Human Molecular Genetics, 2013, Vol. 22, No. 9 ***1903–1910***
doi:10.1093/hmg/ddt033
Advance Access published on January 30, 2013

A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A

Jérôme Lane[1,†], Paul J. McLaren[1,2,3,†], Lucy Dorrell[4], Kevin V. Shianna[5], Amanda Stemke[6], Kimberly Pelak[5], Stephen Moore[4], Johannes Oldenburg[7], Maria Teresa Alvarez-Roman[8], Anne Angelillo-Scherrer[9], Francoise Boehlen[10], Paula H.B. Bolton-Maggs[11], Brigit Brand[12], Deborah Brown[13], Elaine Chiang[14], Ana Rosa Cid-Haro[15], Bonaventura Clotet[16], Peter Collins[17], Sara Colombo[2], Judith Dalmau[16], Patrick Fogarty[18], Paul Giangrande[19], Alessandro Gringeri[20], Rathi Iyer[21], Olga Katsarou[22], Christine Kempton[23], Philip Kuriakose[24], Judith Lin[25], Mike Makris[26], Marilyn Manco-Johnson[27], Dimitrios A. Tsakiris[28], Javier Martinez-Picado[16], Evelien Mauser-Bunschoten[29], Anne Neff[30], Shinichi Oka[31], Lara Oyesiku[19], Rafael Parra[32], Kristiina Peter-Salonen[33], Jerry Powell[34], Michael Recht[35], Amy Shapiro[36], Kimo Stine[37], Katherine Talks[38], Amalio Telenti[2], Jonathan Wilde[39], Thynn Thynn Yee[40], Steven M. Wolinsky[41], Jeremy Martinson[42], Shehnaz K. Hussain[43], Jay H. Bream[44], Lisa P. Jacobson[45], Mary Carrington[46,47], James J. Goedert[48], Barton F. Haynes[6], Andrew J. McMichael[4], David B. Goldstein[5] and Jacques Fellay[1,2,5,*] for the NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI)

[1]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland [2]Institute of Microbiology, University Hospital and University of Lausanne, Lausanne, Switzerland [3]Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA [4]Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK [5]Center for Human Genome Variation, Duke University School of Medicine, Durham, NC, USA [6]Duke Human Vaccine Institute, Duke University, Durham, NC, USA [7]Institute of Experimental Haematology and Transfusion Medicine, University Clinic Bonn, Bonn, Germany [8]Unidad de Coagulopatias Congenitas, La Paz University Hospital, Madrid, Spain [9]Service and Central Laboratory of Hematology, Lausanne University Hospital, Lausanne, Switzerland [10]Division of Angiology and Haemostasis, Department of Specialties of Medicine, University Hospitals, Geneva, Switzerland [11]Manchester Haemophilia Comprehensive Care Centre, Manchester Royal Infirmary, University of Manchester, Manchester, UK [12]Division of Hematology, University Hospital, Zurich, Switzerland [13]Gulf States Hemophilia and Thrombophilia Center, Houston, TX, USA [14]University of Pennsylvania, Penn Comprehensive Hemophilia and Thrombosis Program, Philadelphia, PA, USA [15]Unidad de Coagulopatias Congenitas, La Fe University Hospital, Valencia, Spain [16]AIDS Research Institute IrsiCaixa, Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain [17]Arthur Bloom Haemophilia Centre, School of Medicine, Cardiff University, Wales, UK [18]University of California San Francisco, San Francisco, CA, USA [19]Oxford Haemophilia and Thrombosis Centre, Oxford University Hospitals NHS Trust, Oxford, UK [20]Università degli Studi di Milano and Fondazione IRCCS Ca' Granda Ospedale Policlinico di Milano, Milan, Italy [21]UMHC Clinic for Bleeding and Clotting Disorders, University of Mississippi Medical Center, Jackson, MS, USA [22]Blood Centre and National Reference Centre for Congenital Bleeding Disorders, Laiko General Hospital, Athens, Greece [23]Adult Hemophilia Program, Emory University, Atlanta, GA, USA [24]Henry Ford Health System, Detroit, MI, USA [25]Boston Hemophilia Center, Brigham and Women's Hospital, Boston, MA, USA

*To whom correspondence should be addressed at: EPFL-SV-GHI, Station 19, 1015 Lausanne, Switzerland. Tel: +41 216931849; Fax: +41 216937220; Email: jacques.fellay@epfl.ch
†These authors contributed equally to this work.

[26]Department of Cardiovascular Science, University of Sheffield, Sheffield, UK [27]Department of Pediatrics, University of Colorado and Children's Hospital Colorado, Aurora, USA [28]Hemophilia Comprehensive Care Center, University Hospital, Basel, Switzerland [29]University Medical Center Utrecht, Utrecht, Netherlands [30]Hemostasis–Thrombosis Clinic, Vanderbilt University, Nashville, TN, USA [31]National Center for Global Health and Medicine, Tokyo, Japan [32]Hemophilia Unit, Hospital Vall d'Hebron, Barcelona, Spain [33]Hematology, Inselspital, University Hospital Bern, Bern, Switzerland [34]UC Davis Hemophilia Treatment Center, Sacramento, CA, USA [35]The Hemophilia Center, Oregon Health & Science University, Portland, OR, USA [36]Indiana Hemophilia & Thrombosis Center, Indianapolis, IN, USA [37]Hemophilia Treatment Center of Arkansas, Arkansas Children's Hospital, Little Rock, AR, USA [38]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK [39]New Queen Elizabeth Hospital, Edgbaston, Birmingham, UK [40]Katharine Dormandy Haemophilia Centre & Thrombosis Unit, Royal Free Hospital, London, UK [41]Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, IL, USA [42]Department of Pathology, School of Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, USA [43]Department of Epidemiology, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA [44]Department of Molecular Microbiology and Immunology and [45]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA [46]Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC Frederick, Inc., Frederick National Lab, Frederick, MD, USA [47]Ragon Institute of MGH, MIT and Harvard, Charlestown, MA, USA [48]Division of Cancer Epidemiology and Genetics, NCI, Bethesda, MD, USA

Received November 22, 2012; Revised January 18, 2013; Accepted January 28, 2013

Human genetic variation contributes to differences in susceptibility to HIV-1 infection. To search for novel host resistance factors, we performed a genome-wide association study (GWAS) in hemophilia patients highly exposed to potentially contaminated factor VIII infusions. Individuals with hemophilia A and a documented history of factor VIII infusions before the introduction of viral inactivation procedures (1979–1984) were recruited from 36 hemophilia treatment centers (HTCs), and their genome-wide genetic variants were compared with those from matched HIV-infected individuals. Homozygous carriers of known *CCR5* resistance mutations were excluded. Single nucleotide polymorphisms (SNPs) and inferred copy number variants (CNVs) were tested using logistic regression. In addition, we performed a pathway enrichment analysis, a heritability analysis, and a search for epistatic interactions with *CCR5* Δ32 heterozygosity.

A total of 560 HIV-uninfected cases were recruited: 36 (6.4%) were homozygous for *CCR5* Δ32 or m303. After quality control and SNP imputation, we tested 1 081 435 SNPs and 3686 CNVs for association with HIV-1 serostatus in 431 cases and 765 HIV-infected controls. No SNP or CNV reached genome-wide significance. The additional analyses did not reveal any strong genetic effect.

Highly exposed, yet uninfected hemophiliacs form an ideal study group to investigate host resistance factors. Using a genome-wide approach, we did not detect any significant associations between SNPs and HIV-1 susceptibility, indicating that common genetic variants of major effect are unlikely to explain the observed resistance phenotype in this population.

INTRODUCTION

A single exposure to a transfusion with HIV-contaminated blood or clotting factor concentrates carries a risk of infection of 90%, far greater than that of any other risk exposure (1). This is clearly illustrated by the disastrous consequences of widespread use of contaminated plasma-derived clotting factor concentrates for the treatment of hemophilia in the early days of the current AIDS pandemic.

Hemophilia A is an inherited X-linked bleeding disorder affecting 1 in 5000–1 in 10 000 males. It is caused by mutations in the factor 8 gene (*F8*) on the X chromosome, leading to reduced levels of factor VIII (FVIII) activity in the circulation. Replacement of FVIII is necessary to prevent morbidity and mortality associated with uncontrolled bleeding. The use of donor FVIII concentrates derived from pooled plasma from up to 100 000 donors was the mainstay of hemophilia treatment until recombinant factor products were introduced in the 1990s (2,3). Prior to 1984, factor concentrates were not subjected to viral inactivation processes and as a result, a large proportion of patients with hemophilia A became infected with HIV-1 (4–8). The risk of infection was correlated with the severity of the disease: individuals with severe hemophilia experienced more bleeding episodes requiring treatment with FVIII concentrates or were treated prophylactically two or three times per week, and the quantity

of concentrates correlated directly with the probability of acquiring HIV-1 infection (9). However, infection was not universal, even among patients with severe hemophilia. Individuals, who were likely exposed, yet remain HIV uninfected are here referred to as exposed uninfected (EU).

It is already known that human genetic variation contributes to differences in susceptibility to HIV-1 infection. Homozygosity for a 32 bp deletion in the gene coding for the HIV-1 co-receptor *CCR5* results in the absence of *CCR5* expression at the cell surface, offering protection against R5 strains of HIV-1, the usual infecting virus (10–12). The *CCR5* Δ32/Δ32 homozygous genotype is found in ∼1% of healthy individuals of European descent, but is rare in non-European populations (13). A rarer mutation in *CCR5*, m303, also abrogates expression and confers resistance in homozygotes or compound heterozygotes with *CCR5* Δ32 (14). No other host genetic polymorphism has been consistently shown to protect against HIV-1 acquisition (15). Of note, several studies in Europe and North America have shown that the frequency of *CCR5* Δ32/Δ32 is significantly higher in HIV-uninfected hemophiliacs than in the general population (up to 15% compared with ≤1%), with the highest frequencies in those with severe hemophilia (16–20). In contrast, *CCR5* Δ32/Δ32 is rare or absent in hemophiliacs who acquired HIV infection (20).

We here present a genome-wide association study (GWAS) that aims at discovering additional genetic polymorphisms associated with reduced susceptibility to HIV-1. A clearer understanding of host genetic resistance against HIV-1 infection is of enormous importance to identifying novel prophylactic drug targets as well as correlates of protection for rational HIV-1 vaccine design.

RESULTS

Study participants and genotypes

A total of 483 individuals with hemophilia A were included in the CHAVI014 protocol from 36 hemophilia treatment centers (HTCs) in nine countries (Table 1). Samples from an additional group of 77 highly exposed hemophiliacs were obtained from the Multicenter Hemophilia Cohort Study (MHCS) (6,19). All study participants received potentially contaminated FVIII concentrates between 1979 and 1984 and had at least one documented negative HIV-1 screening test at a later date. Most patients had severe hemophilia and were positive for chronic or resolved hepatitis C virus (HCV) infection, a marker of blood-borne pathogen exposure (Table 1). The control population comprised of 823 HIV positive, ethnically matched individuals from the Multicenter AIDS Cohort Study (MACS).

To reduce heterogeneity, we excluded the only three female subjects who were recruited in CHAVI014. Targeted genotyping of the *CCR5* protective variants identified 35 hemophilia cases as homozygous for the Δ32 deletion and one case as homozygous for the m303 variant. There were no Δ32/m303 compound heterozygotes. We observed a consistent enrichment for *CCR5* Δ32 homozygosity across study sites and countries, closely reflecting the known north–south decreasing cline of Δ32 allele frequency in European populations: 9.4% of individuals of northern European ancestry, 2.6% of central Europeans and 0.9% of southern Europeans were found to be homozygous, versus 2, 0.5 and 0.1% in the respective general populations (13,21). Because the HIV EU phenotype of the 36 subjects carrying *CCR5* homozygous mutations was already explained genetically, they were not included in the GWAS. The frequency of Δ32 heterozygosity was not increased in the EU cohort ($n = 92/560$, 16.4%) in comparison to control populations, implying the absence of additional *CCR5* variants that could form protective compound heterozygotes with the Δ32 deletion.

A total of 521 of the initial sample of 560 EU cases and 823 HIV positive controls were genotyped: 16 samples did not pass initial quality control filtering. An additional 43 individuals were removed due to cryptic relatedness: this high number of related individuals in our study population is unsurprising considering the familial clustering of hemophilia. Finally, we excluded 90 population outliers that were identified through principal component analysis of the genotyping data. The final study population consisted in 430 EU hemophilia cases and 765 HIV positive controls. After all quality control steps, 890 599 single-nucleotide polymorphisms (SNPs) were used for imputation based on the HapMap 3 CEU reference set, resulting in a total of 1 081 435 SNPs used for association testing. The coordinates of nine significant principal component axes were included as covariates in all regression models. Using PennCNV, we identified 2543 deletions and 1143 duplications in 3375 variable genomic regions: the number of copy number variants (CNVs) was consistent with data from the 1000 Genomes Project (22).

Power calculation

With our final numbers of cases and controls, the study had ∼80% power to detect associated variants with a genotype relative risk (GRR) of two or more. Table 2 shows the GRR required for a polymorphism to be significantly associated with resistance against HIV infection at the genome-wide level ($P_{\text{threshold}} = 5\text{E}-08$), under different genetic models and with various minor allele frequencies (MAFs).

Genome-wide association analyses

After imputation, we tested all SNPs for association with HIV resistance under additive, dominant and recessive models using logistic regression. No SNP reached genome-wide significance (Fig. 1A) under any of the genetic models. The distribution of observed *P*-values was very similar to the null expectation, as shown in Figure 1B: a λ value of 1.01 indicated that the association statistics were not confounded by population stratification. As a comparison, the *CCR5* Δ32 variant strongly associated with HIV resistance under a recessive genetic model ($P = 9.4\text{E}-15$) in the initial study population consisting of 560 cases and 823 controls.

Since the only known genetic variant associated with HIV resistance (*CCR5* Δ32) also impacts HIV disease progression (10), we sought to increase power for detecting genetic effects by meta-analyzing the current association results with

Table 1. Characteristics of EU hemophilia cases

Male gender (*n*, %)	557 (99.5)
Caucasian ethnicity (*n*, %)	516 (92.1)
Birth year (median, IQR)	1967 (1954–1976)
Country (*n*, %)	
USA	203 (36.3)
UK	110 (19.6)
Switzerland	61 (10.9)
Italy	60 (10.7)
Netherlands	40 (7.1)
Spain	36 (6.4)
Greece	20 (3.6)
Germany	19 (3.4)
Japan	11 (2)
Severity of hemophilia (*n*, %)[a]	
Severe (<1% normal FVIII activity)	406 (84.1)
Moderate (1% <normal FVIII activity <5%)	77 (15.9)
Number of FVIII infusions 1979–1984 (median, IQR)[a]	51 (7–296)
Hepatitis C status (*n*, %)[a]	
Never infected	29 (6)
Spontaneous clearance	70 (14.5)
Chronically infected	262 (54.2)
Successfully treated	122 (25.3)
Protective *CCR5* genotype: homozygosity Δ32 or m303	36 (6.4)

[a]Information was only available for the 483 CHAVI014 participants.

Table 2. Minimal GRR required for 80% power for variant detection in the genome-wide association analyses

MAF	Genetic model		
	Additive	Dominant	Recessive
5%	2.9	3	36.5
20%	2	2.4	5.3

Power was calculated using the present sample size and a genome-wide significance threshold of $P < 5E-08$.

those from a GWAS on HIV control (23). We observed no genome-wide significant associations after meta-analysis of results under additive models or when combining additive results from the HIV control GWAS with recessive model results from the present study (i.e. imitating the known effects of *CCR5* Δ32).

We further addressed whether combining SNP effects or testing for epistatic interactions between genome-wide SNPs and *CCR5* Δ32 heterozygosity could improve sensitivity for detecting genetic influences on HIV resistance. To achieve this, we performed pathway enrichment, a heritability analysis assessing additive genetic effects and a genome-wide interaction analysis with *CCR5* Δ32. In all cases, no significant evidence for enrichment or interaction was observed further, suggesting a lack of strong genetic effects of common variants on HIV resistance.

DISCUSSION

Individuals with hemophilia, who were exposed to potentially contaminated blood products, yet were not infected by HIV-1 in the early years of the pandemic, form an ideal study group to investigate host resistance factors. Our study represents an unprecedented effort to identify and prospectively recruit such individuals. Through an international collaboration involving 36 HTCs in nine countries and three continents, we obtained informed consent, clinical information and genetic material from 483 subjects. Those were combined with an additional set of EU individuals from the MHCS, resulting in a total number of 560 cases, which were compared with a higher number of ethnically matched controls at more than one million polymorphic sites across the genome.

The possibility of identifying associated variants depends on an actual enrichment of resistance alleles—or depletion of susceptibility alleles—in the case population. The incorrect inclusion of non-exposed individuals (misclassification bias) would decrease study power, because they would most likely be susceptible. Therefore, we applied strict selection criteria to ensure that our EU subjects had a very high likelihood of effective exposure to HIV-1. All included cases had a documented history of treatment with FVIII concentrates with a high likelihood of HIV-1 contamination. Due to the severity of hemophilia, they received a relatively high number of FVIII infusions (median 51), each derived from pooled plasma from thousands of donors. Most subjects were infected by HCV, confirming actual exposure to blood-borne pathogens. The majority of patients treated in the same HTCs were infected before 1984 (4–8). Finally, the observed enrichment of *CCR5* Δ32 homozygosity in cases (6.4% versus ≤1% in European populations) is a clear indicator of effective HIV-1 exposure.

The choice of an adequate control population represented an essential step in the study design. We chose to compare EU cases with HIV-1-infected patients, to make certain that all controls were in fact susceptible. Alternative options would have been to select either unexposed hemophilia subjects or population-level samples as controls. We considered, however, that there was no advantage in matching cases and controls for a monogenic disease that is genetically unrelated to HIV-1 susceptibility, and that using either alternative group could reduce power because potentially resistant subjects would not be excluded. An additional concern was that hemophilia cases were selected on the basis of their resistance to intravenously administered blood products, whereas controls would largely consist of individuals infected through mucosal exposure. This is unlikely to lead to substantial bias, as mechanisms involved in susceptibility or resistance in the intravenous compartment should also impact the likelihood of HIV-1 acquisition after mucosal exposure, as observed for *CCR5*-associated resistance.

Our genome-wide analysis did not reveal any statistically significant association between SNPs or CNVs and resistance against HIV-1 infection. Furthermore, we did not find any evidence for genetic effects in a pathway enrichment analysis, a heritability analysis, and a search for epistatic interactions with *CCR5* Δ32 heterozygosity. Our results strongly suggest that common genetic variants with a major effect do not play a major role in determining susceptibility to HIV-1 in our study population. During the past two decades, several cohorts of EU individuals, identified by different modes of HIV-1 exposure, have been studied for genetic factors that might account for their apparent resistance to infection (24), but only *CCR5* variation has been consistently shown to confer any degree of protection. Additional gene variants

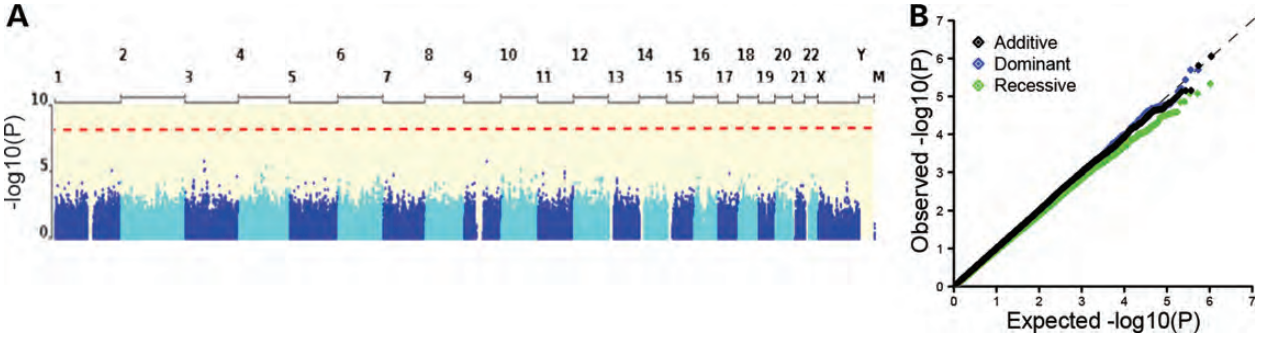

Figure 1. Logistic regression results. (**A**) Manhattan plot of association results from the additive model shows no significant association signal throughout the genome (the dotted red line indicates the significance threshold of 5E−08). (**B**) QQ plot demonstrates that the observed distribution of *P*-values corresponds to the expected distribution under the null hypothesis for all models (additive in black, dominant in blue, recessive in green), indicating that potential confounders are well controlled.

were reported to protect against acquisition or to increase susceptibility to infection, but they are, at best, supported by weak statistical evidence, and none has been convincingly replicated (15). The negative result of our GWAS—one of the largest genetic studies of HIV-1 acquisition to date, performed in the most exposed population, in accordance with the latest standard of genomic research—confirms the absence of common protective variants of large effect in individuals of Western European ancestry, beyond *CCR5*: it also means that similarly sized GWASs are unlikely to reveal any genetic association in lower exposure cohorts (sexual transmission, intravenous drug use, mother-to-child transmission).

The exclusive focus on whites is obviously an important limitation of our study. A small number of Japanese individuals were recruited in the CHAVI014 protocol ($n = 11$), but could not be included in the association analysis due to concerns about stratification and to a lack of ethnically matched controls. To date, the only GWAS of HIV-1 acquisition performed in non-whites did not identify any association in a population from Malawi (25). Additional, large-scale studies of individuals from various ethnic backgrounds are a clear priority in the field.

Our GWAS was designed to identify genetic variants with relatively high MAFs (>5%) and moderate to high GRR (≥2). Consequently, two categories of genetic variants could still be involved in the HIV-1 resistance phenotype observed in epidemiological studies: common variants with weak effects and rare variants (26,27). Several efforts are underway to explore these non-exclusive possibilities. The HIV host genetics community is notably joining forces to run a large meta-analysis of existing genome-wide studies (28). Also, thanks to affordable sequencing technology, it is now feasible to perform exhaustive searches for rare causal variants throughout the exome or the genome of EU individuals. The combination of these approaches in various populations will finally delineate the impact of human genetic diversity on HIV-1 susceptibility.

MATERIALS AND METHODS

Study participants

The CHAVI014 study was set up to obtain peripheral blood specimens from HIV-1 exposed, yet uninfected subjects with hemophilia A to study the genetic factors that may influence susceptibility and resistance to HIV-1 infection. Potentially eligible patients were identified from a search of hemophilia registries, surveillance programme databases or other HTC databases. Individuals 18 years of age or older with moderate or severe hemophilia A (<5 IU/dl or <1 IU/dl normal FVIII activity, respectively) were eligible if they had documented treatment with a plasma-derived FVIII concentrate between 1 January 1979 and 31 December 1984 and documented HIV-negative test. The number of treatment episodes during the high-risk exposure period was recorded. In a separate recruitment effort, retrospective samples from a well-characterized cohort of highly exposed, yet uninfected hemophiliacs were obtained through collaboration with J.J.G. and the MHCS (6,19).

HIV-infected individuals from the MACS that had been genotyped in the context of a previous GWAS (23) were used as controls: none of them were hemophiliacs, and the predominant mode of HIV acquisition was homosexual contact. Briefly, the MACS was established in 1983 and includes 6972 adult homosexual and bisexual men from four metropolitan areas, Baltimore, Chicago, Los Angeles and Pittsburgh, recruited during three recruitment periods, 1984–1985, 1987–1991 and 2001–2003.

Local institutional review boards at each participating center approved the study, and all participants provided informed consent for genetic testing.

Genotypes

Genomic DNA was extracted from 10 ml of whole blood. *CCR5* Δ32 and m303 genotypes were assessed by in-house taqman assays: individuals with known *CCR5* protective genotypes (homozygosity for any of the variant or compound heterozygosity) were excluded from the subsequent GWAS. Genome-wide genotyping was done on Illumina Human 1M or 1Mduo SNP chip, containing 1 072 820 and 1 271 154 SNPs, respectively.

We carried out a series of SNP and sample quality control procedures: polymorphisms were filtered based on missingness (dropped if called in <99% of participants), MAF (dropped if the value was <1%) and severe deviation from Hardy–Weinberg equilibrium (dropped if $P < 5E-08$).

Samples were filtered based on genotyping quality (dropped if call rate <95%) and a gender check (heterozygosity testing). SNPs were then used to identify cryptic relatedness between study participants: we estimated the sharing of genetic information by estimating identity by descent (IBD), and excluded one randomly selected individual in each pair of DNA samples showing >12.5% of estimated IBD, corresponding to first-degree cousins. To control for the possibility of spurious associations resulting from population stratification, we used a modified Eigenstrat method, which derives the principal components of the correlations among gene variants (29): population outliers were discarded, and the coordinates of the significant principal component axes were included in the association tests to correct for residual stratification.

To increase the coverage of genomic variation, we imputed the genotyping data using MACH software with HapMap 3 CEU as a reference set (30); SNPs with r^2 <0.3 and/or an MAF <1% were removed.

CNVs were derived from non-imputed SNPs using PennCNV software (31), separately for the 1M and the 1Mduo chips. To avoid spurious CNV calls, deletions or duplications overlapping centromeric, telomeric and immunoglobulin regions were discarded. Finally, only CNVs present in at least two samples were considered for association analysis.

Association analyses

Logistic regression was used to assess the differences in genotype frequencies of SNPs and CNVs between EU individuals and HIV-infected controls under additive, recessive and dominant genetic models, and for interaction analysis between genome-wide SNPs and *CCR5* Δ32 heterozygosity. To control for population structure, the coordinates of five significant Eigenstrat axes were included as covariates in all models. We used the CaTS Power Calculator for Genome-Wide Association Studies software (http://www.sph.umich.edu/csg/abecasis/CaTS/) for power calculations, PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) for logistic regression analyses and WGAViewer (http://compute1.lsrc.duke.edu/softwares/WGAViewer/) for evaluation and annotation of the association statistics. Bonferroni's correction was applied for multiple testing.

Meta-analysis of association results from an HIV control GWAS

The *CCR5* Δ32 deletion provides proof of concept that genetic variants can impact both HIV acquisition and disease progression. Thus, we sought to improve power for variant detection by combining association results from the HIV resistance analysis with those from a GWAS on HIV control. The HIV control dataset includes 815 members of the Swiss HIV Cohort Study typed on the Illumina HumanHap 550 BeadChip and imputed using the HapMap 3 European sample as reference as previously described (23,32). We meta-analyzed the results from additive genetic models in both studies by combining *z*-scores that incorporated effect direction in both studies (assuming variants that decrease HIV susceptibility also decrease set point viral load) (33). For combining additive association results from the GWAS on HIV control with recessive model results from the recent study (i.e. directly mimicking the observed *CCR5* Δ32 effect), we used Fisher's method for combining *P*-values.

Assessment of enrichment of SNPs association signal in biological pathways

We used MAGENTA (34) to search for abundance of SNPs association signal across pathways using the default parameters to define gene boundaries (mapping to hg18) and to correct for the confounding effects of gene size and linkage disequilibrium between SNPs. Approximately 10 000 gene sets defined by publicly available resources were used. We further added custom gene sets relevant to HIV biology defined by: whole genome siRNA knockdown screens (35), human–HIV protein interactions (36) and a curated list of interferon-stimulated genes (37). We used the false discovery rate *P*-value correction within MAGENTA to assess significance.

Heritability analysis

To investigate a role for measurable, additive genetic contributions to the HIV resistance phenotype, we used GCTA (38). We performed strict sample and SNP quality control as described in (39) and estimated the total genetic variance explained by genome-wide SNPs and the narrow-sense heritability assuming a trait prevalence of 1%. Genetic variance was estimated using the underlying liability scale with the narrow-sense heritability calculated as the proportion of total phenotypic variance that is due to additive genetic effects.

ACKNOWLEDGEMENTS

We would like to thank all study participants and health care workers at the contributing hemophilia treatment centers, as well as Jo Roberts in Oxford for administration and coordination of the CHAVI014 study. The Multicenter AIDS Cohort Study (MACS, http://www.statepi.jhsph.edu/macs/macs.html) has centers (Principal Investigators) at The Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa P. Jacobson), Howard Brown Health Center, Feinberg School of Medicine, Northwestern University and Cook County Bureau of Health Services (John P. Phair, Steven M. Wolinsky), University of California, Los Angeles (Roger Detels) and University of Pittsburgh (Charles R. Rinaldo).

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology (CHAVI) (AI067854). This project has been funded in part with federal funds from the Frederick National Laboratory for Cancer Research (HHSN261200800001E). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products

or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the National Institutes of Health, Frederick National Lab, Center for Cancer Research. The Multicenter AIDS Cohort Study is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute (UO1-AI-35042, UL1-RR025005, UO1-AI-35043, UO1-AI-35039, UO1-AI-35040, UO1-AI-35041).

REFERENCES

1. Baggaley, R., Boily, M., White, R. and Alary, M. (2006) Risk of HIV-1 transmission for parenteral exposure and blood transfusion: a systematic review and meta-analysis. *AIDS*, **20**, 805–812.
2. Fricke, W.A. and Lamb, M.A. (1993) Viral safety of clotting factor concentrates. *Semin. Thromb. Hemost.*, **19**, 54–61.
3. Morgenthaler, J.J. (2001) Securing viral safety for plasma derivatives. *Transfus. Med. Rev.*, **15**, 224–233.
4. AIDS-Hemophilia French Study Group (1985) Immunologic and virologic status of multitransfused patients: role of type and origin of blood products. *Blood*, **66**, 896–901.
5. Ragni, M., Winkelstein, A., Kingsley, L., Spero, J. and Lewis, J. (1987) 1986 update of HIV seroprevalence, seroconversion, AIDS incidence, and immunologic correlates of HIV infection in patients with hemophilia A and B. *Blood*, **70**, 786–790.
6. Kroner, B., Rosenberg, P., Aledort, L., Alvord, W. and Goedert, J. (1994) HIV-1 infection incidence among persons with hemophilia in the United States and western Europe, 1978–1990. Multicenter Hemophilia Cohort Study. *J. Acquir. Immune Defic. Syndr. (1999)*, **7**, 279–286.
7. Darby, S.C., Kan, S.W., Spooner, R.J., Giangrande, P.L., Lee, C.A., Makris, M., Sabin, C.A., Watson, H.G., Wilde, J.T. and Winter, M. (2004) The impact of HIV on mortality rates in the complete UK haemophilia population. *AIDS*, **18**, 525–533.
8. Dietrich, S.L., Mosley, J.W., Lusher, J.M., Hilgartner, M.W., Operskalski, E.A., Habel, L., Aledort, L.M., Gjerset, G.F., Koerper, M.A., Lewis, B.H. *et al.* (1990) Transmission of human immunodeficiency virus type 1 by dry-heated clotting factor concentrates. Transfusion Safety Study Group. *Vox Sang.*, **59**, 129–135.
9. Gjerset, G.F., Clements, M.J., Counts, R.B., Halvorsen, A.S. and Thompson, A.R. (1991) Treatment type and amount influenced human immunodeficiency virus seroprevalence of patients with congenital bleeding disorders. *Blood*, **78**, 1623–1627.
10. Dean, M., Carrington, M., Winkler, C., Huttley, G.A., Smith, M.W., Allikmets, R., Goedert, J.J., Buchbinder, S.P., Vittinghoff, E., Gomperts, E. *et al.* (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*, **273**, 1856–1862.
11. Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A. and Landau, N.R. (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, **86**, 367–377.
12. Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C. *et al.* (1996) Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, **382**, 722–725.
13. Martinson, J.J., Chapman, N.H., Rees, D.C., Liu, Y.-T. and Clegg, J.B. (1997) Global distribution of the CCR5 gene 32-basepair deletion. *Nat. Genet.*, **16**, 100–103.
14. Quillent, C., Oberlin, E., Braun, J., Rousset, D., Gonzalez-Canali, G., Metais, P., Montagnier, L., Virelizier, J.-L., Arenzana-Seisdedos, F. and Beretta, A. (1998) HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. *Lancet*, **351**, 14–18.
15. Telenti, A. and McLaren, P. (2010) Genomic approaches to the study of HIV-1 acquisition. *J. Infect. Dis.*, **202**(Suppl 3), S382–S386.
16. Wilkinson, D.A., Operskalski, E.A., Busch, M.P., Mosley, J.W. and Koup, R.A. (1998) A 32-bp deletion within the CCR5 locus protects against transmission of parenterally acquired human immunodeficiency virus but does not affect progression to AIDS-defining illness. *J. Infect. Dis.*, **178**, 1163–1166.
17. Kupfer, B., Kaiser, R., Brackmann, H.H., Effenberger, W., Rockstroh, J.K., Matz, B. and Schneweis, K.E. (1999) Protection against parenteral HIV-1 infection by homozygous deletion in the C-C chemokine receptor 5 gene. *AIDS*, **13**, 1025–1028.
18. Barretina, J., Blanco, J., Gutierrez, A., Puig, L., Altisent, C., Espanol, T., Caragol, I., Clotet, B. and Este, J.A. (2000) Evaluation of the putative role of C-C chemokines as protective factors of HIV-1 infection in seronegative hemophiliacs exposed to contaminated hemoderivatives. *Transfusion*, **40**, 461–467.
19. Salkowitz, J.R., Purvis, S.F., Meyerson, H., Zimmerman, P., O'Brien, T.R., Aledort, L., Eyster, M.E., Hilgartner, M., Kessler, C., Konkle, B.A. *et al.* (2001) Characterization of high-risk HIV-1 seronegative hemophiliacs. *Clin. Immunol.*, **98**, 200–211.
20. Zhang, M., Goedert, J.J. and O'Brien, T.R. (2003) High frequency of CCR5-delta32 homozygosity in HCV-infected, HIV-1-uninfected hemophiliacs results from resistance to HIV-1. *Gastroenterology*, **124**, 867–868.
21. Novembre, J., Galvani, A.P. and Slatkin, M. (2005) The geographic spread of the CCR5 Δ32 HIV-resistance allele. *PLoS Biol.*, **3**, e339.
22. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K. *et al.* (2011) Mapping copy number variation by population-scale genome sequencing. *Nature*, **470**, 59–65.
23. Fellay, J., Ge, D., Shianna, K.V., Colombo, S., Ledergerber, B., Cirulli, E.T., Urban, T.J., Zhang, K., Gumbs, C.E., Smith, J.P. *et al.* (2009) Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.*, **5**, e1000791.
24. Lederman, M.M., Alter, G., Daskalakis, D.C., Rodriguez, B., Sieg, S.F., Hardy, G., Cho, M., Anthony, D., Harding, C., Weinberg, A. *et al.* (2010) Determinants of protection among HIV-exposed seronegative persons: an overview. *J. Infect. Dis.*, **202**(Suppl 3), S333–S338.
25. Petrovski, S., Fellay, J., Shianna, K.V., Carpenetti, N., Kumwenda, J., Kamanga, G., Kamwendo, D.D., Letvin, N.L., McMichael, A.J., Haynes, B.F. *et al.* (2011) Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS*, **25**, 513–518.
26. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
27. Fellay, J., Shianna, K.V., Telenti, A. and Goldstein, D.B. (2010) Host genetics and HIV-1: the final phase? *PLoS Pathogens*, **6**, e1001033.
28. McLaren, P.J., Zagury, J.-F. and Fellay, J. and The International HIV Acquisition Consortium (2012), Abstract 295, *CROI 2012*, Seattle.
29. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
30. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
31. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. and Bucan, M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
32. The International HIV Controllers Study (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*, **330**, 1551–1557.
33. de Bakker, P.I.W., Ferreira, M.A.R., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.*, **17**, R122–R128.
34. Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J., Altshuler, D., Consortium, D. and investigators, M. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with Type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**, e1001058.

35. Bushman, F.D., Malani, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T.L., Zhou, H., Hazuda, D.J., Espeseth, A.S., Konig, R. *et al.* (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathogens*, **5**, e1000437.
36. Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K. *et al.* (2012) Global landscape of HIV–human protein complexes. *Nature*, **481**, 365–370.
37. Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P. and Rice, C.M. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, **472**, 481–485.
38. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
39. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.