Impact of Single Nucleotide Polymorphisms and of Clinical Risk Factors on New-Onset Diabetes Mellitus in HIV-Infected Individuals


1 Institute of Microbiology, University Hospital and University of Lausanne, 2 Swiss HIV Cohort Study Data Center, and 3 Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, 4 Institute for Social and Preventive Medicine, University of Bern, and 5 Clinic for Infectious Diseases, Bern University Hospital and University of Bern, Bern, 6 Infectious Diseases Service, University Hospital, and 7 Kantonsspital Bruderholz, University of Basel, Basel, 8 Kantonsspital St. Gallen, St. Gallen, 9 Ospedale Regionale, Lugano, 10 University Hospital, Geneva, and 11 University Hospital Zurich, University of Zurich, Zurich, Switzerland

Background. Metabolic complications, including cardiovascular events and diabetes mellitus (DM), are a major long-term concern in human immunodeficiency virus (HIV)–infected individuals. Recent genome-wide association studies have reliably associated multiple single nucleotide polymorphisms (SNPs) to DM in the general population.

Methods. We evaluated the contribution of 22 SNPs identified in genome-wide association studies and of longitudinally measured clinical factors to DM. We genotyped all 94 white participants in the Swiss HIV Cohort Study who developed DM from 1 January 1999 through 31 August 2009 and 550 participants without DM. Analyses were based on 6054 person-years of follow-up and 13,922 measurements of plasma glucose.

Results. The contribution to DM risk explained by SNPs (14% of DM variability) was larger than the contribution to DM risk explained by current or cumulative exposure to different antiretroviral therapy combinations (3% of DM variability). Participants with the most unfavorable genetic score (representing 12% and 19% of the study population, respectively, when applying 2 different genetic scores) had incidence rate ratios for DM of 3.80 (95% confidence interval [CI], 2.05–7.06) and 2.74 (95% CI, 1.53–4.88), respectively, compared with participants with a favorable genetic score. However, addition of genetic data to clinical risk factors that included body mass index only slightly improved DM prediction.

Conclusions. In white HIV-infected persons treated with antiretroviral therapy, the DM effect of genetic variants was larger than the potential toxic effects of antiretroviral therapy. SNPs contributed significantly to DM risk, but their addition to a clinical model improved DM prediction only slightly, similar to studies in the general population.
Table 1. Baseline Characteristics of the Study Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (n = 644)</td>
<td>With new onset DM (n = 94)</td>
<td>Without new onset DM (n = 550)</td>
</tr>
<tr>
<td>Age, median (IQR), years</td>
<td>40 (35–48)</td>
<td>45.5 (38.3–54)</td>
<td>39 (34–46)</td>
</tr>
<tr>
<td>BMI, median (IQR)</td>
<td>23.2 (20.8–25.9)</td>
<td>25.8 (23.1–27.5)</td>
<td>22.8 (20.6–25.2)</td>
</tr>
<tr>
<td>Male sex</td>
<td>512 (79.5)</td>
<td>81 (86.2)</td>
<td>431 (78.4)</td>
</tr>
<tr>
<td>Presumed mode of HIV transmission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men who have sex with men</td>
<td>292 (45.3)</td>
<td>36 (38.3)</td>
<td>256 (46.5)</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>191 (29.7)</td>
<td>35 (37.2)</td>
<td>156 (28.4)</td>
</tr>
<tr>
<td>Injection drug use</td>
<td>137 (21.3)</td>
<td>18 (19.1)</td>
<td>119 (21.6)</td>
</tr>
<tr>
<td>Unknown or other</td>
<td>24 (3.7)</td>
<td>5 (5.3)</td>
<td>19 (3.5)</td>
</tr>
<tr>
<td>ART group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No ART</td>
<td>101 (15.7)</td>
<td>12 (12.8)</td>
<td>89 (16.2)</td>
</tr>
<tr>
<td>NRTI only</td>
<td>77 (12)</td>
<td>15 (16)</td>
<td>62 (11.3)</td>
</tr>
<tr>
<td>NRTI+PI</td>
<td>227 (35.2)</td>
<td>40 (42.6)</td>
<td>187 (34)</td>
</tr>
<tr>
<td>NRTI+NNRTI</td>
<td>198 (30.7)</td>
<td>18 (19.1)</td>
<td>180 (32.7)</td>
</tr>
<tr>
<td>NRTI+NNRTI+PI</td>
<td>41 (6.4)</td>
<td>9 (9.6)</td>
<td>32 (5.8)</td>
</tr>
<tr>
<td>Value during follow-up period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4+ T cells, median (IQR), cells/μL</td>
<td>540 (368–763)</td>
<td>385 (250–594)</td>
<td>562 (400–782)</td>
</tr>
<tr>
<td>HIV viral load &lt;400 copies/mL</td>
<td>427 (66.3)</td>
<td>49 (52.1)</td>
<td>378 (68.7)</td>
</tr>
<tr>
<td>Hepatitis C virus coinfection</td>
<td>185 (28.7)</td>
<td>24 (25.5)</td>
<td>161 (29.3)</td>
</tr>
</tbody>
</table>

NOTE. Data are no. (%) of participants, unless otherwise indicated. ART, antiretroviral therapy; BMI, body mass index; DM, diabetes mellitus; HIV, human immunodeficiency virus; IQR, interquartile range; NNRTI, non-nucleoside reverse-transcriptase inhibitors; NRTI, nucleoside reverse-transcriptase inhibitors; PI, protease inhibitor.

Recent genome-wide association studies (GWAS) have provided a comprehensive inventory of common single nucleotide polymorphisms (SNPs) reproducibly associated with DM in the general population [11–20]. The aim of the present study was to examine the contribution of 22 SNPs identified in GWAS of the general population to the risk of DM in HIV-infected individuals. Because these GWAS have been conducted almost exclusively in the white population, our study population included only participants self-identified as white. We assessed the quantitative impact of genetic background and relevant clinical factors, most notably body mass index (BMI) and ART exposure, on DM risk and compared the relative importance of SNPs and clinical factors.

MATERIALS AND METHODS

Participants, DM diagnosis, and ART and other medication exposure. Study participants were followed up in the Swiss HIV Cohort Study (SHCS) [21] during the study period (1 January 1999 to 31 August 2009). The SHCS Genetics Project was approved by the ethics committees of participating centers. Participants gave written, informed consent for genetic testing. The genotyped study population (n = 644) consisted of all 94 white SHCS participants who developed new-onset DM during the study period, the majority of whom were included in a previous DM epidemiological study [8], and 550 randomly selected white SHCS participants with ≥2 years of ART exposure who did not develop DM. Participants were previously genotyped for a genetics-dyslipidemia study [22]. DM was diagnosed in accordance with the criteria of the Expert Committee on the Diagnosis and Classification of DM [23], with confirmed plasma glucose level cut-off values of >7.0 mmol/L (fasting) or >11.1 mmol/L (nonfasting) [8, 24].

SHCS participants have routine biannual follow-ups with measurements of weight, waist and hip circumference, serum lipids, and glucose. Antiretroviral agents are recorded with start and stop dates in the SHCS database. ART exposure was assessed at the time of DM diagnosis or, for participants without new-onset DM, at the end of the study period. Current exposure to the following ART groups was assessed as described in a previous analysis of the SHCS [8]: (1) NRTI only, (2) NRTI plus PI, (3) NRTI and NNRTI without PI, and (4) NRTI and NNRTI with PI. In an a priori defined sensitivity analysis, cumulative exposure in years to each of the 4 ART groups was...
considered, normalized to each participant’s follow-up duration [5, 6, 9].

**Genetic variants and genotyping.** The 22 genetic variants selected for the study were associated with DM in GWAS that were conducted in the general population and that were published as of February 2009: GCKR, FTO, HNF1B, TCF7L2, SLC30A8, HHEX, EXT2/ALX4/LOC387761, CDKN2A/B, IGF2BP2, CDKAL1, KCNJ11, PPARG, WFS1, KCNQ1, JAZF1, CDC123/CAMK1D, TSPAN8/LGR5, THADA, ADAMTS9, NOTCH2, MTNR1B (2 SNPs) (Table A1, online only). Genotyping was performed by TaqMan allelic discrimination with TaqMan SNP genotyping assays predesigned by Applied Biosystems. Results were entered in the central SHCS genetic database without knowledge of glucose values or DM status.

**Statistical analysis.** The incidence rate of new-onset DM was defined as the number of cases divided by the total number of person-years of follow-up. Follow-up was from the first SHCS visit after 1 January 1999 (baseline) to the date that DM was first diagnosed, death, or last SHCS visit, whichever occurred first. The incidence rate was analyzed by means of a Poisson regression model with log-link function. In a basic model, we considered clinical variables only. Variables were either time dependent, including ART group, BMI category (<18.5, 18.5–25, or >25), CD4+ T cell count (square root of cells/μL) [8], serum high-density lipoprotein (HDL)–cholesterol (mmol/L), and triglycerides (mmol/L), or time fixed, including sex and age (per 10 years) at baseline. Time-dependent variables entered the model as the most recent information available at each follow-up visit. Each SNP was treated as a categorical variable having 2 levels (2 reference alleles versus 1 variant allele). We assumed an additive mode of inheritance. Two separate genetic model–building strategies were compared. First, starting with the basic model and adding the SNPs step-by-step, a final multivariable model 1 (m1) was selected by forward stepwise regression based on the Akaike information criterion. Second, given the prior validation of the SNPs in GWAS [11–20], all SNPs were added to the basic model to build model 2 (m2).

The relative contribution of clinical and genetic variables to DM risk at the level of the study population was assessed by analyzing deviance, a measure for goodness-of-fit when assessing categorical end points, such as DM [25]. To assess the cumulative contribution of SNPs to DM risk at the level of the individual study participant, a genetic score was generated on the basis of the number of DM risk alleles carried. Study participants were then distributed as evenly as possible into 4 risk
Figure 2. Impact of genetic and key nongenetic variables on new-onset diabetes mellitus (DM) in the final, multivariable analysis (model \( m1 \)). Results are represented as the estimated effect and 95% confidence interval on the incidence rate ratio (IRR) of new-onset DM for the genetic variants (black dots), clinical variables (gray triangles), and type of antiretroviral therapy (ART) (gray squares). The estimated effects are relative to a 41.7-year-old woman with normal body mass index (18.5–25), CD4+ T cell count of 484 cells/\mu L, no treatment with ART, a serum high-density lipoprotein (HDL)–cholesterol of 1.23 mmol/L, serum triglycerides of 2.62 mmol/L, and reference alleles at all genetic loci tested. The age effect is with respect to an increase of 10 years, and the CD4 effect is with respect to a unit increase of the number of CD4 cells/\mu L on the square root scale. Serum HDL and triglycerides are with respect to a unit increase (original scale). NNRTI, nonnucleoside reverse-transcriptase inhibitors; NRTI, nucleoside reverse-transcriptase inhibitors; PI, protease inhibitor.

groups, according to the number of risk alleles. The genetic score was computed separately for \( m1 \) and \( m2 \), and the DM incidence rate ratio according to genetic score was adjusted for the variables contained in the basic model. Improvement in area under the receiver operating characteristic (ROC) curves was assessed after adding genetic information to the basic model [26]. For this, new-onset DM was predicted at each time point during the study period by considering clinical variables and the genetic background, which was entered as either the 4 SNPs retained in \( m1 \), all 22 SNPs (\( m2 \)), or the calculated genetic score.

Additional sensitivity analyses included the consideration of current and cumulative azidothymidine (AZT) and stavudine (D4T) exposure as the treatment variable instead of ART group, and hepatitis C virus (HCV) coinfection, defined as a positive serology or detectable HCV RNA during follow-up. All statistical analyses were performed using R, version 2.9.2 [27].

RESULTS

Characteristics of participants, ART, and SNPs. Of 661 study participants, those with unsuccessful genotyping (\( n = 4 \)) or with prevalent DM at the beginning of the study period (\( n = 13 \)) were excluded. Results are therefore based on 644 participants whose characteristics are shown in Table 1. There were 94 participants with new-onset DM. They were older and more likely to be men and they had a higher baseline BMI and a lower CD4+ cell count, compared with the 550 participants without new-onset DM, as previously reported [8]. In addition, they were more likely to be treated with PI-containing ART and less likely to be receiving NNRTI-containing ART. The median number of plasma glucose measurements per participant was 19 (interquartile range [IQR], 15–26), and the median number of ART modifications per participant was 5 (IQR, 3–8), during a median follow-up duration of 9.7 years (IQR, 8.6–9.9 years). Cumulative follow-up for all participants was 6054 person-years, and 13,922 measurements of plasma glucose were analyzed. At the time of DM diagnosis, 12 participants (12.8%) had never been exposed to ART and 11 previously ART-treated participants (12%) were not receiving ART. Minor allelic frequencies were similar to previous reports in ethnically similar populations (Table A1, online only). All SNPs in the participants without DM (\( n = 550 \)) were in Hardy-Weinberg equilibrium (\( P > .001 \)).

Factors contributing to new-onset DM. Increasing BMI, age, and triglycerides were associated with DM risk (\( P > .001 \)), as were decreasing HDL-cholesterol (\( P < .01 \)) and CD4+ cell count (\( P = .02 \)). The DM effects of the 22 interrogated SNPs are shown in Figure 1. In the final, multivariable model \( m1 \), adjusted for all clinical and genetic variables, 4 SNPs were retained (rs5219, rs8050136, rs7903146, and rs1801282) (Figure
Figure 3. Distribution of diabetes mellitus (DM) risk alleles in participants with new-onset DM (black bars) and participants without DM (white bars). A, results according to model m1 in which 4 single nucleotide polymorphisms (SNPs) were retained. B, results according to model m2 in which all 22 SNPs were considered.

2). Sensitivity analysis showed essentially unchanged results when considering cumulative instead of current ART exposure (Appendix A, online only).

**Relative contribution of genetic and clinical factors.** In the basic model, 30% of reduced residual deviance was explained by BMI, 19% by age, 9% by the CD4+ cell count, 15% by serum HDL-cholesterol, 9% by serum triglycerides, and 3% by the ART group. Adding the 4 SNPs retained in m1 to the basic model further reduced the residual deviance by 14% (P < .001). Results were similar when all 22 SNPs were added to the basic model (model m2; data not shown).

**DM risk according to genetic score.** Two separate genetic scores were calculated on the basis of the number of DM risk alleles carried by each participant. First, applying model m1 (4 SNPs), the genetic score ranged from 0 to 8 for a diploid genome. Participants were divided into 4 risk groups according to their genetic score, which corresponded to the presence of 0–3, 4, 5, or 6–8 DM risk alleles (Figure 3A). Applying model m2 (22 SNPs), the genetic score groups corresponded to the presence of 0–22, 23–24, 25–26, or 27–44 risk alleles (Figure 3B). With use of either genetic score, participants with new-onset DM were more likely to be in the upper 2 genetic score groups.

With use of the m1-based genetic score, and compared with participants with 0–3 DM risk alleles (reference; incidence rate ratio [IRR] of 1), participants with 4 risk alleles had an IRR of DM of 2.22 (95% confidence interval [CI], 1.27–3.88; P < .01), participants with 5 alleles had an IRR of DM of 2.24 (95% CI, 1.22–4.11; P < .01), and participants with 6–8 risk alleles had an IRR of DM of 3.80 (95% CI, 2.05–7.06; P < .001) (Figure 4A). With use of the m2-based genetic score, participants in the 2 intermediate risk groups had an only slightly increased DM risk (23–24 risk alleles: IRR 1.15; 95% CI, 0.61–2.16; P = .67; and 25–26 risk alleles: 1.69; 95% CI, 0.93–3.07; P = .09), whereas participants with 27–44 risk alleles had an IRR of DM of 2.74 (95% CI, 1.53–4.88; P < .001) (Figure 4B).

**DM prediction using area under the ROC curve.** The inclusion of genetic data together with clinical risk factors slightly improved the area under the ROC curve (Figure 5). Prediction of DM was similar, regardless of how the genetic data was added to the basic model (area under the ROC curve [AUC], 0.75)—that is, whether considering the 4 SNPs retained in m1 (AUC, 0.78), all 22 SNPs (m2; AUC, 0.77), or the genetic score (AUC, 0.77) (Figure A1, online only).

**Alternative models that consider exposure to thymidine analogue reverse-transcriptase inhibitors and HCV coinfection.** In a further sensitivity analysis, current and cumulative exposure to AZT and D4T as well as HCV coinfection were considered instead of ART group. Current exposure to AZT was associated with DM risk in the basic model (P = .02), whereas current D4T exposure (P = .92) and cumulative exposure to AZT (P = .76) and D4T (P = .60) were not associated with DM risk. However, assessment of the relative contribution of genetic and nongenetic factors, DM risk according to m1- or m2-based genetic score, and DM prediction using area under the ROC curve showed essentially unchanged results when considering current AZT exposure rather than ART group (Appendix A, online only).

HCV coinfection (Table 1) did not contribute significantly to DM risk in the basic model (P = .25). In addition, HCV coinfection explained no residual deviance, and DM risk according to m1- or m2-based genetic score was essentially unchanged when HCV coinfection was included in the models (Appendix A, online only).

**DISCUSSION**

In the present study, we found that common SNPs associated with DM in GWAS in the general population also influenced DM risk in HIV-infected individuals. Four SNPs were retained in the fully adjusted, multivariable analysis. These SNPs are known DM loci identified in multiple GWAS: FTO (rs8050136), KCNJ11 (rs5219), TCF7L2 (rs7903146), and PPARG (rs1801282).
Figure 4. Incidence rate ratios (IRRs) of diabetes mellitus (DM) in participants according to genetic score. A, Results according to model 1, in which 4 single nucleotide polymorphisms were retained. The reference group comprises participants carrying 0–3 DM risk alleles (n = 239). B, Results according to model 2. The reference group comprises participants carrying 0–22 DM risk alleles (n = 180).

[13–16, 19, 20, 28, 29]. These 4 loci also showed the most statistically significant DM association in a large study that assessed 16 genetic variants together with key clinical variables [26]. The SNP effects in our study changed little after adjustment for clinical factors. For example, the DM effect of rs8050136 (FTO), which in GWAS correlates with its effect on obesity [14, 28], remained independently associated with DM after adjusting for BMI in the present data set.

At the level of the study population, the contribution of common SNPs to DM risk was similar to the effect of other well-established DM risk factors, such as older age, but the effect of the SNPs was smaller than the effect of BMI. It is notable that the effect of SNPs on DM risk was larger than the potential toxic effects of ART on DM risk. Whether long-term exposure or the acute effect of certain ART agents is more important in regard to DM risk is currently being debated [5, 6, 8, 9]. Importantly, the SNP effects on DM risk were little affected when we considered either current or cumulative exposure to ART groups previously defined in the SHCS [8] and other studies [6, 9, 30] or current and cumulative exposure to thymidine analogue NRTIs.

We followed the trend toward summarizing the effects of multiple SNPs into a single, clinically useful genetic score [31, 32]. Consistent with the results from GWAS, the DM risk attributable to an individual risk allele was modest, whereas the combination of multiple unfavorable SNPs contributed substantially to DM risk. Applying either of 2 genetic scores, participants with the most unfavorable genetic background had ≥2.7-fold increased DM risk. In model 1, only those 4 SNPs with a strong DM association were retained. We also applied an alternative model 2, in which all SNPs were considered (including some with weak DM associations), because the goal of the study was to study all SNPs with a previously identified DM association in the general population. Therefore, m2 was more likely than m1 to assign relatively high genetic scores to participants without DM. However, irrespective of the genetic
Figure 5. Area under the receiver operating characteristic curve (AUC) for basic and basic-plus-genetic models predicting new-onset diabetes mellitus in individual study participants.

model applied, and consistent with previous cohort studies from the general population [26, 33], the addition of genetic data to clinical risk factors increased DM prediction only slightly. The main reason for this is likely that the non-DM control subjects in this study and in reports from the general population [26, 33] may have undiagnosed impaired glucose tolerance without meeting formal DM diagnostic criteria. This notion is consistent with the much stronger predictive capacity of SNPs (AUC = 0.86 for a 15-SNP basic-plus-genetic model) observed in a recent study in the general population that compared DM participants with highly selected control groups of patients who had normal glucose tolerance [34].

Strengths of this study include the prospective, longitudinal data collection; the analysis restricted to SNPs previously identified in GWAS; exploration of different modeling strategies; and inclusion of all participants who developed new-onset DM during enrollment in an established, large observational study over a median follow-up period of >9 years. Our study was limited mainly by sample size. Not all evaluated SNPs were retained in the final multivariable model, most likely because of the generally small effect estimates of the genetic variants. Previous GWAS have relied on the exploitation of data from thousands of individuals to document such modest genetic effects [11–20]. However, several SNPs not retained in the final model m1 were associated with DM in only one GWAS. We have observed in a previous genetic-dyslipidemia study [22] that SNPs identified in GWAS were more likely to be replicated in HIV-infected participants according to the number of GWAS that independently confirmed their effect [35]. Therefore, large data sets will be important in future genetic-DM studies in HIV-infected individuals. This should also permit the stratification of participants according to genetic score and exposure to individual antiretroviral drugs, as in genetic-dyslipidemia studies [22, 36, 37]. Additional limitations of the current study include the relatively small number of women (25% of study participants). Because previous GWAS were conducted in essentially white populations, we restricted our study to white SHCS participants. Because of differences in the tagging characteristics of the array SNPs, however, our findings may not be applicable to other populations.

In summary, our study explored GWAS-identified SNPs that contribute to new-onset DM, a major metabolic complication in HIV-infected individuals. It is likely that some of the DM heritability unexplained by GWAS-identified SNPs can be uncovered [35, 38] by interrogating additional common SNPs associated with DM [39] and with fasting plasma glucose levels [11, 14, 40, 41] or by identifying rare variants through extensive resequencing of the loci identified in GWAS. The potential for
comprehensive genetic and pharmacogenetic prediction of key outcomes of clinical HIV care is now emerging, including prediction of viral load set point and CD4 decline, dyslipidemia, antiretroviral hypersensitivity, response to chronic HCV infection treatment, and lipodystrophy. We are currently conducting a multicohort, GWAS-based genetic study of acute coronary artery disease end points in HIV-infected individuals [35].

Appendix A, online only, provides further details of our study [42–56].

MEMBERS OF THE SHCS


Acknowledgments

We are grateful to Dr Marcel Zwahlen for helpful statistical discussions.

Financial support. SHCS, which is financed by the Swiss National Science Foundation.

Potential conflicts of interest. All authors: no conflicts.

References