The Journal of Infectious Diseases

MAJOR ARTICLE

A Multiancestry Sex-Stratified Genome-Wide Association Study of Spontaneous Clearance of Hepatitis C Virus

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

[1]Johns Hopkins University, Bloomberg School of Public Health, Baltimore, Maryland, USA, [2]Johns Hopkins University, School of Medicine, Baltimore, Maryland, USA, [3]Universidad Pontificia Bolivariana, Medellín, Colombia, [4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA, [5]Liver Unit IRCCS "Casa Sollievo della Sofferenza," San Giovanni Rotondo, Italy, [6]RTI International, Research Triangle Park, North Carolina, USA, [7]Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA, [8]Liver Center and Gastrointestinal Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA, [9]Division of Gastroenterology, Department of Medicine, School of Medicine, University of California, San Francisco, California, USA, [10]University of Southampton, Southampton General Hospital, Southampton, United Kingdom, [11]Department of Internal Medicine and Digestive Diseases, CHU Rangueil, UMR 152 IRD, Toulouse 3 University, France, [12]South West Liver Unit, Plymouth, United Kingdom, [13]Rho, Inc., Chapel Hill, North Carolina, USA, [14]SUNY Downstate College of Medicine, Brooklyn, New York, USA, [15]University of California and Vitalant Research Institute, San Francisco, California, USA, [16]UCL Institute for Liver and Digestive Health, Royal Free Hospital, Hampstead, London, United Kingdom, [17]University of Colorado, Aurora, Colorado, USA

Background. Spontaneous clearance of acute hepatitis C virus (HCV) infection is more common in women than in men, independent of known risk factors.

Methods. To identify sex-specific genetic loci, we studied 4423 HCV-infected individuals (2903 male, 1520 female) of European, African, and Hispanic ancestry. We performed autosomal, and X chromosome sex-stratified and combined association analyses in each ancestry group.

Results. A male-specific region near the adenosine diphosphate–ribosylation factor–like 5B (*ARL5B*) gene was identified. Individuals with the C allele of rs76398191 were about 30% more likely to have chronic HCV infection than individuals with the T allele (OR, 0.69; $P = 1.98 \times 10^{-07}$), and this was not seen in females. The *ARL5B* gene encodes an interferon-stimulated gene that inhibits immune response to double-stranded RNA viruses. We also identified suggestive associations near septin 6 and ribosomal protein L39 genes on the X chromosome. In box sexes, allele G of rs12852885 was associated with a 40% increase in HCV clearance compared with the A allele (OR, 1.4; $P = 2.46 \times 10^{-06}$). Septin 6 facilitates HCV replication via interaction with the HCV NS5b protein, and ribosomal protein L39 acts as an HCV core interactor.

Conclusions. These novel gene associations support differential mechanisms of HCV clearance between the sexes and provide biological targets for treatment or vaccine development.

Keywords. HCV; GWAS; Sex; X chromosome; *ARL5B*; Septin 6; Host-genetics; infection; immune.

Hepatitis C virus (HCV) is a blood-borne pathogen that results in either spontaneous clearance or persistence of the virus with chronic hepatitis after an acute infection. Between 15% and 45% of individuals acutely infected with HCV will spontaneously clear the virus, and the others will have persistence of the HCV infection, which is a major cause of cirrhosis and hepatocellular carcinoma. Globally, >70 million people have persistent infection with HCV, and these infections are a major contribution to global mortality rates [1, 2].

Females are twice as likely as males to spontaneously clear HCV infection, and black individuals are 5-fold less likely than white individuals [3–6]. An estimated 6%–15% of the variation in spontaneous HCV clearance is explained by autosomal markers in the human genome [7–9]. The interferon lambda 3 and interferon lambda 4 (*IFNL3–IFNL4*) locus, the major histocompatibility complex (MHC) class II region and G protein–coupled receptor 158 (*GPR158*) gene are all autosomal regions associated with spontaneous clearance of HCV infection and explain in part differences between ancestral populations [7, 9]. However, the genetic explanation for the sex differences is unknown, because autosomal variants have not been tested for sex-specific effects and the X chromosome has not been systematically evaluated for association with HCV clearance despite the noted sex differences. We performed an autosomal genome-wide sex-stratified analysis and X chromosome analyses in the Extended HCV Genetic Consortium [9] to identify sex-specific loci associated with HCV clearance in populations of diverse ancestry.

Received 27 July 2020; editorial decision 19 October 2020; accepted 28 October 2020; published online October 29, 2020.

Presented in part: American Society of Human Genetics Virtual Meeting 2020. October 27-30 2020.

Correspondence: Candelaria Vergara, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 725 N Wolfe St, Baltimore, MD 21205 (cvergar2@jhmi.edu).

The Journal of Infectious Diseases® 2021;223:2090–8

DOI: 10.1093/infdis/jiaa677

METHODS

Study Population

The analyzed population included 4423 individuals (2903 male, 819 with clearance, and 1520 female, 500 with clearance) participating in the HCV Extended Genetics Consortium, as described elsewhere [9]. The distribution of the population by ancestry, sex, and HCV infection status (Table 1) is detailed elsewhere [9]. Based on principal components, we genetically determined participants' ancestry, which was assigned to 4 data sets as having predominantly European ancestry (n = 1736; 701 with HCV clearance), African ancestry (n = 2201; 445 with HCV clearance), distributed as 2 independent data sets depending on the genotyping platform (African ancestry [n = 1869; 340 with HCV clearance] and African Ancestry Group 2 [n = 332; 105 with HCV clearance]), and Hispanics with multiancestry components (n = 486; 173 with HCV clearance), as described elsewhere [9]. Each individual study obtained consent for genetic testing from the governing institutional review board, and the Johns Hopkins School of Medicine Institutional Review Board approved the overall analysis.

Genotyping and Quality Control

Genotyping and Imputation

Genotypes on chromosomes 1–22 and the X chromosome pseudo-autosomal and non–pseudo-autosomal regions were derived from a genome-wide association study (GWAS) performed in the HCV Extended Genetics Consortium, as described elsewhere [9]. Briefly, individuals were genotyped using the Illumina Omni1-Quad BeadChip or Infinium Multi-Ethnic Global BeadChip (MEGA) (Illumina), and autosomal genotypes were processed using standard GWAS protocols for quality control for each platform. X chromosome–specific quality control was done using the protocol described in the XWAS tools [10, 11] and included removing variants in the pseudo-autosomal regions and filtering out variants with significantly different minor allele frequency or missingness between male and female controls and those that were not in Hardy-Weinberg equilibrium in females. Genotype imputation was performed using Minimac3 software [12] for each platform through the publicly available Michigan Imputation Server [13], as described elsewhere [9]. Variants with imputation quality $r^2 < 0.3$ [12, 14, 15] and a minor allele frequency <5% for each sex category of each ancestry group were excluded. After quality control of the data sets in each ancestry and sex category, we included in the analysis 5 595 785 markers on chromosomes 1–22 and 134 908 on the X chromosome. These markers were consistently genotyped/imputed across the data sets.

Population Stratification

We assessed and corrected for potential population stratification via either autosomal-derived or X chromosome–derived principal component analysis, and we studied the inflation of test statistics in each association analysis to evaluate the most accurate correction for each model. This approach was implemented given the sex-biased history of the populations, which is expected to lead to differential population structure on the X chromosome and the autosomes. The population structure on the X chromosome captures a 1:2 male-to-female contribution, while on the autosomes males and females contribute equally to the observed structure [10]. Principal component analysis was done using EIGENSOFT software, version 7.2.1 [16, 17], using a set of markers pruned for linkage disequilibrium and removing large linkage disequilibrium blocks and regions of high linkage disequilibrium in PLINK software (version 2.0 alpha) [18, 19]. For the sex-stratified association tests in autosomes in all ancestry groups and in the X chromosome in Hispanics, we included 10 autosomal-derived principal components. For the X chromosome sex-stratified analysis in the European and African ancestry populations, we included 10 X chromosome–derived principal components because we obtained

Table 1. Distribution of Anayzed Populations by Ancestry Groups, Sex, Human Immunodeficiency Virus Infection Status, and Hepatitis C Virus Clearance or Persistence

Study Participants by Ancestry Group	Participants, Total (Male:Female), No.	HCV Infection, No. of Participants (% HIV Positive)					
		Male Participants		Female Participants		Total	
		HCV Persistence	HCV Clearance	HCV Persistence	Clearance	HCV Clearance	HCV Persistence
Extended HCV Multi-Cohort Study							
African ancestry	1869 (1245:624)	1054 (35.1)	191 (34.0)	475 (42.3)	149 (57.8)	1529 (37.3):	340 (44.4)
European ancestry	1736 (1192:544)	703 (17.1)	489 (14.6)	332 (10.5)	212 (15.1)	1035 (14.9)	701 (14.7)
Multiancestry Hispanic	486 (288:198)	197 (27.9)	91 (14.2)	116 (46.5)	82 (56.1)	313 (34.8)	173 (34.1)
African Ancestry Group 2							
African ancestry	332 (178:154)	130 (8.5)	48 (22.9)	97 (11.3)	57 (14.0)	227 (9.7)	105 (18.1)
Total	4423 (2903:1520)	2084 (26.7)	819 (19.5)	1020 (29.5)	500 (34.4)	3104 (26.7)	1319 (19.5)

Abbreviations: HCV, hepatitis C virus; HIV, human immunodeficiency virus.

better correction of the inflation of test statistics, as observed in quantile-quantile (QQ) plots.

Statistical Analysis

Single-Nucleotide Polymorphism–Based Association Analysis

Association Tests in Autosomes. In each of the ancestry groups and sex categories (8 individual datasets), we performed an association analysis of dosage of the variants using an additive logistic regression model, adjusting for population stratification and human immunodeficiency virus (HIV) infection status using Mach2dat [20] in order to detect sex-based differences in allelic frequencies. Single-nucleotide polymorphism (SNP)–based association testing compares the frequency of alleles (dosage) between HCV clearance and persistence groups to determine whether a specific allele co-occurs with a phenotype more often than would be expected by chance.

Results of each sex in each ancestry group (4 data sets per sex) were meta-analyzed using a fixed-effect (inverse variance-weighted) model, where the effect size estimates, are weighted by their estimated standard errors [21] yielding meta-analyzed sex-specific *P* values for association and corresponding sex-specific effect estimates. Miami and QQ plots were used to visualize results using EasyStrata software, version 8.6 (www.genepiregensburg.de/easystrata). 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, as described in detail elsewhere [9]. To estimate the GWAS threshold we used the widely implemented method recommended by Sobota et al and others [22–26], which estimates the effective number of independent tests in a genetic data set after accounting for linkage disequilibrium between SNPs using the linkage disequilibrium pruning function in the PLINK 2.0 alpha software package [18, 19]. In the fixed effects meta-analysis for each sex, the GWAS significance level threshold was set at 2.05×10^{-07} and the suggestive significance was set to *P* values between 5×10^{-05} and 2.05×10^{-07}.

Association Tests in the X Chromosome. We applied 2 approaches for the analysis of markers on the X chromosome: (1) sex-stratified analyses with meta-analysis across ancestry groups for each sex category and also in the total sample including males and females from all ancestry groups; (2) sex-combined analysis in the total sample incorporating biological data of X chromosome inactivation. Allelic dosages on the X chromosome were transformed to hard call genotypes using PLINK 2.0 alpha software [18, 19], and imputed markers with uncertainty greater than 0.1 were treated as missing.

For the sex-stratified analysis in each ancestry and sex category (8 individual data sets), we performed an association analysis of hard call genotypes using an additive logistic regression model including HIV infection status and 10 principal components. Males were coded as either having 0 or 2 copies of an allele and females were coded as having 0, 1, or 2 alleles of each marker. Standard errors and *P* values from each sex and ancestry group were genomic control (GC) corrected using lambda factors (λGC) and meta-analyzed using a fixed-effect (inverse variance–weighted) model using GWAMA software [21] yielding meta-analyzed sex-specific *P* values for association and corresponding sex-specific effect estimate (obtained from 4 data sets in each sex category). Miami and QQ plots were used to visualize results of the sex-stratified meta-analysis for males and females using EasyStrata software (www.genepiregensburg.de/easystrata) and R software, version 3.5.2 (https://cran.r-project.org/). This method is not affected by the allele coding in males (eg, 0/2 copies or 0/1 copies of the effect allele) and aims to capture the information about markers that have similar effect size and direction across populations in each sex category (sex-specific associations).

Because this is the first association analysis of markers on the X chromosome in this multiancestry study, in addition to the sex-specific association, we evaluated the association in the complete population. Thus, summary statistics of the associations on the X chromosome in each sex and ancestry group (8 data sets) were meta-analyzed using the method described by Stouffer et al [27], as implemented with METAL software [28] to produce a significance value for each SNP in the entire population (*P* value for all individuals). This method accommodates the possibility of differential effect size and direction between males and females and is not affected by the allele coding in males.

In the sex-combined analysis, we implemented an X chromosome–specific approach that allows for modeling SNP effects depending on X chromosome inactivation status (inactivation vs escape) [29]. We used previously published biological data of inactivation status across multiple studies and multiple tissue types for genes on the X chromosome [29, 30] to define regions likely to experience X chromosome inactivation or escape and model SNP effects. This approach assumes that under a given coding scheme, the SNP effect is the same in males and females. We used logistic regression models including hard call genotypes, HIV infection status and 10 principal components as covariates, similar to the sex-stratified analysis. However, in this approach we included males and females in a single analysis for each ancestry group (4 data sets, each containing females and males). For females, SNPs on the X chromosome were coded as the number of copies of the minor allele (0, 1, 2) irrespective of the biological inactivation status and, for males the coding depended of the biological inactivation status (0 or 2 copies of the minor allele for regions under inactivation and 0 or 1 copy for regions with escape of inactivation) [29].

For regions with no biological data, we modeled the genotype variable under both coding schemes for males, and the Akaike information criterion was used to determine which inactivation/

escape status led to the better fitting model in each ancestry group (lower Akaike information criterion indicates better model fit). If the better model fit was not consistent across the ancestry groups, then we assumed the region to be under inactivation, because most of the X chromosome is subject to inactivation [29]. The results of the sex-combined analysis for each ancestry group (4 data sets) were combined using Stouffer's method [27] and METAL software [28] to produce a sex-combined *P* value.

Conditional Analysis and Evaluation of SNP × Sex Interaction

To determine whether the observed associations in autosomes or the X chromosome were independent and also to detect interaction between the loci, we performed conditional analysis using logistic regression models in each sex category and ancestry data set (8 data sets), adjusting for population stratification, HIV infection status, and the associated loci using Mach2dat software [20]. Results of the conditional analysis in each data set were meta-analyzed, following the methods described before.

In addition, to evaluate whether the effect of the variants depended on the sex, we evaluated the SNP × sex interaction by performing association analysis, using an additive logistic regression model in each ancestry data set (4 data sets, each including males and females), including population stratification, HIV infection status, sex, and the SNP × sex interaction term. Results of the SNP × sex interaction term were meta-analyzed with a fixed-effect model, as described elsewhere.

RESULTS

In the sex-stratified autosomal analysis we identified a region on Chr10p12. 31 that was associated with spontaneous clearance of HCV in males with genome-wide significance (λGC, 1.003) (Figure 1 and Supplementary Figure 1). This region is located 137 kb downstream of the 3' end of the gene adenosine diphosphate–ribosylation factor–like 5B (*ARL5B*) and 229 kb upstream of the MAM and LDL receptor class A domain–containing 1 (*MALRD1*) gene (Figure 2). The lead marker is rs76398191 (C > T), and in males each copy of the allele C confers a 31% reduced likelihood of clearance compared with the T allele (OR, 0.69; $P = 1.98 \times 10^{-07}$) but this association was not seen in females (OR, 1.01; $P = .94$) (Table 2). The SNP × sex interaction confirmed the stratified analysis that males with the C allele rs76398191 (C > T) were less likely than females to spontaneously clear HCV infection ($P = 3.4 \times 10^{-04}$ for rs76398191 × sex interaction) (Supplementary Table 1).

In addition, 2 previously observed associated loci on Chr19q13.2 (*IFNL4*) and Chr6p21.32 (*MHC*) [9] were confirmed in both sexes independently. These associations were consistent in effect size, direction, and strength of association (Figure 1 and Supplementary Table 2). In

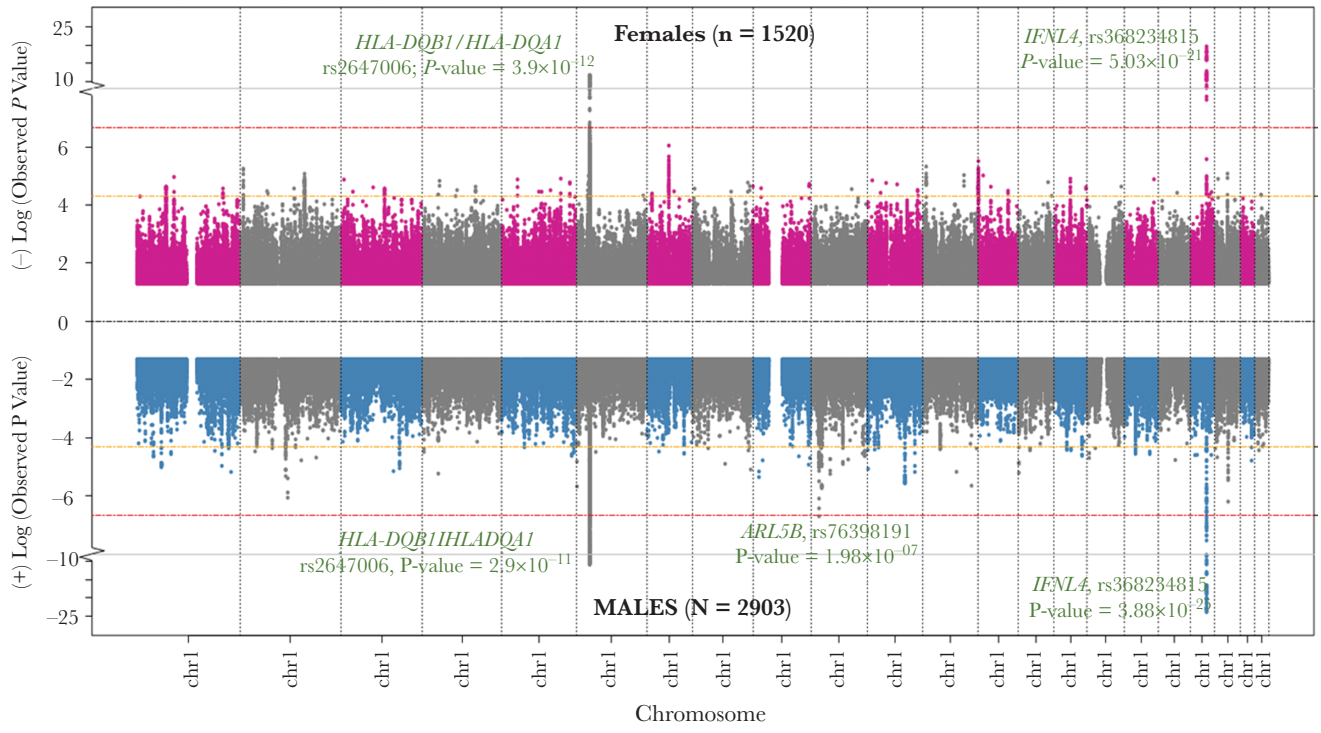

Figure 1. Miami plot showing results of sex-stratified multiancestry genome-wide fixed-effects meta-analyses in chromosomes 1–22, highlighting a new significant locus in Chr10p12.31 in males and confirming known loci on chromosome 19 and 6 in both sexes. Association *P* values are shown for the female-specific (*top*) and male-specific (*bottom*) meta-analyses. The red line indicates a genome-wide association study (GWAS) significant threshold ($P = 2.05 \times 10^{-07}$), and the yellow line, to a GWAS suggestive threshold ($P = 5 \times 10^{-05}$). Abbreviations: *ARL5B,* adenosine diphosphate–ribosylation factor–like 5B; Chr, chromosome; *IFNL4,* interferon lambda 4.

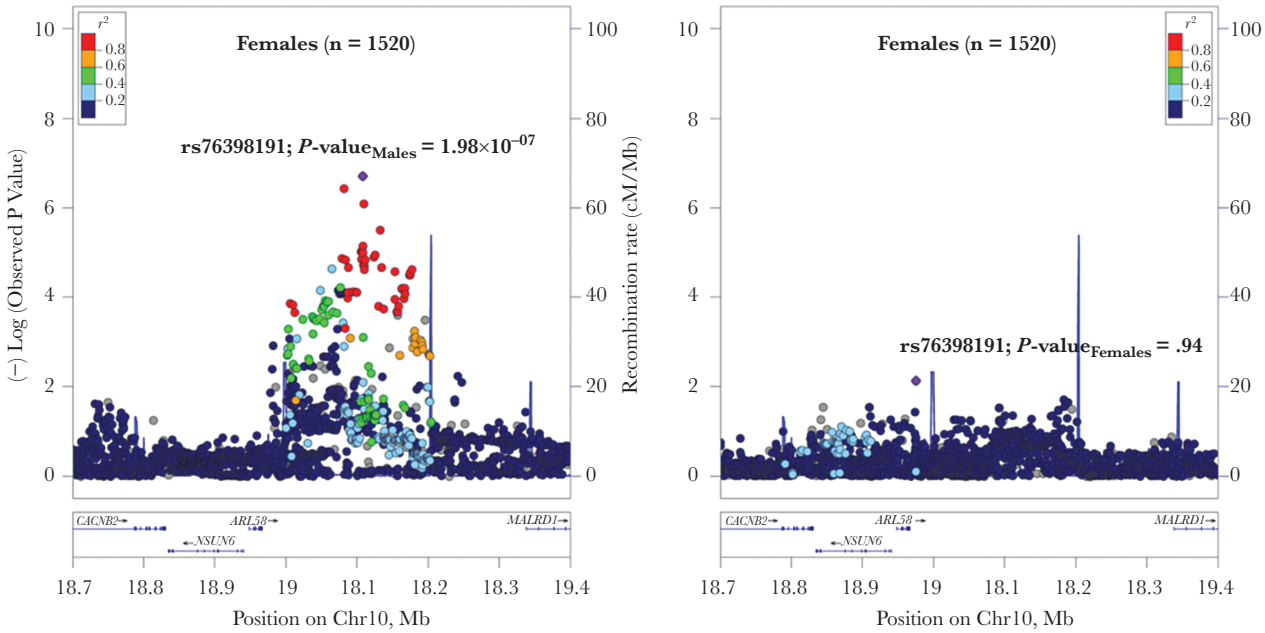

Figure 2. Locus zoom plot presenting the results of the fixed-effects multiancestry meta-analysis in Chr10p12.31 region in males and females. Each point represents a single-nucleotide polymorphism (SNP) passing quality control plotted with its *P* value for males or females as a function of genomic position (Genome Reference Consortium Human genome build 37 assembly). The lead SNP is represented by the purple symbol. The color coding of all other SNPs indicates linkage disequilibrium with the lead SNP estimated by r_2 values for European populations of the 1000 Genomes Project (*red,* $r_2 \geq 0.8$; *gold,* $0.6 \leq r_2 < 0.8$; *green,* $0.4 \leq r_2 < 0.6$; *cyan,* $0.2 \leq r_2 < 0.4$; *blue,* $r_2 < 0.2$; *gray,* r_2 unknown). Abbreviations: *ARL5B,* adenosine diphosphate–ribosylation factor–like 5B; *CACNB2,* calcium voltage-gated channel auxiliary subunit beta 2; *NSUN6,* NOP2/Sun RNA methyltransferase 6; *MALRD1,* MAM and LDL receptor class A domain containing 1.

males, the Chr10p12.31 locus association and the effect size of the associated allele was not modified when conditioned on either the *IFNL4* (OR for rs76398191 conditioned on rs368234815, 0.68; $P = 3.4 \times 10^{-07}$) or the *MHC* loci (OR for rs76398191 conditioned on rs2647006, 0.69; $P = 3.61 \times 10^{-07}$).

We also identified 3 suggestive regions associated with a 40%–60% higher likelihood of spontaneous clearance of HCV in males. The previously described locus in *GPR158* on Chr10p12.1 [9], as well as 2 new regions, on Chr2q14.1 (rs1346763; T > C) upstream of the dipeptidyl peptidase–like 10 (*DPP10*) gene and on Chr20q11.22 (rs6087561; C > T) upstream of the Agouti signaling protein (*ASIP*) gene (Supplementary Material, Supplementary Figure 2, and Supplementary Table 2).

In the female sex-stratified autosomal analysis, except for chromosome 19 (*IFNL4*) and chromosome 6 (*MHC*), we did not find any genome-wide significant associations (λGC, 0.99). However, a suggestive association of 60%

Table 2. Top Associated Variants With Hepatitis C Virus Spontaneous Clearance in Multiancestry Meta-analyses in Male and Female Participants

Nearby Gene and Associated SNPs	Chromosome: Position (GRCh37/hg19)	Effect allele	Males Participants (n = 2903)			Female Participants (n = 1520)		
			Effect Allele Frequency	OR	*P* Value	Effect Allele Frequency	OR	*P* Value
HLA-DQB1, HLA-DQA2								
rs2647006	6:32 660 582	C	0.64	1.69	2.9×10^{-11}	0.63	1.97	3.93×10^{-12}
rs2858319	6:32 661 294	C	0.64	1.68	4.0×10^{-11}	0.63	1.98	3.0×10^{-12}
rs9275521	6:32 674 952	C	0.64	1.69	2.6×10^{-11}	0.64	1.96	6.9×10^{-12}
ARL5B								
rs76398191	10:19 108 059	C	0.31	0.69	1.9×10^{-07}	0.32	1.01	0.94
rs12573426	10:19 081 778	A	0.30	0.69	3.7×10^{-07}	0.31	1.04	0.65
IFNL3–IFNL4								
rs12979860	19:39 738 787	C	0.54	2.48	5.8×10^{-24}	0.52	2.46	2.5×10^{-20}
rs778086851	19:39 739 153	CT	0.52	2.52	7.4×10^{-25}	0.51	2.50	4.6×10^{-21}
rs74597329	19:39 739 155	T	0.52	2.53	3.9×10^{-25}	0.51	2.50	5.0×10^{-21}

Abbreviations: *ARL5B,* adenosine diphosphate–ribosylation factor–like 5B; GRCh37/hg19, Genome Reference Consortium Human genome build 37/Human Genome version 19; *IFNL4,* interferon lambda 4; *IFNL3,* interferon lambda 3; OR, odds ratio for hepatitis C virus spontaneous clearance; SNP, single-nucleotide polymorphism.

higher likelihood of clearance was observed on Chr7p12.1 (rs1649709; G > A), in an intergenic region upstream of the gene POM121 transmembrane nucleoporin–like 12 (*POM121L12*). This signal is not present in males even though the minor allelic frequency of the lead SNP is the same. Other potential signals in females were observed, as described in the Supplementary Materials (Supplementary Figure 3 and Supplementary Table 2).

In the sex-stratified analysis of X chromosome, the meta-analysis of the 4 ancestry groups in each sex category, did not reveal any significant sex-specific associated SNPs (Supplementary Figure 4). However, in the X chromosome meta-analysis of the complete sample we found a suggestive association with rs12852885 (G > A; ChrX:118 886 370 [$P = 2.46 \times 10^{-06}$ in all individuals]; allele G: OR, 1.46 in males [$P = 5.7 \times 10^{-04}$] and 1.36 in females [$P = 3.4 \times 10^{-03}$]) along with markers in high linkage disequilibrium spanning a 20.2-kb region on ChrXq24 (ChrX: 118 883 490–118 903 763), as displayed in Figure 3. Rs12852885 is located in an intergenic region 77.6 kb upstream of the septin 6 (*SEPT6*) gene and 34 kb downstream from the ribosomal protein L39 (*RPL39*) gene (Supplementary Figure 5) (λGC, 0.99).

The suggestive association with rs12852885 was also observed in the sex-combined analysis incorporating biological data of X inactivation. However, the strength of the significance was lower using this model (sex combined analysis for rs12852885, $P = 1.21 \times 10^{-05}$) (Supplementary Figure 6).

DISCUSSION

In this analysis, we present a sex-stratified genome-wide analysis of spontaneous HCV clearance in autosomes and the X chromosome in multiancestry populations. We demonstrate that previously identified associations in the *MHC* region, *IFNL4,* and *GPR158* [9] have a similar magnitude and direction of effect in males and females; therefore, they do not explain differences in spontaneous clearance rates between the sexes. We discovered a novel association in males near *ARL5B* on chromosome 10. *ARL5B* is an interferon (IFN)–stimulated gene [31] up-regulated in murine hepatocytes after treatment with IFN-λ1 and IFN-α2a [32]. Arl5B binds to the C-terminal domain of melanoma differentiation–associated protein 5 (MDA5) inhibiting its activation by double-stranded RNA [5, 33]. Because MDA5 stimulates antiviral IFNs in response to viral RNAs, the net effect of Arl5B appears to be to restrain that cascade. Kitai et al [34] reported that overexpression of Arl5B repressed IFN-β promoter activation by MDA5 in HEK293 cells and Arl5B-knockdown enhanced MDA5-mediated responses. In addition, Arl5B-deficient mouse embryonic fibroblast cells exhibited increased type I IFN expression in response to MDA5 agonists such as poly (I:C) and encephalomyocarditis virus. Mechanistically, Arl5B binds to the C-terminal domain MDA5 and prevents its interaction with double-stranded RNA, acting as a negative regulator for MDA5 [34].

In this context, our finding suggests a model in which differences in *ARL5B* or its expression contribute to the net antiviral response to HCV in men, but not women. Because our finding

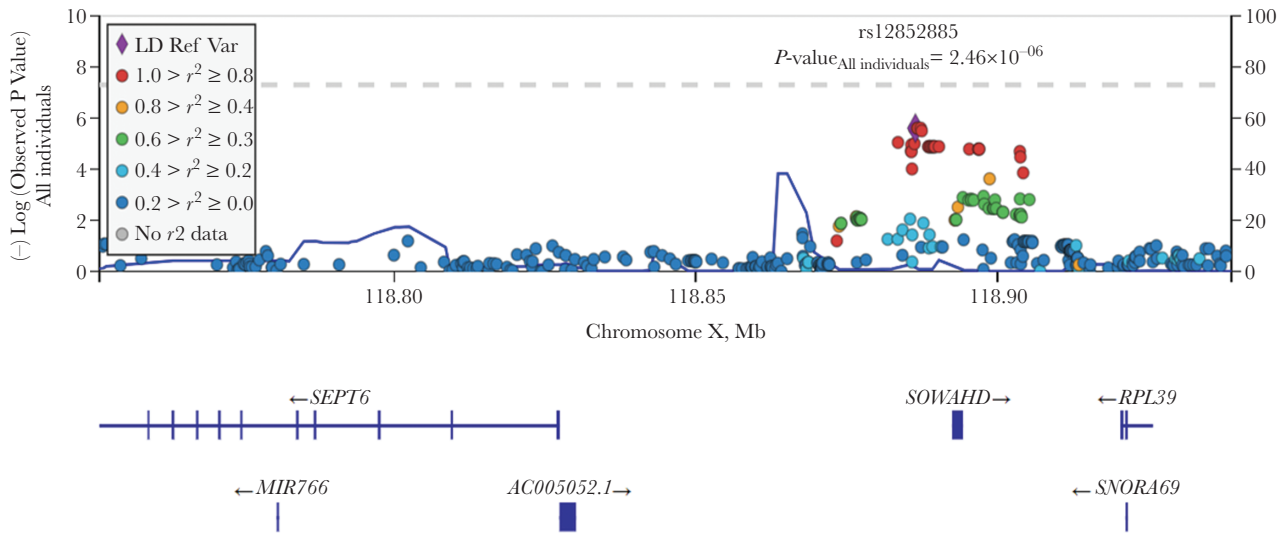

Figure 3. Locus zoom plot of the significance *P* value for the X chromosome multiancestry meta-analysis of all individuals (male and female), showing suggestively associated markers in the ChrXq24 region. Each point represents a single-nucleotide polymorphism (SNP) passing quality control, plotted with its *P* value as a function of genomic position (Genome Reference Consortium Human genome build 37 assembly). The lead SNP is represented by the purple symbol. The color coding of all other SNPs indicates linkage disequilibrium with the lead SNP estimated by r_2 values for European populations of the 1000 Genomes Project (*red,* $r_2 \geq 0.8$; *gold,* $0.6 \leq r_2 < 0.8$; *green,* $0.4 \leq r_2 < 0.6$; *cyan,* $0.2 \leq r_2 < 0.4$; *blue,* $r_2 < 0.2$; *gray,* r_2 unknown. Abbreviations: *AC005052.1, Homo sapiens* BAC clone CTB-38K21 from Xq23; *MIR766,* microRNA 766; *RPL39,* ribosomal protein L39; *SEPT6,* septin 6; *SNORA69,* small nucleolar RNA; *SOWAHD,* Sosondowah Ankyrin repeat domain family member D, H/ACA box 69.

was not in the gene itself, differences in regulation may explain the association. In fact, differences in expression of *ARL5B* have already been linked to regional SNPs. Specifically, dendritic cells from peripheral blood monocytes of healthy individuals homozygous for the allele A of rs11015435 had higher expression of *ARL5B* after stimulation with *Escherichia coli* lipopolysaccharide, influenza, or IFN-β [25]. The identified chromosome 10 locus in our study led by the association of rs76398191 is 133 kb upstream from rs11015435, which was neither associated with HCV clearance in males in our study (P = .43; minor allele frequency, 0.21) nor is in linkage disequilibrium with rs76398191 (r_2 = 0.001–0.002). Nonetheless, while the finding is consistent with the notion that host genetic differences in *ARL5B* expression may exist, the HCV effect appears distinct. In addition, our model implies that the differential regulation of *ARL5B* occurs only in men, which might occur if the specific regulatory region was silenced only in women, resulting in a differential effect. Our model is consistent with the recent work of Schmiedel and colleagues [35], who not only linked SNPs with expression of immune genes but who also demonstrated that such findings were cell specific and, especially, sex specific.

Another interesting finding is a suggestive association of HCV clearance with a region harboring *SEPT6* and *RLP39* genes on the X chromosome. Septins belong to a family of guanosine triphosphate–binding proteins that function as a dynamic scaffold recruiting other proteins [36–39]. Kim et al [40] described that NS5b, septin 6, and heterogenous nuclear ribonucleoproteins A1 form a complex that facilitates HCV replication. Thus, *SEPT6* codes for a host protein playing an important role in the replication of HCV through RNA-protein and protein-protein interactions. Moreover, knockdown of septin 6 and overexpression of an N-terminally truncated septin 6 inhibited HCV replication [40].

Similarly, *RPL39* encodes a protein that has been identified as an HCV core interactor [41] with differential expression in human hepatic cells that are highly permissive versus nonpermissive to HCV infection [42]. Other functionally related genes in the region include NF-κB repressing factor (*NKRF*) and NF-κB activating protein (*NKAP*). *NKRF* is a transcriptional silencer implicated in the basal silencing of specific NF-κB targeting genes, including *INOS*, *IFN-β*, interleukin 8 (*IL-8*)/*CXCL8*, and interferon gamma-induced protein 10 (*IP-10*) [43, 44]. *NKAP* encodes a transcriptional repressor in T regulatory cells required for conventional T-cell maturation [45]. Interestingly, the position of the associated markers completely overlaps with a regulatory element (GH0XJ119753) of those nearby genes. The direction and size of the effect for the associated allele was the same in males and females, but the statistical significance appeared higher in males despite loss of power when analyzing haploid genotypes. The different analytical models used in the evaluation of the X chromosome helped identify variants that may be under different states of chromosome inactivation or escape in a model that is more biologically accurate than the standard statistical models without biological inactivation data. Further analysis is needed to determine whether the promoter/enhancer that harbors the signal modulates the expression of the genes and whether the variant or variants in linkage disequilibrium in the region alter the functionality of this regulatory element.

One limitation of this study is that we did not include a separate replication cohort. However, this is the first genome-wide study to examine a host genetic basis for sex-specific differences in spontaneous HCV clearance and uses the largest group of genotyped HCV-infected individuals. This limitation is also offset by the careful characterization and homogeneity of hepatitis C infection, the large sample size of both men and women, and the diverse ancestry of the cohorts. Another potential limitation is the inclusion of individuals with HIV-HCV–coinfection, which commonly occurs owing to routes of transmission. However, by including HIV infection status as a covariate in the models, the potential confounding effect of HIV infection in the observed sex-specific association is decreased. We also did not have any information about hepatitis B infection status or specific age at infection, so we were not able to evaluate the effect of these SNPs with hepatitis B virus coinfection or age.

In summary, we identified new genetic associations that support mechanisms of spontaneous HCV clearance that differ between males and females. These associations are hypothesis generating, and further studies that include men and women in other populations of different ancestral background are warranted. In addition, studies that translate these results to functional models in vitro may contribute to the identification of biological mechanisms of clearance of HCV infection that are sex specific.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. Acquisition and preparation of data and samples: C. L. T., J. J. G., A. M., V. P., E. J., A. H. K., T. R. O., S. H. M., G. D. K., A. Y. K., G. M. L., R. T. C., A. L. C., M. G. P., S. I. K., L. A., M. E. C., S. M. D., B. R. E., M. P. B., G. A., H. R. R., E. L. M., and D. L. T.. Statistical analyses: C. V., A. V., G. L. W., and P. D. Interpretation of data: C. V., A. V., C. L. T., G. L. W., M. A. T., D. L. T., and P. D. Study concept and design: C. V., C. L. T., D. L. T., and P. D. Manuscript writing: C. V., A. V., C. L. T.,

D. L. T., and P. D. All authors contributed to the critical revision of the manuscript for important intellectual content.

Financial support. This work was supported by the National Institutes of Drug Abuse (grants AI0148049, DA 033541, DA 04334, R01 DA12568, and U01DA036297), the National Institutes of Health (grant U01 AI131314, U19 AI066345, grant R01 HL076902 to the REVELL cohort, and grants R01 DA16159, R01 DA21550, and UL1 RR024996 to the Swann cohort); AIDS Research through the Center for Inherited Diseases at Johns Hopkins University; the National Institute of Allergy and Infectious Diseases (funding to the Multicenter AIDS Cohort Study [MACS]; grants U19 AI088791 and AI082630; and grants UO1 AI35004, UO1 AI31834, UO1 AI34994, UO1 AI34989, UO1 AI34993, and UO1 AI42590 to the Women's Interagency HIV Study [WIHS]); the Frederick National Laboratory for Cancer Research (contract HHSN261200800001E); the National Cancer Institute (grants UO1 AI35042, UL1 RR025005, UO1 AI35043, UO1 AI35039, UO1 AI35040, and UO1 AI35041 to the MACS and grant R01 HD41224 to the Hemophilia Growth and Development Study); the National Institute of Child Health and Human Development (grant UO1 HD32632 to the WIHS) .

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. Global hepatitis report 2017. License CC BY-NC-SA 3.0 IGO. Geneva, Switzerland: World Health Organization; **2017**.
2. Polaris Observatory HCV Collaborators. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. Lancet Gastroenterol Hepatol **2017**; 2:161–76.
3. Ikezaki H, Furusyo N, Hiramine S, et al. Association of IL28B rs8099917 genotype and female sex with spontaneous clearance of hepatitis C virus infection: a Japanese cross-sectional study. Arch Virol **2016**; 161:641–8.
4. Thomas DL, Astemborski J, Rai RM, et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. JAMA **2000**; 284:450–6.
5. Hajarizadeh B, Grebely J, Dore GJ. Epidemiology and natural history of HCV infection. Nat Rev Gastroenterol Hepatol **2013**; 10:553–62.
6. Micallef JM, Kaldor JM, Dore GJ. Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies. J Viral Hepat **2006**; 13:34–41.
7. Duggal P, Thio CL, Wojcik GL, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann Intern Med **2013**; 158:235–45.
8. Wojcik GL, Thio CL, Kao WH, et al. Admixture analysis of spontaneous hepatitis C virus clearance in individuals of African descent. Genes Immun **2014**; 15:241–6.
9. Vergara C, Thio CL, Johnson E, et al. Multi-ancestry genome-wide association study of spontaneous clearance of hepatitis C virus. Gastroenterology **2019**; 156:1496–1507. e7.
10. Chang D, Gao F, Slavney A, et al. Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. PLoS One **2014**; 9:e113684.
11. Gao F, Chang D, Biddanda A, et al. XWAS: A software toolset for genetic data analysis and association studies of the X chromosome. J Hered **2015**; 106:666–71.
12. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet **2012**; 44:955–9.
13. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. Nat Genet **2016**; 48:1284–7.
14. Sung YJ, Wang L, Rankinen T, Bouchard C, Rao DC. Performance of genotype imputations using data from the 1000 Genomes Project. Hum Hered **2012**; 73:18–25.
15. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science **2007**; 316:1341–5.
16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet **2006**; 38:904–9.
17. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PloS Genet **2006**; 2:e190.
18. 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**; 4:7.
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet **2007**; 81:559–75.
20. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol **2010**; 34:816–34.
21. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics **2010**; 11:288,2105-11-288.
22. White MJ, Risse-Adams O, Goddard P, et al. Novel genetic risk factors for asthma in African American children: precision medicine and the SAGE II Study. Immunogenetics **2016**; 68:391–400.
23. Mak ACY, White MJ, Eckalbar WL, et al; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. Whole-genome sequencing of pharmacogenetic drug

response in racially diverse children with asthma. Am J Respir Crit Care Med **2018**; 197:1552–64.
24. Verma A, Lucas A, Verma SS, 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**; 102:592–608.
25. Lee BD, Gonzalez S, Villa E, 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**; 174:683–90.
26. Sobota RS, Shriner D, Kodaman N, et al. Addressing population-specific multiple testing burdens in genetic association studies. Ann Hum Genet **2015**; 79:136–47.
27. Stouffer S, Suchman E, DeVinney L, Star S, Williams R. Adjustment during army life. Princeton, New Jersey: Princeton University Press; **1949**.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics **2010**; 26:2190–1.
29. Jons WA, Colby CL, McElroy SL, Frye MA, Biernacka JM, Winham SJ. Statistical methods for testing X chromosome variant associations: application to sex-specific characteristics of bipolar disorder. Biol Sex Differ **2019**; 10:57.
30. Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol Sex Differ **2015**; 6:35.
31. Rusinova I, Forster S, Yu S, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res **2013**; 41:D1040–6.
32. Dickensheets H, Sheikh F, Park O, Gao B, Donnelly RP. Interferon-lambda (IFN-λ) induces signal transduction and gene expression in human hepatocytes, but not in lymphocytes or monocytes. J Leukoc Biol **2013**; 93:377–85.
33. Quicke KM, Diamond MS, Suthar MS. Negative regulators of the RIG-I-like receptor signaling pathway. Eur J Immunol **2017**; 47:615–28.
34. Kitai Y, Takeuchi O, Kawasaki T, et al. Negative regulation of melanoma differentiation-associated gene 5 (MDA5)-dependent antiviral innate immune responses by Arf-like protein 5B. J Biol Chem **2015**; 290:1269–80.
35. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of genetic polymorphisms on human immune cell gene expression. Cell **2018**; 175:1701,1715.e16.
36. Hall PA, Russell SE. The pathobiology of the septin gene family. J Pathol **2004**; 204:489–505.
37. Hall PA, Jung K, Hillan KJ, Russell SE. Expression profiling the human septin gene family. J Pathol **2005**; 206:269–78.
38. Field CM, Kellogg D. Septins: cytoskeletal polymers or signalling GTPases? Trends Cell Biol **1999**; 9:387–94.
39. Kinoshita M. Assembly of mammalian septins. J Biochem **2003**; 134:491–6.
40. Kim CS, Seol SK, Song OK, Park JH, Jang SK. An RNA-binding protein, hnRNP A1, and a scaffold protein, septin 6, facilitate hepatitis C virus replication. J Virol **2007**; 81:3852–65.
41. Ngo HT, Pham LV, Kim JW, Lim YS, Hwang SB. Modulation of mitogen-activated protein kinase-activated protein kinase 3 by hepatitis C virus core protein. J Virol **2013**; 87:5718–31.
42. Shirasago Y, Sekizuka T, Saito K, et al. Isolation and characterization of an Huh.7.5.1-derived cell clone highly permissive to hepatitis C virus. Jpn J Infect Dis **2015**; 68:81–8.
43. Huang KH, Wang CH, Lin CH, Kuo HP. NF-κB repressing factor downregulates basal expression and *Mycobacterium tuberculosis* induced IP-10 and IL-8 synthesis via interference with NF-κB in monocytes. J Biomed Sci **2014**; 21:71.
44. Huang KH, Wang CH, Lee KY, Lin SM, Lin CH, Kuo HP. NF-κB repressing factor inhibits chemokine synthesis by peripheral blood mononuclear cells and alveolar macrophages in active pulmonary tuberculosis. PloS One **2013**; 8:e77789.
45. Dash B, Shapiro MJ, Chung JY, Romero Arocha S, Shapiro VS. Treg-specific deletion of NKAP results in severe, systemic autoimmunity due to peripheral loss of Tregs. J Autoimmun **2018**; 89:139–48.