The American Journal of Human Genetics, Volume *109*

Supplemental information

Trans-ancestral fine-mapping of MHC reveals key amino acids associated with spontaneous clearance of hepatitis C in HLA-DQβ1

Ana Valencia, Candelaria Vergara, Chloe L. Thio, Nicolas Vince, Venceslas Douillard, Alba Grifoni, Andrea L. Cox, Eric O. Johnson, Alex H. Kral, James J. Goedert, Alessandra Mangia, Valeria Piazzolla, Shruti H. Mehta, Gregory D. Kirk, Arthur Y. Kim, Georg M. Lauer, Raymond T. Chung, Jennifer C. Price, Salim I. Khakoo, Laurent Alric, Matthew E. Cramp, Sharyne M. Donfield, Brian R. Edlin, Michael P. Busch, Graeme Alexander, Hugo R. Rosen, Edward L. Murphy, Genevieve L. Wojcik, Mary Carrington, Pierre-Antoine Gourraud, Alessandro Sette, David L. Thomas, and Priya Duggal

Trans-Ancestral Fine-Mapping of MHC Reveals Key Amino Acids Associated with Spontaneous Clearance of Hepatitis C in HLA-DQβ1

Supplemental Materials and Methods

Tables and Figures

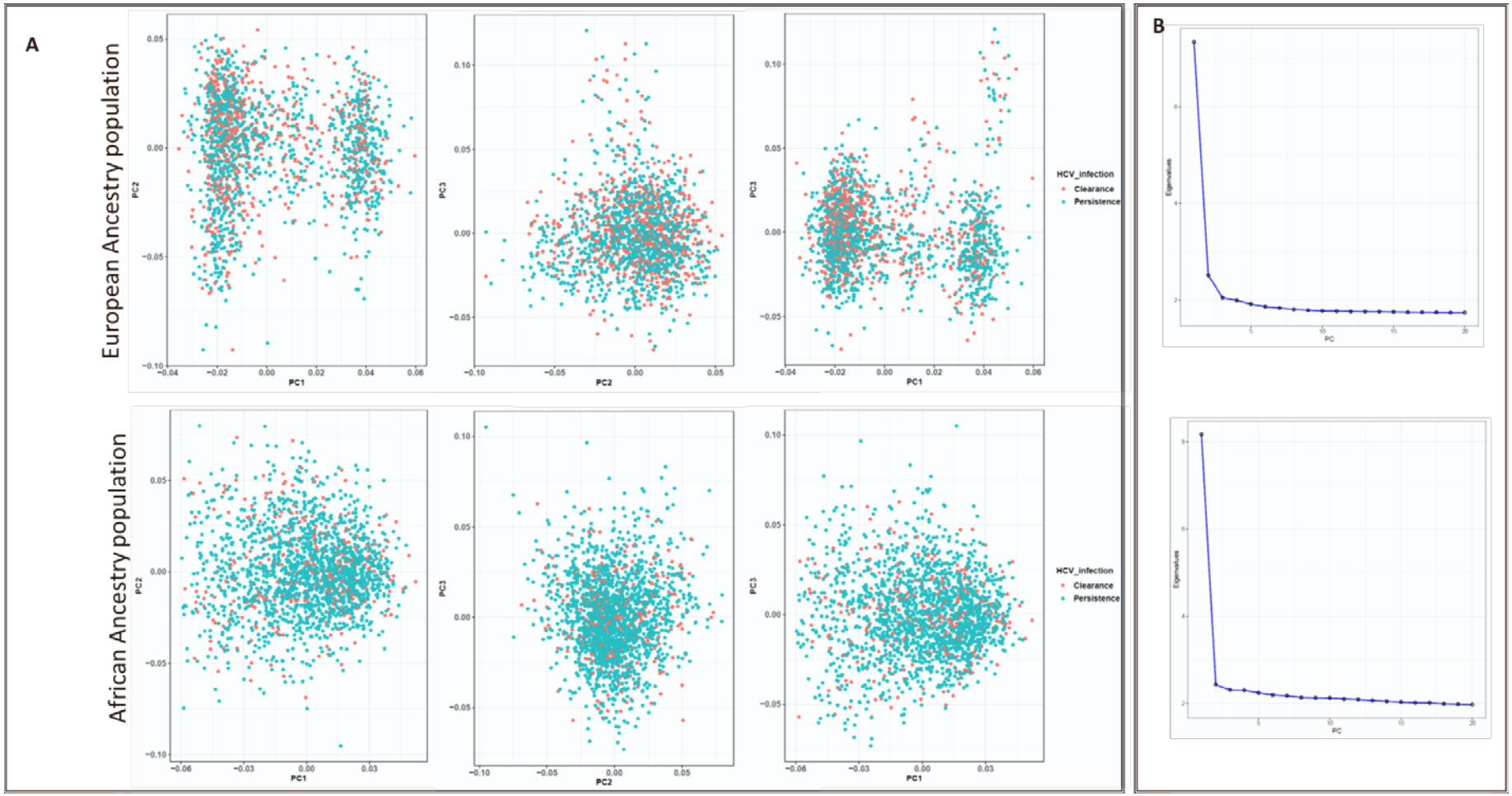

Figure S1: Principal components analysis of the European and African ancestry populations. (A) Scatterplot of PC1, PC2 and PC3. (B) Eigenvalues of each principal component. PC1 and PC2 explains most of the variance in the data for the European and African ancestry populations.

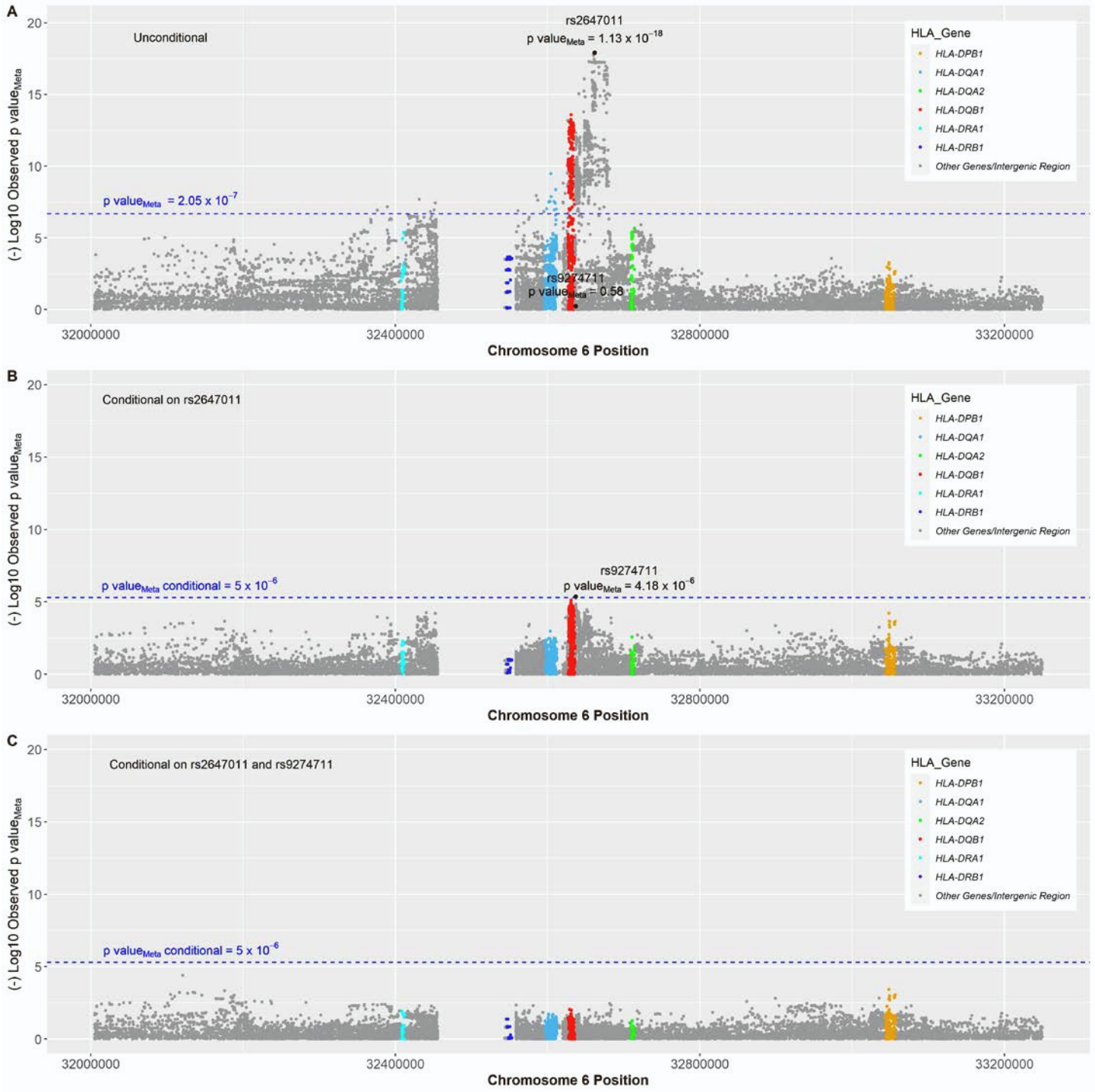
A
Unconditional
rs2647011
p value$_{Meta}$ = 1.13 x 10^{-18}
p value$_{Meta}$ = 2.05 x 10^{-7}
rs9274711
p value$_{Meta}$ = 0.58
(-) Log10 Observed p value$_{Meta}$
Chromosome 6 Position
HLA_Gene
HLA-DPB1
HLA-DQA1
HLA-DQA2
HLA-DQB1
HLA-DRA1
HLA-DRB1
Other Genes/Intergenic Region
B
Conditional on rs2647011
rs9274711
p value$_{Meta}$ = 4.18 x 10^{-6}
p value$_{Meta}$ conditional = 5 x 10^{-6}
C
Conditional on rs2647011 and rs9274711
p value$_{Meta}$ conditional = 5 x 10^{-6}

Figure S2. Results of the meta-analysis of the association of SNPs in the MHC class II region with spontaneous clearance of HCV infection in European and African ancestry in the unconditional analysis and after several steps of conditional analysis on the top SNPs. (A) Association results of the unconditional analysis. (B) Conditional analysis on rs2647011. (C) Conditional analysis on rs2647011 and rs9274711. After conditioning on rs2647011 and rs9274711 no SNPs showed significant association with HCV spontaneous clearance. Each point corresponds to the p $value_{Meta}$ for each SNP. The blue line and the p $value_{Meta}$ colored in blue represent the threshold level of fixed effects meta-analysis significance corrected by multiple comparisons in the unconditional and conditional analyses.

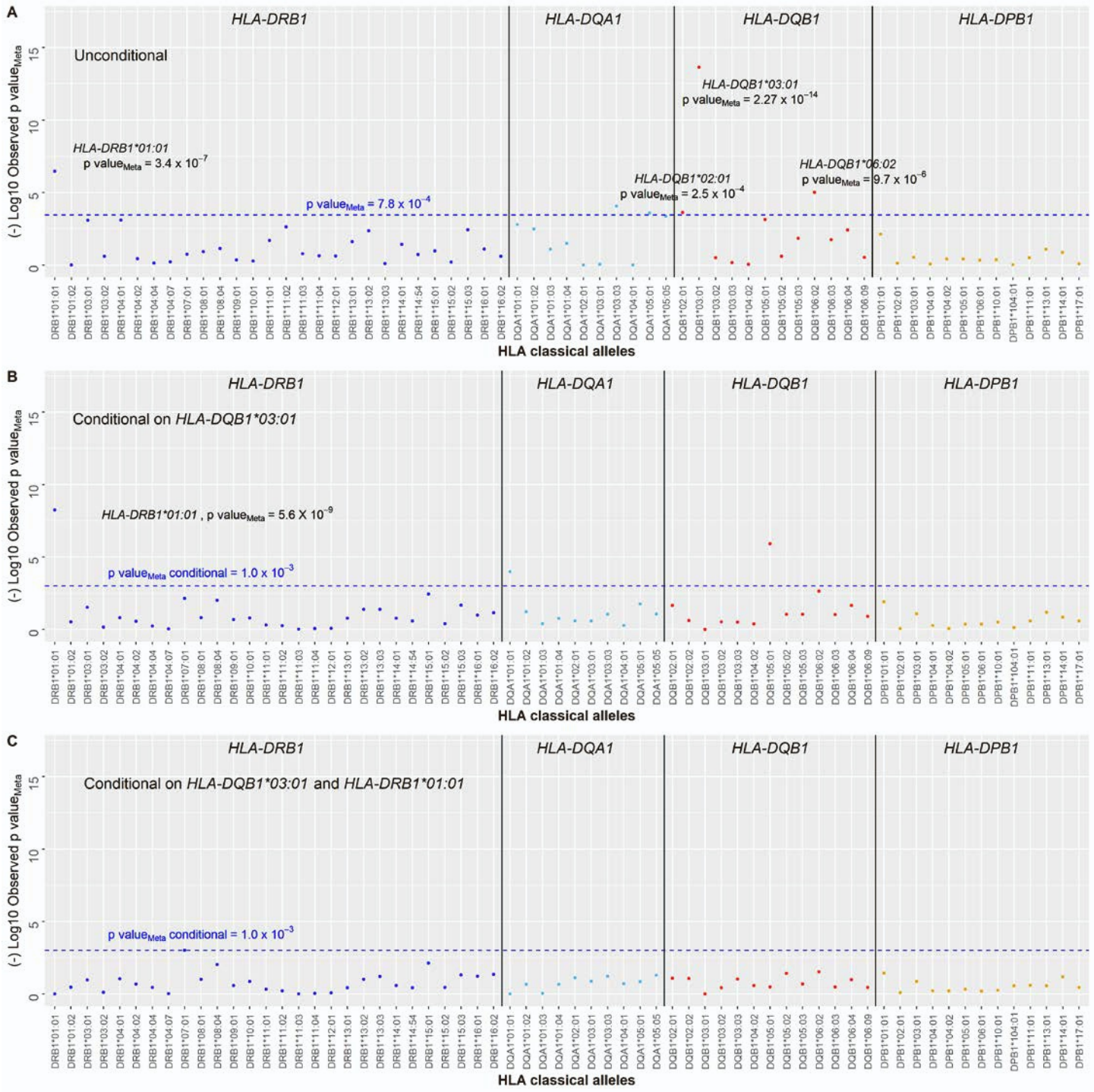
A
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Unconditional
HLA-DQB1*03:01
p value_Meta = 2.27 x 10^-14
HLA-DRB1*01:01
p value_Meta = 3.4 x 10^-7
HLA-DQB1*02:01
p value_Meta = 2.5 x 10^-4
HLA-DQB1*06:02
p value_Meta = 9.7 x 10^-6
p value_Meta = 7.8 x 10^-4
(-) Log10 Observed p value_Meta
HLA classical alleles
B
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Conditional on HLA-DQB1*03:01
HLA-DRB1*01:01 , p value_Meta = 5.6 X 10^-9
p value_Meta conditional = 1.0 x 10^-3
(-) Log10 Observed p value_Meta
HLA classical alleles
C
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Conditional on HLA-DQB1*03:01 and HLA-DRB1*01:01
p value_Meta conditional = 1.0 x 10^-3
(-) Log10 Observed p value_Meta
HLA classical alleles

Figure S3. Results of the meta-analysis of the association of HLA class II alleles with spontaneous clearance of HCV infection in European and African ancestry in the unconditional analysis and after several steps of conditioning on top HLA classical alleles. (A) Association results of the unconditional analysis. (B) Conditional analysis on *HLA-DQB1*03:01*. (C) Conditional analysis on *HLA-DQB1*03:01* and *HLA-DRB1*01:01*. After conditioning on *HLA-DQB1*03:01* and *HLA-DRB1*01:01* no HLA alleles showed significant association with HCV spontaneous clearance. Each point corresponds to the p $value_{Meta}$ for each HLA classical allele. The blue line and the p $value_{Meta}$ colored in blue represent the threshold level of fixed effects meta-analysis significance corrected by multiple comparisons in the unconditional and conditional analyses.

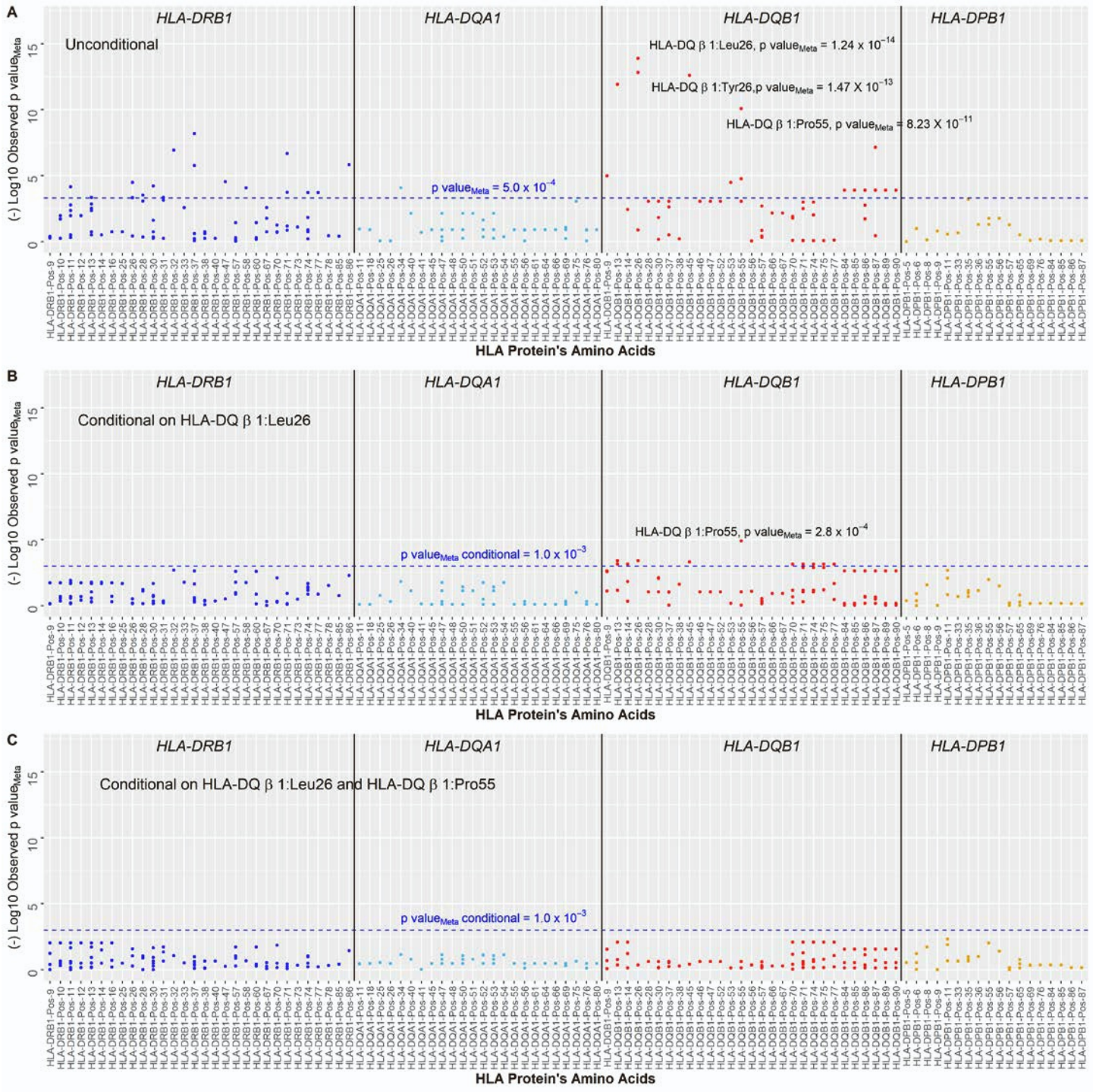
A
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Unconditional
HLA-DQ β 1:Leu26, p value$_{Meta}$ = 1.24 x 10^{-14}
HLA-DQ β 1:Tyr26,p value$_{Meta}$ = 1.47 X 10^{-13}
HLA-DQ β 1:Pro55, p value$_{Meta}$ = 8.23 X 10^{-11}
p value$_{Meta}$ = 5.0 x 10^{-4}
(-) Log10 Observed p value$_{Meta}$
HLA Protein's Amino Acids
B
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Conditional on HLA-DQ β 1:Leu26
HLA-DQ β 1:Pro55, p value$_{Meta}$ = 2.8 x 10^{-4}
p value$_{Meta}$ conditional = 1.0 x 10^{-3}
(-) Log10 Observed p value$_{Meta}$
HLA Protein's Amino Acids
C
HLA-DRB1
HLA-DQA1
HLA-DQB1
HLA-DPB1
Conditional on HLA-DQ β 1:Leu26 and HLA-DQ β 1:Pro55
p value$_{Meta}$ conditional = 1.0 x 10^{-3}
(-) Log10 Observed p value$_{Meta}$
HLA Protein's Amino Acids

Figure S4. Results of the meta-analysis of the association of amino acid residues in HLA class II proteins with HCV spontaneous clearance in European and African ancestry populations in the unconditional analysis and after several steps of conditioning on top amino acid residues. (A) Association results of the unconditional analysis. (B) Conditional analysis on HLA-DQβ1Leu26. (C) Conditional analysis on HLA-DQβ1Leu26 and HLA-DQβ1Pro55. After conditioning on HLA-DQβ1Leu26 and HLA-DQβ1Pro55 no amino acid residue showed significant association with HCV spontaneous clearance. Each point corresponds to the p $value_{Meta}$ for each amino acid. The blue line and the p $value_{Meta}$ colored in blue represent the threshold level of fixed effects meta-analysis significance corrected by multiple comparisons in the unconditional and conditional analyses.

	Imputed HLA class II genes			
European ancestry (N=1,600)	***HLA-DRB1***	***HLA-DQA1***	***HLA-DQB1***	***HLA-DPB1***
No. of individuals with imputed HLA alleles after quality control procedures (% from the total of imputable genotypes)	1,440 (90)	1,582 (98.9)	1,600 (100)	1,543 (96.4)
HCV persistent infection (controls)	912	997	1006	967
HCV spontaneous clearance (cases)	528	585	594	576
Number of unique HLA alleles to test	19	10	12	13
Number of amino acid positions/number of residues or combinations tested	28/63	24/39	29/74	18/29
African ancestry (N=1,864)				
No. of individuals with imputed HLA alleles after quality control procedures (% from the total of imputable genotypes)	1,429 (76.7)	0	1,792 (96.2)	0
HCV persistent infection (controls)	1,179	0	1,467	0
HCV spontaneous clearance (cases)	254	0	330	0
Number of unique HLA alleles to test	20	0	13	0
Number of amino acid positions/number of residues or combinations tested	30/84	0	29/71	0

Table S1. Efficiency and quality of imputation of HLA class II classical alleles and amino acids residues per ancestry group.

Table S2 (Excel file)

Table S2. SNPs significantly associated with HCV spontaneous clearance in the MHC class II region. Abbreviations: Chr: Chromosome; BP: position in base pairs; OR: odds ratio for HCV spontaneous clearance; CI: Confidence Interval; p value$_{Meta}$: p value of the meta-analysis of European and African ancestry populations.

HLA Gene	Effect Allele	European Ancestry (N=1,600)				African Ancestry (N=1,869)				p value$_{Meta}$
		Frequency Effect Allele		OR (95%CI)	p value	Frequency Effect Allele		OR (95%CI)	p value	
		Clearance (n=594)	Persistence (n=1,006)			Clearance (n=340)	Persistence (n=1,529)			
DRB1	*DRB1*01:01*	0.10	0.06	1.68 (1.3-2.2)	2.0×10^{-4}	0.06	0.03	2.34 (1.5-3.8)	4.7×10^{-4}	3.4×10^{-7}
DQA1	*DQA1*03:03*	0.08	0.04	1.85 (1.4-2.5)	8.5×10^{-5}	NA	NA	NA	NA	8.5×10^{-5}
	*DQA1*05:01*	0.09	0.13	0.61 (0.5-0.8)	2.6×10^{-4}	NA	NA	NA	NA	2.6×10^{-4}
	*DQA1*05:05*	0.19	0.16	1.41(1.2-1.7)	4.4×10^{-4}	NA	NA	NA	NA	4.4×10^{-4}
DQB1	*DQB1*02:01*	0.09	0.13	0.60 (0.5-0.8)	1.9×10^{-4}	0.06	0.07	0.76 (0.5-1.1)	1.4×10^{-1}	2.5×10^{-4}
	*DQB1*03:01*	0.27	0.20	1.61 (1.3-1.9)	7.6×10^{-8}	0.27	0.18	1.75 (1.4-2.1)	5.9×10^{-8}	2.3×10^{-14}
	*DQB1*05:01*	0.13	0.10	1.39 (1.1-1.7)	4.2×10^{-3}	0.13	0.1	1.24 (1.0-1.6)	5.5×10^{-2}	7.3×10^{-4}
	*DQB1*06:02*	0.10	0.12	0.71 (0.6-0.9)	3.4×10^{-3}	0.15	0.21	0.7 (0.5-0.9)	8.9×10^{-4}	9.7×10^{-6}

Table S3. HLA alleles class II significantly associated with HCV spontaneous clearance in European and African ancestry populations. Abbreviations: OR: odds ratio for HCV spontaneous clearance; CI: Confidence Interval; p value$_{Meta}$: p value of the meta-analysis of European and African ancestry populations; NA: Not available.

Table S4 (excel file)

Table S4. Results of the association analysis and meta-analysis of amino acids residues on HLA class II genes with spontaneous clearance of HCV infection in European and African ancestry groups. Amino acids are named in a single-letter code, those with asterisk (*) are significantly associated with HCV spontaneous clearance. Abbreviations: OR: odds ratio; CI: confidence interval; p $value_{Meta}$: p value of the meta-analysis of both ancestry groups; NA: Not available. Single letter and three letter amino acid code: A:Ala, C:Cys, D:Asp, E:Glu, F:Phe, G:Gly, H:His, I:Ile, K:Lys, L:Leu, M:Met, N:Asn, P:Pro, Q:Gln, R:Arg, T:Thr, S:Ser, Y:Tyr, V:Val, W:Trp.

Marker/Gene (Effect Allele)	HIV infection status	European Ancestry n (HIV -) = 1,342 n (HIV +) = 258		African Ancestry n (HIV -) = 1,142 n (HIV +) = 722		p value$_{Meta}$
		OR	p value	OR	p value	
rs2647011 (A)	HIV (-)	1.23	2.89×10^{-8}	1.26	2.06×10^{-5}	3.28×10^{-12}
	HIV (+)	1.43	1.30×10^{-4}	1.31	2.78×10^{-5}	2.69×10^{-8}
rs9274711 (T)	HIV (-)	0.97	7.39×10^{-1}	0.88	3.74×10^{-1}	3.96×10^{-1}
	HIV (+)	1.08	7.16×10^{-1}	1.02	9.12×10^{-1}	7.79×10^{-1}
*DQB1(DQB1*03:01)*	HIV (-)	1.52	3.03×10^{-5}	1.90	1.95×10^{-6}	3.23×10^{-10}
	HIV (+)	1.96	2.94×10^{-3}	1.61	3.49×10^{-3}	5.36×10^{-5}
*DQB1(DQB1*06:02)*	HIV (-)	0.70	7.18×10^{-3}	0.73	3.76×10^{-2}	7.06×10^{-4}
	HIV (+)	0.54	6.77×10^{-2}	0.60	5.70×10^{-3}	9.28×10^{-4}
*DRB1 (DRB1*01:01)*	HIV (-)	1.65	1.43×10^{-3}	1.77	9.41×10^{-2}	4.42×10^{-4}
	HIV (+)	1.69	1.46×10^{-1}	3.02	1.92×10^{-3}	6.82×10^{-4}
HLA-DQβ1 Leu26	HIV (-)	0.84	1.45×10^{-5}	0.57	5.53×10^{-5}	3.24×10^{-9}
	HIV (+)	0.75	3.27×10^{-3}	0.52	8.10×10^{-5}	9.18×10^{-7}
HLA-DQβ1 Pro55	HIV (-)	1.42	1.86×10^{-4}	1.64	4.50×10^{-4}	3.00×10^{-7}
	HIV (+)	1.77	6.06×10^{-3}	1.57	1.21×10^{-2}	3.15×10^{-4}

Table S5. Unconditional association analysis of independently associated SNPs, MHC classical alleles and amino acids by HIV infection status. Abbreviations: OR: odds ratio; p value$_{Meta}$: p value of the meta-analysis of both ancestry groups.

Table S6 (excel file)

Table S6. Results of the unconditional association analysis and fixed effect meta-analysis of the top associated SNPs, HLA class II classical alleles and amino acid residues and after conditioning with the same type of variant and across different type of variants. Abbreviations: p $value_{Meta}$: p value of the meta-analysis of both ancestry groups. Single letter and three letter amino acid code: A:Ala, C:Cys, D:Asp, E:Glu, F:Phe, G:Gly, H:His, I:Ile, K:Lys, L:Leu, M:Met, N:Asn, P:Pro, Q:Gln, R:Arg, T:Thr, S:Ser, Y:Tyr, V:Val, W:Trp.

Table S7 (excel file)

Table S7. Results of the unconditional association analysis and fixed effect meta-analysis of SNPs, HLA alleles and amino acid residues in the MHC class II and after conditioning with amino acids associated with high significance. Abbreviations: p $value_{Meta}$: p value of the meta-analysis of both ancestry groups. Single letter and three letter amino acid code: A:Ala, C:Cys, D:Asp, E:Glu, F:Phe, G:Gly, H:His, I:Ile, K:Lys, L:Leu, M:Met, N:Asn, P:Pro, Q:Gln, R:Arg, T:Thr, S:Ser, Y:Tyr, V:Val, W:Trp.

Table S8 (excel file)

Table S8. List of peptides spanning the HCV proteome and their relative IC_{50} predicted values for each HLA class II alleles associated with viral clearance or persistence. Peptide sequences are listed in combination with localization, derived protein and predicted IC_{50} value for each of the HLA class II analyzed. Peptides are defined as epitopes if the IC_{50} values was $\leq$ 1000nM in at least one of the HLA class II alleles analyzed. The ratio is calculated only on defined epitopes considering the geometric mean of the IC_{50} values calculated in persistence alleles versus clearance alleles. The higher the ratio, the stronger the binding affinity in clearance HLA alleles versus persistence HLA alleles. Single letter and three letter amino acid code: A:Ala, C:Cys, D:Asp, E:Glu, F:Phe, G:Gly, H:His, I:Ile, K:Lys, L:Leu, M:Met, N:Asn, P:Pro, Q:Gln, R:Arg, T:Thr, S:Ser, Y:Tyr, V:Val, W:Trp.

Gene	All			African Ancestry (N=18)			European Ancestry (N=13)		
	Total alleles (N)	**Concordant alleles (N)**	**Accuracy (%)**	**Total alleles (N)**	**Concordant alleles (N)**	**Accuracy (%)**	**Total alleles (N)**	**Concordant alleles (N)**	**Accuracy (%)**
HLA-DRB1	42	41	97.6	26	25	96.2	16	16	100.0
HLA-DQA1	22	22	100.0	NA	NA	NA	22	22	100.0
HLA-DQB1	50	49	98.0	32	31	96.9	18	18	100.0
HLA-DPB1	24	23	98.0	NA	NA	NA	24	23	95.8

Table S9. Results of accuracy of the imputation with HIBAG and HLA-HD in all individuals and in individuals of European and African ancestry.

Materials and Methods

Samples

Study Cohorts

Boston Area HCV Study: Transmission, Immunity, Outcomes Network (BAHSTION). Subjects were enrolled in Massachusetts General Hospital, Brigham and Women's Hospital and Lemuel Shattuck Hospital (USA). HCV spontaneous clearance was determined by 2 negative tests separated by at least six months, and no treatment was provided prior to case status determination.

Cramp et al. These individuals were recruited from patients seen at the South West Liver Unit and at King's College Hospital (London). HCV spontaneous clearance was determined by being both HCV antibody positive and HCV RNA seronegative over a period of at least 18 months with no treatment. HCV persistence was determined by both HCV RNA and HCV antibody positive. The clearance and persistence groups were comparable in terms of age, sex, estimated duration of infection and route of infection[1].

Hemophilia Growth and Development Study (HGDS). The HGDS is a multicenter natural history study of adolescents and children with hemophilia and its complications. HCV spontaneous clearance was defined as HCV antibody positive and HCV RNA non-detectable for two or more study measurements separated by six months. HCV persistence was defined as HCV antibody positive and HCV RNA above the limit of detection for two or more study measurements. HCV treatments given prior to the case determination were not considered[2].

Mangia et al. This cohort corresponds to HCV mono-infected patients with HCV positive antibody test followed up at the Research laboratory of IRCCS "Casa Sollievo della Sofferenza" (Italy). Determination of HCV spontaneous clearance was made by repeated HCV RNA assays

performed every 6 months during an observational period of 18 months with no treatment The original study was approved by the Ethic Committee of the above IRCCS in Italy[3].

AIDS Link to the Intravenous Experience (ALIVE). The ALIVE cohort is based in Baltimore (USA) and includes individuals who inject drugs with follow up visits every six months. HCV spontaneous clearance was determined by has having two serum samples in which HCV RNA was <50 IU/ml and having HCV antibody positive; subjects with HCV antibodies and RNA in serum were considered as persistently infected. No HCV treatments were given prior to the case determination. Approval was granted by the Johns Hopkins Bloomberg School of Public Health[4].

Baltimore Before and After Acute Study of Hepatitis (BBAASH). This study is based in Baltimore (USA) and comprises a prospective cohort of people who inject drugs and at risk of HCV infection[5]. In this study, HCV infection was confirmed by HCV antibody seroconversion for all subjects. HCV spontaneous clearance was defined as the absence of detectable viremia in at least two serum or plasma specimens obtained >300 days after initial viremia and >60 days apart. HCV persistence was defined as continuous viremia <300 days after initial viremia. No subjects were included who underwent therapy prior to determination of HCV outcome. The study was approved by the Institutional Review Board at the Johns Hopkins School of Medicine[5].

Multicenter Hemophilia Cohort Studies (MHCS I and II). The MHCS consists of hemophilic subjects (mostly with hemophilia A) recruited in 16 sites located in the United States, Greece, Germany, and Austria with a high percentage of HCV positive individuals (70%). The Second Multicenter Hemophilia Cohort Study (MHCS-II) includes subjects >13 years of age with an inherited coagulation disorder such as hemophilia A or B, deficiencies in other factors such as V or XI, and von Willebrand who were exposed to HCV[6].

Correlates of Resolved Versus Low-Level Viremic Hepatitis C Infection in Blood Donors study (REVELL). This cohort is a case control study of serologically confirmed HCV positive blood donors identified within a network of 17 blood banks in Western and Southern USA. HCV persistence was defined as being HCV antibody positive and RNA+ both at an index seropositive donation and on 1-2 follow-up samples collected 1-3 years later[7]. HCV spontaneous clearance was determined by negative test by replicate. No HCV therapy was administered while determining the diagnosis. The Committee on Human Research at the University of California, San Francisco, approved the study protocol[7].

The Swan Project. This study recruits persons who inject illicit drugs at a community-based field site on the Lower East Side of Manhattan (USA). The determination of HCV spontaneous clearance/persistence was made with a single test using the discriminatory HCV transcription mediated amplification assay (GenProbe Inc., San Diego, CA). No HCV treatments were given prior to case determination. The study is approved by the institutional review boards of SUNY Downstate College of Medicine and Weill Cornell Medical College. Written informed consent is obtained from all study participants[8].

Toulouse, France cohort. This study includes individuals born in the south of France with HCV spontaneous clearance/persistence determined by antibody and RNA determinations as described extensively in reference[9].

Women's Interagency HIV Study (WIHS). This is a multicenter, collaborative, prospective study for women at risk for HIV infection with several recruiting centers at USA including New York City/Bronx, Brooklyn, NY; Washington DC, Northern California; Los Angeles County/Southern California; Chicago. The study includes HIV+, HIV- women and those previously diagnosed with clinical AIDS or low CD4+ cell counts. HCV spontaneous clearance

was defined as one HCV RNA at entry into WIHS with positive HCV antibodies. No HCV treatments were given prior to case determination[10].

United Kingdom Drug Use cohort. This cohort is based in London at a Hepatology clinic at the Kings College. Subjects identified with HCV spontaneous clearance had two negative PCR reactions at least six months apart and those with HCV persistence had at least two positive PCR reactions. No treatment was given prior to this assessment. Most individuals are self-reported Caucasian, HIV negative, and injection drug users[11].

Urban Health Study. This is a serial, cross-sectional, sero-epidemiological study of injection drug users in the San Francisco Bay Area (USA). HCV spontaneous clearance or persistence was determined based on antibodies and Hepatitis C Viral RNA test was done on those individuals with positive antibody tests. The present analysis included mostly genetically determined African Ancestry individuals[12, 13].

Genotyping, imputation of SNPs, classical HLA alleles and amino acids

The reference population to build the pre-fit classifiers for the European ancestry population used in the software HIBAG included between 1,624 and 2,572 individuals (depending on the HLA locus) from diverse, though mainly European, or European-ancestry individuals from HLARES database obtained from GlaxoSmithKline clinical trials as described in reference 14. SNP from the MHC region in the reference population were typed using the Illumina 1M and 1M Duo platform. 7,976 SNPs with less than 10% missing data were used to build the classifiers. Classical HLA alleles data was generated by Conexio Genomics, HistoGenetics (Ossining, NY, USA) and LabCorp (Burlington, NC, USA) using the SBT, SSO and SSP methodologies for *HLA-A, B, C, DRB1, DQA1, DQB1* and *DPB1* genes[14].

Classifiers used for the imputation of classical HLA alleles in the African ancestry population were built based on whole genome sequencing data from 880 participating in the "Consortium on Asthma among African-ancestry Populations in the Americas" (CAAPA)[15] with SNP genotyped in the MHC region and typed for classical *HLA* alleles such as *HLA-A, HLA-B, HLA-C, HLA-DRB1 and HLA-DQB1*[16]. As input for the HIBAG algorithm we selected independent SNPs in this reference population determined by pruning out the markers having a linkage disequilibrium (LD) r^2 value > 0.8. After pruning, we obtained 24,504 SNPs with a genotyping rate > 98% and minor allele frequency > 0.001 to use for the construction of the classifiers. On the other hand, to select SNP of our African ancestry group to input in HIBAG, we used genotyped SNPs and imputed SNPs. Allelic dosages of the imputed SNPs were transformed to hard calls using PLINK 2.00 alpha version[17, 18] and imputed markers with uncertainty greater than 0.1 were treated as missing. Out of the 25,565 imputed/genotyped SNPs, 14,620 were in the intersection with the reference panel and were used as input for building the classifiers. For each population we imputed classical HLA alleles and translated to amino acids in cases and controls together.

Statistical analysis

Association analyses and meta-analysis of each type of variants

For the association analysis we encoded each HLA classical allele and amino acid residue as a bi-allelic marker, corresponding to the presence or absence of each allele and for multi allelic amino acid positions we also generated a variable for each possible combination of the residues for that position using HIBAG[14].

Identification of independent signals for each type of variants

In the model of the conditional analyses for each type of variant, we generated dummy variables for each classical HLA allele and each residue in amino acid positions to which we added the top associated variants of each type one by one in an additive logistic regression model including HIV infection status as well as the first two principal components after PCA. We then evaluated the effect of this conditioning on the association of the other variants of the same type. Data formatting of the classical HLA alleles and amino acids was implemented with customized scripts in R[19] and conditional association analyses were performed using PLINK 2.00 alpha version[17, 18].

Epitope prediction of associated classical HLA class II alleles containing causal amino acids

In order to determine pairs of *HLA-DQA1/HLA-DQB1* classical HLA alleles to include in this analysis, we estimated haplotypes of these 2 genes in the European ancestry population using a maximum likelihood estimate of haplotype probabilities using the haplo.stats package (https://www.mayo.edu/research/labs/statistical-genetics-genetic-epidemiology/software) implemented in R[19]. Haplotypes with highest frequencies containing classical HLA alleles with positive and negative effect in HCV spontaneous clearance were included in the analysis. Predicted epitopes associated with clearance or persistence have been selected based on the ratio calculated as following, where *p* are the persistence alleles and *c* are the clearance alleles:

$$Geomean\ Ratio = \frac{\sqrt[pn]{Allele\ p1 * Allele\ p2 * Allele\ pn}}{\sqrt[cn]{Allele\ c1 * Allele\ c2 * Allele\ cn}}$$

Results

Unconditional and conditional association analyses and meta-analyses

Unconditional and conditional association analysis with amino acids residues.

Besides HLA-DQβ1Leu26 and HLA-DQβ1Tyr26, other amino acids were also associated with HCV spontaneous clearance with similar significance, including HLA-DQβ1Glu45 (p $value_{Meta}$= 2.4 x 10^{-13}), HLA-DQβ1Ala13 (p $value_{Meta}$= 1.18 x 10^{-12}). HLA-DQβ1Leu53, HLA-DQβ1Pro55 and HLA-DQβ1Phe87 were also associated with less strength of association (Table S4). Their association was notably diminished in the conditional analysis with HLA-DQβ1Leu26 excepting HLA-DQβ1Pro55 and they were not significantly associated after double conditional analysis with HLA-DQβ1Leu26 and HLA-DQβ1Pro55 (Figure S4).

Conditional analysis on associated residues in amino acid positions 13, 45, 53, 55 and 87 did not have a major effect on the association of HLA-DQβ1Leu26. Only HLA-DQβ1Phe9 was able to eliminate the significance of the association of HLA-DQβ1Leu26 (Table S7).

Amino acid residues explain SNP and HLA classical allele associations with HCV clearance

The individual conditional models for other amino acid residues associated with high significance are presented in Table S7. The change in the association of rs2647011 and rs9274711 observed when conditioning with other amino acids was not as significant as it was in the conditional model with HLA-DQβ1Leu26 (Tables S6 and S7). The association of *HLA-DQB1*03:01* was eliminated in the conditional analysis of HLA-DQβ1Ala13 and HLA-DQβ1Glu45 and decreased with HLA-DQβ1 Tyr9 and HLA-DQβ1Pro55 (Table S7).

The association with the classical allele *HLA-DQB1*06:02* was abrogated when conditioning in each of the associated amino acid positions: HLA-DQβ1Tyr9, HLA-DQβ1Ala13, HLA-DQβ1Glu45, HLA-DQβ1Pro55 and HLA-DQβ1Phe87. The association of *HLA-DRB1*01:01* was only reduced with conditional analysis with amino acid residues in HLA-DRβ1 but was widely reduced when conditioning in HLA-DQβ1Leu26 (Table S7). The association of

*HLA-DQB1*02:01* was reduced in the conditional analysis with amino acid residues in positions 13, 26, 45, 55 of HLA-DQβ1.

Epitope prediction of classical HLA class II alleles associated with HCV clearance.

The list of peptide sequences used for the prediction analyses with the IC_{50} prediction values for each HLA class II allele analyzed as well as additional information related to the location, specific protein and ratio between persistence/clearance alleles calculated on peptides predicted to be epitopes in at least one of the alleles studied is presented in Table S8.

References

1. Kim, A.Y., Kuntzen, T., Timm, J., Nolan, B.E., Baca, M.A., Reyor, L.L., Berical, A.C., Feller, A.J., Johnson, K.L., Schulze zur Wiesch, J. et al. (2011). Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. Gastroenterology 140, 686-696.e1.

2. Hilgartner, M.W., Donfield, S.M., Willoughby, A., Contant, C.F.,Jr, Evatt, B.L., Gomperts, E.D., Hoots, W.K., Jason, J., Loveland, K.A., McKinlay, S.M. (1993). Hemophilia growth and development study. Design, methods, and entry data. Am. J. Pediatr. Hematol. Oncol. 15, 208-218.

3. Mangia, A., Gentile, R., Cascavilla, I., Margaglione, M., Villani, M.R., Stella, F., Modola, G., Agostiano, V., Gaudiano, C., Andriulli, A. (1999). HLA class II favors clearance of HCV infection and progression of the chronic liver damage. J. Hepatol. 30, 984-989.

4. Vlahov, D., Munoz, A., Anthony, J.C., Cohn, S., Celentano, D.D., Nelson, K.E. (1990). Association of drug injection patterns with antibody to human immunodeficiency virus type 1 among intravenous drug users in Baltimore, Maryland. Am. J. Epidemiol. 132, 847-856.

5. Cox, A.L., Netski, D.M., Mosbruger, T., Sherman, S.G., Strathdee, S., Ompad, D., Vlahov, D., Chien, D., Shyamala, V., Ray, S.C. et al. (2005). Prospective evaluation of community-acquired acute-phase hepatitis C virus infection. Clin. Infect. Dis. 40, 951-958.

6. Goedert, J.J., Chen, B.E., Preiss, L., Aledort, L.M., Rosenberg, P.S. (2007). Reconstruction of the hepatitis C virus epidemic in the US hemophilia population, 1940-1990. Am. J. Epidemiol. 165, 1443-1453.

7. Tobler, L.H., Bahrami, S.H., Kaidarova, Z., Pitina, L., Winkelman, V.K., Vanderpool, S.K., Guiltinan, A.M., Cooper, S., Busch, M.P., Murphy, E.L. (2010). A case-control study of factors

associated with resolution of hepatitis C viremia in former blood donors (CME). Transfusion 50, 1513-1523.

8. Edlin, B.R., Shu, M.A., Winkelstein, E., Des Jarlais, D.C., Busch, M.P., Rehermann, B., O'Brien, T.R., Talal, A.H., Tobler, L.H., Zeremski, M. et al. (2009). More rare birds, and the occasional swan. Gastroenterology 136, 2412-2414.

9. Alric, L., Fort, M., Izopet, J., Vinel, J.P., Charlet, J.P., Selves, J., Puel, J., Pascal, J.P., Duffaut, M., Abbal, M. (1997). Genes of the major histocompatibility complex class II influence the outcome of hepatitis C virus infection. Gastroenterology 113, 1675-1681.

10. Kuniholm, M.H., Gao, X., Xue, X., Kovacs, A., Marti, D., Thio, C.L., Peters, M.G., Greenblatt, R.M., Goedert, J.J., Cohen, M.H. et al. (2011). The relation of HLA genotype to hepatitis C viral load and markers of liver fibrosis in HIV-infected and HIV-uninfected women. J. Infect. Dis. 203, 1807-1814.

11. Khakoo, S.I., Thio, C.L., Martin, M.P., Brooks, C.R., Gao, X., Astemborski, J., Cheng, J., Goedert, J.J., Vlahov, D., Hilgartner, M. et al. (2004). HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. Science 305, 872-874.

12. Kral, A.H., Bluthenthal, R.N., Lorvick, J., Gee, L., Bacchetti, P., Edlin, B.R. (2001). Sexual transmission of HIV-1 among injection drug users in San Francisco, USA: risk-factor analysis. Lancet 357, 1397-1401.

13. Kral, A.H., Lorvick, J., Gee, L., Bacchetti, P., Rawal, B., Busch, M., Edlin, B.R. (2003). Trends in human immunodeficiency virus seroincidence among street-recruited injection drug users in San Francisco, 1987-1998. Am. J. Epidemiol. 157, 915-922.

14. Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R., Weir, B.S. (2014). HIBAG--HLA genotype imputation with attribute bagging. Pharmacogenomics J. 14, 192-200.

15. Daya, M., Rafaels, N., Brunetti, T.M., Chavan, S., Levin, A.M., Shetty, A., Gignoux, C.R., Boorgula, M.P., Wojcik, G., Campbell, M. et al. (2019). Association study in African-admixed populations across the Americas recapitulates asthma risk loci in non-African populations. Nat. Commun. 10, 880-7.

16. Vince, N., Limou, S., Daya, M., Morii, W., Rafaels, N., Geffard, E., Douillard, V., Walencik, A., Boorgula, M.P., Chavan, S. et al. (2020). Association of HLA-DRB1 *09:01 with tIgE levels among African-ancestry individuals with asthma. J. Allergy Clin. Immunol. 146, 147-155.

17. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7-8. eCollection 2015.

18. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559-575.

19. R Core Team. (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.