

Liver gene expression signature to predict response to pegylated interferon plus ribavirin combination therapy in patients with chronic hepatitis C

T Asselah,^{1,2} I Bieche,^{2,3} S Narguet,² A Sabbagh,^{2,3} I Laurendeau,² M-P Ripault,¹ N Boyer,¹ M Martinot-Peignoux,¹ D Valla,¹ M Vidaud,^{2,3} P Marcellin¹

¹ Service d'Hépatologie, Pôle des Maladies de l'Appareil Digestif and INSERM U773, CRB3, University of Paris VII, AP-HP Beaujon Hospital, Clichy, France; ² INSERM UMR745, University of Paris Descartes, Paris, France; ³ Service de Biochimie et Génétique moléculaire, Beaujon Hospital, Clichy, France

Correspondence to: Dr T Asselah, Service d'Hépatologie and INSERM CRB3, Beaujon Hospital, 92 110 Clichy, France; tarik.asselah@bjn.aphp.fr

TA and IB contributed equally.

Revised 3 August 2007
Accepted 28 August 2007
Published Online First
28 September 2007

ABSTRACT

Background and Aims: The gold standard treatment of chronic hepatitis C (CHC) is combined pegylated interferon and ribavirin. Considering side effects and treatment cost, prediction of treatment response before therapy is important. The aim of this study was to identify a liver gene signature to predict sustained virological response in patients with CHC.

Methods: Group A (training set) comprised 40 patients with CHC including 14 non-responders (NRs) and 26 sustained virological responders (SVRs). Group B (validation set) comprised 29 patients including 9 NRs and 20 SVRs. Eleven responder-relapsers were also included. A total of 58 genes associated with liver gene expression dysregulation during CHC were selected from the literature. Real-time quantitative RT-PCR assays were used to analyse the mRNA expression of these 58 selected genes in liver biopsy specimens taken from the patients before treatment.

Results: From the Group A data, three genes whose expression was significantly increased in NRs compared with SVRs were identified: *IFI-6-16/G1P3*, *IFI27* and *ISG15/G1P2*. These three genes also showed significant differences in their expression profiles between NRs and SVRs in the independent sample (Group B). Supervised class prediction analysis identified a two-gene (*IFI27* and *CXCL9*) signature, which accurately predicted treatment response in 79.3% (23/29) of patients from the validation set (Group B), with a predictive accuracy of 100% (9/9) and of 70% (14/20) in NRs and SVRs, respectively. The expression profiles of responder-relapsers did not differ significantly from those of NRs and SVRs, and 73% (8/11) of them were predicted as SVRs with the two-gene classifier.

Conclusion: NRs and SVRs have different liver gene expression profiles before treatment. The most notable changes occurred mainly in interferon-stimulated genes. Treatment response could be predicted with a two-gene signature (*IFI27* and *CXCL9*).

Chronic hepatitis C (CHC) is among the leading causes of chronic liver disease worldwide, with a prevalence of approximately 170 million people. The severity of disease varies from asymptomatic chronic infection to cirrhosis and hepatocellular carcinoma.¹⁻³ The main treatment goal in CHC is the prevention of cirrhosis and hepatocellular carcinoma by eradicating the virus. Recently, advances have been made in treatment with the combination of pegylated interferon (PEG-IFN) and ribavirin.⁴⁻⁸ Despite this progress, treatment failure still occurs in about half of the patients.

Furthermore, therapy results in several side effects. Premature withdrawal due to adverse events was required in 10–14% of participants in registration trials.⁵⁻⁸ Considering side effects and treatment cost, prediction of virological non-response before therapy with more reliable markers is mandatory.^{1,2}

The recent development of effective tools for the large-scale analysis of gene expression has provided new insights into the involvement of gene networks and regular pathways in various disease processes.⁹ These methods include microarrays to analyse the expression of thousand of genes at a time, and real-time RT-PCR assays for more accurate and quantitative expression analysis of smaller numbers of candidate genes.¹⁰ Using large-scale real-time quantitative RT-PCR, we have shown that several altered molecular pathways are involved in CHC compared with normal liver¹¹ and in the progression of fibrosis.¹⁰ Real-time RT-PCR is also more appropriate than microarrays for analysing weakly expressed genes such as genes encoding cytokines.

We hypothesised that non-responders (NRs) and sustained virological responders (SVRs) might have different liver gene expression patterns prior to treatment. The aim of our study was to predict treatment outcome using real-time PCR in patients with CHC before treatment. We selected 58 genes from the literature involved in various cellular and molecular mechanisms associated with liver gene expression during CHC.¹⁰⁻¹⁷ We then built prediction models using supervised class prediction algorithms for gene signature discovery. A two-gene subset (*IFI27* and *CXCL9*) accurately predicted treatment response (79.3%) in the validation set. This signature might help predict response to PEG-IFN plus ribavirin therapy in CHC.

PATIENTS AND METHODS

Chronic hepatitis C patients

Percutaneous liver biopsy specimens were selected from a cohort of adult patients with CHC followed at Beaujon Hospital (Clichy, France). Pretreatment liver biopsies from patients who were subsequently treated and for whom treatment responses were known were used for the study. In each case, both immediately frozen liver tissue (stored at -80°C) and fixed paraffin-embedded tissue (for histology) were available. Participants gave their informed consent. The study conformed to the ethical guidelines of the 1975 Declaration of Helsinki.



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://gut.bmj.com/info/unlocked.dtl>

Patients were included in this study if they met the following criteria.

- ▶ An established diagnosis of CHC with detectable anti-hepatitis C virus (HCV) antibodies, and detectable serum HCV RNA with RT-PCR (HCV Amplicor 2.0; Roche Diagnostics, Mannheim, Germany).
- ▶ Absence of other causes of chronic liver disease (undetectable hepatitis B surface antigen (HBs-Ag), no excessive alcohol consumption (<30 g/day), haemochromatosis, autoimmune hepatitis, Wilson disease, α -1 antitrypsin deficiency, primary sclerosing cholangitis or primary biliary cirrhosis).
- ▶ Standard treatment regimen: all were naïve patients and received the same complete treatment of PEG-IFN α -2b (Viraferon-peg, 1.5 μ g/kg/week; Schering Corp.) and ribavirin (dose adjusted to body weight: <65 kg, 800 mg/day; 65–85 kg, 1000 mg/day; >85 kg, 1200 mg/day). Duration of treatment was 24 weeks for genotypes 2 or 3, and 48 weeks for genotypes 1 or 4. NRs were treated for a full course of therapy.
- ▶ Adequate follow-up: detection of serum HCV RNA by RT-PCR was performed at week 24, at the end, and 6 months after the end of treatment.

Patients were selected based on their treatment response. Sustained virological response was defined as undetectable HCV RNA 6 months after completion of treatment.^{1,2} Non-response was defined as detectable serum HCV RNA at the end of treatment.

We first built a gene signature on a training set with 40 patients with CHC including 14 NRs and 26 SVRs (Group A). We then validated the gene signature on an independent validation set of 29 patients including 9 NRs and 20 SVRs (Group B), to confirm that it would generalise to new data. Baseline characteristics of groups A and B are shown in tables 1 and 2.

A group of 11 responder–relapsers (RRs) was also included in the study. Relapse was defined as the reappearance of detectable serum HCV RNA after treatment cessation in patients with no detectable serum HCV RNA at the end of treatment. They have been included with the same criteria as previous CHC patients.

Table 1 Characteristics of patients from the training set (Group A)

Variable	Patients	NRs	SVRs
No.	40	14	26
Gender (male, n (%))	26 (65)	9 (64)	17 (65)
Age (years)	47.4 (8.6)	46.5 (7.1)	47.9 (9.4)
Source of infection (n (%))			
Blood transfusion	11 (27.5)	3 (21.4)	8 (30.8)
Intravenous drug use	13 (32.5)	5 (35.7)	8 (30.8)
Unknown	16 (40.0)	6 (42.9)	10 (38.5)
ALT (IU/l, median)	120 (57)	132 (86)	114 (43)
HCV genotype (n)			
1	24	13	11
2	3	0	3
3	8	0	8
4	4	1	3
5	1	0	1
Stage (n)			
F1	13	1	12
F2	15	7	8
F3	8	4	4
F4	4	2	2

ALT, alanine aminotransferase; HCV, hepatitis C virus; NRs, non-responders; SVRs, sustained virological responders.

In particular, they were all naïve patients and received the same complete treatment of PEG-IFN α -2b (Viraferon-peg, 1.5 μ g/kg/week; Schering Corp.) and ribavirin (dose adjusted to body weight: <65 kg, 800 mg/day; 65–85 kg, 1000 mg/day; >85 kg, 1200 mg/day). Baseline characteristics of RR patients are shown in table 3.

Histologically normal controls

Percutaneous liver biopsy specimens were taken from eight adults with mildly elevated alanine aminotransferase (ALT) activity with no cause of liver disease (medication, alcohol, chronic viral hepatitis, autoimmune processes and metabolic disease). All these adults gave their informed consent for the study. All these eight liver tissue specimens were histologically normal (ie, absence of inflammation, fibrosis and pathological pattern).

Real-time RT-PCR

In previous studies using the same technological approach, we have shown that several altered molecular pathways are involved in hepatitis C compared with the normal liver¹¹ and in the progression of liver fibrosis.¹⁰

Theoretical basis

The theoretical and practical aspects of real-time quantitative RT-PCR using the ABI Prism 7900 Sequence Detection System (Perkin-Elmer Applied Biosystems, Foster City, CA) have been described in detail elsewhere.¹⁰

The precise amount of total RNA added to each reaction mix (based on optical density) and its quality (ie, lack of extensive degradation) are both difficult to assess. We therefore also quantified transcripts of two endogenous RNA control genes involved in two cellular metabolic pathways, namely *TBP* (GenBank accession no. NM_003194), which encodes the TATA box-binding protein (a component of the DNA-binding protein complex TFIID), and *RPLP0* (also known as 36B4; NM_001002), which encodes human acidic ribosomal phosphoprotein P0. Each sample was normalised on the basis of its *TBP* (or *RPLP0*) content. Results, expressed as n-fold differences in target gene expression relative to the *TBP* (or *RPLP0*) gene, and termed “Ntarget”, were determined as $N_{target} = 2^{\Delta Ct_{sample}}$, where the ΔCt value of the sample was determined by subtracting the average Ct value of the target gene from the average Ct value of the *TBP* (or *RPLP0*) gene.

The Ntarget values of the samples were subsequently normalised such that the median of the normal histological liver Ntarget values was 1.

Selection of genes

By studying the literature, we selected 58 genes involved in various cellular and molecular mechanisms associated with liver gene expression dysregulation during CHC in humans.^{10–17} We especially focused on genes whose expression was dysregulated during HCV infection compared with the normal histological liver (n = 37),¹¹ genes involved in the progression of fibrosis in CHC (n = 7),¹⁰ those recently associated with treatment outcome (n = 18)¹² and genes associated with the immune response to HCV.¹⁷ These genes mainly encode proteins involved in the immune response and interferon signalling pathways (table 4).

Primers and controls

Primers for *TBP*, *RPLP0* and the 58 target genes were chosen with the assistance of the Oligo 5.0 computer program

Hepatitis

Table 2 Characteristics of patients from the validation set (Group B)

Variable	Patients	NRs	SVRs
No.	29	9	20
Gender (male, n (%))	22 (76)	8 (89)	14 (70)
Age (years)	45.1 (9.1)	48.2 (10.6)	43.6 (8.3)
Source of infection (n (%))			
Blood transfusion	3 (10.3)	1 (11)	2 (10)
Intravenous drug use	9 (31.1)	3 (33)	6 (30)
Unknown	17 (58.6)	5 (66)	12 (60)
ALT (IU/l, median)	129 (70)	120 (49)	132 (79)
HCV genotype (n)			
1	17	8	9
2	2	0	2
3	6	0	6
4	4	1	3
5	0	0	0
Stage (n)			
F1	7	1	6
F2	13	4	9
F3	7	2	5
F4	2	2	0

ALT, alanine aminotransferase; HCV, hepatitis C virus; NRs, non-responders; SVRs, sustained virological responders.

(National Biosciences, Plymouth, MN). We conducted searches in the dbEST and nr databases to confirm the total gene specificity of the nucleotide sequences chosen as primers, and the absence of single nucleotide polymorphisms. In particular, the primer pairs were selected to be unique in relation to the sequences of closely related family member genes or of the corresponding retropseudogenes. To avoid amplification of contaminating genomic DNA, one of the two primers was placed at the junction between two exons, if possible. In general, amplicons were between 70 and 120 nucleotides long. Gel electrophoresis was used to verify the specificity of PCR amplicons.

For each primer pair, no-template control (NTC) and no-reverse transcriptase control (RT negative) assays were performed and produced negligible signals (usually >40 in Ct value), suggesting that primer dimer formation and genomic DNA contamination effects were negligible. The RNA extraction, cDNA synthesis and PCR conditions are described elsewhere.¹⁰

Significance testing, clustering and supervised class prediction

Genes differentially expressed between SVRs and NRs were identified using the signal-to-noise (SNR) and Student *t* test statistics. *p* Values were estimated using 10 000 permutations of the sample status (SVR/NR), and genes with an adjusted *p* value <0.05 were considered to be significantly differentially expressed. *p* Values were corrected using the Benjamini and Hochberg false discovery rate (FDR) method¹⁸ to control false-positive results from multiple testing.

Unsupervised hierarchical clustering of expression data was performed using average linkage clustering with cosine distance as the similarity metric.

Supervised class prediction analysis was performed using *k* nearest neighbour (KNN) and weighted voting (WV) prediction algorithms to achieve optimal predictive accuracy for sample classification with a minimal set of predictors. Briefly, a KNN classifier¹⁹ is based on a distance function between pairs of observations and identifies the class of a testing sample by determining its *k* nearest neighbours in the training data set and

Table 3 Characteristics of the responder–relapser patients

Variable	Patients
No.	11
Gender (male, n (%))	8 (73)
Age (years)	50.7 (6.2)
Source of infection (n (%))	
Blood transfusion	3 (27)
Intravenous drug use	3 (27)
Unknown	5 (45)
ALT (IU/l, median)	123 (54)
HCV genotype (n)	
1	3
2	1
3	5
4	2
5	0
Stage (n)	
F1	3
F2	4
F3	4
F4	0

ALT, alanine aminotransferase; HCV, hepatitis C virus.

by choosing the most common class of those *k* neighbours. Our KNN classifier used a distance-weighted voting scheme with cosine distance. The number of *k* neighbours was chosen by cross-validation (CV) on the training set, and the *k* yielding the smallest CV error rate was retained for use on the test set. The WV classifier makes a weighted linear combination of relevant marker genes obtained in the training set to provide a classification scheme for new samples. The selection of classifier input features was performed by computing the SNR statistic.

Prediction models were built using gene expression data from Group A samples, which served as a training set. Models containing increasing numbers of genes were evaluated by leave-one-out cross-validation (LOOCV) to identify the smallest predictive model yielding the most accurate class assignments. In LOOCV, one sample from the training set is excluded, and the rest of the training samples are used to build the classifier. Then the classifier is used to predict the class of the one that was left out, and this is repeated for each sample in the training set. The LOOCV estimate of classification accuracy is the overall number of correctly classified samples, divided by the number of samples in the training set. Thereafter, a validation analysis of the predictive classification model was performed using an independent test set (Group B), and the overall accuracy of class assignment for samples in the test set was reported.

All analyses were performed using GenePattern version 2.0 (<http://www.broad.mit.edu/cancer/software/genepattern>).²⁰

RESULTS

Identifying significantly differentially expressed genes between SVRs and NRs

We first examined whether there were significant differences in individual gene expression between SVR and NR liver tissue specimens. Table 5 shows the results of the Student *t* tests when applied to Group A. Similar results were obtained with the SNR statistic, pointing to the same differentially expressed genes, although not ranked exactly in the same order. Genes correlated with SVR/NR status were identified by sorting all of the genes according to the *t* test statistic. Ten genes displayed a non-adjusted *p* value <0.05. Although this does not strictly

Table 4 List of the 58 genes studied

Gene	Protein
Interferon signalling pathway:	
<i>IRF7</i>	Interferon regulatory factor 7
<i>ISGF3G</i>	Interferon-stimulated transcription factor 3, gamma 48 kDa
<i>STAT1</i>	Signal transducer and activator of transcription 1
<i>STAT2</i>	Signal transducer and activator of transcription 2
Interferon-inducible proteins:	
<i>ISG15</i>	Interferon alpha-inducible protein 2 (clone IFI -15K)
<i>IFI-6-16</i>	Interferon alpha-inducible protein 3 (clone IFI-6-16)
<i>IFI27</i>	Interferon alpha-inducible protein 27
<i>IFI35</i>	Interferon alpha-inducible protein 35
<i>IFIT1</i>	Interferon-induced protein with tetratricopeptide repeats 1
<i>IFIT4</i>	Interferon-induced protein with tetratricopeptide repeats 4
<i>IFITM1</i>	Interferon-induced transmembrane protein 1
<i>LGALS3BP</i>	Lectin galactoside-binding soluble 3
<i>MX1</i>	Myxovirus (influenza virus) resistance 1
<i>OAS1</i>	2'-5'-Oligoadenylate synthetase 1
<i>OAS2</i>	2'-5'-Oligoadenylate synthetase 2
<i>OAS3</i>	2'-5'-Oligoadenylate synthetase 3
<i>PLSCR1</i>	Phospholipid scramblase 1
<i>PRKR</i>	Protein kinase IFN-inducible double-stranded RNA dpt
<i>PSMB9</i>	Proteasome (prosome, macropain) subunit beta type 9
<i>HERC5</i>	Hect domain and RLD 5
<i>RSAD2</i>	Radical S-adenosyl methionine domain-containing 2
Growth factor receptors:	
<i>CXCR3</i>	Chemokine (C-X-C motif) receptor 3
Growth factors; chemokines; interleukins:	
<i>CXCL6</i>	Chemokine (C-X-C motif) ligand 6 (GCP-2)
<i>CCL8</i>	Chemokine (C-C motif) ligand 8 (MCP-2)
<i>CXCL9</i>	Chemokine (C-X-C motif) ligand 9 (MIG)
<i>CXCL10</i>	Chemokine (C-X-C motif) ligand 10 (IP-10)
<i>CXCL11</i>	Chemokine (C-X-C motif) ligand 11 (IP-9)
<i>FGF7</i>	Fibroblast growth factor 7 (FGF)
<i>IFNG</i>	Interferon, gamma
<i>IL6</i>	Interleukin 6
<i>IL8</i>	Interleukin 8
<i>IL10</i>	Interleukin 10
<i>MDK</i>	Midkine (NEGF2)
<i>TGFB2</i>	Transforming growth factor beta 2
<i>TNF</i>	Tumour necrosis factor
<i>TSLP</i>	Thymic stromal lymphopoietin
Angiogenesis:	
<i>FIGF</i>	VEGF-D
Signal transduction:	
<i>DUSP1</i>	Dual specificity phosphatase 1 (MKP-1)
<i>EEF1G</i>	Eukaryotic translation elongation factor 1 g

Continued

Table 4 Continued

Gene	Protein
Cell cycle:	
<i>CDKN2A</i>	CDK inhibitor 2A (p16 protein)
<i>SFN</i>	Stratifin (protein 14-3-3 sigma)
Transcription factors:	
<i>ATF5</i>	Activating transcription factor 5
Cytoskeletal:	
<i>STMN2</i>	Stathmin-like 2 (SCG 10)
Iron metabolism:	
<i>HAMP</i>	Hepcidin antimicrobial peptide
Oxidative stress:	
<i>CA9</i>	Carbonic anhydrase IX
<i>CYP2E1</i>	Cytochrome P450 CYP2E1
Cell adhesion and cell junction:	
<i>ITGA2</i>	Integrin, alpha 2
Matrix proteases:	
<i>MMP9</i>	Matrix metalloproteinase 9
<i>ANPEP</i>	Alanyl aminopeptidase (CD13)
Various:	
<i>CD14</i>	CD14 antigen
<i>HLA-B</i>	Major histocompatibility complex, class I, B
<i>LGP1</i>	D11lgp1e-like
<i>STXBP5</i>	Syntaxin-binding protein 5 (Tomosyn)
<i>PIK3AP1</i>	Phosphoinositide-3-kinase adaptor protein 1
<i>RPLP2</i>	Ribosomal protein large P2
<i>RPS28</i>	Ribosomal protein S28
<i>LAP3</i>	Leucine aminopeptidase 3
<i>USP18</i>	Ubiquitin-specific protease 18

assess significance, these non-adjusted p values are useful to identify genes that should be studied further. From the Group A data, we identified three genes whose expression differed significantly between SVRs and NRs (FDR <0.05): *IFI-6-16* (previously named *G1P3*), *IFI27* and *ISG15* (previously named *G1P2*) (table 5). Note that *ISG15* and *IFI-6-16* were already identified as differentially expressed genes between SVRs and NRs (p<0.001) in an independent study.¹²

When the gene expression profiles of the 10 genes displaying a non-adjusted p value <0.05 in Group A were compared between NR and SVR patients from Group B, 6 out of 10 genes (*IFI27*, *ISG15*, *IFI-6-16*, *OAS2*, *HERC5* and *IFIT1*) were found to be differentially expressed in this independent data set (FDR <0.05) (table 6). Among these six genes, five (*IFI27*, *ISG15*, *IFI-6-16*, *OAS2* and *IFIT1*) were also identified as differentially expressed in a previous study.¹² The three genes shown to be significant (FDR <0.05) in Group A (*IFI-6-16*, *IFI27* and *ISG15*) also showed significant differences in their expression profiles between SVRs and NRs in Group B and displayed the highest level of significance.

Selection of the most useful gene subsets for gene expression-based classification

Exploratory data analysis using unsupervised hierarchical clustering was performed on both Group A and Group B using the whole set of genes (n=58). No distinctive clusters were identified between the SVR and NR groups by this analysis. This was expected since most of the genes considered were not correlated with NR/SVR status. Hence, the SVR and NR patients could not be differentiated by unsupervised data analysis.

Hepatitis

Prediction models were built using two supervised learning classifiers, KNN and WV, for gene signature discovery. Since good predictors do not necessarily have low type 1 errors when the null hypothesis of no association with the clinical outcome is tested for each of the investigated variables,²¹ the analysis was not restricted to the statistically significant discriminatory genes but was performed using the entire set of genes. The KNN algorithm identified a two-gene classifier (*IFI27* and *CXCL9*, using $k = 3$) that gave the highest overall accuracy (77.5%) in distinguishing between SVR and NR gene expression profiles as evaluated by LOOCV of Group A, used as a training set (fig 2). Increasing the size of the classifier set did not increase accuracy. This two-gene signature correctly classified 23 of 26 SVR patients (88.5%) and 8 of 14 NR patients (57.1%) (table 7). It was validated on an independent test set (Group B) consisting of 29 patients, 23 of whom were correctly classified (79.3%): treatment response was correctly predicted for all NR patients (100%) and 14 of 20 SVR patients (70%). With the WV classifier, the highest classification accuracy (80.0%) was obtained using a three-gene subset (*IFI27*, *CXCL9* and *IFI-6-16*) in Group A, and the results did not improve when additional genes were included (fig 2 and table 7). This predictive gene signature also showed good predictive accuracy in the independent test set (Group B), with 72.4% (21/29) of samples correctly assigned. Interestingly, among the genes selected as good predictors for gene expression-based classification of treatment outcomes, *CXCL9* was not shown to be differentially expressed between SVR and NR patients (FDR >0.05).

When we tested the two-gene (*IFI27* and *CXCL9*) and the three-gene (*IFI27*, *CXCL9* and *IFI-6-16*) signatures in the group of patients composed only of individuals with genotype 1 infection, both signatures correctly classified 31 of 41 patients (75.6%). The predictive accuracy was of 71.4% (15/21) and 80% (16/20) in NR and SVR patients, respectively, with the two-gene signature, and it was of 66.7% (14/21) and 85% (17/20) in NR and SVR patients, respectively, with the three-gene signature.

The 69 patients from Group A and Group B were grouped into two distinct samples according to their liver fibrosis stage, and the performance of the proposed gene signatures was evaluated separately in the “mild/moderate fibrosis” group (F1–F2; $n = 48$) and in the “severe fibrosis/cirrhosis” group (F3–F4, $n = 21$)(Metavir).²² In our experience, the majority of patients with CHC have mild to moderate fibrosis (87%) in comparison

with a minority having advanced fibrosis/cirrhosis (13%).²³ Both signatures accurately predicted treatment outcome in either group (table 8).

It is interesting to note that when we built predictive models using the entire data set (69 samples from Group A and Group B pooled together), the best results were obtained using a two-gene (*IFI27* and *CXCL9*, with $k = 3$) and a three-gene (*IFI27*, *CXCL9* and *ISG15*) subset using the KNN and WV classifier, respectively. Both predictive models accurately predicted treatment response for 56 of 69 patients (81.2%) with CHC. However, while the prediction errors within the NR and SVR groups of patients were similar with the two-gene signature KNN classifier (26.1 and 15.2%, respectively), they were markedly unbalanced when training the WV algorithm with the three-gene subset (39.1 and 8.7%, respectively) (fig 2).

Results obtained with the 11 RR samples

Besides the 69 NR and SVR patients included in the two independent Groups A and B, we analysed 11 additional RR patients. Compared with the normal liver controls, the three genes *IFI27*, *IFI-6-16* and *CXCL9* were all upregulated in the 11 RR samples: their liver expression values were 29.1 (SD 25.0), 32.6 (29.4), and 49.4 (46.8), respectively, after normalisation such that the median of the normal histological liver values was 1. Neither unsupervised hierarchical clustering nor supervised class prediction analysis could distinguish the RR samples from NRs or SVRs: instead of being assigned to a new group, they all clustered with either the NR or the SVR samples. Once trained on the pooled data from Group A and Group B, the two-gene classifier (*IFI27* and *CXCL9*) predicted 8 out of 11 RRs (73%) as SVRs, and 10 out of 11 RRs (91%) were predicted as SVRs with the three-gene classifier (*IFI27*, *CXCL9* and *IFI-6-16*). The sole RR sample classified as NR with the three-gene signature was included in the three RR samples classified as NRs with the two-gene signature.

DISCUSSION

In this study, we used large-scale real-time quantitative RT-PCR to predict treatment outcome in patients with CHC receiving PEG-IFN plus ribavirin. We examined the expression level of 58 genes known to be involved in various cellular and molecular mechanisms associated with response to treatment. All patients were well defined; all were naïve patients and received the same complete treatment of PEG-IFN α -2b with ribavirin.

Real-time quantitative RT-PCR is a promising alternative to microarrays for gene profiling. In particular, real-time RT-PCR is far more precise, reproducible and quantitative than microarrays. Using this approach, we have shown that several altered molecular pathways are involved in the progression of fibrosis in CHC.¹⁰ Real-time RT-PCR is also more useful for analysing weakly expressed genes such as cytokines in the present study.

The best classification results were obtained using a two-gene (*IFI27* and *CXCL9*) and a three-gene (*IFI27*, *CXCL9* and *IFI-6-16*) subset when performing KNN and WV analyses, respectively (fig 1). The identified two-gene and three-gene signatures accurately predicted treatment response for 79.3 and 72.4% of patients from the validation set, respectively (table 7).

As patients with genotype 1 infection are less likely to respond to antiviral therapy, we verified that the selected gene signatures still perform well in this subgroup of patients and observed that 75.6% (31/41) of the genotype 1 patients were correctly classified. The signatures also performed well whatever the degree of fibrosis: they predicted treatment outcome

Table 5 List of the genes that differ between NRs and SVRs in Group A

Gene	NR/normal	SVR/normal	NR/SVR	p Value*	FDR
<i>IFI-6-16</i>	126.5 (84.3)	35.6 (48.1)	3.5	0.002	0.039
<i>IFI27</i>	141.1 (107.3)	33.6 (40.7)	4.2	0.002	0.039
<i>ISG15</i>	88.5 (80.3)	24.1 (33.0)	3.7	0.002	0.039
<i>MX1</i>	40.1 (33.9)	14.9 (19.2)	2.7	0.006	0.059
<i>HERC5</i>	11.6 (9.1)	5.4 (5.3)	2.2	0.006	0.059
<i>TGFB2</i>	6.2 (7.6)	2.3 (1.7)	2.7	0.006	0.059
<i>OAS2</i>	27.0 (15.4)	14.9 (16.0)	1.8	0.016	0.118
<i>VEGFD</i>	5.5 (5.0)	2.3 (2.9)	2.4	0.020	0.118
<i>IL8</i>	49.2 (63.9)	15.4 (15.5)	3.2	0.020	0.118
<i>IFIT1</i>	387.8 (1399.3)	7.0 \pm 8.2	55.3	0.020	0.118

The gene expression values of the samples were normalised such that the median of the normal histological liver values was 1. Gene expression ratios were compared among NR and SVR liver gene expression values. Statistics are calculated using the Student t test.

*Not adjusted for multiple testing.

FDR, false discovery rate; NR, non-responder; SVR, sustained virological responder.

Table 6 List of the genes that differ between NRs and SVRs in Group B among the 10 genes with a non-adjusted p value below 0.05 in Group A

Gene	NR/normal	SVR/normal	NR/SVR	p Value*	FDR
<i>IFI-6-16</i>	443.2 (569.6)	78.3 (89.0)	5.7	0.002	0.026
<i>IFI27</i>	486.1 (370.7)	95.8 (108.9)	5.1	0.002	0.026
<i>ISG15</i>	341.3 (321.6)	43.0 (39.2)	7.9	0.002	0.026
<i>OAS2</i>	79.5 (77.9)	24.9 (26.5)	3.2	0.004	0.026
<i>HERC5</i>	24.6 (28.2)	5.5 (3.9)	4.5	0.004	0.026
<i>IFIT1</i>	24.3 (20.8)	9.1 (9.3)	2.7	0.012	0.047

The gene expression values of the samples were normalised such that the median of the normal histological liver values was 1. Gene expression ratios were compared among NR and SVR liver gene expression values. Statistics are calculated using the Student t test.

*Not adjusted for multiple testing.

FDR, false discovery rate; NR, non-responder; SVR, sustained virological responder.

with a high classification accuracy in both the “mild/moderate fibrosis” (F1 and F2; n = 48) and the “severe fibrosis/cirrhosis” (F3 and F4, n = 21) groups of patients (table 8). Note that the *IFI-6-16* (*G1P3*) liver gene expression has been shown to be increased in patients with moderate fibrosis (F2) compared with those with mild fibrosis (F1).¹⁰ This gene, in combination with others, also appears in the present report to enable the prediction of treatment response.

In a large randomised study, the proportion of patients with a sustained virological response was 54% among 511 CHC patients treated with PEG-IFN α -2b (1.5 μ g/kg/week) plus ribavirin, and the proportion of patients with response-relapse was found to be low (<18%).⁵ In the present study, the RR patients could not be distinguished from SVR and NR patients on the basis of their liver gene expression profiles, and most of them clustered with the SVR patients. Once trained on the pooled data from Group A and Group B, the two-gene classifier (*IFI27* and *CXCL9*) predicted 8 out of 11 RRs (73%) as SVRs, and 10 out of 11 RRs (91%) were predicted as SVRs with the three-gene classifier (*IFI27*, *CXCL9* and *IFI-6-16*). Interestingly, in a previous study which evaluated the efficacy of IFN-PEG α -2b plus ribavirin in the retreatment of 154 unselected patients with CHC who previously failed to respond or relapsed after a standard combination therapy, relapsers achieved higher response rates than non-responders (58.5% vs 13%).²⁴

Since a significant number of patients will fail to respond to treatment and will develop significant side effects, it is of major

interest, both for patient care and economically, to predict non-response as early as possible, ideally before treatment. For this purpose, the best signature to predict non-response is the two-gene (*IFI27*, *CXCL9*) subset, which correctly classified 100% of NR patients in the validation set, using the KNN classifier. This two-gene signature also displayed the highest overall predictive accuracy in the validation set (79.3%). Both *IFI27* and *CXCL9* genes belong to the IFN-stimulated genes family.

Many of the genes that were found to be differentially expressed between SVRs and NRs in both the training and the validation set (*IFI-6-16*, *IFI27*, *ISG15*, *OAS2* and *IFIT1*) belong to IFN-inducible genes. Interestingly, all these genes were upregulated in both SVR and NR samples compared with normal liver. It was proposed that IFN-inducible genes are expressed during CHC via direct activation of *IRF3*, *NFKB1* or *JUN* through IFN-dependent signalling events in host cells.¹⁴ Interestingly, the three genes included in the identified signatures (*IFI27*, *CXCL9* and *IFI-6-16*) were found to be strongly upregulated in CHC (with mild disease) in comparison with normal liver.¹¹ *IFI27* and *IFI-6-16* are IFN α / β -inducible genes with unknown function.²⁵ *CXCL9* is an IFN γ -inducible gene.

In a recent study, 18 genes were found to be differentially expressed in the liver between 15 NRs and 16 SVRs.¹² Interestingly, five of these genes (*IFI27*, *ISG15*, *IFI-6-16*, *OAS2* and *IFIT1*) were also identified in our study as being differentially expressed between NRs and SVRs. Using a number of independent classifier analyses, the authors¹² identified an eight-gene subset that predicted treatment response for 30 of the 31 patients; and another analysis from the same data using different methods found that the best prediction results were obtained using a 24-gene signature,²⁶ which correctly classified 28 of 31 patients. The predictive models identified in the present study involved a smaller number of genes which could be used as useful predictors. Moreover, the predictive accuracy of the signatures identified in our study was assessed using an independent validation set, and not only by performing a supervised LOOCV analysis on a single data set.

In another study, liver tissue samples prior to treatment by IFN or IFN/ribavirin were analysed by microarray.²⁷ In the IFN treatment group, the differentially expressed genes were mainly IFN-, lipid metabolism-, complement- and oxidoreductase-related genes. For the IFN/ribavirin combination treatment, a

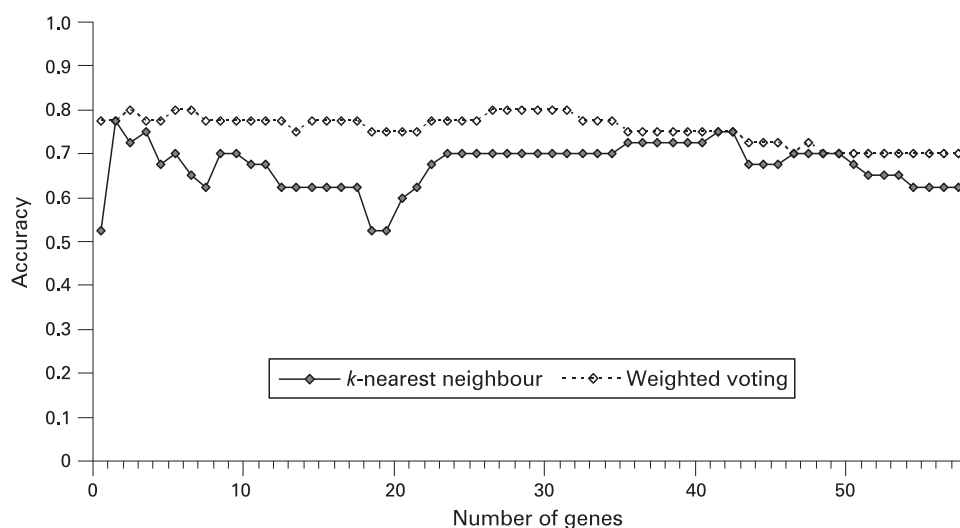
Figure 1 LOOCV (leave-one-out cross-validation) classification accuracies of supervised training models of increasing size in Group A (40 patients) using *k*-nearest neighbour and weighted voting classifiers.

Table 7 Supervised class prediction of NR and SVR using gene expression profiles

	Best selected model	NR predictive accuracy	SVR predictive accuracy	Overall predictive accuracy
<i>k</i> -nearest neighbour				
Group A (training set, n = 40)	<i>IFI27, CXCL9</i> *	57.1% (8/14)	88.5% (23/26)	77.5% (31/40)
Group B (test set, n = 29)		100% (9/9)	70.0% (14/20)	79.3% (23/29)
Weighted voting				
Group A (training set, n = 40)	<i>IFI27, CXCL9, IFI-6-16</i> *	64.3% (9/14)	88.5% (23/26)	80.0% (32/40)
Group B (test set, n = 29)		88.9% (8/9)	65.0% (13/20)	72.4% (21/29)

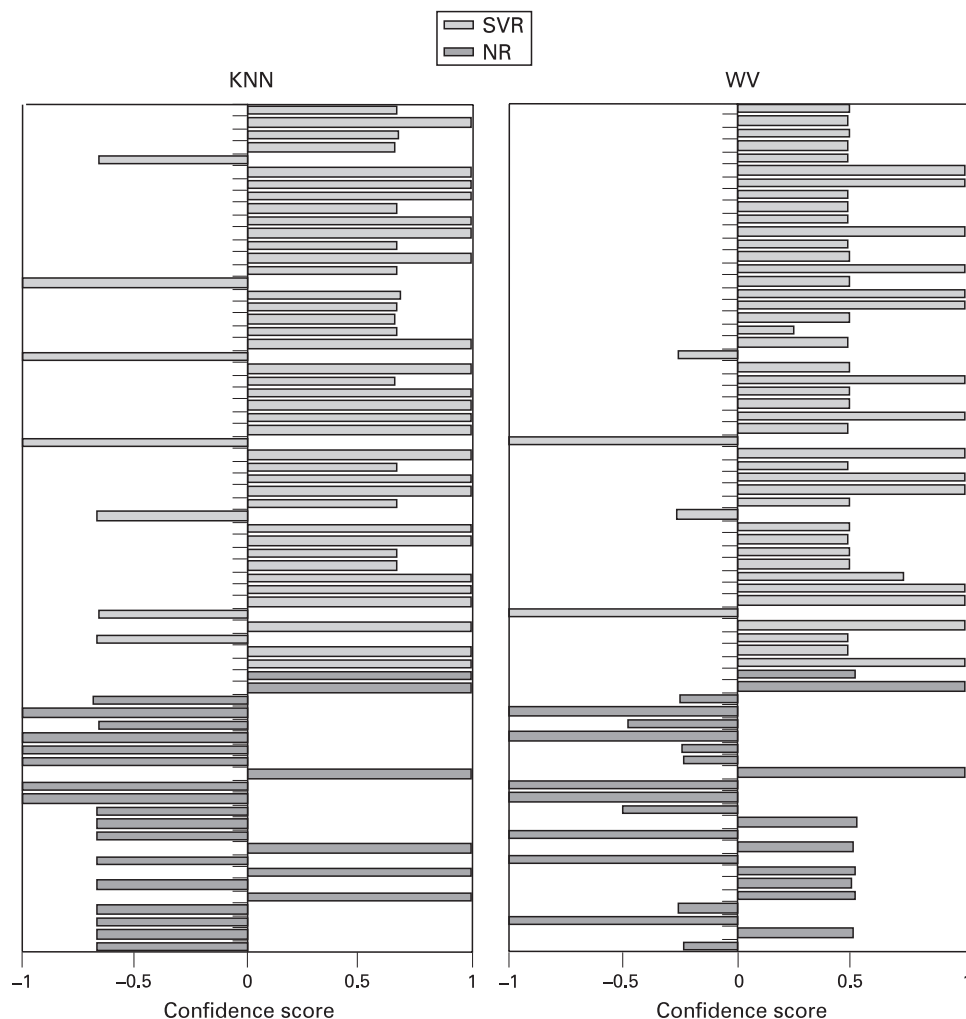
*p Value <0.005 (estimated using 200 randomly permuted data sets).
NR, non-responder; SVR, sustained virological responder.

Table 8 Performance of the two-gene signatures according to liver fibrosis stage

	NR predictive accuracy	SVR predictive accuracy	Overall predictive accuracy
Two-gene signature (<i>IFI27, CXCL9</i>)*			
Group F1–F2 (n = 48)	69.2% (9/13)	82.9% (29/35)	79.2% (38/48)
Group F3–F4 (n = 21)	90.0% (9/10)	81.8% (9/11)	85.7% (18/21)
Three-gene signature (<i>IFI27, CXCL9, IFI-6-16</i>)*			
Group F1–F2 (n = 48)	69.2% (9/13)	88.6% (31/35)	83.3% (40/48)
Group F3–F4 (n = 21)	60.0% (6/10)	100% (11/11)	81.0% (17/21)

*The predictive accuracy of the two-gene and of the three-gene signatures was evaluated using the *k*-nearest neighbour classifier and the weighted voting classifier, respectively.
NR, non-responder; SVR, sustained virological responder.

Figure 2 Results of the *k*-nearest neighbour (KNN) and weighted voting (WV) classifier analyses on the whole data set (69 patients) using the two-gene (*IFI27, CXCL9*) and the three-gene (*IFI27, CXCL9, ISG15*) predictor sets, respectively. Confidence scores in favour of non-responders (NRs) are presented as negative values, and confidence scores of class assignments in favour of sustained virological responders (SVRs) are presented as positive values. The actual origins of the patients are indicated (SVR, light grey; NR, dark grey). The overall accuracy of class assignment using the two-gene predictor with KNN was 81.2% (56/69) in the whole sample, in which 17 of 23 NRs were correctly classified as NRs, and 39 of 46 SVRs were correctly classified as SVRs. The overall classification accuracy of the three-gene predictor using WV was 81.2% (56/69) in the whole sample, in which 14 of 23 NRs were correctly classified as NRs, and 42 of 46 SVRs were correctly classified as SVRs.



different set of genes was identified including cyclophilin A and multidrug resistance protein. Surprisingly the discriminatory family genes that were identified were different in the two groups receiving either IFN or IFN/ribavirin.

Finally, another study analysing gene expression patterns in peripheral blood mononuclear cells during IFN therapy confirmed upregulation of genes thought to be IFN-stimulated genes as well as genes involved in antigen processing and presentation.²⁸

These three studies have some limitations. First, only a few patients were analysed without a validation set. Secondly, patients were heterogeneous according to treatment (IFN monotherapy, PEG-IFN, ribavirin, etc.). Thirdly, compliance with treatment was not assessed, and NR groups included patients without a complete course of treatment. Moreover, the diversity of microarray platforms used and the variability of microarray data emphasise the need for quality assurance. High quality RNA samples are essential, and must be rigorously assessed. Careful measures must be taken throughout RNA extraction to prevent the RNA from degrading. Furthermore, improved analytical procedures and the use of large numbers of patients are needed for validation.²⁹

Many of the genes found to be upregulated between NRs and SVRs encode molecules secreted in the serum (cytokines, IFN pathway, *IFI27* and *ISG15*) and provide a logical functional approach for the development of serum markers to predict treatment response. For instance, one study examined the levels of chemokines (*CXCL9*, *CXCL10* and *CXCL11*) that bind to CXC chemokine receptor 3 (*CXCR3*) to determine whether these chemokines might play a role in the failure of the immune system to clear HCV infection.³⁰ The baseline level of *CXCL10* (before treatment) was highest in patients who did not respond to therapy. These results suggest that plasma concentrations of immunoreactive *CXCL10* may predict non-responsiveness to antiviral therapy. Another recent study revealed that pretreatment IP-10 levels might help to predict sustained virological response in patients with HCV genotype 1 infection.³¹

In conclusion, our study demonstrated that NR and SVR patients have different gene expression profiles prior to treatment. The most notable changes in gene expression were mainly observed in the IFN-stimulated genes. We used two independent groups of patients (training set and validation set) and we could predict treatment response with a two-gene signature (*IFI27* and *CXCL9*). The majority of RR patients clustered with SVR patients. Interestingly, the basal liver levels of expression of IFN-stimulated genes were higher in NRs in comparison with SVRs. In NRs, the failure to respond to exogenous PEG-IFN could indicate a blunted response to IFN. This raises the possibility that, in NRs, the IFN-stimulated genes are already maximally induced. The genes included in the signature encode molecules secreted in the serum and provide a logical functional approach for the development of serum markers to predict response to treatment.

Funding: This study was supported by the French National Agency for Aids and Viral Hepatitis (ANRS). TA was supported by the French Association for the Study of the Liver (AFEF). AS was supported by the French Fondation pour la Recherche Médicale.

Competing interests: None.

REFERENCES

1. **Seeff LB**, Hoofnagle JH. The National Institutes of Health Consensus Development Conference Management of Hepatitis C 2002. *Clin Liver Dis* 2003;**7**:261–287.
2. **EASL International Consensus Conference on Hepatitis C**. Consensus Statement. *J Hepatol* 1999;**30**:956–61.
3. **Marcellin P**, Asselah T, Boyer N. Fibrosis and disease progression in hepatitis C. *Hepatology* 2002;**36**:S47–56.
4. **Craxi A**, Licata A. Clinical trial results of peginterferons in combination with ribavirin. *Semin Liver Dis* 2003;**23**(Suppl 1):35–46.
5. **Manns MP**, McHutchison JG, Gordon SC, *et al*. Peginterferon alpha-2b plus ribavirin compared with IFN-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001;**358**:958–65.
6. **Fried MW**, Shiffman ML, Reddy KR, *et al*. Peginterferon alpha-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002;**347**:975–82.
7. **Hadziyannis SJ**, Sette H Jr, Morgan TR, *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**:346–55.
8. **Bronowicki JP**, Ouzan D, Asselah T, *et al*. Effect of ribavirin in genotype 1 patients with hepatitis C responding to pegylated interferon alpha-2a plus ribavirin. *Gastroenterology* 2006;**131**:1040–8.
9. **DeRisi J**, Penland L, Brown PO, *et al*. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996;**14**:457–60.
10. **Asselah T**, Bieche I, Laurendeau I, *et al*. Liver gene expression signature of mild fibrosis in patients with chronic hepatitis C. *Gastroenterology* 2005;**129**:2064–75.
11. **Bieche I**, Asselah T, Laurendeau I, *et al*. Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection. *Virology* 2005;**332**:130–44.
12. **Chen L**, Borozan I, Feld J, *et al*. Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection. *Gastroenterology* 2005;**128**:1437–44.
13. **Honda M**, Kaneko S, Kawai H, *et al*. Differential gene expression between chronic hepatitis B and C hepatic lesion. *Gastroenterology* 2001;**120**:955–66.
14. **Gale M Jr**, Foy EM. Evasion of intracellular host defence by hepatitis C virus. *Nature* 2005;**436**:939–45.
15. **Helbig KJ**, Lau DT, Semendric L, *et al*. Analysis of ISG expression in chronic hepatitis C identifies viperin as a potential antiviral effector. *Hepatology* 2005;**42**:702–10.
16. **Mihm S**, Frese M, Meier V, *et al*. Interferon type I gene expression in chronic hepatitis C. *Lab Invest* 2004;**84**:1148–59.
17. **Smith MW**, Yue ZN, Korh MJ, *et al*. Hepatitis C virus and liver disease global transcriptional profiling and identification of potential markers. *Hepatology* 2003;**38**:1458–67.
18. **Benjamini Y**, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;**57**:289–300.
19. **Cover TM**, Hart PE. Nearest neighbor pattern classification. *IEEE Trans Info Theory* 1967;**IT-13**:21–7.
20. **Reich M**, Liefeld T, Gould J, *et al*. GenePattern 2.0. *Nat Genet* 2006;**38**:500–1.
21. **Devlin B**, Roeder K, Wasserman L. Analysis of multilocus models of association. *Genet Epidemiol* 2003;**25**:36–47.
22. **Bodossa P**, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;**24**:289–93.
23. **Asselah T**, Boyer N, Guimont MC, *et al*. Liver fibrosis is not associated with steatosis but with necroinflammation in French patients with chronic hepatitis C. *Gut* 2003;**52**:1638–43.
24. **Moucari R**, Ripault M-P, Oules V, *et al*. High predictive value of early viral kinetics in retreatment with peginterferon and ribavirin of chronic hepatitis C patients non-responders to standard combination therapy. *J Hepatol* 2007;**46**:596–604.
25. **Li Q**, Zhang M, Kumar S, *et al*. Identification and implantation stage-specific expression of an interferon-alpha-regulated gene in human and rat endometrium. *Endocrinology* 2001;**142**:2390–400.
26. **Kim IJ**, Kang HC, Park JG. Evaluation of microarray analysis for predicting treatment responsiveness in patients with chronic hepatitis C viral infection. *Gastroenterology* 2005;**129**:1803.
27. **Hayashida K**, Daiba A, Sakai A, *et al*. Pretreatment prediction of interferon-alpha efficacy in chronic hepatitis C patients. *Clin Gastroenterol Hepatol* 2005;**3**:1253–9.
28. **Ji X**, Cheung R, Cooper S, *et al*. Interferon alpha regulated gene expression in patients initiating interferon treatment for chronic hepatitis C. *Hepatology* 2003;**37**:610–21.
29. **Asselah T**, Bièche I, Paradis V, *et al*. Genetics, genomics, and proteomics: implications for the diagnosis and the treatment of chronic hepatitis C. *Semin Liver Dis* 2007;**27**:13–27.
30. **Butera D**, Marukian S, Iwamaye AE, *et al*. Plasma chemokine levels correlate with the outcome of antiviral therapy in patients with hepatitis C. *Blood* 2005;**106**:1175–82.
31. **Lagging M**, Romero AI, Westin J, *et al*. IP-10 predicts viral response and therapeutic outcome in difficult-to-treat patients with HCV genotype 1 infection. *Hepatology* 2006;**44**:1617–25.



Liver gene expression signature to predict response to pegylated interferon plus ribavirin combination therapy in patients with chronic hepatitis C

T Asselah, I Bieche, S Narguet, A Sabbagh, I Laurendeau, M-P Ripault, N Boyer, M Martinot-Peignoux, D Valla, M Vidaud and P Marcellin

Gut 2008 57: 516-524 originally published online September 25, 2007
doi: 10.1136/gut.2007.128611

Updated information and services can be found at:
<http://gut.bmj.com/content/57/4/516>

References

These include:

This article cites 31 articles, 2 of which you can access for free at:
<http://gut.bmj.com/content/57/4/516#BIBL>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections

Articles on similar topics can be found in the following collections
[Hepatitis C](#) (160)

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>