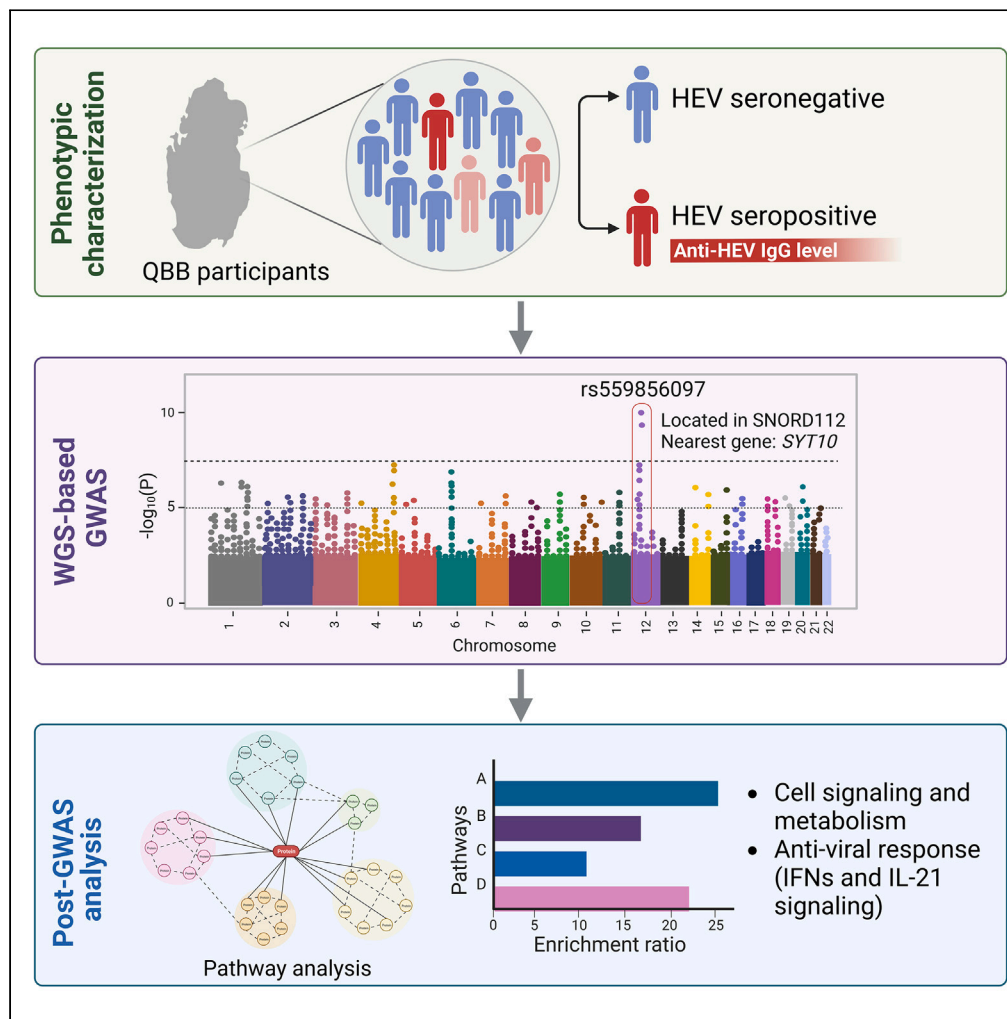


Article

Genome-wide association study identifies several loci for HEV seropositivity



Maria K. Smatti,
Yasser A. Al-Sarraj,
Omar Albagha,
Hadi M. Yassine

oalbagha@hbku.edu.qa (O.A.)
hyassine@qu.edu.qa (H.M.Y.)

Highlights

The impact of host genetics on HEV remains largely unknown

GWAS identified a locus (12p11.1, rs559856097) associated with HEV antibody level

GWAS genes were significantly enriched for anti-viral response pathways



Article

Genome-wide association study identifies several loci for HEV seropositivity

Maria K. Smatti,^{1,5} Yasser A. Al-Sarraj,^{1,2} Omar Albagha,^{1,3,*} and Hadi M. Yassine^{4,*}

SUMMARY

Hepatitis E viral (HEV) infection imposes a heavy global health burden. The variability in the prevalence of serological markers of HEV infection between different ethnic groups proposes a host genetic influence. Here, we report genetic polymorphisms associated with anti-HEV antibody positivity and level using binary- and quantitative-trait genome-wide association studies (GWAS) on a population from Qatar (n = 5829). We identified a region in 12p11.1 (lead SNP: rs559856097, allele: A, $p = 2.3 \times 10^{-10}$) significantly associated with anti-HEV antibodies level. This intergenic variant is located near SNORD112, a small nucleolar RNA (snoRNA). Additional gene-set and pathway enrichment analyses highlighted a strong correlation with anti-viral response-related pathways, including IFNs (alpha/beta) and interleukin-21 (IL-21) signaling. This is the first GWAS on the response to HEV infection. Further replication and functional experimentation are warranted to validate these findings.

INTRODUCTION

Hepatitis E virus (HEV) is a long-neglected RNA virus belonging to the *Hepeviridae* family. It is a major causative agent of acute viral hepatitis and jaundice cases worldwide.¹ HEV is an emerging zoonotic virus with a broad host range including humans, pigs, dogs, rhesus monkeys, and many others.² Hepatitis E is a serious public health concern especially in developing countries. The virus primarily causes acute infections in young adults, with a mortality rate reaching 2%. Infection with HEV is usually asymptomatic or self-limited in immunocompetent individuals, but can progress to chronicity and lead to fulminant hepatitis in immunocompromised patients and other risk groups such as pregnant women.³ High HEV-related morbidity and mortality rates are reported among pregnant women, infants <2 years old, and solid organ transplant recipients.⁴ Up to 20–25% of pregnant women can die if they get hepatitis E in the third trimester according to the WHO estimates.⁵

Globally, at least 20 million HEV infections occur every year, accounting for approximately 3.3 million symptomatic cases of acute illness. In a relatively recent study, it has been estimated that 110 million individuals have recent/current infection based on anti-HEV IgM antibody positivity, while 15 million carry an active infection according to the HEV RNA positivity.⁶ The prevalence of HEV varies between different geographical regions. Developed countries tend to have a lower HEV seroprevalence (5–20%) compared to developing countries, reaching almost 70% in HEV endemic regions.^{7–9} Similarly, variable rates of HEV were recorded in the Middle East and North Africa (MENA) region, ranging from 10% in Saudi Arabia and Tunisia to 90% in Sudan.^{10,11} In Qatar, a seroprevalence rate of approximately 21% was reported in blood donors, indicating a high HEV infection rate in the general population.¹² The diversity of the people residing in Qatar and the high influx of foreign expatriates contribute to this high prevalence. In Qatari nationals, a prevalence of 11.5% was reported.¹² Further, HEV IgM antibodies and viral RNA were detected in about 0.6%, supporting reported data from other countries about the risk of blood transfusion-associated HEV, particularly to pregnant women or immunocompromised patients.

Besides the environmental, behavioral, and socioeconomic factors, the variability in the prevalence of serological markers of HEV infection between different ethnic groups proposes a genetic basis for HEV susceptibility. A study in the US indicated that Hispanic ethnicity is associated with a higher seroprevalence of HEV compared to non-Hispanic blacks and non-Hispanic whites.¹³ Interestingly, it was also reported that non-Hispanic blacks carry $\epsilon 3$ and $\epsilon 4$ alleles in the *Apolipoprotein (APOE)* gene, which is associated with lower seroprevalence of HEV in this racial group, compared to non-Hispanic whites or Mexican Americans.¹⁴ In a replication study conducted on the Chinese Han population, Gu et al. reported that *APOE* variant was significantly associated with lower odds of anti-HEV seropositivity.¹⁵ Nonetheless, these findings require additional large-scale replication with particular attention to various confounders, especially since the role of *APOE* in HEV infection was not validated by *in vitro* experiments.^{16,17} A polymorphism in *Toll-like receptor 4 (TLR-4)* Thr399Ile (rs4986791) was also associated with HEV infection in an Indian cohort.¹⁸ Although these reports, to a limited extent, provided the first evidence that HEV infection could be

¹College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

²Qatar Genome Program, Qatar Foundation Research, Development and Innovation, Qatar Foundation, Doha, Qatar

³Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

⁴Biomedical Research Center, Qatar University, Doha, Qatar

⁵Lead contact

*Correspondence: oalbagha@hbku.edu.qa (O.A.), hyassine@qu.edu.qa (H.M.Y.)

<https://doi.org/10.1016/j.isci.2023.107586>



Table 1. HEV seroprevalence in the studied cohort

Category	No. (%)	Positive anti-HEV IgG	p value
Gender			
Male	2593 (43.2)	284 (10.9)	0.018
Female	3407 (56.8)	442 (13)	
Age groups			
15-30	1694 (28.2)	49 (2.9)	<0.0001
31-40	1624 (27.1)	62 (3.8)	
41-50	1378 (22.9)	144 (10.4)	
>50	1304 (21.7)	478 (36.7)	
Total	6000 (100)	733 (12.2)	

modulated by host genetics, they were limited by the modest sample sizes and experimental design. A very small number of SNPs (around 1–500) were genotyped and correlated to HEV. Until now, no genome-wide studies have analyzed variants related to HEV infection at a whole genome level. The impact of host genetics in HEV infection remains largely understudied compared to other hepatic viruses, including Hepatitis C and B, which were extensively studied in previous GWASs. Most importantly, genomic studies are ancestry specific and need to be conducted on different populations to identify ancestry-relevant risk variants. In this study, we conducted a GWAS to identify genomic variants in the genomes of 6000 Qatari individuals that associate with antibody response to HEV infection.

RESULTS

Seropositivity of anti-HEV IgG

Serological screening indicated a 12.2% (733/5998) positivity rate of anti-HEV IgG. HEV positivity was significantly higher among females (13%) compared to males (10.9%) ($p = 0.018$). In addition, a statistically significant positive correlation was found between anti-HEV IgG levels and age ($p < 0.0001$) (Table 1). Further classification of positive samples ($n = 733$) according to the ELISA reading indicated that most samples ($n = 497$) had an intermediate antibody level (groups 2 and 3), while 142 and 94 samples were considered strong and weak positives, respectively (Figure S1).

SNP-level association analysis of HEV seropositivity

The binary trait association, which correlates anti-HEV IgG positivity with genomic variants, revealed 18 suggestive loci ($p < 10^{-5}$). The strongest association was detected at 4q35.1 locus, where two SNPs (rs10002421 and rs4861528) showed suggestive evidence of association with HEV status ($p = 2.2 \times 10^{-7}$ and $p = 8.2 \times 10^{-6}$, respectively). The lead-independent SNP at this locus is rs10002421 (effect allele: G). Another association signal was also detected at chromosome 2, precisely at 2q32.2 (top lead SNP: rs7608839, $p = 2.67 \times 10^{-7}$), in addition to rs60915953, which was an independently associated variant with a p value of 1.05×10^{-6} . The genomic locus 3q26.32 included two independent SNPs, rs143844407 (top lead SNP, $p = 4.73 \times 10^{-6}$) and rs190707108 ($p = 9.1 \times 10^{-6}$). Remarkably, the highest number of GWAS-tagged candidate SNPs was found in the genomic locus 6p12.1 (top lead SNP: rs12176566, $p = 4.34 \times 10^{-6}$, with 101 tagged SNPs). Moreover, one of the top SNPs of this association was derived by a single variant on chromosome 8 (rs766747207 at q24.21, $p = 3.33 \times 10^{-7}$). Similarly, we observed a relatively high association signal on chromosome 12 at p11.1 from two independent lead SNPs: rs559856097 and rs77221465 ($p = 7.16 \times 10^{-7}$ and $p = 8.92 \times 10^{-6}$, respectively). The q24.31 locus on chromosome 12 had also shown other suggestive association signals, with rs112973617 being the top variant ($p = 1.1 \times 10^{-6}$). Other variants at 15q25.1 (rs150987782), 7q34 (rs116867893), 7q33 (rs4728340), 20p12.3 (rs59491562), and 11q23.3 (rs756724039) showed suggestive evidence of association with HEV seropositivity (p value $< 5 \times 10^{-6}$). Moreover, several loci with a marginal significant associations ($p < 1 \times 10^{-5}$) were detected at different chromosomes (Table 2; Figure 1).

SNP-level association analysis of quantitative HEV seropositivity

Genomic association with anti-HEV antibody level identified a single locus at chromosome 12 (p11.1) with genome-wide significance ($p < 5 \times 10^{-8}$). Three SNPs were detected at this locus, rs559856097 (the top lead independent SNP, $p = 2.3 \times 10^{-10}$), in addition to rs374241700 and rs377585811 ($p = 2.3 \times 10^{-10}$ and $p = 1.09 \times 10^{-9}$, respectively). Allele A of this lead variant (rs559856097) showed a negative correlation with anti-HEV antibody levels (Beta = -0.28). Importantly, SNPs at this locus were suggestively significant in the binary trait analysis, where rs559856097 was also detected as a lead SNP. Moreover, 16 unique candidate SNPs in LD with the independent significant SNP (rs559856097) were identified in this region. Of those, two SNPs were detected from this GWAS study (rs371520150 and rs372352546), while the remaining ones are known SNPs in the 1000G data (1KG/Phase3 reference panel) (Table S1; Figure 2).

Multiple other loci harbored variants close to the GWAS significance ($p < 5 \times 10^{-7}$), of which the majority have also been detected in the binary analysis. This includes the high signals observed at chromosome 4q35.1 locus (lead SNP: rs10002421, $p = 6.08 \times 10^{-8}$), and chromosome 6p12.1 (rs12176566, $p = 1.44 \times 10^{-7}$). The latter demonstrated the highest number of tagged SNPs ($n = 101$) compared to all other

Table 2. Top and independent SNPs at all genomic loci associated with HEV seropositivity binary trait

No	rsID	Chr	Pos	A1	A2 (effect)	AF_Allele2	AF (Cases)	AF (Controls)	Beta	Odds ratio	p-value	Nearest gene	Type of gene
1	rs7608839	2	190633678	T	C	0.82	0.77	0.83	-0.40	0.67	2.67E-07	NAB1	protein coding
2	rs60915953	2	190466578	C	G	0.85	0.80	0.85	-0.40	0.67	1.05E-06	MFSD6	protein coding
3	rs13413799	2	137888729	T	C	0.99	1.00	0.98	1.16	3.20	9.12E-06	AC020601.1	lincRNA
4	rs143844407	3	176225247	T	A	0.97	0.95	0.98	-0.82	0.44	4.73E-06	RP11-78E6.1	pseudogene
5	rs190707108	3	176213819	G	C	0.97	0.94	0.97	-0.74	0.48	9.18E-06	RP11-78E6.1	pseudogene
6	rs10002421	4	182853949	A	G	0.81	0.75	0.81	-0.39	0.68	2.22E-07	AC114798.1	miRNA
7	rs4075197	4	64697960	A	G	0.51	0.55	0.50	0.27	1.31	8.21E-06	RP11-158O16.1	pseudogene
8	rs793985	4	13926469	A	G	0.99	0.97	0.99	-1.06	0.35	8.24E-06	LINC01182	lincRNA
9	rs12176566	6	55741580	C	T	0.94	0.91	0.94	-0.56	0.57	4.34E-06	BMP5	protein coding
10	rs116867893	7	138958412	A	T	0.99	0.97	0.99	-1.29	0.28	2.35E-06	KIAA1549	protein coding
11	rs4728340	7	134850946	A	T	0.75	0.70	0.76	-0.32	0.72	3.76E-06	CALD1	protein coding
12	rs766747207	8	128517347	T	A	0.99	0.97	0.99	-1.36	0.26	3.33E-07	LINC00824	lincRNA
13	rs756724039	11	116068428	A	G	0.98	0.96	0.98	-0.93	0.39	4.24E-06	RPL5	pseudogene
14	rs238929	11	114101717	G	A	0.53	0.49	0.54	-0.27	0.76	5.91E-06	ZBTB16	protein coding
15	rs74355568	11	114453338	T	A	0.93	0.90	0.94	-0.52	0.59	9.33E-06	RP11-212D19.5	pseudogene
16	rs559856097	12	33204614	G	A	0.99	0.97	0.99	-1.37	0.25	7.16E-07	SNORD112	snoRNA
17	rs112973617	12	124927340	G	T	0.98	0.96	0.98	-0.97	0.38	1.11E-06	RPL22P19	pseudogene
18	rs77221465	12	33946298	T	C	0.99	0.98	0.99	-1.31	0.27	8.92E-06	RNU6-472P	snRNA
19	rs67692894	14	49337833	G	A	0.81	0.76	0.81	-0.34	0.71	6.71E-06	RP11-326E7.1	pseudogene
20	rs150987782	15	79305805	T	A	0.99	0.98	0.99	-1.50	0.22	1.49E-06	TMED3	protein coding
21	rs1609687	16	81402544	T	C	0.83	0.80	0.84	-0.37	0.69	7.01E-06	GAN	protein coding
22	rs190110595	19	44362376	C	T	0.98	0.99	0.98	0.92	2.51	7.08E-06	ZNF112:CTC-512J12.6	protein coding
23	rs59491562	20	5893391	A	AC	0.99	0.97	0.99	-1.25	0.29	3.18E-06	RNU1-55P	snRNA
24	rs73593909	20	7714039	T	A	0.82	0.78	0.83	-0.35	0.70	9.80E-06	AL031653.1	miRNA
25	rs28502553	21	41822335	C	G	0.77	0.72	0.78	-0.31	0.73	8.50E-06	PRDM15	protein coding

Chr: Chromosome number; Pos: Genomic position in GRCh38 hg38; A1: Allele1; A2, Allele2 that is the effect allele.AF: Allelic frequency.

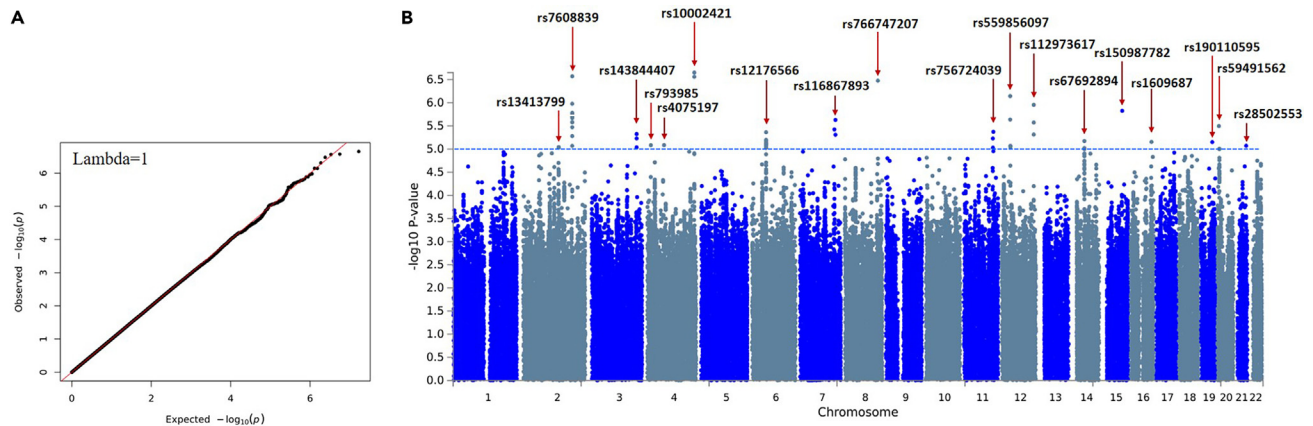


Figure 1. Quantile-Quantile (QQ) and Manhattan plots of the GWAS on HEV seropositivity binary trait

(A) QQ plot of the association test statistics. The regression line (in red) represents the expected distribution of p values under the null hypothesis. Lambda value is the genomic inflation factor calculated as the ratio of the median of the observed distribution of the test statistic to the expected values.

(B) Observed $-\log_{10}$ p values (y axis) are shown for all SNPs on each autosomal chromosome (x axis). The blue dotted line indicates suggestive significant of p value $<1 \times 10^{-5}$. The top SNPs at each suggestive genomic loci are labeled with rs identifiers.

associated loci, in a similar pattern to that observed in the binary GWAS. Additionally, 61 suggestively significant SNPs ($p < 1 \times 10^{-5}$) were found at 56 genomic loci (Table S2).

Gene mapping of the top GWAS SNPs

SNP annotation according to the physical position on the genome (positional mapping) at a 100 kb genomic window indicated that our GWAS SNPs are located within or near protein-coding genes, long intergenic non-coding RNAs (lincRNAs), microRNAs (miRNAs), pseudogenes, small nucleolar RNAs (snoRNAs), or small nuclear RNAs (snRNAs). The top hit of the binary trait association, rs10002421, is an intergenic variant located around 27.5 kb from AC114798.1 miRNA, and 64 kb from *DCTD* protein-coding gene. Additionally, the two SNPs at 2q32.2 were both linked to protein-coding genes. The variant rs7608839 is 13 kb close to *NAB1*, while rs60915953 is located within *MFS5D6* genes. Other intronic SNPs positioned at protein-coding genes are rs116867893 and rs4728340 on chromosome 7, which were at *KIAA1549* and *CALD1*, respectively. Also, rs238929 is located within *ZBTB16*, rs190110595 in two overlapping genes, *ZNF112* and *CTC-512J12.6*, and rs28502553 is located in the *PRDM15* gene. Within a distance of 13 kb or less from a protein-coding gene, rs150987782, rs1609687, and rs12176566 intergenic variants were identified near *TMED3*, *GAN*, and *BMP5* genes, respectively (Table 2). The regional plots for all suggestive loci are presented in Figure S2.

On the other hand, rs559856097, the top SNP associated with quantitative HEV antibody response, is an intergenic variant located in *SNORD112*, a snoRNA on 12p11.1. The nearest protein-coding gene to this variant is *SYT10*, located after approximately 170 kb. The two independent SNPs at this locus, rs371520150 and rs372352546, were mapped to two RP11-313F23.4 (lincRNA) and RP13-359K18.1 (pseudogene), respectively (Figure 3).

In addition to positional mapping, eQTL and chromatin interaction mapping were performed. The complete list of mapped genes to all GWAS SNPs and SNPs in LD is presented in Tables S3 and S4. A list of 552 and 68 genes were mapped in the binary and quantitative GWASs, respectively. While no eQTLs were identified from the quantitative HEV GWAS, 26 out of 552 genes were mapped by eQTLs as shown in Table S5.

Gene-set based analysis

To further investigate the mechanisms underlying genetic signals, we performed gene-set-based analysis. We utilized the FUMA tool based on the Multi-marker Analysis of GenoMic Annotation (MAGMA). Initially, analysis of the binary trait GWAS genes revealed that among the top gene sets identified by MAGMA are inflammatory response, regulation SMAD2/3 signaling, antigen response, and PI3K/ErbB pathway gene sets (Table S6). Additionally, the top genes identified by MAGMA analysis included several genes involved in the immune system pathways, specifically, the innate immune response. This includes *IL2RA*, *KPNA7*, *CFHR4*, *KLHL3*, and *CFHR2* (Table S7). Moreover, these genes had a distinct expression in the liver compared to other tissues (Figure 4A). However, none of the identified pathways or genes had a false discovery rate (FDR) below 0.05. On the other hand, analysis of the quantitative GWAS mapped genes by MAGMA showed that the top gene sets are those related to PI3K/ErbB pathway, followed by nuclear import signal receptor activity pathway, and IRF3-mediated induction of type I IFN pathway (Table S8). While the first two pathways are generally related to cell signaling and general cell metabolism functions, the latter (IRF3 pathway) is one of the central pathways in regulating type 1-interferon during bacterial and viral infections. Additionally, among the top genes identified by MAGMA are genes closely related to interferon, class I MHC-mediated antigen processing and presentation, and complement pathways. This includes *MAP3K3*, *KLHL3*, *CFHR4*, and *CFHR2* genes. The list of the top ten MAGMA genes is presented in Table S9, while the tissue expression pattern is demonstrated in Figure 4B.

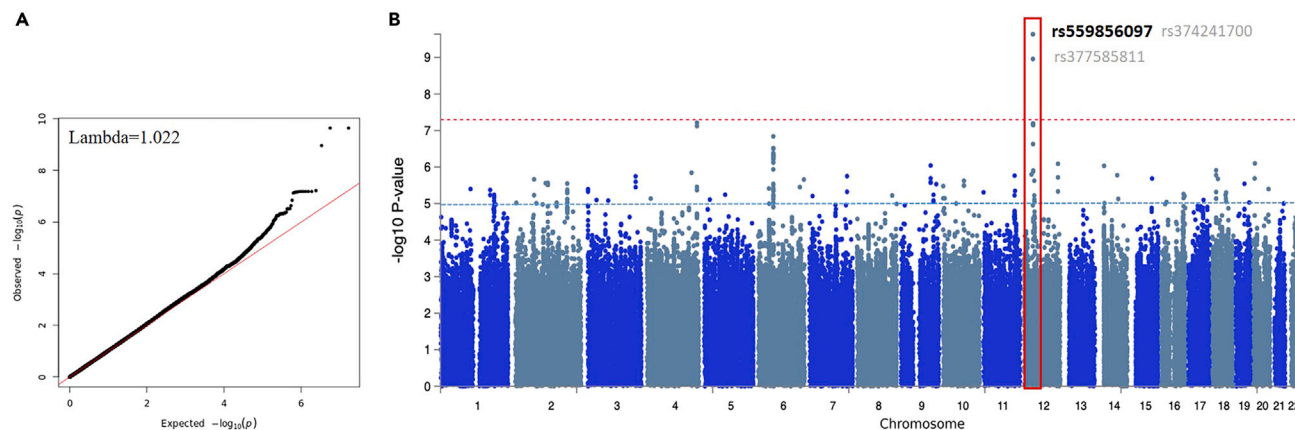


Figure 2. Quantile-Quantile (QQ) and Manhattan plots of the GWAS on anti-HEV antibodies level quantitative trait

(A) QQ plot of the association test statistics. The regression line (in red) represents the expected distribution of p values under the null hypothesis. Lambda value is the genomic inflation factor calculated as the ratio of the median of the observed distribution of the test statistic to the expected values.

(B) Observed $-\log_{10}$ p values (y axis) are shown for all SNPs on each autosomal chromosome (x axis). The red dotted line indicates GWAS significant of p value $< 1 \times 10^{-8}$, while the black line indicates suggestive significant at p value $< 1 \times 10^{-5}$. The GWAS significant genomic locus (12p11.1) is bordered with red, and the top SNPs are labeled with the rs identifiers. The top lead SNP is in black, while the two SNPs in the locus (dependent) are in light gray.

Genes enrichment analysis

To further analyze the biological roles of the mapped genes in our GWAS, we utilized WebGestalt, a functional enrichment analysis tool.¹⁹ We performed one of the main well-established methods for enrichment analysis, the over-representation analysis (ORA), in addition to gene ontology (GO) analysis. GO annotates genes based on cellular components (CC), biological processes (BP), and molecular functions (MF). GO enrichment analysis of the candidate protein-coding genes mapped from the binary GWAS ($n = 189$ genes mapped from 18 genomic loci), revealed several significantly enriched pathways, mostly related to metabolism and cell adhesion ($p < 1 \times 10^{-6}$, FDR < 0.05) (Figure 5A). The ORA pathway analysis on the Reactome database showed that the most enriched pathways were those related to cell surface interactions at the vascular wall and Ca^{2+} activated K^{+} channels ($p < 6 \times 10^{-5}$, FDR = 0.054). The interleukin-21 signaling pathway was also one of the top enriched gene sets, with a suggestive significance (p value = 0.004, FDR = 0.8). The top enriched pathways of ORA Reactome analysis are presented in Figure 5B.

In the quantitative GWAS genes enrichment analysis, 441 candidate genes covering the significant and suggestive GWAS loci were included. Of these, 379 genes were unambiguously mapped to the database. GO pathway analysis revealed multiple significantly enriched pathways (Figure 6A). While various gene sets were related to metabolic pathways, type I interferon signaling and cellular response signaling pathways were again significantly enriched (Both p value and FDR < 0.05). Similarly, Reactome pathway analysis showed that the Interferon alpha/beta signaling pathway is enriched substantially ($p = 1.4 \times 10^{-5}$, FDR = 0.025), compared to interleukin-36, compliment, and metabolic pathways, which were among the top pathways but with a suggestive enrichment ($p < 0.05$, FDR > 0.05) (Figure 6B).

Considering that our GWAS is based on the phenotypic characterization (antibody response) of cases based on their immune status with regard to HEV, we further analyzed the results focusing exclusively on immune genes and pathways. Initially, candidate protein-coding gene sets identified from both binary and quantitative GWASs were combined. This resulted in a list of 537 unique genes included in the InnateDB database search. Of these, 43 genes were directly related to the immune response and were therefore subjected to further gene enrichment analysis. Starting with the GO gene sets, we found a highly significant enrichment in genes related to the innate immune response particularly ($p = 1. \times 10^{-16}$, FDR = 1.36×10^{-12}). Other immune-related pathways were also markedly enriched, including those pertaining to viral processes (Figure 7A). Reactome pathway analysis indicated that our candidate genes are enriched for several interferon and cytokine signaling, as well as anti-viral mechanisms pathways ($p < 1 \times 10^{-4}$, FDR < 0.01) (Figure 7B). Besides, we utilized the GLAD4U database to explore the pattern of gene enrichment in disease-related pathways. The most significant enrichment was found in pathways predominantly related to viral infections (FDR = 2.9×10^{-9}). Moreover, significant enrichment of pathways involving different infections was observed, including dengue, influenza, and sexually transmitted diseases. Notably, hepatitis-associated pathways were also enriched with high statistical significance (FDR < 0.00005) (Figure 7C).

DISCUSSION

The role of host genetic factors in the susceptibility or response to infections is increasingly recognized. Nevertheless, genome-wide assessment of human genetic determinants of antibody responses to many viral infections is still scarce, including the response to HEV. In this present study, we mapped genetic loci associated with HEV infection in a 6,000 well-characterized cohort. This was accompanied by gene network analysis and gene overrepresentation to explore possible underlying molecular mechanisms.

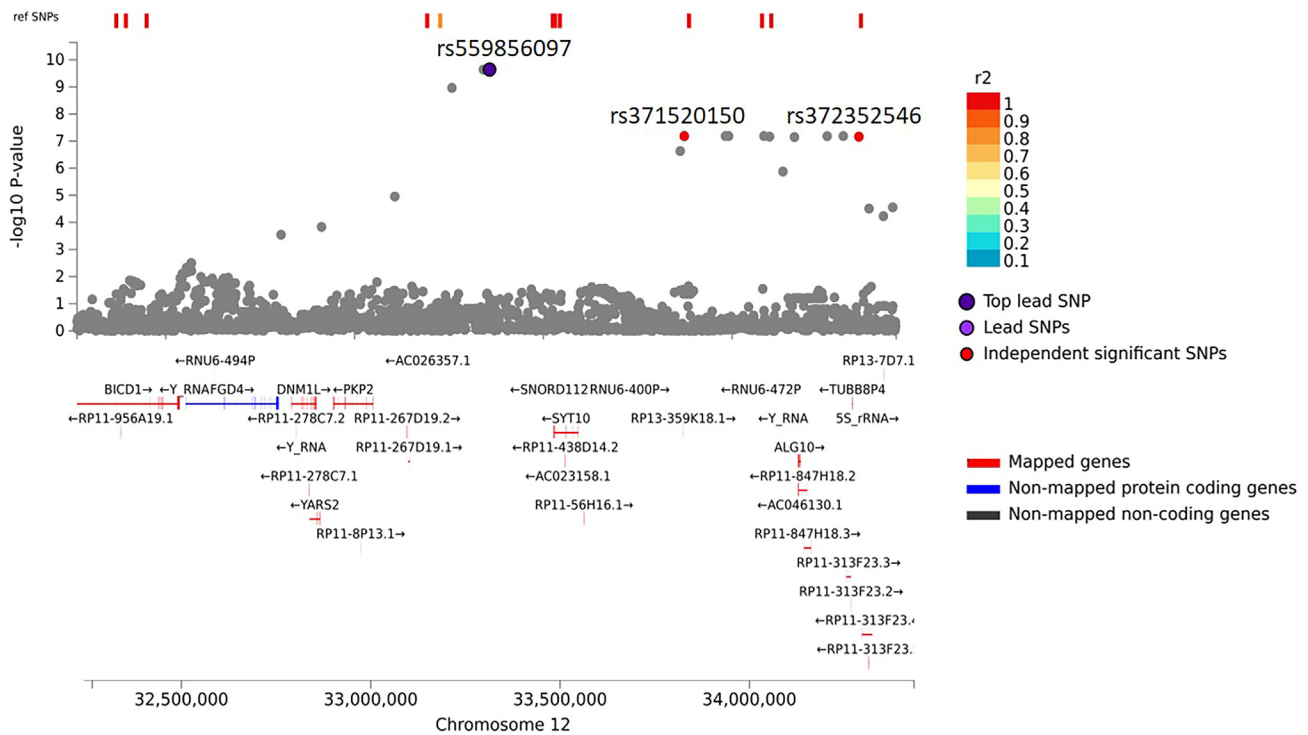


Figure 3. Regional association plot for the top signal at rs559856097

Genes were mapped to this locus by positional, eQTLs, and chromatin interaction mapping using FUMA platform (<https://fuma.ctglab.nl/>). The top lead and the independent significant SNPs are labeled with rs identifier.

We initially determined the HEV serostatus at a 12.2% positivity rate among the adult Qatari population. This prevalence is in line with previous reports, where an 11.5% seropositivity rate was reported in Qatari blood donors,¹² and neighboring counties such as Saudi Arabia (11% in apparently healthy population). However, higher seroprevalence of anti-HEV antibodies was reported in other Middle East and North Africa (MENA) region, such as Egypt (47% in pregnant women) and Sudan 90%.^{10,11} We based our phenotypic classification on ELISA assay because of its sensitivity in determining the immune status. Importantly, the commercial kit (MP Diagnostics HEV ELISA) that was utilized in the study had a recorded sensitivity of 99.2%. Following the serostatus determination, we used two GWAS approaches, a binary and a quantitative phenotype, to pinpoint genomic regions associated with HEV seropositivity or antibodies level, respectively.

Prior exposure to HEV evidenced by a positive anti-HEV IgG test result was suggestively associated with 18 genomic loci. Most of the signals at these loci map to non-coding regions (15 out of 25 variants), which is expected for whole-genome association analyses.²⁰ Additionally, ten variants were intronic or intergenic and mapped to a protein-coding gene. The nearest protein-coding genes to intergenic variants were *NAB1*, *BMP5*, *TMED3*, and *GAN*. On the other hand, intronic variants within protein-coding genes were found in *MFSD6*, *KIAA1549*, *CALD1*, *ZBTB16*, *ZNF112*, and *PRDM15* genes. Notably, most of these genes have a functionally validated or predicted role in regulating the immune response. For instance, *NAB1* is involved in the regulation of *IFNGR1*, the receptor for type II interferon ($IFN\gamma$), and thus, associated with susceptibility to various infections. Interestingly, previous studies demonstrated that polymorphisms in *ifngr1* promoter correlated with hepatitis B infection susceptibility.²¹ Hence, it is possible that variants of *NAB1*, a regulator of *IFNGR1*, could also play a role in susceptibility and/or immune response to HEV. *MFSD6*, *GAN*, and *ZBTB16* genes were also linked to class I MHC (MHC-1) antigen processing and presentation. In fact, the role of $CD8^+$ T cells in combating viral infections is well studied. These immune cells function by recognizing viral-derived peptides presented by MHC-1 on the surface of infected cells.²² Manipulation of the MHC-1 presentation pathway has been found to limit the MHC-1 expression, reduce the ability of $CD8^+$ T to recognize viral peptides, and thus, impair viral clearance.²³ While the exact role of *MFSD6*, *GAN*, and *ZBTB16* genes in MHC-1 pathway is still not clear, mutations in genes involved in any process of MHC-1 expression or antigen processing and presentation pathways could alter the subsequent immune responses. Given the important role of MHC-1 molecules in detecting virally infected cytotoxic T lymphocytes (CTLs), multiple viruses evolved proteins that inhibit the MHC-1 pathway through diverse mechanisms for host immune impairment and viral evasion.²⁴

Two ZNF protein-coding genes were also prioritized in our GWAS, *PRDM15*, and *ZNF112*. The ZNF protein family is known for its essential function in regulating immune responses at transcriptional and post-transcriptional levels.²⁵ *PRDM15* gene, tagged by rs28502553 intronic SNP, is part of the immune response cluster as classified by the Human Protein Atlas. It is expressed in immune cells, overexpressed in B-cell lymphomas, and plays an essential role in B-cell lymphomagenesis.²⁶ *ZNF112* gene, on the other hand, has been linked to immunoglobulin resistance in Kawasaki Disease (KD) possibly due to its regulatory contribution to inflammation.²⁷ Importantly, KD could be triggered

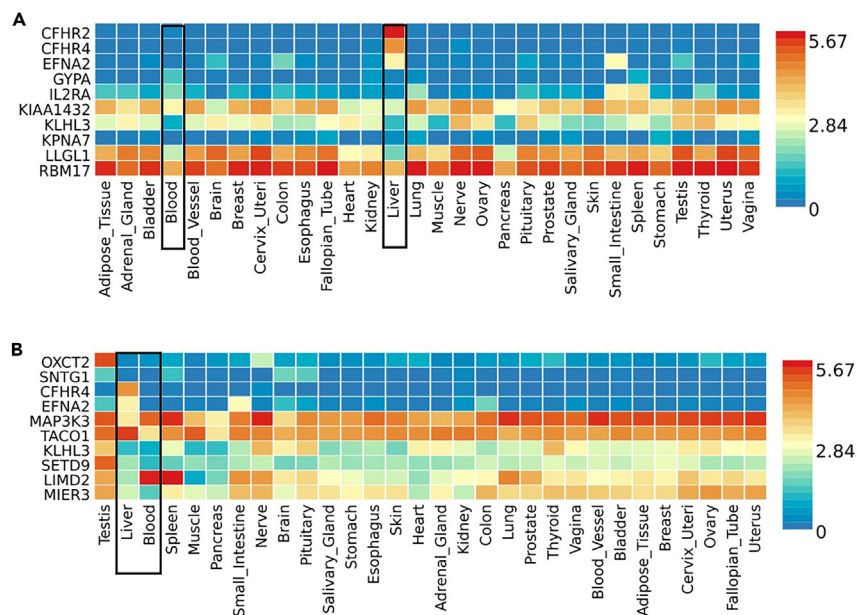


Figure 4. Heatmaps of the average gene expression per tissue (GTEx v8 30 general tissue types) for the top genes (highest p value) identified by MAGMA analysis

(A) Top 10 MAGMA genes from the binary trait GWAS.

(B) Top 10 MAGMA genes from the quantitative trait GWAS. The expression in blood and liver tissues are particularly bordered in black. FUMA platform (<https://fuma.ctglab.nl/>) was used for gene annotation and representation.

by various infectious agents in genetically susceptible individuals.²⁸ Still, the role of *PRDM15* and *ZNF112* in the immune response to infections has never been studied. Therefore, further studies are needed before we understand the mechanistic role of polymorphisms in *PRDM15* and *ZNF112* on HEV antibody response.

BMP5 and *TMED3*, the nearest protein-coding genes to rs12176566 and rs150987782, also have immune-related functions. While *BMP5* acts primarily as a growth factor and is linked to cytokine activity, *TMED3* belongs to the *TMED* family, an important player in innate immune signaling.²⁹ Besides the involvement in immunity, other genes with diverse functions were also tagged by several suggestive SNPs. This includes *KIAA1549* and *CALD1*, which are associated with brain tumors and regulation of muscle contraction, respectively.

We noticed multiple signals from SNPs associated with HEV seropositivity at non-coding RNAs (ncRNAs), specifically the long intergenic ncRNAs. Generally, this class of non-coding RNAs has been recently described as critical regulators in anti-viral responses and virus-host interactions.³⁰ Moreover, in a recent study on HBV integration sites, it was reported that the HBV genome was highly integrated into several non-coding RNAs, including LINC00824.³¹ This particular lincRNA was tagged by rs766747207, one of the top SNPs in our binary HEV GWAS. Although HBV is a DNA virus with the potential basis for host genome integration that resembles retroviral integration, non-retroviral RNA viruses can incidentally integrate.³² Nevertheless, the definitive role of LINC00824 and other ncRNAs identified in our study in HEV infection necessitates additional research.

In addition to long non-coding RNAs, we identified four SNPs in short non-coding RNAs regions, of which two were located at miRNAs (AC114798.1 and AL031653.1). Host-encoded miRNAs are known for their ability to modulate the host-virus interplay, given their pivotal role in various cellular processes. These miRNAs could either promote or repress viral replication by suppressing or upregulating the IFN- α/β signaling pathway.³³ In fact, the link between cellular miRNAs and viruses emerged after the observation that HCV infection affects the expression of miRNA-122, which enhances HCV replication and the stability of its RNA.^{34,35} In HEV infection, several miRNAs (such as: miR-450b, miR-125b-5p, miR-192-5p, and miR-99a) have been identified and associated with HEV progression to chronicity or acute liver failure.^{36,37} However, until now, there have been no studies on the role of AC114798.1 and AL031653.1 miRNAs in HEV or any other viral infection.

Next, we investigated whether host genetic variants contribute to the level of antibodies in HEV infection. To count for the possible differences related to the time of exposure due to demographic differences, we adjusted our model for age and gender before conducting the quantitative GWAS. We identified a region in 12p11.1 (Lead SNP: rs559856097) significantly associated with anti-HEV levels. Carrying A allele negatively correlated with the level of anti-HEV antibodies. This strong GWAS signal came from a variant near SNORD112, a snoRNA. SNORDs are a highly expressed class of non-coding RNAs that play a prominent role in rRNA modification and act as a scaffold for protein complexes.³⁸ Solid evidence has accumulated regarding the clinical role of SNORDs in disease outcomes, including viral infections. These nucleolar RNAs are found to interact with RNA viruses. In particular, a specific C/D box snoRNAs is needed for optimal viral replication, as observed with dengue fever virus, influenza A virus, human rhinovirus 16, herpes simplex virus 2, and others.³⁹ In a similar way, C/D box 126 (SNORD126) has been recently reported to enhance HCV replication by activating the PI3K-AKT signaling pathway.⁴⁰ However,

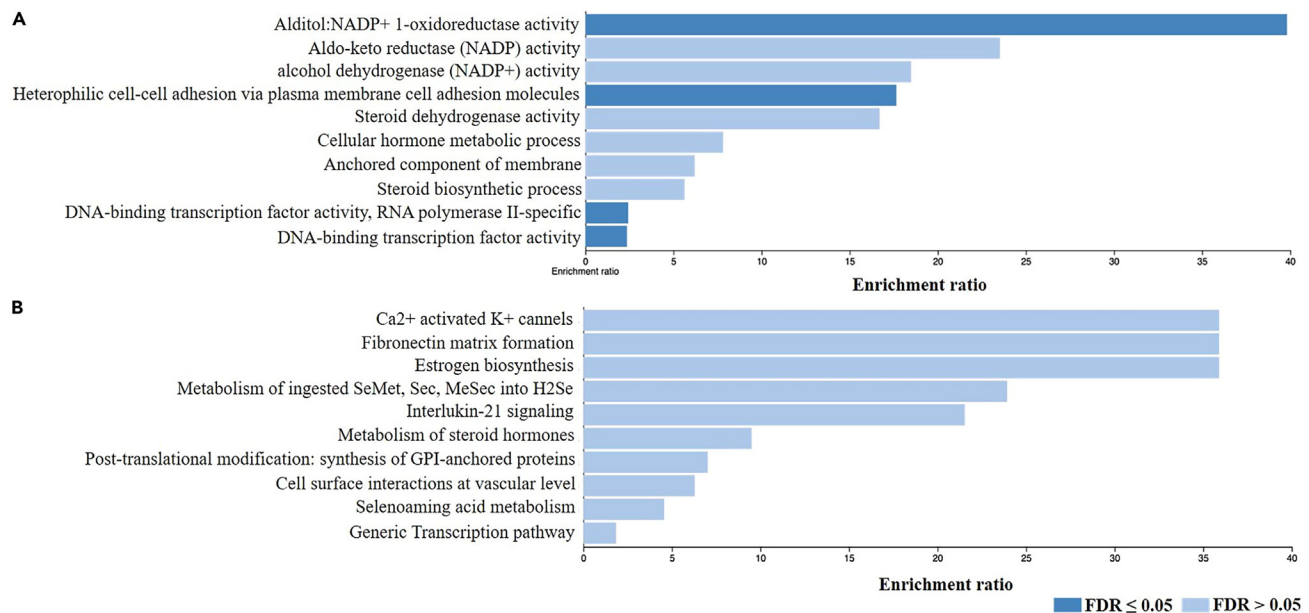


Figure 5. Gene set enrichment analysis of the binary trait HEV GWAS using WebGestalt

(A) GO enrichment analysis on the candidate protein coding genes mapped by FUME on the binary GWAS gene set (n = 189 genes mapped from 18 genomic loci).

(B) ORA pathway analysis using Reactome database on the same set of genes (n = 189). The top ten enriched pathways are presented in each bar graph. Categories with significant enrichment (FDR < 0.05) are colored in dark blue.

SNORD112, is one of the orphan snoRNAs that exist with unknown complementarity to any cellular RNA and, thus, unknown function.⁴¹ Overexpression of SNORD112, in addition to SNORD113 and SNORD114, was documented in acute promyelocytic leukemia (APL).⁴¹ However, the precise role of SNORD112 and its involvement in viral infections remains open for investigation. The nearest protein-coding gene to the lead SNP (rs559856097) at 12p11.1 was *SYT10* gene. This gene acts as a calcium sensor needed explicitly for the Ca(2+)- dependent exocytosis of secretory vesicles containing insulin-like growth factor 1 (IGF1) in neurons of the olfactory bulb.⁴²

Despite the diverse functions of the protein-coding genes mapped by the binary and quantitative GWASs, most of these genes carry polymorphisms associated with liver diseases and viral infections, as demonstrated in the GWASs PhenoScanner database. For instance, variants in the *BMP5* gene have been linked to acute respiratory infections, while *TMED3* gene variants were strongly linked to HBV. On the other hand, *MFSD6*, *CALD1*, *GAN*, and *PRDM15* genes were associated with UTIs. Theoretically, genetic variants in common innate and adaptive immune-related genes could interfere with the host response to different viruses regardless of their site of infection. On the other hand, liver cirrhosis, fatty liver, and hepatic failure phenotypes were associated with variants in *NAB1*, *CALD1*, *ZBTB16*, *GAN*, *TMED3*, and/or *PRDM15* genes. Of note, the progression of HEV infection to chronicity, leading to liver damage and cirrhosis has been previously reported.⁴³

In order to employ a more comprehensive analysis of genetic markers that simultaneously contribute to a complex phenotype, we applied MAGMA gene and gene-set analyses. Several gene sets prioritized by MAGMA were closely related to the immune response processes. This included genes of IRF3-mediated induction of type I IFN and SMAD2/3 signaling, which are essential pathways in regulating type 1 interferons during bacterial and viral infections.^{44,45} In fact, treatment by different types of interferons (type I, II, and III IFNs) demonstrated an anti-viral response to HEV infection.⁴⁶ Moreover, HEV and other hepatitis viruses express molecules that target signal transductions across IFN pathways as a defensive mechanism to block the production of IFNs and impair the host anti-viral response.^{47–49} Interestingly, specific polymorphisms in *IFNλ4* gene were found to correlate with anti-HEV IgG positivity among hemodialysis patients.⁵⁰ It is, therefore, not surprising that variants in genes involved in IFN signaling pathways could modulate the susceptibility, seropositivity, or response to HEV infection.

Further, we performed an overrepresentation enrichment analysis using several databases to identify pathways whose perturbation may contribute to the HEV infection. Remarkably, in addition to involvement in metabolism, our candidate genes were significantly enriched for interferons (alpha/beta) signaling and cellular response to IFN-1 pathways. This again highlights the possible key role of IFNs in HEV response. Furthermore, IL-21 signaling pathway was suggestively enriched in the quantitative GWAS candidate gene set. IL-21 is a major cytokine produced by T follicular helper (Tfh) cells. Along with other cytokines, IL-21 modulates antibody production by T-cell-dependent activated B-cells. It also enhances somatic hypermutation and production of high-affinity antibodies from plasma cells.⁵¹ Most importantly, the observed enrichment in IFNs and antibodies production pathways strengthens the validity of our GWAS findings since it was mainly designed to detect variants associated with antibodies positivity or level.

As a complementary approach to explore enriched pathways exclusively related to immunity, we further performed enrichment analysis on genes directly involved in the immune response. Across a list of 43 immune genes filtered from our candidate genes list, we identified highly

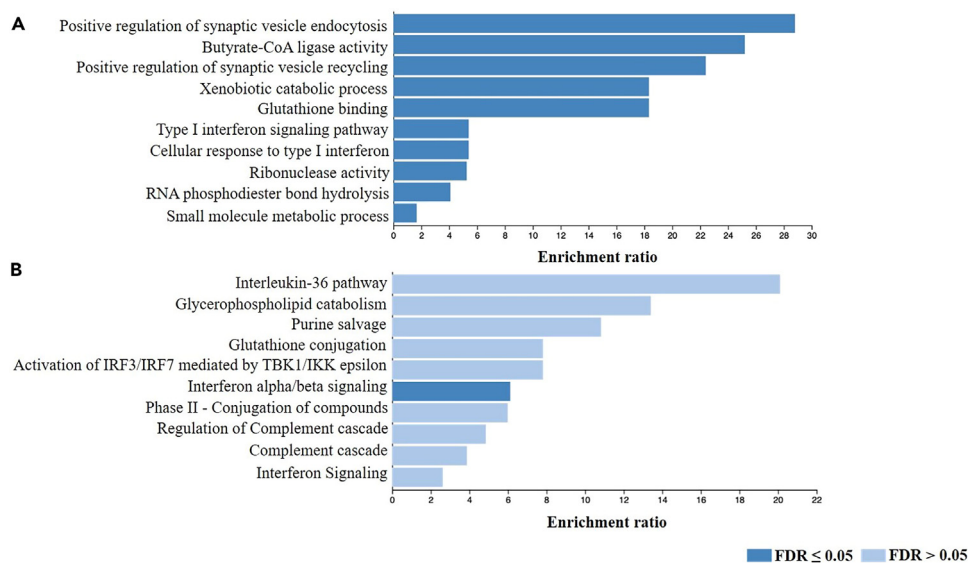


Figure 6. Gene set enrichment analysis of the quantitative trait HEV GWAS using WebGestalt

(A) GO enrichment analysis on the candidate protein coding genes mapped by FUME on quantitative GWAS gene set (n = 441 candidate genes covering the significant and suggestive GWAS loci).

(B) ORA pathway analysis using Reactome database on the same set of genes (n = 441). The top ten enriched pathways are presented in each bar graph. Categories with significant enrichment (FDR < 0.05) are colored in dark blue.

enriched gene sets contributing primarily to anti-viral responses. This included anti-viral responses to various hepatic (e.g., HCV and HBV) and non-hepatic viruses (e.g., measles, influenza). Notably, this points out that polymorphisms in genes related to common anti-viral signaling pathways could modulate the response to distinct viruses from different groups. This also highlights the need for more genomic studies and, most importantly, meta-analyses combining GWAS findings on different infectious diseases. Determining common variants that disturb overlapping anti-viral response pathways could provide potential preventive, prognostic, and therapeutic targets for genetically predisposed individuals.

The strength of this study comes from the uniqueness of the Qatari population, which provided an opportunity to detect variants that contribute to HEV infection in genomes representing the Qatari and Arab regions, an extremely underrepresented population in genomic studies. A recent report on the genetic composition of the Qatari population included the same cohort as in the present study and identified six major ancestries: General Arabs (38%), Peninsular Arabs (17%), Arabs of Western Eurasia and Persia (22%), South Asian Arabs (1%), African Arabs (3%), and Admixed Arabs (19%).⁵² The richness in the genomic diversity of the population makes it a perfect representative of the whole Middle Eastern region. In fact, the unique population structure could partially explain why our results did not replicate previous reports on the role of *TLR-4* and *APOE* polymorphisms in HEV infection in non-Hispanic whites, non-Hispanic blacks, Mexican Americans, Indians, and Chinese populations.^{14,15,18} Zhang et al. study found that non-Hispanic blacks were the only population that showed an association between anti-HEV seropositivity and *APOE* variants, compared to other studied populations.¹⁴ Interestingly, *APOE* ε3 and ε4 variants were found to be significantly associated with protection against HEV infection in non-Hispanic blacks, who had the lowest seroprevalence of anti-HEV (15.3%) compared to other groups (>20%).¹⁴ This emphasizes the importance of ancestry-specific genomic studies to understand infections' susceptibility and pathogenesis at a population level. Additionally, this study was based on whole genome sequencing data rather than SNP array or targeted genotyping, and thus, was able to detect variants beyond the protein-coding regions. This allowed for a comprehensive representation of regulatory variants that associate with RNA expression (eQTL variants), or chromatin interaction and, consequently, improved our candidate gene mapping and pathway analysis. In addition, the consistency in the results between the GWAS of the binary and quantitative traits further confirms our findings. Most of the suggestive loci identified in the binary GWAS were also detected in the quantitative GWAS. Although our binary GWAS did not succeed in catching significant hits, additional categorization of samples based on anti-HEV antibody levels seemed to improve the phenotype precision, resulting in GWAS significant associations.

In conclusion, we reported several host genetic variants associated with anti-HEV antibody response. This study provides new evidence on the role of human genetic variation in HEV infection and emphasizes the importance of further replication and functional validation.

Limitations of the study

There are several main limitations in this study. First, our classification of cases and controls was based on anti-HEV IgG positivity only. This could have led to the misclassification of individuals previously exposed to HEV, whose antibody level has waned to undetectable levels. Until now, the longevity of immunological response following HEV infection remains unclear. Multiple small-scale early and recent reports showed that anti-HEV antibodies could persist from 5 to 30 years post-infection in an age-dependent manner.⁵³ Importantly, 80% of individuals were

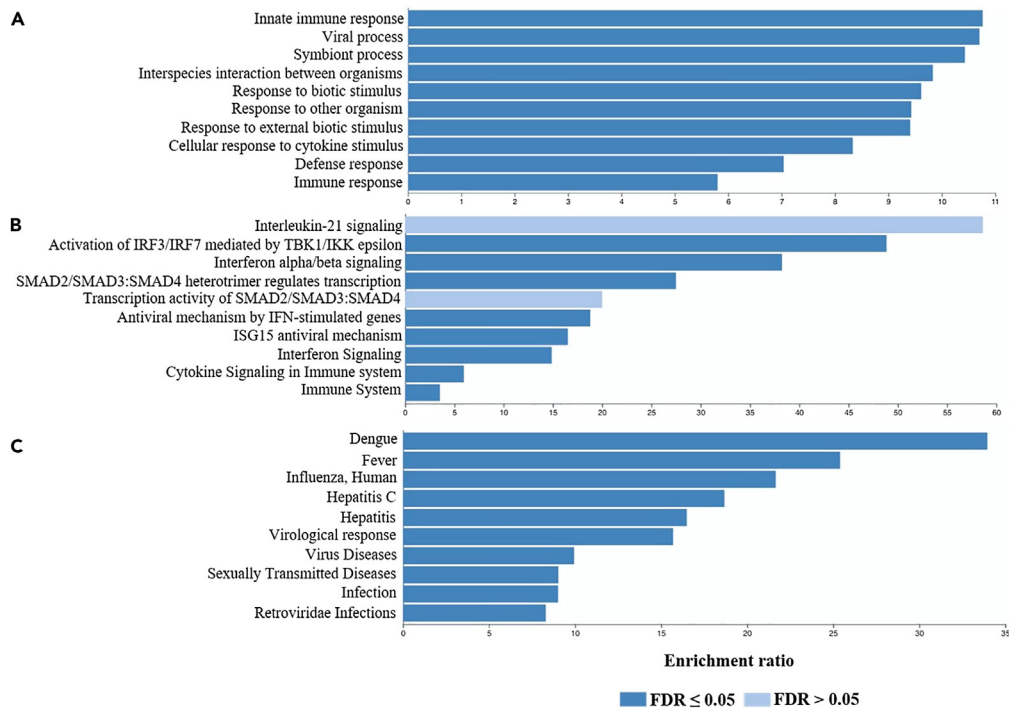


Figure 7. Gene set enrichment analysis of immune-related genes from the binary and quantitative trait HEV GWASs using WebGestalt

(A–C) A list of 43 immune genes filtered from our candidate genes list was used for: A.GO enrichment analysis. B.ORA pathway analysis. C. ORA in GLAD4U disease database. The top ten enriched pathways are presented in each bar graph. Categories with significant enrichment (FDR <0.05) are colored in dark blue.

still positive for anti-HEV antibodies at least 10 years post-infection.⁵⁴ Nonetheless, this does not eliminate the possibility of having false negative cases in our control group. Our results, therefore, provide insights on variants related to the antibody response and perhaps correlate with durability but cannot be translated into susceptibility due to the lack of solid evidence of exposure. Moreover, although we performed two GWASs, our results require replication in an entirely different cohort. It was also unfeasible to perform a meta-analysis due to the unavailability of other GWASs on HEV infection.

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
 - Sample selection criteria
 - Ethical approval
- METHOD DETAILS
 - Genomic data and quality control
 - Binary and quantitative phenotypes classification
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Genome-wide association analysis
 - Functional annotation
 - Gene enrichment analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.107586>.

ACKNOWLEDGMENTS

This study was funded by Qatar University High Impact Grant (Grant number: QUHI-BRC-20_21-1). The work makes use of data generated by the Qatar Genome Program (QGP) and Qatar Biobank (QBB), which are funded by Qatar foundation for Education, Science and Community.

AUTHOR CONTRIBUTIONS

Conceptualization, H.M.Y. and O.A.; Methodology, O.A., Y.A.S., and M.K.S.; Formal Analysis, Y.A.S. and M.K.S.; Investigation, M.K.S. and Y.A.S.; Resources, H.M.Y. and O.A.; Data Curation, Y.A.S. and O.A.; Writing – Original Draft, M.K.S.; Writing – Review and Editing, H.M.Y., O.A., and Y.A.S.; Funding Acquisition, H.M.Y.; Supervision, H.M.Y. and O.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 16, 2023

Revised: June 22, 2023

Accepted: August 4, 2023

Published: August 10, 2023

REFERENCES

1. Wißing, M.H., Brüggemann, Y., Steinmann, E., and Todt, D. (2021). Virus-Host Cell Interplay during Hepatitis E Virus Infection. *Trends Microbiol.* 29, 309–319. <https://doi.org/10.1016/j.tim.2020.07.002>.
2. Kenney, S.P. (2019). The Current Host Range of Hepatitis E Viruses. *Viruses* 11, 452. <https://doi.org/10.3390/v11050452>.
3. Wedemeyer, H., Pischke, S., and Manns, M.P. (2012). Pathogenesis and treatment of hepatitis e virus infection. *Gastroenterology* 142, 1388–1397.e1. <https://doi.org/10.1053/j.gastro.2012.02.014>.
4. Ditah, I., and Charlton, M. (2015). Genetic susceptibility to hepatitis E viral infection: An enigmatic virus gives up a secret. *Hepatology* 62, 1337–1338. <https://doi.org/10.1002/hep.28106>.
5. (2022). Hepatitis E. <https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-e>.
6. Li, P., Liu, J., Li, Y., Su, J., Ma, Z., Bramer, W.M., Cao, W., de Man, R.A., Peppelenbosch, M.P., and Pan, Q. (2020). The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis. *Liver Int.* 40, 1516–1528. <https://doi.org/10.1111/liv.14468>.
7. Capai, L., Falchi, A., and Charrel, R. (2019). Meta-Analysis of Human IgG anti-HEV Seroprevalence in Industrialized Countries and a Review of Literature. *Viruses* 11. <https://doi.org/10.3390/v11010084>.
8. Abebe, M., Ali, I., Ayele, S., Overbo, J., Aseffa, A., and Mihret, A. (2017). Seroprevalence and risk factors of Hepatitis E Virus infection among pregnant women in Addis Ababa, Ethiopia. *PLoS One* 12, e0180078. <https://doi.org/10.1371/journal.pone.0180078>.
9. Kaur, M., Sidhu, S.K., Singh, K., Devi, P., Kaur, M., and Singh, N.J. (2017). Hepatitis E virus: A leading cause of waterborne viral hepatitis in Northwest Districts of Punjab, India. *J. Lab. Physicians* 9, 121–124. <https://doi.org/10.4103/0974-2727.199636>.
10. Babanejad, M., Izadi, N., Najafi, F., and Alavian, S.M. (2016). The HBsAg Prevalence Among Blood Donors From Eastern Mediterranean and Middle Eastern Countries: A Systematic Review and Meta-Analysis. *Hepat. Mon.* 16, e35664. <https://doi.org/10.5812/hepatmon.35664>.
11. Yazbek, S., Kreidieh, K., and Ramia, S. (2016). Hepatitis E virus in the countries of the Middle East and North Africa region: an awareness of an infectious threat to blood safety. *Infection* 44, 11–22. <https://doi.org/10.1007/s15010-015-0807-5>.
12. Nasrallah, G.K., Al Absi, E.S., Ghandour, R., Ali, N.H., Taleb, S., Hedaya, L., Ali, F., Huwaidy, M., and Hussein, A. (2017). Seroprevalence of hepatitis E virus among blood donors in Qatar (2013-2016). *Transfusion* 57, 1801–1807. <https://doi.org/10.1111/trf.14116>.
13. Ditah, I., Ditah, F., Devaki, P., Ditah, C., Kamath, P.S., and Charlton, M. (2014). Current epidemiology of hepatitis E virus infection in the United States: low seroprevalence in the National Health and Nutrition Evaluation Survey. *Hepatology* 60, 815–822. <https://doi.org/10.1002/hep.27219>.
14. Zhang, L., Yesupriya, A., Chang, M.-H., Teshale, E., and Teo, C.-G. (2015). Apolipoprotein E and protection against hepatitis E viral infection in American non-Hispanic blacks. *Hepatology* 62, 1346–1352. <https://doi.org/10.1002/hep.27938>.
15. Gu, M., Qiu, J., Guo, D., Xu, Y., Liu, X., Shen, C., and Dong, C. (2018). Evaluation of candidate genes associated with hepatitis A and E virus infection in Chinese Han population. *Viral J.* 15, 47. <https://doi.org/10.1186/s12985-018-0962-2>.
16. Pischke, S., Hartl, J., Roque-Afonso, A.M., and Mallet, V. (2016). Apolipoprotein E $\epsilon 3$ and $\epsilon 4$ are associated with a lower exposure to hepatitis E virus in American non-Hispanic blacks. *Hepatology* 64, 687–688. <https://doi.org/10.1002/hep.28355>.
17. Weller, R., Todt, D., Engelmann, M., Friesland, M., Wedemeyer, H., Pietschmann, T., and Steinmann, E. (2016). Apolipoprotein E polymorphisms and their protective effect on hepatitis E virus replication. *Hepatology* 64, 2274–2276. <https://doi.org/10.1002/hep.28788>.
18. Arya, R.P., Mishra, N., Biswas, K., and Arankalle, V.A. (2018). Association of Toll-like receptor 4 polymorphism with hepatitis E virus-infected Indian patients. *J. Viral Hepat.* 25, 1617–1623. <https://doi.org/10.1111/jvh.12980>.
19. 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>.
20. Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. Genome-wide Association Studies
21. Zhou, J., Chen, D.Q., Poon, V.K.M., Zeng, Y., Ng, F., Lu, L., Huang, J.D., Yuen, K.Y., and Zheng, B.J. (2009). A regulatory polymorphism in interferon-gamma receptor 1 promoter is associated with the susceptibility to chronic hepatitis B virus infection. *Immunogenetics* 61, 423–430. <https://doi.org/10.1007/s00251-009-0377-8>.
22. Croft, N.P., Smith, S.A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Wickett, M.J., Sebastian, P., Flesch, I.E.A., Heading, S.L., et al. (2019). Most viral peptides displayed by class I MHC on infected cells are immunogenic. *Proc. Natl. Acad. Sci. USA* 116, 3112–3117. <https://doi.org/10.1073/pnas.1815239116>.
23. Koutsakos, M., McWilliam, H.E.G., Aktepe, T.E., Fritzlar, S., Illing, P.T., Mifsud, N.A., Purcell, A.W., Rockman, S., Reading, P.C., Vivian, J.P., et al. (2019). Downregulation of MHC Class I Expression by Influenza A and B Viruses. *Front. Immunol.* 10, 1158. <https://doi.org/10.3389/fimmu.2019.01158>.
24. Hewitt, E.W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology* 110, 163–169. <https://doi.org/10.1046/j.1365-2567.2003.01738.x>.
25. Rakhra, G., and Rakhra, G. (2021). Zinc finger proteins: insights into the transcriptional and post transcriptional regulation of immune response. *Mol. Biol. Rep.* 48, 5735–5743. <https://doi.org/10.1007/s11033-021-06556-x>.
26. Giallourakis, C.C., Benita, Y., Molinie, B., Cao, Z., Despo, O., Pratt, H.E., Zukerberg, L.R., Daly, M.J., Rioux, J.D., and Xavier, R.J. (2013). Genome-wide analysis of immune system genes by expressed sequence Tag profiling.

- J. Immunol. 190, 5578–5587. <https://doi.org/10.4049/jimmunol.1203471>.
27. Lu, Z., Zheng, Z., Xu, Y., Wang, C., Lin, Y., Lin, K., Fu, L., Zhou, H., Pi, L., Che, D., and Gu, X. (2022). The Associated of the Risk of IVIG Resistance in Kawasaki Disease with. *J. Inflamm. Res.* 15, 5053–5062. <https://doi.org/10.2147/JIR.S378080>.
 28. Yıldırım, D.G., Demirdağ, T.B., Tokgöz, S.T., Tapisız, A., Bakkaloğlu, S.A., and Buyan, N. (2021). Viral infections associated with Kawasaki disease. *Turk. Arch. Pediatr.* 56, 272–274. <https://doi.org/10.5152/TurkArchPediatr.2021.20250>.
 29. Connolly, D.J., O'Neill, L.A.J., and McGettrick, A.F. (2013). The GOLD domain-containing protein TMED1 is involved in interleukin-33 signaling. *J. Biol. Chem.* 288, 5616–5623. <https://doi.org/10.1074/jbc.M112.403899>.
 30. Liu, W., and Ding, C. (2017). Roles of LncRNAs in Viral Infections. *Front. Cell. Infect. Microbiol.* 7, 205. <https://doi.org/10.3389/fcimb.2017.00205>.
 31. He, Z., Zhu, J., Mo, J., Zhao, H., and Chen, Q. (2020). HBV DNA integrates into upregulated ZBTB20 in patients with hepatocellular carcinoma. *Mol. Med. Rep.* 22, 380–386. <https://doi.org/10.3892/mmr.2020.11074>.
 32. Desfarges, S., and Ciuffi, A. (2012). Viral Integration and Consequences on Host Gene Expression. In *Viruses: Essential Agents of Life* (© Springer Science+Business Media Dordrecht 2012), pp. 147–175. https://doi.org/10.1007/978-94-007-4899-6_7.
 33. Wong, R.R., Abd-Aziz, N., Affendi, S., and Poh, C.L. (2020). Role of microRNAs in antiviral responses to dengue infection. *J. Biomed. Sci.* 27, 4. <https://doi.org/10.1186/s12929-019-0614-x>.
 34. Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. (2005). Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science* 309, 1577–1581. <https://doi.org/10.1126/science.1113329>.
 35. Schult, P., Roth, H., Adams, R.L., Mas, C., Imbert, L., Orlik, C., Ruggieri, A., Pyle, A.M., and Lohmann, V. (2018). microRNA-122 amplifies hepatitis C virus translation by shaping the structure of the internal ribosomal entry site. *Nat. Commun.* 9, 2613. <https://doi.org/10.1038/s41467-018-05053-3>.
 36. Harms, D., Choi, M., Allers, K., Wang, B., Pietsch, H., Papp, C.P., Hanisch, L., Kurreck, J., Hofmann, J., and Bock, C.T. (2020). Specific circulating microRNAs during hepatitis E infection can serve as indicator for chronic hepatitis. *Sci. Rep.* 10, 5337. <https://doi.org/10.1038/s41598-020-62159-9>.
 37. Trehanpati, N., Sehgal, R., Patra, S., Vyas, A., Vasudevan, M., Khosla, R., Khanam, A., Kumar, G., Maiwall, R., Ramakrishna, G., et al. (2017). miRNA signatures can predict acute liver failure in hepatitis E infected pregnant females. *Heliyon* 3, e00287. <https://doi.org/10.1016/j.heliyon.2017.e00287>.
 38. Stamm, S., and Lodmell, J.S. (2019). C/D box snoRNAs in viral infections: RNA viruses use old dogs for new tricks. *Noncoding. RNA Res.* 4, 46–53. <https://doi.org/10.1016/j.ncrna.2019.02.001>.
 39. Murray, J.L., Sheng, J., and Rubin, D.H. (2014). A role for H/ACA and C/D small nucleolar RNAs in viral replication. *Mol. Biotechnol.* 56, 429–437. <https://doi.org/10.1007/s12033-013-9730-0>.
 40. Qian, X., Xu, C., Wu, B., Tang, H., Zhao, P., and Qi, Z. (2020). SNORD126 Promotes Hepatitis C Virus Infection by Upregulating Claudin-1 via Activation of PI3K-AKT Signaling Pathway. *Front. Microbiol.* 11, 565590. <https://doi.org/10.3389/fmicb.2020.565590>.
 41. Valleron, W., Laprevotte, E., Gautier, E.F., Quelen, C., Demur, C., Delabesse, E., Agirre, X., Prósper, F., Kiss, T., and Brousset, P. (2012). Specific small nucleolar RNA expression profiles in acute leukemia. *Leukemia* 26, 2052–2060. <https://doi.org/10.1038/leu.2012.111>.
 42. Yates, D. (2011). Molecular neuroscience: Synaptotagmin 10-mediated release. *Nat. Rev. Neurosci.* 12, 307. <https://doi.org/10.1038/nrn3051>.
 43. Thakur, V., Ratho, R.K., Kumar, S., Saxena, S.K., Bora, I., and Thakur, P. (2020). Viral Hepatitis E and Chronicity: A Growing Public Health Concern. *Front. Microbiol.* 11, 577339. <https://doi.org/10.3389/fmicb.2020.577339>.
 44. Tanaka, Y., and Chen, Z.J. (2012). STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci. Signal.* 5, ra20. <https://doi.org/10.1126/scisignal.2002521>.
 45. Pokharel, S.M., Shil, N.K., and Bose, S. (2016). Autophagy, TGF- β , and SMAD-2/3 Signaling Regulates Interferon- β Response in Respiratory Syncytial Virus Infected Macrophages. *Front. Cell. Infect. Microbiol.* 6, 174. <https://doi.org/10.3389/fcimb.2016.00174>.
 46. Todt, D., François, C., Anggakusuma, Behrendt, P., Behrendt, P., Engelmann, M., Knegeendorf, L., Vieyres, G., Wedemeyer, H., Hartmann, R., Pietschmann, T., et al. (2016). Antiviral Activities of Different Interferon Types and Subtypes against Hepatitis E Virus Replication. *Antimicrob. Agents Chemother.* 60, 2132–2139. <https://doi.org/10.1128/AAC.02427-15>.
 47. Nan, Y., Yu, Y., Ma, Z., Khattar, S.K., Fredericksen, B., and Zhang, Y.J. (2014). Hepatitis E virus inhibits type I interferon induction by ORF1 products. *J. Virol.* 88, 11924–11932. <https://doi.org/10.1128/JVI.01935-14>.
 48. Tan, G., Song, H., Xu, F., and Cheng, G. (2018). When Hepatitis B Virus Meets Interferons. *Front. Microbiol.* 9, 1611. <https://doi.org/10.3389/fmicb.2018.01611>.
 49. Qu, L., and Lemon, S.M. (2010). Hepatitis A and hepatitis C viruses: divergent infection outcomes marked by similarities in induction and evasion of interferon responses. *Semin. Liver Dis.* 30, 319–332. <https://doi.org/10.1055/s-0030-1267534>.
 50. Grzegorzewska, A.E., Świdarska, M.K., Niepolski, L., Bura, M., Mostowska, A., Łągiedo-Żelazowska, M., and Jagodziński, P.P. (2018). Interferon- γ gene polymorphisms, circulating interferon λ 3, and clinical variables in hemodialysis patients exposed to hepatitis E virus. *Pol. Arch. Intern. Med.* 128, 344–353. <https://doi.org/10.20452/pamw.4281>.
 51. Crotty, S. (2019). T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. *Immunity* 50, 1132–1148. <https://doi.org/10.1016/j.immuni.2019.04.011>.
 52. Mbarek, H., Devadoss Gandhi, G., Selvaraj, S., Al-Muftah, W., Badji, R., Al-Sarraj, Y., Saad, C., Darwish, D., Alvi, M., Fadd, T., et al. (2022). Qatar genome: Insights on genomics from the Middle East. *Hum. Mutat.* 43, 499–510. <https://doi.org/10.1002/humu.24336>.
 53. Chadha, M.S., Walimbe, A.M., and Arankalle, V.A. (1999). Retrospective serological analysis of hepatitis E patients: a long-term follow-up study. *J. Viral Hepat.* 6, 457–461. <https://doi.org/10.1046/j.1365-2893.1999.00190.x>.
 54. Kmush, B.L., Zaman, K., Yunus, M., Saha, P., Nelson, K.E., and Labrique, A.B. (2018). A Ten Year Immunopersistence Study of Hepatitis E Antibodies in Rural Bangladesh. *Am. J. Epidemiol.* 187, 1501–1510. <https://doi.org/10.1093/aje/kwy044>.
 55. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>.
 56. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* 50, 1335–1341. <https://doi.org/10.1038/s41588-018-0184-y>.
 57. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826. <https://doi.org/10.1038/s41467-017-01261-5>.
 58. Thareja, G., Al-Sarraj, Y., Belkadi, A., Almotawa, M., Qatar Genome Program Research OQPR Consortium, Suhre, K., and Albagha, O.M.E. (2021). Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nat. Commun.* 12, 1250. <https://doi.org/10.1038/s41467-021-21381-3>.
 59. FastQC
 60. Bwakit. <https://github.com/lh3/bwa/tree/master/bwakit>.
 61. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
 62. Data Processing and Output of the Qatar Genome Project. Qatar Biobank
 63. Lai, Y., Feldman, K.L., and Clark, R.S.B. (2005). Enzyme-linked immunosorbent assays (ELISAs). *Crit. Care Med.* 33, S433–S434. <https://doi.org/10.1097/01.ccm.0000191717.44652.c9>.
 64. 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
Biological samples		
Serum samples	Qatar Biobank (QBB)	N/A
Critical commercial assays		
HEV ELISA-IgG	MP Biomedicals	SKU: 0721150096T
Software and algorithms		
PLINK-1.9.	Purcell et al. ⁵⁵	http://pngu.mgh.harvard.edu/purcell/plink/
SAIGE (Scalable and Accurate Implementation of GEneralized mixed model)	Zhou et al. ⁵⁶	https://github.com/weizhouUMICH/SAIGE/
Functional Mapping and Annotation of genetic associations (FUMA)	Watanabe et al. ⁵⁷	https://fuma.ctglab.nl/
Web-based Gene SeT AnaLysis Toolkit	Wang et al. ¹⁹	http://www.webgestalt.org
R (version 4.2.1)	The R Foundation	www.R-project.org
Other		
Whole genome sequences	Qatar Biobank (QBB)	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maria K. Smatti (mismatti@qu.edu.qa).

Materials availability

The materials used in this paper (biological samples and whole genome sequences) are owned by Qatar Biobank (<https://www.qatarbiobank.org.qa>).

Data and code availability

- Whole-genome sequence data are not publicly available in repositories. To obtain access, an application must be submitted to the Qatar Biobank.
- This paper does not report original code.
- Results of the study (raw summary statistics) and any additional information required to reanalyze the data reported in this paper are available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Sample selection criteria

The present study included a cohort from Qatar Biobank (QBB) participants. Whole-genome sequences of 6,218 Qatari nationals who had previously participated in phase 1 of the Qatar Genome Project (QGP) were obtained. This cohort represented the first data release from QBB. However, the corresponding sera samples were available for only 6000 individuals. A detailed demographic characterization of this cohort has been previously documented.⁵⁸ QBB recruits adults (age ≥ 18 years) following specific inclusion/exclusion criteria to comprehensively include the permanent heterogeneous population in Qatar. Only Qatari nationals or long-term residents (≥ 15 years living in Qatar) are eligible to participate. All study participants were of Qatari origin. Male participants represented 43.2% ($n=2593$) of the study population compared to 56.8% ($n=3407$) females. The mean age of enrolled subjects was 40.5 years ($SD \pm 12$). All QBB participants signed informed consent before their participation.

Ethical approval

The study was approved by Qatar University Institutional Review Board (IRB, QU-IRB 1287-EA/20) and Qatar Biobank IRB (E-2020-QBB-RES-ACC-0184-0110).

METHOD DETAILS

Genomic data and quality control

Sequencing read data were generated by Illumina HiSeq X Ten sequencers and converted from the native BCL format to paired-end FASTQ format using *bcl2fastq*. The quality of the raw data was assessed using *fastqc*.⁵⁹ Data passing quality control was then aligned to the reference genome sequence (build GRCh37 (hs37d53) using the *bwa-kit6* aligner [v0.7.12].⁶⁰ Variant calling was performed using GATK haplotype caller [v3.3].^{61,62} The genetic variant data was then converted to PLINK file format using PLINK-1.9.⁵⁵ Standardized quality-assurance and quality control (QA/QC) methods were followed to generate high quality and confidence on both SNPs and sample levels, as previously described.⁵⁸ Briefly, SNPs with genotype call rate <90%, Hardy-Weinberg *p*-value <1×10⁻⁶ were removed. Additionally, subjects with excess heterozygosity, duplicates, call rate <95%, and gender ambiguity were removed. We calculated the Pairwise identity by-state (IBS) matrix based on a pruned set of independent autosomal SNPs (N = 62,475) using a window size of 200 SNPs and LD threshold of *r*² = 0.05 to identify and remove duplicated samples. Moreover, multidimensional scaling (MDS) was performed to identify and remove population ancestry outliers. Samples with ±4 SD from the mean of the two MDS components were considered population outliers and removed from the analysis.^{52,58} The final set of sequences used for the GWAS comprised 8,408,727 SNPs. The pass QC number of samples was 6045.

Binary and quantitative phenotypes classification

Sera samples (n=6000) were tested for HEV IgG antibodies using a commercial ELISA assay (HEV ELISA-IgG, MP Biomedicals) according to the manufacturer's protocol. For the binary trait GWAS, samples were grouped into cases or controls based on the anti-HEV IgG positivity. For the quantitative GWAS, samples were semi-quantitatively categorized into five groups based on their ELISA absorbance value, which correlates with the anti-HEV antibody level.⁶³ We used the positive control value as a reference because it is plate/experiment-specific. Samples were classified as follows: Group1= samples with absorbance higher than 1.5*(positive control); Group2= samples with absorbance higher than the positive control but less than 1.5*(positive control); Group3= samples with absorbance higher than the positive control and higher than 0.5*(positive control); Group5= positive samples with absorbance less than or equal to 0.5*(positive control); and Group 6= negative samples. Two samples were excluded due to indeterminate ELISA results obtained from two separate experiments. As per the phenotypic data availability and genomic QC filtration, the final total number of subjects included in the GWAS was 5829 individuals.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-wide association analysis

Genome-wide association testing was performed using the SAIGE (Scalable and Accurate Implementation of GEneralized mixed model) implemented in the R package.⁵⁶ This tool is based on a generalized mixed model association test that uses the saddlepoint approximation to calibrate the distribution of score test statistics for binary and quantitative traits. Importantly, this method provides better association accuracy because it accounts for case-control imbalance and sample relatedness.⁵⁶ For all tested traits, we included age, sex, and the first four principal components (PCs) as covariates in the model. Principal component analysis was performed using PLINK. The genome-wide significance threshold was set as *p*<5×10⁻⁸.²⁸ Quantile-Quantile plots, Manhattan plots, and regional association plots were all generated using Functional Mapping and Annotation of genetic associations (FUMA).⁵⁷

Functional annotation

Functional annotation was performed using the FUMA platform (<https://fuma.ctglab.nl/>), an integrative web-based tool that processes GWAS summary statistics for functional annotation and gene prioritization. First, a pre-calculated LD structure based on the 1000G reference population was used to define a genomic risk locus. Independent significant variants (*p*<5×10⁻⁸) which are independent of each other at *r*² < 0.6 were also defined. Additionally, SNPs that are independent at *r*² < 0.1 were considered independent lead SNPs. Moreover, all SNPs in LD with an independent SNP (even if not in the GWAS results) are added to the annotation.⁵⁷ Variant annotation with biological functionality and gene mapping was then carried out according to three main strategies: physical position, expression quantitative trait loci (eQTL), and chromatin interaction. FUMA includes 18 biological data repositories for gene-set analysis by MAGMA (Multi-marker Analysis of GenoMic Annotation). MAGMA applies multiple linear regression models to examine whether genes in different gene sets are strongly associated with a trait. This was followed by gene-property analysis to investigate further the relationship between tissue-specific gene expression profiles and gene associations.⁶⁴ Gene analysis was performed using MAGMA (v1.6) with default settings. Significant *p*-value threshold set at 0.05 / (number of tested genes).

Gene enrichment analysis

Gene enrichment analysis was performed utilizing WebGestalt (Web-based Gene SeT Analysis Toolkit, <http://www.webgestalt.org>). This toolkit integrates functional categories from different databases and computational analyses to provide a comprehensive yet easy way to evaluate gene sets for biological context interpretation.¹⁹ We used the over-representation analysis (ORA) method for Gene ontology

(GO), pathways (Reactome), and disease (GLAD4U) databases. For all analyses, the number of genes in each pathway was set from 5 to 2000, and Benjamini-Hochberg FDR was used for multiple test adjustment. Furthermore, InnateDB database (<https://www.innatedb.com>) was used to determine which genes are involved in immune-relevant pathways. Briefly, the combined set of mapped genes from the quantitative and qualitative GWASs (n=1866) was filtered to select protein-coding genes only (n=537). This list was then annotated using InnateDB annotation tool to explore which genes are associated with immune responses. Subsequently, the resulting list of genes that have an “immune response” function (n=44) was subjected to functional enrichment analysis.