The Influence of host genetic factors on HIV-1 infection

Manen, D.

Citation for published version (APA):
GENOME-WIDE ASSOCIATION STUDY ON HIV-1 SUSCEPTIBILITY IN DUTCH HIGH-RISK SERO-NEGATIVE INDIVIDUALS

Daniëlle van Manen¹, Sebastiaan M. Bol¹, Brigitte Boeser-Nunnink¹, Ineke Stolte², Leonard H. van den Berg³, Sophie Limou⁴, Jean-François Zagury⁴, Jan H. Veldink³, Angélique B. van ’t Wout¹ and Hanneke Schuitemaker¹

¹Dept Experimental Immunology, Sanquin research, Landsteiner Laboratory, and Center for Infectious Diseases and Immunity Amsterdam (CINIMA) at the Academic Medical Center of the University of Amsterdam, the Netherlands
²Cluster of Infectious Diseases, Amsterdam Public Health Service, the Netherlands ³Rudolf Magnus Institute of Neuroscience, Department of Neurology, University Medical Center Utrecht, The Netherlands ⁴Chaire de Bioinformatique, Conservatoire National des Arts et Métiers and ANRS Genomic Group, Paris, France

Submitted for publication
ABSTRACT

Some individuals have remained persistently HIV-1-negative, despite multiple high-risk exposures to the virus. Previously identified host genetic factors cannot explain all resistance to HIV-1 infection. To reveal additional genetic markers that associate with HIV-1 acquisition, we performed a genome-wide association study (GWAS) in High Risk Sero-Negatives (HRSN). We tested the association of SNP alleles with HIV-1 susceptibility in 60 HRSN and 434 HIV-1 positive individuals. Subsequently, we analyzed interesting signals in an association analysis between HRSN or HIV-1 infected individuals and a HIV-negative control population (n=1057).

The HRSN were enriched for homozygosity of the CCR5-Δ32 allele when compared to HIV-positive individuals, emphasizing that the population consisted of exposed-uninfected individuals. P-values for the top 10 associations between SNP genotypes and HIV-1 susceptibility in this same comparison varied from 8.15 x 10⁻⁵ – 8.99 x 10⁻⁶, but lacked genome-wide significance. Interestingly however, the minor allele frequency (MAF) of these 10 SNPs was also significantly enriched in HRSN when compared with healthy controls. The MAF of four of these SNPs were significantly reduced in HIV-1 positive individuals when compared to healthy controls, two of these SNPs were also associated with HIV-1 acquisition in the French Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort when compared with healthy controls (combined P-values: P=6.00 x 10⁻⁴ and P=2.00 x 10⁻³).

This GWAS on host genetic markers of HIV-1 susceptibility in a population of European descent has revealed potentially important SNPs for HIV-1 resistance. Larger populations and meta-analyses are needed to confirm findings and identify novel host factors.
INTRODUCTION
In the AIDS pandemic, rare individuals have remained persistently HIV-1-negative despite multiple high-risk exposures to HIV-1. Several viral and host factors are associated with acquisition of HIV-1. Differences in virus subtype and virus load in the donor influences the likelihood of transmission to a naive host [1]. The presence of host genetic factors in the seronegative recipient associated with HIV-1 infection may also play a role. A proven host genetic factor to be correlated with HIV-1 susceptibility in several Western-European and Asian cohorts, is the 32 bp deletion in the CCR5 gene, which codes for a coreceptor for HIV-1. Individuals being homozygous for this 32 bp deletion are resistant to infection with CCR5-using viruses [2-4]. High expression levels of chemokines Rantes, MIP-1α and MIP-1β, all natural ligands for CCR5, have also been associated with resistance to HIV infection [5,6]. Furthermore, High Risk Sero-Negative (HRSN) individuals were found to carry a higher frequency of HLA-A11, A31 and Cw15 alleles when compared to HIV-positive individuals [6]. However, with the knowledge of host genetic markers today, we cannot explain all resistance to HIV-1 infection.

Genome-wide association studies (GWAS) have recently become a widely accepted approach to search for common genetic variation involved in HIV-1 control and disease progression. Several studies have been published, which mainly demonstrated the primary role of the MHC region in HIV-1 control. These studies only focused on host genetic factors in the control of viral load and disease progression in HIV-1 infected individuals [7-14]. Until now, only one GWAS on HIV-1 susceptibility to infection and performed in an African population, has been published [15]. Here we performed a GWAS in homosexual men from the Amsterdam Cohort Studies (ACS) who were selected based on the fact that they had remained seronegative despite high-risk sexual behavior and compared them with participants from the same cohort studies who had become infected with HIV-1 during follow-up in the same calendar time. Interesting signals were subsequently compared for allele frequencies between HRSN or HIV-1 infected individuals and an HIV-negative control population from the Netherlands, and for association with HIV-1 susceptibility in an independent cohort study. Studying these extreme phenotypes may identify genetic loci that are associated with HIV-1 susceptibility, thereby revealing potentially novel mechanisms of resistance to HIV-1 infection.

METHODS
Study population
In total we selected 64 HIV-seronegative homosexual participants of the ACS with high-risk sexual behavior. These men had an HIV-seronegative follow-up of more than five years despite unprotected receptive anogenital sex with at least 2 different non-steady partners and/or a reported episode of syphilis [16]. DNA for genotyping was available for all 64 HRSN.
We compared the HRSN with a total of 455 HIV-1 positive individuals who participate in the prospective Amsterdam Cohort Studies on HIV infection and AIDS on men who have sex with men (MSM, n=335) and drug users (DU; n=120). MSM were enrolled in the cohort between October 1984 and March 1986 (205 seroprevalent at entry and 130 seroconverters during follow-up). DU were enrolled from 1985 onwards. From all participants a DNA sample was available. More details are described in van Manen et al [17].

The ACS has been conducted in accordance with the ethical principles set out in the declaration of Helsinki and all participants provided written informed consent. The study was approved by the Academic Medical Center institutional Medical Ethics Committee of the University of Amsterdam.

Minor allele frequencies of potentially interesting SNPs that associated with HIV-1 susceptibility were compared between HRSN or HIV-1 positive individuals and 1067 Dutch seronegative individuals. The Dutch control population was described previously in a GWAS on amyotrophic lateral sclerosis (ALS) [18], and consisted of 450 control individuals who were unrelated, healthy volunteers who accompanied ALS patients to the UMC Utrecht neurology outpatient clinic and 617 healthy individuals who were recruited from an ongoing, prospective population-based study on ALS in The Netherlands.

2nd stage analysis
The Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort (n=360) is composed of French HIV-1 seroprevalent long-term non-progressors (LTNP, n=275) [9] and rapid progressors (RP, n=85) [10]. The HIV-1 positive individuals were compared to a French HIV-1 seronegative control group (n=697) from the DESIR program [19].

GWAS scan genotyping and quality control
DNA samples were genotyped using either Illumina’s Infinium HumanHap300 or Human 370CNV BeadChips. Genotype clustering was performed using the Infinium Beadstudio program. PLINK was used for quality control; SNPs were excluded based on call rate (<99%), minor allele frequency (MAF) (<0.01) and Hardy-Weinberg Equilibrium deviation (<1E-6). Samples were excluded based on call rate (<95%), gender check and heterozygosity.

Population stratification
Population stratification was addressed by PLINK by conducting identity-by-state (IBS) analysis. A selection was made of unlinked autosomal SNPs in both the HRSN, HIV-positive and Dutch control population. This selection of SNPs was also downloaded of the HapMap samples. The first four dimensions of a multidimensional scaling (MDS) analysis of IBS distances for each sample were calculated. We plotted the first two dimensions for each sample and removed the non-caucasian samples. After removal
GWAS IN HIV-EXPOSED SERONEGATIVE INDIVIDUALS

of 4 HRSN, 21 HIV-positive individuals and 10 Dutch healthy controls we recalculated the first four dimensions of the MDS analysis. These four dimensions were used as covariates to correct for any remaining population structure.

Analysis
Association analysis between SNP genotypes and HIV-1 susceptibility was tested using logistic regression models in PLINK, thereby correcting for gender and the first four dimensions of the multidimensional scaling analysis. LD was characterized using WGA Viewer [20].

RESULTS
A total of 64 HRSN and 455 HIV-1-positive individuals were genotyped, with Illumina Beadchip Arrays. After SNP and sample quality control (see Methods), analysis could be performed with 308,960 SNPs from 60 HRSN and 434 HIV-positive individuals. Association for HIV-1 susceptibility was tested by logistic regression model with gender and the first four dimensions of the multidimensional scaling analysis as covariates.

In Table 1, the SNPs that ranked in the top 10 for association between HIV-1 resistance are shown, all with a p-value smaller than 8.15 x10^{-5}. The association for SNP rs6491147, which is the top SNP that is located 20 kilo bases (kb) downstream of the gene encoding for G protein-coupled receptor 12 (GPR12), has a P-value of 8.99 x 10^{-6} (Table 1), which means that none of the SNPs reached genome-wide significant values. For all of these SNPs the minor allele frequency (MAF) was also significantly different between HRSN and healthy Dutch controls (Table 2). SNP rs1338442, which

<table>
<thead>
<tr>
<th>SNP</th>
<th>P-value</th>
<th>Chr</th>
<th>Coordinate</th>
<th>Gene</th>
<th>Position</th>
<th>MAF_HRSN</th>
<th>MAF_HIV+</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6491147</td>
<td>8.99E-06</td>
<td>13</td>
<td>27306551</td>
<td>GPR12</td>
<td>20kb 3'</td>
<td>0.44</td>
<td>0.26</td>
</tr>
<tr>
<td>rs2367070</td>
<td>2.29E-05</td>
<td>10</td>
<td>24914386</td>
<td>ARHGAP21</td>
<td>Intron</td>
<td>0.28</td>
<td>0.20</td>
</tr>
<tr>
<td>rs4662294</td>
<td>4.39E-05</td>
<td>2</td>
<td>143387612</td>
<td>Intergenic</td>
<td></td>
<td>0.40</td>
<td>0.25</td>
</tr>
<tr>
<td>rs7943401</td>
<td>4.39E-05</td>
<td>11</td>
<td>104018088</td>
<td>PDGFD</td>
<td>Intron</td>
<td>0.38</td>
<td>0.22</td>
</tr>
<tr>
<td>rs10240140</td>
<td>7.04E-05</td>
<td>7</td>
<td>158367161</td>
<td>PTPRN2</td>
<td>Intron</td>
<td>0.38</td>
<td>0.20</td>
</tr>
<tr>
<td>rs12671268</td>
<td>7.04E-05</td>
<td>7</td>
<td>158369430</td>
<td>PTPRN2</td>
<td>Intron</td>
<td>0.38</td>
<td>0.20</td>
</tr>
<tr>
<td>rs4680266</td>
<td>7.22E-05</td>
<td>3</td>
<td>156156285</td>
<td>KCNAB1</td>
<td></td>
<td>0.52</td>
<td>0.35</td>
</tr>
<tr>
<td>rs6743405</td>
<td>7.70E-05</td>
<td>2</td>
<td>179983572</td>
<td>SESTD1</td>
<td>Intron</td>
<td>0.49</td>
<td>0.32</td>
</tr>
<tr>
<td>rs1374378</td>
<td>8.07E-05</td>
<td>2</td>
<td>179984117</td>
<td>SESTD1</td>
<td>Intron</td>
<td>0.49</td>
<td>0.33</td>
</tr>
<tr>
<td>rs7132834</td>
<td>8.15E-05</td>
<td>12</td>
<td>22108595</td>
<td>ABCC9</td>
<td>20kb 5'</td>
<td>0.56</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Chr; chromosome, MAF; minor allele frequency, HRSN; High Risk Sero-Negative, HIV+; HIV-1 positive
is located intronic of T-box 15 (TBX15), has even a smaller P-value when HRSN were compared with healthy Dutch controls, compared to the discovery analysis (P=3.61 \times 10^{-6} and P=9.53 \times 10^{-3} respectively, Table 2). Interestingly, in the case of four of the SNPs that ranked in the top 10 for potential association with HIV-1 susceptibility, the minor allele was significantly reduced in HIV-1 positive individuals when compared to healthy Dutch controls (Table 2).

Moreover, we analyzed whether these top 10 SNPs were also associated with infection in another independent cohort. A total of 360 individuals from the French Genomics of Resistance to Immunodeficiency Virus (GRIV) cohort were compared with a French HIV-1 seronegative control group (n=697). Three SNPs in this 2nd stage analysis had a P-value < 0.05 when HIV-1-positives were compared with healthy controls. Overall, SNPs rs2367070 and rs7132834 seemed to be strongly associated with HIV-1 infection, when we looked at the combined p-values on the comparison of HIV-1 positive individuals with healthy controls, between the ACS and GRIV cohorts (P=6.00 \times 10^{-4} and P=2.00 \times 10^{-3} respectively, Table 2).

### Table 2. P-values of SNPs that are potentially associated with HIV-1 susceptibility in other study populations.

<table>
<thead>
<tr>
<th>SNP</th>
<th>MAF</th>
<th>ACS HRSN</th>
<th>Dutch Ctrl</th>
<th>ACS HIV+</th>
<th>ACS HRSN vs Ctrl</th>
<th>ACS HIV+ vs Ctrl</th>
<th>GRIV HIV+ vs Ctrl</th>
<th>combined HIV+ vs Ctrl</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6491147</td>
<td>0.44</td>
<td>0.28</td>
<td>0.26</td>
<td>3.93E-04</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>rs2367070</td>
<td>0.28</td>
<td>0.23</td>
<td>0.20</td>
<td>8.86E-03</td>
<td>1.66E-02</td>
<td>3.51E-03</td>
<td>6.00E-04</td>
<td></td>
</tr>
<tr>
<td>rs4662294</td>
<td>0.40</td>
<td>0.32</td>
<td>0.25</td>
<td>2.00E-02</td>
<td>1.77E-03</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rs7943401</td>
<td>0.38</td>
<td>0.22</td>
<td>0.22</td>
<td>1.37E-04</td>
<td>ns</td>
<td>4.76E-02</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>rs10240140</td>
<td>0.38</td>
<td>0.22</td>
<td>0.20</td>
<td>1.47E-03</td>
<td>ns</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rs12671268</td>
<td>0.38</td>
<td>0.22</td>
<td>0.20</td>
<td>1.43E-03</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>rs4680266</td>
<td>0.52</td>
<td>0.41</td>
<td>0.35</td>
<td>1.00E-02</td>
<td>2.39E-03</td>
<td>ns</td>
<td>1.40E-02</td>
<td></td>
</tr>
<tr>
<td>rs6743405</td>
<td>0.49</td>
<td>0.32</td>
<td>0.32</td>
<td>1.62E-05</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>rs1374378</td>
<td>0.49</td>
<td>0.32</td>
<td>0.33</td>
<td>1.68E-05</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>rs7132834</td>
<td>0.56</td>
<td>0.41</td>
<td>0.39</td>
<td>1.08E-03</td>
<td>4.21E-02</td>
<td>5.16E-03</td>
<td>2.00E-03</td>
<td></td>
</tr>
</tbody>
</table>

MAF: minor allele frequency, ACS: Amsterdam Cohort Studies, HRSN: High Risk Sero-Negative, Ctrl: healthy control, HIV+: HIV-1 positive, GRIV: Genomics of Resistance to Immunodeficiency Virus cohort, ns: not significant (P>0.05) NA: not applicable.

### DISCUSSION

We here describe the first GWAS on HIV-1 susceptibility in a population from European descent, in which we compared HRSN with HIV-1 positive individuals. We report here the top 10 SNPs for which the minor allele was strongly enriched within HRSN when compared to HIV-1 positive individuals, albeit not genome-wide significant.
For all SNPs that rank in the top 10 associations, the MAF is lower in HIV-positives and higher in HRSN when compared to the MAF in healthy controls, indicating that the minor alleles of the top 10 SNPs associate with HIV-1 resistance. Furthermore, we analyzed whether the MAF of SNPs that were identified as potentially important for HIV-1 susceptibility, differed between HIV-positives and healthy Dutch controls as well, to emphasize a potential genetic basis for the HIV-1 resistance phenotype in HRSN. Indeed, the MAF of four of these SNPs that are potentially associated with protection from infection was significantly lower in HIV-1 positive individuals when compared with healthy Dutch controls. The association of two of these four SNPs with HIV-1 susceptibility could be confirmed by comparing the MAF of these SNPs in the independent French HIV-1 positive GRIV cohort and HIV-1 seronegative individuals also from France. In this comparison as well, the MAF of these two SNPs (rs2367070 and rs7132834) was significantly lower in the HIV-1 positive population.

Interestingly, although not confirmed in the GRIV cohort, two independent gene regions that ranked in the top 10 for being potentially associated with HIV-1 susceptibility have previously been reported to encode for so-called HIV dependency factors. SNPs rs10240140 and rs12671268, which are in complete linkage disequilibrium (LD; \( r^2=1 \)) are both located intronic of protein tyrosine phosphatase, receptor type, N polypeptide 2 (PTRPN2) on chromosome 7, which was identified as an HIV-1 dependency factor for infection and early stage HIV-1 replication in an in vitro siRNA screen [21]. The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family and treatment with a PTP inhibitor, which induces signal transduction pathways similar to those induced upon antigenic costimulation, resulted in the activation of the HIV-1 LTR in T-cells, partly through NF-κB [22].

The other gene region that is potentially involved in HIV-1 susceptibility and tagged by SNPs rs6743405 and rs1374378 which are also in high LD with each other \( (r^2>0.9) \), is located in an intron of the gene SEC14 and spectrin domains 1 (SESTD1). The SESTD1 gene was described to be essential for in vitro HIV-1 infection and replication [23].

Data on risk behavior and level of pathogen exposure were only known for a fraction of the HRSN and HIV-1 positive individuals that were enrolled in this study. A larger proportion of HRSN were known to have had at least one reported episode of syphilis (24/42 = 57.1 %) when compared to HIV-1 positive individuals (149/338 = 44.1%). An enrichment of homozygosity for the CCR5-Δ32 allele was observed in the HRSN when compared to HIV-1 positive individuals in the ACS (3.0% vs 0.0% respectively, \( P=2.0 \times 10^{-3}, \) data not shown), emphasizing the exposed HIV-negative status of the HRSN.

The HIV-positive individuals from the ACS consist of both MSM and DU, the difference in transmission route may have an influence on association with HIV-1 susceptibility. However, the genotype prevalence of the top 15 SNPs analyzed in this study was similar between MSM and DU (data not shown).

The SNPs identified in this study do not map to gene regions found in a recently reported GWA study on HIV-1 susceptibility that was performed in an African population [15]. The lack of association is most likely due to the difference in ancestry
of individuals included in both studies. However, there may be other explanations that contribute to this difference, like gender, exposure level and transmission route of infection. Therefore, larger study populations and meta-analyses are necessary to confirm involvement of previously identified gene regions and to identify novel gene regions that are associated with resistance or reduced susceptibility to HIV-1 infection.

ACKNOWLEDGEMENTS

The Amsterdam Cohort Studies on HIV infection and AIDS, a collaboration between the Public Health Service of Amsterdam, the Academic Medical Center of the University of Amsterdam, the Sanquin Blood Supply Foundation, Jan van Goyen Medical Center, and the University Medical Center Utrecht, are part of the Netherlands HIV Monitoring Foundation and financially supported by the Center for Infectious Disease Control of the Netherlands National Institute for Public Health and the Environment. We acknowledge funding from the Netherlands Organization for Scientific Research (TOP, registration number 9120.6046).

REFERENCES


