A Novel Approach to Identify Candidate Prognostic Factors for Hepatitis C Treatment Response Integrating Clinical and Viral Genetic Data

Alicia Amadoz[1,2] and Fernando González-Candelas[3,4]

[1]Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia. [2]Computational Genomics Department, Centro de Investigatión Príncipe Felipe. [3]Unidad Mixta Infección y Salud Pública FISABIO-Universidad de Valencia, Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Valencia, Spain. [4]CIBER en Epidemiología y Salud Pública, Spain.

ABSTRACT: The combined therapy of pegylated interferon (IFN) plus ribavirin (RBV) has been for a long time the standard treatment for patients infected with hepatitis C virus (HCV). In the case of genotype 1, only 38%–48% of patients have a positive response to the combined treatment. In previous studies, viral genetic information has been occasionally included as a predictor. Here, we consider viral genetic variation in addition to 11 clinical and 19 viral populations and evolutionary parameters to identify candidate baseline prognostic factors that could be involved in the treatment outcome. We obtained potential prognostic models for HCV subtypes la and lb in combination as well as separately. We also found that viral genetic information is relevant for the combined treatment assessment of patients, as the potential prognostic model of joint subtypes includes 9 viral-related variables out of 11. Our proposed methodology fully characterizes viral genetic information and finds a combination of positions that modulate inter-patient variability.

KEYWORDS: hepatitis C virus, treatment response, prognostic model, data integration, genetic variability, evolutionary genetics

Introduction

Hepatitis C virus (HCV) causes a disease that affects the human liver in more than 185 million people worldwide, 170 millions of whom are estimated to be chronic patients with increasing risk of developing cirrhosis and cancer of the liver.[1] The World Health Organization (WHO) estimates that between 3 and 4 million people are infected each year and about 70% of them will develop chronic hepatitis.[2] HCV belongs to the genus *Hepacivirus* from the *Flaviviridae* family and includes seven genotypes and more than 60 subtypes identified.[3] The distribution of HCV genotypes varies geographically,[4] with genotype la being (the prototype genotype) common in the United States and Western Europe and genotype 1b being the widely distributed genotype worldwide.

Despite several strategies against HCV being developed,[5] the combined therapy of pegylated interferon (IFN) plus ribavirin (RBV) has been for a long time the standard treatment for patients infected with HCV. IFN treatment is effective in 39% of patients,[6] but when combined with RBV, it is effective in more than 60% of patients.[7] However, there are differences in the response to treatment among the viral genotypes. In the case of genotype 1, about 48% of patients have a positive response to the combined treatment,[8] but in the case of genotypes 2 and 3, it is about 80%.[7] Moreover, the cost of HCV treatment is high,[9] it has numerous side effects, and it might not be appropriate for some patients.[10]

The selective pressure of drugs, the high replication rate of HCV, and its low replication fidelity are the main viral causes of treatment resistance. It is estimated that, on average, a nucleotide change is produced per replication cycle.[11] The identification of specific mutations and genetic patterns responsible for clinical phenotypes would improve diagnosis and treatment of patients.[12] Some studies have revealed that HCV's genetic variability contributes to its escape from the patient's immune response.[13,14]

HCV variability is not distributed evenly along its genome, and it affects differently to treatment in each genome region.[15] It has been established that the greater the immune pressure in a region, the higher the genetic variability,[16] and, therefore, most studies of treatment–variability relationship have focused on these genome regions. On the other hand, it has been suggested that it is the overall genome variability that influences treatment response.[17]

Treatment response. Different HCV treatment response types have been established depending on the number of weeks

until HCV-RNA levels in serum or plasma are not detectable.[7] A rapid virological response (RVR) appears at treatment week 4, an early virological response (EVR) appears at treatment week 12, an end-of-treatment response (ETR) appears at the end of 24 or 48 weeks of treatment, and a sustained virological response (SVR) appears at 24 weeks after cessation of treatment. It has been observed that the latter type of response depends mainly on the viral genotype.[18] In the case of HCV genotype 1, patients treated during 48 weeks have an SVR rate of 38%–48%.[8] On the other hand, a patient is considered a *non-responder* if HCV-RNA clearing from serum fails after 24 weeks of therapy, or *relapse* when HCV-RNA reappears in serum after therapy is discontinued.

Personalized therapy of HCV infection is a common practice due to the diversity of disease progression.[19] Identifying those patients that will respond or not to the treatment before starting it would increase therapeutic efficacy and reduce personal suffering. Viral, environmental, treatment, and host factors play important roles in the outcome of HCV infection and treatment response.[20–22] Several studies on treatment–outcome prediction have taken into account factors measured before treatment.[23–27] In general, these studies take into account variables that describe the patient from clinical (alanine transaminase levels, viral load in serum, kidney biopsy, etc) and demographic (age, sex, habits, etc) points of view and also include some viral variables such as genotype, core and interferon sensitivity determining region (ISDR) region substitutions, and some variability parameters of the E1E2 region. It has been shown that the most consistent factors with treatment outcome are the viral genotype and viral load.[28]

Phylogenetic predictors. A methodology to detect candidate genetic polymorphisms influencing clinical outcome from pathogen genomes[29] uses well-supported clades in a phytogeny as statistical predictors. Differences between clades were not well defined at the viral subtype level in our dataset; therefore, the statistical power of the previous methodology was diluted. Here, we propose an alternative methodology that overcomes the lack of statistical support at the phylogenetic subtype level but considers major determinants of genetic variation in the viral genome to help in the prediction of treatment response.

Materials and Methods

Clinical and epidemiological data were retrieved together with viral sequences from the local epiPATH bioinformatics platform.[30]

Patients. Serum samples were obtained from 49 patients infected with HCV genotype 1a (17 patients, of which 11 had a positive response and 6 had a negative response) and genotype 1b (32 patients, of which 12 had a positive response and 20 had a negative response).[31,32] In summary, our sample included 23 patients with a positive response and 26 patients with a negative response. All patients provided written consent to be included in the study, which was approved by the corresponding ethics

committees of the institutions involved (Hospital General de Valencia, Hospital Clínico Universitario de Valencia, and Hospital General de Alicante). The research was conducted in accordance with the principles of the Declaration of Helsinki.

Treatment response assessment was done by the institutions involved with the following criteria:

Positive response: absence of HCV-RNA in serum or a >2log viral load decline relative to the basal viral load at week 12.

Negative response: when there was no positive response or presence of HCV-RNA in serum after week 12.

The following demographic, clinical, and treatment variables were included in this study (Table 1): age, sex, Knodell index,[33] the ratio between Glutamic-oxaloacetic transaminase and Glutamic-pyruvate transaminase (GOT/GPT) serum levels, alanine transaminase (ALT) serum levels, treatment duration, completed treatment, number of treatment, IFN dose and RBV dose.

HCV sequences. Host's immune pressure affects HCV variants but not equally through all the viral genome regions.[32] Two of the most studied regions in relation to genetic variability and treatment response are the E1E2 and NS5A viral regions. Independently of the genotype, viral factors included in our study were calculated using a high number of partial sequences from both regions described elsewhere[34] to capture the amount of variability of the viral quasispecies. Specifically, we used 100 sequences per patient of 472 nucleotides from the E1E2 region (nucleotides 1310–1781 of HCV genome reference sequence with accession number D50481), which includes three hypervariable regions (HVR-1, HVR-2, and HVR-3) but does not include the E2-PePHD region. Regarding the NS5A region, we used between 25 and 96 sequences per patient of 743 nucleotides (nucleotides 6742–7484 in the HCV genome reference sequence), which includes the ISDR, protein kinase R binding domain (PKR-BD), and variable region 3 (V3) regions that have been related to the combined treatment outcome.[35] Therefore, viral genome variability from about 7,500 sequences were summarized in the viral factors.

RNA extraction, reverse transcription, amplification, cloning, and sequencing, were explained in detail elsewhere.[36] Briefly, after viral RNA extraction (High Pure Viral RNA Kit; Roche, Mannheim, Germany), reverse transcription reactions were performed with random hexadeoxynucleotides in order to prevent any bias during reactions due to unspecific oligonucleotides. Amplified DNA products for each region were purified with High Pure PCR product Purification Kit (Roche) and directly cloned into EcoRV-digested pBluescript II SK (+) phagemid (Stratagene, Heidelberg, Germany). Plasmid DNA was purified with High Pure Plasmid Isolation Kit (Roche). Cloned products for E1E2 region or NS5A region were sequenced using vector-based primers KS and SK (Stratagene).

The following intra-patient viral variables per region were included in this study (Table 2): genotype, number of non-synonymous substitutions per non-synonymous site (dN), number of synonymous substitutions per synonymous

Table 1. Demographic, clinical, and treatment factors.

VARIABLE	VALUES	PATIENTS (N = 49)
Outcome	Positive (%)	23 (46.94%)
	Negative (%)	26 (53.06%)
Age	Years	43.61 ± 12.122 (23;73)
Sex	Male (%)	34 (69.4%)
	Female (%)	15 (30.6%)
Knodell index		8 ± 3.403 (1;17)
GOT/GPT		0.601 ± 0.323 (0.3;2.3)
ALT		122.67 ± 74.463 (24;361)
Treatment duration	months	11.27 ± 1.987 (6;12)
Completed treatment	Yes (%)	43 (87.8%)
	No (%)	6 (12.2%)
Number of treatment	1 (%)	28 (57.1%)
	2 (%)	20 (40.8%)
	3 (%)	1 (2%)
IFN dose	3 mU/3tpw (%)	40 (81.6%)
	5 mU/3tpw (%)	3 (6.1%)
	90 g/day (%)	1 (2%)
	100 g/day (%)	3 (6.1%)
	120 g/day (%)	2 (4.1%)
RBV dose	mg/day	1040.82 ± 122.336 (800;1200)

Notes: Age, Knodell index, GOT/GPT, and ALT were measured at baseline. Data are shown as mean ± standard deviation unless stated otherwise.

site (dS), dN/dS ratio, total number of mutations (η), number of segregating sites (S), GC content (GC), haplotype diversity that accounts for the allele combinations of a genetic region (H), H^2, H^3, nucleotide diversity (π), π^2, π^3, mean number of nucleotide differences (κ), number of sites under positive selection (pts), and the most relevant parameters of neutrality tests: Tajima's D, Fu and Li's D^*, Fu and Li's F^*, and Fu's Fs. Genotype information was retrieved from the local epiPATH bioinformatics platform; polymorphism parameters were calculated with DnaSP,[37] and sites under positive selection were identified with CODEML.[38]

Sequence similarity. In addition to population and evolutionary viral parameters, we obtained a sequence similarity measure at the molecular level. We used a *consensus* viral protein sequence per region and patient (72 patients, including 49 whose clinical data were available) obtained before treatment. Then, we aligned and filtered the variable positions from the *consensus* sequences per genotype and viral regions. We applied a multiple correspondence analysis (MCA) per dataset with SPSS 13.0 statistical software[39] considering each amino acid position as a variable and its amino acid type as its value in terms of 20 characters. Therefore, patients were grouped using the functional-related polymorphisms detected in the viral sequence. This analysis allowed us to reduce the viral sequence information into fewer variables while preserving

its patient variability. We selected seven dimensions for E1E2 and 12 dimensions for NS5A following Cattell's criteria.[40] Dimensions were included as viral variables in the treatment–response modeling methodology.

Statistical methodology. The treatment response was modeled using a logistic regression with 49 patients and 66 variables, including viral, demographic, and clinical data. We applied two methodologies in obtaining the regression model: (a) using variable subgroups, and (b) using all variables together. In an epidemiological context, logistic regression coefficients are interpreted as the odds ratio (OR) logarithm, ie, the effect of a unit of change in its corresponding variable on having a positive treatment response. OR is an association measure between the treatment outcome and the variables included in the regression model.

In the subgroup-based method (a), we created four groups of variables depending on which environment they were related to: patients, E1E2 region, NS5A region, and MCA dimensions. Then, a subgroup-based model was balanced among different treatment response factors. It was obtained following a generalized linear model (GLM) approach with a stepwise selection process in R using a logit transformation. First, we applied the following methodology to each subgroup of variables. We generated a *minimum model* without variables and a *maximum model* with all variables. Then, variables were added from the *minimum model* to the *maximum model* and evaluated with their Chi-squared value. We filtered variables with a Chi-squared significance level value ≤0.05 and generated a *model 1*. Next, a backward–forward stepwise selection method was applied to *model 1*, and Akaike information criterion (AIC)[41] was used to evaluate models. Finally, subgroup-based models were joined in a new model, and interactions between variables were studied with the previous method. This methodology was applied to genotypes la and lb jointly (1a+1b) and separately.

Once the subgroup-based models were obtained, we used them to predict the treatment outcome of three datasets: (1) the 49 patients used to obtain the models, (2) 8 new patients with complete clinical information, and (3) 10 new patients with some missing clinical factors which were estimated following four methods: the expectation-maximization algorithm (EM) implemented both in R and SPSS, mean substitution, and SPSS regression estimation. Predictions were made in R using the corresponding subgroup-based model, and prediction results were compared with the observed treatment outcome.

In the all-variables method (b), we applied the LASSO methodology[42] implemented in R also to genotypes la and lb jointly and separately. The LASSO method selects variables penalizing regression coefficients so that, if the coefficients are not greater than a given threshold, the corresponding variables are not included in the model. This method is also used to study how subgroup variables are affecting each other. We used minimum estimated *lambda* value and its standard deviation as thresholds.

Table 2. Viral variables.

GENOTYPE	SUBTYPES	PATIENTS (N = 49)
1	1a (%)	17 (34.7%)
	1b (%)	32 (65.3%)
VARIABLE	**E1E2**	**NS5A**
dN	0.017 ± 0.013 (0;0.044)	0.005 ± 0.005 (0;0.022)
dS	0.041 ± 0.034 (0;0.122)	0.0379 ± 0.032 (0;0.128)
dN/dS	0.671 ± 1.238 (0;8.277)	0.147 ± 0.103 (0;0.479)
η	77.224 ± 43.394 (2;153)	77.449 ± 53.659 (2;177)
S	70.939 ± 38.698 (2;134)	74.224 ± 50.569 (2;163)
GC	0.594 ± 0.011 (0.569;0.621)	0.603 ± 0.008 (0.588;0.621)
H	0.863 ± 0.23 (0.039;0.999)	0.833 ± 0.285 (0.066;1)
H^2	0.797 ± 0.285 (0.002;0.998)	0.774 ± 0.332 (0.004;1)
H^3	0.75 ± 0.315 (0;0.997)	0.736 ± 0.351 ($3e^{-4}$;1)
π	0.022 ± 0.016 (0;0.058)	0.013 ± 0.011 ($9e^{-5}$;0.04)
π^2	0.001 ± 0.001 (0;0.003)	$2.93e^{-4} \pm 4.02e^{-4}$ ($8.1e^{-9}$;0.002)
π^3	$3.01e^{-5} \pm 4.45e^{-5}$ ($0;2e^{-4}$)	$7.95e^{-6} \pm 1.51e^{-5}$ ($0;7.33e^{-5}$)
k	10.458 ± 7.74 (0.039;27.33)	9.782 ± 8.183 (0.07;31.099)
D	-0.993 ± 0.985 (-2.58;0.89)	-1.418 ± 0.678 (-2.43;0.07)
D^*	-1.908 ± 1.627 (-5.45;1.88)	-2.567 ± 1.363 (-5.24;0.27)
F^*	-1.847 ± 1.507 (-5.21;1.76)	-2.546 ± 1.218 (-4.92;-0.17)
Fs	-46.341 ± 32.161 (-124.662;1.256)	-26.647 ± 19.965 (-85.549;1.662)
pts	3.388 ± 3.523 (0;13)	0.878 ± 1.409 (0;5)

Note: Data are shown as mean ± standard deviation (range) unless stated otherwise.

Results

Subgroup-based models. The best model obtained for HCV subtypes 1a and 1b combined (AIC = 24) included the following parameters: treatment duration and ALT levels from the patients' variables subgroup; some parameters related to the NS5A region subgroup, including H and the number of sites under positive selection; some parameters related to the E1E2 region subgroup including dS, H^3 and π^2; some MCA dimensions related to both viral regions: the 11th dimension of NS5A region and the 7th dimension of E1E2 region; and interactions between NS5A 11th dimension and E1E2 7th dimension, and between E1E2 dS and E1E2 H^3 (Table 3).

Table 3. Subgroup-based model of 1a+1b HCV genotypes.

PARAMETER	COEFFICIENT	STD. ERROR	X-STANDARDIZED COEFFICIENT	ODDS RATIO
(Intercept)	$-3.19e^3$	$3.66e^5$		
Treatment duration	$5.89e^2$	$6.14e^4$	1,170.26	$5.7e^{255}$
ALT	3.526	$3.88e^2$	262.56	$3.4e^1$
NS5A 11th	$-1.62e^2$	$1.75e^4$	$-2,310.48$	$5.9e^{-71}$
E1E2 7th	$2.68e$	$4.68e^3$	382.37	$4.2e^{11}$
E1E2 dS	$-1.43e^5$	$1.46e^7$	$-2,039,800.00$	0
E1E2 H^3	$-9.2e^2$	$1.09e^5$	$-11,065.69$	0
E1E2 pi^2	$1.94e^5$	$2.33e^7$	2,770,577.00	*inf*
NS5A H	$-3.01e^3$	$3.08e^5$	$-35,762.37$	0
NS5A pts	$8.57e$	$9.83e^3$	120.69	$1.6e^{37}$
NS5A 11th:E1E2 7th	$3.14e^2$	$3.29e^4$		$1.4e^{136}$
E1E2 dS:E1E2 H^3	$1.32e^5$	$1.35e^7$		*inf*

Note: AIC = 24.

With respect to the best model obtained for subtype la separately (AIC = 8), the following parameters were included: H^2 and GC content related to the E1E2 region subgroup, and S related to the NS5A region subgroup (Supplementary Table 1). On the other hand, the best model obtained for subtype lb separately (AIC = 14) included the following parameters: treatment duration included in the patients' variables subgroup; some parameters related to the NS5A region subgroup including dS, η, κ and dN; and the 11th dimension of NS5A region included in the MCA dimensions subgroup (Supplementary Table 2).

The goodness of fit of the best models was quite good despite the low sample size (Table 4). All predictions were correct using patients included in the development of prediction models (test dataset), and about 66%–70% of predictions were correct using a new dataset of patients: 6 new patients for the best model of subtype la, 9 new patients for the best model of subtype lb, and 10 new patients for the joint subgroup-based model. The selection of new patients for each validation dataset was performed taking into account the parameters included in each model, in such a way that not all the patients from separate subtypes had information about parameters required in the combined subtypes model. Therefore, the number of new patients used in the validation dataset of the combined subtypes model was lower than that of the separate subtypes model.

After obtaining the dimensional variables in the final models, we tried to find positions that modulate interpatient variability. We found that, in the NS5A region, the 11th dimension retained more than 95% of the variability in 10/69 polymorphisms of subtype la and in 100/127 of those in subtype lb. With regard to the E1E2 region, the seventh dimension retained more than 95% of the total variability in 27/65 polymorphisms of subtype la and 44/93 of lb polymorphisms. A summary of the positions that contribute with more than 3% individually can be found in Tables 5 and 6. We did not find a clear relationship between the type of amino acid in each position and treatment response (results not shown). Nevertheless, when comparing the positions with the highest individual contribution to global variability with the substitution patterns reported by Enomoto et al.[43,44], we found that IFN-sensitive amino acid substitutions in 386 and 388 codons of 1b E1E2 region had a greater variability contribution than IFN-resistant ones (Table 7).

Table 4. Goodness of fit from response predictions.

GENOTYPE	PATIENTS	
	INCLUDED	NOT INCLUDED
1a	17/17 (100%)	4/6 (66.66%)
1b	32/32 (100%)	6/9 (66.66%)
1a+1b	49/49 (100%)	7/10 (70%)

Notes: Included: patients used to obtain models; not included: patients not used to obtain models. Results shown as correct/total.

Table 5. Contribution of amino acid positions to the total variability of NS5A region.

SUBTYPE	SUBREGION	POSITION	CONTRIBUTION
1a		2143	0.169
		2198	0.169
		2203	0.084
	ISDR/PKR-BD	2221	0.042
		2304	0.042
		2306	0.084
		2350	0.038
	V3	2368	0.032
	V3	2376	0.071
		2381	0.169
		2382	0.084
1b		2146	0.08
		2169	0.032
	ISDR/PKR-BD	2229	0.031
	ISDR/PKR-BD	2233	0.032
	PKR-BD	2264	0.079
	PKR-BD	2265	0.049
		2353	0.08
	V3	2376	0.043

Notes: Selected positions that contribute individually with a >3% to the total genetic variation in the MCA analysis. Position is given as the corresponding amino acid position in the HCV reference sequence (D50481).

All-variables models. The main disadvantage of the subgroup-based approximation is that the influence between the parameters of different subgroups is not evaluated. Therefore, we applied a secondary methodology to check the influence between all variables included in this study. We used minimum λ (min) and its standard deviation (1 se) as threshold coefficients, the former being the most conservative approach (Fig. 1). The prediction model obtained for HCV subtypes la and lb combined with GLMNET methodology includes parameters from different subgroups defined for the subgroup-based methodology (Table 8). Therefore, it can be concluded that there is no relevant influence among subgroups of variables.

Discussion

The aim of this study was to identify candidate baseline prognostic factors that could be involved in the response of patients infected with HCV subtypes la and lb to combined treatment with IFN and RBV. Although HCV drug therapies have experienced a recent change with the availability of new antiviral drugs,[5] the time required to design and conduct treatment–response studies led us to outline a retrospective

Table 6. Contribution of amino acid positions to the total variability of E1E2 region.

SUBTYPE	SUBREGION	POSITION	CONTRIBUTION	POSITIVE SELECTION
1a	E1	357	0.031	
	E2-HVR1	384	0.030	True
	E2-HVR1	393	0.070	
	E2-HVR1	394	0.048	True
	E2-HVR1	397	0.030	True
	E2-HVR1	399	0.098	True
	E2-HVR1	401	0.048	True
	E2-HVR1	406	0.064	
	E2-HVR1	407	0.070	True
	E2-HVR1	410	0.033	True
	E2	418	0.030	
	E2	419	0.033	
	E2-HVR3	436	0.030	
	E2-HVR3	440	0.058	
	E2-HVR3	444	0.037	True
	E2	452	0.041	
1b	E2-HVR1	386	0.058	True
	E2-HVR1	388	0.054	True
	E2-HVR1	392	0.038	True
	E2-HVR1	393	0.034	
	E2-HVR1	397	0.076	True
	E2-HVR1	407	0.031	True
	E2-HVR2	478	0.062	True
	E2-HVR2	480	0.042	True

Notes: Selected positions that contribute individually with a >3% to the total genetic variation in the MCA analysis. Position is given as the corresponding amino acid position in the HCV reference sequence (D50481). Positive selection indicates whether the position has a significant positive selection or not.[45]

Table 7. Common substitution patterns with Enomoto et al.[43,44]

REGION	SUBREGION	POSITION	TOTAL	AMINO ACID	INDIVIDUAL
E1E2	E2-HVR1	386	0.058	T^{sen}	0.029
				D	0.024
				G	0.002
				N	$1.55e^{-03}$
				S	$0.76e^{-03}$
				E	$0.55e^{-03}$
				R^{sen}	$8.02e^{-05}$
				Q	$4.01e^{-05}$
				$H^{res}*$	$1.13e^{-06}$
				Y^{res}	-
	E2-HVR1	388	0.054	H	0.024
				R	0.024
				Y	0.004
				$T^{sen}*$	$0.60e^{-03}$
				N	$0.52e^{-03}$
				V	$4.77e^{-05}$
				ores	-
NS5A		2169	0.032	s	0.011
				$A^{sen/res}*$	0.005
				E	0.004
				H	$0.82e^{-03}$
				$T^{sen/res}$	-

Notes: Codon positions that contribute individually with a >3% to the total variability of the seventh dimension in the MCA analysis. Substitutions refer to HCV lb subtype. Selected positions were found in common with the substitution patterns reported by Enomoto et al.[43,44] Position is given as the corresponding amino acid position in the HCV reference sequence (D50481). *Total* column indicates the total contribution of the position and *Individual* column indicates the individual contribution of the specific amino acid. *sen* indicates IFN-sensitive, *res* indicates IFN-resistant, and *indicates HCV-J as reported in Enomoto et al.[43,44]
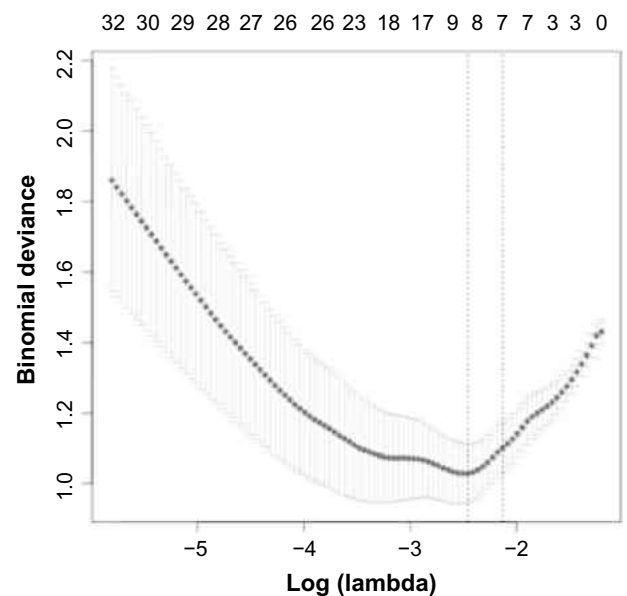
study with data previously generated in our group.[31,32] We used treatment and patient variables along with more viral variables than similar previous studies.[23–27] A new viral factor included was a multidimensional measure of sequences similarity that accounts for inter-patient viral variability. The hypothesis based on which this study was designed is that the integration of treatment response with viral sequences data would provide new insights into the interaction between different viral genomic regions and the treatment outcome, which eventually would improve our understanding of the viral evolution role towards patient therapy. In this sense, our proposed methodology could be applied in future studies, which include the recently developed drugs.

Figure 1. Parameter estimates based on the LASSO method for the HCV subtypes 1a+1b combined.

Table 8. LASSO-based model of 1a+1b HCV genotypes.

1SE		MIN	
PARAMETER	**COEFFICIENT**	**PARAMETER**	**COEFFICIENT**
(Intercept)	0.893	(Intercept)	0.715
NS5A H	−0.057	E1E2 Tajima's D	0.001
NS5A H^2	−0.003	NS5A S	−0.003
NS5A H^3	−0.0004	NS5A H	−0.059
NS5A 11th	−0.010	NS5A H^2	−0.004
Treatment duration	0.029	NS5A H^3	−0.0004
IFN dose	−0.111	NS5A 11th	−0.019
Treatment number	0.412	Treatment duration	0.053
		IFN dose	−0.250
		Treatment number	0.872

Different models were obtained with two distinct methodologies (Table 3, Supplementary Tables 1 and 2 with subgroup-based methodology; and Table 8 with LASSO-based methodology). The main difference between the two methods is that the LASSO method studies how subgroup variables affect each other. Results from LASSO-based method demonstrate that variables included belong to different groups defined in subgroup-based methodology. Therefore, it can be concluded that there is no relevant influence between subgroups of variables. The subgroup-based method weights all subgroups equally and evaluates parameters thoroughly.

Subgroup-based model for both viral genotypes is balanced regarding variable subgroups, because it includes variables from the different subgroups defined for this methodology (Table 3). It includes treatment duration and ALT levels as patient variables. The current recommended duration of treatment for genotype 1-infected patients is 48 weeks,[7] although an extension to 72 weeks in patients without RVR has been proposed[46] and a reduction to 24 weeks in those patients with RVR.[47] Before treatment, ALT levels could represent the patient's immune system activity, as they include the elimination of infected cells by natural killer (NK) cells and cytotoxic T lymphocytes (CTL).[48,49] The fact that ALT levels appear in the model could indicate that the higher the immune activity, the more effective the treatment. In this respect, ALT levels have also been obtained in other models.[24,27] Nevertheless, it has been shown that the SVR rate in patients with normal levels of ALT is equivalent to patients with higher ALT levels[50] and there are other factors that could influence before-treatment ALT levels such as an imbalance of fatty acids and carbohydrates metabolism,[51] alcohol abuse,[52] and other drugs.[53]

Sequence similarity measures at the molecular level are obtained in our results. NS5A 11th and E1E2 7th dimensions

account for specific amino acids in certain sequence positions. There has been extensive discussion about the relationship between treatment response and mutations in both regions. There is a correlation between treatment response and NS5A substitutions in Japanese patient cohorts,[54,55] but it has not been reported in European[56,57] or American patients.[58] However, some European studies have found a correlation between the ISDR sequence of genotype lb and treatment response.[59,60] After a long controversy, a correlation between NS5A's ISDR region and treatment response has been reported in three different meta-analyses[61–63] and in more recent studies.[64,65] Moreover, substitutions in other NS5A regions seem to be related to treatment response,[35,65,66] and recently an association between an SVR and combined mutations in NS5A and core regions has been shown.[67]

With regard to the genetic variation in the E1E2 region, some studies have found no relationship between the E2-PePHD region and treatment response because it is a highly conserved region and its variability levels do not differ between responder and non-responder patients,[35,68] although some studies have found this relationship.[69,70] Despite the fact that the E2-PePHD region was not included in our study, we found that some originally reported IFN-sensitive codon substitutions[43] had greater contribution to the total variability than IFN-resistant in specific E2-HVR1 positions (Table 7). An interesting aspect of our results is that, in addition to including sequence similarity variables in the final la+lb HCV genotype subgroup-based model, we also obtained the interaction of both NS5A 11th and E1E2 7th dimensions. This result could indicate that a joint substitution profile of both regions might increase the chances of responding to treatment. In this respect, a correlation between the number of mutations in the E2 and ISDR regions has been found[70]; and another study obtained significant results for the correlation between nucleotide diversity of both E1E2 and NS5A regions.[34] A possible explanation of these observations could be that variants in both viral regions interact in an epistatic way due to the functional and/or structural relationship between them, as it has been previously suggested with E2, NS2 y NS5A regions.[71]

Some variables related to the viral response to selection were included in the joint subgroup-based model for HCV subtypes 1a+1b. This is the case for the rate of synonymous substitutions (dS) in the E1E2 region and the number of positions under positive selection in the NS5A region. The possibility of positive selection affecting viral escape from the patient's immune system has been previously studied[45,72,73] but without any evidence of relationship with treatment response. The HCV E1E2 region includes some hypervariable regions that tend to accumulate amino acid changes during viral infection, and it is known that the main evolutionary force that affects this region is purifying selection due to the functional restrictions.[74] In this sense, E1E2 dS obtained in our results would indicate the presence of purifying selection because the higher the dS, the lower the ω, causing a decrease

in the probability of a treatment response. Regarding the NS5A region, a high number of positions under positive selection could cause a wrong PKR inhibition and reduce viral replication, which would lead to a higher probability of treatment response. However, conflicting results on the action of positive selection in the NS5A region of SVR patients have been reported.[66,72]

Intra-patient viral variability factors that are included in the final 1a+1b HCV genotypes subgroup-based model are E1E2 H^3, E1E2 π^2, and NS5A H. Viral genetic diversity before treatment is higher in nonresponsive patients than in responsive patients,[17,34] and a higher haplotype diversity could promote the appearance of treatment-resistant variants that allow the escape from the immune system, as seen previously.[71] On the other hand, there seems to exist a diversity of nucleotide positions among patients that would disallow the virus from entering the cell, new viral particle assembly, antibody neutralization, and/or their interaction with PKR, despite the fact that a high variability would decrease the patient's immune response.[75] Recently, it has been shown that a higher NS5A variability is correlated with a positive response to treatment.[76] Though genetic diversity is not usually considered as a good SVR predictor,[77] a possible explanation of its inclusion in our model is that we have transformed these variables and detected the amount of information related to the treatment outcome. Another interesting aspect of our results is that they include the interaction between E1E2 dS and E1E2 H^3, which could indicate that a certain amount of purifying selection together with the presence of some E1E2 haplotypes would increase the probability of response to treatment.

Models obtained for 1a and 1b genotypes separately are completely different (Supplementary Tables 1 and 2). Variables included in genotype 1a results are related to viral polymorphisms, but variables obtained for genotype 1b are related to treatment, viral polymorphisms, and sequence similarity measures. The latter variable appears also in the model for both viral genotypes but not in the 1a model. In this sense, a significant correlation between ISDR mutations and treatment response has been observed in 1b but not in 1a patients.[60]

It is not always easy to provide a biological explanation when obtaining models under statistical criteria. We have provided a biological interpretation for every variable included in the final 1a+1b HCV genotype subgroup-based model; however, all variables should be considered together in the interpretation of our results. The methodologies that we used in this study include variables depending on their statistical significance; and because of high standard errors in the subgroup-based methodology, more tests would be necessary before applying these results to personalized therapy. Moreover, coefficients and OR values shown are the final result from the complete statistical methodology applied and should not be interpreted as their final individual significance. In this respect, we have used statistical methods for the identification of relevant candidate prognostic factors for HCV treatment response and not for the quantification of their individual effect in treatment response.

Nevertheless, our results indicate that viral genetic information is essential for the IFN–RBV combined treatment assessment of patients. In this sense, identifying a profile of combined mutations along viral genome regions that modulate treatment outcome could help treatment management, reducing costs and side effects. In general, our methodology can be applied to identify different joint-substitution patterns that could arise comparing the new HCV therapies that are currently being developed, eg, different patients, drug combinations, different time points of the treatment, in order to assess the best therapy approach for each case. Moreover, it could also be applied in the study of different viruses as well as in co-infections.

Conclusions

Population and evolutionary parameters quantify genetic sequences in terms of variability and its functional effect. In a high replication rate and a low replication fidelity scenario such as RNA viruses, the action of selective pressures could modulate the treatment response. Therefore, the relevance and interest of studying viral populations under an evolutionary perspective has a direct application on therapy improvement. As far as we know, population and evolutionary parameters together with the complete sequence variability have not been used before to study HCV treatment response.

In our study, we have found that these kinds of parameters are relevant prognostic factors, as they have been included in the best prognostic models obtained from different datasets. The best prognostic model of 1a and 1b joint subtypes includes 9 out of 11 variables related to population and evolutionary parameters. We have discussed the interpretation of each of them separately, and we have also given some insights of the applicability of our proposed new measurement to related studies.

The integration of clinical and viral genetic data is an important issue for the evaluation of different factors related to HCV treatment response. In this study, a new viral factor that accounts for inter-patient viral variability was suggested. One of the advantages that we have found by using the sequence similarity measure is that we were able to reduce the number of parameters for regression analyses. Moreover, it preserved patient variability in terms of complete viral sequences and could be integrated with clinical data. These characteristics make our multidimensional measure of sequence similarity useful for the identification of joint substitutions profiles that might modulate the chances of responding to treatment.

Our proposed new methodology could be applied in related studies to identify viral positions involved in treatment response and also in a comparison of new therapies at different time points to study the evolution of viral joint-mutation profiles regarding treatment outcome.

Author Contributions

Conceived and designed the experiments: AA, FG-C. Analyzed the data: AA. Wrote the first draft of the manuscript: AA. Contributed to the writing of the manuscript: AA, FG-C. Agree with manuscript results and conclusions: AA, FG-C. Jointly developed the structure and arguments for the paper: AA, FG-C. Made critical revisions and approved final version: AA, FG-C. Both authors reviewed and approved of the final manuscript.

Supplementary Data

Supplementary Table 1. Subgroup-based model of la HCV genotype.

Supplementary Table 2. Subgroup-based model of lb HCV genotype.

REFERENCES

1. Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology*. 2013;57(4):1333–42.
2. Hauri AM, Armstrong GL, Hutin YJ. The global burden of disease attributable to contaminated injections given in health care settings. *Int J STD AIDS*. 2004;15(1):7–16.
3. Smith DB, Bukh J, Kuiken C, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology*. 2014;59(l):318–27.
4. Simmonds P, Bukh J, Combet C, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*. 2005;42(4):962–73.
5. Pawlotsky JM. New hepatitis C therapies: the toolbox, strategies, and challenges. *Gastroenterology*. 2014;146(5):1176–92.
6. Zeuzem S, Feinman SV, Rasenack J, et al. Peginterferon alfa-2a in patients with chronic hepatitis C. *N Engl J Med*. 2000;343(23):1666–72.
7. Ghany MG, Strader DB, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology*. 2009;49(4):1335–74.
8. Druyts E, Mills EJ, Nachega J, O'Regan C, Cooper CL. Differences in clinical outcomes among hepatitis C genotype 1-infected patients treated with peginterferon alpha-2a or peginterferon alpha-2b plus ribavirin: a meta-analysis. *Clin Exp Gastroenterol*. 2012;5:11–21.
9. Shepherd J, Brodin HFT, Cave CB, Waugh NR, Price A, Gabbay J. Clinical- and cost-effectiveness of pegylated interferon alfa in the treatment of chronic hepatitis C: a systematic review and economic evaluation. *Int J Technol Assess Health Care*. 2005;21(1):47–54.
10. Manns MP, Wedemeyer H, Cornberg M. Treating viral hepatitis C: efficacy, side effects, and complications. *Gut*. 2006;55(9):1350–9.
11. Cuevas JM, González-Candelas F, Moya A, Sanjuán R. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J Virol*. 2009;83(11):5760–4.
12. González-Candelas F, López-Labrador FX. Clinical relevance of genetic heterogeneity in HCV. *Future Virol*. 2010;5(1):33–49.
13. López-Labrador FX, He XS, Berenguer M, et al. Genetic variability of hepatitis C virus non-structural protein 3 and virus-specific CD8+ response in patients with chronic hepatitis C. *J Med Virol*. 2004;72(4):575–85.
14. Sklan EH, Charuworn P, Pang PS, Glenn JS. Mechanisms of HCV survival in the host. *Nat Rev Gastroenterol Hepatol*. 2009;6(4):217–27.
15. Maekawa S, Enomoto N. Viral factors influencing the response to the combination therapy of peginterferon plus ribavirin in chronic hepatitis C. *J Gastroenterol*. 2009;44(10):1009–15.
16. Reed KE, Rice CM. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. *Curr Top Microbiol Immunol*. 2000;242:55–84.
17. Cuevas JM, Torres-Puente M, Jiménez-Hernández N, et al. Refined analysis of genetic variability parameters in hepatitis C virus and the ability to predict antiviral treatment response. *J Viral Hepat*. 2008;15(8):578–90.
18. Hadziyannis SJ, Sette H Jr, Morgan TR, PEGASYS International Study Group, et al. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med*. 2004;140(5):346–55.
19. Brown RS. Customizing treatment to patient populations. *Nat Clin Pract Gastroenterol Hepatol*. 2007;4(suppl l):S3–9.
20. Gao B, Hong F, Radaeva S. Host factors and failure of interferon-alpha treatment in hepatitis C virus. *Hepatology*. 2004;39(4):880–90.
21. Hofmann WP, Zeuzem S, Sarrazin C. Hepatitis C virus-related resistance mechanisms to interferon alpha-based antiviral therapy. *J Clin Virol*. 2005;32(2):86–91.
22. Mallat A, Hezode C, Lotersztajn S. Environmental factors as disease accelerators during chronic hepatitis C. *J Hepatol*. 2008;48(4):657–65.
23. Shirakawa H, Matsumoto A, Joshita S, Nagano Interferon Treatment Research Group, et al. Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology*. 2008;48(6):1753–60.
24. Trapero-Marugan M, Marin M, Pivel JP, et al. Predictive graphical model, network-based medical tool for the prognosis of chronic hepatitis C patients treated with peg-interferon plus ribavirin. *Aliment Pharmacol Ther*. 2008;28(4):468–74.
25. Okanoue T, Itoh Y, Hashimoto H, et al. Predictive values of amino acid sequences of the core and NS5 A regions in antiviral therapy for hepatitis C: a Japanese multi-center study. *J Gastroenterol*. 2009;44(9):952–63.
26. Kurosaki M, Matsunaga K, Hirayama I, et al. A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis. *Hepatol Res*. 2010;40(3):251–60.
27. Saludes V, Bascuñana E, Jordana-Lluch E, et al. Relevance of baseline viral genetic heterogeneity and host factors for treatment outcome prediction in hepatitis C virus lb-infected patients. *PLoS One*. 2013;8(8):e72600.
28. Lee SS. Indicators and predictors of response to anti-viral therapy in chronic hepatitis C. *Aliment Pharmacol Ther*. 2003;17(5):611–21.
29. Szmaragd C, Nichols RA, Balloux F. A novel approach to characterise pathogen candidate genetic polymorphisms involved in clinical outcome. *Infect Genet Evol*. 2006;6(1):38–45.
30. Amadoz A, González-Candelas F. epiPATH: an information system for the storage and management of molecular epidemiology data from infectious pathogens. *BMC Infect Dis*. 2007;7(1):32.
31. Jiménez-Hernández N. *Evolucion del virus de la hepatitis C en muestras hospitalarias de la Comunidad Valenciana* [PhD thesis]. Valencia: Universidad de Valencia; 2004.
32. Torres-Puente M. *Variabilidad genetica y respuesta al tratamiento antiviral en el virus de la Hepatitis C (VHC)* [PhD thesis]. Valencia: Universidad de Valencia; 2004.
33. Desmet Valeer J, Knodell RG, Ishak KG, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis [Hepatology 1981;1:431435]. *J Hepatol*. 2003;38(4):382–6.
34. Torres-Puente M, Cuevas JM, Jiménez-Hernández N, et al. Genetic variability in hepatitis C virus and its role in antiviral treatment response. *J Viral Hepat*. 2008;15(3):188–99.
35. Muñoz de Rueda P, Casado J, Patón R, et al. Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol*. 2008;82(13):6644–53.
36. Jiménez-Hernández N, Torres-Puente M, Bracho MA, et al. Epidemic dynamics of two coexisting hepatitis C virus subtypes. *J Gen Virol*. 2007;88(1):123–33.
37. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003;19(18):2496–7.
38. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 1997;13(5):555–6.
39. SPSS Inc. *SPSS for Windows, Version 13.0*. Chicago: SPSS Inc; 2004.
40. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res*. 1966;1(2):245–76.
41. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716–23.
42. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
43. Enomoto N, Sakuma I, Asahina Y, et al. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus lb. Sensitivity to interferon is conferred by amino acid substitutions in the NS5 A region. *J Clin Invest*. 1995;96(1):224–30.
44. Enomoto N, Sakuma I, Asahina Y, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus lb infection. *N Engl J Med*. 1996;334(2):77–82.

45. Torres-Puente M, Cuevas JM, Jiménez-Hernández N, et al. Using evolutionary tools to refine the new hypervariable region 3 within the envelope 2 protein of hepatitis C virus. *Infect Genet Evol*. 2008;8(1):74–82.

46. Pearlman BL, Ehleben C, Saifee S. Treatment extension to 72 weeks of peginterferon and ribavirin in hepatitis c genotype 1infected slow responders. *Hepatology*. 2007;46(6):1688–94.

47. Zeuzem S, Buti M, Ferenci P, et al. Efficacy of 24 weeks treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C infected with genotype 1 and low pretreatment viremia. *J Hepatol*. 2006;44(1):97–103.

48. Neumann-Haefelin C, Blum HE, Chisari FV, Thimme R. T cell response in hepatitis C virus infection. *J Clin Virol*. 2005;32(2):75–85.

49. Golden-Mason L, Rosen HR. Natural killer cells: primary target for hepatitis C virus immune evasion strategies? *Liver Transpl*. 2006;12(3):363–72.

50. Zeuzem S, Diago M, Gane E, PEGASYS Study NR16071 Investigator Group, et al. Peginterferon alfa-2a (40 kilodaltons) and ribavirin in patients with chronic hepatitis C and normal aminotransferase levels. *Gastroenterology*. 2004;127(6):1724–32.

51. Prati D, Shiffman ML, Diago M, et al. Viral and metabolic factors influencing alanine aminotransferase activity in patients with chronic hepatitis C. *J Hepatol*. 2006;44(4):679–85.

52. Sorbi D, Boynton J, Lindor KD. The ratio of aspartate aminotransferase to alanine aminotransferase: potential value in differentiating nonalcoholic steatohepatitis from alcoholic liver disease. *Am J Gastroenterol*. 1999;94(4):1018–22.

53. Kaplowitz N. Drug-induced liver injury. *Clin Infect Dis*. 2004;38(suppl 2): S44–8.

54. Nakano I, Fukuda Y, Katano Y, Nakano S, Kumada T, Hayakawa T. Why is the interferon sensitivity-determining region (ISDR) system useful in Japan? *J Hepatol*. 1999;30(6):1014–22.

55. Murayama M, Katano Y, Nakano I, et al. A mutation in the interferon sensitivity-determining region is associated with responsiveness to interferonribavirin combination therapy in chronic hepatitis patients infected with a Japan-specific subtype of hepatitis C virus genotype 1b. *J Med Virol*. 2007;79(1): 35–40.

56. Zeuzem S, Lee JH, Roth WK. Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. *Hepatology*. 1997;25(3):740–4.

57. Duverlie G, Khorsi H, Castelain S, et al. Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *J Gen Virol*. 1998;79(pt 6):1373–81.

58. Chung RT, Monto A, Dienstag JL, Kaplan LM. Mutations in the NS5A region do not predict interferon-responsiveness in American patients infected with genotype 1b hepatitis C virus. *J Med Virol*. 1999;58(4):353–8.

59. Sarrazin C, Berg T, Lee JH, et al. Improved correlation between multiple mutations within the NS5 A region and virological response in European patients chronically infected with hepatitis C virus type 1b undergoing combination therapy. *J Hepatol*. 1999;30(6):1004–13.

60. Torres-Puente M, Cuevas JM, Jiménez-Hernández N, et al. Hepatitis C virus and the controversial role of the interferon sensitivity determining region in the response to interferon treatment. *J Med Virol*. 2008;80(2):247–53.

61. Witherell GW, Beineke P. Statistical analysis of combined substitutions in nonstructural 5 A region of hepatitis C virus and interferon response. *J Med Virol*. 2001;63(1):8–16.

62. Pascu M, Martus P, Höhne M, et al. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut*. 2004;53(9):1345–51.

63. Schinkel J, Spaan WJ, Kroes AC. Meta-analysis of mutations in the NS5A gene and hepatitis C virus resistance to interferon therapy: uniting discordant conclusions. *Antivir Ther*. 2004;9(2):275–86.

64. Yen YH, Hung CH, Hu TH, et al. Mutations in the interferon sensitivity-determining region (nonstructural 5A amino acid 22092248) in patients with hepatitis C-1b infection and correlating response to combined therapy of pegylated interferon and ribavirin. *Aliment Pharmacol Ther*. 2008;27(1):72–9.

65. Kumthip K, Pantip C, Chusri P, et al. Correlation between mutations in the core and NS5 A genes of hepatitis C virus genotypes 1a, 1b, 3a, 3b, 6f and the response to pegylated interferon and ribavirin combination therapy. *J Viral Hepat*. 2011;18(4):e117–25.

66. Queiróz ATL, Maracaja-Coutinho V, Jardim ACG, Rahal P, Carvalho-Mello IMVG, Matioli SR. Relation of pretreatment sequence diversity in NS5A region of HCV genotype 1 with immune response between pegylated-INF/ribavirin therapy outcomes. *J Viral Hepat*. 2011;18(2):142–8.

67. Hayashi K, Katano Y, Ishigami M, et al. Mutations in the core and NS5A region of hepatitis C virus genotype 1b and correlation with response to pegylated-interferon-alpha 2b and ribavirin combination therapy. *J Viral Hepat*. 2011;18(4):280–6.

68. Gaudy C, Lambelé M, Moreau A, Veillon P, Lunel F, Goudeau A. Mutations within the hepatitis C virus genotype 1b E2-PePHD domain do not correlate with treatment outcome. *J Clin Microbiol*. 2005;43(2):750–4.

69. Gupta R, Subramani M, Khaja MN, et al. Analysis of mutations within the 5 untranslated region, interferon sensitivity region, and PePHD region as a function of response to interferon therapy in hepatitis C virus-infected patients in India. *J Clin Microbiol*. 2006;44(3):709–15.

70. Ukai K, Ishigami M, Yoshioka K, et al. Mutations in carboxy-terminal part of E2 including PKR/eIF2alpha phosphorylation homology domain and interferon sensitivity determining region of nonstructural 5A of hepatitis C virus 1b: their correlation with response to interferon monotherapy and viral load. *World J Gastroenterol*. 2006;12(23):3722–8.

71. Xu Z, Fan X, Xu Y, Di Bisceglie AM. Comparative analysis of nearly full-length hepatitis C virus quasispecies from patients experiencing viral breakthrough during antiviral therapy: clustered mutations in three functional genes, E2, NS2, and NS5a. *J Virol*. 2008;82(19):9417–24.

72. Jiménez-Hernández N, Sentandreu V, Castro JA, et al. Effect of antiviral treatment and host susceptibility on positive selection in hepatitis C virus (HCV). *Virus Res*. 2008;131(2):224–32.

73. Cuevas JM, Gonzalez M, Torres-Puente M, et al. The role of positive selection in hepatitis C virus. *Infect Genet Evol*. 2009;9(5):860–6.

74. Brown RJ, Juttla VS, Tarr AW, et al. Evolutionary dynamics of hepatitis C virus envelope genes during chronic infection. *J Genl Virol*. 2005;86(7):1931–42.

75. Taylor DR, Shi ST, Romano PR, Barber GN, Lai MM. Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science*. 1999; 285(5424):107–10.

76. Donlin MJ, Cannon NA, Aurora R, Virahep-C Study Group, et al. Contribution of genome-wide HCV genetic differences to outcome of interferon-based therapy in Caucasian American and African American patients. *PLoS One*. 2010;5(2):e9032.

77. Cuevas JM, Torres-Puente M, Jiménez-Hernández N, et al. Combined therapy of interferon plus ribavirin promotes multiple adaptive solutions in hepatitis C virus. *J Med Virol*. 2009;81(4):650–6.