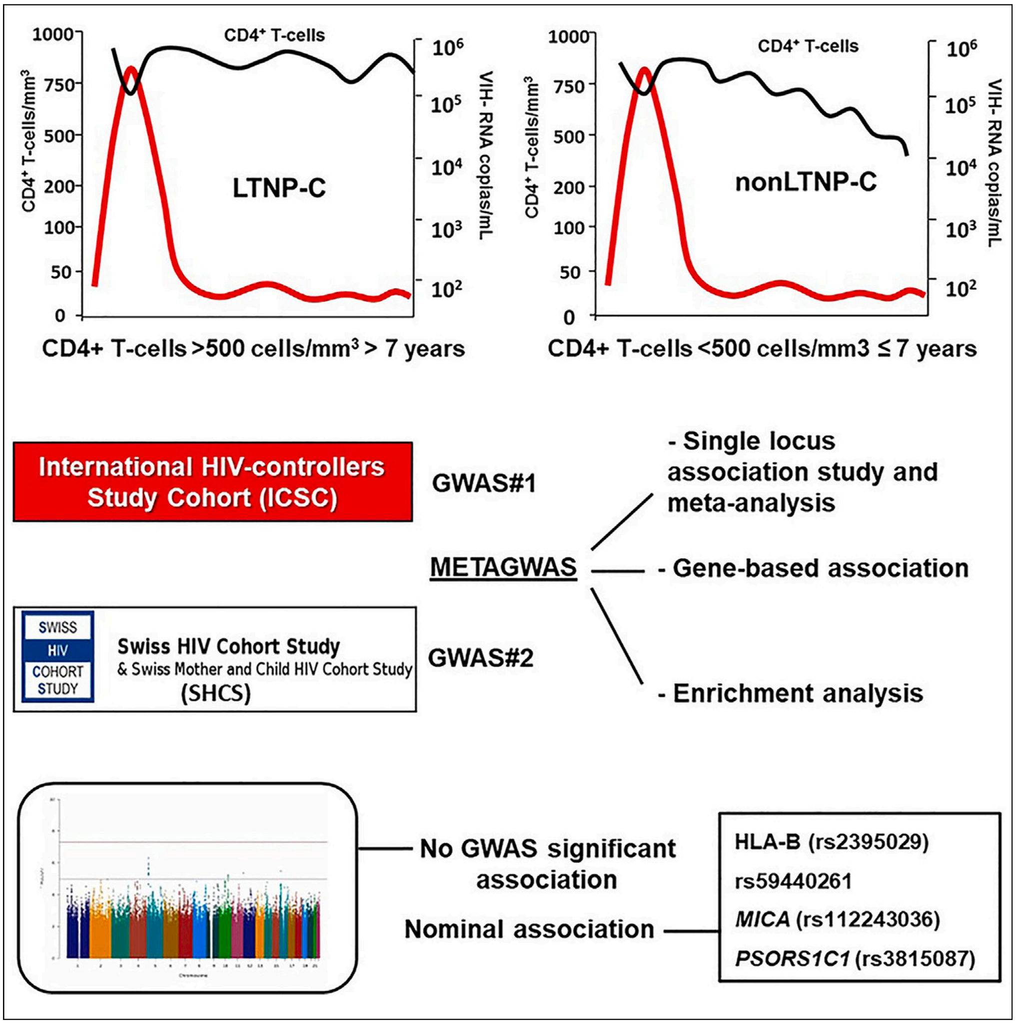
iScience

CellPress
OPEN ACCESS

Article

A metagenome-wide association study of HIV disease progression in HIV controllers



Luis Miguel Real, María E. Sáez, Anais Corma-Gómez, ..., Mathias Lichterfeld, Ezequiel Ruiz-Mateos, the Swiss HIV Cohort Study

eruizmateos-ibis@us.es

Highlights

No SNP or gene was associated with the long-term non-progressor HIV control phenotype

SNPs linked to LOC285696, *RMI2*, and chromosome 5 region showed suggestive association

Process related to metallopeptidase activity showed suggestive significance

SNPs previously associated with natural HIV control showed nominal association

Real et al., iScience *26*, 107214
July 21, 2023
https://doi.org/10.1016/j.isci.2023.107214

iScience

CellPress OPEN ACCESS

Article

A metagenome-wide association study of HIV disease progression in HIV controllers

Luis Miguel Real,[1,2,3,17] María E. Sáez,[4,17] Anais Corma-Gómez,[1,3] Antonio Gonzalez-Pérez,[4] Christian Thorball,[5,6] Rocío Ruiz,[7] María Reyes Jimenez-Leon,[8] Alejandro Gonzalez-Serna,[1,3] Carmen Gasca-Capote,[8] María José Bravo,[2] José Luis Royo,[2] Alberto Perez-Gomez,[8] María Inés Camacho-Sojo,[8] Isabel Gallego,[8] Joana Vitalle,[8] Sara Bachiller,[8,9] Alicia Gutierrez-Valencia,[8] Francisco Vidal,[3,10,11,12] Jacques Fellay,[5,6] Mathias Lichterfeld,[13,14,15,16] Ezequiel Ruiz-Mateos,[8,18,*] and the Swiss HIV Cohort Study

SUMMARY

Some HIV controllers experience immunologic progression with CD4$^+$ T cell decline. We aimed to identify genetic factors associated with CD4$^+$ T cell lost in HIV controllers. A total of 561 HIV controllers were included, 442 and 119 from the International HIV controllers Study Cohort and the Swiss HIV Cohort Study, respectively. No SNP or gene was associated with the long-term non-progressor HIV spontaneous control phenotype in the individual GWAS or in the meta-analysis. However, SNPs previously associated with natural HIV control linked to HLA-B (rs2395029 [p = 0.005; OR = 1.70], rs59440261 [p = 0.003; OR = 1.78]), *MICA* (rs112243036 [p = 0.011; OR = 1.45]), and *PSORS1C1* loci (rs3815087 [p = 0.017; OR = 1.39]) showed nominal association with this phenotype. Genetic factors associated with the long-term HIV controllers without risk of immunologic progression are those previously related to the overall HIV controller phenotype.

INTRODUCTION

HIV controllers are an extraordinary subset of individuals who are able to naturally control HIV replication in the absence of antiretroviral therapy.[1] In fact, this phenotype has been proposed as a model of functional cure.[2]

The natural control of HIV infection has been considered a complex condition where host genetic factors could play a key role. For this reason, several genome-wide association studies (GWASs) have been published focused on this specific phenotype as well as in related ones such as viral load at set point and disease progression.[3,4] The SNPs more frequently associated with these related phenotypes in Caucasian individuals were rs2395029,[5–11] linked to *HLA-B**5701 allele, and rs9264942,[6,7,10] linked to HLA-C locus. These findings pointed out the importance of the HLA region in all these phenotypes.

However, different studies afterward revealed how some of these subjects showed clinical, immunological, and virological progression.[12,13] There are individuals eventually losing the virological control, the so-called transient controllers, opposed to persistent controllers who maintained virus control permanently.[14,15] Persistent controllers are characterized by having high HIV-specific T cell response, low viral diversity and HIV reservoir, low frequency of viral blips, and a peculiar proteomic, lipidomic, and metabolomic profiles compatible with low levels of inflammation.[15–18]

In terms of immunological progression in HIV controllers, different immunological factors have been involved in CD4$^+$ T cell loss, such as disturbance in T cell homeostasis,[19] with T cell activation[20,21] being one of the main proposed mechanisms. In relation to genetic factors involved in this specific phenotype, data are scarce. We have previously shown in a local and in a validation cohort from the International HIV controllers Study[22] that discrete genetic markers, commonly used in the clinic, such as *HLA-B*57* and interferon lambda-4-related polymorphisms, were associated with protection against CD4$^+$ T cell loss in controllers. However, to our knowledge, there are no prior GWAS focused on this issue only taking into account the HIV controller population.

[1]Unidad Clínica de Enfermedades Infecciosas y Microbiología, Hospital Universitario Virgen de Valme, Instituto de Biomedicina de Sevilla, IBIS, Sevilla, Spain

[2]Departamento de Especialidades Quirúrgicas, Bioquímica e Inmunología, Facultad de Medicina, Universidad de Málaga, Málaga, Spain

[3]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas (CIBERINFEC)-Instituto de Salud Carlos III, Madrid, Spain

[4]Centro Andaluz de Estudios Bioinformáticos (CAEBI, SL), Sevilla, Spain

[5]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[6]Precision Medicine Unit, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland

[7]Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Sevilla/Instituto de Biomedicina de Sevilla-Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, 41012 Sevilla, Spain

[8]Institute of Biomedicine of Seville, IBiS/Virgen del Rocío University Hospital/CSIC/University of Seville, Clinic Unit of Infectious Diseases, Microbiology and Preventive Medicine, Seville, Spain

[9]Department of Medical Biochemistry, Molecular Biology and Immunology, School of Medicine,

Continued

Table 1. Main characteristics of subjects selected from the International HIV controllers Study Cohort

Variables	LTNP-C (n = 233)	Non-LTNP-C (n = 209)	p; OR (95%CI)
Female, n (%)	42 (18.0)	38 (18.2)	0.95; 1.02 [0.63–1.65]
Age at HIV diagnosis, years[a]	32 [26–28]	35 [29–41]	<0.001; 0.96 [0.94–0.98]
Calendar year of HIV diagnosis, n (%)			
1980–1989	97 (41.6)	19 (9.1)	Ref.
1990–1999	100 (42.9)	43 (20.7)	<0.001; 20.71 [11.22–38.20]
2000–2008	36 (15.5)	146 (70.2)	<0.001; 9.43 [5.66–15.72]
PWID, n (%)	24 (10.3)	14 (6.7)	0.19; 1.59 [0.80–3.17]
Elite, n (%)	83 (35.6)	51 (24.4)	0.01; 1.70 [1.13–2.60]

PWID, people who have ever injected drugs; LTNP-C, long-term non-progress controllers; OR, Odds ratio; CI, confidence interval.

Ref., reference variable.

[a] Median (quartile 1 – quartile 3). Data available in 440 individuals.

Identifying biological markers of HIV-1 controllers who are at risk for HIV-1 disease progression as defined by declining CD4$^+$ T cell counts is an important objective, both for improving clinical care of these persons and for extending the conceptual understanding of immune mechanisms involved in persistent long-term non-progressor HIV spontaneous control (LTNP-C) as the right model of functional cure. Because of this, the aim of this study was to analyze genetic determinants, at a genome-wide association level, related to the LTNP-C phenotype in HIV controllers.

RESULTS

Study populations

A total of 443 and 120 individuals from the International HIV controllers Study Cohort (ICSC) and the Swiss HIV Cohort Study (SHCS), respectively, fulfilled the inclusion criteria. Among them, 233 (52.7%) subjects in the ICSC and 59 (49.2%) subjects in the SHCS were LTNP-Cs. After quality controls, 442 and 119 subjects remained in the ICSC and SHCS cohorts, respectively. These individuals constituted the study populations. Among them, 233 (52.7%) individuals in the ICSC and 58 (48.7%) individuals in the SHCS were LTNP-Cs. Main characteristics of both populations are depicted in Tables 1 and 2.

Single-locus association analyses and *meta*-GWAS

A total of 477,986 and 9,829,427 SNPs were available in the ICSC and SHCS datasets, respectively. Among them, 5,766,421 SNPs passed the quality controls in the SHCS dataset. Regarding the ISCS dataset, 477,952 SNPs remained after the quality controls. Overall, a total of 7,597,066 SNPs were available for association studies after imputation.

Principal component (PC) analysis performed among each dataset did not reveal population admixture (Figures S1A and S1B). Moreover, there was no overall inflation of the test statistic ($\lambda \leq 1.05$) on each of them (Figures S2A and S2B), supporting that systematic confounding factors were unlikely in both populations.

Results of single-locus genetic association analysis adjusted by PC vectors, age, sex, and the elite controller condition on each dataset did not reveal any marker associated with the LTNP-C phenotype in any of the two datasets analyzed. However, 28 suggestive signals ($p < 10^{-5}$) were found in the ICSC population (Table S1 and Figure S3).

We analyzed combined data from both datasets using the meta-analysis tool in Plink. A total of 4,562,723 SNPs were common in both datasets. Again, there was no overall inflation of the test statistic ($\lambda = 1.00$) in this analysis (Figure S4). No SNP was associated with the LTNP-C phenotype at GWAS significant p value (Figure 1). Nevertheless, 17 SNPs, linked to *LOC285696*, *RMI2* loci, and to an intergenic region within chromosome 5, showed suggestive association (Table 3).

University of Seville, Seville, Spain

[10] Hospital Universitari de Tarragona Joan XXIII (HJ23), Tarragona, Spain

[11] Institut Investigació Sanitària Pere Virgili (IISPV), Tarragona, Spain

[12] Universitat Rovira i Virgili (URV), Tarragona, Spain

[13] Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA

[14] Harvard Medical School, Boston, MA, USA

[15] Infectious Disease Division, Massachusetts General Hospital, Boston, MA, USA

[16] Infectious Disease Division, Brigham and Women's Hospital, Boston, MA, USA

[17] These authors contributed equally

[18] Lead contact

*Correspondence:
eruizmateos-ibis@us.es

https://doi.org/10.1016/j.isci.2023.107214

Table 2. Main characteristics of subjects selected from the Swiss HIV Cohort Study

Variables	LTNP-C (n = 58)	Non-LTNP-C (n = 61)	p; OR (95%CI)
Female, n (%)	23 (39.7)	25 (41.0)	0.88; 1.06 [0.51–2.12]
Age at HIV diagnosis, years[a]	31 [24–37]	30 [26–38]	0.33; 0.98 [0.95–1.02]
Calendar year of HIV diagnosis, n (%)			
1980–1989	14 (24)	8 (13.1)	Ref.
1990–1999	24 (41.4)	34 (55.7)	0.35; 1.66 [0.57–4.86]
2000–2008	20 (34.5)	19 (31.1)	0.34; 0.67 [0.30–1.52]
PWID, n (%)	15 (25.9)	20 (32.8)	0.41; 1.40 [0.63–3.10]
Elite, n (%)	18 (31)	13 (21.3)	0.23; 1.70 [0.73–3.80]

PWID, people who have ever injected drugs; LTNP-C, long-term non-progress controllers; OR, Odds ratio; CI, confidence interval.

Ref., reference variable.

[a] Median (quartile 1 – quartile 3).

Gene-based association and enrichment analyses

The results obtained in the meta-analysis were analyzed with the Magma software for carrying a gene-based association analysis. A total of 18,173 genes were ranked, but none of them reached the threshold p value stablished after multiple testing correction. Despite this, the *RMI2* gene reached suggestive statistical significance (p_SNPwise_mean value = 7.52×10^{-6}) (Table S2).

We analyzed if the best 200 ranked genes obtained in the gene-based association study (Table S2) were significantly aggregated in specific categories of Gene Ontology for biological processes. The process related to metallopeptidase activity reached the FDR p value threshold stablished (Table 4).

Analysis of SNPs previously associated with the HIV control or related phenotypes

Table 5 shows those SNPs previously associated at GWAS p value significant level with HIV controllers, viral load at set point, and/or disease progression reported in GWAS performed in Caucasians that were

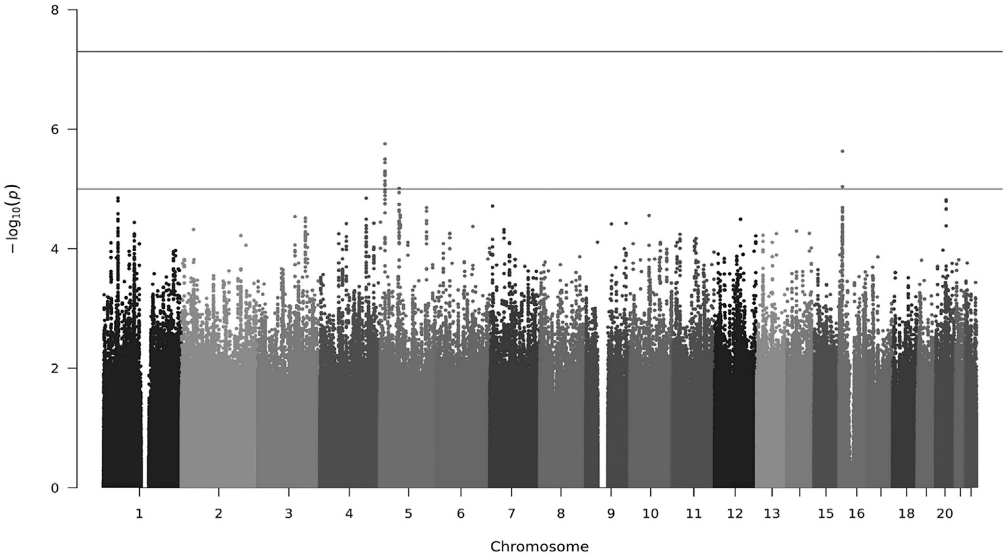


Figure 1. Manhattan plot of meta-analysis of the two GWAs carried out in selected subjects from the Swiss HIV Cohort Study and from the International HIV controllers Study Cohort

Horizontal lines correspond to 1×10^{-5} and 5×10^{-8} p values, respectively.

Table 3. Best single locus results ($p < 10^{-5}$) using the meta-analysis tool in Plink

CHR	SNP	BP	A1	P	P(R)	OR	OR(R)	Q	I	ANNOT
5	rs2261514	17145620	A	1.75×10^{-6}	4.58×10^{-6}	2.19	2.18	0.306	4.33	*LOC285696*
16	rs35578928	11467777	A	2.33×10^{-6}	2.33×10^{-6}	2.28	2.28	0.432	0	*RMI2*[a]
5	rs11739746	17147118	C	3.15×10^{-6}	3.15×10^{-6}	2.17	2.17	0.571	0	*LOC285696*
5	rs2624417	17148318	A	3.15×10^{-6}	3.15×10^{-6}	2.17	2.17	0.571	0	*LOC285696*
5	rs2731804	17146261	C	3.60×10^{-6}	3.60×10^{-6}	2.13	2.13	0.354	0	*LOC285696*
5	rs2652674	17152228	G	5.01×10^{-6}	5.01×10^{-6}	2.12	2.12	0.476	0	*LOC285696*
5	rs2170525	17132500	T	5.33×10^{-6}	2.71×10^{-6}	2.04	1.95	0.201	38.84	*LOC285696*
5	rs2731796	17153618	G	5.63×10^{-6}	4.58×10^{-5}	2.09	2.06	0.289	10.93	*LOC285696*
5	rs2731795	17154626	A	5.63×10^{-6}	4.58×10^{-5}	2.09	2.06	0.289	10.93	*LOC285696*
5	rs2652666	17119879	T	5.64×10^{-6}	0.077	1.99	1.77	0.057	72.29	*LOC285696*[a]
5	rs2731788	17130597	T	5.94×10^{-6}	0.017	2.04	1.88	0.134	55.25	*LOC285696*
5	rs2624429	17142201	G	7.32×10^{-6}	0.009	2.03	1.90	0.167	47.43	*LOC285696*
5	rs2624431	17136611	G	8.63×10^{-6}	6.81×10^{-5}	2.02	2.01	0.290	10.38	*LOC285696*
5	rs2731806	17141956	A	8.63×10^{-6}	6.81×10^{-6}	2.02	2.00	0.290	10.38	*LOC285696*
16	rs4508435	11478703	G	9.11×10^{-6}	9.11×10^{-6}	1.99	1.99	0.740	0	*RMI2*[a]
5	rs67834917	62282963	G	9.70×10^{-6}	9.70×10^{-6}	0.49	0.49	0.446	0	.
5	rs72756740	62286523	T	9.70×10^{-6}	9.70×10^{-6}	0.49	0.49	0.446	0	.

CHR, Chromosome; SNP, Single-nucleotide polymorphism; BP, Base pair position according to UCSC genome browser (NCBI37/hg19) and dbSNP build 142; A1, Reference allele (minor allele); p, Fixed-effects p value; p(R), Random-effects p value; OR, Fixed-effects odds ratio (for LTNP-C condition); OR(R), Random-effects OR; Q, p value for heterogeneity of OR; I, effect size for heterogeneity of OR. Gene names are shown in italics.

[a] Closer gene within 200 kilobases.

available in our meta-analysis. Four of them showed nominal association with the LTNP-C phenotype in the same direction that previously was reported (Table 5).

DISCUSSION

This study shows that those genetic factors previously associated with the spontaneous HIV control in the overall HIV controller population seem to be also involved in the long-term maintenance of $CD4^+$ T cells in HIV controllers. However, neither single-locus association analyses nor gene-based association analyses yielded statistically significant results at the p value thresholds stablished.

Taken together, these facts could suggest that non-genetic factors or genetic-environment interactions could have a more important role in the protection against the $CD4^+$ T cell loss in HIV controllers than genetic factors. According to this hypothesis, we previously identified differences in virological and immune factors between subjects with persistent control of viral replication and controllers who lost virological control.[13,15,17,18] Moreover, it has been reported that viral genetic variation has also a key role in the progression of the disease in Caucasian subjects.[26]

Despite this, some suggestive signals were observed in our meta-analysis that deserve attention. Most of them were linked to *LOC285696*, also known as *BASP1 antisense RNA 1* (*BASP1-AS1*). *BASP1-AS1* regulates the expression of its adjacent coding gene, *BASP1*, which seems to be involved in Th17 cells differentiation.[27] This subtype of $CD4^+$ T cells are especially susceptible to HIV infection and, for that reason, this subpopulation appear depleted in infected individuals, but not in elite controllers.[28] In addition, it has also been suggested that the expression of *BASP1* might be involved in modulation of the transcriptional program during T cell apoptosis.[29] Similarly, some of the better signals were reached by SNPs linked to *RMI2* gene. Moreover, this gene was the top signal in the gene-based association analyses. RMI2 is a component of a protein complex together with BLM and topoisomerase IIIα (OMIM * 612426). This complex is involved in the regulation of genome integrity, a key process in lymphocyte proliferation.[30] Interestingly, the metalloprotease activity was the only biological processes associated with the LTNP-C phenotype in our study. The HIV infection has been related to the matrix metalloproteinase dysregulation.

Table 4. Categories of Gene Ontology (GO) (http://www.geneontology.org/) for biological processes that were enriched using the top 200 ranked genes obtained in the gene-based association analysis

GO category	Description	Category Size	Genes included	p value	FDR	Genes included
GO:0008237	metallopeptidase activity	185	9	1.29×10^{-5}	0.015	*TRHDE; NAALAD2; ASTL; ZMPSTE24; CPO; KEL; ADAMTS10; MMP20; ADAM9*

FDR, False discovery rate. Gene names are shown in italics.

In fact, the increment of metalloprotease activity has been associated with both the HIV dissemination and HIV-related pathology progression.[31] However, its specific role in the LTNP-C phenotype is not known. All these results warrant future validations in independent studies.

Among those SNPs previously associated with the natural HIV control or related phenotypes analyzed in our study, four were associated with the LTNP-C phenotype. One of the strongest associations was observed with the rs2395029 marker within the *HCP5* gene and linked to the *HLA*-B*5701 allele. This SNP was previously associated with viral load at set point,[6,7] HIV-1 disease progression,[6] and HIV-1 controller phenotype.[5,9–11] Likewise, other markers linked to the *HLAB* locus such as rs59440261, in partial linkage disequilibrium with rs2395029,[25] and rs9266409, located in the 3′ region of *HLA-B*, all of them previously associated with viral load at set point,[6] were also associated, or tended to be associated, with the LTNP-C phenotype in our study. These results are in agreement with our previous study that found an association between the *HLA-B**57 allele and protection against CD4$^+$ T cell loss in controllers.[22]

Interestingly, rs112243036 within *MICA* gene also showed association with the LTNP-C phenotype in our study. This marker was previously independently associated with HIV controllers.[23] It has been proposed that the rs2395029G-rs112243036A-rs9264942C haplotype significantly favors the viral load control and the non-progressor phenotype.[23] However, the SNP rs9264942, linked to HLA-C locus and previously associated with viral load at set point and with the non-progressor phenotype,[6,7,10] was not associated with the LTNP-C phenotype in our study. Therefore, our results partially support the hypothesis that the effect of rs9264942 could be only observed in the context of rs2395029G-rs112243036A haplotype as Le Clerc et al. proposed.[23] Taken together, all these data pointed out the possible role of the *HLAB* and *MICA* loci in the maintenance of CD4$^+$ T cell levels within HIV controllers.

The fact that the most replicated genetic associations with the HIV controller phenotype are also observed in our work suggests that in previous studies, the overall HIV controller population was enriched in the LTNP-C phenotype analyzed herein. Therefore, these genetic factors could help to refine the true HIV controller phenotype characterized by the long-term HIV remission or low viral loads and absence of HIV disease progression which in the case of elite participants is similar to those previously defined as exceptional or persistent HIV controllers.[15,18,32,33] In this sense, non-LTNP-C phenotype may behave as non-controllers justifying the association of the LTNP-C phenotype with classical SNPs associated with spontaneous HIV control, such as those in *HLAB* and *MICA* loci.

In conclusion, we have performed the first *meta*-GWAS focused in the long-term maintenance of CD4$^+$ T cells in HIV controllers. Our results suggest that genetic factors previously associated with the overall HIV controller phenotype, mainly those linked to *PSORS1C1*, *HLAB*, and *MICA* loci, seem to have a role in this complex trait. In this study, we also have identified suggestive signals that deserve a validation in the future. However, GWAS and *meta*-GWAS analysis in larger samples, deep-sequencing of candidate genes as well as both gene-gene and genetic-environment interactions analyses will be needed to identify all the genetic variants involved in the maintenance of the non-immunological progressor HIV controller phenotype and what their real effects are. The results of the present study could help us to refine the best definition for HIV controllers. Consequently, these HIV controllers should be considered as the best model of functional cure. The study of this specific phenotype could help us to design new strategies for preventing HIV progression, not necessarily associated with antiretroviral treatment, in those subjects who show CD4$^+$ T cell loss.

Table 5. Main SNPs previously associated at GWAS-p value significant with natural HIV control, HIV disease progression, or viral load set point in Caucasian populations that has been analyzed in our study

CHR	Linked Gene	SNP	Allele	Reported effect	Reference	P	p (R)	OR	OR (R)	Q	I	ICSC p; OR (95%CI)	SHCS p; OR (95%CI)
6	*HLA-C*	rs9264942	C	Protective	Fellay et al., International HIV Controllers Study, et al.[6,7,10]	0.089	0.826	1.24	1.07	0.060	71.52	0.020; 1.38 (1.05–1.83)	0.317; 0.73 (0.40–1.34)
6	***HCP5***	**rs2395029**	G	Protective	Dalmasso et al.; Fellay et al.; Le Clerc et al.; Limou et al.; International HIV Controllers Study et al.; Guergnon et al.[5–11]	0.005	0.005	1.70	1.70	0.561	0	0.006; 1.81 (1.18–2.77)	0.442; 1.37 (0.60–3.22)
6	***MICA***	**rs112243036**	A	Protective	Le Clerc et al.[23]	0.011	0.011	1.45	1.45	0.582	0	0.012; 1.52 (1.09–2.11)	0.515; 1.24 (0.64–2.38)
6	*TRIM10*	rs9468692	T	Not Known	Fellay et al.[6]	0.883	0.883	0.96	0.96	0.427	0	0.584; 0.84 (0.47–1.52)	0.553; 1.32 (0.47–3.35)
6	*HLA-B*	rs9266409	C	Protective	Fellay et al.[6]	0.051	0.051	1.33	1.33	0.719	0	0.124; 1.29 (0.93–1.80)	0.211; 1.49 (0.80–2.69)
6	*PSORS1C3*	rs3131018	A	Protective	International HIV Controllers Study et al.[10]	0.898	0.898	0.98	0.98	0.693	0	0.763; 0.94 (0.67–1.34)	0.776; 1.09 (0.58–2.05)
6	***PSORS1C1***	**rs3815087**	T	Protective	Limou et al.[9]	0.017	0.017	1.39	1.39	0.316	0,55	0.091; 1.29 (0.95–1.76)	0.051; 1.84 (0.99–3.40)
6	*ZNRD1*	rs9261174	C	Not Known	Fellay et al.[6]	0.761	0.761	1.06	1.06	0.960	0	0.796; 1.05 (0.71–1.54)	0.867; 1.08 (0.44–2.61)
6	*RNF39*	rs2074480	G	Not Known	Fellay et al.[6]	0.760	0.760	1.06	1.06	0.960	0	0.795; 1.05 (0.71–1.54)	0.867; 1.08 (0.44–2.61)
6	*RNF39*	rs2301753	T	Not Known	Fellay et al.[6]	0.760	0.760	1.06	1.06	0.960	0	0.795; 1.05 (0.71–1.54)	0.867; 1.08 (0.44–2.61)
6	*ZNRD1*	rs9261129	C	Not Known	Fellay et al.[6]	0.760	0.760	1.06	1.06	0.960	0	0.795; 1.05 (0.71–1.54)	0.867; 1.07 (0.45–2.61)
3	*CXCR6*	rs2234358	T	Non Protective	Limou et al.[24]	0.188	0.188	0.84	0.84	0.330	0	0.470; 0.89 (0.65–1.99)	0.142; 0.66 (0.38–1.15)
6	***HLA-B***	**rs59440261**	A	Protective	McLearn et al.[25]	0.003	0.003	1.78	1.78	0.657	0	0.005; 1.87 (1.21–2.90)	0.305; 1.52 (0.68–3.92)
3	*CCRL2*	rs1015164	A	Protective	McLearn et al.[25]	0.516	0.516	0.91	0.91	0.618	0	0.422; 0.89 (0.67–1.20)	0.866; 1.05 (0.55–2.00)

CHR. Chromosome; SNP. Single Nucleotide Polymorphism; A1. Reference allele (minor allele); P. Fixed-effects p value; P(R): Random-effects p value; OR. Fixed-effects Odds Ratio (for being LTNP-C); SHCS. Swiss HIV Cohort Study; ICSC. International HIV controllers Study Cohort; Q: p value for heterogeneity of OR; I. effect size for heterogeneity of OR; OR(R); Random-effects Odds ratio; CI, Confidence interval. Common SNPs found significant in our dataset are shown in bold. Gene names are shown in italics.

Limitations of the study

Our work has some limitations. First, due to the low number of individuals included on each dataset, the power for detecting GWAS significant signals could be low. Given the sample size and the case-control ratio in the ICSC and a minor allele frequency of a causative SNP of 0.25, this GWAS had an 80% power to detect OR>3.8 at the p value stablished in these studies. However, our meta-analysis had 80.0% power to detect OR = 3.2 at the same p value. Therefore, our study was powered to detect robust signals but not genetic factors with modest effect. However, the fact of our study was performed on a rare phenotype using two very well characterized cohorts are strengths of our work. Second, since meta-analysis was performed with those SNPs genotyped in both datasets, it is possible that some SNPs associated with the LTNP-C phenotype were excluded in our study. It was the case of the rs12980275 marker linked to *IL28B* gene, a SNP in strong linkage disequilibrium with the causative variant rs368234815 in Caucasian individuals,[34] that was previously associated with LTNP-C by us.[22] This marker was not included or imputed in the SHCS dataset and it could not be analyzed in the meta-analysis. Similarly, some of the markers previously associated with related genotypes could not be analyzed because they have not been genotyped or imputed in one or both datasets (Table S3). However, although the possibility to exclude important SNPs exists, the inclusion of more than 4 million of SNPs in the meta-analysis minimalizes this risk.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Genotype quality controls and genotype imputation
  - Single locus association study and meta-analysis
  - Gene-based association study and enrichment analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107214.

ACKNOWLEDGMENTS

This work was supported by grants from the Consejería de Salud de la Junta de Andalucía (PI-0001/2017); Plan Nacional de I + D + I cofinanced by ISCIII-Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER) co-funded by the European Union (www.red.es/redes/inicio) (RD16/0025/0040, RD12/0017/0012). E.R.-M. was supported by Consejo Superior de Investigaciones Científicas (CSIC) and Consejería de Salud y Bienestar Social, Junta de Andalucía through the Nicolás Monardes Program (C-0032/17) and Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Fondos Europeos para el Desarrollo Regional, FEDER, co-funded by the European Union grants PI16/00684, PI19/01127, PI22/01796, PI10/02635, PI13/0796, PI16/00503 and PI19/01337, ''A way to make Europe''/''Investing in your future''; RETICS, Red de Investigación en SIDA co-funded by the European Union (RD16/0025/0020, RD16/0025/0006-ISCIII-FEDER (Spain). Programa de Suport als Grups de Recerca AGAUR (2017SGR948); Gilead Fellowship Program GLD14/293 and GLD22/00147. F.V. is supported by grants from the Programa de Intensificación de Investigadores (I)-ISCIII. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

This study would not have been possible without the collaboration of the patients, medical and nursery staff, and data managers who took part in the project.

Members of the Swiss HIV Cohort Study: Anagnostopoulos A, Battegay M, Bernasconi E, Böni J, Braun DL, Bucher HC, Calmy A, Cavassini M, Ciuffi A, Dollenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H, Fux CA, Günthard HF (President of the SHCS), Haerry D (deputy of ''Positive Council''), Hasse B, Hirsch HH, Hoffmann M, Hösli I, Huber M, Kahlert CR (Chairman of the Mother & Child Substudy), Kaiser L, Keiser O, Klimkait T, Kouyos RD, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner KJ, Müller N, Nicca D, Paioni P, Pantaleo G, Perreau M, Rauch A (Chairman of the Scientific Board), Rudin C, Scherrer AU (Head of Data Center), Schmid P, Speck R, Stöckle M (Chairman of the Clinical and Laboratory Committee), Tarr P, Trkola A, Vernazza P, Wandeler G, Weber R, Yerly S.

AUTHOR CONTRIBUTIONS

Study design: L.M.R., M.E.S., E.R.M.; raw genomic-data management: C.T., J.F., M.L., E.R.M.; data quality controls and statistical analysis: L.M.R., M.E.S., A.G.P., R.R., E.R.M.; data interpretation: L.M.R., M.E.S., A.C.G., M.R.J.L., A.G.S., C.G.C., M.J.B., J.L.R., A.P.G., I.C.S., I.G., J.V., S.B., A.G.V., F.V., E.R.M.; writers: L.M.R., M.E.S., E.R.M. with contribution of all authors. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: April 4, 2023
Revised: May 19, 2023
Accepted: June 22, 2023
Published: June 26, 2023

REFERENCES

1. Lambotte, O., Boufassa, F., Madec, Y., Nguyen, A., Goujard, C., Meyer, L., Rouzioux, C., Venet, A., and Delfraissy, J.F.; SEROCO-HEMOCO Study Group (2005). HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication. Clin. Infect. Dis. *41*, 1053–1056. https://doi.org/10.1086/433188.
2. Autran, B., Descours, B., Avettand-Fenoel, V., and Rouzioux, C. (2011). Elite controllers as a model of functional cure. Curr. Opin. HIV AIDS *6*, 181–187. https://doi.org/10.1097/COH.0b013e328345a328.
3. Le Clerc, S., Limou, S., and Zagury, J.F. (2019). Large-Scale "OMICS" Studies to Explore the Physiopatholgy of HIV-1 Infection. Front. Genet. *10*, 799. https://doi.org/10.3389/fgene.2019.00799.
4. Limou, S., and Zagury, J.F. (2013). Immunogenetics: Genome-Wide Association of Non-Progressive HIV and Viral Load Control: HLA Genes and Beyond. Front. Immunol. *4*, 118. https://doi.org/10.3389/fimmu.2013.00118.
5. Dalmasso, C., Carpentier, W., Meyer, L., Rouzioux, C., Goujard, C., Chaix, M.L., Lambotte, O., Avettand-Fenoel, V., Le Clerc, S., de Senneville, L.D., et al. (2008). Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. PLoS One *3*, e3907. https://doi.org/10.1371/journal.pone.0003907.
6. Fellay, J., Ge, D., Shianna, K.V., Colombo, S., Ledergerber, B., Cirulli, E.T., Urban, T.J., Zhang, K., Gumbs, C.E., Smith, J.P., et al. (2009). Common genetic variation and the control of HIV-1 in humans. PLoS Genet. *5*, e1000791. https://doi.org/10.1371/journal.pgen.1000791.
7. Fellay, J., Shianna, K.V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. Science *317*, 944–947. https://doi.org/10.1126/science.1143767.
8. Le Clerc, S., Limou, S., Coulonges, C., Carpentier, W., Dina, C., Taing, L., Delaneau, O., Labib, T., Sladek, R., ANRS Genomic Group, et al. (2009). Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). J. Infect. Dis. *200*, 1194–1201. https://doi.org/10.1086/605892.
9. Limou, S., Le Clerc, S., Coulonges, C., Carpentier, W., Dina, C., Delaneau, O., Labib, T., Taing, L., Sladek, R., Deveau, C., et al. (2009). Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). J. Infect. Dis. *199*, 419–426. https://doi.org/10.1086/596067.
10. International HIV Controllers Study, Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., de Bakker, P.I.W., Walker, B.D., Ripke, S., Brumme, C.J., Pulit, S.L., et al. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science *330*, 1551–1557. https://doi.org/10.1126/science.1195271.
11. Guergnon, J., Dalmasso, C., Broet, P., Meyer, L., Westrop, S.J., Imami, N., Vicenzi, E., Morsica, G., Tinelli, M., Zanone Poma, B., et al. (2012). Single-nucleotide polymorphism-defined class I and class III major histocompatibility complex genetic subregions contribute to natural long-term nonprogression in HIV infection. J. Infect. Dis. *205*, 718–724. https://doi.org/10.1093/infdis/jir833.
12. Leon, A., Perez, I., Ruiz-Mateos, E., Benito, J.M., Leal, M., Lopez-Galindez, C., Rallon, N., Alcami, J., Lopez-Aldeguer, J., Viciana, P., et al. (2016). Rate and predictors of progression in elite and viremic HIV-1 controllers. AIDS *30*, 1209–1220. https://doi.org/10.1097/QAD.0000000000001050.
13. Chereau, F., Madec, Y., Sabin, C., Obel, N., Ruiz-Mateos, E., Chrysos, G., Fidler, S., Lehmann, C., Zangerle, R., Wittkop, L., et al. (2017). Impact of CD4 and CD8 dynamics and viral rebounds on loss of virological control in HIV controllers. PLoS One *12*, e0173893. https://doi.org/10.1371/journal.pone.0173893.
14. Dominguez-Molina, B., Leon, A., Rodriguez, C., Benito, J.M., Lopez-Galindez, C., Garcia, F., Del Romero, J., Gutierrez, F., Viciana, P., Alcami, J., et al. (2016). Analysis of Non-AIDS-Defining Events in HIV Controllers. Clin. Infect. Dis. *62*, 1304–1309. https://doi.org/10.1093/cid/ciw120.
15. Pernas, M., Tarancón-Diez, L., Rodríguez-Gallego, E., Gómez, J., Prado, J.G., Casado, C., Dominguez-Molina, B., Olivares, I., Coiras, M., León, A., et al. (2018). Factors Leading to the Loss of Natural Elite Control of HIV-1 Infection. J. Virol. *92*, e01805-17. https://doi.org/10.1128/JVI.01805-17.
16. Canouï, E., Lécuroux, C., Avettand-Fenoël, V., Gousset, M., Rouzioux, C., Saez-Cirion, A., Meyer, L., Boufassa, F., Lambotte, O., Noël, N., et al. (2017). A Subset of Extreme Human Immunodeficiency Virus (HIV) Controllers Is Characterized by a Small HIV Blood Reservoir and a Weak T-Cell Activation Level. Open Forum Infect. Dis. *4*, ofx064. https://doi.org/10.1093/ofid/ofx064.
17. Rodríguez-Gallego, E., Tarancón-Diez, L., García, F., Del Romero, J., Benito, J.M., Alba, V., Herrero, P., Rull, A., Dominguez-Molina, B., Martinez-Madrid, O., et al. (2019). Proteomic Profile Associated With Loss of Spontaneous Human Immunodeficiency Virus Type 1 Elite Control. J. Infect. Dis. *219*, 867–876. https://doi.org/10.1093/infdis/jiy599.
18. Tarancon-Diez, L., Rodríguez-Gallego, E., Rull, A., Peraire, J., Viladés, C., Portilla, I., Jimenez-Leon, M.R., Alba, V., Herrero, P., Leal, M., et al. (2019). Immunometabolism is a key factor for the persistent spontaneous elite control of HIV-1 infection. EBioMedicine *42*,

86–96. https://doi.org/10.1016/j.ebiom.2019.03.004.

19. Benito, J.M., Ortiz, M.C., León, A., Sarabia, L.A., Ligos, J.M., Montoya, M., Garcia, M., Ruiz-Mateos, E., Palacios, R., Cabello, A., et al. (2018). Class-modeling analysis reveals T-cell homeostasis disturbances involved in loss of immune control in elite controllers. BMC Med. *16*, 30. https://doi.org/10.1186/s12916-018-1026-6.

20. Bansal, A., Sterrett, S., Erdmann, N., Westfall, A.O., Dionne-Odom, J., Overton, E.T., and Goepfert, P.A. (2015). Normal T-cell activation in elite controllers with preserved CD4+ T-cell counts. Aids *29*, 2245–2254. https://doi.org/10.1097/QAD.0000000000000860.

21. Hunt, P.W., Brenchley, J., Sinclair, E., McCune, J.M., Roland, M., Page-Shafer, K., Hsue, P., Emu, B., Krone, M., Lampiris, H., et al. (2008). Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy. J. Infect. Dis. *197*, 126–133. https://doi.org/10.1086/524143.

22. Dominguez-Molina, B., Tarancon-Diez, L., Hua, S., Abad-Molina, C., Rodriguez-Gallego, E., Machmach, K., Vidal, F., Tural, C., Moreno, S., Goñi, J.M., et al. (2017). HLA-B*57 and IFNL4-related polymorphisms are associated with protection against HIV-1 disease progression in controllers. Clin. Infect. Dis. *64*, 621–628. https://doi.org/10.1093/cid/ciw833.

23. Le Clerc, S., Delaneau, O., Coulonges, C., Spadoni, J.L., Labib, T., Laville, V., Ulveling, D., Noirel, J., Montes, M., Schächter, F., et al. (2014). Evidence after imputation for a role of MICA variants in nonprogression and elite control of HIV type 1 infection. J. Infect. Dis. *210*, 1946–1950. https://doi.org/10.1093/infdis/jiu342.

24. Limou, S., Coulonges, C., Herbeck, J.T., van Manen, D., An, P., Le Clerc, S., Delaneau, O., Diop, G., Taing, L., Montes, M., et al. (2010). Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. J. Infect. Dis. *202*, 908–915. https://doi.org/10.1086/655782.

25. McLaren, P.J., Coulonges, C., Bartha, I., Lenz, T.L., Deutsch, A.J., Bashirova, A., Buchbinder, S., Carrington, M.N., Cossarizza, A., Dalmau, J., et al. (2015). Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. Proc. Natl. Acad. Sci. USA *112*, 14658–14663. https://doi.org/10.1073/pnas.1514867112.

26. Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., Hall, M., Hillebregt, M., Ong, S.H., Albert, J., et al. (2017). Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. PLoS Biol. *15*, e2001855. https://doi.org/10.1371/journal.pbio.2001855.

27. Tuomela, S., Salo, V., Tripathi, S.K., Chen, Z., Laurila, K., Gupta, B., Äijö, T., Oikari, L., Stockinger, B., Lähdesmäki, H., and Lahesmaa, R. (2012). Identification of early gene expression changes during human Th17 cell differentiation. Blood *119*, e151–e160. https://doi.org/10.1182/blood-2012-01-407528.

28. Renault, C., Veyrenche, N., Mennechet, F., Bedin, A.S., Routy, J.P., Van de Perre, P., Reynes, J., and Tuaillon, E. (2022). Th17 CD4+ T-Cell as a Preferential Target for HIV Reservoirs. Front. Immunol. *13*, 822576. https://doi.org/10.3389/fimmu.2022.822576.

29. Ohsawa, S., Watanabe, T., Katada, T., Nishina, H., and Miura, M. (2008). Novel antibody to human BASP1 labels apoptotic cells post-caspase activation. Biochem. Biophys. Res. Commun. *371*, 639–643. https://doi.org/10.1016/j.bbrc.2008.04.056.

30. Mönnich, M., Hess, I., Wiest, W., Bachrati, C., Hickson, I.D., Schorpp, M., and Boehm, T. (2010). Developing T lymphocytes are uniquely sensitive to a lack of topoisomerase III alpha. Eur. J. Immunol. *40*, 2379–2384. https://doi.org/10.1002/eji.201040634.

31. Mastroianni, C.M., and Liuzzi, G.M. (2007). Matrix metalloproteinase dysregulation in HIV infection: implications for therapeutic strategies. Trends Mol. Med. *13*, 449–459. https://doi.org/10.1016/j.molmed.2007.09.001.

32. Casado, C., Galvez, C., Pernas, M., Tarancon-Diez, L., Rodriguez, C., Sanchez-Merino, V., Vera, M., Olivares, I., De Pablo-Bernal, R., Merino-Mansilla, A., et al. (2020). Permanent control of HIV-1 pathogenesis in exceptional elite controllers: a model of spontaneous cure. Sci. Rep. *10*, 1902. https://doi.org/10.1038/s41598-020-58696-y.

33. Mendoza, D., Johnson, S.A., Peterson, B.A., Natarajan, V., Salgado, M., Dewar, R.L., Burbelo, P.D., Doria-Rose, N.A., Graf, E.H., Greenwald, J.H., et al. (2012). Comprehensive analysis of unique cases with extraordinary control over HIV replication. Blood *119*, 4645–4655. https://doi.org/10.1182/blood-2011-10-381996.

34. Real, L.M., Neukam, K., Herrero, R., Guardiola, J.M., Reiberger, T., Rivero-Juarez, A., Salazar, J., Mandorfer, M., Merino, D., Soriano, V., et al. (2014). IFNL4 ss469415590 variant shows similar performance to rs12979860 as predictor of response to treatment against Hepatitis C Virus genotype 1 or 4 in Caucasians. PLoS One *9*, e95515. https://doi.org/10.1371/journal.pone.0095515.

35. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

36. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287. https://doi.org/10.1038/ng.3656.

37. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283. https://doi.org/10.1038/ng.3643.

38. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

39. Turner, S.D. (2018). qqman: an R package for visualizing GWAS results usingQ-Q and manhattan plots. J. Open Source Softw. *3*, 731. https://doi.org/10.1101/005165.

40. Miettinen, O.S. (1969). Individual matching with multiple controls in the case of all-or-none responses. Biometrics *25*, 339–355.

41. Jackson, D., and Turner, R. (2017). Power analysis for random-effects meta-analysis. Res. Synth. Methods *8*, 290–302. https://doi.org/10.1002/jrsm.1240.

42. Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. *45*, W130–W137. https://doi.org/10.1093/nar/gkx356.

43. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

44. Real, L.M., Fernández-Fuertes, M., Sáez, M.E., Rivero-Juárez, A., Frías, M., Téllez, F., Santos, J., Merino, D., Moreno-Grau, S., Gómez-Salgado, J., et al. (2019). A genome-wide association study on low susceptibility to hepatitis C virus infection (GEHEP012 study). Liver Int. *39*, 1918–1926. https://doi.org/10.1111/liv.14177.

45. Barsh, G.S., Copenhaver, G.P., Gibson, G., and Williams, S.M. (2012). Guidelines for genome-wide association studies. PLoS Genet. *8*, e1002812. https://doi.org/10.1371/journal.pgen.1002812.

46. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
PLINK 2.0 software	Chang et al.[35]	https://www.cog-genomics.org/plink2
Michigan server	Das et al.[36]	https://imputationserver.sph.umich.edu
Sanger Imputation Service	McCarthy et al.[37]	https://imputation.sanger.ac.uk
Variant Effect Predictor tool	McLaren et al.[38]	https://www.ensembl.org/Homo_sapiens/Tools/VEP
qqman R package	Turner SD[39]	https://CRAN.R-project.org/package=qqman
Episheet software	Miettinen, OS[40]	https://www.drugepi.org/dope/software#Episheet
metapower package	Jackson et al.[41]	https://rdrr.io/github/jasonwgriffin/metapower
Webgestalt software	Wang et al.[42]	www.webgestalt.org
Gene Ontology	Ashburner et al.[43]	http://www.geneontology.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagent should be directed to and will be fulfilled by the Lead Contact, Dr. Ezequiel Ruiz-Mateos (eruizmateos-ibis@us.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In this retrospective case-control study we used the genetic data coming from two large cohorts: the International HIV controllers Study Cohort (ICSC)[10] and the Swiss HIV Cohort Study (SHCS).[6,7]

All individuals who were Caucasian HIV-1 controllers, defined as subjects with plasma HIV viral load (VL) < 2000 HIV-RNA copies/mL for at least 1 year in the absence of anti-retroviral treatment (ART),[22] were included in the present analysis. These subjects were classified as LTNP-Cs if CD4$^+$ T-cell counts were higher than 500 cells/mm^3 for more than 7 years after HIV-infection diagnosis, and those who did not fulfil this criterion and progressed to CD4$^+$ T-cell counts <500 cells/mm^3 in less than 7 years were considered as non-LTNP-Cs. Characteristics of the participants, including sex and age are described in Tables 1 and 2.

This study was in compliance with the national legislation and it was performed according to the ethical guidelines of the Declaration of Helsinki. The study was approved by the Ethics Committee of the Hospital Universitario Virgen del Rocío (Sevilla, Spain) (Code: 1036-N-15). All patients gave written informed consent before being recruited in the cohorts.

METHOD DETAILS

Genotype quality controls and genotype imputation

Only autosomal chromosomes were present in the data-sets. On each data-set, genotype quality controls were performed as previously described.[44] Briefly, samples with a call rate lower that 97% were excluded.

SNPs with a call rate <95% or with a minor allele frequency below 0.01 were removed. In addition, those individuals with heterozygosity rates greater than 0.35, or those who were related to other individuals in the sample (Identity by state (IBS) > 0.1875), were excluded. All these analyses were carried out using PLINK 2.0 software (https://www.cog-genomics.org/plink2). PC analysis was run together with other genotype data of other populations obtained from phase 3 of the 1000 Genomes Project (https://www.internationalgenome.org). Only individuals of Caucasian origin (using a threshold of 6 standard deviations from mean Caucasian PC values) were kept for further analyses.

Imputation of new SNPs in the ICSC data-set was performed at the Michigan server (https://imputationserver.sph.umich.edu) using minimac4 and the HRC r1.1 population as reference. All those SNPs that showed above imputation quality threshold (R^2>0.30) and MAF >0.01 were retained for association analysis. Imputation of the SHCS cohort was performed using the Sanger Imputation Service (https://imputation.sanger.ac.uk) using EAGLE2 for phasing and Positional Burrows Wheeler Transform (PBWT) for imputation together with The 1000 Genomes Project phase 3 reference panel. Only SNPs with an imputation quality score (INFO > 0.8) and MAF > 0.01 were retained for later analyses.

Single locus association study and meta-analysis

Association analyses on each data set were carried out using genetic additive models. In these analyses, the results were adjusted by the first 4 PC vectors, age (continuous variable), sex and the elite controller condition using the logistic regression procedures included in PLINK 2.0 software. Combined data from both data-sets was analysed using the meta-analysis tool in Plink. In this meta-analysis fixed effect models were taken into account when no evidence of heterogeneity was found. Otherwise, random effects models were considered.

In all these studies, a GWAS significant p-value was stablished at 5×10^{-8},[45] whereas a p-value$<10^{-5}$ was considered as suggestive of statistical significance. These variants were annotated using the Variant Effect Predictor tool (https://www.ensembl.org/Homo_sapiens/Tools/VEP).[38] Plink was also used to estimate the genomic inflation factor (λ). The Software qqman R package (https://CRAN.R-project.org/package=qqman) was used for graphical representation of the GWAS single locus analysis results (Manhattan plot).

The estimation of individual GWAS statistical power was performed by the Episheet software (https://www.drugepi.org/dope/software#Episheet). Power estimation for the meta-GWAS was performed in R using the metapower package (https://rdrr.io/github/jasonwgriffin/metapower).

SNPs previously reported as associated with natural HIV-control, or related phenotypes, by mean of GWAS in Caucasian individuals were considered associated with the LTNP-C condition in our study if their effect direction was the same to that originally reported and if the p value was <0.05.

Gene-based association study and enrichment analyses

Gene-wise statistics were computed using the Magma software. This software takes into account physical distance and linkage disequilibrium between SNPs for detecting multi-marker effects.[46] These analyses used 50-kb upstream and downstream window around each gene in order to capture potential regulatory variants of these genes. For gene-based association analyses, the p_SNPwise_mean value calculated by the software was corrected by the number of genes analyzed. In this study the p-value threshold was stablished at 2.7×10^{-6}, whereas a $p < 10^{-4}$ was considered as suggestive of statistical significance. In these analyses, ranked genes that were included in the genetic region of other tops genes were not considered as independently associated with the LTNP-C phenotype.

Webgestalt software[42] (www.webgestalt.org) implemented in the R statistical package (WebgestaltR) was used for exploring enrichment in GO categories using top genes obtained from the gene-based association analyses. An overrepresentation analysis was performed using the database of Gene Ontology[43] (http://www.geneontology.org) for biological processes. Multiple testing correction was applied using the Benjamini–Hochberg method implemented in the software. We considered significant those processes with false discovery rate (FDR) p value <0.05.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed on each of the STAR Methods sections.