

Novel host genetic variations associated with spontaneous clearance of a single-source outbreak of HCV1b infections

Hong You,¹ Sandu Liu,² Yong Xie,³ Rui Cong,¹ Yameng Sun,¹ Jingjing Ren,⁴ Kangfei Wei,⁴ Xin Jin,⁴ Yujian Shi,⁴ Haiying Zhang,⁵ Jie Li,⁶ Lai Wei,⁵ Hui Zhuang,⁶ Mingliang Cheng,⁷ Jidong Jia¹

To cite: You H, Liu S, Xie Y, *et al.* Novel host genetic variations associated with spontaneous clearance of a single-source outbreak of HCV1b infections. *BMJ Open Gastro* 2014;1:e000010. doi:10.1136/bmjgast-2014-000010

HY, SL and YX are the first coauthors.

The abstract of this paper has been presented in the 22nd Congress of the Asian Pacific Association for the Study of the Liver (APASL 2012, PS04-03), and published in the abstract book.

Received 27 May 2014
Revised 24 August 2014
Accepted 29 August 2014

For numbered affiliations see end of article.

Correspondence to
Dr Mingliang Cheng;
chengml@21cn.com,
Dr Jidong Jia;
jia_jd@ccmu.edu.cn

ABSTRACT

Background and aims: A total of 105 patients were identified as accidentally infected with hepatitis C virus genotype 1b (HCV1b) through blood transfusion from a single blood donor. This group provides a unique patient population to study host factors involved in the spontaneous clearance of HCV and disease progression.

Methods: Clinical markers, HCV RNA and eight single nucleotide polymorphisms (SNPs) of interleukin-28B (IL-28B) were detected. Exome capture and sequencing were analysed for association with HCV clearance.

Results: Among the 85 patients with the positive HCV antibody, 27 cases (31.8%) were HCV RNA negative over a period of 9–12 years. Of the 58 patients with positive HCV RNA, 22.4% developed chronic hepatitis, and 5.2% developed cirrhosis. Age was found to be associated with HCV1b clearance. IL-28 rs10853728 CC showed the trend. By exon sequencing, 39 SNPs were found to be significantly different in spontaneous clearance patients ($p < 0.001$). Two SNPs in the tenascin receptor (TNR), five in the transmembrane protease serine 11A (TMPRSS11A), and one in the serine peptidase inhibitor kunitz type 2 (SPINT2) showed the closest associations ($p < 10^{-5}$).

Conclusions: Host genetic analyses on the unique, single source HCV1b-infected patient population has suggested that age and mutations in TNR, TMPRSS11A and SPINT2 genes may be factors associated with HCV clearance.

INTRODUCTION

Hepatitis C virus (HCV) infection affects hundreds of millions of people worldwide. It has been reported that about 20% of HCV-infected adults can spontaneously clear the virus, while 30% of patients with chronic infection progress to cirrhosis and hepatocellular carcinoma (HCC).¹

Viral and host factors are involved in HCV spontaneous clearance and disease progression. Virus factors include HCV genotypes, quasispecies, viral load and co-infection. Host factors include gender, age at infection

SUMMARY BOX

What is already known about this subject?

Host interleukin-28B (IL-28B) polymorphisms were known to be associated with spontaneous hepatitis C virus (HCV) clearance and also response to treatment. HCV is the other factor contributing to clearance. When both host and viral factors are mixed involving in HCV clearance and disease progression, it is difficult to tell the important factors.

What are the new findings?

This is a the study on a unique group of patients with HCV1b-infection (n=105) accidentally transmitted from a single blood donor infected with genotype 1b in Guizhou province, southwest China. With the sole resource of the virus, the clear-known infected time, the similar ethnicity and environments, it is better to understand the host factors for HCV spontaneous clearance and disease progress.

How might it impact on clinical practice in the foreseeable future?

Add the knowledge of how the host factors may affect HCV clearance.

or the ageing process, race, alanine aminotransferase (ALT) elevation and genetic factors.² Recently, interleukin-28B (IL-28B) polymorphisms have been reported to be associated with spontaneous HCV clearance and also response to treatment.^{3–7}

The purpose of the current study was to analyse a group of patients infected with the same HCV genotype 1b (HCV1b) source in order to focus on host parameters that may be involved in resolution or persistence of HCV infection. These patients in the current study are unique for several reasons. First, the sole resource of the HCV1b virus excludes virus genotypic differences. HCV1b is a difficult-to-treat genotype with interferon-based therapy. Second, the known date of infection provides data on the natural history of HCV infection over the course of 9–12 years. Third,

the common ethnicity and similar environments of the patients reduce some variables into the analysis of factors involved in HCV spontaneous clearance and disease progression. Lastly, the broad age range of patients is helpful to study the importance of host age.

MATERIALS AND METHODS

Study subjects

All patients had received blood transfusions, from 1998 to 2002, from a single blood donor who was subsequently found to have had HCV1b. All recipients were Chinese from Pingtang, Guizhou province, southwest China. Inclusion criteria were transfusion of blood or blood-products from the identified contaminated batches of the same donor. Patients who died from causes other than HCV-related liver disease, and patients we were unable to contact, were excluded. All patients with positive HCV RNA were tested and found to be genotype 1b.

Patients were identified and blood samples were collected from 2010 to 2011, 9–12 years postinfection (median 10 years).

The study was approved by the ethics committee from Guiyang Medical College and conformed to the ethical guidelines of the Declaration of Helsinki; informed consent had been obtained from each individual included in the study.

Serum HCV antibody and RNA assays

Serum biochemical parameters including ALT levels were measured by routine automated methods according to the manufacturer's instructions. Anti-HCV antibody levels were measured by electrochemiluminescence immunoassay (ECLIA) using Abbott Architect i2000 (ABBOTT, Wiesbaden, Germany) according to the manufacturer's instructions. HCV RNA was detected by the commercial quantitative reverse transcription PCR (RT-PCR; COBAS AMPLICOR, Roche Diagnostic Systems, Indianapolis, USA) according to the manufacturer's instructions. The lower limit of detection was 15 IU/mL. The HCV genotype was determined by Versant HCV Genotype 2.0 (LiPA; Siemens Healthcare Diagnostics, Tarrytown, New York, USA).

Fibroscan detection

Transient elastography was performed using FibroScan (Echosens, France). The examination was performed on the right lobe of the liver through the seventh or eighth intercostal space. The measurement depth was between 25 and 65 mm. As suggested by the manufacturer, only results obtained with 10 valid measurements, with a success-rate of at least 60% and with an IQR \leq 30%, were considered reliable, as described previously.^{8–10}

PCR amplification and sequencing of IL-28B polymorphisms

Genomic DNAs were isolated from 0.5 mL whole blood using the QIAamp DNA Mini Kit (Qiagen, Hilden,

Germany). IL-28 rs12979860, rs8099917, rs10853728, rs12980275, rs4803219, rs4803223, rs8105790 and rs28416813 were amplified by PCR. The PCR protocol involved initial denaturation at 95°C for 10 min, 35 cycles of denaturation for 30 s at 95°C, annealing of primers for 30 s at 55°C and extension for 40 s at 72°C, followed by final extension at 72°C for 10 min. The amplified DNA fragments were separated on a 2% agarose gel, and purified with the QIAquick gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Nucleotide sequences were determined by Sanger sequencing using the Applied Biosystems Automated 3730 DNA Analyzer.

Exome capture and sequencing assay

Exome capture

Purified genomic DNA samples were randomly fragmented by Covaris, with the size of library fragments mainly distributed between 150 and 200 bp.^{11–13} Adapters were ligated to both ends and purified by the Agencourt AMPure SPRI beads according to the manufacturer's instructions. Fragments with insert sizes of about 250 bp were excised. Extracted DNA was amplified by ligation-mediated PCR (LM-PCR), purified and hybridised to the SureSelect Biotinylated RNA Library (BAITS) for enrichment. Hybridised fragments were bound to the streptavidin beads, whereas non-hybridised fragments were washed out after 24 h. Captured LM-PCR products were analysed using an Agilent 2100 Bioanalyzer to estimate the magnitude of enrichment. Each captured library was then loaded on a HiSeq2000 platform, to ensure that each sample meets the desired average sequencing depth for high-throughput sequencing. Raw image files were processed by Illumina basecalling Software V.1.7 with default parameters and the sequences of each individual were generated as 90 bp pair-end reads.¹¹

Read mapping

SOAPaligner (V.2.21) was used to align the sequencing reads to the NCBI human genome reference assembly (build 36.3) with a maximum of 3 mismatches and the parameters were set as -a -b -D -o -u -p -2 -m -x -s 40 -l 35 -v 3. To evaluate exome capture efficiency, the proportions of reads mapping to target regions and to their flanking regions (within 200 bp) were calculated for each individual. Reads that aligned to the designed target region (TR \pm 200 bp) were collected for single nucleotide polymorphism (SNP) identification and subsequent analysis.¹¹

Individual genotype calling

On the basis of SOAP alignment results, the software SOAPsnv was used to call genotypes. The following parameters were set: -r 0.0005 -e 0.001 -t -u -2 -i -d -o -M -L 90 -s -T (<http://soap.genomics.org.cn/> for details). To obtain an accurate genotype set, the genotype was filtered on the basis of the following criteria: the SNP should be observed in at least one individual with a

quality score >20, and in a way that the number of reads containing mutant alleles was larger than the reads containing reference alleles. At the same time, at least 90% of all individuals got a quality score >20 and coverage >4.^{12 13} Our next steps are all based on this genotype set.

Principal component analysis

To check the population stratification, we did a principal component analysis with EIGENSTRAT software. We chose the SNP from 1000 genomes (<http://www.1000genomes.org/>) to check our samples.

Determination of associations

We checked the genome-wide association results using PLINK software (the <http://pngu.mgh.harvard.edu/~purcell/plink/>) to perform a standard case/control association analysis using: `plink -file mydata -assoc`. All the top sites SNPs with statistical significance were also checked for Hardy-Weinberg equilibrium, and sites which had $p < 0.001$ were filtered.^{12 13}

Statistical analysis

Clinical statistical analyses were performed using the Statistical Program for Social Sciences (V.11.5; SPSS). Continuous variables such as ALT, HCV RNA and time were presented as medians (range), and categorical variables as frequencies. Categorical variables were tested by χ^2 test or Fisher's exact test. p Values <0.05 were considered to be statistically significant.

RESULTS

Patient demographic data

The demographic features of the patients are shown in figure 1 and table 1. A total of 105 receipts were found with the single same donor from medical records of the blood transfusions from 1998 to 2002. Twenty of them with a negative HCV antibody could not be contacted or died from causes other than liver disease. Among the 85 patients with a positive HCV antibody, the male-to-female ratio was 45:40 with a median age of 32 years (9–71 years). Twenty-seven cases (31.8%) were HCV RNA negative, and were considered to have achieved spontaneous clearance of HCV. The other 58 cases (68.2%) were HCV RNA positive with levels ranging from 3.0 log to 7.1 log IU/mL, median 4.8 log IU/mL.

Characteristics of HCV spontaneous clearance

Among the 85 patients with a positive HCV antibody, host factors including the gender, age and IL-28 allele were analysed for an association with spontaneous clearance. It was shown that an age less than 20 years had the most significant association with spontaneous clearance (OR=2.04, 95% CI (1.13 to 3.69), $p=0.028$), while other ages did not have any significant relation to virus clearance (table 1). Gender had no relation with viral clearance ($p=0.462$; figure 2).

Eight IL-28B polymorphisms, alleles rs12979860, rs8099917, rs10853728, rs12980275, rs4803219, rs4803223, rs8105790 and rs28416813 were PCR amplified and sequenced. As shown in table 2, IL-28B rs10853728 showed a strong trend, which fell just short of statistical significance ($p=0.058$). For the most well-known SNP, IL-28B rs12979860, 22 of the 24 (91.7%) patients who had spontaneous clearance had the CC allele, while 31 of the 40 (77.5%) in the non-clearance group had the CC allele ($p=0.132$). The TT allele was not found in any of the patients (figure 3).

Risk factors for HCV disease progression

Nine to 12 years after transfusion, none of the 27 patients with spontaneous clearance had disease progression by liver function tests, abdominal ultrasonography and fibroscan tests. Among the 58 patients with positive HCV RNA, 9 cases (15.5%) had elevated serum ALT ranging from 41 to 192 IU/mL. A total of 13 cases (22.4%) developed chronic hepatitis with mild to moderate fibrosis as determined by clinical manifestations, fibroscan values higher than 7.1 kPa, and enhanced and coarse echogenicity of the liver by ultrasonography. A total of three patients (5.2%) developed cirrhosis with decreased albumin, fibroscan values higher than

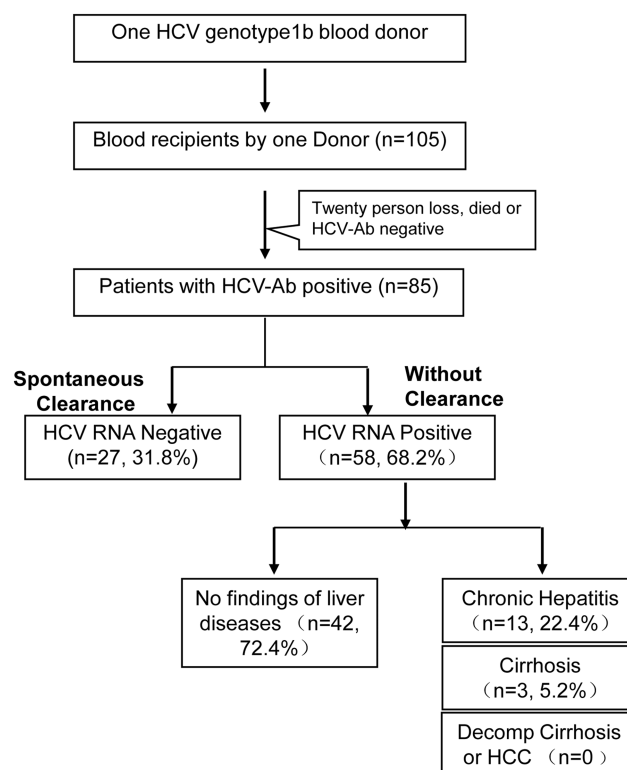


Figure 1 Data for the special HCV-infected group from one single blood donor. A total of 105 receipts, which were accidentally infected by a single HCV genotype 1b donor, from 1998 to 2002. AB, antibody; HCV, hepatitis C virus; HCC, hepatocellular carcinoma.

Table 1 Host characteristics of spontaneous clearance for HCV

Characteristics	Total (case, %)	Spontaneous clearance	Non-clearance	OR (95% CI)	p Value
Case (case, %)	85	27, 31.8	58, 68.2	—	—
Gender (case, %)					0.462
Male	45, 52.9	15, 55.6	30, 51.7	1.11 (0.59 to 2.08)	
Female	40, 47.1	12, 44.4	28, 48.3	0.95 (0.71 to 1.27)	
Age, median (range) (case, %)	32 (9–71),	30 (9–61),	34 (10–71),		0.160
≤20 years	19, 22.4	10, 37.0	9, 15.5	2.04 (1.13 to 3.69)	0.028*
21–40 years	39, 45.9	11, 40.7	28, 48.3	0.81 (0.43 to 1.54)	0.340
41–60 years	21, 24.7	5, 18.5	16, 27.6	0.69 (0.30 to 1.60)	0.267
≥61 years	6, 7.0	1, 3.7	5, 8.6	0.51 (0.08 to 3.12)	0.375
IL-28 allele frequency detected (case, %)	64	24	40		
rs10853728 CC	36, 56.3	17, 70.8	19, 47.5	1.89 (0.91 to 3.91)	0.058
rs12979860 CC	53, 82.8	22, 91.7	31, 77.5	2.28 (0.63 to 8.32)	0.132
rs8099917 TT	53, 82.8	22, 91.7	31, 77.5	2.28 (0.63 to 8.32)	0.132
rs12980275 AA	53, 82.8	22, 91.7	31, 77.5	2.28 (0.63 to 8.32)	0.132
rs4803219 CC	55, 85.9	22, 91.7	33, 82.5	1.80 (0.51 to 6.37)	0.264
rs4803223 AA	47, 73.4	16, 66.7	31, 77.5	0.72 (0.38 to 1.38)	0.254
rs8105790 TT	54, 84.3	22, 91.7	32, 80.0	2.04 (0.57 to 7.33)	0.189

*p<0.05 was considered to be statistically significant.
HCV, hepatitis C virus; IL, interleukin.

9.5 kPa, and splenomegaly as determined by ultrasound.^{8–10} Neither decompensated cirrhosis nor HCC was found. According to the infected time and duration, it was estimated that the rate of HCV progression to mild or moderate fibrosis was 2.2% per year, and to cirrhosis was 0.5% per year.

Multivariate regression analysis showed that gender (p=0.393), HCV RNA level (p=0.262) and IL-28B allele frequencies (p=0.565) were not statistically associated with disease progression. An age less than 40 years (OR=0.13, 95% CI (1.13 to 3.69), p=0.020) had a negative association with disease progression.

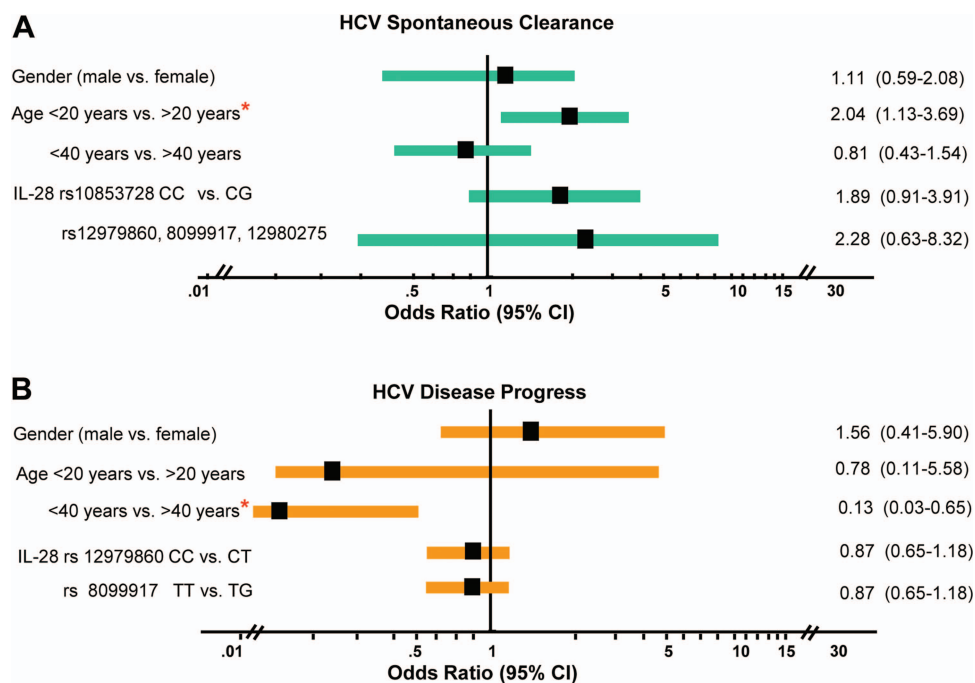


Figure 2 Association of host factors including gender, age and interleukin-28 (IL-28) polymorphisms with hepatitis C virus (HCV) spontaneous clearance or disease progression by OR. (A) Association of age of infection less than 20 years and HCV spontaneous clearance. (B) Association of age of infection less than 40 years and HCV disease progression.

Table 2 Characteristics of patients without HCV clearance and persisted infection with HCV RNA detectable

Characteristics	Patients without HCV clearance
Case number (%)	58
HCV RNA, median (range) (number, %)	4.8 (3.0–7.1), log IU/mL
≤3.0 log IU/mL	2, 3.4
3.1–5.0 log IU/mL	29, 50.0
5.1–7.0 log IU/mL	25, 43.2
≥7.1 log IU/mL	2, 3.4
ALT, median (range) (number, %)	30, (12–192), U/L
≤1.0 ULN	49, 84.5
1.1–3.0 ULN	3, 5.2
3.1–5.0 ULN	6, 10.3
Fibroscan, median (range) (number, %)	5.9, (3.7–35.3), kPa
Failure/undetected	12, 20.7
≤5.0 kPa	12/46, 26.1
5.0–7.3 kPa	21/46, 45.6
7.3–9.5 kPa	10/46, 21.7
>9.6 kPa	3/46, 6.6
Ultrasonography (number, %)	
Non-specific/undetected	41, 70.7
Enhanced/coarse echo of liver	13, 22.4
Splenomegaly	4, 6.9

ALT, alanine aminotransferase; HCV, hepatitis C virus; ULN, upper limit of normal value.

Whole-exome capture and sequencing

Using exome capture and sequencing, 64 449 SNPs were identified in the sample population. A total of 17 081 coding genes were sequenced with coverage of each individual exome at an average depth of 33.9-fold. On average, about 95% of the target regions were covered by at least one read. More than 86% of the target

regions were covered by more than 4 reads (figure 4). The population analysis showed that samples used for the association analysis had no significant population stratification.

The top 20 SNPs which had the closest association are listed in table 3. There were SNPs from 11 exons from IL-28B rs2239818 and rs34842046, two from the tenascin receptor (TNR), one (rs3745948) from the serine peptidase inhibitor kunitz type 2 (SPINT2), one (rs7627615) from the 5-hydroxytryptamine receptor 3 family member E (HTR3E), four (rs1370840, rs11930532, rs28437478 and rs6552134) from the transmembrane protease serine 11A (TMPRSS11A), one (rs1263810) from the sal-like protein 2 (SALL2), three (rs9901726, rs2291604 and rs9900543) from the spermatogenesis associated 22 (SPATA 22), one (rs607332) from the nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2), two (novel) from the NCK-associated protein 1 (NCKAP1), one (rs2303225) from the MARVEL domain containing 3 (MARVELD3), three (novel) from the zinc finger protein 491, 440 and 439 (ZNF491, ZNF 440 and ZNF 439), and one (rs2307075) from the carbonate dehydratase II (CA2; $p < 10^{-4}$).

From the function analysis, there were five SNPs of receptors, TNR, HTR3E and MARVELD3, and four SNPs from the transmembrane protease TMPRSS11A, which may affect HCV binding to the receptors and entry into hepatocytes ($p < 10^{-4}$).

DISCUSSION

The differences between HCV genotypes coupled with the large diversity of host factors make an analysis of favourable and morbid outcomes in populations difficult to determine and analyse. Our study focused on a specific group of patients with HCV1b-infection due to a unique sole source reservoir of virus, the clear known time of infection, the similar ethnic and environmental background, and the broad scale of 85 patients including both sexes and almost all ages. With the same HCV source, this information may help us to better study the host factors involved in virus clearance and/or disease progress, in exclusion of HCV viral differences.

There are only two studies comparable to the current study, one from Ireland and one from Germany. Both were reports of single-source infections with HCV1b involving homogeneous women of childbearing age who had received HCV-contaminated anti-D immunoglobulin to prevent Rh isoimmunisation from 1977 to 1979. The Irish group had 376 women with a mean age of 28 ±6 years, who had been infected for about 17 years.¹⁴ The German cohort had 1018 women with a median age of 24 years, who had been followed for 20 years.¹⁵ Both studies provided very important information on HCV spontaneous clearance and disease progression in large groups of patients with known durations of infection.^{16–22} However, owing to the relatively uniform age

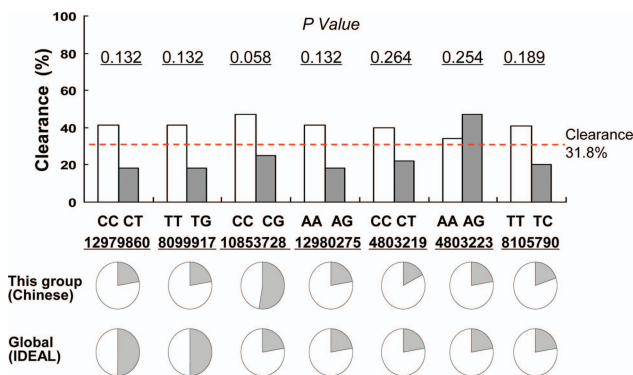


Figure 3 Interleukin-28 (IL-28) polymorphisms with hepatitis C virus (HCV) spontaneous clearance. IL-28 rs10853728 CC and HCV clearance ($p=0.058$). IL-28 single nucleotide polymorphisms (SNPs) and associations with spontaneous clearance of HCV. IL-28 SNPs, rs12979860 CC, rs8099917 TT and rs10853782 prevalence in Chinese patients.

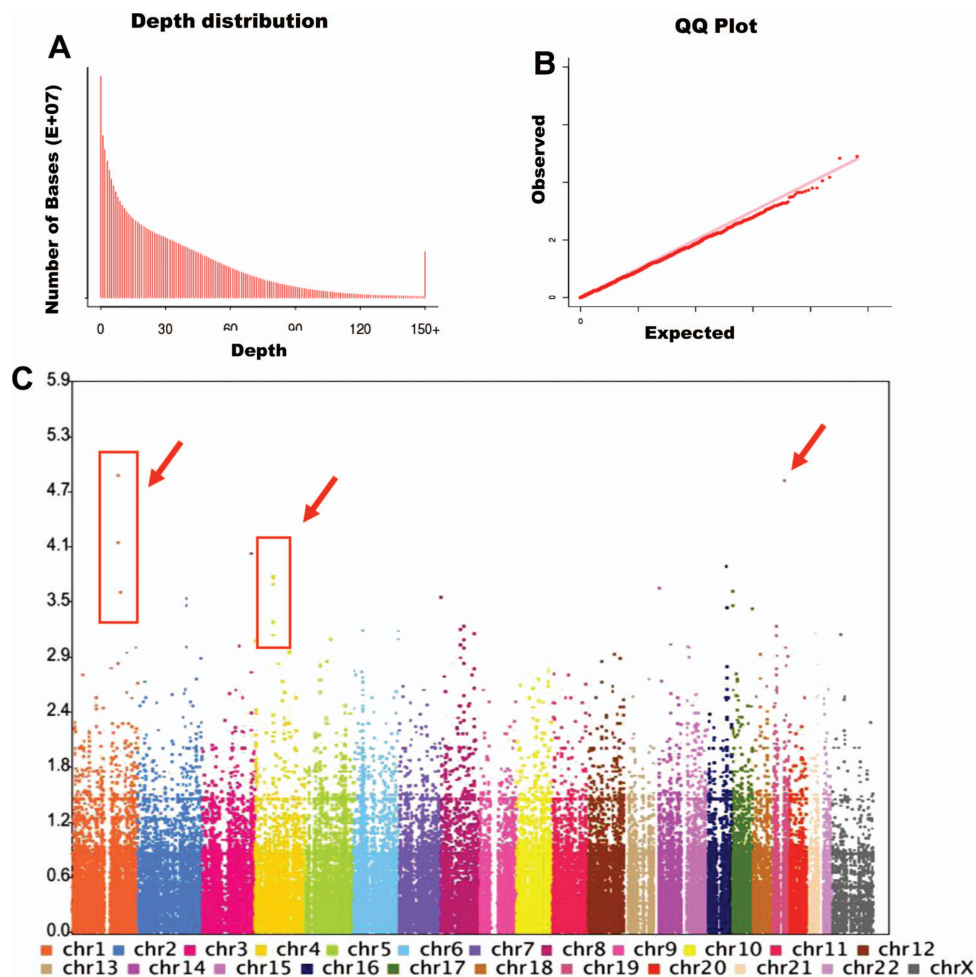


Figure 4 Exome capture and sequencing assay showing that single nucleotide polymorphisms (SNPs) are associated with spontaneous clearance of hepatitis C virus (HCV). (A) Depth distribution. (B) QQ plot to assess the discrepancy between the predicted value and the observed value. (C) A total of 64 449 SNPs were called from individuals, of which 400 were found to be associated with viral clearance by individual genotype calling. Two SNPs in tenascin-R (TNR), four in transmembrane protease serine 11A (TMPRSS11A), and one in serine peptidase inhibitor kunitz type 2 (SPINT2) showed the closest association ($p < 10^{-5}$).

and gender, these two studies provide less data on how age and gender affect HCV spontaneous clearance and disease progression.

IL-28B has recently been found to be a promising gene marker associated with treatment response and spontaneous clearance.^{3–7} However, data on the association between IL-28B genetic variants and spontaneous viral clearance in a Chinese study involving 376 HCV-infected paid plasma donors did not show an association with rs12979860. The other four SNPs, rs8099917, rs8105790, rs12980275 and rs10853728, were significantly associated with spontaneous HCV clearance.²³ In our study, IL-28 rs10853728 showed a stronger association than the other seven SNPs, but the association fell short of statistical significance ($p=0.058$). This may be because the number of patients in this cohort was not large enough to provide statistical power to the differences. Also, in the highly prevalent IL-28B, favourable genotype area and, in particular, the Chinese

population (>80%) may be different from the Caucasian population (40–50%).^{3,4}

TNR rs2239818 and rs34842046 are involved with receptor binding, extracellular matrix organisation and negative regulation of cell adhesion in the cell surface or extracellular region.^{24–25} Transmembrane protease TMPRSS11A rs1370840, rs11930532, rs28437478, rs6552134 and rs977728 are expressed in the normal liver, oesophagus, colon and lung, but downregulated in tumours.^{26–27} Neither TNR nor TMPRSS11A had been reported to be associated with spontaneous clearance of HCV. SPINT2 rs3745948 functions as an endopeptidase inhibitor within the plasma membrane, cytoplasm or extracellular region. Methylated SPINT2 and SRD5A2, combined with AFP and PIVKA-II, have been reported to be the most satisfactory panel to detect HCC in patients with chronically infected HCV.^{28–29} The mechanisms by which these three categories of SNPs result in the associations are not known. However, data

Table 3 List of top 20 SNP differences in clearance and non-clearance patients

	Gene name	SNP ID	SNP in chromosome	Minor allele	Major allele	Frequency in clearance	Frequency in controls	p Value	OR	Function
1	TNR	rs2239818	chr1_173642258	A	G	0.01	0.22	1.28E-05	0.04	Receptor binding; negative regulation of cell adhesion
2	SPINT2	rs3745948	chr19_43471696	T	C	0.05	0.33	1.46E-05	0.12	Serine-type endopeptidase inhibitor activity
3	TNR	rs34842046	chr1_173559295	T	A	0.02	0.22	6.67E-05	0.08	Receptor binding; negative regulation of cell adhesion
4	HTR3E	rs7627615	chr3_185301110	G	A	0.35	0.04	8.81E-05	11.69	Receptor activity; extracellular ligand-gated ion channel activity
5	TMPRSS11A	rs1370840	chr4_68492841	A	G	0.31	0.02	0.000157	18.97	TMPRSS11A expression in normal oesophagus, liver, colon and lung, with downregulated expression in tumours
6	TMPRSS11A	rs11930532	chr4_68494878	C	T	0.29	0.02	0.000159	18.91	TMPRSS11A expression in normal oesophagus, liver, colon and lung, with downregulated expression in tumours
7	TMPRSS11A	rs28437478	chr4_68493016	A	T	0.29	0.02	0.000189	18.53	TMPRSS11A expression in normal oesophagus, liver, colon and lung, with downregulated expression in tumours
8	SALL2	rs1263810	chr14_21061466	G	C	0.16	0.45	0.000206	0.23	DNA binding; sequence-specific DNA binding transcription factor activity
9	SPATA 22	rs9901726	chr17_3290284	T	C	0.02	0.19	0.000223	0.09	SPATA22
10	SPATA 22	rs2291604	chr17_3299188	G	C	0.02	0.19	0.000223	0.09	SPATA 22
11	NMNAT2	rs607332	chr1_181519836	A	G	0.18	0.48	0.000227	0.24	Nucleotide binding; nicotinamide-nucleotide adenyltransferase activity
12	NCKAP1	Novel	chr2_183540400	T	C	0.03	0.22	0.000266	0.11	Protein binding; plasma membrane; apoptosis
13	SPATA 22	rs9900543	chr17_3296435	C	T	0.02	0.19	0.000316	0.09	SPATA 22
14	NCKAP1	Novel	chr2_183501325	T	C	0.03	0.22	0.000319	0.12	Protein binding; plasma membrane; apoptosis
15	MARVELD3	rs2303225	chr16_70232383	C	G	0.42	0.11	0.000333	5.61	Membrane; integral to membrane;
16	TMPRSS11A	rs6552134	chr4_68511802	G	A	0.27	0.02	0.00048	16.25	TMPRSS11A expression in normal oesophagus, liver, colon and lung, with downregulated expression in tumours
17	ZNF491	Novel	chr19_11778746	A	C	0.01	0.15	0.000525	0.06	DNA binding; zinc ion binding
18	ZNF440	Novel	chr19_11804552	A	G	0.01	0.15	0.000525	0.06	DNA binding; zinc ion binding
19	ZNF439	Novel	chr19_11840305	G	C	0.01	0.15	0.000525	0.06	DNA binding; zinc ion binding
20	CA2	rs2307075	chr8_86575480	C	A	0.34	0.66	0.000528	0.27	Kidney development; morphogenesis of an epithelium; carbonate dehydratase activity; protein binding

CA2, carbonate dehydratase II; HTR3E, 5-hydroxytryptamine receptor 3 family member E; MARVELD3, MARVEL domain containing 3; NCKAP1, NCK-associated protein 1; NMNAT2, nicotinamide mononucleotide adenyltransferase 2; SALL2, sal-like protein 2; SNP, single nucleotide polymorphism; SPATA 22, spermatogenesis associated 22; SPINT2, serine peptidase inhibitor, kunitz type 2; TMPRSS11A, transmembrane protease, serine 11A; TNR, tenascin receptor; ZNF, zinc finger protein;

from functional analyses suggest that they may affect HCV binding to the receptors and entering hepatocytes.

There were two main limitations to the study. The first was the small number of patients. Though we cannot increase the sample size of this unique group since HCV were accidentally infected, the limited statistical power made us interpret the data more conservatively. Future studies, based on another larger scale cohort of patients with HCV and controls, are ongoing to verify those observations. Since this was a retrospective study, the other limitation was that we could not get the clinical data from the acute HCV phase about 10 years ago. The development of jaundice, other symptoms and laboratory findings during the acute phase could not be evaluated for spontaneous HCV clearance.

In summary, this unique single-source HCV1b-infected patient population allows analysis of HCV1b with spontaneous clearance from a new perspective. Host gene SNPs within *Tenascin-R*, *TMPRSS11A*, and *SPINT2* and *IL-28* most likely play roles in the HCV spontaneous clearance and disease progression.

Author affiliations

¹Liver Research Center, Beijing Key Laboratory of Translational Medicine in Liver Cirrhosis, Beijing Friendship Hospital, Capital Medical University, Beijing, China

²Department of Infectious Diseases, Qiannan People's Hospital, Guizhou, China

³Department of Infectious Diseases, Pingtang People's Hospital, Guizhou, China

⁴Beijing Genomic Institute, Shenzhen, Guangdong, China

⁵Hepatology Institute, Peking University People's Hospital, Beijing, China

⁶Department of Microbiology, Peking University Health Science Center, Beijing, China

⁷Department of Infectious Diseases, Guiyang Medical College, Guizhou, China

Contributors HY contributed to patient management and analysis of the risk factors. SL was involved in local patient management. YX conducted local patients' follow-up and maintained records. RC conducted the IL-28 SNPs analysis. YaS contributed to IL-28 detection. JR conducted the Exome capture and sequencing experiments and analysis. KW conducted SNPs data analysis. XJ contributed to SNPs comparison and analysis. YuS conducted bioinformatics analysis. HaZ was involved in HCV RNA detection. JL conducted patient follow-up and is a consultant. LW was involved in HCV RNA analysis and disease stage confirmation. HuZ was involved in study design and is a consultant. MC contributed to sample management and patient management. JJ contributed to the study design and final approval of the paper.

Funding This work was supported by the Program for National Science and Technology Major Project (2013ZX10002004, 2012ZX10002003), Key Project from the Education Bureau of Beijing (KZ201210025024).

Competing interests None.

Patient consent Obtained.

Ethics approval The study was approved by the Ethics Committee from Guiyang Medical College, Guizhou, China.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided

the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Seeff LB. Natural history of chronic hepatitis C. *Hepatology* 2002;36: S35–46.
2. Mosley JW, Operskalski EA, Tobler LH, *et al.* Viral and host factors in early hepatitis C virus infection. *Hepatology* 2005;42:86–92.
3. Thomas DL, Thio CL, Martin MP, *et al.* Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 2009;461:798–801.
4. Ge D, Fellay J, Thompson AJ, *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009;461:399–401.
5. Suppiah V, Moldovan M, Ahlenstiel G, *et al.* IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet* 2009;41:1100–4.
6. Tanaka Y, Nishida N, Sugiyama M, *et al.* Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet* 2009;41:1105–9.
7. Rauch A, Kutalik Z, Descombes P, *et al.* Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* 2010;138:1338–45.
8. Foucher J, Chanteloup E, Vergniol J, *et al.* Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study. *Gut* 2006;55:403–8.
9. Ziol M, Handra-Luca A, Kettaneh A, *et al.* Non-invasive assessment of liver fibrosis by stiffness measurement: a prospective multicentre study in patients with chronic hepatitis C. *Hepatology* 2005;41:48–54.
10. Castéra L, Vergniol J, Foucher J, *et al.* Prospective comparison of transient elastography, Fibrotest, APRI and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 2005;128:343–50.
11. Ng SB, Turner EH, Robertson PD, *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–6.
12. Qi XP, Ma JM, Du ZF, *et al.* RET germline mutations identified by exome sequencing in a Chinese multiple endocrine neoplasia type 2A/familial medullary thyroid carcinoma family. *PLoS ONE* 2011;6: e20353.
13. Gui Y, Guo G, Huang Y, *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 2011;43:875–8.
14. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. *N Engl J Med* 1999;340:1228–33.
15. Tillmann HL, Thompson AJ, Patel K, *et al.* A polymorphism near IL28B is associated with spontaneous clearance of acute hepatitis C virus and jaundice. *Gastroenterology* 2010;139:1586–92.
16. Duffy M, Salemi M, Sheehy N, *et al.* Comparative rates of nucleotide sequence variation in the hypervariable region of E1/E2 and the NS5b region of hepatitis C virus in patients with a spectrum of liver disease resulting from a common source of infection. *Virology* 2002;301:354–64.
17. Fanning LJ, Levis J, Kenny-Walsh E, *et al.* Viral clearance in hepatitis C (1b) infection: relationship with human leukocyte antigen class II in a homogeneous population. *Hepatology* 2000;31:1334–7.
18. Barrett S, Goh J, Coughlan B, *et al.* The natural course of hepatitis C virus infection after 22 years in a unique homogenous cohort: spontaneous viral clearance and chronic HCV infection. *Gut* 2001;49:423–30.
19. Mosbrugger TL, Duggal P, Goedert JJ, *et al.* Large-scale candidate gene analysis of spontaneous clearance of hepatitis C virus. *J Infect Dis* 2010;201:1371–80.
20. Wiese M, Berr F, Lafrenz M, *et al.* Low frequency of cirrhosis in a hepatitis C (genotype 1b) single-source outbreak in Germany: a 20-year multicenter study. *Hepatology* 2000;32:91–6.
21. Wiese M, Grüngreiff K, Güthoff W, *et al.*; East German Hepatitis C Study Group. Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany—a 25-year multicenter study. *J Hepatol* 2005;43:590–8.
22. Casiraghi MA, De Paschale M, Romanò L, *et al.* Long-term outcome (35 years) of hepatitis C after acquisition of infection through mini transfusions of blood given at birth. *Hepatology* 2004;39:90–6.

23. Rao HY, Sun DG, Jiang D, *et al.* IL28B genetic variants and gender are associated with spontaneous clearance of hepatitis C virus infection. *J Viral Hepat* 2012;19:173–81.
24. Tucker RP, Chiquet-Ehrismann R. The regulation of tenascin expression by tissue microenvironments. *Biochim Biophys Acta* 2009;1793:888–92.
25. Huang W, Zhang L, Niu R, *et al.* Tenascin-R distinct domains modulate migration of neural stem/progenitor cells in vitro. *In Vitro Cell Dev Biol Anim* 2009;45:10–14.
26. Li LW, Li YY, Li XY, *et al.* A novel tumor suppressor gene ECRG4 interacts directly with TMPRSS11A (ECRG1) to inhibit cancer cell growth in esophageal carcinoma. *BMC Cancer* 2011;11:52.
27. Yueying W, Jianbo W, Hailin L, *et al.* ECRG1, a novel esophageal gene, cloned and identified from human esophagus and its inhibition effect on tumors. *Carcinogenesis* 2008;29:157–60.
28. Kongkham PN, Northcott PA, Ra YS, *et al.* An epigenetic genome-wide screen identifies SPINT2 as a novel tumor suppressor gene in pediatric medulloblastoma. *Cancer Res* 2008;68:9945–53.
29. Iizuka N, Oka M, Sakaida I, *et al.* Efficient detection of hepatocellular carcinoma by a hybrid blood test of epigenetic and classical protein markers. *Clin Chim Acta* 2011;412:152–8.