Infection, Genetics and Evolution 20 (2013) 418–421

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Viral evolution explains the associations among hepatitis C virus genotype, clinical outcomes, and human genetic variation

Rebecca Rose *, Peter V. Markov, Tommy T. Lam, Oliver G. Pybus

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom

ARTICLE INFO

Article history:
Received 7 August 2013
Received in revised form 23 September 2013
Accepted 26 September 2013
Available online 16 October 2013

Keywords:
Co-evolution
Hepatitis C virus
Phylogeny
SNPs
IFNL3

ABSTRACT

Specific human polymorphisms, most commonly found in Central Africa, can predict the success of drug treatment against the hepatitis C virus (HCV), a significant and globally-distributed human pathogen. However, this association is only found for a subset of HCV genotypes. Despite receiving considerable attention in the medical and virological literature, no evolutionary explanation for this curious pattern has been put forward. Here we suggest that the 'drug treatment resistance' phenotype exhibited today by some HCV genotypes evolved hundreds to thousands of years ago in response to human genetic variation local to Central Africa: an adaptation that has since accrued a new function in the era of anti-viral drug treatment. This could represent one of the oldest known examples of viral exaptation at the population level.

1. Introduction

The hepatitis C virus (HCV) infects hundreds of millions of people worldwide and can lead to severe liver disease and cancer. HCV is classified into seven genetically distinct genotypes (1–7) and has infected humans for at least 1000 years, although very likely much longer (Pybus et al., 2009; Simmonds, 2004) . The rapid transmission and global spread of HCV during the 20th century was preceded by the endemic circulation of HCV genotypes in distinct geographic regions: genotypes 1 and 4 in Central and North Africa, 2 in West Africa, 5 in South Africa, 3 in South Asia, and 6 in East Asia (Pybus et al., 2009; Simmonds, 2004; Markov et al., 2009; Pybus et al., 2007). It is currently unclear whether all HCV genotypes originated from a single emergence in humans with subsequent diversification, or if two or more HCV genotypes arose from separate cross-species transmissions from a currently unidentified source population (Pybus and Gray, 2013). The recent discovery of many highly-diverse viruses related to HCV in a small number of bat and rodent species (Quan et al., 2013; Kapoor et al., 2013; Drexler et al., 2013) suggests that future investigations may uncover one or more potential precursors of HCV in wild animals.

People infected with HCV can be treated with anti-viral drugs. Standard treatment currently comprises a combination of two drugs, interferon alpha and ribavirin. However the ability of this treatment to cure HCV infection varies considerably among HCV genotypes. HCV genotypes 1 and 4 are harder to treat than genotypes 2 and 3: less than 50% of patients infected with genotypes 1 or 4 will attain a successful outcome, typically defined as no virus detected in the patient's blood 24 weeks after cessation of therapy ("sustained virological response" or SVR24). On the other hand, upto 80% patients infected with genotypes 2 and 3 attain SVR24, usually after a drug regimen of 24 weeks instead of the 48 weeks recommended for genotypes 1 and 4 (e.g., Manns et al., 2001; Kamal and Nasser, 2008). The chance of treatment success is also affected by ethnicity, with African–Americans less likely than European–Americans to attain SVR24 (Muir et al., 2004). In 2009 the influence of ethnicity was partially clarified by the discovery of single nucleotide polymorphisms (SNPs) close to the human *IFNL3* gene on chromosome 19 that can predict the success of drug treatment (Ge et al., 2009; Tanaka et al., 2009) against HCV genotype 1 (the genotype that causes the majority of HCV infections in the developed world). These SNPs are also associated with an increased chance of resolving genotype 1 infection spontaneously, i.e., without requiring treatment (Muir et al., 2004). This finding is notable for the clear clinical relevance to an infectious disease resulting from genome-wide association studies.

The frequency of the *IFNL3*-linked variant that is most strongly associated with poor treatment outcome (*T* at position rs12979860) varies among populations and is highest in Central Africa and lowest in East Asia (Thomas et al., 2009) (Fig. 1a). The association between *IFNL3*-linked SNPs and HCV treatment success has been confirmed by numerous studies for genotype 1, and a

* Corresponding author. Address: Tinbergen Building, South Parks Road, Oxford OX31PS, United Kingdom. Tel.: +44 (0) 1865 271212.
E-mail address: rebecca.rose@zoo.ox.ac.uk (R. Rose).

1567-1348/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.meegid.2013.09.029

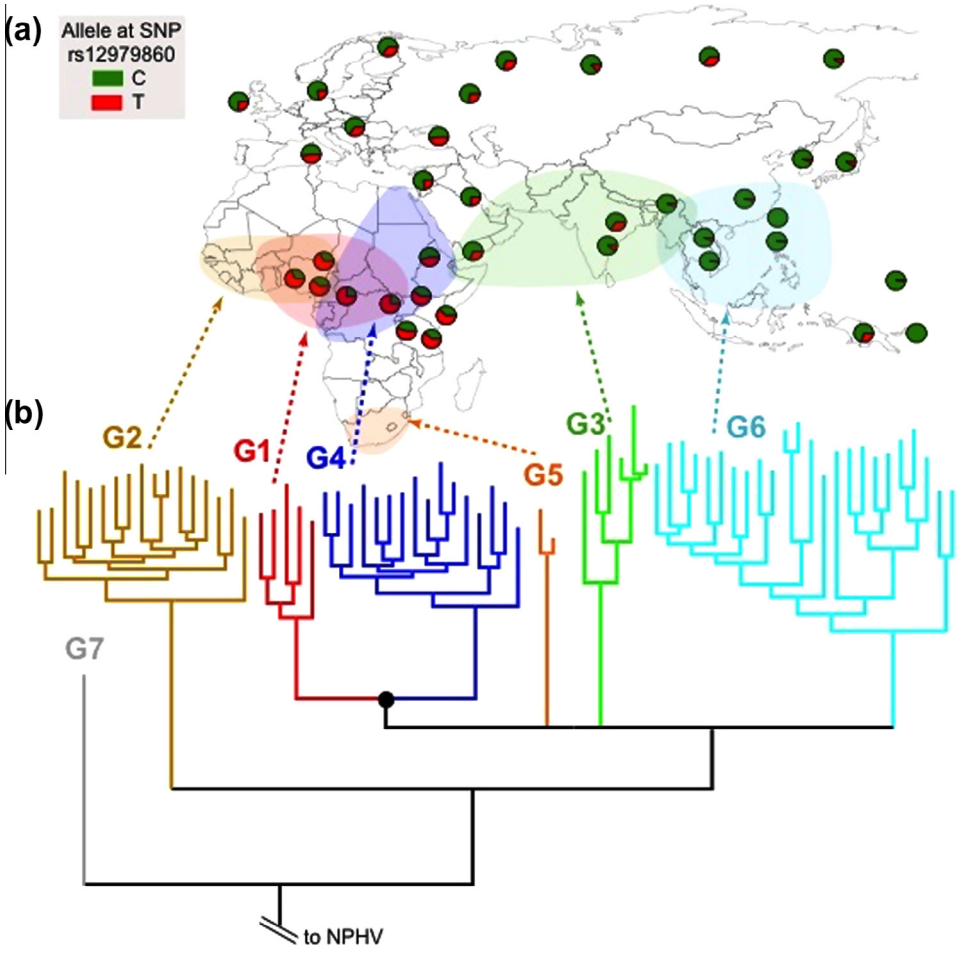

Fig. 1. Linking the spatial distribution of endemic HCV to viral and human genetic variation. (a) Map of Eurasia showing how allele frequencies at locus rs12979860 vary among locations (data from Thomas et al. (2009), population locations from http://alfred.med.yale.edus). Shaded areas show estimated regions of endemic HCV origin and persistence for each genotype (data adapted from Pybus et al. (2007); red = *G1*; brown = *G2*; green = *G3*; dark blue = *G4*; orange = *G5*; cyan = *G6*). These regions ignore endemic lineages that were trans-located to others areas during the colonial era (see Simmonds, 2004; Pybus et al., 2007 for details). The endemic origin of G7 is unknown. (b) Molecular phylogeny of HCV estimated from whole-genome nucleotide sequences, rooted with non-primate hepacivirus (NPHV) genomes. Genotypes are colored as above and arrows link each genotype to its estimated region of endemic origin. Among-genotype branches with weak bootstrap support (<70%) are shown as unresolved polytomies. The filled circle denotes the common ancestor of *G1* and *G4*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

recent meta-analysis reported the same association for genotype 4 (Chen et al., 2012). However, evidence that these human genetic markers predict treatment response is poor to non-existent for infections with genotypes 2 and 3 (Chen et al., 2012). Furthermore, limited data to date also suggest poor association for infections with HCV genotypes 5 (Antaki et al., 2012) or 6 (Seto et al., 2011). Although studies that include non-genotype 1 infections often suffer from low statistical power due to the small number of patients that fail treatment, lack of association for genotypes 2 and 3 is clear when multiple studies are combined in meta-analyses (Ge et al., 2009).

It therefore appears that the success of drug treatment against HCV infection depends on an interaction between host and viral genetics for some viral genotypes (1 and 4) but not others (2 and 3). This observation leads to important evolutionary questions that have not been addressed in the virological or clinical literature; specifically, how did the association between HCV treatment success and human SNPs evolve, and why should it vary among HCV genotypes?

The phylogeny of HCV provides an intriguing answer. Fig. 1b shows a maximum likelihood phylogeny estimated from representative full genome amino acid sequences of all HCV genotypes (see Figure legend for details). HCV genotypes 1 and 4 uniquely cluster together with significant bootstrap support, indicating these strains share a more recent common ancestor than any other pair of genotypes (Fig. 1b, Salemi and Vandamme, 2002). Crucially, genotypes 1 and 4 also share overlapping geographic distributions of endemic origin in Central Africa, precisely where the genetic diversity of SNPs linked to *IFNL3* is highest (Thomas et al., 2009). The extraordinary spatial coincidence of human and viral genetic variation associated with SVR24 can be parsimoniously explained by positing that, during its evolutionary past, HCV in Central Africa evolved in the context of a different immune environment to HCV elsewhere. In this geographic region the majority of individuals carry *IFNL3*-linked SNPs that may cause interferon-stimulated anti-viral genes to be differently expressed upon viral infection (explained in more detail below). Thus this host environment could have imposed a selective pressure on the common ancestor of HCV genotypes 1 and 4, favoring the past acquisition of mutations that increased viral fitness in the presence of this altered immune environment. Although these viral adaptations likely evolved centuries ago in response to human genetic variation in Central Africa, they have since acquired a new function following the development and widespread use of interferon-based anti-viral

drug treatment. We suggest this represents an unusual example of a viral *exaptation* (Gould and Vrba, 1982). By contrast, endemic HCV lineages of other genotypes originated in regions where *IFNL3* diversity was low and thus had no need to evolve this phenotype.

The evolutionary hypothesis outlined above does not depend on the specific mechanism that resulted in the imposition of a selective pressure on the common ancestor of HCV genotypes 1 and 4 by human genetic variability near the *IFNL3* locus. However, recently published work points towards a possible molecular cause. Prokunina-Olsson et al. (2013) showed that a novel protein is produced by individuals with a frame-shift variant (*ΔG*) at locus ss469415590 (closely linked to the *T* allele at rs12979860) (Prokunina-Olsson et al., 2013). The ΔG variant is even more strongly associated with treatment failure than the SNPs reported previously (Tanaka et al., 2009). The novel protein appears to affect the induction of host anti-viral genes (interferon-stimulated genes; ISGs) (Prokunina-Olsson et al., 2013). It is possible that individuals with the *ΔG* variant have increased or altered interferon responses to viral infection which impair the body's ability to clear the virus during acute infection, thus rendering interferon-based drug treatment less effective. This is compatible with observed associations among the unfavorable *T* allele at rs12979860, poor treatment outcome and pre-activation and/or transient expression of treatment-induced genes (Younossi et al., 2012). This hypothesis is also consistent with the observation that higher expression of ISGs before anti-viral drug therapy is, somewhat paradoxically, associated with lower rates of treatment success (Sarasin-Filipowicz et al., 2008). However, even among people with the beneficial SNPs, the chance of treatment success is still less for individuals infected with genotypes 1 and 4 than for those infected with genotypes 2 and 3 (Thompson et al., 2010; Mangia et al., 2010).

Viral mutations that contribute to the 'drug treatment resistance' phenotype are most likely to be found among the set of amino acid changes that are shared by genotypes 1 and 4; i.e., they evolved along the ancestral lineage basal to the two genotypes. Analysis using phylogenetic methods of ancestral sequence reconstruction reveals that there are 82 amino acids that change along the phylogenetic branch ancestral to genotypes 1 and 4 (Table 1). Although functionally relevant mutations are likely among this set, we caution that many or most will not be associated with SVR24, nor have arisen through positive selection. This set does however provide a starting point for experimental virologists who wish to investigate the molecular determinants of variation in HCV clearance and treatment response. Multiple challenges face such studies, including potentially complex epistatic interactions among viral mutations and significant differences between *in vivo* and *in vitro* viral replication. *IFNL3*-linked SNPs also appear to correlate with treatment outcome for the new classes of anti-HCV drugs currently being tested, hence identifying the evolutionary and genetic causes of this association remains important (Ghany et al., 2011).

One aspect of our hypothesis that is still highly uncertain relates to HCV genotype 2. Unlike genotypes 3 and 6, which appear to be endemic to Asia, genotype 2 is thought to have originated in Western Africa (Jeannel et al., 1998). Few data are available on the frequency of SNPs linked to *IFNL3* in West Africa but it seems likely that SNP frequencies there will resemble those in Central Africa, given the shared histories of the respective populations in the two regions. If this is the case, why does HCV genotype 2 not also exhibit the same 'drug treatment resistance' exaptation as genotypes 1 and 4? The simplest explanation is that genotype 2 is genetically very divergent from genotypes 1 and 4, hence the mutations in the latter that were selected in the presence of *IFNL3*-linked SNPs were specific to the particular genetic background of the virus. Genotype 2 viruses may have accumulated different adaptive mutations in response to the historical selection

Table 1
The position and nature of amino acid changes that evolved on the branch ancestral to the common ancestor of *G*1 and *G*4.

	Region	H77[a]	Ancestral[b]	Derived
1	CORE	72	T	E
2	E1	241	P	A
3	E1	242	V	L
4	E1	248	A/V	A
5	E1	301	V	T
6	E1	343	A/I/L/V	A/V
7	E1	345	I	V
8	E1	347	I	M
9	E1	357	A/F/G/I/L	A
10	E2	466	D	A
11	E2	501	S	E/T
12	E2	522	L/R	L
13	E2	546	Q	G/Q
14	E2	667	F/H/L	L
15	E2	673	A	Q
16	E2	674	I	V
17	E2	682	M/L	L
18	E2	731	I	V
19	P7	766	L	V
20	P7	767	W	S
21	P7	777	H	Y
22	P7	787	T	A
23	P7	791	F	Y
24	NS2	814	E	V
25	NS2	820	V	G
26	NS2	825	I	F
27	NS2	829	F	L
28	NS2	832	T	S
29	NS2	834	G	R
30	NS2	837	K	R
31	NS2	904	D	A
32	NS2	919	L	I
33	NS2	928	L	M
34	NS2	997	L	I
35	NS2	1010	V	I
36	NS2	1017	D	G
37	NS2	1019	R	T
38	NS3	1136	N	H
39	NS3	1142	A	V
40	NS3	4450	A	G
41	NS3	1169	V	L
42	NS3	1176	V	A
43	NS3	1193	L	V
44	NS3	1200	T	S
45	NS3	1226	G	A
46	NS3	1227	Y	H
47	NS3	1272	Y	H
48	NS3	1341	V	A
49	NS3	1456	V	T
50	NS3	1457	A	S
51	NS3	1516	V	S
52	NS3	1598	K	Q
53	NS3	1619	T	H
54	NS3	1635	I	V
55	NS3	1656	I	V
56	NS4a	1695	V	I
57	NS4b	1722	Q	L
58	NS4b	1724	I	L
59	NS4b	1732	V	A
60	NS4b	1746	I	A
61	NS4b	1747	K	T
62	NS4b	1769	V	I
63	NS4b	1787	V	I
64	NS4b	1804	H	Q
65	NS4b	1822	P	T
66	NS4b	1842	I	V
67	NS4b	1941	S	A
68	NS4b	1943	K	R
69	NS5a	2034	K	E
70	NS5a	2099	Y	F
71	NS5a	2130	I	L
72	NS5a	2187	A	R
73	NS5b	2439	E	A

Table 1 (*continued*)

	Region	H77[a]	Ancestral[b]	Derived
74	NS5b	2493	V	A
75	NS5b	2568	D	E
76	NS5b	2570	A	E
77	NS5b	2599	T	I
78	NS5b	2609	P	A
79	NS5b	2709	M	F
80	NS5b	2874	V	I
81	NS5b	2963	G	A
82	NS5b	3000	T	S

[a] Numbering relative to the H77 reference genome.
[b] Ancestral sequences were reconstructed using maximum likelihood joint method in HYPHY (<http://hyphy.org>).

pressure imposed by human immunological variation. However these changes may not have conferred the 'drug treatment-resistance' phenotype exhibited today by genotypes 1 and 4.

We posit here a simple evolutionary explanation for the otherwise puzzling variation among HCV strains in the association between patient ethnicity and clinical outcome. It highlights the importance of considering an evolutionary perspective when interpreting clinical phenomena, as it can help to synthesize and explain seemingly unconnected observations and provide a clearer path to functional experiments. Such scenarios are likely to be encountered with increasing frequency in the era of 'personalized medicine'. Understanding the evolution and function of the HCV 'drug treatment-resistance' phenotype in its original, historical context may also help to determine whether the world-wide spread and dominance of HCV genotype 1 was due in part to a selective advantage, or was merely a consequence of historical and epidemiological founder effects. If our hypothesis is correct, the evolution of HCV in response to *IFNL3*-linked variability represents an important example of viral adaptation to the genetic composition of its host population; a similar argument has been made for the adaptation of HIV and HCV in response to human variation at HLA loci, albeit over much more recent timescales (Gaudieri et al., 2006; Kawashima et al., 2009). The SNPs linked to *IFNL3* discussed here are possibly thousands of years old and could represent one of the oldest viral adaptations to human genetic variation yet identified.

Acknowledgments

The authors thank Charles Rice, Paul Klenerman, and Ludmila Prokunina for helpful discussion. RR is supported by the United Kingdom Medical Research Council.

References

Antaki, N., Bibert, S., Kebbewar, K., Asaad, F., Baroudi, O., et al., 2012. IL28B polymorphisms do not predict response to therapy in chronic hepatitis C with HCV genotype 5. Gut 61, 1640–1641.
Chen, Y., Xu, H.X., Wang, L.J., Liu, X.X., Mahato, R.I., et al., 2012. Meta-analysis: IL28B polymorphisms predict sustained viral response in HCV patients treated with pegylated interferon-α and ribavirin. Aliment. Pharmacol. Ther. 36, 91–103.
Drexler, J.F., Corman, V.M., Müller, M.A., Lukashev, A.N., Gmyl, A., et al., 2013. Evidence for novel hepaciviruses in rodents. PLoS Pathog. 9, e1003438.
Gaudieri, S., Rauch, A., Park, L.P., Freitas, E., Herrmann, S., et al., 2006. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. J. Virol. 80, 11094–11104.
Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., et al., 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461, 399–401.
Ghany, M.G., Nelson, D.R., Strader, D.B., Thomas, D.L., Seeff, L.B., et al., 2011. An update on treatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases. Hepatology 54, 1433–1444.
Gould, S., Vrba, E., 1982. Exaptation – a missing term in the science of form. Paleobiology, 4–15.
Jeannel, D., Fretz, C., Traore, Y., Kohdjo, N., Bigot, A., et al., 1998. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. J. Med. Virol. 55, 92–97.
Kamal, S.M., Nasser, I.A., 2008. Hepatitis C genotype 4: what we know and what we don't yet know. Hepatology 47, 1371–1383.
Kapoor, A., Simmonds, P., Scheel, T.K., Hjelle, B., Cullen, J.M., et al., 2013. Identification of rodent homologs of hepatitis C virus and pegiviruses. mBio. 4, e00216–00213.
Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., et al., 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458, 641–645.
Mangia, A., Thompson, A.J., Santoro, R., Piazzolla, V., Tillmann, H.L., et al., 2010. An IL28B polymorphism determines treatment response of hepatitis C virus genotype 2 or 3 patients who do not achieve a rapid virologic response. Gastroenterology 139, 821–827 (827.e821).
Manns, M.P., McHutchison, J.G., Gordon, S.C., Rustgi, V.K., Shiffman, M., et al., 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet 358, 958–965.
Markov, P., Pepin, J., Frost, E., Deslandes, S., Labbé, A., et al., 2009. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. J. Gen. Virol. 90, 2086–2096.
Muir, A., Bornstein, J., Killenberg, P., 2004. Peginterferon alfa-2b and ribavirin for the treatment of chronic hepatitis C in blacks and non-Hispanic whites. N. Engl. J. Med. 350, 2265–2271.
Prokunina-Olsson, L., Muchmore, B., Tang, W., Pfeiffer, R.M., Park, H., et al., 2013. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat. Genet. 45, 164–171.
Pybus, O.G., Gray, R.R., 2013. The virus whose family expanded. Nature 498, 310.
Pybus, O.G., Markov, P.V., Wu, A., Tatem, A.J., 2007. Investigating the endemic transmission of the hepatitis C virus. Int. J. Parasitol. 37, 839–849.
Pybus, O.G., Barnes, E., Taggart, R., Lemey, P., Markov, P., et al., 2009. Genetic history of hepatitis C virus in East Asia. J. Virol. 83, 1071–1082.
Quan, P.L., Firth, C., Conte, J.M., Williams, S.H., Zambrana-Torrelio, C.M., et al., 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proc. Natl. Acad. Sci. U.S.A. 110, 8194–8199.
Salemi, M., Vandamme, A., 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. J. Mol. Evol. 54, 62–70.
Sarasin-Filipowicz, M., Oakeley, E.J., Duong, F.H., Christen, V., Terracciano, L., et al., 2008. Interferon signaling and treatment outcome in chronic hepatitis C. Proc. Natl. Acad. Sci. U.S.A. 105, 7034–7039.
Seto, W.K., Tanaka, Y., Liu, K., Lai, C.L., Yuen, M.F., 2011. The Effects of IL-28B and ITPA polymorphisms on treatment of hepatitis C virus genotype 6. Am. J. Gastroenterol. 106, 1007–1008.
Simmonds, P., 2004. Genetic diversity and evolution of hepatitis C virus – 15 years on. J. Gen. Virol. 85, 3173–3188.
Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., et al., 2009. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat. Genet. 41, 1105–1109.
Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., et al., 2009. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 461, 798–801.
Thompson, A.J., Muir, A.J., Sulkowski, M.S., Ge, D., Fellay, J., et al., 2010. Interleukin-28B polymorphism improves viral kinetics and is the strongest pretreatment predictor of sustained virologic response in genotype 1 hepatitis C virus. Gastroenterology 139 (120–129), e118.
Younossi, Z.M., Birerdinc, A., Estep, M., Stepanova, M., Afendy, A., et al., 2012. The impact of IL28B genotype on the gene expression profile of patients with chronic hepatitis C treated with pegylated interferon alpha and ribavirin. J. Transl. Med. 10, 25.