NIH Public Access
Author Manuscript
J Infect Dis. Author manuscript; available in PMC 2011 May 1.

Published in final edited form as:
J Infect Dis. 2010 May 1; 201(9): 1371–1380. doi:10.1086/651606.

Large-Scale candidate gene analysis of spontaneous hepatitis C virus clearance

Timothy L Mosbruger[1], **Priya Duggal**[1], **James J. Goedert**[2], **Gregory D Kirk**[1], **W. Keith Hoots**[3], **Leslie H Tobler**[4], **Michael Busch**[4], **Marion G. Peters**[5], **Hugo R Rosen**[6], **David L Thomas**[1], and **Chloe L Thio**[1]

Timothy L Mosbruger: tmosbru1@jhmi.edu; Priya Duggal: pduggal@jhsph.edu; James J. Goedert: goedertj@mail.nih.gov; Gregory D Kirk: gkirk@jhsph.edu; W. Keith Hoots: hootswk@nhlbi.nih.gov; Leslie H Tobler: ltobler@bloodsystems.org; Michael Busch: mbusch@bloodsystems.org; Marion G. Peters: mpeters@itsa.uscf.edu; Hugo R Rosen: hugo.rosen@uchsc.edu; David L Thomas: dthomas@jhmi.edu; Chloe L Thio: cthio@jhmi.edu

[1]Johns Hopkins University

[2]Infections & Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute

[3]Division of Blood Diseases and Resources, National Heart Lung and Blood Institute

[4]Blood Systems Research Institute

[5]University of California, San Francisco

[6]University of Colorado

Abstract

Human genetic variation is a determinant of recovery from an acute hepatitis C virus (HCV) infection, but, to date, single nucleotide polymorphisms (SNPs) in a limited number of genes have been studied with respect to HCV clearance. We determined whether SNPs in 112 selected immune-response genes are important for HCV clearance by genotyping 1536 SNPs in a cohort of 343 persons with natural HCV clearance and 547 persons with HCV persistence. PLINK and Haploview software packages were used to perform association, permutation, and haplotype analyses stratified by African-American (AA) and European-American (EA) race. Of the 1536 SNPs tested, 1426 were successfully genotyped (92.8%). In AAs, we identified 18 SNPs located in 11 gene regions that were associated with HCV outcome (empirical p-value < 0.01). In EAs, there were 20 SNPs located in eight gene regions associated with HCV outcome. Four of the gene regions studied (*TNFSF18, TANK*, *HAVCR1* and *IL18BP*) contained SNPs with empirical p-values < 0.01 in both of the race groups.

Conclusion—In this large-scale analysis of 1426 genotyped SNPs in 112 candidate genes, we identified four gene regions that are likely candidates for a role in HCV clearance or persistence in both AAs and EAs.

Keywords

TANK; TNFSF18; HAVCR1; IL18BP; genetics

Contact Info: Chloe L Thio, MD, 855 N. Wolfe St Room 533, Baltimore, MD 21205, Phone: (410)-614-6088, Fax: (410)-614-7564, cthio@jhmi.edu.

Conflict of interest statement:
None of the others have a commercial or other association that would pose a conflict of interest with this study.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Introduction

Hepatitis C virus (HCV) infection results in viral persistence in the majority of cases, but approximately 30% of individuals mount an immune response that successfully eliminates infection (1,2). The factors required for generating this effective immune response are largely unknown but such information could lead to the development of more effective therapeutics and vaccinations. Since host genetic variation is a determinant of the immune response, comparing host genetic differences in immune response genes between those with HCV clearance and persistence could provide clues to these factors. Several studies have demonstrated genetic differences in immune response genes in those who have mounted a successful HCV-specific immune response compared to those who did not (3–7). These studies have used a candidate gene approach in which polymorphisms in one or a few genes are studied for an association with outcome. Since HCV clearance is likely to be polygenic, i.e. involving many genes, studying a few genes at a time limits the number of discoveries. Alternative approaches are now available to screen a large number of genes rapidly using small quantities of DNA.

One approach is to perform a genome-wide scan where single nucleotide polymorphisms (SNPs) at regular intervals through the entire genome are examined. The advantages of this approach are that a priori knowledge about pathogenesis is not needed since candidate genes do not need to be specified and that selection bias in gene selection does not occur. The disadvantage is that it is expensive and that SNP coverage within each gene is more limited, so associations within important candidates may be missed. The costs of such an approach can be decreased by studying pools of people, but then the ability to construct haplotypes, which are groups of SNPs that are in the same chromosomal region and inherited together, is lost. The second approach is to test many candidate genes simultaneously incorporating large-scale genotyping technologies to examine SNPs. The hits from this large-scale candidate gene approach can focus the gene or gene region for further exploration.

In order to determine whether selected immune-response genes are important for HCV clearance, we used the candidate gene approach and large-scale genotyping methods to test 1536 variants in 112 candidate immune-response genes throughout the genome. These genes were tested in a consortial cohort with well-defined HCV clearance and persistence.

Methods

Study population

The subjects in the cohort were participants in one of the following studies: (i) AIDS Link to Intravenous Experience (ALIVE) study (300 subjects), which is an ongoing study of 2,921 injection drug users enrolled in Baltimore, MD, from February 1988 to March 1989, as previously described (8); (ii) Multicenter Hemophilia Cohort Study (MHCS) (320 subjects), which is a prospectively-followed cohort of patients with hemophilia, von Willebrand disease, or a related coagulation disorder from 16 comprehensive hemophilia treatment centers enrolled between 1982 and 1986, as previously described (9); (iii) a cohort of blood donors throughout the United States (85 subjects); (iv) an HCV clinic cohort in Portland, Oregon (11 subjects); (v) Women’s Interagency Health Study (WIHS) (60 subjects), which is a cohort of female injection drug users, described previously (10);and (vi) Hemophilia Growth and Development Study (HGDS) (115 subjects), which is a continuing study of 333 children and adolescents with hemophilia, von Willebrand disease, or other coagulation disorder enrolled between March 1989 and May 1990, as previously described (11).

Prior HCV infection was established by detection of HCV antibody (anti-HCV) by enzyme immunoassay (EIA) and recombinant immunoblot assay (RIBA). Individuals with HCV

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

recovery had anti-HCV (confirmed by RIBA) and undetectable HCV RNA in serum or plasma without any HCV therapy. Persistently infected individuals had anti-HCV and HCV RNA in serum or plasma prior to any HCV therapy. For each individual with HCV clearance we matched one or two individuals with HCV persistence from the same cohort based on ethnicity and gender with the exception of the HCV clinic cohort from Portland which were all HCV clearance subjects.

Informed consent for genetic testing was obtained from all participants and the study was approved by the institutional review boards at all participating institutions.

Serologic testing

All serum or plasma specimens were stored at −70° C until testing. HIV type 1 (HIV-1) antibody testing was done by EIA with reactive results confirmed as positive by Western blotting as previously reported (8,9,12,13). Anti-HCV testing was done by Ortho HCV 2.0 or 3.0 EIA (Ortho Diagnostic Systems, Raritan, N.J.). HCV RNA was assessed by a branched DNA (bDNA) assay (Quantiplex HCV RNA 2.0 assay; Chiron Corporation, Emeryville, CA), qualitative HCV COBAS AMPLICOR system (COBAS AMPLICOR HCV; Roche Diagnostics, Branchburg, N.J.) or by transcription-mediated amplification (Novartis, Emeryville and Gen Probe, San Diego, CA.) Those subjects with a sample below the limit of detection by the bDNA assay (potential subjects with HCV recovery) had another sample tested with the qualitative COBAS, and their antibody status was confirmed by RIBA (RIBA 3.0) (Novartis). All assays were performed according to the manufacturer’s specifications.

SNP Selection and Genotyping

The 112 candidate genes selected for this study were either known or hypothesized to be involved in the immune response to HCV (Table 1). A Scientific Advisory Board selected candidate genes based on their known or suspected role in the pathogenesis of HCV clearance (see acknowledgement). A gene region, which was defined as the candidate gene plus 20 kb of flanking sequence both 5’ and 3’ to encompass regulatory regions, was determined for each candidate gene. If the flanking sequence from one candidate gene contained another candidate gene, the two gene regions were combined into one large gene region. Since the flanking sequences were large, a gene region often contained genes not selected specifically for this study. We initially selected all Phase I HapMap (www.hapmap.org) SNPs in each of these gene regions. The goal of the the first phase of the International HapMap Consortium was to genotype a common SNP every 5kb (14). Since our study population was comprised of both European-American (EA) and African-American (AA) individuals, allele frequencies from HapMap for both the Yoruba in Ibadan, Nigeria (YRI) and the Utah residents with ancestry from northern and western Europe (CEU) datasets were considered during SNP selection. If a SNP was in complete linkage disequilibrium (D’=1) with another SNP being tested in both the CEU and YRI datasets, only one was included. Only SNPS with minor allele frequencies (MAF) 5% in both the CEU and YRI datasets were considered.

Variants in each selected SNP were determined using the Illumina platform at the Johns Hopkins Genetic Resources Core Facility (JHGRCF). As a quality control measure, 20 known duplicates provided by JHGRCF and 20 blind duplicate samples from our test set were included. The JHGRCF did not report any data from samples that failed genotyping or SNPs that had: 1) poorly defined clusters, 2) excessive replicate errors, 3) more than 50% missing data, or 4) all heterozygote genotype calls. For each SNP, the JHGRCF reported individual raw data as well as allele frequencies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Statistical Analysis

Genotype quality analysis and filtering was done using PLINK. The criteria to exclude SNPs from statistical analysis included 1) minor allele frequencies (MAF) < 5%, 2) > 10% missing genotypes, and 3) deviation from Hardy Weinberg Equilibrium (HWE) (P-values < 0.001).

Population stratification is the presence of multiple ancestry groups in a study population that can lead to type I or type II errors due to inherent allele frequency differences. We used the genomic control (GC) (15) procedure to correct for population stratification. This method is based on the idea that under population stratification, the distribution of observed chi-squared test statistics from association tests will be inflated by a genomic inflation factor (λ) compared to the true chi-squared distribution under the null hypothesis. λ is calculated by dividing the median of the observed chi-squared distribution by the median of the expected chi-squared distribution (0.456). A λ of 1.0 or less suggests there is no inflation due to stratification, so no correction is made. Values greater than 1.0 suggest stratification is present and GC is applied by dividing the observed test statistics by λ. In candidate gene studies, λ can be calculated using a set of unlinked SNPs that are not related to the disease of interest. However, since we did not choose SNPs specific for GC analysis, we used an alternative method to calculate λ. In this method, we started with the set of all successfully genotyped SNPs. We dropped SNPs if they were in linkage disequilibrium with another SNP in the set ($r^2 > 0.5$) or if they had an association test P-value < 0.05. These two groups of SNPs were not included because they would artificially inflate λ. This procedure was done for both the EA and AA data, resulting in a λ for each group. Individuals self-identified their ethnicity. This self-identification was accurate as determined by a multidimensional scaling (MDS) analysis using the software program PLINK version 1.05 (http://pngu.mgh.harvard.edu/~purcell/plink/) (16) (Figure 1). Thus, samples self-identified as EA or AA were analyzed separately. Samples that did not self-identify as one of these two race groups were not included in this analysis (n=91).

Testing for association and Controlling Type 1 Error

Statistical analysis was performed in the software package PLINK. The allele frequencies of HCV clearance and persistence subjects were compared and the odds of clearing HCV infection were calculated for each SNP. Linkage disequilibrium (LD) in a genetic region was calculated for SNPs with allelic association test P-values < 0.01. SNPs in LD (r2 > 0.5) were grouped together as one signal.

Haplotype blocks were defined for all gene regions containing at least one SNP with an allelic association test p-value < 0.05. Haploview (http://www.broad.mit.edu/haploview/haploview) (17) was used to define the haplotype blocks using the "Solid spine of LD" algorithm. This definition requires all SNPs within the haplotype block to be in strong LD (D'>0.8) with the first and last SNP in the block. Haplotypes were inferred in Haploview using an E-M algorithm.

To control the Type 1 error, we performed 25,000 permutations for every SNP. Permutation testing calculates an empirical p-value by determining how frequently the association identified would occur by chance. For each permutation, the case-control labels (phenotype) are shuffled and the maximum chi-squared test statistic observed is compared to the experimental test-statistics for each SNP. An empirical p-value is calculated that provides a pointwise estimate for the significance of each SNP. Empirical p-values were calculated for every allelic association and p-values <0.01 were considered statistically significant. In addition, for the gene regions with at least one SNP meeting the pointwise p-value threshold of 0.01 for statistical significance in both populations, we calculated a gene region specific empirical p-value that accounted for the intra-correlation of SNPs and LD structure within the candidate gene and potential Type 1 error (18). We performed these permutations using the max(T) permutation procedure in PLINK.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Power

Power calculations were determined using the statistical program Quanto (http://hydra.usc.edu/gxe/). Assuming a population–wide HCV clearance rate of 30%, this study had 80% power to detect an associated SNP with a frequency of 20% and a relative risk of >1.5 when the analysis was not stratified by race. For a race-stratified analysis, there was 80% power to detect an associated SNP with a frequency of 20% and a relative risk >1.7.

Results

Study subjects

There were 352 individuals included in the AA group of whom 133 (37.8%) had HCV clearance and 219 (62.2%) had HCV persistence. There were 441 individuals included in the EA group of whom 169 (38.3%) had HCV clearance and 272 (61.7%) had HCV persistence (Table 2). The 91 individuals in the 'Other' group were not used in this analysis because the racial background of this group was very diverse including Hispanic, Asian and Native American leaving limited power for evaluation of these ethnic groups.

Data Quality

Of the 1536 SNPs tested, 110 were excluded because they did not reach the standards of JHGRCF quality control, leaving 1426 SNPs that were successfully genotyped (92.8%). Of the 890 individuals, 6 (0.7%) individuals had repeat genotyping failures and were removed from the dataset by the JHGRCF for poor DNA quality. The final data set generated by JHGRCF had a missing data rate of 0.074%. The error rate for the known and blinded duplicates was 0%.

Prior to analysis, 98 (7%) SNPs from the AA group and 258 (18%) from the EA were removed because the minor allele frequency < 5%. An additional 5 SNPs in the AA group and 2 SNPs in the EA group were eliminated because the SNPs deviated from Hardy-Weinberg equilibrium

Population stratification

Population stratification was assessed using the genomic control method in the AA (n=818 SNPs) and EA (n=526 SNPs) separately. The λ values were 1.01 and 0.931 for AA and EA, respectively suggesting no population stratification in either group.

Association Tests

Allelic association in the AA population identified 18 SNPs in 11 genes regions with empirical p-values < 0.01 (Supplementary Table 1). The same test in the EA samples resulted in 20 SNPs in eight gene regions with empirical p-values < 0.01 (Supplementary Table 2). Since many of these SNPs were in linkage disequilibirum within the same gene, we reduced the number of unique signals to 10 SNPs in AAs and 8 SNPs in EAs (Table 3). Interestingly, there were 4 gene regions, *TNFSF18, TANK, HAVCR1* and *IL18BP*, which contained SNPs significantly associated with HCV outcome (empirical pvalue < 0.01) in both the AAs and EAs. However, the specific SNPs within these genes that were associated differed between the racial groups (Table 4).

Haplotypes were generated for AAs and EAs for each gene region with at least one SNP with *p*-value < 0.05. None of these haplotypes were more significantly associated with outcome than the individual SNPs tested (data not shown).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Discussion

In this large-scale study, over 1400 SNPs in 112 gene regions that are candidates for playing a role in HCV clearance were genotyped. We found 18 SNPs in African-Americans and 20 SNPs in European-Americans with an empirical $p < 0.01$. Interestingly, we identified 4 gene regions that had an empirical p-value < 0.01 in both of the major races studied: *TNFSF18, TANK, HAVCR1* and *IL18BP*. Two of the SNPs tested, *HAVCR1* (rs1553316) and *IL18BP* (rs5743673), are coding SNPs, however, neither is known to have functional consequences. The replication of four gene regions in two independent populations is encouraging and suggests that these gene regions should be considered leading candidates for a role in HCV clearance. Although the exact SNPs were not necessarily replicated in each population this may be due to differences in allele frequencies, LD structure, or true allelic heterogeneity.

The TNFSF18 (Tumor Necrosis Factor (ligand) Superfamily, member 18), also known as GITRL, gene region is found on chromosome 1. TNFSF18 is expressed on CD4+CD28+ Regulatory T-cells (TRegs). TRegs can suppress other immune responses, providing a negative feedback on the immune system and preventing autoimmune responses. The binding of TNFSF18 to its receptor results in a down-regulation of TReg regulatory function and thus can lead to an increase in immune response, which would be favorable for HCV clearance (19).

The TANK (TRAF family member-associated NFKB activator) region is located on chromosome 2. TANK has been found to be important in type 1 interferon production through its interaction with both the RIG-I and toll-like receptor dependent (TLR) pathways (20), both of which are important in the innate immune response to HCV. TANK also plays a role in inducing a cellular response to tumor necrosis factor-alpha (21), and it has been described as an adaptor protein that is required for IRF3 activation (22). Thus, if a SNP alters the function of TANK, then either the innate or adaptive immune response to HCV could be affected.

The HAVCR1 (Hepatitis A Virus Cellular Receptor 1), also known as TIM1, gene region is found on chromosome 5. It belongs to a family of cell surface glycoproteins and appears to act as a costimulatory molecule in vitro leading to enhancement of T cell proliferation as well as Th1 and Th2 cytokine production. Interestingly, polymorphisms in HAVCR1 including a six-amino-acid insertion at residue 157 (157insMTTTVP), are linked to asthma and autoimmune diseases suggesting that these variants may affect HAVCR1 function (23). Thus, it is also possible that such functional variants could alter the immune response to HCV

The IL18BP (Interleukin-18 Binding Protein) gene region is found on chromosome 11. IL18BP is a secreted protein that can bind to and neutralize IL18, which prevents IL18-induced IFN-gamma production (24). Polymorphisms in both IFN-gamma and IL18 have been implicated in HCV infection outcome (25,26), and IL18 is up regulated in persons with chronic HCV infection (27). It is possible variants in IL18BP could affect the activity or production of IL18 and IFN gamma altering HCV outcome.

In addition to these four gene regions, one of the top-scoring SNPs in the EA group, rs1804027, was also significant in a study (listed as IMS-JST013416) investigating natural clearance of HCV in a Japanese population (28). This SNP results in a non-synonymous mutation in nuclear body protein SP110. The function of SP110 has not been well described, but it has been shown that HCV core protein can bind an isoform of SP110, SP110b, which results in the activation of Retinoic Acid Receptor (RARα)-mediated transcription (29).

It is important to consider the limitations of this study when interpreting the results. First, the size of the study makes it difficult to detect weak associations in frequent polymorphisms and any associations in rare variants. Second, deletion or insertion polymorphisms that may alter function are unlikely to be discovered unless they are tightly linked to one of the tested SNPs.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Third, SNPs were selected for coverage of genes and not for specific function therefore this study was not designed to identify causal alleles, but genes that may influence HCV clearance. Fourth, this study included many of the leading candidate gene regions potentially associated with HCV clearance at the time it was designed. However, since it was not intended to be an exhaustive survey of all interesting gene regions, additional studies based on these data should also consider the most recent data in HCV pathogenesis and include other relevant gene regions. For example, recently we and others reported a polymorphism in Il28B associated with HCV clearance and treatment response (30,31). Lastly, epistatic interactions between variants in different genes (such as ligand-receptor pairs) were not considered, because this study did not have enough power for such a large number of comparisons. Such interactions can be important in HCV pathogenesis as has been demonstrated for HLA and KIR genes (3).

By providing data on over 1400 SNPs in 112 candidate gene regions for HCV clearance or persistence, this study is an important first step since it reveals SNPs in four gene regions that warrant further investigation as a possible genetic basis for the natural clearance of HCV in multiple populations. Furthermore, this study provides the stimulus for confirmatory studies of our top scoring SNPs in other large, independent cohorts in order to determine the causal gene regions involved in the outcome of an acute HCV infection. These gene regions then need to be further dissected in order to determine the specific polymorphisms involved in HCV clearance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the Scientific Advisory Board including: Steven O’Brien, Robert Lanford, Stanley Lemon, Hugo Rosen, Christopher Karp, Michael Gale and Mary Carrington.

Funding: NIH-R01-DA13324, NHLBI-R01HL076902, NIH, National Institute of Child Health and Human Development grant 1 R01-HD41224, NIH-R01-DA0334, NIH-R01-DA12568, NIAID (UO1-AI-35004, UO1-AI-31834, UO1-AI-34994, UO1-AI-34989, UO1-AI-34993, and UO1-AI-42590), NICHHD (UO1-HD-32632 National Cancer Institute contract N02-CP-91027 with RTI International. Also in part by the Intramural Research Program, National Institute of Health.

Reference List

1. Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA 2000 Jul 26;284(4):450–456. [PubMed: 10904508]
2. Barrett S, Goh J, Coughlan B, Ryan E, Stewart S, Cockram A, et al. The natural course of hepatitis C virus infection after 22 years in a unique homogenous cohort: spontaneous viral clearance and chronic HCV infection. Gut 2001 Sep;49(3):423–430. [PubMed: 11511566]
3. Khakoo SI, Thio CL, Martin MP, Brooks C, Gao X, Astemborski J, et al. HLA and NK Cell Inhibitory Receptor Genes in Resolving Hepatitis C Virus Infection. Science 2004 Aug;305(5685):872–874. [PubMed: 15297676]
4. Thio CL, Gao X, Goedert JJ, Vlahov D, Nelson KE, Hilgartner MW, et al. HLA-Cw*04 and hepatitis C virus persistence. J Virol 2002;76:4792–4797. [PubMed: 11967296]
5. Thio CL, Thomas DL, Goedert JJ, Vlahov D, Nelson KE, Hilgartner MW, et al. Racial differences in HLA class II associations with hepatitis C virus outcomes. J Infect Dis 2001 Jul 1;184(1):16–21. [PubMed: 11398104]
6. Knapp S, Yee LJ, Frodsham AJ, Hennig BJ, Hellier S, Zhang L, et al. Polymorphisms in interferon-induced genes and the outcome of hepatitis C virus infection: roles of MxA, OAS-1 and PKR. Genes Immun 2003 Sep;4(6):411–419. [PubMed: 12944978]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

7. Hennig BJ, Hellier S, Frodsham AJ, Zhang L, Klenerman P, Knapp S, et al. Association of low-density lipoprotein receptor polymorphisms and outcome of hepatitis C infection. Genes Immun 2002 Sep;3 (6):359–367. [PubMed: 12209363]
8. Vlahov D, Anthony JC, Muñoz A, Margolik J, Celentano DD, Solomon L, et al. The ALIVE Study: a longitudinal study of HIV-1 infection in intravenous drug users: description of methods. J Drug Issues 1991;21:759–776.
9. Goedert JJ, Kessler CM, Aledort LM, et al. A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. N Engl J Med 1989;321:1141–1148. [PubMed: 2477702]
10. Barkan SE, Melnick SL, Preston-Martin S, Weber K, Kalish LA, Miotti P, et al. The Women's Interagency HIV Study. WIHS Collaborative Study Group. Epidemiology 1998 Mar;9(2):117–125. [PubMed: 9504278]
11. Hilgartner MW, Donfield SM, Willoughby A, Contant CF Jr, Evatt BL, Gomperts ED, et al. Hemophilia growth and development study. Design, methods, and entry data. Am J Pediatr Hematol Oncol 1993 May;15(2):208–218. [PubMed: 8498644]
12. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 1987;126(2):310–318. [PubMed: 3300281]
13. Chmiel JS, Detels R, Kaslow RA, Van Raden M, Kingsley LA, Brookmeyer R. Factors associated with prevalent human immunodeficiency virus (HIV) infection in the Multicenter AIDS Cohort Study. Am J Epidemiol 1987 Oct;126(4):568–577. [PubMed: 3651095]
14. A haplotype map of the human genome. Nature 2005 Oct 27;437(7063):1299–1320. [PubMed: 16255080]
15. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999 Dec;55(4):997–1004. [PubMed: 11315092]
16. 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 Sep;81 (3):559–575. [PubMed: 17701901]
17. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005 Jan 15;21(2):263–265. [PubMed: 15297300]
18. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005 Feb;6(2):95–108. [PubMed: 15716906]
19. Chattopadhyay K, Ramagopal UA, Mukhopadhaya A, Malashkevich VN, Dilorenzo TP, Brenowitz M, et al. Assembly and structural properties of glucocorticoid-induced TNF receptor ligand: Implications for function. Proc Natl Acad Sci U S A 2007 Dec 4;104(49):19452–19457. [PubMed: 18040044]
20. Guo B, Cheng G. Modulation of the interferon antiviral response by the TBK1/IKKi adaptor protein TANK. J Biol Chem 2007 Apr 20;282(16):11817–11826. [PubMed: 17327220]
21. Bonif M, Meuwis MA, Close P, Benoit V, Heyninck K, Chapelle JP, et al. TNFalpha- and IKKbeta-mediated TANK/I-TRAF phosphorylation: implications for interaction with NEMO/IKKgamma and NF-kappaB activation. Biochem J 2006 Mar 15;394(Pt 3):593–603. [PubMed: 16336209]
22. Ryzhakov G, Randow F. SINTBAD, a novel component of innate antiviral immunity, shares a TBK1-binding domain with NAP1 and TANK. EMBO J 2007 Jul 11;26(13):3180–3190. [PubMed: 17568778]
23. McIntire JJ, Umetsu SE, Macaubas C, Hoyte EG, Cinnioglu C, Cavalli-Sforza LL, et al. Immunology: hepatitis A virus link to atopic disease. Nature 2003 Oct 9;425(6958):576. [PubMed: 14534576]
24. Novick D, Kim SH, Fantuzzi G, Reznikov LL, Dinarello CA, Rubinstein M. Interleukin-18 binding protein: a novel modulator of the Th1 cytokine response. Immunity 1999 Jan;10(1):127–136. [PubMed: 10023777]
25. Huang Y, Yang H, Borg BB, Su X, Rhodes SL, Yang K, et al. A functional SNP of interferon-gamma gene is important for interferon-alpha-induced and spontaneous recovery from hepatitis C virus infection. Proc Natl Acad Sci U S A 2007 Jan 16;104(3):985–990. [PubMed: 17215375]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

26. An P, Thio CL, Kirk GD, Donfield S, Goedert JJ, Winkler CA. Regulatory polymorphisms in the interleukin-18 promoter are associated with hepatitis C virus clearance. J Infect Dis 2008 Oct 15;198 (8):1159–1165. [PubMed: 18781864]
27. Schvoerer E, Navas MC, Thumann C, Fuchs A, Meyer N, Habersetzer F, et al. Production of interleukin-18 and interleukin-12 in patients suffering from chronic hepatitis C virus infection before antiviral therapy. J Med Virol 2003 Aug;70(4):588–593. [PubMed: 12794721]
28. Saito T, Ji G, Shinzawa H, Okumoto K, Hattori E, Adachi T, et al. Genetic variations in humans associated with differences in the course of hepatitis C. Biochem Biophys Res Commun 2004 Apr 30;317(2):335–341. [PubMed: 15063762]
29. Watashi K, Hijikata M, Tagawa A, Doi T, Marusawa H, Shimotohno K. Modulation of retinoid signaling by a cytoplasmic viral protein via sequestration of Sp110b, a potent transcriptional corepressor of retinoic acid receptor, from the nucleus. Mol Cell Biol 2003 Nov;23(21):7498–7509. [PubMed: 14559998]
30. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009 Oct 8;461(7265):798–801. [PubMed: 19759533]
31. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009 Sep 17;461(7262):399–401. [PubMed: 19684573]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

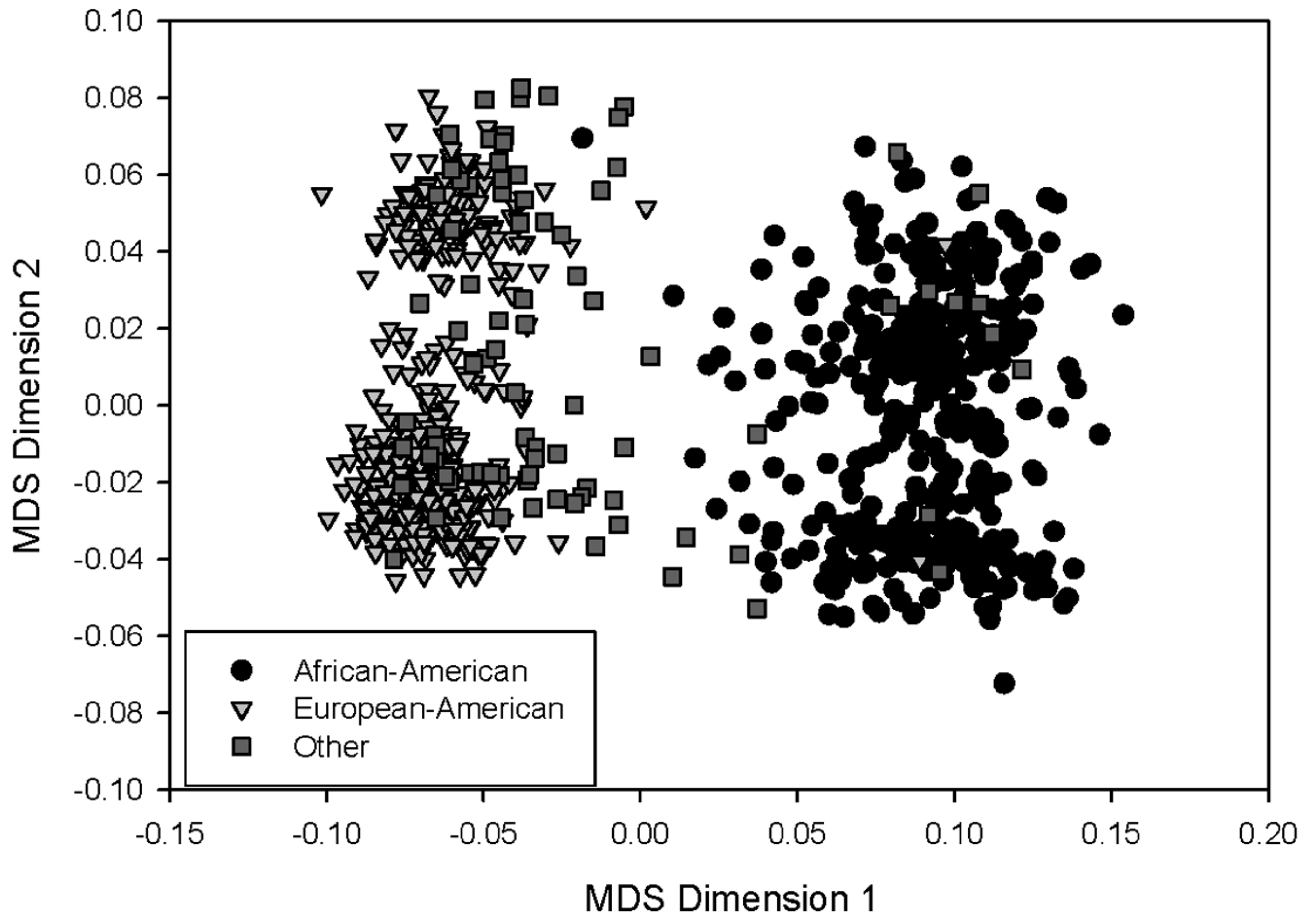

Figure 1. Multidimensional scaling plot of study cohort

Identity-by-state (IBS) distance measurements were reduced to two dimensions using a multidimensional scaling (MDS) procedure. The two resulting dimensions are used in a scatterplot. Individuals with similar genotypes cluster will together, showing the structure of the study population. Individuals self-identified as African-American largely clustered together as did individuals self-identified as European-American suggesting self-identification was accurate.

NIH-PA Author Manuscript

Table 1

Gene regions tested in this study

Candidate Gene Symbol	Candidate Gene Names	Chrom. Location
APOA2	Apolipoprotein A-II	1q21-q23
APOA5	Apolipoprotein A-V	11q23
APOBEC1	Apolipoprotien B mRNA editing enzyme	12p13.1
APOC3, APOA1	Apolipoprotein C-III, A-I	11q23.1-q23.2
APOC4, APOC2	Apolipoprotein C-IV, C-II	19q13.2
APOE, APOC1	Apolipoprotein E	19q13.2
BTLA	B and T lymphocyte associated	3q13.2
CCL5 (RANTES)	Chemokine ligand 5	17q11.2-q12
CCR5	Chemokine receptor 5	3p21
CD180	CD180	5q12
CD209 (DCSIGN), CD209L	CD209, CD209L	19p13
CD28	CD28	2q33
CD80	CD80	3q13.3-q21
CD81	CD81	11p15.5
CD86	CD86	3q21
CHUK	Conserved helix-loop-helix ubiquitous kinase	10q24-25
CTLA4	Cytotoxic T-lymphocyte associated protein 4	2q33
DDX58 (RIG-I)	Dead box polypeptide 58	9p12
DHX58 (LGP-2)	DEXH box polypeptide 58	17q21.2
EIF2AK2	Eukaryotic translation initiation factor 2-alpha kinase 2	2p22-p21
HAVCR1	Hepatitis A virus cellular receptor 1 (TIM-1)	5q33.2
HAVCR2	Hepatitis A virus cellular receptor 2 (TIM-3)	5q33.3
HLAA	Major histocompatibility complex, class I A	6p21.3
HLAB	Major histocompatibility complex, class I B	6p21.3
HLAC	Major histocompatibility complex, class I C	6p21.3
ICAM1, ICAM4, ICAM5	Intracellular adhesion molecule 1, 4 & 5	19p13.3-p13.2
ICAM3	Intracellular adhesion molecule 3	19p13.3-p13.2
ICOS	Inducible T-cell costimulator	2q33
IFITM1	Interferon induced transmembrane protein 1	11p15.5
IFNA1,IFNA10,IFNA7,IFNA17	Interferon-alpha 1, 10, 7, 17	9p22
IFNA14	Interferon-alpha 14	9p22
IFNA2	Interferon-alpha 2	9p22
IFNA21	Interferon-alpha 21	9p22
IFNA5	Interferon-alpha 5	9p22
IFNA6	Interferon-alpha 6	9p22
IFNA8	Interferon-alpha 8	9p22
IFNAR1	Interferon receptor 1	21q22.11

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Candidate Gene Symbol	Candidate Gene Names	Chrom. Location
IFNAR2	interferon receptor 2	21q22.1
IFNB1	Interferon beta 1, fibroblast	9p21
IFNG	Interferon gamma	12q14
IFNGR1	Interferon gamma receptor 1	6q23-24
IFNW1	Interferon omega 1	9p22
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	8p11.2
IL10RA	IL 10 receptor alpha	11q2.3
IL12A	IL-12A	3q12-q13.2
IL12B	IL-12B	5q31.1-q33.1
IL18BP	IL-18 binding precursor	11q22.2-q22.3
IL1RL1	IL-1 receptor-like 1	2q12
IL2	IL-2	4q26-q27
INDO	Indoleamine-pyrrole 2,3 dioxygenase	8p12-p11
INPP5D	Inositol polyphosphate-5-phosphatase	2q37.1
IRAK3	Interleukin-1 receptor associated kinase 3	12q14.3
IRF3	Interferon regulatory factor 3	19q13.3-q13.4
IRF7	Interferon regulatory factor 7	11p15.5
IRF9	Interferon regulatory factor 9	14q11.2
JAG1	Jagged 1	20p12.1-p11.23
LAG3	Lymphocyte activation gene 3	12p13.32
LILRA1-LILRA6	Leukocyte, Ig-like receptor subfamily A, member 1 – 6	19q13.4
LMB7	Proteasome subunit, beta type 8	6p21.3
LRP8	Low density lipoprotein receptor-related protein 8	1p34
MBL	Mannose binding lectin	10q11.2-q21
MX1, MX2	Myxovirus resistance 1 and 2	21q22.3
MYD88	Myeloid differentiation primary response gene 88	3p22
NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B cells 1	4q24
NOTCH4	Notch homolog 4	6p21.3
PIK3CA	Phosphoinositide-3-kinase catalytic, alpha polypeptide	3q26.3
PIK3CB	Phosphoinositide-3-kinase catalytic, beta polypeptide	3q22.3
PIK3CG	Phosphoinositide-3-kinase catalytic, gamma polypeptide	7q22.3
RNASEL	Ribonuclease L	1q25
SCARB1	Scavenger receptor class B, member 1	12q24.31
SIGIRR	Single immunoglobulin and toll-interleukin 1 receptor domain	11p15.5
SLC26A9	Solute carrier family 26, member 9	1q31-q32

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Candidate Gene Symbol	Candidate Gene Names	Chrom. Location
SLC4A10	Solute carrier family 4, sodium bicarbonate transporter-like, member 10	2q23-q24
SOCS1	Suppressor of cytokine signaling 1	16p13.3
SP110	SP110 nuclear body protein	2q37.1
STAT1	Signal transducer and activator of transcription 1	2q32.2
STAT2, APOF, IL23	Signal transducer and activator of transcription 2, Apolipoprotein F, IL-23	12q13.2
STAT3, STAT5A	Signal transducer and activator of transcription 3, 5A	17q21.31
TANK	TRAF family member-associated NFKB activator	2q24-q31
TBK1	Tank binding kinase 1	12q14.1
TICAM1 (TRIF)	Toll-like receptor adaptor molecule 1	19p13.3
TLR3	Toll-like receptor 3	4q35
TLR9	Toll-like receptor 9	3p21.3
TNFA	TNF alpha	6p21.3
TNFRSF1A	TNF receptor superfamily, member 1	12p13.2
TNFSFR18 (GITL), TNFRSF4	TNF receptor superfamily, member 18, 4	1p36.3
TNFSF18 (GITRL)	TNF ligand superfamily, member 18	1q23
TNFSF4	TNF ligand superfamily, member 4	1q25
TRAF5	TNF receptor-associated factor 5	1q32
TRAF6	TNF receptor associated factor 6	11p12
TRAM1	Translocation associated membrane protein 1	8q13.3
TSG101	Tumor susceptibility gene 101	11p15
VDR	Vitamin D receptor	12q13.11

Headings indicate the following information: **Candidate Gene Symbol,** official symbol of the tested candidate gene. If more than one candidate gene was tested in a region, both official symbols are listed. Symbols in parenthesis indicate a common alternate name; **Candidate Gene Name,** official full name of the tested candidate gene; **Chromosome Region,** location of the candidate gene region.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Demographics of study subjects

	African-American (N=352)		European-American (N=441)	
	Clearance subjects (N=133)	Persistence Subjects (N=219)	Clearance subjects (N=169)	Persistence Subjects (N=272)
Mean age (years)	34	33	30	27
Male (%)	52	64	79	91
Risk factor				
IDU (%)	89	88	11	5
Transfusion (%)	11	12	61	85
Unknown (%)	0	0	27	10
HIV-infected (%)	52	40	23	29

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

African-American and European-American SNPs with empirical p-values < 0.01

SNP	Odds Ratio	Empirical p-value	# LD	Gene Region Tested	Genes Containing SNPs
AFRICAN-AMERICAN					
SNP	**Odds Ratio**	**Empirical p-value**	**# LD**	**Gene Region Tested**	**Genes Containing SNPs**
rs6880589	0.41	0.0007	0	*HAVCR1*	
rs2295977	057	0.0009	0	*IRF9*	*IPO4*
rs723858	2.74	0.002	0	*TNFSF18*	
rs1168773	0.58	0.003	0	*IRAK3*	
rs7586242	0.59	0.003	4	*TANK*	
rs2715268	2.25	0.004	1	*CD86*	
rs7011101	0.55	0.005	1	*TRAM1*	
rs3750912	2.44	0.005	0	*IL18BP*	*NUMA1*
rs1424860	1.92	0.005	1	*IFNB1*	
rs7851560	0.62	0.009	1	*DDX58*	*TOPORS*
EUROPEAN-AMERICAN					
SNP	Odds Ratio	Empirical p-value	# LD	Gene Region Tested	Genes Containing SNPs
rs919178	1.72	0.0002	4	*SP110*	
rs1267059	0.51	0.0008	4	*TANK*	
rs2298455	1.85	0.003	0	*IL18BP*	
rs12104272	0.64	0.003	1	*IRF3*	*SCAF1,IRF3,BCL2L12*
rs2236876	1.58	0.004	2	*TNFSF18*	
rs4571051	1.53	0.005	0	*INPP5D*	
rs7303577	0.48	0.006	1	*TBK1*	
rs953569	0.68	0.007	0	*HAVCR1*	

Allelic association analysis with 25,000 permutations was performed on AA and EA data separately. Headings indicate the following information: **SNP,** top scoring SNP in each LD cluster; **Odds-Ratio (OR),** odds ratio of the minor allele of the top scoring SNP being associated with clearance. OR>1, minor allele is associated with viral clearance. OR<1, minor allele is associated with viral persistence; **Empirical p-value,** empirical p-value of the top scoring snp; **#LD,** number of SNPs with empirical p-value < 0.01 and in linkage disequilibrium with the top scoring SN;. **Gene Region Tested**, region tested including the candidate gene and 20kb of flanking sequence 3’ and 5’; **Genes Containing SNPs,** The gene containing the significant SNPs, which may or may not be the candidate gene tested.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Empirical p-values of all SNPs tested in the *TNFSF18, TANK, HAVCR1* and *IL18B*P Gene Regions

Gene Region	CHR	Position BP	SNP	African-American empirical p-value	European-American empirical p-value
TNFSF18	1	169736971	rs12067458	0.43	NA
TNFSF18	1	169744461	rs6425201	0.33	NA
TNFSF18	1	169745424	rs975074	0.07	0.23
TNFSF18	1	169749143	rs2236876	0.11	***0.004 (0.04)***
TNFSF18	1	169749920	rs723858	***0.002 (0.03)***	0.02
TNFSF18	1	169753006	rs7537126	0.40	NA
TNFSF18	1	169758508	rs7546619	0.43	***0.007***
TNFSF18	1	169764274	rs957544	0.02	***0.009***
TNFSF18	1	169768905	rs2223517	0.17	0.30
TANK	2	161801194	rs13430514	0.02	0.86
TANK	2	161803154	rs2884364	0.66	0.77
TANK	2	161805389	rs1267082	0.22	0.079
TANK	2	161810654	rs4664406	0.02	0.86
TANK	2	161813607	rs7586242	***0.003***	NA
TANK	2	161813990	rs1267070	0.28	***0.005(0.04)***
TANK	2	161820633	rs3820998	***0.007***	NA
TANK	2	161828644	rs13402291	0.05	NA
TANK	2	161837642	rs1267068	0.05	0.07
TANK	2	161841338	rs1559526	0.51	0.26
TANK	2	161844150	rs1267059	0.12	***0.0008 (0.004)***
TANK	2	161848350	rs3769980	***0.005***	NA
TANK	2	161854620	rs7568498	NA	0.79
TANK	2	161858169	rs889916	0.39	***0.002(0.03)***
TANK	2	161864079	rs1267080	0.59	***0.004(0.03)***
TANK	2	161873088	rs1267072	0.05	0.07
TANK	2	161877080	rs1267074	0.04	0.06
TANK	2	161882525	rs3754972	***0.005***	NA

Gene Region	CHR	Position BP	SNP	African-American empirical p-value	European-American empirical p-value
TANK	2	161889871	rs10186736	0.02	NA
TANK	2	161898245	rs1267034	0.02	0.04
TANK	2	161911654	rs10189790	0.52	NA
TANK	2	161914481	rs12471074	***0.007***	NA
TANK	2	161923662	rs1267037	0.19	***0.002 (0.01)***
TANK	2	161926585	rs3820996	0.05	NA
TANK	2	161930291	rs1146030	0.31	0.31
TANK	2	161932802	rs3731769	0.53	0.37
TANK	2	161936739	rs1267053	0.19	NA
TANK	2	161939567	rs1882364	0.59	0.36
HAVCR1	5	156369257	rs1995377	0.08	0.79
HAVCR1	5	156382326	rs7707445	0.37	0.20
HAVCR1	5	156390564	rs6555810	0.91	NA
HAVCR1	5	156392267	rs7732478	0.89	NA
HAVCR1	5	156392673	rs2277025	0.80	0.06
HAVCR1	5	156393018	rs6880589	***0.0007 (0.01)***	NA
HAVCR1	5	156396820	rs2036402	0.28	0.28
HAVCR1	5	156402361	rs2270926	0.53	0.14
HAVCR1	5	156409978	rs953569	0.14	***0.007***
HAVCR1	5	156411901	rs1553318	0.15	0.13
HAVCR1	5	156412087	rs1553316	0.10	0.86
HAVCR1	5	156421481	rs6873777	0.44	0.04
HAVCR1	5	156423731	rs4704843	0.47	0.04
HAVCR1	5	156432559	rs2116787	0.65	0.67
HAVCR1	5	156436363	rs7725390	0.06	0.25
HAVCR1	5	156437811	rs11134551	0.54	0.10
IL18BP	11	71367189	rs6592455	0.90	1
IL18BP	11	71372293	rs4945392	0.74	NA
IL18BP	11	71379171	rs2276384	1.00	NA

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Gene Region	CHR	Position BP	SNP	African-American empirical p-value	European-American empirical p-value
IL18BP	11	71381789	rs7124429	0.70	NA
IL18BP	11	71388126	rs2298455	0.71	***0.003 (0.004)***
IL18BP	11	71390101	rs5743673	0.65	NA
IL18BP	11	71391681	rs5743679	0.74	NA
IL18BP	11	71403860	rs3750912	***0.005 (0.04)***	NA
IL18BP	11	71405070	rs1057992	0.76	NA
IL18BP	11	71407954	rs949326	0.91	NA
IL18BP	11	71408363	rs2298456	0.75	1
IL18BP	11	71411859	rs2298457	0.75	1

Allelic association analysis with 25,000 permutations was performed on AA and EA data separately. Headings indicate the following information: **Gene Region,** gene region containing the tested SNP; **CHR,** chromosome where the tested SNP is located; **Position,** chromosomal position, in base pairs, of the tested SNP; **SNP,** rs number of the tested SNP. **AA and EU Empirical P-value,** empirical p-value of the tested SNPs in AA and EA populations. Italicized SNPs have empirical p-values < 0.01. Parenthesis denotes gene specific permuted p-values. NA indicates that the MAF was < 0.05 and a p-value was not calculated for this population

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript