UCSF
UC San Francisco Previously Published Works

Title

Multi-Ancestry Genome-Wide Association Study of Spontaneous Clearance of Hepatitis C Virus

Permalink

Journal

ISSN

Authors

Vergara, Candelaria
Thio, Chloe L
Johnson, Eric
et al.

Publication Date

DOI

Peer reviewed

Multi-Ancestry Genome-Wide Association Study of Spontaneous Clearance of Hepatitis C Virus

Candelaria Vergara[1], **Chloe L. Thio**[1], **Eric Johnson**[2], **Alex H. Kral**[2], **Thomas R. O'Brien**[3], **James J. Goedert**[4], **Alessandra Mangia**[4], **Valeria Piazzolla**[4], **Shruti H. Mehta**[5], **Gregory D. Kirk**[1,5], **Arthur Y. Kim**[6], **Georg M. Lauer**[6], **Raymond T. Chung**[6], **Andrea L. Cox**[1], **Marion G. Peters**[7], **Salim I. Khakoo**[8], **Laurent Alric**[9], **Matthew E. Cramp**[10], **Sharyne M. Donfield**[11], **Brian R. Edlin**[12], **Michael P. Busch**[13], **Graeme Alexander**[14], **Hugo R. Rosen**[15], **Edward L. Murphy**[13], **Rachel Latanich**[1], **Genevieve L. Wojcik**[16], **Margaret A. Taub**[5], **Ana Valencia**[1,17], **David L. Thomas**[1], **Priya Duggal**[5]

[1]Johns Hopkins University, School of Medicine, Baltimore, MD. USA

[2]RTI International, Research Triangle Park, NC, Atlanta, GA, San Francisco, CA. USA

[3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD. USA.

[4]Liver Unit IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy.

[5]Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD. USA.

[6]Liver Center and Gastrointestinal Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA. USA

[7]Division of Gastroenterology, Department of Medicine, School of Medicine, University of California, San Francisco.CA. USA.

[8]University of Southampton, Southampton General Hospital, Southampton, UK.

[9]Department of Internal Medicine and Digestive Diseases, CHU Purpan, UMR 152, IRD Toulouse 3 University, France.

[10]South West Liver Unit, Plymouth, PL6 8DH, UK

[11]Rho, Inc., Chapel Hill, NC.USA

[12]SUNY Downstate College of Medicine, Brooklyn, NY. USA

[13]University of California and Vitalant Research Institute, San Francisco, CA, USA.

Correspondence: Priya Duggal, PhD, MPH, 615 N. Wolfe Street, Room W6511, Baltimore, Maryland 21205, Phone: 410-955-1213,410-955-0863, pduggal@jhu.edu.

[14]UCL Institute for Liver and Digestive Health, The Royal Free Hospital, Pond St, Hampstead, London NW3 2QG.

[15]University of Colorado, Aurora, Colorado, 90045, United States

[16]Department of Genetics, Stanford University School of Medicine, Stanford, CA. USA

[17]Universidad Pontificia Bolivariana. Medellin, Colombia.

Abstract

Background & Aims: Spontaneous clearance of hepatitis C virus (HCV) occurs in approximately 30% of infected persons and less often in populations of African ancestry. Variants in major histocompatibility complex (MHC) and in interferon lambda genes are associated with spontaneous HCV clearance but there have been few studies of these variants in persons of African ancestry. We performed a dense multi-ancestry genome-wide association study of spontaneous clearance of HCV, focusing on individuals of African ancestry.

Methods: We performed genotype analyses of 4423 people from 3 ancestry groups: 2201 persons of African ancestry (445 with HCV clearance and 1756 with HCV persistence), 1739 persons of European ancestry (701 with HCV clearance and 1036 with HCV persistence), and 486 multi-ancestry Hispanic persons (173 with HCV clearance and 313 with HCV persistence). Samples were genotyped using Illumina arrays and statistically imputed to 1000 Genomes Project. For each ancestry group, the association of single nucleotide polymorphisms with HCV clearance was tested by log-additive analysis and then a meta-analysis was performed.

Results: In the meta-analysis, significant associations with HCV clearance were confirmed at the interferon lambda gene locus IFNL4–IFNL3 (19q13.2; $P=5.99\times10-50$) and the MHC locus 6p21.32 ($P=1.15\times10-21$). We also associated HCV clearance with polymorphisms in the G-protein-coupled receptor 158 gene (GPR158) at 10p12.1 ($P=1.80\times10-07$). These 3 loci had independent, additive effects of HCV clearance, and account for 6.8% and 5.9% of the variance of HCV clearance in persons of European and African ancestry, respectively. Persons of African or European ancestry carrying all 6 variants were 24-fold and 11-fold, respectively, more likely to clear HCV infection compared to individuals carrying none or 1 of the clearance-associated variants.

Conclusions: In a meta-analysis of data from 3 studies, we found variants in MHC genes, IFNL4–IFNL3, and GPR158 to increase odds of HCV clearance in patients of European and African ancestry. These findings could increase our understanding of immune response to and clearance of HCV infection.

Keywords

Introduction

Hepatitis C virus (HCV) is a major cause of chronic liver disease, hepatocellular carcinoma and end stage liver disease (1,2). Approximately 50–70% HCV-infected persons develop chronic hepatitis C, whereas the remainder spontaneously eliminate the virus (3–8). The

prevalence of spontaneous clearance is lower in African ancestry individuals compared to populations of European ancestry (9). Similarly, people form African ancestry have a higher prevalence of chronic HCV infection (1,2) and a lower response to anti-HCV therapy with interferon (10) and with some direct-acting antivirals (11).

The identification of genetic factors participating in the pathogenesis of a disease is pivotal to understanding disease mechanisms and identifying biological targets for treatments and vaccines (12). Development of vaccines that protect against persistent HCV infection remains a public health priority since even extensive use of highly effective direct-acting antivirals is unlikely to achieve HCV elimination without vaccines to limit transmission (13–15).

Genome-wide association studies (GWAS) have revealed that spontaneous HCV clearance and persistence is associated with genetic variants in the MHC region and *IFNL4* (16–18). However, these two genes do not explain all of the variation of spontaneous HCV clearance (16), suggesting that other genetic variants may contribute to the HCV immune response. Additionally, these studies were chiefly based on European or Asian ancestry individuals, while the persistent phenotype is more common in persons of African ancestry (2–4).

To identify additional genetic factors for spontaneous HCV clearance, particularly in persons of African ancestry, we performed a dense multi-ancestry GWAS comparing individuals with spontaneous clearance of HCV infection to those with persistent infection.

Methods

Study populations

There are three major sources of data for this analysis as described in Table 1 and the Supplementary Materials. One is derived from a study initially conducted to understand the host genetic contributions to Human Immunodeficiency Virus (HIV) infection (19,20). By testing stored plasma for HCV antibodies and RNA, we phenotyped 1661 individuals who had human genotyping done as part of the Urban Health Study (UHS): a serial, cross-sectional, sero-epidemiological study of people who inject drugs in the San Francisco Bay Area (19,20). Of these 1536 individuals were included in this study with complete phenotype and genotype data. A second data source was our prior HCV multi-cohort GWAS, described before (16) and in detail in the Supplementary Materials. The genotypic and phenotypic data of the HCV multi-cohort study and UHS were consolidated for this analysis in a unique dataset referred here as the "extended HCV multi-cohort study". Individuals were assigned to one of three genetically determined ancestry groups consisting of 1736 European Ancestry, 1869 African Ancestry Group 1 and 486 multi-Ancestry Hispanic and analyzed separately for association with HCV clearance.

A third source of data came from 332 individuals of African ancestry from either the ALIVE or UHS cohort (21) recruited and diagnosed under the same criteria previously described but not included in the prior HCV multi-cohort GWAS (19,20,22). All subjects with spontaneously cleared infection were included and frequency matched to controls with persistent infection on the basis of self-reported ethnicity. Because a different genotyping

array was performed on these subjects, the group is separately named (African Ancestry Group 2) to clarify when our analytical procedures differed.

All study procedures were approved by an IRB of Johns Hopkins University and those related to each cohort were approved by the local IRBs including RTI International, the National Cancer Institute and the Committee on Human Subjects Research at the University of California, San Francisco.

Genome-wide Genotyping and Imputation

Participants in the extended HCV multi-cohort study were genotyped using the Illumina Omni1-Quad BeadChip array (Illumina, San Diego, CA) on genomic DNA; independently for the UHS cohort and the HCV multi-cohort study. For this analysis, we merged the genotyped datasets after standard quality control and used 612,887 markers in the intersection for imputation after removing SNPs that were only genotyped on one cohort and applying quality control protocols to the intersected data (Supplementary Materials). Individuals in the African Ancestry Group 2 study were genotyped using the Infinium Multi-Ethnic AMR/AFR-8v1.0 array (Illumina, San Diego, CA). There were data for 1,020,897 SNPs in 332 unrelated individuals used for imputation after applying quality control protocols as described in Supplementary Materials. Imputation was performed for chromosomes 1 to 22 using the Minimac3 software through the publicly available Michigan Imputation Server (23) (Supplementary Materials). Coordinates of the genotyped and imputed data are based on the Human Feb. 2009 (GRCh37/hg19) assembly.

Principal component analysis and individual ancestry determination

Genetic ancestry and population structure for the extended HCV multi-cohort study and the African ancestry Group 2 study was determined by principal component analysis using the smartpca program in EIGENSOFT (24) as well as the evaluation of estimates of individual ancestry using a variational Bayesian framework implemented in fast-STRUCTURE (25) as described in detail in Supplementary Materials and Supplementary Figures S1–S4.

Genome wide association analysis and fixed effect meta-analyses.

For the extended HCV multi-cohort study, mach2dat was used to test the association of allelic imputed dosage with HCV clearance under a log-additive model (26). HIV infection status and 20 principal components were included as covariates, separately for each genetically determined ancestry group to correct for population structure (24). For the African ancestry Group 2 we ran a stratification-score matching analysis with the aim of improving the correction for confounding by population stratification (27). For this group, we tightly matched cases and controls into matched sets having similar genetic ancestry based on their values for 20 PCs and tested for association between HCV clearance and the markers using a Cochran-Mantel-Haenszel test implemented in the package PLINK (28,29) (version 1.9, URL: www.coggenomics.org/plink/1.9/). After quality control (Supplementary Materials), the final dataset used for the analysis (Table 1) comprised 4423 individuals (3104 clearance and 1319 persistence): 1869 African ancestry Group 1 individuals (340 clearance/ 1529 persistence), 1736 European ancestry (701 clearance/1035 persistence), 486 multi-ancestry Hispanic (173 clearance/313 persistence) and 332 African ancestry Group 2 (105

clearance/227 persistence). Results of the ancestry groups were then combined in a fixed-effects sample size-weighted meta-analysis implemented in METAL (30).

We calculated the population-specific effective number of independent tests for each genetically determined ancestral group included in the current study and for the fixed effects meta-analysis. We used the method recommended by Sobota et al. (31), which estimates the effective number of independent tests in a genetic dataset after accounting for linkage disequilibrium (LD) between SNPs using the LD pruning function in the PLINK 1.9 software package (28,29) (Supplementary Materials). This approach has been successfully implemented in recent genome-wide, phenome-wide, linkage and whole genome sequencing analyses of complex traits in populations with similar ancestral background to those analyzed in the current study (32–35). The GWAS significance threshold was set at 2.05×10^{-07} for the fixed effects meta-analysis, 4.44×10^{-07} for European ancestry, 2.92×10^{-07} for African ancestry, 3.95×10^{-07} for Multi-Ancestry Hispanics. These values are similar to those estimated based on the 1000 Genomes Project dataset (31). Suggestive significance level for the fixed effects meta-analysis was set to P values between 5×10^{-05} and 2.05×10^{-07}.

Trans-ethnic meta-analysis and calculation of credible sets of variants

To identify the credible regions around each of the the associated loci and to capture loci with effect heterogeneity across the analyzed populations, we performed a trans-ethnic meta-analysis using MANTRA (Meta-ANalysis of Transethnic Association studies) software (36). MANTRA accounts for the shared similarity in closely related populations using Bayesian partition model assuming the same underlying allelic effect. It models the effect heterogeneity among distant populations by clustering according to the shared ancestry and allelic effects. For this meta-analysis results, a Log_{10} Bayes factor ($Log_{10}BF$)   5 were considered statistically significant (21).

Within each significantly associated locus, we also the estimated posterior probabilities for each variant in the region (37,38). Posterior probabilities are the ratio of evidence for each variant versus all others, which makes it a useful comparator for fine-mapping. Assuming there is exactly one putative causal variant in a region, and all variants are included in the analysis, then for any 'credible set (CSs)' of variants, we can state that the causal variant will be included in the set with confidence equal to the sum of the posteriors of the SNPs in the set. CSs were created by sorting the SNPs in each region in descending order based on their Bayes factors starting with the index SNP since this SNP has the region's largest BF by definition. Going down the sorted list, the SNPs' posterior probabilities were summed until the cumulative value exceeded 95% of the total cumulative posterior probability for all SNPs in the region. Credible sets can be interpreted in a way similar to confidence intervals in a frequentist statistical framework. For example, assuming that a locus harbors a single causal variant that is reported in the meta-analysis, the probability that it will be contained in the 95% credible set is 0.95. Smaller credible sets, in terms of the number of SNPs they contain or the genomic interval they cover, thus correspond to fine mapping at higher resolution. In constructing credible sets, it is assumed that there is a single causal variant at each locus, so formal conditioning of the locus adjusting for genotypes at each top SNP in turn is needed

before construction of the credible set for each underlying causal variant. In view of that, and also to determine whether a single SNP explained the association of a genetic region, we did a conditional analysis for each locus. We included the top SNP in the model (for some regions, more than 1 SNP was included consecutively to explain the significance of a region) and then reevaluated the association plots. In combined models, we included several SNPs different loci in the analysis. The length of credible sets for each region is expressed in Kilobases (Kb).

Estimation of combined effect of GWAS associated variants

To estimate the effect of carrying any risk allele of the associated variants, we constructed a model where individuals of European ancestry and African ancestry were grouped based on the number of alleles they carried from 0 to 6 alleles for the three significantly associated loci. The odds ratio of spontaneous HCV clearance conferred by carrying 2–6 risk alleles was compared to those having 0–1 risk variants for all individuals. The analysis was performed using customized scripts in R version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria). We also calculated the positive and negative predictive value for the presence of the favorable genotype in each of the top associated variants and for the combination of the favorable genotypes.

Estimation of the percentage of variance explained by associated variants.

The percentage of phenotypic variance explained by each of the three top associated SNPs was estimated by constructing logistic regression models including sex, HIV infection status and each of the SNPs as the predictors for HCV clearance. To calculate the total variance explained by them, we included all three SNPs in a unique model. We calculated the change of deviance for the model with and without the SNPs and then divided by total deviance from the null model. The analysis were performed for European and African Ancestry Group 1 using R, version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria).

Results

We analyzed a total of 4423 individuals: 1319 individuals with HCV clearance and 3104 individuals with persistence in 4 independent study groups for association with genetic variants across the autosomes. Overall, 34% were women, and 28% were HIV infected (Table 1).

Significantly associated regions in meta-analysis

The fixed effects meta-analysis identified three chromosomal regions with marked differences in allele frequency between individuals with HCV clearance and persistence at the meta-analysis GWAS significance levels (2.05×10^{-07}) (Figure 1, Table 2 and Supplementary Figure S5). We replicated the associations for loci in chr19q13.2 (rs74597329, Allele T OR: 2.14 – 3.28, P value= 5.99×10^{-50}) and chr6p21.32 (rs2647006, Allele C OR: 1.71 –1.78, P value = $1.15 \times 10 \times ^{-21}$), which were previously associated with HCV clearance (16,17,39,40), and identified a novel locus on chr10p12.1 (rs1754257, Allele A OR : 1.06–1.55, P value =1.8×10^{-07}).

Chr19q13.2 region: We highlight 13 of 100 SNPs in a 71.9 Kb region that were significantly associated ($P < 2.03 \times 10^{-07}$) in the chr19q13.2 locus harboring the interferon lambda genes (Figure 1–2 and Table 2, Supplementary Table S1–S4). The most significant association was for the imputed SNP rs74597329 (imputation $r^2 = 0.92$), a G/T change located in exon 1 of *IFNL4* gene (chr19:39739155, Figure 2). This association was present and had the same direction of effect across the European ancestry group, African ancestry Group 1 and Multi-Ancestry Hispanics reaching GWAS significance levels in each population (Table 2, Supplementary Figure S6–S7). The T allele is associated with higher odds of HCV clearance in all four study groups (OR: 2.14 – 3.28 depending on the population). In conjunction with the -/T variant (rs11322783), this SNP forms the G/TT polymorphism (rs368234815) which has been previously associated with HCV clearance (41,42). The positive and negative predictive value (95% CI) for the presence of the TT genotype in HCV clearance is 0.39 (0.33, 0.45) and 0.85 (0.83, 0.87) in the African Ancestry Group 1 and 0.53 (0.49, 0.56) and 0.71 (0.68, 0.74) in the European Ancestry population, respectively. In conditional analysis, the lead SNP in this region, rs74597329, accounted for the signal at this locus (Supplementary Figure S8). However, in European Ancestry, rs4803221 located 26 bp upstream of rs74597329 but not in high LD with each other (LD $r^2 = 0.52$), had attenuated significance but remained present (Allele C OR: 1.98, P Value: 6.4×10^{-6}) after conditioning on rs74597329 (Supplementary Figure S9). The 95% credible set extended 1642 base pairs from exon1 to exon 5 of *IFNL4* (19:39737513–19:39739155) (Table 3, Figure 2).

Chr6p21.32 region: There were 1,210 significantly associated SNPs ($P < 2.03 \times 10^{-07}$) spanning a region of 305.2 kb across the class II MHC genes (*HLA-DQB1/HLA-DQA1* and *HLA-DQA2*). Seven of the SNPs are presented in Table 2 (additional SNPs in Supplementary Table S5–S7). The lead SNP, rs2647006, is located 48.5 Kb upstream of *HLA-DQA2* and 33.3kb from the *HLA-DQB1* gene. For each C allele of rs2647006, there is an increase of 1.71–1.78 in the odds of HCV clearance depending on the population (fixed effects meta-analysis P value = 1.1×10^{-21}) (Table 2). The positive and negative predictive value (95% CI) for the presence of the CC genotype on HCV clearance is 0.25 (0.21, 0.28) and 0.86 (0.84, 0.88) in the African Ancestry Group 1 and 0.48 (0.45, 0.52) and 0.65 (0.62, 0.68) in the European Ancestry group, respectively. After conditioning on this SNP, no other SNPs remained significantly associated at GWAS level ($P < 2.05 \times 10^{-7}$) (Supplementary Figure S10–S11). The credible sets narrows the region to 14.3kb intergenic region between *HLA-DQB1* and *HLA-DQA2* genes (Table 3).

Chr10p12.1 region: We identified 4 significantly associated SNPs (P value $< 1.91 \times 10^{-07}$) spanning a region of 15.4 kb harboring the *G Protein-coupled receptor 158* (*GPR158)* (Table 2, Figure 2). Each of the ancestry groups contributed to this association in the meta-analysis. Similar to what was observed in the other regions, in the conditional analyses, the lead SNP, rs1754257 explained the significant findings at this locus (Supplementary Figures S12–S13). The top variant: rs1754257 (chr10:25674614: A>G, P value 1.8×10^{-07}) is located in intron 2 of *GPR158*; the A allele (effect allele) has a frequency of 0.7–0.8 in the analyzed groups (Table 2) which is consistent with the frequency observed in European (0.73) and African (0.77) populations of the 1000 Genomes Project (43). The presence of each copy of

this allele increases the odds of clearance in each group 24–55 % (Table 2) with the largest effect in the African ancestry group. The positive and negative predictive value (95% CI) for the presence of the AA genotype in HCV clearance is 0.21 (0.19, 0.24) and 0.86 (0.83, 0.88) in the African Ancestry Group 1 and 0.43 (0.40, 0.46) and 0.63 (0.59, 0.66) in the European Ancestry group, respectively. The credible sets narrows the implicated region to 12.5Kb region in the *GPR158* (Table 3). The gene has high expression in the brain, but no significant expression quantitative loci (eQTLs) are described for this gene in the publicly available Genotype-Tissue Expression (GTEx) project dataset (44).

These three significant regions were independently associated with HCV clearance and the conditioning on one region did not affect the signal observed in the other regions (Supplementary Figure S14). The association in those significant regions was not modified HIV infection status or sex (Supplementary Table S9).

Additional Regions of Interest

In the meta-analysis, we identified several additional regions with suggestive associations including chr11q13.4 (P value = 6.09×10^{-07}). This region was significantly associated at ancestry specific GWAS level in the African ancestry Group 1 where we identified 13 associated variants spanning a 110.03 Kb region (chr11: 73108664–73218694) with the main signal (rs11235775, A/G, chr11:73198087, Allele G OR: 1.82, SE: 0.11; P Value= 1.07×10^{-07}) located in the *Family with Sequence Similarity 168, Member A* (*FAM168A*) gene and downstream from the *Receptor Expressed In Lymphoid Tissues* (*RELT*) gene (Figure 3 and Supplementary Table S8–S9). rs11235775 has a minor allele frequency (G allele) of 0.16 in African ancestry and 0.06 in multi-ancestry Hispanic but is monomorphic in European ancestry individuals where the G allele is not present and allele A is fixed (Supplementary Figure S15). The variants had consistent direction of the effect across both African ancestry and the Multi-Ancestry Hispanic group (Supplementary Table S10) and one variant (rs11235778; chr11:73205982) located 7.9 Kb upstream of rs11235775 was significantly associated in the trans-ethnic meta-analysis (Log_{10} BF=5.0).

Other loci in the chr5q23.1, chr6q14.1, chr6p22.1, chr10q22.2, chr11q13.4, chr19p12 and chr19q13.32 regions were suggestive of association in the African ancestry 1 group or in the meta-analysis (Supplementary Table S11). The meta-analysis results were consistent in both the fixed effects and trans-ethnic meta-analysis (Supplementary Figure S16, Supplementary Table S12).

Combined effect of GWAS associated variants

In an analysis evaluating the effect of the presence of favorable alleles for the top associated SNPs rs74597329, rs2647006 and rs1754257, there was a stepwise increase in the percentage of spontaneous clearance with the inclusion of each favorable allele, indicating a cooperative additive effect of the protective variants (Figure 4). The combination of 5 favorable alleles had an OR of 7.5 (95% CI: 2.6–21.2) and 5.6 (95% CI: 1.6–19.4) in the African Ancestry group and the European ancestry group, respectively, compared to those individuals carrying 0–1 favorable alleles. Moreover, six favorable alleles conferred by the three genotypes of the top SNPs in the homozygous state (rs74597329 TT, rs2647006 CC

and rs1754257 AA genotypes) had an odds ratio of 23.6 (95% CI: 7.7–71.7) in the African Ancestry and 10.8 (95% CI: 3.0–38.2) in the European Ancestry Group compared to those individuals carrying 0–1 favorable alleles. This increasing additive effect was not modified by either sex or HIV infection status (Supplementary Table S13). The positive predictive value (95% CI) for the combination of the favorable genotypes in the *IFNL3-IFNL4* and MHC loci is 0.52 (0.42, 0.62) in African Ancestry Group 1. When incorporating the AA genotype of *GPR158* rs1754257, these values increase slightly to 0.55 (0.43, 0.67). In the European Ancestry population, the combination of the 2 favorable genotypes in the IFNL3-INFL4 and MHC loci have a positive predictive value of 0.62 (95% CI: 0.57, 0.68) and for the three favorable genotypes (including the GPR158 locus) they are 0.67 (95% CI: 0.60, 0.73).

Percentage of variance explained by associated variants

The total variance explained by these three variants is 5.92% for African Ancestry and 6.75% for European Ancestry. An estimate of 3.04% of the variance in spontaneous HCV clearance is explained by the top SNP at chromosome 19 (rs74597329) for persons of African ancestry and 4.23% for those of European ancestry. The top SNP at chromosome 6 (rs4273729) explained 2.02% of variance for persons of African ancestry and 1.84 % for those of European. The top SNP in chromosome 10, explained 1.02% of variance for persons of African ancestry and 0.46% for those of European ancestry.

Discussion

In this study, we evaluated the association of genetic variants and spontaneous clearance of HCV infection in individuals representing a diverse ancestry background including a large population of African Ancestry individuals. In the meta-analysis of four studies with 3 ancestry groups, we confirmed the association of two known loci (*IFNL4-IFNL3* and MHC) and identified a novel locus, *GPR158* and a suggestive African ancestry specific locus *FAM168A/RELT*.

The association of HCV clearance with the *IFNL4-IFNL3* locus been extensively described in GWAS and associated in multiple gene candidate studies in several populations of different ancestries (16–18,41,42). We now observe this association with genome wide significance in the largest African ancestry and multi-ancestry Hispanic participants analyzed for spontaneous clearance. We also further narrowed the region potentially harboring the causal variants using credible sets to 1,642 base pairs extending from exon1 to exon 5 of *IFNL4*. rs74597329 drove the association of this region in all ancestry groups and was within the credible set. This variant is part of the G/TT polymorphism (rs368234815) previously described as associated with HCV clearance (41) and is also in linkage disequibrium (LD) with the variant described in our previous GWAS (rs12979860) (16) in the analyzed groups (LD r^2=0.88 for African Ancestry, LD r^2=0.99 for European Ancestry and LD r^2=0.94 for Multi-Ancestry Hispanics) and in the African (YRI LD r^2=0.76) and European (CEU LD r^2=0.97) populations of the 1000 Genomes Project (43). We found a second variant in this region with suggestive association in European Ancestry after conditioning on the top variant. The presence of several signals in an associated region is not

uncommon in complex diseases and warrants detailed sequencing across ancestry populations to disentangle the relationship and contributions of variants at this locus to identify the putative causal alleles.

Similarly, genetic variants in the MHC region were previously identified associated with HCV clearance (16). The top SNP of the previous GWAS, rs4273729, was also significant in this analysis (P value = 2.7×10^{-19}) and it is located within the credible set for this loci, 18 Kb from our top SNP, rs2647006. These two SNPs are in high LD and the association of rs4273729 decreased when conditioned on rs2647006. Our current top variant and credible sets of the variants in this region map to a 14.3 Kb intergenic-noncoding sequence between *HLA-DQB1* and *HLA-DQA2*, implying that the association with HCV clearance might be mediated through gene regulation.

In a novel finding, variants in the chromosome 10 region containing the *GPR158* gene, were associated with HCV clearance in this study. *GPR158* belongs to the glutamate receptor subfamily of G-protein-coupled receptors (GPCRs), a large and versatile superfamily of proteins that transmit signals from extracellular messenger molecules and sensory stimuli to intracellular signaling pathways (45,46). GPCRs can be activated by a diverse array of ligands including neurotransmitters, chemokines as well as sensory stimuli. Recently, it was reported that they play an important role in the immune system as they are expressed in T cells and involved with T cell-mediated immunity (46). Variants located downstream of this gene have been identified as genetic risk factors for response to smallpox vaccine in African Americans (47) and others have been associated in GWAS studies of obesity associated phenotypes (energy expenditure, resting metabolic rate, BMI and percent body fat) in American Indians (48) and a trait associated with *Trypanosoma cruzi* induced cardiomyopathy. Although the odds of clearance ranged from 24–55% which is substantial and consistent with complex traits, they were lower than the nearly 2-fold increased odds for both *IFNL4-IFNL3* and *HLA* seen in replicated studies. Thus, this study was likely the first powered study to identify this novel locus and benefitted from the effect spanning multiple ancestry populations.

We also identified a region in chromosome 11 that was significantly associated with HCV clearance in the African ancestry participants. We were unable to replicate the finding in the other populations because the top SNP is fixed in European Ancestry and has a low minor allele frequency in multi-ancestry Hispanics which precluded contribution in the meta-analysis. However, this could be an African ancestry specific finding with a large effect that should be further investigated. This region contains the genes *RELT* and *FAM168A* which mediates resistance to cisplatin in oral squamous cell carcinoma cells (49). *RELT* is a type I transmembrane glycoprotein with a cysteine-rich extracellular domain, possessing significant homology to other members of the tumor necrosis factor receptor superfamily. It activates the NF-kappa B pathway and selectively bind tumor necrosis factor receptor-associated factor 1 (50). Strong expression has been detected in adult human peripheral blood leukocytes, lymph node, spleen, and bone marrow. Genetic variants are associated with blood protein levels (65 ,66) but no significant cis- expression quantitative trait loci (eQTL) have been described in the GTEx database. These findings also highlight the importance of both trans-ethnic findings and individual ancestry findings that each may

contribute to clearance of infection and may reflect historic immune or environmental pressures that now contribute to the genetic architecture of HCV and might also define the distribution of the identified HCV risk alleles across different populations (Supplementary Figure S15). Future studies need to continue to include and expand participants from multiple ancestral backgrounds.

The stepwise increase in the size of the combined effect of GWAS associated variants support that additive genetic effects contribute to the response to HCV infections. These independent effects are most prominent when favorable genotypes are combined. Specifically, in those carrying 5 and 6 favorable alleles the odds of clearance range from 10–24x as compared to those with up to 1 favorable allele. Although the sample size is small for this effect size, the addition of the *GPR158* favorable genotype increased the positive predictive value of the *INFL4-INFL3*+MHC favorable genotypes in both African and European populations. Similarly, we found that the total variance explained by only three variants is high for all three ancestry groups as compared with other complex diseases. This highlights the importance of these genes in the pathogenesis of the disease and potential targets for vaccine development. These genes are involved in cytokine synthesis and the immune response which are key factors in the clearance of the infection. Understanding the genetic basis of host-limiting infection of HCV and particularly the identification of genetic components with high impact as those described here, provides a step forward to predict disease risk and to find clues to the prevention of the chronic infection state of the disease. The mechanisms need to be characterized, but we can use our results as a base to investigate the potential involvement of new genes in the modulation of the HCV clearance in populations of several ancestries.

This is the first genome-wide study of HCV clearance involving a cohort of more than 1,500 African ancestry individuals. The major strengths of our study design include the inclusion of a large number of subjects of several ancestries, the well characterized phenotype and the use of a dense set of markers provided by high quality imputation. Our study was powerful enough to identify one additional loci with small effect, and consequently explain more of the overall phenotypic variance and, an additional African ancestry specific locus. However, an additional expansion of the sample size will clarify the role of the detected suggestive regions in HCV clearance. The clear functional relevance of the molecular pathways for the new and known genes identified, warrants further attention to better illuminate the mechanisms of disease. Discoveries in the HLA region might guide antigen presentation choices as has occurred with the tight connection between elite suppression of HIV and HLA B*57 and KIR3DL (51). Similar knowledge derived from our current findings will augment our understanding of immune mechanisms and define biomarkers of immunity that can help in optimizing the development of a vaccine to provide protection from hepatitis C virus infection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Abbreviations

1000 Genomes	The Thousand Genomes Project
ALIVE	AIDS Linked to the IntraVenous Experience
BF	Bayes Factor
CS	Credible Set
eQTLs	Expression Quantitative Trait Loci
GPCR	Glutamate Receptor subfamily of G-protein-coupled receptor
GPR158	G protein-coupled receptor 158
GTEx	Genotype-Tissue Expression
GWAS	Genome Wide Association Studies
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
IFNL	Interferon Lambda
Kb	Kilobases
LD	Linkage Disequilibrium
MHC	Major Histocompatibility complex
UHS	Urban Health Study
OR	Odds Ratio
SE	Standard Error

SNP   Single Nucleotide Polymorphism

References

(1). Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. N Engl J Med 1999 8 19;341(8):556–562. [PubMed: 10451460]

(2). Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med 2006 5 16;144(10):705–714. [PubMed: 16702586]

(3). 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 9;49(3):423–430. [PubMed: 11511566]

(4). Seaberg EC, Witt MD, Jacobson LP, Detels R, Rinaldo CR, Margolick JB, et al. Spontaneous Clearance of the Hepatitis C Virus Among Men Who Have Sex With Men. Clin Infect Dis 2015 11 1;61(9):1381–1388. [PubMed: 26175521]

(5). Wiese M, Grungreiff K, Guthoff W, Lafrenz M, Oesen U, Porst H, et al. Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany--a 25-year multicenter study. J Hepatol 2005 10;43(4):590–598. [PubMed: 16237783]

(6). Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, et al. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science 1998 10 2;282(5386):103–107. [PubMed: 9756471]

(7). Villano SA, Vlahov D, Nelson KE, Cohn S, Thomas DL. Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. Hepatology 1999 3;29(3): 908–914. [PubMed: 10051497]

(8). Micallef JM, Kaldor JM, Dore GJ. Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies. J Viral Hepat 2006 1;13(1):34–41. [PubMed: 16364080]

(9). Mir HM, Stepanova M, Afendy M, Kugelmas M, Younossi ZM. African americans are less likely to have clearance of hepatitis C virus infection: the findings from recent U.S. population data. J Clin Gastroenterol 2012 Sep;46(8):e62–5.

(10). Conjeevaram HS, Fried MW, Jeffers LJ, Terrault NA, Wiley-Lucas TE, Afdhal N, et al. Peginterferon and ribavirin treatment in African American and Caucasian American patients with hepatitis C genotype 1. Gastroenterology 2006 8;131(2):470–477. [PubMed: 16890601]

(11). Benhammou JN, Dong TS, May FP, Kawamoto J, Dixit R, Jackson S, et al. Race affects SVR12 in a large and ethnically diverse hepatitis C-infected patient population following treatment with direct-acting antivirals: Analysis of a single-center Department of Veterans Affairs cohort. Pharmacol Res Perspect 2018 2 22;6(2):e00379. [PubMed: 29484189]

(12). Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017 7 6;101(1):5–22. [PubMed: 28686856]

(13). Bailey JR, Barnes E, Cox AL. Approaches, Progress, and Challenges to Hepatitis C Vaccine Development. Gastroenterology 2018 9 26.

(14). Shoukry NH. Hepatitis C Vaccines, Antibodies, and T Cells. Front Immunol 2018 6 28;9:1480. [PubMed: 30002657]

(15). Guo X, Zhong JY, Li JW. Hepatitis C Virus Infection and Vaccine Development. J Clin Exp Hepatol 2018 6;8(2):195–204. [PubMed: 29892184]

(16). Duggal P, Thio CL, Wojcik GL, Goedert JJ, Mangia A, Latanich R, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann Intern Med 2013 2 19;158(4):235–245. [PubMed: 23420232]

(17). Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology 2010 4;138(4):1338–45, 1345.e1–7. [PubMed: 20060832]

(18). Miki D, Ochi H, Takahashi A, Hayes CN, Urabe Y, Abe H, et al. HLA-DQB1*03 confers susceptibility to chronic hepatitis C in Japanese: a genome-wide association study. PLoS One 2013 12 20;8(12):e84226. [PubMed: 24376798]

(19). 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 5 5;357(9266):1397–1401. [PubMed: 11356437]

(20). 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 5 15;157(10):915–922. [PubMed: 12746244]

(21). Vlahov D, Munoz A, Anthony JC, Cohn S, Celentano DD, Nelson KE. Association of drug injection patterns with antibody to human immunodeficiency virus type 1 among intravenous drug users in Baltimore, Maryland. Am J Epidemiol 1990 11;132(5):847–856. [PubMed: 2239899]

(22). Tseng FC, O'Brien TR, Zhang M, Kral AH, Ortiz-Conde BA, Lorvick J, et al. Seroprevalence of hepatitis C virus and hepatitis B virus among San Francisco injection drug users, 1998 to 2000. Hepatology 2007 9;46(3):666–671. [PubMed: 17657818]

(23). Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016 10;48(10):1284–1287. [PubMed: 27571263]

(24). 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 8;38(8): 904–909. [PubMed: 16862161]

(25). Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 2014 6;197(2):573–589. [PubMed: 24700103]

(26). Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet 2009;10:387–406. [PubMed: 19715440]

(27). Epstein MP, Duncan R, Broadaway KA, He M, Allen AS, Satten GA. Stratification-score matching improves correction for confounding by population stratification in case-control association studies. Genet Epidemiol 2012 4;36(3):195–205. [PubMed: 22714934]

(28). Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015 2 25;4:7-015-0047-8. eCollection 2015.

(29). 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 9;81(3):559–575. [PubMed: 17701901]

(30). Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010 9 1;26(17):2190–2191. [PubMed: 20616382]

(31). Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao YT, et al. Addressing population-specific multiple testing burdens in genetic association studies. Ann Hum Genet 2015 3;79(2): 136–147. [PubMed: 25644736]

(32). White MJ, Risse-Adams O, Goddard P, Contreras MG, Adams J, Hu D, et al. Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. Immunogenetics 2016 7;68(6–7):391–400. [PubMed: 27142222]

(33). Mak ACY, White MJ, Eckalbar WL, Szpiech ZA, Oh SS, Pino-Yanes M, et al. Whole-Genome Sequencing of Pharmacogenetic Drug Response in Racially Diverse Children with Asthma. Am J Respir Crit Care Med 2018 6 15;197(12):1552–1564. [PubMed: 29509491]

(34). Verma A, Lucas A, Verma SS, Zhang Y, Josyula N, Khan A, et al. PheWAS and Beyond: The Landscape of Associations with Medical Diagnoses and Clinical Measures across 38,662 Individuals from Geisinger. Am J Hum Genet 2018 4 5;102(4):592–608. [PubMed: 29606303]

(35). Lee BD, Gonzalez S, Villa E, Camarillo C, Rodriguez M, Yao Y, et al. A genome-wide quantitative trait locus (QTL) linkage scan of NEO personality factors in Latino families segregating bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2017 10;174(7):683–690. [PubMed: 28556497]

(36). Morris AP. Transethnic meta-analysis of genomewide association studies. Genet Epidemiol 2011 12;35(8):809–822. [PubMed: 22125221]

(37). Wellcome Trust Case Control Consortium, Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 2012 12;44(12):1294–1301. [PubMed: 23104008]

(38). Morris AP. Fine mapping of type 2 diabetes susceptibility loci. Curr Diab Rep 2014;14(11):549–014–0549–2. [PubMed: 25239271]

(39). 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 10 8;461(7265):798–801. [PubMed: 19759533]

(40). 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 9 17;461(7262):399–401. [PubMed: 19684573]

(41). Prokunina-Olsson L, Muchmore B, Tang W, Pfeiffer RM, Park H, Dickensheets H, et al. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat Genet 2013 2;45(2):164–171. [PubMed: 23291588]

(42). Bibert S, Roger T, Calandra T, Bochud M, Cerny A, Semmo N, et al. IL28B expression depends on a novel TT/-G polymorphism which improves HCV clearance prediction. J Exp Med 2013 6 3;210(6):1109–1116. [PubMed: 23712427]

(43). 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015 10 1;526(7571):68–74. [PubMed: 26432245]

(44). Consortium GTEx. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013 6;45(6):580–585. [PubMed: 23715323]

(45). Bjarnadottir TK, Fredriksson R, Schioth HB. The gene repertoire and the common evolutionary history of glutamate, pheromone (V2R), taste(1) and other related G protein-coupled receptors. Gene 2005 12 5;362:70–84. [PubMed: 16229975]

(46). Wang D. The essential role of G protein-coupled receptor (GPCR) signaling in regulating T cell immunity. Immunopharmacol Immunotoxicol 2018 6;40(3):187–192. [PubMed: 29433403]

(47). Ovsyannikova IG, Kennedy RB, O'Byrne M, Jacobson RM, Pankratz VS, Poland GA. Genome-wide association study of antibody response to smallpox vaccine. Vaccine 2012 6 13;30(28):4182–4189. [PubMed: 22542470]

(48). Piaggi P, Masindova I, Muller YL, Mercader J, Wiessner GB, Chen P, et al. A Genome-Wide Association Study Using a Custom Genotyping Array Identifies Variants in GPR158 Associated With Reduced Energy Expenditure in American Indians. Diabetes 2017 8;66(8):2284–2295. [PubMed: 28476931]

(49). Gu Y, Fan S, Liu B, Zheng G, Yu Y, Ouyang Y, et al. TCRP1 promotes radioresistance of oral squamous cell carcinoma cells via Akt signal pathway. Mol Cell Biochem 2011 11;357(1–2):107–113. [PubMed: 21603883]

(50). Sica GL, Zhu G, Tamada K, Liu D, Ni J, Chen L. RELT, a new member of the tumor necrosis factor receptor superfamily, is selectively expressed in hematopoietic tissues and activates transcription factor NF-kappaB. Blood 2001 5 1;97(9):2702–2707. [PubMed: 11313261]

(51). Martin MP, Naranbhai V, Shea PR, Qi Y, Ramsuran V, Vince N, et al. Killer cell immunoglobulin-like receptor 3DL1 variation modifies HLA-B*57 protection against HIV-1. J Clin Invest 2018 5 1;128(5):1903–1912. [PubMed: 29461980]

Author Manuscript
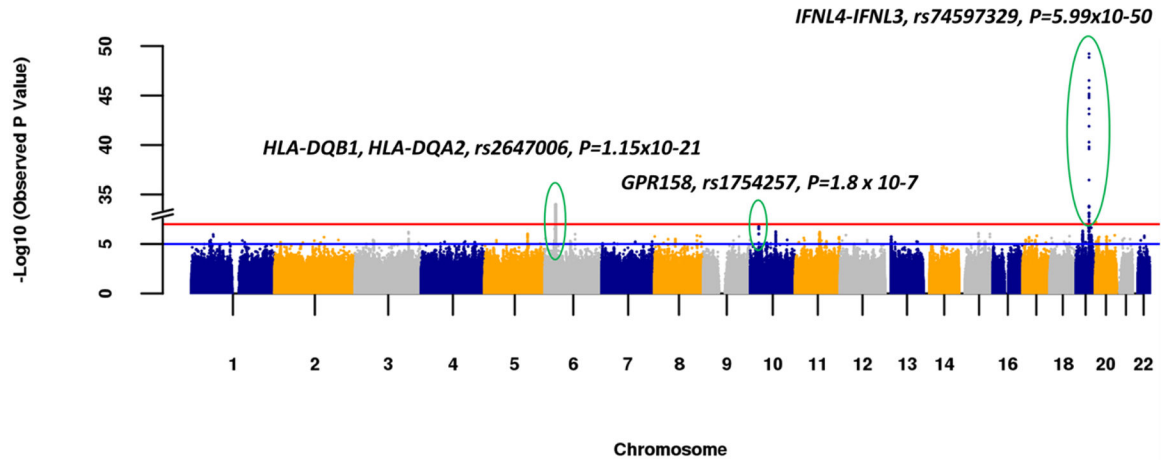
Figure 1.
Manhattan plots summarizing the fixed-effects meta-analysis results in 1319 persons with spontaneous clearance of HCV infection and 3104 persons with persistent HCV infection. Each point corresponds to the P value for a single nucleotide polymorphism. The P values are plotted by location of the variants across the genome. The red line represents the level of fixed effects meta-analysis genome-wide significance (P value = 2.05×10^{-7}) and the blue line represents the suggestive association (P value 5×10^{-5}). Variants in the Major Histocompatibility complex (*MHC*), Interferon Lambda (*IFNL4-IFNL3*) and G Protein-Coupled Receptor 158 gene (*GPR158*) regions exceed the significance threshold. P values in the chr5q23.1, chr6q14.1, chr6p22.1, chr10q22.2, chr11q13.4, chr19p12 and chr19q13.32, regions are suggestive of association.
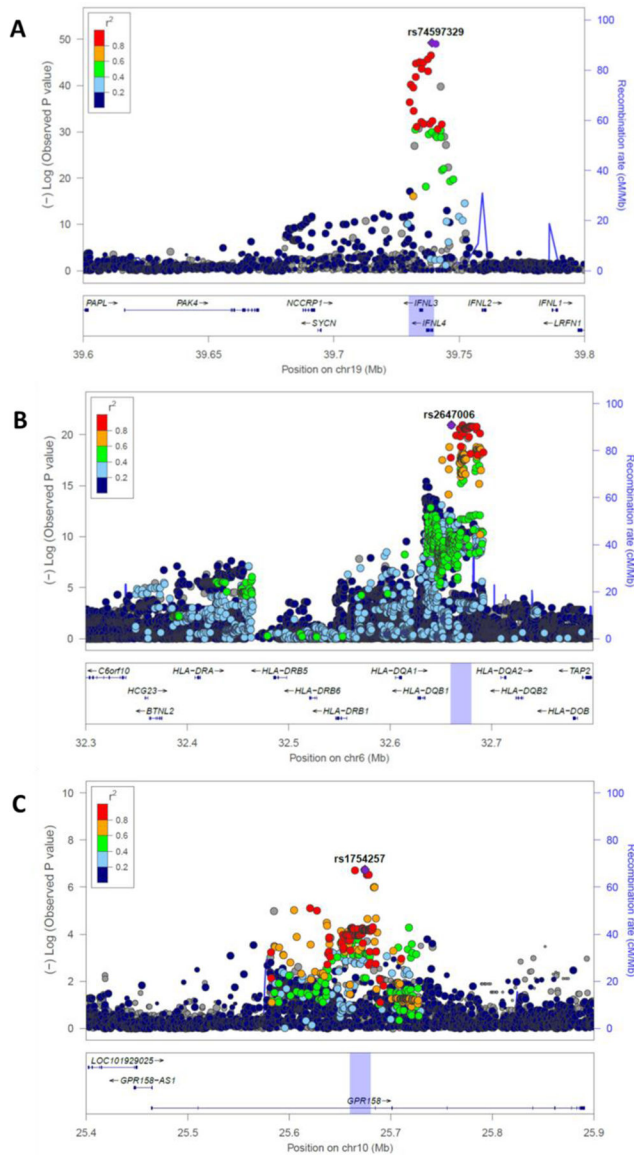
Figure 2.
Locus zoom plots presenting the results of the fixed-effects meta-analysis depicting the 95% credible sets of SNPs at the associated loci in A) Chr19q13.2; B) Chr6p21.32 and C) Chr10p12.1 regions. In each plot, each point represents a SNP passing quality control in the fixed-effects meta-analysis plotted with its P value as a function of genomic position (GRCh37 Assembly). The lead SNP is represented by the purple symbol. The color coding of all other SNPs indicates LD with the lead SNP estimated by r2 values of EUR populations of 1000 Genomes Project: red, $r2 \geq 0.8$; gold, $0.6 \leq r2 < 0.8$; green, $0.4 \leq r2 < 0.6$; cyan, $0.2 \leq r2 < 0.4$; blue, $r2 < 0.2$; gray, $r2$ unknown. Recombination rates are estimated from Phase II HapMap. The 95% credible sets were constructed on the basis of the trans-ethnic meta-analysis of individuals of European Ancestry, African ancestry Group 1, Multi-Ancestry Hispanic and African Ancestry Group 2 (1319 clearance /3104 persistence) and corresponds to the genomic region highlighted in light purple.
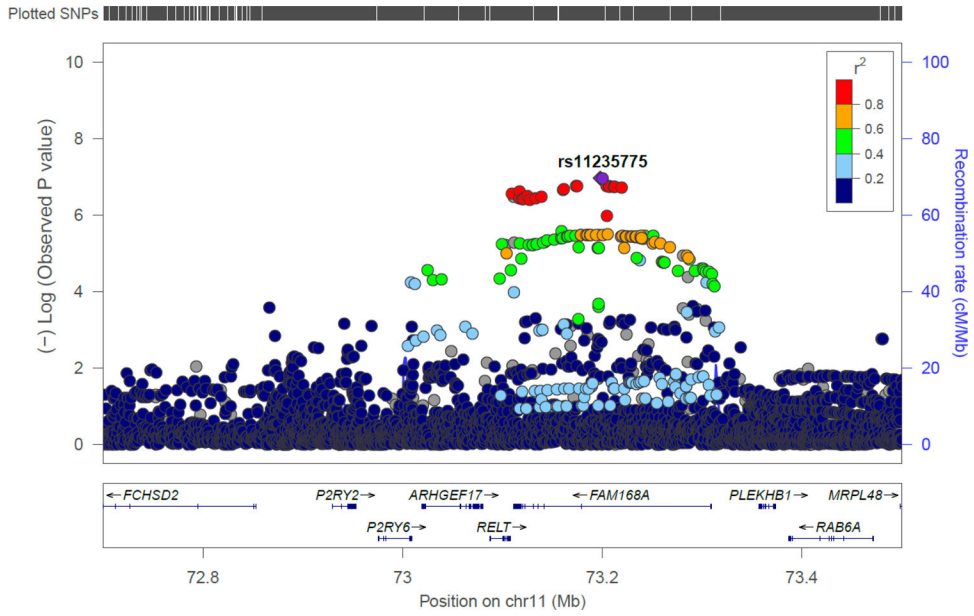
Figure 3.
Locus zoom plot presenting significant variants (at ancestry specific GWAS level) in chr11q13.4 in African Ancestry Group 1 (1529 clearance /340 persistence). In each plot, each point represents a SNP passing quality control in the logistic regression analysis of imputed dosage plotted with its P value as a function of genomic position (GRCh37 Assembly). The lead SNP is represented by the purple symbol. The color coding of all other SNPs indicates LD with the lead SNP estimated by r2 values of AFR populations of 1000 Genomes Project : red, $r2 \geq 0.8$; gold, $0.6 \leq r2 < 0.8$; green, $0.4 \leq r2 < 0.6$; cyan, $0.2 \leq r2 < 0.4$; blue, $r2 < 0.2$; gray, $r2$ unknown. Recombination rates are estimated from Phase II HapMap (Yoruba in Ibadan -YRI-population).
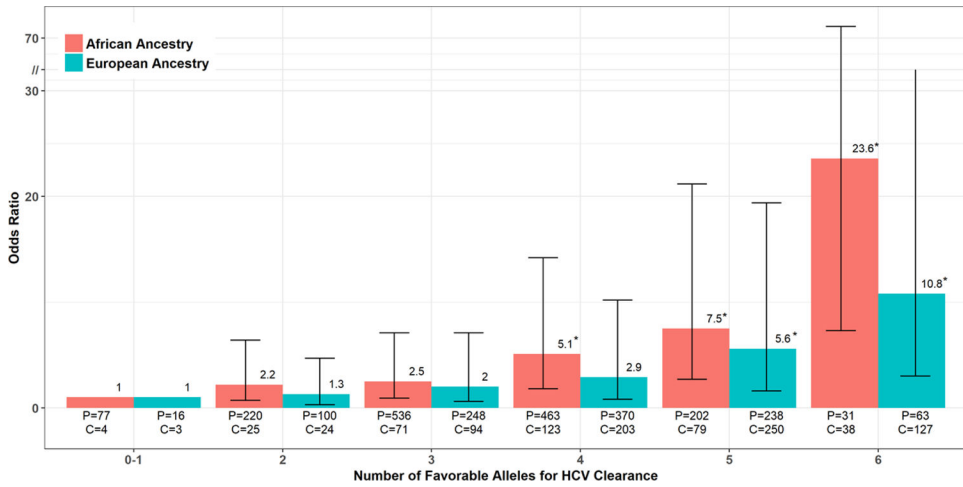
Figure 4.
Cooperative additive effect of polymorphisms on MHC, IFNL4-IFNL3 and GPR158 loci on the risk of HCV spontaneous clearance in European ancestry and African ancestry. The values for each bar represent the odds ratio (95% CI) for HCV clearance for carrying 2 to 6 favorable alleles for the top SNPs: rs74597329, rs2647006 and rs1754257 compared to carrying 0–1 favorable allele. The OR were calculated for independently for each genetically determined ancestry group. C= Clearance, P=Persistence, * P Value < 0.005

Table 1.

Demographic characteristics and genotyping arrays of the analyzed studies by genetically determined ancestry groups. Abbreviations: AMR: Americas; AFR: African.

Study Participants	Genetically Determined Ancestry Group	Genotyping Array	N	HCV infection Persistence:Clearance	(+) HIV infection	Female Sex
Extended HCV Multi-Cohort Study	African ancestry	Illumina Omni1-Quad BeadChip	1869	1529:340	38%	33%
	European ancestry	Illumina Omni1-Quad BeadChip	1736	1035:701	16%	31%
	Multi-ancestry Hispanic	Illumina Omni1-Quad BeadChip	486	313:173	12%	41%
African Ancestry Group 2	African ancestry	Infinium Multi-Ethnic AMR/AFR-8v1.0 array	332	227:105	12%	46%
	Total		4423	3104:1319	28%	34%

Table 2.

Top SNPs associated in the fixed-effects meta-analysis for Hepatitis C Virus spontaneous clearance and persistence by genetically determined ancestry group. Abbreviations: OR: Odds Ratio; Freq: frequency; SNP: Single Nucleotide Polymorphism. GRCh37/h19: Genome Reference Consortium Human Build 37. IFNL: Interferon Lambda; HLA: Human Leukocyte Antigen. A: Adenine; C: cytosine; G: Guanine; T: Thymine.

Near gene	Lead SNP	Effect Allele	Position (GRCh37/hg19)	Functional Annotation	African Ancestry Group 1			European Ancestry			Multi-Ancestry Hispanics			African Ancestry Group 2			Meta Analysis
					Freq Effect Allele	OR	P Value	Freq Effect Allele	OR	P Value	Freq Effect Allele	OR	P Value	Freq Effect Allele	OR	P Value	P value
HLA-DQB1, HLA-DQA2	rs2647006	C	6:32660582	Intergenic	0.62	1.79	8.9×10^{-10}	0.64	1.78	2.1×10^{-09}	0.69	1.71	1.2×10^{-03}	0.63	1.79	5.3×10^{-03}	1.1×10^{-21}
	rs2858319	C	6:32661294	Intergenic	0.62	1.76	2.4×10^{-09}	0.64	1.79	1.4×10^{-09}	0.69	1.75	6.9×10^{-04}	0.63	1.79	5.3×10^{-03}	1.3×10^{-21}
	rs2858318	G	6:32661302	Intergenic	0.62	1.76	2.4×10^{-09}	0.64	1.79	1.4×10^{-09}	0.69	1.75	6.9×10^{-04}	0.63	1.79	5.3×10^{-03}	1.3×10^{-21}
	rs9274748	G	6:32670308	Intergenic	0.63	1.75	3.8×10^{-09}	0.64	1.79	1.1×10^{-09}	0.7	1.71	1.1×10^{-03}	0.64	1.86	3.2×10^{-03}	1.7×10^{-21}
	rs2647051	C	6:32670897	Intergenic	0.63	1.75	3.8×10^{-09}	0.64	1.79	1.1×10^{-09}	0.7	1.71	1.1×10^{-03}	0.64	1.86	3.2×10^{-03}	1.7×10^{-21}
	rs2856699	G	6:32671765	Intergenic	0.63	1.75	3.8×10^{-09}	0.64	1.79	1.1×10^{-09}	0.7	1.71	1.1×10^{-03}	0.64	1.86	3.2×10^{-03}	1.7×10^{-21}
	rs9275521	C	6:32674952	Intergenic	0.63	1.75	3.8×10^{-09}	0.64	1.79	1.1×10^{-09}	0.7	1.71	1.1×10^{-03}	0.64	1.86	3.2×10^{-03}	1.7×10^{-21}
IFNL4-IFNL3	rs35790907	A	19:39730755	Intergenic	0.35	1.93	7.6×10^{-12}	0.70	2.29	7.9×10^{-22}	0.57	3.16	1.4×10^{-11}	0.38	1.59	1.4×10^{-02}	5.1×10^{-41}
	rs12980275	A	19:39731783	Intergenic	0.48	1.79	1.4×10^{-11}	0.70	2.27	1.5×10^{-22}	0.63	2.84	2.0×10^{-10}	0.50	1.56	2.0×10^{-02}	2.5×10^{-40}
	rs8105086	A	19:39732712	Intergenic	0.44	2.04	4.6×10^{-13}	0.69	2.31	3.1×10^{-22}	0.61	3.01	3.2×10^{-10}	0.46	1.77	3.5×10^{-02}	1.3×10^{-42}
	rs688187	G	19:39732752	Intergenic	0.41	2.05	1.9×10^{-14}	0.69	2.28	3.3×10^{-22}	0.59	3.16	2.2×10^{-11}	0.42	1.85	1.4×10^{-02}	1.7×10^{-45}
	rs4803217	C	19:39734220	Intergenic	0.41	2.03	4.4×10^{-14}	0.69	2.28	5.2×10^{-22}	0.60	3.19	1.7×10^{-11}	0.40	2.09	1.3×10^{-04}	6.8×10^{-46}
	rs8103142	T	19:39735106	Missense	0.38	1.97	6.8×10^{-13}	0.69	2.29	2.3×10^{-22}	0.58	2.98	1.4×10^{-10}	0.37	2.20	5.8×10^{-05}	2.1×10^{-44}
	rs28416813	C	19:39735644	Intergenic	0.41	2.03	6.2×10^{-14}	0.69	2.27	1.1×10^{-21}	0.60	3.23	1.4×10^{-11}	0.36	2.18	7.4×10^{-05}	1.1×10^{-45}
	rs12979731	C	19:39737513	Intergenic	0.40	2.09	1.2×10^{-14}	0.69	2.28	3.6×10^{-22}	0.59	3.23	1.1×10^{-11}	0.40	2.04	3.2×10^{-04}	1.6×10^{-46}

Near gene	Lead SNP	Effect Allele	Position (GRCh37/ hg19)	Functional Annotation	African Ancestry Group 1			European Ancestry			Multi – Ancestry Hispanics			African Ancestry Group 2			Meta Analysis
					Freq Effect Allele	OR	P Value	Freq Effect Allele	OR	P Value	Freq Effec Allele	OR	P Value	Freq Effect Allele	OR	P Value	P value
	rs11882871	A	19:39737610	Intergenic	0.42	2.01	2.8×10^{-13}	0.69	2.28	4.9×10^{-22}	0.60	3.29	1.5×10^{-11}	0.42	1.85	2.1×10^{-03}	7.3×10^{-44}
	rs12979860	C	19:39738787	Intronic	0.40	2.10	5.8×10^{-15}	0.69	2.27	5.7×10^{-22}	0.59	3.23	7.7×10^{-12}	0.41	2.18	1.1×10^{-03}	3.0×10^{-47}
	rs778086851	CT	19:39739153	Inframe deletion	0.37	2.13	1.4×10^{-15}	0.69	2.29	1.9×10^{-22}	0.58	3.27	3.2×10^{-12}	0.39	2.43	8.8×10^{-06}	1.41×10^{-49}
	rs74597329	T	19:39739155	Missense	0.37	2.14	9.8×10^{-16}	0.69	2.29	1.3×10^{-22}	0.58	3.28	2.8×10^{-12}	0.39	2.43	8.8×10^{-06}	5.9×10^{-50}
	rs553620818	T	19:39742683	Intergenic insertion	0.50	1.93	3.8×10^{-12}	0.70	2.22	5.0×10^{-21}	0.63	3.08	3.9×10^{-11}	0.49	1.64	1.1×10^{-03}	1.5×10^{-40}
GPR158	rs1410932	A	10:25664699	Intronic	0.76	1.55	4.7×10^{-05}	0.74	1.29	1.6×10^{-03}	0.70	1.24	0.18	0.80	1.10	0.68	1.9×10^{-07}
	rs1591250	A	10:25674081	Intronic	0.76	1.55	4.6×10^{-05}	0.74	1.30	1.2×1^{-03}	0.70	1.24	0.18	0.80	1.06	0.79	1.8×10^{-07}
	rs1754257	A	10:25674614	Intronic	0.76	1.55	4.6×10^{-05}	0.74	1.30	1.3×10^{-03}	0.70	1.24	0.18	0.80	1.06	0.79	1.8×10^{-07}
	rs36012813	GT/G	10:25676339	Intronic insertion	0.76	1.56	4.4×10^{-05}	0.74	1.30	1.5×10^{-03}	0.70	1.26	0.15	0.80	1.07	0.79	1.8×10^{-07}

Table 3.

Chromosomal position for the 95% credible set obtained in the meta-analysis for Hepatitis C Virus spontaneous clearance and persistence in European Ancestry, two groups of African Ancestry and Multi-Ancestry Hispanics (1319 clearance /3104 persistence). Positions are based on the Genome Reference Consortium Human Build 37 (GRCh37/hg19). Abbreviations: SNP: Single Nucleotide Polymorphism. IFNL: Interferon Lambda; GPR58: G-Protein-Receptor 58; HLA: Human Leukocyte Antigen. A: Adenine; C: cytosine; G: Guanine; T: Thymine.

Locus	Lead SNP	Effect allele/Non Effect allele	Region containing significant SNPs (Kb)	95% credible set region (length in Kb)	Significant SNPs in credible set
IFNL4-IFNL3	rs74597329	G/T	19: 39680914–39752820 (71.9)	19:39737513–39739155 (1.6)	9
HLA-DQB1, HLA-DQA1, HLA-DQA2	rs2647006	G/T	6:31391401–32681631 (305.2)	6:32660582–32674952 (14.3)	173
GPR158	rs1754257	T/C	10:25664699–25680150 (15.4)	10:25664699–25677287 (12.5)	5