Published in final edited form as:

AIDS. 2015 April 24; 29(7): 767–777. doi:10.1097/QAD.0000000000000604.

Associations of Common Variants in the *BST2* Region with HIV-1 Acquisition in African American and European American People who Inject Drugs

Dana B. HANCOCK^a, Nathan C. GADDIS^b, Joshua L. LEVY^b, Laura J. BIERUT^c, Alex H. KRAL^d, and Eric O. JOHNSON^e

^aBehavioral Health Epidemiology Program, Behavioral Health and Criminal Justice Division, Research Triangle Institute (RTI) International, Research Triangle Park, NC

^bResearch Computing Division, RTI International, Research Triangle Park, NC

^cDepartment of Psychiatry, Washington University School of Medicine, St. Louis, MO

^dUrban Health Program, Behavioral Health and Criminal Justice Division, RTI International, San Francisco, CA

^eFellow Program and Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC

Abstract

Objective—The bone marrow stromal cell antigen 2 (*BST2*) gene encodes a host restriction factor that acts as an innate immune sensor of HIV-1 exposure and suppresses release of HIV-1 particles. We aimed to identify associations of variants in the *BST2* gene region with HIV-1 acquisition and disease progression.

Design/methods—Using HIV+ cases and HIV- controls from the Urban Health Study (N=3,136 African Americans and European Americans who inject drugs), we tested 470 variants in *BST2* and its flanking regions for association with HIV-1 acquisition and log-transformed viral load.

Results—We found that the single nucleotide polymorphism (SNP) rs113189798 surpassed the *P* value threshold corrected for multiple testing. The rs113189798-G allele (frequency=16% in African Americans, 4% in European Americans) was associated with increased HIV-1 acquisition risk (meta-analysis $P=1.43 \times 10^{-4}$): odds ratio (95% confidence interval) of 1.22 (1.01-1.49) in African Americans and 2.17 (1.43-3.33) in European Americans. We also found that the previously reported rs12609479-A allele (frequency=35% in African Americans, 81% in European Americans) was nominally associated with decreased risk of acquiring HIV-1 in our study (meta-analysis $P=0.036$). Rs12609479-A is predicted to increase *BST2* expression and thereby decrease risk of acquiring HIV-1. Rs113189798 and rs12609479 were only weakly correlated ($r^2=0.2-0.4$) and represented distinct association signals. None of our tested variants were significantly associated with log-transformed viral load among the HIV+ cases.

Conclusions—Our findings support *BST2* as a genetic susceptibility factor for HIV-1 acquisition: identifying a novel SNP association for rs13189798 and linking the previously reported regulatory SNP rs12609479 to HIV-1 acquisition.

Keywords

genetic association study; HIV-1; innate immunity; single nucleotide polymorphisms; tetherin

INTRODUCTION

An estimated 50% of the variability in acquiring HIV-1 upon exposure is attributable to host genetics [1, 2]. The only genetic variant conclusively associated with HIV-1 acquisition is a 32-base pair deletion in the chemokine (C-C motif) receptor 5 (*CCR5*) gene [3]. Because the deletion allele occurs at low frequency in populations of European descent (>2%) and rarely occurs in non-European populations [4], it accounts for little of the variability in HIV-1 acquisition. Identification of other genetic variants underlying HIV-1 susceptibility is needed to better understand the disease pathogenesis.

Genes encoding host restriction factors that interfere with viral life cycles are strong candidates for influencing risk of HIV-1 acquisition and disease progression. In 2008, bone marrow stromal cell antigen 2 (*BST2*), also known as tetherin, was established as a transmembrane protein that suppresses the release of HIV-1 and other enveloped viral particles into the cell by tethering them to the cell surface [5, 6]. Accumulating evidence has shown that *BST2* also acts as an innate immune sensor of HIV-1 that elicits an inflammatory response upon exposure [7-10]. Both activities, signaling innate immune responses and restricting viral release, are counteracted by the HIV-1 accessory protein *Vpu*, which facilitates the spread of the virus by sequestering *BST2* from the cell surface and thereafter promoting its degradation [6, 11, 12]. Given these indications, *BST2* is recognized as a potential therapeutic target [5, 13].

A few *BST2* variants, mostly single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) located in the promoter, have been studied for influencing HIV-1 acquisition and disease progression. Laplana et al. reported two promoter variants that likely influence *BST2* expression (indel rs3217318 and its tag SNP rs10415893), and using a Spanish population, they found associations with disease progression in 185 HIV-positive cases but no associations with HIV-1 acquisition when comparing the cases to two control sets: 171 HIV-1 exposed and 188 HIV-1 unexposed individuals [14]. Skelton et al. identified seven SNPs and indels by sequencing the *BST2* promoter region in 581 Africans and assessed their regulatory potential; two SNPs (rs73921425 and rs12609479) were shown to potentially alter transcription factor binding sites, but their associations with HIV-1 acquisition and disease progression were not tested [13]. Beyond the promoter, coding SNPs and indels may also influence risk of HIV-1 acquisition and viral load by altering *BST2* function, but only rare coding SNPs have been studied for their functional effects [15].

The aim of our study was to test a dense set of SNPs and indels spanning *BST2* and its flanking regions for association with risk of acquiring HIV-1 (955 HIV+ cases and 2,181 HIV- controls, the largest sample to date) and log-transformed viral load among the cases.

Our study participants were drawn from the Urban Health Study (UHS) of European American (EA) or African American (AA) people who inject drugs. To enhance detection of genetic differences based on HIV status, HIV+ cases were frequency-matched to highly exposed HIV- controls based on several drug use and sexual risk behaviors. Our findings identified a novel and statistically significant SNP association with HIV-1 acquisition and further showed nominal associations for several other SNPs in the region, including a previously reported SNP predicted to influence *BST2* expression.

MATERIALS AND METHODS

HIV-1 cases and controls

The UHS recruited more than 10,000 people who injected drugs in the San Francisco Bay Area from 1986 to 2005 using targeted sampling methods [16-18]. Individuals aged 18 or older, who injected an illicit drug in the past 30 days (verified by signs of venipuncture) and were able to provide informed consent and speak English or Spanish, were invited to participate. Study participants were asked to report key demographics, drug use, and sexual risk behavior and provide blood samples. The Institutional Review Boards at RTI International and the University of California, San Francisco (UCSF) approved all UHS protocols, and all study participants provided written informed consent.

A subset of self-reported AA and EA HIV+ cases and HIV- controls with stored serum samples were selected for genotyping [19]. HIV+ cases were frequency-matched to approximately two HIV- controls, based on HIV risk class, self-reported ancestry, sex, age, and year of ascertainment. The HIV risk class was derived from a latent class analysis revealing three groups, as shown in Supplemental Figure 1, with distinct patterns of drug and sexual exposure risks (receptive needle sharing, multiple sex partners in the past 30 days, sex work, having had a sexually transmitted infection [STI], and receptive anal sex). Class 1, consisting of 15.8% of participants, had a large proportion of people indicating having multiple sex partners, engaging in sex work, and having had an STI. They were also fairly likely to report receptive needle sharing, and had the highest probability of endorsing receptive anal sex. Classes 2 (43.8%) and 3 (40.4%) both had low endorsement of multiple sex partners, receptive anal sex, and sex work but differed on probabilities of having an STI and receptive needle sharing. Corresponding to these differences in HIV risk behavior, particularly probability of needle sharing, Classes 1 and 2 were significantly more likely to be HIV+ than Class 3, even among people who inject drugs (odds ratio [95% confidence interval] = 2.04 [1.74-2.40] and 1.48 [1.29-1.87], respectively). For matching purposes, participants were assigned to only one group based upon their highest posterior probability of group membership.

Viral load for each HIV+ case was determined using serum samples stored at the UCSF AIDS Specimen Bank. Viral load testing was conducted at the Duke Human Vaccine Institute Immunology and Virology Quality Assessment Center using the FDA approved Roche COBAS Ampliprep/COBAS Taqman 48 system. Although plasma is the standard biospecimen for determining viral load, viral load is robustly detectable from serum samples [20], and there is a high correlation of quantified viral load between serum and plasma samples across multiple studies regardless of assay used [21-25]. Thus, serum is a valid

source for viral load within studies where all participants had the same type of biospecimen used for determining viral load. In our protocol, 1 mL samples were added to the assay system, from which viral RNA was extracted and reverse-transcribed. The assay then amplified and detected a portion of the HIV-gag gene, and the viral load results were generated as the number of viral copies per mL. Each run included three kit controls (high positive, low positive, and negative) and an internal control that was tracked over many runs for monitoring the within-laboratory variability. Thus, viral load was determined from a single time point rather than determining viral load set point from multiple measures. However, it is unlikely that many participants were ascertained during the acute phase of HIV infection [17], and we have excluded those participants reporting treatment for HIV, missing treatment data, or having AIDS. The median $\log_{10}$ viral load among treatment naïve UHS participants analyzed here is 4.34 with an inter-quartile range (IQR) of 2.31-6.45, which is similar to other studies of treatment naïve viral load set point using multiple measures (median=4.27, IQR=3.58-4.73 [26]; median=4.45, IQR=4.32-5.14 [27]).

Genotyping, quality control, and imputation

While genotyping whole genome amplified DNA from serum samples has proved problematic [28], genotyping of serum-derived genomic DNA (gDNA) is valid but technique dependent [29, 30]. As described elsewhere for this study [19], the Rutgers Cell and DNA Repository extracted gDNA from serum without additional whole genome amplification. The Center for Inherited Disease Research (CIDR) at the Johns Hopkins University, performed restoration of the samples using the Illumina Formalin-Fixed Paraffin-Embedded kit to improve gDNA quality and then performed genotyping using the Illumina HumanOmni1Quad BeadChip with a reagent substitute from Illumina. Quality statistics were very high for the restored samples. The overall missing rate was 0.97%. Genotype concordance rates were high among HapMap control samples (99.7%) and blind duplicates (99.9%), indicating overall high genotyping quality [31].

The STRUCTURE program [32] was used to corroborate self-identified ancestry for almost all participants via comparisons to HapMap reference populations of West Africans (denoted YRI), European Americans (denoted CEU), and Chinese (denoted CHB) from the International HapMap phase III project [33] using 10,000 SNPs randomly distributed across the genome. The ancestries of these original HapMap populations have been highly characterized. We classified AAs as having African ancestry >25% and EAs as having African ancestry <25%. All other quality control steps were implemented using PLINK [34] unless otherwise stated. Genotyped SNPs were excluded due to call rate <90% or Hardy-Weinberg equilibrium $p < 0.0001$. Across *BST2* and its 50kb flanking regions, there remained 23 SNPs genotyped with high quality: their call rates ranging from 96.1% to 100% (Supplemental Table 1).

Participant-level quality control began with 2,444 AA and 1,288 EA samples that were successfully genotyped with call rate >90%, representing 90.5% of the 4,126 attempted samples. Genotyped participants were excluded due to unexpected duplication resulting from the serial nature of the UHS design (identity-by-state estimates >90%; 322 AAs, 97 EAs), other sources of sample duplication (42 AAs, 30 EAs), first-degree relatedness (58

AAs based on identity-by-descent estimates >40% and kinship coefficient >0.177 and 5 EAs based on identity-by-descent estimates >40%), or sex discordance (12 AAs, 5 EAs). At the duplication and relatedness steps, the participant with the highest call rate from each pair or cluster was retained. The additional kinship coefficient at the relatedness step in AAs, derived using the KING program, was used to circumvent the inflation of identity-by-descent estimates due to population stratification [35]. None of the genotyped participants had excessive homozygosity. STRUCTURE analyses were repeated to corroborate the ancestral estimates for the 2,010 AAs and 1,151 EAs passing all aforementioned quality control steps. After removing two EA participants with high Chinese ancestry (>80%) but little European ancestry (8-12%), there were 2,010 AAs and 1,149 EAs.

SNP and indel genotype imputation was conducted in AAs and EAs, separately, using IMPUTE2 [36] with reference to the 1000 Genomes Phase I integrated variant set release (v3) ALL haplotype panel (accessed on April 19, 2012 at https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html). Our imputation procedure was chosen based on our prior optimization of imputation performance for studies with AAs [37]. Imputations were preceded by prephasing the study genotypes with the ShapeIT program [38] to estimate haplotypes. SNPs and indels with MAF<0.5% in either ancestry group were filtered out, and the IMPUTE2 probabilities for each of the three genotypes were converted into a single imputed genotype dosage value (a fractional value between 0 and 2 indicating the expected number of minor allele copies). Using the genotype dosage as the predictor in regression models testing for disease association has been recommended as an effective approach to account for imputation uncertainty [39, 40]. Following the association testing, we used the IMPUTE2 info metric to evaluate the SNP/indel imputation quality rather than imposing an imputation quality filter and possibly missing truly associated SNPs/indels [8, 41].

Statistical analyses

We tested 470 genotyped and imputed SNPs and indels, located from 50kb upstream to 50kb downstream of the *BST2* gene, for association with HIV-1 case/control status using logistic regression models stratified by ancestry and adjusted for age, sex, HIV risk class (based on the latent class analysis), survey year (1986-1994 or 1995-2002 to account for pre/post antiretroviral therapy availability), and the first 10 principal component eigenvectors to minimize bias due to population stratification. The eigenvectors were computed using genotyped SNPs in linkage equilibrium (pairwise $r^2 < 0.2$) in the EIGENSTRAT program [42]. The final analytic data set included 3,136 participants who were successfully genotyped, passed all quality control procedures, and had complete covariate data: 2,004 AAs (628 HIV+ cases and 1,376 HIV- controls) and 1,132 EAs (327 HIV+ cases and 805 HIV- controls).

Among the HIV+ cases, 479 AAs and 211 EAs were used to test SNP associations with log-transformed viral load ($\log_{10}$ copies of virus/mL) using linear regression models with adjustment for age, sex, survey year, and the 10 eigenvectors. Participants who had viral load data available, who reported no current treatment, and whose disease had not progressed to AIDS were included. The HIV+ cases without viral load data available

occurred due to their sample not amplifying through PCR, failing to detect virus, or presenting viral copies below the quantifiable range. This rate of sample failure was expected given the systematically lower viral load counts observed in serum samples, as compared to plasma samples [20].

The ancestry-specific association results for both HIV acquisition and log-transformed viral load were combined via fixed-effects sample size-weighted meta-analysis, as done in prior multi-ancestry meta-analyses [43, 44], using the METAL program [45]. This strategy enhances statistical power with a larger sample size and emphasizes associations present across ancestry groups, but it reduces the detection of ancestry-specific associations [46, 47]. $P < 0.05$ was used to declare nominal significance, and $P < 3.85 \times 10^{-4}$, which corresponded to correction for 130 independent tests taking into account the correlations among the 470 tested SNPs and indels [48, 49], was used to declare statistical significance.

Bioinformatics analyses

Pairwise SNP linkage disequilibrium patterns were discerned using LocusZoom [50] and Haploview [51] with reference to the 1000 Genomes populations of European (denoted EUR) or African (denoted AFR) ancestry; r^2 (square of the correlation coefficient) measures the degree to which one SNP can predict another, whereas D' (the standardized coefficient of disequilibrium) is useful for identifying blocks of adjacent linked SNPs without being constrained by varied allele frequencies. SNP regulatory potential was assessed using four bioinformatics databases: (1) RegulomeDB [52], (2) SNPinfo [53], (3) HaploReg v2 [54], and (4) University of Chicago expression quantitative trait loci (eQTL) browser [55]. Each database provided unique information on predicted regulatory states from the Encyclopedia of DNA Elements, the Roadmap Epigenome Mapping Consortium, the Genotype-Tissue Expression project, large-scale QTL studies, and elsewhere.

RESULTS

Statistically significant association between rs113189798 and HIV-1 acquisition

We tested 470 genotyped and 1000 Genomes-imputed SNPs and indels across *BST2* and its flanking regions for association with HIV-1 acquisition. SNPs and indels associated at meta-analysis $P < 0.05$ are presented in Table 1. Results of all SNPs and indels are presented in Supplemental Table 2. The intergenic SNP rs113189798, located 17.8 kb downstream of *BST2* and 7.8 kb upstream of the plasmalemma vesicle associated protein (*PLVAP*) gene (Figure 1), was significantly associated with HIV-1 acquisition at meta-analysis $P = 1.43 \times 10^{-4}$, surpassing the P value threshold of 3.85×10^{-4} (overall $\alpha = 0.05/130$ independent tests among the 470 variants). The rs113189798-G allele, which occurred at frequencies of 16% in AAs and 4% in EAs, was significantly associated in EAs and nominally associated in AAs with increased risk of HIV-1 acquisition (Table 1). Several other SNPs in moderate to strong LD with rs113189798 were associated with HIV-1 acquisition (Figure 1 and Supplemental Figure 2) (smallest $P = 6.7 \times 10^{-5}$ for rs149275055, $r^2 = 1$ with rs113189798).

Rs113189798 was not associated with log-transformed viral load among HIV-1 infected cases (meta-analysis $P=0.60$). None of the other 469 tested SNPs or indels were significantly associated with log-transformed viral load at meta-analysis $P<3.85\times 10^{-4}$ (Supplemental Table 3).

Nominal association between the previously reported rs12609479 and HIV-1 acquisition

Of the 9 SNPs and indels previously reported by Laplana et al. for HIV progression [14] or Skelton et al. [13] (associated with differences in HIV prevalence across African regions), 4 were not available in 1000 Genomes: -411 i19/ 19 (rs3217318), -160 i1/ 1 (rs28413175), -187 A>G, and +26 i6/ 6 (rs28413176). The other 5 variants were captured in our association testing with HIV-1 acquisition and log-transformed viral load: -193 G>A (or C>T, rs28413174), -199 G>A (or C>T, rs73921425), -201 C>T (or G>A, rs12609479), -225 C>T (or G>A, rs112492472), and +4016 G>A (rs10415893). Although none of these SNPs were significantly associated in our study, the rs12609479-A allele, which occurred at frequency=65% in AAs and frequency=19% in EAs, was nominally associated with decreased risk of acquiring HIV (Table 2, meta-analysis $P=0.036$). Rs12609479-A was nominally associated with log-transformed viral load in AAs at $P=0.042$ (β [standard error] = 0.12 [0.058]), but the association was not corroborated in EAs ($P=0.21$, β [standard error] = -0.12 [0.098]). None of the other previously reported SNPs were associated with HIV-1 acquisition or log-transformed viral load at $P<0.05$.

Linkage disequilibrium patterns among the previously reported and newly associated SNPs from our study (meta-analysis $P<0.05$) are presented in reference to the 1000 Genomes AFR (Supplemental Figure 3A-B) and EUR (Supplemental Figure 3C-D) populations.

Rs12609479 was correlated with some SNPs associated at nominal significance in our study, but rs12609479 was not correlated with our significantly associated SNP rs113189798, especially in African populations ($r^2=0.001$ and $D'=0.078$). The linkage disequilibrium between rs12609479 and rs113189798 was constrained by varied allele frequencies in the EUR panel ($r^2=0.21$, $D'=1$). We repeated the association analyses additionally adjusting for rs12609479. The rs113189798-G association with HIV-1 acquisition (meta-analysis $P=4.65\times 10^{-4}$) did not surpass the $P<3.85\times 10^{-4}$ threshold used in the primary association testing, but there was little change in its magnitude of association in both ancestry groups: odds ratio (95% confidence interval) of 1.20 (0.99-1.47) in AAs and 2.27 (1.41-3.57) in EAs.

Regulatory potential of HIV-associated SNPs and indels

Rs12609479 was reported as a putative QTL that alters a transcription factor binding site in the *BST2* promoter region [13]. We evaluated the regulatory potential of rs12609479, the significantly associated SNP rs113189798, and all other SNPs and indels associated with HIV-1 acquisition at meta-analysis $P<0.05$, by using information contained in RegulomeDB [52], SNPinfo [53], and Haploreg v2 [54]. These results are presented in Supplemental Table 4. Rs12609479 was predicted to alter transcription factor binding sites and several other regulatory motifs, serve as an active promoter and a strong enhancer, reside in a DNase hypersensitivity site (indicative of transcriptional activity), and bind numerous

proteins. Rs113189798 was implicated as a weak enhancer located in a DNase hypersensitivity site that alters three regulatory motifs.

We also queried the University of Chicago QTL browser [55], which revealed 6 SNPs in *BST2* and its 50kb flanking regions as QTLs. Four of the 6 SNPs were identified by Degner et al. [56] as DNase sensitivity QTLs in lymphoblast cell lines from African subjects: (1) rs73921425, shown as chr19.17377626 based on its NCBI build 36 position, was reported by Skelton et al. [13] and located 2 base pairs away from rs12609479; (2) rs28413174, shown as chr19.17377620, was also reported by Skelton et al. [13] but was not associated with HIV-1 acquisition in our study (meta-analysis $P=0.098$); (3) rs8102791, shown as chr19.17366933, was associated with HIV-1 acquisition at meta-analysis $P=8.18 \times 10^{-3}$; and (4) rs8106139, shown as chr19.17367176, was associated with HIV-1 acquisition at meta-analysis $P=4.72 \times 10^{-3}$. The two other identified QTLs – rs919265 [57] and rs34984238, shown as chr19.17415857 [56] – were not associated with HIV-1 acquisition (meta-analysis $P=0.57$ and 0.14 , respectively). None of the SNPs identified as QTLs were associated with log-transformed viral load (smallest meta-analysis $P=0.53$ for rs8102791).

DISCUSSION

The functional roles of *BST2* in signaling innate immune responses upon exposure and inhibiting viral particle release gives it strong biological plausibility for influencing HIV-1 acquisition and disease progression and makes it a potential therapeutic target. However, few genetic studies have tested *BST2* and flanking variants for association with HIV-1 outcomes. Our study identified a novel and statistically significant association between rs113189798 and HIV-1 acquisition: its minor allele (G) conferring an increased risk for acquiring HIV-1 across two ancestry groups. Several variants correlated with rs113189798 showed nominally significant associations with HIV-1 acquisition, and bioinformatics analyses revealed strong evidence of regulatory roles for many of these newly implicated SNPs and indels on *BST2* expression. Rs113189798 and its correlated variants are located between the *PLVAP* and *BST2* genes, and we cannot rule out the possibility that the associations may reflect involvement of *PLVAP*, which has no prior implications for HIV-1.

Our study also found that the rs12609479-A allele, previously reported by Skelton et al. as a putative regulatory SNP that alters key transcription factor binding sites [13], was nominally associated with decreased risk of acquiring HIV-1 across EAs and AAs. This direction of association is consistent with the Skelton et al. findings; the rs12609479-A allele is predicted to increase *BST2* expression, which in turn is expected to decrease risk of acquiring HIV-1 due to more *BST2* being available to enhance innate immune sensing and hinder dissemination of HIV-1 particles [13]. The protective rs12609479-A allele occurs as the major allele in EAs (frequency=81%) but the minor allele in AAs (frequency=35%). These frequencies are consistent with 1000 Genomes EUR (frequency=77%) and AFR (frequency=31%) populations. Skelton et al. suggested that this allele frequency disparity may contribute to differences in HIV-1 disease prevalence across ancestry groups [13]. The protective alleles for rs12609479 (A) and rs113189798 (A) occur frequently together (77.4% and 25.0% of the haplotypes in the 1000 Genomes European and African populations, respectively). However, their r^2 values are low, and follow-up analyses of rs113189798,

conditioned on rs12609479, indicated that our novel SNP association represents an independent signal.

Not all of the previously reported *BST2* variants were available for testing in our study. Missense SNPs identified by Sauter et al. were not available due to their rare frequencies [15]. Laplana et al. reported two promoter variants associated with disease progression. The low frequency indel rs3217318 was not available for imputation in 1000 Genomes; its tag SNP rs10415893 was available but not associated with HIV-1 acquisition (meta-analysis $P=0.11$) or log-transformed viral load (meta-analysis $P=0.58$) in our study [14]. Overall, our study did not detect any significant SNP or indel associations with log-transformed viral load. Our lack of findings for viral load may be explained by the suggestion that *BST2* inhibits transmission between individuals more than it limits within-individual disease progression [14, 58]. Alternatively, having a single measurement of viral load in this study, rather than determining viral load set point from multiple measures over time (see Methods), may have weakened our ability to detect associations. That said, however, we observed that the median and range of viral load in this study is comparable to other studies of treatment naïve cohorts [26, 27], and we corroborated associations between known variants in *HCP5* (meta-analysis $P=4.25\times 10^{-6}$ for rs2255221) and *HLA-B* (meta-analysis $P=1.26\times 10^{-3}$ for rs9264942) [59, 60]. Additional adjustment for rs2255221 and rs9264942 resulted in only negligible changes in our observed SNP/indel associations with viral load (results not shown).

In this largest genetic association study focusing on *BST2* and HIV-1 to date, we identified rs113189798 and other correlated SNPs/indels as novel variants associated with HIV-1 acquisition and extended prior findings linking the putative regulatory SNP rs12609479 to HIV-1 acquisition. Identification of these *BST2* SNP associations was enabled by using 1000 Genomes imputation, moderate sample size ($N=3,159$), and study design accounting for HIV exposure in the control group; consideration of all these facets are needed in future *BST2* genetic association studies. The SNPs identified here for HIV-1 acquisition were not associated with viral load among HIV+ cases, and no other SNP associations with viral load surpassed our statistical significance threshold. This apparent discrepancy in association findings between HIV-1 acquisition and disease progression may be explained by the dual roles of *BST2* being “genetically separable,” meaning that the DNA sequence encoding functional elements for innate immune signaling are distinct from the sequence underlying the restriction of viral release [8-10]. Our study provides support to *BST2* as a genetic susceptibility factor specifically for acquiring HIV-1. Our newly implicated SNPs and indels in the *BST2* gene region merit further study to determine whether they exert important regulatory effects or tag other regulatory variants that underlie their associations with HIV-1 acquisition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by the National Institute on Drug Abuse (NIDA) grants R01 DA026141 and R01 DA038632. Storage and processing of the UHS serum samples for DNA extraction and genotyping was conducted by the Rutgers University Cell and DNA Repository was supported by the NIDA Center for Genetics (contract number N01DA-09-7770). Genotyping was conducted at CIDR and supported by NIH contracts HHSN268201100011I and HHSN268200782096C. These UHS data are available in the database of Genotypes and Phenotypes via accession number phs000454.v1.p1.

L.J.B., A.H.K., and E.O.J. obtained funding and contributed to the overall study design; D.B.H., N.C.G., and J.L.L. conducted statistical analyses; D.B.H. and E.O.J. drafted the article; and all authors interpreted the results and contributed to the final version of the article. L.J.B. is listed as an inventor on Issued U.S. Patent 8,080,371, “Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Otherwise, the authors declare that there are no potential conflict of interest related to this manuscript.

Source of Funding: This work was supported by the National Institute on Drug Abuse (NIDA) grant numbers R01 DA026141 and R01 DA038632. Storage and processing of the Urban Health Study serum samples for DNA extraction and genotyping was conducted by the Rutgers University Cell and DNA Repository was supported by the NIDA Center for Genetics (contract number N01DA-09-7770). Genotyping was conducted by the Center for Inherited Disease and supported by NIH contracts HHSN268201100011I and HHSN268200782096C.

REFERENCES

1. Telenti A, Johnson WE. Host genes important to HIV replication and evolution. *Cold Spring Harb Perspect Med*. 2012; 2:a007203. [PubMed: 22474614]
2. Loeuillet C, Deutsch S, Ciuffi A, Robyr D, Taffe P, Munoz M, et al. In vitro whole-genome analysis identifies a susceptibility locus for HIV-1. *PLoS Biol*. 2008; 6:e32. [PubMed: 18288889]
3. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, et al. Resistance to -1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine HIV receptor gene. *Nature*. 1996; 382:722–725. [PubMed: 8751444]
4. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet*. 1997; 16:100–103. [PubMed: 9140404]
5. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*. 2008; 451:425–430. [PubMed: 18200009]
6. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, et al. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe*. 2008; 3:245–252. [PubMed: 18342597]
7. Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y, et al. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene*. 2003; 22:3307–3318. [PubMed: 12761501]
8. Galao RP, Le Tortorec A, Pickering S, Kueck T, Neil SJ. Innate sensing of HIV-1 assembly by Tetherin induces NFkappaB-dependent proinflammatory responses. *Cell Host Microbe*. 2012; 12:633–644. [PubMed: 23159053]
9. Tokarev A, Suarez M, Kwan W, Fitzpatrick K, Singh R, Guatelli J. Stimulation of NF-kappaB activity by the HIV restriction factor BST2. *J Virol*. 2013; 87:2046–2057. [PubMed: 23221546]
10. Hotter D, Sauter D, Kirchhoff F. Emerging role of the host restriction factor tetherin in viral immune sensing. *J Mol Biol*. 2013; 425:4956–4964. [PubMed: 24075872]
11. Douglas JL, Viswanathan K, McCarroll MN, Gustin JK, Fruh K, Moses AV. Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a {beta}TrCP-dependent mechanism. *J Virol*. 2009; 83:7931–7947. [PubMed: 19515779]
12. Hauser H, Lopez LA, Yang SJ, Oldenburg JE, Exline CM, Guatelli JC, et al. HIV-1 Vpu and HIV-2 Env counteract BST-2/tetherin by sequestration in a perinuclear compartment. *Retrovirology*. 2010; 7:51. [PubMed: 20529266]
13. Skelton MM, Kampire EE, Wonkam AA, Mhandire KK, Kumwenda JJ, Duri KK, et al. Frequency Variation Among Sub-Saharan Populations in Virus Restriction Gene, BST-2 Proximal Promoter Polymorphisms: Implications for HIV-1 Prevalence Differences Among African Countries. *OMICS*. 2014

14. Laplana M, Caruz A, Pineda JA, Puig T, Fibla J. Association of BST-2 gene variants with HIV disease progression underscores the role of BST-2 in HIV type 1 infection. *J Infect Dis.* 2013; 207:411–419. [PubMed: 23148293]
15. Sauter D, Hotter D, Engelhart S, Giehler F, Kieser A, Kubisch C, et al. A rare missense variant abrogates the signaling activity of tetherin/BST-2 without affecting its effect on virus release. *Retrovirology.* 2013; 10:85. [PubMed: 23937976]
16. Kral AH, Bluthenthal RN, Lorvick J, Gee L, Bacchetti P, Edlin BR. Sexual transmission of HIV-1 among injection drug users in San Francisco, USA: risk-factor analysis. *Lancet.* 2001; 357:1397–1401. [PubMed: 11356437]
17. Kral AH, Lorvick J, Gee L, Bacchetti P, Rawal B, Busch M, et al. Trends in human immunodeficiency virus seroincidence among street-recruited injection drug users in San Francisco, 1987–1998. *Am J Epidemiol.* 2003; 157:915–922. [PubMed: 12746244]
18. Watters JK, Biernacki P. Targeted Sampling: Options for the Study of Hidden Populations. *Social Problems.* 1989; 36:416–430.
19. Johnson EO, Hancock DB, Gaddis NC, Levy JL, Page GP, Novak SP, et al. Novel genetic locus implicated for HIV-1 acquisition with putative regulatory links to HIV replication and infectivity: a genome-wide association study. *PLoS One.* in press.
20. Lew J, Reichelderfer P, Fowler M, Bremer J, Carrol R, Cassol S, et al. Determinations of levels of human immunodeficiency virus type 1 RNA in plasma: reassessment of parameters affecting assay outcome. TUBE Meeting Workshop Attendees. Technology Utilization for HIV-1 Blood Evaluation and Standardization in Pediatrics. *J Clin Microbiol.* 1998; 36:1471–1479. [PubMed: 9620364]
21. Vahey, M.; Wong, MT. A quantitative liquid hybridization polymerase chain reaction methodology employing storage phosphor technology.. In: Diffenbach, C., editor. Cold Spring Harbor Laboratory manual for PCR. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y.: 1995.
22. Dickover RE, Herman SA, Saddiq K, Wafer D, Dillon M, Bryson YJ. Optimization of specimen-handling procedures for accurate quantitation of levels of human immunodeficiency virus RNA in plasma by reverse transcriptase PCR. *J Clin Microbiol.* 1998; 36:1070–1073. [PubMed: 9542939]
23. Rodriguez RJ, Dayhoff DE, Chang G, Cassol SA, Birx DL, Artenstein AW, et al. Comparison of serum and plasma viral RNA measurements in primary and chronic human immunodeficiency virus type 1 infection. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1997; 15:49–53. [PubMed: 9215654]
24. Holodniy M, Mole L, Yen-Lieberman B, Margolis D, Starkey C, Carroll R, et al. Comparative stabilities of quantitative human immunodeficiency virus RNA in plasma from samples collected in VACUTAINER CPT, VACUTAINER PPT, and standard VACUTAINER tubes. *J Clin Microbiol.* 1995; 33:1562–1566. [PubMed: 7650187]
25. Griffith BP, Rigsby MO, Garner RB, Gordon MM, Chacko TM. Comparison of the Amplicor HIV-1 monitor test and the nucleic acid sequence-based amplification assay for quantitation of human immunodeficiency virus RNA in plasma, serum, and plasma subjected to freeze-thaw cycles. *J Clin Microbiol.* 1997; 35:3288–3291. [PubMed: 9399536]
26. Helleberg M, Kronborg G, Larsen CS, Pedersen G, Pedersen C, Obel N, et al. No change in viral set point or CD4 cell decline among antiretroviral treatment-naïve, HIV-1-infected individuals enrolled in the Danish HIV Cohort Study in 1995–2010. *HIV Med.* 2013; 14:362–369. [PubMed: 23433434]
27. Kumwenda JJ, Makanani B, Taulo F, Nkhoma C, Kafulafula G, Li Q, et al. Natural history and risk factors associated with early and established HIV type 1 infection among reproductive-age women in Malawi. *Clin Infect Dis.* 2008; 46:1913–1920. [PubMed: 18462100]
28. Croft DT Jr, Jordan RM, Patney HL, Shriver CD, Vernalis MN, Orchard TJ, et al. Performance of whole-genome amplified DNA isolated from serum and plasma on high-density single nucleotide polymorphism arrays. *J Mol Diagn.* 2008; 10:249–257. [PubMed: 18403606]
29. Ivarsson MI, Dillner J, Carlson J. Validity of maternal genotypes in DNA from archival pregnancy serum samples. *Clin Chem.* 2009; 55:842–843. [PubMed: 19325020]

30. Hoque A, Goodman P, Ambrosone CB, Figg WD, Price DK, Kopp W, et al. Extraction of DNA from serum for high-throughput genotyping: findings from pilot studies within the Prostate Cancer Prevention Trial. *Urology*. 2008; 71:967–970. [PubMed: 18267333]
31. Romm, J.; Pottinger, S.; Aker, L.; Hilton, C.; Ongaco, H.; Doheny, K. Experience with Illumina Infinium FFPE DNA Restoration Solution. American Society of Human Genetics; San Francisco: 2012.
32. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
33. International HapMap C. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010; 467:52–58. [PubMed: 20811451]
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
35. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010; 26:2867–2873. [PubMed: 20926424]
36. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011; 1:457–470. [PubMed: 22384356]
37. Hancock DB, Levy JL, Gaddis NC, Bierut LJ, Saccone NL, Page GP, et al. Assessment of Genotype Imputation Performance using 1000 Genomes in African American Studies. *PLoS One*. 2012; 7:e50610. [PubMed: 23226329]
38. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012; 44:955–959. [PubMed: 22820512]
39. Zheng J, Li Y, Abecasis GR, Scheet P. A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genet Epidemiol*. 2011; 35:102–110. [PubMed: 21254217]
40. Liu K, Luedtke A, Tintle N. Optimal methods for using posterior probabilities in association testing. *Hum Hered*. 2013; 75:2–11. [PubMed: 23548776]
41. Beecham GW, Martin ER, Gilbert JR, Haines JL, Pericak-Vance MA. APOE is not associated with Alzheimer disease: a cautionary tale of genotype imputation. *Ann Hum Genet*. 2010; 74:189–194. [PubMed: 20529013]
42. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
43. Guo Y, Lanktree MB, Taylor KC, Hakonarson H, Lange LA, Keating BJ. Gene-centric meta-analyses of 108 912 individuals confirm known body mass index loci and reveal three novel signals. *Hum Mol Genet*. 2013; 22:184–201. [PubMed: 23001569]
44. Lanktree MB, Guo Y, Murtaza M, Glessner JT, Bailey SD, Onland-Moret NC, et al. Meta-analysis of Dense Genecentric Association Studies Reveals Common and Uncommon Variants Associated with Height. *Am J Hum Genet*. 2011; 88:6–18. [PubMed: 21194676]
45. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–2191. [PubMed: 20616382]
46. Pulit SL, Voight BF, de Bakker PI. Multiethnic genetic association studies improve power for locus discovery. *PLoS One*. 2010; 5:e12600. [PubMed: 20838612]
47. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol*. 2011; 35:809–822. [PubMed: 22125221]
48. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004; 74:765–769. [PubMed: 14997420]
49. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*. 2005; 95:221–227. [PubMed: 16077740]

50. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26:2336–2337. [PubMed: 20634204]
51. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc*. 2009; 2009 pdb ip71.
52. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012; 22:1790–1797. [PubMed: 22955989]
53. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009; 37:W600–605. [PubMed: 19417063]
54. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012; 40:D930–934. [PubMed: 22064851]
55. Wright FA, Shabalin AA, Rusyn I. Computational tools for discovery and interpretation of expression quantitative trait loci. *Pharmacogenomics*. 2012; 13:343–352. [PubMed: 22304583]
56. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*. 2012; 482:390–394. [PubMed: 22307276]
57. Veyrieras JB, Kudravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet*. 2008; 4:e1000214. [PubMed: 18846210]
58. Sauter D, Schindler M, Specht A, Landford WN, Munch J, Kim KA, et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe*. 2009; 6:409–421. [PubMed: 19917496]
59. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. 2010; 330:1551–1557. [PubMed: 21051598]
60. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science*. 2007; 317:944–947. [PubMed: 17641165]

Figure 1.

Association plots for the *BST2* gene region on chromosome 19. The $-\log_{10}(P)$ values correspond to the meta-analysis for HIV-1 acquisition across African Americans and European Americans. The r^2 correlation values between the significantly associated SNP (rs113189798 in purple) and other SNPs in *BST2* and its 50kb flanking regions were based on 1000 Genomes populations of (A) European or (B) African ancestry. Circles indicate imputed SNPs, and squares indicated SNPs that were genotyped and used as input for imputation.

(A)

(B)

Table 1 *BS72 and flanking single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) associated with HIV-1 acquisition in the Urban Health Study at meta-analysis $p < 0.05$. SNPs and indels are sorted by the meta-analysis P value.*

SNP/indel	Position (NCBI build 37)	Coded allele	African Americans (N=2,004)				European Americans (N=1,132)				Meta-analysis P value
			Coded allele frequency	a info	P value	OR (95 CI%)	Coded allele frequency	a info	P value	OR (95% CI)	
rs113189798	17,495,949	G	0.16	0.82	0.043	1.22 (1.01-1.49)	0.04	0.96	2.75×10^{-4}	2.17 (1.43-3.33)	1.43×10^{-4}
rs17457472	17,493,610	C	0.93	0.91	0.15	0.82 (0.62-1.08)	0.96	0.97	2.16×10^{-4}	0.45 (0.30-0.69)	7.44×10^{-4}
rs116444576	17,492,625	G	0.94	0.90	0.19	0.83 (0.62-1.10)	0.96	0.97	2.04×10^{-4}	0.45 (0.30-0.69)	1.02×10^{-3}
rs8106139	17,506,176	T	0.57	0.94	0.026	1.18 (1.02-1.36)	0.15	0.98	0.080	1.25 (0.97-1.61)	4.72×10^{-3}
rs76197715	17,496,141	A	0.98	0.77	0.54	0.86 (0.53-1.39)	0.96	0.97	1.80×10^{-4}	0.44 (0.29-0.68)	6.18×10^{-3}
rs113964659	17,495,958	A	0.97	0.78	0.54	0.88 (0.58-1.33)	0.96	0.96	1.92×10^{-4}	0.44 (0.29-0.68)	6.43×10^{-3}
rs8102791	17,505,933	A	0.58	0.94	0.038	1.16 (1.01-1.34)	0.15	0.98	0.10	1.24 (0.96-1.59)	8.18×10^{-3}
rs12972417	17,525,314	T	0.68	-	0.064	1.15 (0.99-1.33)	0.18	-	0.079	1.23 (0.98-1.55)	0.011
rs2886110	17,529,525	T	0.65	0.92	0.066	1.15 (0.99-1.35)	0.20	0.92	0.073	1.23 (0.98-1.55)	0.011
rs140112879	17,495,332	G	0.99	0.70	0.77	0.90 (0.44-1.83)	0.96	0.97	1.69×10^{-4}	0.44 (0.29-0.68)	0.013
rs11667840	17,509,006	C	0.69	0.93	0.031	1.19 (1.02-1.39)	0.21	0.88	0.20	1.17 (0.92-1.47)	0.013
rs2162889	17,464,411	G	0.51	0.91	0.15	1.11 (0.96-1.28)	0.29	0.94	0.036	1.25 (1.01-1.54)	0.016
rs8109902	17,509,460	G	0.41	0.84	0.092	1.14 (0.98-1.32)	0.19	0.86	0.081	1.25 (0.97-1.60)	0.017
rs8107250	17,532,350	T	0.65	0.90	0.067	1.16 (0.99-1.35)	0.21	0.90	0.12	1.20 (0.95-1.51)	0.017
rs11668883	17,527,370	G	0.69	0.98	0.058	1.16 (0.99-1.34)	0.19	0.96	0.16	1.18 (0.94-1.49)	0.018
rs4808632	17,518,544	C	0.48	0.93	0.075	0.88 (0.77-1.01)	0.85	0.97	0.12	0.82 (0.63-1.06)	0.019
rs9783971	17,536,982	T	0.69	0.80	0.075	0.86 (0.73-1.02)	0.86	0.86	0.13	0.80 (0.61-1.06)	0.019
rs2364551	17,526,009	C	0.69	0.98	0.064	1.15 (0.99-1.34)	0.20	0.96	0.16	1.18 (0.94-1.49)	0.020
rs11882366	17,518,018	C	0.49	0.93	0.049	0.87 (0.75-1.00)	0.86	0.96	0.24	0.85 (0.66-1.11)	0.022
rs12985272	17,543,681	G	0.99	0.59	0.52	0.74 (0.29-1.85)	0.95	0.67	3.04×10^{-3}	0.48 (0.29-0.78)	0.022
rs148422811	17,490,987	A	0.99	0.83	0.96	1.02 (0.45-2.31)	0.96	0.97	1.25×10^{-4}	0.44 (0.29-0.67)	0.024
rs117262143	17,491,099	T	0.99	0.83	0.96	1.02 (0.45-2.31)	0.96	0.97	1.26×10^{-4}	0.44 (0.29-0.67)	0.024
rs11671286	17,467,180	T	0.32	0.90	0.19	0.90 (0.77-1.05)	0.71	0.93	0.044	0.81 (0.65-0.99)	0.025
rs1560091	17,469,034	G	0.34	0.89	0.18	0.90 (0.77-1.05)	0.72	0.94	0.052	0.81 (0.66-1.00)	0.025

SNP/indel	Position (NCBI build 37)	Coded allele	African Americans (N=2,004)				European Americans (N=1,132)				Meta-analysis <i>P</i> value
			Coded allele frequency	<i>a</i> info	<i>P</i> value	OR (95 CI)%	Coded allele frequency	<i>a</i> info	<i>P</i> value	OR (95% CI)	
rs8101151	17,554,927	G	0.84	0.89	0.048	0.82 (0.68-1.00)	0.80	0.93	0.27	0.87 (0.68-1.11)	0.028
rs186203383	17,492,314	A	0.99	0.81	0.93	1.03 (0.49-2.19)	0.96	0.97	1.41×10 ⁻⁴	0.44 (0.29-0.67)	0.028
rs12971834	17,516,689	T	0.65	0.97	0.082	1.14 (0.98-1.32)	0.18	0.97	0.17	1.18 (0.93-1.49)	0.028
rs9300058	17,511,492	G	0.64	0.96	0.077	1.14 (0.99-1.32)	0.18	0.99	0.19	1.17 (0.93-1.48)	0.028
rs12979773	17,516,764	C	0.65	0.96	0.085	1.14 (0.98-1.32)	0.18	0.97	0.17	1.18 (0.93-1.49)	0.028
rs2278233	17,517,174	T	0.61	0.94	0.091	0.89 (0.77-1.02)	0.87	0.97	0.16	0.82 (0.63-1.08)	0.028
rs149275055	17,490,668	A	0.99	0.81	0.76	1.14 (0.49-2.66)	0.96	0.96	6.72×10 ⁻⁵	0.41 (0.27-0.64)	0.031
rs200791202	17,509,459	AG	0.40	0.89	0.23	1.09 (0.95-1.26)	0.14	0.96	0.046	1.30 (1.00-1.69)	0.031
rs8107534	17,508,678	A	0.37	0.92	0.27	1.09 (0.94-1.26)	0.14	0.97	0.035	1.32 (1.02-1.72)	0.032
rs140677148	17,497,859	T	0.99	0.84	0.90	1.05 (0.47-2.38)	0.96	0.97	2.04×10 ⁻⁴	0.45 (0.29-0.68)	0.033
rs200474195	17,493,728	TTTG	0.58	0.59	0.023	1.23 (1.03-1.47)	0.53	0.59	0.61	1.07 (0.84-1.36)	0.034
rs4808633	17,518,877	C	0.47	0.90	0.095	0.88 (0.77-1.02)	0.84	0.95	0.19	0.84 (0.66-1.09)	0.034
rs56009618	17,510,453	G	0.64	0.94	0.062	1.15 (0.99-1.33)	0.19	0.96	0.31	1.13 (0.89-1.43)	0.036
rs12609479	17,516,628	G	0.65	0.96	0.10	1.13 (0.98-1.31)	0.19	0.97	0.19	1.17 (0.92-1.48)	0.036
rs143487272	17,521,839	G	0.97	0.60	0.11	1.51 (0.91-2.50)	0.88	0.6	0.18	1.29 (0.89-1.88)	0.037
rs2303679	17,531,650	C	0.68	0.84	0.21	0.90 (0.77-1.06)	0.86	0.91	0.075	0.78 (0.60-1.02)	0.037
rs10422271	17,525,503	C	0.54	0.92	0.15	0.90 (0.78-1.04)	0.87	0.97	0.12	0.81 (0.62-1.06)	0.038
rs17272023	17,467,214	G	0.95	0.77	0.46	0.88 (0.62-1.25)	0.85	0.94	0.014	0.72 (0.56-0.94)	0.039
rs62125237	17,549,824	G	0.71	0.95	0.069	1.16 (0.99-1.35)	0.61	0.96	0.32	1.10 (0.91-1.33)	0.04
rs73923943	17,565,714	C	0.82	0.92	0.11	0.86 (0.72-1.03)	0.99	0.64	0.21	0.53 (0.20-1.42)	0.041
rs35822396	17,512,364	AG	0.49	0.94	0.16	1.11 (0.96-1.28)	0.17	0.97	0.13	1.20 (0.94-1.54)	0.043
rs10416855	17,525,029	A	0.54	0.92	0.13	0.90 (0.78-1.03)	0.87	0.98	0.19	0.84 (0.64-1.09)	0.045
rs10420449	17,525,420	G	0.54	0.92	0.13	0.90 (0.78-1.03)	0.87	0.98	0.19	0.84 (0.65-1.09)	0.045
rs10421592	17,525,921	G	0.54	0.92	0.13	0.90 (0.78-1.03)	0.87	0.98	0.19	0.84 (0.65-1.09)	0.045
rs11672798	17,508,855	T	0.64	0.93	0.094	1.13 (0.98-1.32)	0.21	0.90	0.30	1.13 (0.90-1.43)	0.049

CI, confidence interval; OR, odds ratio

^aThe info metric, as provided by IMPUTE2, is indicative of imputation quality. It ranges from 0 to 1, with values near 1 indicating that the SNP was imputed with high certainty. A dash (-) indicates that the variant was genotyped and used as input for imputation.

Table 2

Associations of the 5 previously reported *BST2* SNPs with HIV-1 acquisition in our study, sorted by *P* value of the meta-analysis across African Americans and European Americans from the Urban Health Study.

SNP (ref.)	Coded allele	African Americans (N=2,004)				European Americans (N=1,132)				Meta-analysis <i>P</i> value	
		Coded allele frequency	<i>a</i> info ^a	<i>P</i> value	OR (95% CI)	Coded allele frequency	<i>a</i> info ^a	<i>P</i> value	OR (95% CI)		
rs12609479 [13]	A	0.35	0.96	0.10	0.88 (0.76-1.02)	0.81	0.97	0.19	0.85 (0.68-1.09)	0.036	
rs28413174 [13]	C	0.70	0.94	0.36	0.93 (0.80-1.08)	0.87	0.97	0.12	0.81 (0.62-1.06)	0.098	
rs10415893 [14]	G	0.74	-	0.45	0.94 (0.81-1.10)	0.88	-	0.10	0.80 (0.60-1.05)	0.11	
rs112492472 [13]	G	0.96	0.82	0.16	0.77 (0.53-1.11)	1.00	NA	NA	NA	0.36	
rs73921425 [13]	C	0.91	0.85	0.74	0.96 (0.75-1.23)	1.00	NA	NA	NA	0.38	

^aThe info metric, as provided by IMPUTE2, is indicative of imputation quality. It ranges from 0 to 1, with values near 1 indicating that the SNP was imputed with high certainty. A dash (-) indicates that the variant was genotyped and used as input for imputation. NA, not available due to the SNP being monomorphic in the respective ancestry group