;C</td>
<td>3’UTR</td>
<td>0.085</td>
<td>0.000</td>
<td>NA</td>
<td>0.000</td>
</tr>
<tr>
<td>rs2737835</td>
<td>c.620G&gt;T</td>
<td>3’UTR</td>
<td>0.033</td>
<td>0.129</td>
<td>0.108</td>
<td>0.076</td>
</tr>
<tr>
<td>rs61306098</td>
<td>c.684G&gt;A</td>
<td>3’UTR</td>
<td>0.032</td>
<td>0.092</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>rs2665515</td>
<td>c.775C&gt;T</td>
<td>3’UTR</td>
<td>0.142</td>
<td>0.094</td>
<td>0.067</td>
<td>0.066</td>
</tr>
<tr>
<td>rs13906294</td>
<td>c.790A&gt;G</td>
<td>3’UTR</td>
<td>0.031</td>
<td>NA</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>rs536106976</td>
<td>c.815G&gt;T</td>
<td>3’UTR</td>
<td>0.000</td>
<td>NA</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>rs74366322</td>
<td>c.1079T&gt;C</td>
<td>3’UTR</td>
<td>0.186</td>
<td>0.200</td>
<td>NA</td>
<td>0.000</td>
</tr>
<tr>
<td>rs13248</td>
<td>c.1232G&gt;A</td>
<td>3’UTR</td>
<td>0.611</td>
<td>0.500</td>
<td>0.004</td>
<td>0.015</td>
</tr>
<tr>
<td>rs41268959</td>
<td>c.1268G&gt;A</td>
<td>3’UTR</td>
<td>0.000</td>
<td>0.025</td>
<td>0.016</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Genetic variants of PSIP1 in intronic regions

In the flanking regions of the coding sequences, four intronic SNPs were detected, of which one previously unknown heterozygous intronic variant in one patient (ss536106972) (Table 3).

Intronic SNP rs2737828, upstream of exon 4, was significantly less abundant in HIV-positive Caucasians compared to Hapmap and dbSNP randomly selected patient population (p<0.0001). There was no significant clustering in patients with long-term non-progression, nor an association with CD4 slope (p = 0.688) or average viral load (p = 0.702) (Figure 3A). A non-significant trend towards lower LEDGF/p75 mRNA expression levels (p = 0.053) was detected in these patients compared to the wild-type genotype. HRP2 expression was not correlated with rs2737828 minor alleles (p = 0.171) (Figure 3B).

For intronic SNP rs16933270, mainly observed in the African subgroup and only in one Caucasian, the MAF was comparable with expected frequencies in Africans, but was relatively more prevalent in the LTNP patients. This variant was associated with slower CD4 decline (p = 0.020) but not with viral load levels, nor with LEDGF/p75 or HRP2 mRNA expression (p = 0.181 and p = 0.150, respectively) (Figure 3A and 3B). A multivariate analysis taking the viral load set-point and LEDGF/p75 mRNA expression into account, confirmed this impact on CD4 slope (p = 0.025). For intronic SNP rs2795128 the observed MAFs were in accordance with the expected frequencies both for Africans as for Caucasians. The minor alleles did not cluster in LTNP groups.

The allele frequencies of two intronic tagSNPs in Africans, i.e. rs2277191 and rs12339417, were additionally determined [37]. Both tagSNPs were abundant in the African subgroup and MAFs were in line with expected frequencies. However, no correlation with CD4 decline or LEDGF/p75 mRNA expression could be confirmed in the cohort. It is interesting to notice that the minor allele of rs12339417 is considered the wild-type allele in Africans.

**Genetic variants in the 3’untranslated region**

The 3’ untranslated region (UTR) harbored 12 different genetic variants, one of which was not described previously (Table 3). For rs2737835, a non-significant trend towards association of minor SNP alleles with slower CD4 decline (p = 0.053) was observed in...
Caucasian + African

Figure 2. Phenotypic impact of observed genetic variants in the PSIP1 coding region. Box-plots showing the association of individual observed SNPs in the coding region with CD4 decline (top), average log viral load (middle) and LEDGF/p75 mRNA expression (bottom). The data are combined for Africans and Caucasians. SNPs not in accordance with Hardy-Weinberg law (rs35678110) were excluded. In case of insufficient data to create a boxplot (limited amount of data points) a bar representing the mean of the values is shown. doi:10.1371/journal.pone.0050204.g002

Caucasian but not in African patients (Figure 3A). There was no correlation with average viral load (p = 0.931). The expression levels of both LEDGF/p75 (p = 0.093) and HRP2 (p = 0.317) mRNA were not associated with the presence of these minor alleles (Figure 3B). The SNP rs2737835 is located in a 3’UTR region with higher variability, although no linkage disequilibrium with other variants in this region could be determined.

The other 11 variants did not reveal MAFs aberrant from expected frequencies, eight variants were infrequent both in Africans and in Caucasians. None of these clustered in LTNP subgroups.

LEDGF/p75 and HRP2 mRNA expression levels

Gene expression analysis was performed on a subset of patients from the Ghent cohort (n = 104). Validation of reference genes for normalization with the GeNorm, NormFinder and BestKeeper software gave congruent results (Figure S1). The geometric mean of the two most stable genes (B2M and YMAHAZ) was used for normalization of gene expression data.

We could not demonstrate significant differences in expression levels between patients who received cART and those who were therapy-naïve. No statistical significant differences in either LEDGF/p75 or HRP2 mRNA expression were observed between the five disease progression groups with ANOVA. There was no correlation between LEDGF/p75 expression and the major disease outcome parameters (CD4 decline and average viral load). We could not establish an inverse correlation between LEDGF/p75 and HRP2 expression. [Pearson r = 0.490; p < 0.001] (Figure 4A). The biological variability of LEDGF/p75 expression was determined in 24 patients by analyzing samples obtained from two different time-points. Pearson test indicated correlation (r = 0.427; p = 0.033) and the paired T-test could not determine a significant difference in the means between the groups (p = 0.937) (Figure 4B). Inter-patient variability ranged from 7.7-fold expression for LEDGF/p75 till 37-fold for HRP2.

Discussion

Small molecules inhibiting the interaction between host-factor LEDGF/p75 and the viral enzyme IN form a promising new class of antiretroviral drugs targeting the integration step of HIV-1 replication cycle [28]. Increasing data indicate that genetic variation in host genes can influence HIV-1 disease susceptibility, evolution or therapy response [57]. Since there is little in vivo knowledge of the LEDGF/p75-IN interplay, the present study focused on a comprehensive characterization of this host-factor in a large and diverse patient cohort, on the levels of genetic background and mRNA expression.

The data indicate that the coding region of LEDGF/p75 is highly conserved. Of the 35 annotated genetic variants in HapMap (most of them extremely infrequent in a reference population), only five were detected in a cohort of 325 HIV-1 positive patients. Three non-synonymous SNPs were low-abundant and there was no in silico indication that they had a major impact on the phenotype. Functional evaluation needs to be performed to confirm these predictions. For rs61744944 (Q472L), experimental data previously showed no alteration of the LEDGF/p75-IN binding affinity and the near-complete rescue of HIV-1 infection by mutant LEDGF/p75 [37]. However, this does not exclude the possibility of changes in other functions such as integration site distribution. The low MAFs of these variants in the present cohort were insufficient to detect a more subtle impact on disease progression. Patients harboring these variants had normal levels of LEDGF/p75 mRNA expression. One homozygous missense variant (L478V, rs35678110) was detected in one LTNP viremic controller. Unfortunately, no further expression data could be obtained for this patient.

Alignment of consensus sequences of PSIP1 gene products from different species (Chimpanzee, Gorilla, Gibbons) revealed a highly conserved protein along evolutionary lines, suggesting the important biological function of this gene (Figure 1).

Three SNPs were described with indications for a weak association with HIV disease outcome, situated either in intronic sequences (rs16933270, rs2737828) or in the 3’UTR (rs2737835). In HIV-infected Caucasian patients, rs2737828 was significantly underrepresented and showed a non-significant trend towards lower LEDGF/p75 mRNA expression. This might suggest a protective role of this variant in acquiring HIV-1 without affecting disease evolution, but the cross-sectional study design and the lack
of a highly-exposed sero-negative or healthy control arm do not allow to establish susceptibility associations. The underrepresentation of this variant in Caucasian HIV-1 patients needs to be confirmed in different cohorts, since only limited data is available on the geographic distribution of this SNP. In African patients, there was a clustering of rs16933270 minor alleles in patients with slower CD4 decline but no impact on viral load set-point or LEDGF/p75 expression levels. Although similar viral load levels are maintained, some LEDGF/p75 haplotypes could result in a better CD4+ T-cell survival. Low sample size in the African sub-cohort however can introduce bias in these results. The investigation of this variant in a larger cohort of African HIV-1 patients to further assess its impact is recommended. Splice-site finders could not exclude that rs2737828 and rs16933270 had the potential to create a branch point sequence and alter splicing, although no minor splice variants were detected. Variant rs2737835 showed a non-significant tendency towards slower CD4 decline but not towards a lower average viral load. Based on miRNA databases (Patrocles finder), this 3’UTR SNP could be the target for miRNA binding (hsa-mir-1274a), leading to mRNA destabilization and lower protein levels.

The present data did not confirm the association of African tagSNP rs12339417 with delayed CD4 decline in a chronic infection phase. It must be noted that the small sample size of the African sub-cohort and larger genetic diversity in Africans from different regions could explain part of these findings. Besides this, lower LEDGF/p75 expression levels associating with the minor allele were previously only seen in a sero-converter/early infection cohort and it might be possible that LEDGF/p75 levels mainly affect initial HIV infection risk and early replication events. Alternatively, the use of only one reference gene (GAPDH) as normalization strategy in the gene expression assays might have

![Figure 3. Phenotypic impact of three genetic variants in the PSIP1 non-coding region.](image)

(A) Differential CD4 decline for patients carrying wild-type or variant alleles of rs2737828 and rs2737835 (Caucasian) and rs16933270 (African), based on the mean CD4 slope and a similar starting point. P values to estimate significance of difference are indicated. (B) Differential expression of LEDGF/p75 mRNA and HRP2 mRNA in PBMCs from patients carrying wild-type or variant alleles of rs2737828 and rs2737835 (Caucasian) and rs16933270 (African). Horizontal bars indicate the mean expression levels.

doi:10.1371/journal.pone.0050204.g003

![Figure 4. Biological variability and correlation of LEDGF/p75 mRNA with HRP2 mRNA expression.](image)

(A) Scatter-plot showing LEDGF/p75 mRNA and HRP2 expression in identical patient samples (n = 68). Pearson r-values and p-values are indicated. (B) Scatter-plot showing the biological variation of LEDGF/p75 mRNA expression in two samples from 24 patients at different time points. t1 = time point 1; t2 = time point 2.

doi:10.1371/journal.pone.0050204.g004
introduced additional bias and hampers comparison [52]. In our assays, normalization with only one reference gene resulted in a 27% increase in variability compared to the standard procedure with two validated reference genes (data not shown).

Relatively stable inter-individual LEDGF/p75 mRNA expression was detected in vivo. This variance was not related to genetic polymorphisms in the coding region or 3’UTR. Low levels of LEDGF/p75 expression, linked with decreased integration in transcriptional active regions in vitro, did not result in lower viral load [23]. In addition, there was no correlation between low LEDGF/p75 mRNA expression and CD4 decline. Previous in vitro studies revealed that LEDGF/p75 knockdown might be rescued by HRP2, which harbors a similar IN-binding domain [29]. Consequently, the expression of HRP2 was investigated to assess whether low levels of LEDGF/p75 are correlated with higher levels of HRP2 and to see if this rescue mechanism is also present in vivo, but this was not observed. In vitro studies revealed that low levels of LEDGF/p75 are sufficient to completely rescue HIV-1 integration [58]. Other still unknown factors or a more important effect of LEDGF/p75 during early replication might provide further explanation for the observed lack of correlation between LEDGF/p75 expression levels and disease outcome in chronically infected patients. It must be noted that mRNA was extracted from PBMCs, containing not only CD4+ but also CD8+ and B-lymphocytes, and therefore representing average expression levels of these cells. The obtained results did not motivate us to further analyze the 5’ UTR and promoter region of PSIP1.

Because of the limited phenotypic effect of non-coding SNPs (on instance for tertiary structure) and the observed completeness of the correct coding region of LEDGF/p75, our data can provide further validation of the LEDGF/p75 – IN interaction as a promising target for antiretroviral treatment. The relatively stable expression of their target shows that LEDGINs could be broadly applicable and provide a good and durable inhibitory effect.

In general, the results of this study underscore the importance of detecting rare variants in genes with a high probability of influencing disease outcome in distinct populations. In contrast with Genome-wide association studies, which are designed to reveal genotype-phenotype relations with common variants, rare variants which show an impact in well-chosen and defined patient populations can help elucidate functional roles of the gene of interest.

Supporting Information

Table S1 Overview of primers, cycling conditions and reference sequences. Overview of primers and cycling conditions for the different PSIP1 gene fragments, including all reference genes used for gene expression analysis. The Ensembl transcript and protein ID of the different PSIP1 gene products from humans and four primates is listed as well.

(TDOC)

Figure S1 Average expression stability of reference genes. Average expression stability of the reference genes used in the gene expression assays. Values are calculated with the GeNorm software on at least 15 samples. The gene stability measure M (Y-axis) is the average pairwise variation V of a gene (as indicated in X-axis) with all other genes.

(TIF)

Acknowledgments

We want to acknowledge the LTNP-RIS cohort and the HIV Biobank (Spanish AIDS Research Network) for providing clinical data and the generous gift of clinical samples used in this work.

Author Contributions

Conceived and designed the experiments: P. Messiaen WDS JA BV BP CV. Performed the experiments: P. Messiaen WDS KV PVA P. Performed the experiments: P. Messiaen WDS JA BV BP CV. Contributed reagents/materials/analysis tools: JA BV BP CV LV. Performed the experiments: P. Messiaen WDS JA BV BP CV.

References


