Impact of virus subtype and host *IFNL4* genotype on large-scale RNA structure formation in the genome of hepatitis C virus

PETER SIMMONDS,[1] LIZE CUYPERS,[2] WILL L. IRVING,[3] JOHN MCLAUCHLAN,[4] GRAHAM S. COOKE,[5] ELLIE BARNES,[1] STOP-HCV CONSORTIUM[6] and M. AZIM ANSARI[1]

[1]Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, OX1 3SY, Oxford, United Kingdom

[2]University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Research, BE 3000, Leuven, Belgium

[3]Faculty of Medicine and Health Sciences, University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham, NG7 2UH, United Kingdom

[4]MRC-University of Glasgow Centre for Virus Research, Glasgow, G61 1QH, United Kingdom

[5]Imperial College London, London, W2 1PG, United Kingdom

ABSTRACT

Mechanisms underlying the ability of hepatitis C virus (HCV) to establish persistent infections and induce progressive liver disease remain poorly understood. HCV is one of several positive-stranded RNA viruses capable of establishing persistence in their immunocompetent vertebrate hosts, an attribute previously associated with formation of large-scale RNA structure in their genomic RNA. We developed novel methods to analyze and visualize genome-scale ordered RNA structure (GORS) predicted from the increasingly large data sets of complete genome sequences of HCV. Structurally conserved RNA secondary structure in coding regions of HCV localized exclusively to polyprotein ends (core, NS5B). Coding regions elsewhere were also intensely structured based on elevated minimum folding energy difference (MFED) values, but the actual stem–loop elements involved in genome folding were structurally poorly conserved, even between subtypes 1a and 1b. Dynamic remodeling was further evident from comparison of HCV strains in different host genetic backgrounds. Significantly higher MFED values, greater suppression of UpA dinucleotide frequencies, and restricted diversification were found in subjects with the TT genotype of the rs12979860 SNP in the *IFNL4* gene compared to the CC (nonexpressing) allele. These structural and compositional associations with expression of interferon-λ4 were recapitulated on a larger scale by higher MFED values and greater UpA suppression of genotype 1 compared to genotype 3a, associated with previously reported HCV genotype-associated differences in hepatic interferon-stimulated gene induction. Associations between innate cellular responses with HCV structure and further evolutionary constraints represent an important new element in RNA virus evolution and the adaptive interplay between virus and host.

Keywords: HCV; RNA secondary structure; evolution; interferon lambda; UpA dinucleotide; RNaseL

INTRODUCTION

Hepatitis C virus (HCV) is a major pathogen of humans infecting more than 71 million individuals. An estimated 400,000 deaths per year occur as a direct consequence of progressive inflammatory disease and hepatocellular carcinoma over many decades of persistent infection (Westbrook and Dusheiko 2014; Polaris Observatory HCV Collaborators 2017). HCV is one of a small number of RNA viruses capable of establishing long term infections in humans and other vertebrates (Hoofnagle 2002). This contrasts with the more typical outcomes of RNA virus infections—acute infection and rapid clearance mediated through the actions of innate and adaptive immune responses and subsequent protective immunity. How HCV is able to counteract the otherwise powerful host responses mounted against viruses has remained unclear for several decades since the concept of RNA virus persistence

---

became recognized (Oldstone 2009; Randall and Griffin 2017).

Persistence is the observed outcome of infections with several other mammalian viruses including human pegivirus (HPgV) and related viruses in the *Hepacivirus* and *Pegivirus* genera of *Flaviviridae* infecting other mammals (Ploss et al. 2009; Baechlein et al. 2015; Pfaender et al. 2015; Trivedi et al. 2017). It is also documented in several picornaviruses (e.g., foot-and-mouth disease virus; FMDV; Condy et al. 1985; Cortey et al. 2019) and caliciviruses (murine norovirus [MNV] and some vesiviruses) (Forrester et al. 2003; Coyne et al. 2006; Hsu et al. 2006; Thackray et al. 2007). Collectively, they differ substantially in many aspects of their host interactions and disease outcomes. Infections with several are asymptomatic, as in the case of HPgV and, where known, in other animal pegiviruses and in MNV. While FMDV may cause severe and often fatal vesicular disease in cows and other ruminants, infections in buffalo, its natural host, is often clinically inapparent (Thomson et al. 1992). Chronic and progressive liver disease is observed in infections with HCV and with some hepaciviruses infecting other host species, such as GBV-B in tamarins (Beames et al. 2001; Bukh et al. 2001) and equine hepaciviruses in horses (Pfaender et al. 2015; Ramsay et al. 2015). However, bovine hepacivirus infections of cows have been reported as entirely apathogenic (Baechlein et al. 2019).

While these features of infection are disparate in both pathogenicity and target tissues, a unifying characteristic of these persistent RNA viruses is their intensely structured RNA genomes, where a high degree of internal sequence complementarity creates elaborate tandem arrays of stem–loops and potentially tertiary structure elements spanning most of the genomic RNA (Simmonds et al. 2004, 2008; Davis et al. 2008; McFadden et al. 2013). Existing bioinformatic and physicochemical analysis of this genome attribute, that we termed genome-scale ordered RNA structure (GORS) reveals many differences from the better characterized discrete elements of folded RNA found in RNA virus genomes. The latter may serve as replication elements or mediate ribosomal interactions in translation initiation or control (e.g., frame shifting). Contrastingly, the distributed and extensive nature of GORS leads to a major difference in the configuration of the RNA in terms of its overall shape and accessibility to hybridization to external probes (Davis et al. 2008). How this globally folded configuration of genomic RNA contributes to the interaction of the virus and its host (and its ability to establish persistence) remains unknown.

In the current study, we have reinvestigated many aspects of GORS in HCV infections now that there are many thousands of accurately determined whole-genome sequences of HCV available. Furthermore, RNA structure information on pairing interactions is now available for several whole-genome RNA molecules using recently developed SHAPE methods (Mauger et al. 2015; Pirakitikulr et al. 2016). In the current study, we have analyzed the nature and phylogenetic conservation of RNA structures in HCV and their commonality in extents and constraints with RNA folding in the genomes of other persistent viruses. We show that the formation of RNA structure may be a much more dynamic process that may further evolve in response to genetic background of the host it infects.

RESULTS

RNA structure prediction in HCV genomes

The propensity of RNA to internally base pair and its structural configuration is dependent on both the order of bases and the G + C content of the sequence. The sequence order component of RNA structure formation in HCV genomes was estimated by comparison of minimum folding energies (MFEs) of native sequences with those of the same sequence scrambled in base order while maintaining native dinucleotide frequencies using the algorithm NDR. The resulting values (MFEDs) were generated for consecutive 240 base fragments incrementing by nine bases through the coding region of whole-genome sequences of HCV genotypes 1a ($n = 388$), 1b ($n = 106$), and 3a ($n = 855$). MFED values greater than zero were observed throughout the genomes of each genotype analyzed (Fig. 1A), although there were some differences in the degree of MFED elevation predicted for different genome regions. These differences were mirrored in a parallel plot of Z-scores representing the MFE of the native sequence in the distribution of MFEs of the sequence order randomized controls (Supplemental Fig. S1A; Supplemental Data).

Overall, however, elevated mean MFED values for whole coding regions sequences were observed in sequences from all genotypes (6.0%–10.9% range in MFED values) and more strikingly, between variants of the same genotype (Fig. 1B). Although there was a scatter of MFED values within each genotype, MFED values were significantly higher in genotype 1a than 1b and 3a (Fig. 1B). Genotype-associated differences in predicted RNA structure formed by the three genotypes were observed across the genome (Supplemental Fig. S1B; Supplemental Data) although not confined to individual stem–loops. Increased MFED values were also observed irrespective of the underlying degree of sequence diversity of HCV in different parts of the genome. For example, they remained elevated throughout the E1/E2 region despite the substantially greater sequence variability of the envelope protein gene sequences, particularly around the hypervariable region of E2 (Supplemental Fig. S1C; Supplemental Data).
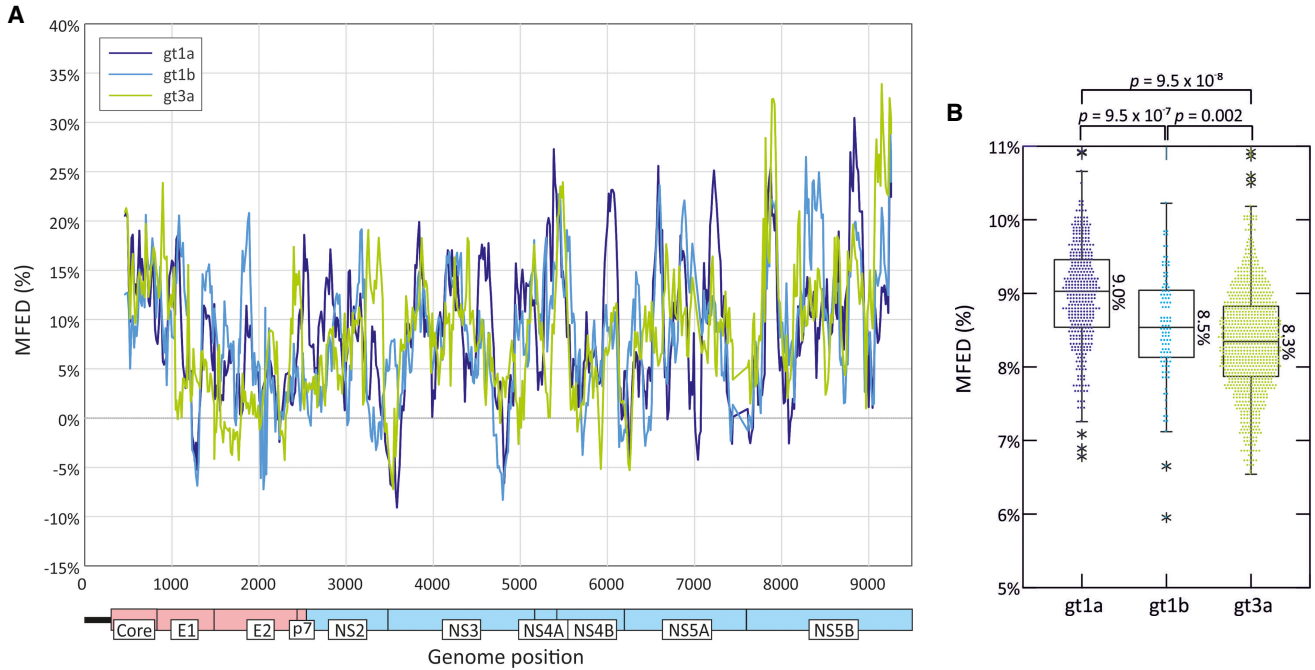
FIGURE 1. MFED differences in HCV coding regions of genotypes 1a, 1b, and 3a. MFED values were calculated for consecutive 240 base fragments, incrementing by 15 bases across the HCV coding region (569 fragments). MFED values were calculated by subtraction of the MFE of the native sequence by the mean value of 49 sequence order randomized controls. (*A*) Mean values of genotypes 1a (*n* = 388), 1b (*n* = 106), and 3a (*n* = 855) sequences plotted by genome position. (*B*) Mean values of sequence fragments for individual polyprotein sequences of gts 1a, 1b, and 3a. The box plots show (from the *top*) 2 standard deviations (SDs) above mean, 1 SD above mean, mean, 1 SD below mean, and 2 SDs below mean; stars represent outliers outside this range. Genome positions were normalized by reference to the H77 prototype sequence, AF011751 (coding region positions 342–9377). Distributions of MFED values of different genotypes were compared by Kruskall–Wallace nonparametric test; *P*-values on comparing 1a with 1b, 1a with 3a, and 1b with 3a are shown *above* graph.

The use of mean values of bulk MFED values for collections of sequences of the same subtype or genotype provides a very coarse-grained indication of the RNA structural variability of HCV at the genotype or individual sequence level. To visualize structural heterogeneity more effectively, we developed a new plotting method for RNA structures predicted by RNAFOLD. These were based on sampling an ensemble of suboptimal folds generated by the program SubOpt.exe using predictions for individual sequences and recording pairing predictions supported by 50% or more of the ensemble. For calculating stem–loop heights, unpaired bases in terminal loops of each secondary structure were identified in the ensemble consensus connect file for each sequence and plotted as a height of zero (z-axis) with their genome positions and sequence number in x- and y-axes. Neighboring bases were successively plotted according to a color scale that reflects their distance in the stem from the terminal loop. Examples of a standard stem–loop, an interrupted stem–loop and a more complex clover leaf structure are shown in their MFOLD representation and three- and two-dimensional contour plots (Figs. 2A,B,C, respectively). Contour plots therefore provide an approximate visualization of the positions, shapes and sizes of RNA structure elements across whole alignments of potentially large numbers of se-

quences. Applying this method to a short section of the HCV genome in the core/E1 gene region (Fig. 2D), the three-dimensional displays sequence positions (x-axis), sequence number in the alignment (y-axis) and relative heights of predicted stem–loops. Its transformation to a two-dimensional representation and associated depiction of structural heterogeneity between sequences greatly facilitates an analysis of RNA structure conservation between and within genotypes.

Much larger scale contour plots of the coding region sequences of samples obtained in the current study from gt1a, gt1b, gt2a, and gt3a (supplemented with published sequences for gt2a) were created from 13 sequential 1600 base fragments of each alignment incrementing by 400 bases between fragments. The fragment size of 1600 bases chosen for individual contour plots reproduces pairings predicted from shorter and longer sequence fragments (Supplemental Fig. S2A; Supplemental Data); this represents a necessary compromise between computational time and enabling the method to detect longer range-base pairings. Using these settings, this composite genome-wide representation revealed the existence of highly conserved structure elements across genotypes located specifically in the core and NS5B regions (Fig. 3A,B). These were highly concordant with RNA pairings detected by
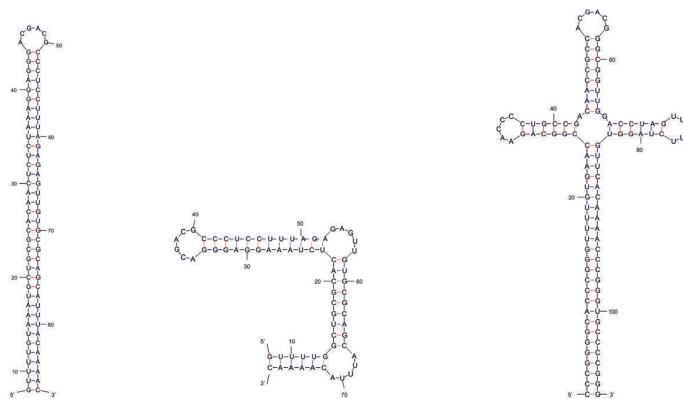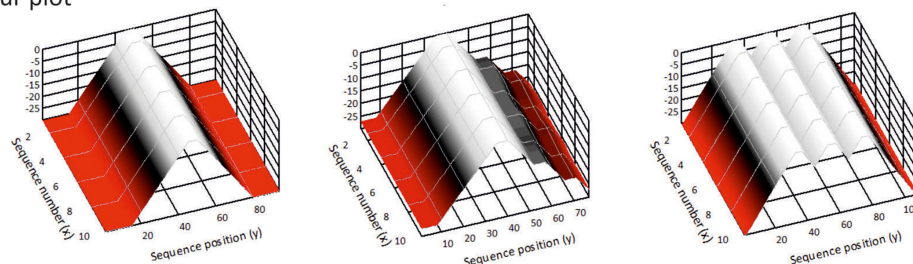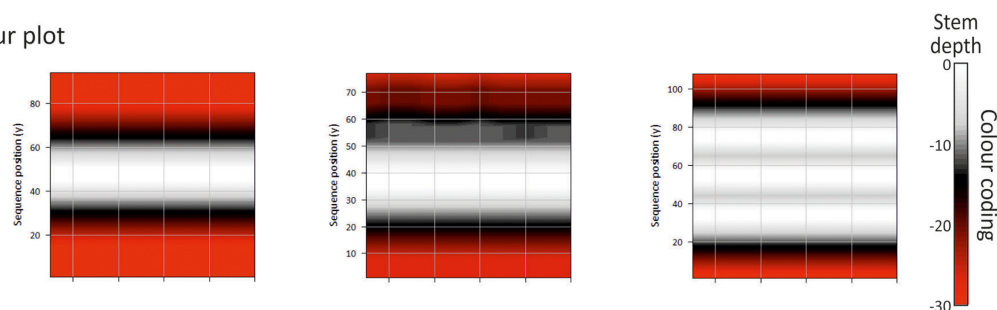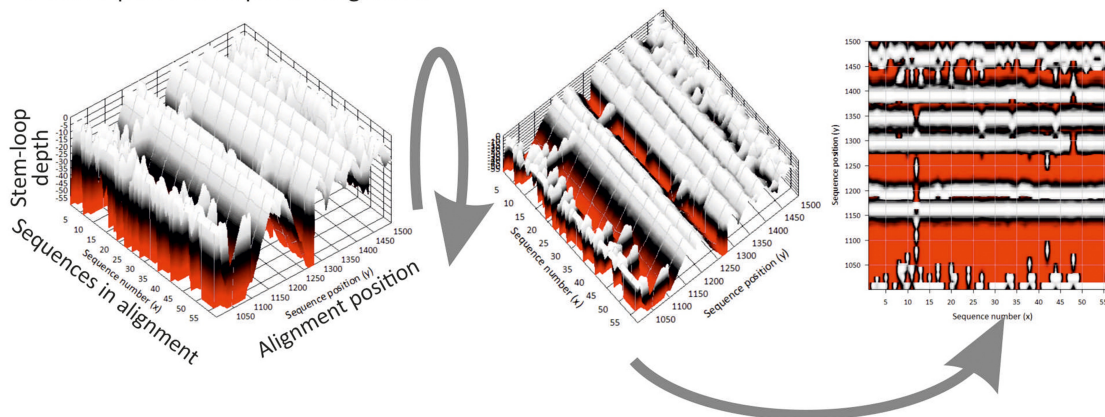
FIGURE 2. Visualization of RNA structure conservation in a sequence alignment. Comparison of RNA structure visualization using (*A*) MFOLD, (*B*) 3D, and (*C*) 2D contour plots produced by StructureDist of example RNA secondary structure elements (stem–loop, interrupted stem–loop, and clover leaf). Predicted consensus positions of terminal loops in RNAFold RNAsubopt output were aligned and plotting depths based on pairings either side calculated. They were depicted as canyons corresponding to duplex lengths (*B*) and then rotated to create a 2D representation of sequences in the alignment (*x*-axis) and alignment position (*y*-axis) and color-coded depth. (*D*) Application of contour plotting to an alignment of HCV sequences in the core/E1 region in three- and two-dimensional representations. Pairing predictions represent as majority role consensus derived from the ensemble of suboptimal folds produced by the RNAsubopt program; similar results are obtained irrespective of whether suboptimal folds are sampled from a defined range of MFE values from the optimum, suboptimal structures sampled based on their Boltzmann weights in a partition function, or Zuker suboptimals (Supplemental Fig. S2B; Supplemental Data).
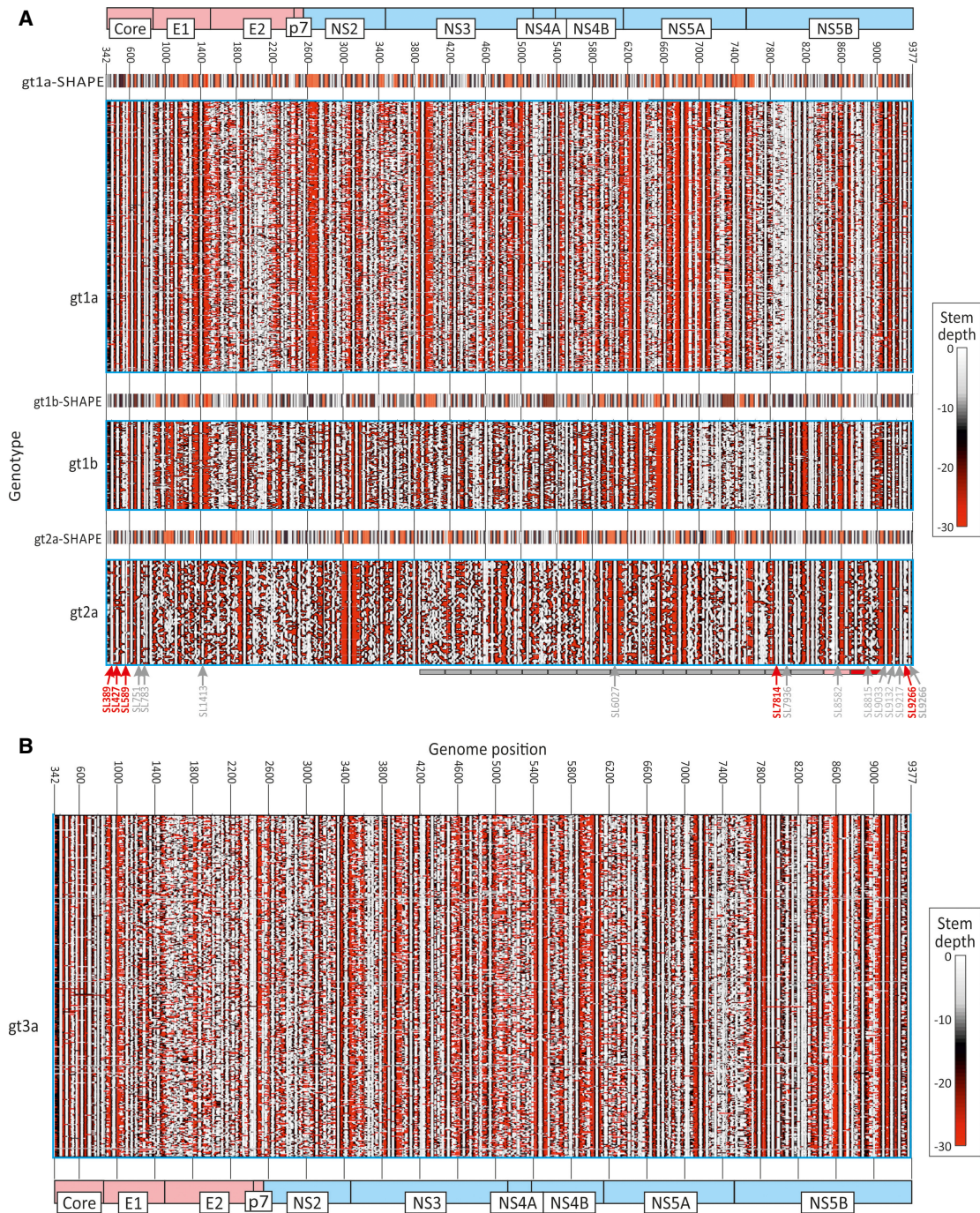
FIGURE 3. Comparison of RNA structure prediction for HCV genotypes 1a, 1b, 2a, and 3a sequences shown as contour plots. (*A*) HCV genotypes 1a, 1b, and 2a with SHAPE data; (*B*) Genotype 3a. Contour plots of RNA structure predictions by RNAFOLD of the coding regions of (*A*) 388 gt1a, 106 gt1b, 66 gt2a, and (*B*) 855 gt3a sequences obtained in the current study. *Inset* plots show RNA structures previously determined for gt1a (H77), gt1b (Con1), and gt2a (JFH1) RNA transcripts by SHAPE (Mauger et al. 2015) plotted to the same scale. The contour plot scale and genome positions of RNA structure elements were converted to those in the H77 (gt1a) prototype sequence (AF011751). Each figure was assembled from graphical output from StructureDist of 13 sequential 1600 base fragments of the alignment incrementing by 400 bases between fragments. The positions of functionally investigated RNA structures of genome segments in previous studies (You et al. 2004; McMullan et al. 2007; Diviney et al. 2008; Chu et al. 2013; Mauger et al. 2015; Pirakitikulr et al. 2016) are indicated by arrows *underneath* the contour plot of gt2a; numbering and arrows point to the first (5′) bases of the stem–loop. Those showing little or no phenotypic effect on disruption are depicted in gray; those leading to substantial (>1 log) reductions in replication shown in red. Phenotypes of viruses with 17 systematically mutated segments across the NS region (Chu et al. 2013) indicated in gray (no effect), gray (mild effect—segment 16), and red (severe effect—segment 17). These functional depictions are a simplification and interactions between RNA structures with other elements in coding and noncoding regions have been excluded; for this information please consult the original publications.

SHAPE mapping for genotypes 1a, 1b, and 2a in a previous study (Mauger et al. 2015) replotted in the same contour format (Fig. 3). RNA structure determination by SHAPE (Mauger et al. 2015; Pirakitikulr et al. 2016) indeed largely verifies previously described RNA structure prediction programs and analysis of covariance (Tuplin et al. 2004) and RNase mapping (Tuplin et al. 2004). Several of the RNA secondary structures have been functionally characterized through investigation of effects on replication when disrupted (You et al. 2004; McMullan et al. 2007; Diviney et al. 2008; Mauger et al. 2015; Pirakitikulr et al. 2016) or through measurement of replication effects of systematic large-scale sequence mutation in different coding region segments (Chu et al. 2013). The simplified summary in Figure 3A highlights the concentration of investigated structures in the genome ends. It also demonstrates the relative infrequency with which identified structures influence replication capacity in cell culture.

However, the most striking aspect of the structure predictions is the extensive variability in the position and intensity of structure formation between genotypes and subtypes of HCV. In regions outside of the core and 3′terminal NS5B genes, structure conservation was evident at subtype level only, with only a small number of predicted stem–loops shared across two or more subtypes. There were additionally regions that showed extensive predicted structural heterogeneity within subtypes, including the envelope genes and parts of NS5A and NS5B. Despite these structural differences, areas without sequence conservation between subtypes nevertheless showed elevated MFED values (Fig. 1A), indicating that in the period in which the four subtypes had diversified, there has been a substantial degree of RNA structure reinvention and appearance of quite different structured elements while at the same time, maintaining similar levels of overall structure (Fig. 1B). Further structural diversity of HCV is apparent on examination of representative examples of each currently classified HCV subtype (Supplemental Fig. S3; Supplemental Data), where the only shared structural elements are the stem–loops in the core and 3′ NS5B regions. Despite the sequence divergence (>30%) and structural heterogeneity evident from the contour plots, each currently classified HCV subtype showed elevated MFED values, but with some variability between genotypes in mean values and ranges (Supplemental Fig. S4; Supplemental Data).

The ability of contour plots to localize areas of RNA secondary structure was investigated by extension of the analysis to virus groups with previously documented structured genomes and for which full genome sequences from a range of strains has been previously obtained (foot-and-month disease virus [FMDV] type O, human pegivirus type 1 [HPgV-1] and murine norovirus type 3 [MNV3] analyzed in the current study) (Table 1; Simmonds et al. 2004, 2008; Davis et al. 2008; McFadden et al. 2013). Contour

TABLE 1. Sequence data sets used in the study

Virus group	N	MFED	Het. score[a]	MPD[b]
HCV gt1a	388	8.96%	6.33	8.25%
HCV gt1b	106	8.53%	6.32	9.38%
HCV gt2a	67	7.70%	7.42	10.91%
HCV gt3a	855	8.35%	6.00	8.51%
EV-A71	138	0.68%	9.75	14.77%
HPeV-3	38	1.62%	9.12	9.17%
JEV	65	1.40%	8.72	8.76%
FMDV-O	71	11.71%	5.75	10.60%
MNV	63	7.09%	7.59	10.19%
HPgV-1	100	11.99%	4.93	12.24%

[a]Mean difference in predicted pairing depth between sequences at each site within a data set; serves as a metric of heterogeneity of pairing predictions between sequences.
[b]Mean pairwise nucleotide sequence P distances (all sites) between members of each sequence group.

plots of these were compared with example data sets of unstructured virus groups (enterovirus A71 [EV-A71], human parechovirus type 3 [HPeV-3] and Japanese encephalitis virus [JEV]) (Fig. 4). These virus groups were selected to possess similar degrees of naturally occurring sequence diversity as found within the gt1a, 1b, 2a, and 3a data sets (Table 1) which might otherwise influence degrees of RNA structure conservation. Similarly to HCV, the structured viruses (FMDV-O, HPgV-1, and MNV-3) showed evidence for conserved areas of RNA structure formation throughout their genomes, with substantial ordering of stem–loop structures in large parts of the genome. HPgV-1, which possesses the highest MFED values, shows a series of very large and quite regularly spaced stem–loops throughout the entire coding region. FMDV similarly shows packed RNA structures although with shorter stem–loops than observed in HPgV-1. MNV has a lower overall MFED value than the other structured viruses and possesses fewer regions of ordered structure than FMDV and HPgV-1. It is also apparent that some conserved loops (e.g., at positions 5045) may play functional roles as elements of the subgenomic promoters for the capsid gene (Simmonds et al. 2008). Further structures may contribute to the expression of ORF3 genes and frameshifting for ORF4. As with HCV, extension of structure prediction to a wider range of variants of FMDV variants and types introduced substantially greater heterogeneity into the structure predictions, with marked differences in structure predictions between FMDV serotypes A, C and O in the more divergent structural genes VP4, VP2, VP3, and VP1 (Supplemental Fig. S5; Supplemental Data), despite there being elevated MFED values throughout this region.

The contour plots for HCV and the three structured virus data sets (MFED values 7.7%–12.0%) were quite distinct from those of unstructured viruses (MFED values ranging
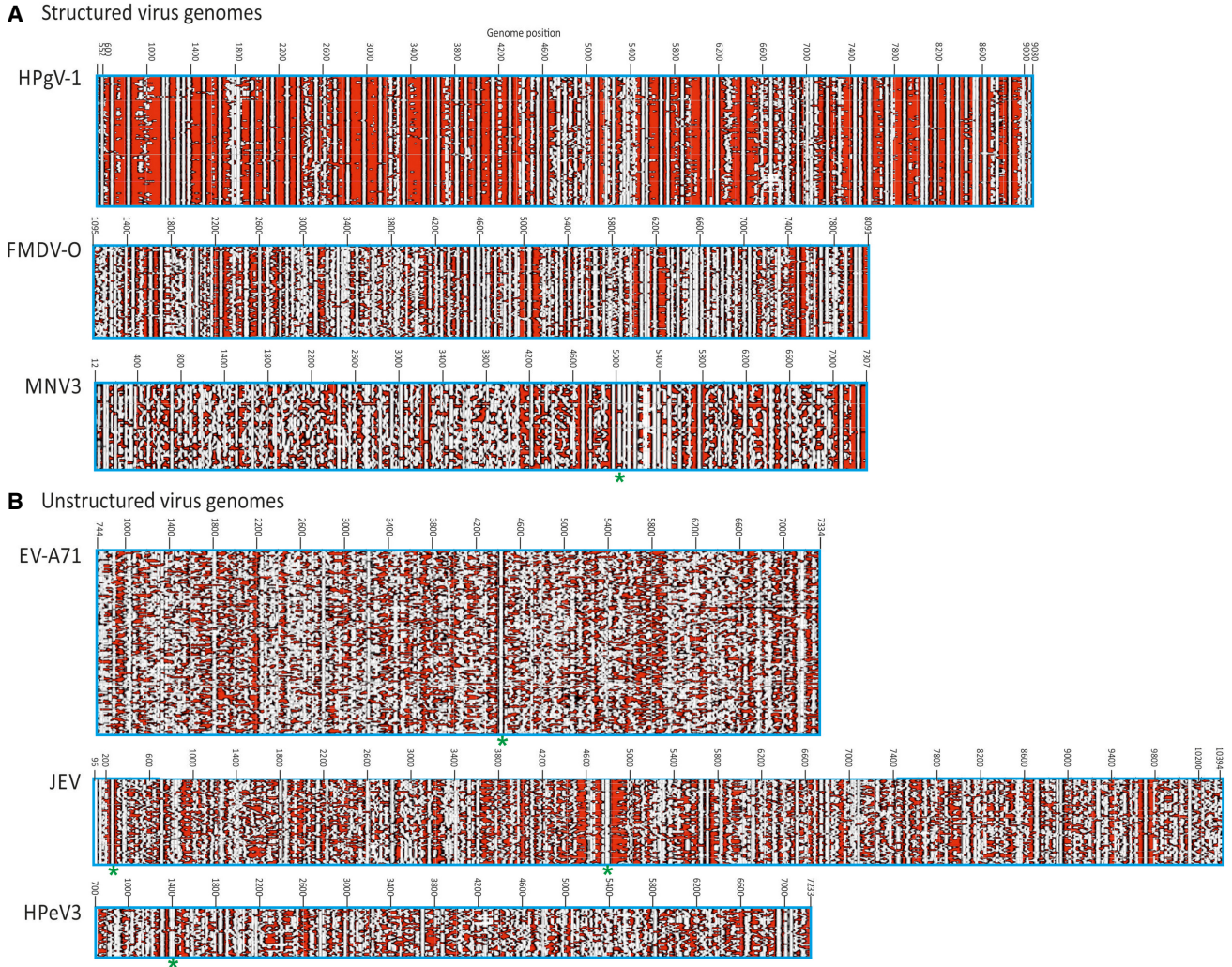
A Structured virus genomes

B Unstructured virus genomes

FIGURE 4. Contour plots of other structured and unstructured virus genomes. Comparison of contour plots of (*A*) structured (MFED > 5%) and (*B*) unstructured (MFED < 2%) virus groups plotted to scale. These show similar levels of sequence diversity (8%–15%; Table 1). Previously documented structures indicated by asterisks—EV-A71: *cis*-replication element (Rieder et al. 2000; Goodfellow et al. 2003); JEV: core region stem–loops; MNV: subgenomic promoter (Simmonds et al. 2008); HPeV: *cis*-replication element (Al Sunaidi et al. 2007).

from 0.7%–1.6%) (Table 1). For EV-A71, large parts of the coding region showed no conserved structure formation, the exception being the *cis*-replicating element at position 4442 (Rieder et al. 2000; Goodfellow et al. 2003). The contour plot for JEV is similarly largely unstructured but core loops associated with translation and a large structure at position 4400 (no documented function) are evident. The HPeV-3 CRE (Al Sunaidi et al. 2007) is similarly evident at position 1400 but the genome possesses few other conserved stem–loops.

The contour plot program was also used to quantify the degree of heterogeneity in folding between different sequences in an alignment. The degree of folding heterogeneity can be calculated as the mean difference in folding depth on pairwise comparisons of sequences in an alignment. Sequences in areas of conserved pairings show the same folding depths and a calculated heterogeneity

of zero; pairing depths in unstructured RNA are arbitrary and therefore heterogeneous. Mean values for the whole alignments of unstructured viruses were around nine while structured viruses showed degrees of conservation related to their MFED values (Table 1; Fig. 5). As a control, sequences of HCV genotype 1a were permuted by the algorithm CDLR to scramble codon order while retaining amino acid sequence of the encoded polyprotein and preserving native dinucleotide frequencies. It did, however, largely disrupt RNA secondary structures within the mutated region, with a reduction in mean MFED value from 9.0% to 0.5%. This change was paralleled by a large increase in heterogeneity in pairing predictions (Fig. 5). Contrastingly, little change in MFED values or pairing heterogeneity was observed on CDLR scrambled EV-A71 sequences, verifying the virtual absence of RNA secondary structure in the native sequences of this virus. The agreement between
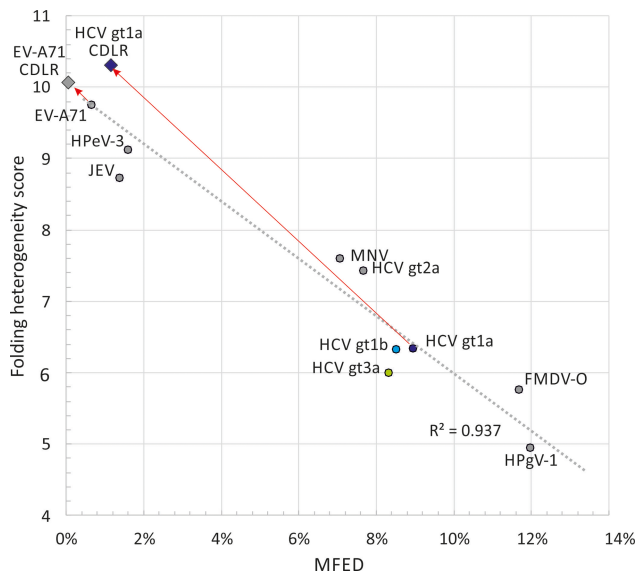
FIGURE 5. Association of MFED scores with folding prediction heterogeneity. Comparison of MFED scores with folding prediction heterogeneity for complete polyprotein sequences of HCV, other predicted structured viruses (HPgV-1, FMDV-O, and MNV), and unstructured viruses (JEV, HPeV-3, and EV-A71). Folding heterogeneity was expressed as the difference in folding depth between different sequences within an alignment averaged over the whole polyprotein sequence. As controls, sequence order randomized sequence data sets were generated by the algorithm CDLR for structured (HCV-1a) and unstructured (EV-A71) genome sequences and their folding energies (MFED score; x-axis) and structural heterogeneity (y-axis) reestimated.

these two distinct metrics of RNA folding—sequence order-dependence and heterogeneity of predicted pairings ($R^2 = 0.93$) supports the value of MFED values as a bulk metric of RNA structure formation in viral sequences.

Influence of host on HCV RNA structure formation

Although different HCV genotypes and subtypes showed distinct distributions of MFED values (Fig. 1B; Supplemental Fig. S4; Supplemental Data), there remained a substantial variability and overlap in MFED values between categories. To investigate whether host factors also influenced RNA structure formation, the relationship between several host demographics and clinical features with MFED values were investigated by multivariate analysis. This was based upon available data from genotype 3a individuals ($n = 503$) in the BOSON cohort (Table 2). We additionally used information on naturally occurring differences in interferon λ4 expression from the *IFNL4* gene inferred from the single nucleotide polymorphism (SNP) rs12979860, where CC, CT and TT alleles associated with no expression, medium and high-level expression respectively (Thomas et al. 2009). The SNP was included in the analysis because of its previously reported effects on disease progression, outcomes of treatment and viral loads (Suppiah et al. 2009; Tanaka et al. 2009; Thomas

et al. 2009) and associations with potential drivers of sequence diversification in this data set (Ansari et al. 2019).

Multivariate analysis indeed demonstrated that *IFNL4* SNP rs12979860 was the strongest predictor of MFED values, followed by viral load but no other significant host predictive factors other than minor effects of patient age. The combination of factors accounted for ~25% of the variability in MFED values (multiple $R = 0.267$). Mean MFED values were significantly different between host *IFNL4* genotypes in both HCV genotypes 1 and 3 (Fig. 6). The direction and effect sizes were also consistent with approximately a 0.4% increase in MFED values from CC to TT genotypes in both HCV genotypes. *IFNL4* genotypes also showed substantial effects on viral diversity and UpA dinucleotide frequencies (Figs. 7, 8). There were generally highly significant differences between subjects with CC, CT, and TT alleles in the sequence divergence of their infecting strains from a reconstructed ancestral sequence of each genotype, both at nonsynonymous sites and synonymous sites for all three genotypes (Fig. 7).

The previously reported difference in UpA dinucleotide frequencies between *IFNL4* SNP genotypes in HCV genotype 3a (Ansari et al. 2019) was reproduced in the larger data sets used in the current study for all three genotypes, although there was little systematic difference in UpU frequencies between different alleles (Fig. 8). No differences in CpG frequencies were identified (data not shown).

To determine whether population geographic differences of HCV influenced the association between *IFNL4* SNP rs12979860 with MFED values and UpA frequencies, we reanalyzed associations in genotype 3a using the BOSON cohort, which includes study subjects from the UK, USA, Canada, Australia and New Zealand. As controls, we selected 500 SNPs from across the human genome that were frequency matched to *IFNL4* SNP rs12979860 (listed in Supplemental Table S2; Supplemental Data—avoiding X and Y chromosomes and SNPs within 200 kb from the *IFNL4* gene on chromosome 19). Linear regression was

TABLE 2. Analysis of host factors influencing MFED values by multivariate analysis[a]

Variable	Values/range	F-ratio	P-value
IFNL4 genotype	CC, CT, TT	9.688	8×10^{-5}
SVR	Yes, No	1.1	0.295
Race	White, Black, Asian[b]	0.51	0.801
Cirrhosis	Yes, No	0.315	0.575
Gender	Male, Female	0.075	0.784
Age	19–70 yr	4.036	0.045
Log VL	$10^{3.9}$–$10^{7.6}$ RNA/mL	8.7	0.003

[a]Analyzed by the General Linear Model program in the SYSTAT v13 package.
[b]Additional categories were indigenous North American and Indigenous Hawaiian and unknown.
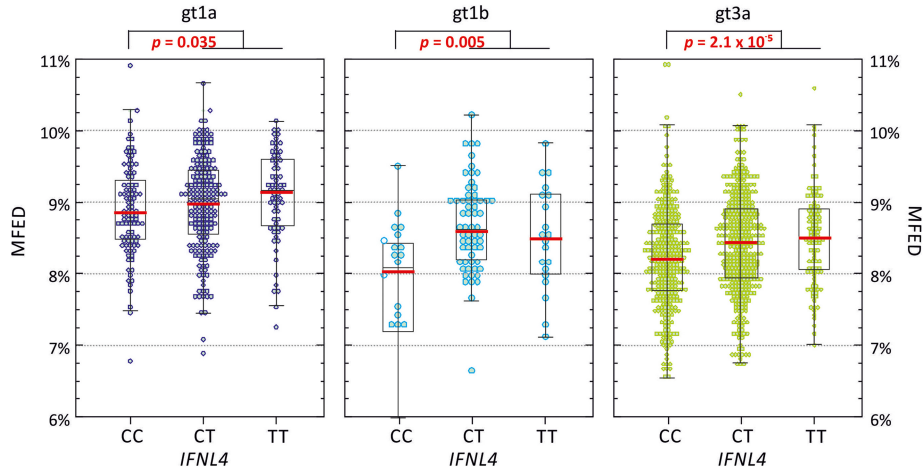
FIGURE 6. MFED values of gt1a, gt1b, and gt3a genomes recovered from study subjects with different *IFNL4* genotypes. MFED values for gt1a, gt1b, and gt3a full genome sequences divided into the *IFNL4* genetic background of the host. The box plots show (from the *top*): 2 SDs above mean, 1 SD above mean, mean (red bar), 1 SD below mean, and 2 SDs below mean. Differences in the distribution of MFED vales were calculated using the Kruskall–Wallace nonparametric test.

to test for associations between MFED and UpA frequency for each SNP and t-statistics for effects plotted for all 500 SNPs and rs12979860 (Fig. 9). We observed a normal distribution of *t*-statistics for the control SNPs, consistent with an absence of associations with MFED and UpA frequencies while accounting for possible effects of population structure. Contrastingly, the t-statistic for *IFNL4* SNP rs12979860 with MFED and UpA frequencies lies far outside the null distribution.

DISCUSSION

Visualization methods for RNA secondary structure

A diagrammatic representation of an individual stem–loop and its potential kissing loop/pseudoknot tertiary pairings represent the standard means for depicting RNA secondary and tertiary structures. Their utility, however, is limited as a means to incorporate information on RNA structural variability, such as different identities of the bases paired
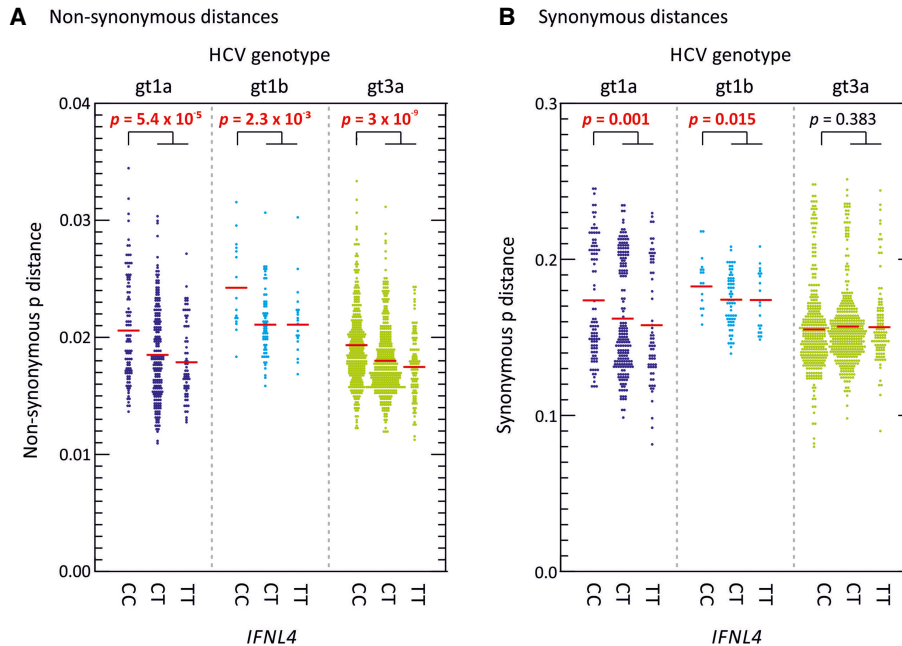
FIGURE 7. Nonsynonymous and synonymous site sequence divergence of polyprotein sequences of different HCV genotypes. Sequence distances between each study polyprotein sequence of gts 1a, 1b, and 3a from a reconstructed consensus (ancestral) sequence of each genotype at (*A*) nonsynonymous, and (*B*) synonymous sites. Red bars show mean values; differences in the distribution of MFED values were calculated using the Kruskall–Wallace test; significant values shown in red bold type.

FIGURE 8. Distributions of UpA and UpU frequencies. Observed to expected frequencies of (*A*) UpA, and (*B*) UpU dinucleotides targeted by RNaseL calculated for polyprotein seque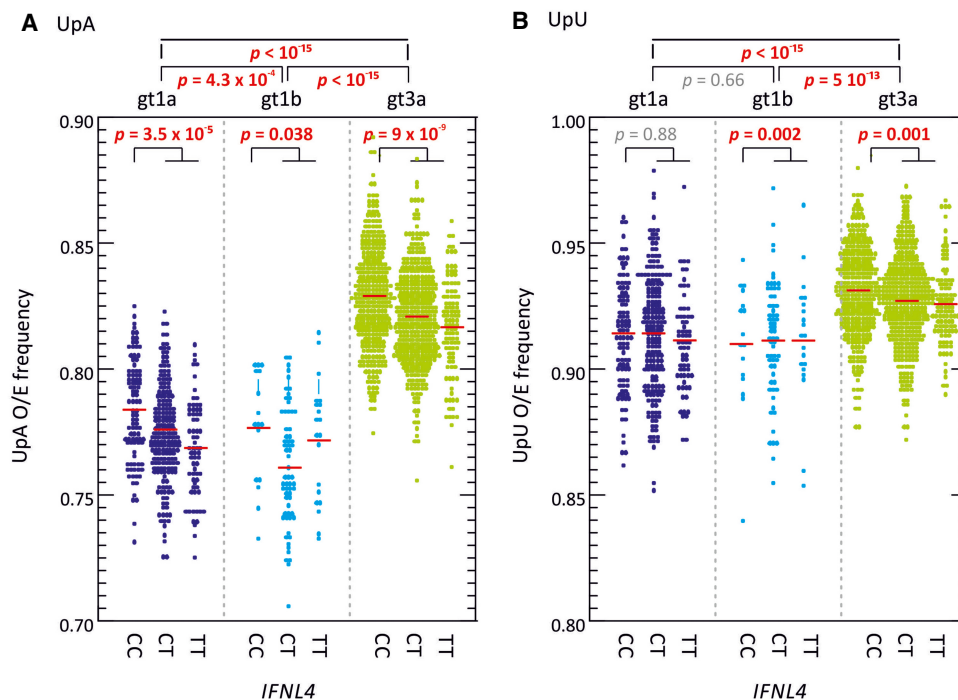nce of gts 1a, 1b, and 3a and different IFNL4 alleles. Red bars show mean values; differences in the distribution of O/E ratios were calculated using the Kruskall–Wallace test; significant values shown in red bold type.

and their relative positions within a structure. RNA structure diagrams become increasingly difficult to construct for longer sequences, such as whole virus genomes analyzed in the current study. To address these restrictions in our study of RNA structural diversity, we used a form of three-dimensional representation of multiple sequences in alignments spanning the lengths of viral genomes. The depiction of stem–loop tips at a standard height (0) highlights the principal structural elements of an RNA folding prediction for each sequence while the size of the stem–loop is depicted of depth. In the current study, we color-coded the StructureDist output but the three-dimensional coordinate file produced by the program can be used for alternative forms of representation, such as three-dimensional surface visualizations combined with object rotations.

The advantages of this form of whole-genome representation is apparent when comparing predicted RNA structures for HCV genotypes 1–3, other viruses with GORS and viruses with unstructured genomes (Fig. 2). For the latter, the plethora of arbitrary, short and variable pairings in alignments of unstructured RNA sequences appeared entirely distinct from the ordered sets of stem–loops visualized for those that possess sequence order-dependent pairings (high MFED values). The ability to display both pairing and conservation at each position in the genome provides a highly effective way to visualize which elements and structures are conserved across a data set and the ex-

tent to which RNA structures may systematically vary between genotypes or serotypes. The calculation of a heterogeneity score for pairing predictions between individual sequences clearly distinguishes between viruses with unstructured and structured genomes (Fig. 5).

The accuracy of the depictions of RNA structure in contour plots is obviously only as good as the underlying algorithm used to predict RNA structure pairings. The analysis described in the current study used RNAFold, a widely used and validated structure prediction method based upon folding energy minimization (Lorenz et al. 2011). The underlying algorithm outperforms other energy minimization algorithms in terms of speed and accuracy for structure prediction (Lorenz et al. 2011) and is able to generate numeric characteristics of RNA folds, such as MFE values that are used in MFED calculations (Fig. 1). For contour plots, the accuracy of the predictions was increased through generating an ensemble of suboptimal folds for each sequence fragment (Wuchty et al. 1999) and deriving a consensus prediction based on a simple majority rule (>50%). Sites with multiple conflicting predictions, typically those in regions without organized RNA folding, were therefore excluded from the plots and sharpen the differentiation between structured and unstructured regions (Fig. 4). The calling framework in SSE can however be adapted to future bioinformatic developments in RNA structure prediction, including the use of algorithms that address the knotty problem of
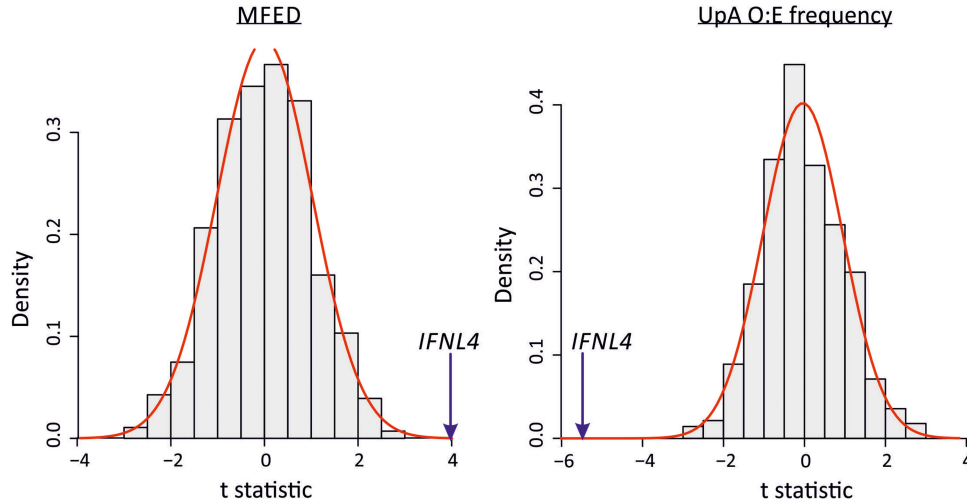
FIGURE 9. Association of MFED values and UpA frequencies with the *IFNL4* SNP rs12979860. The distribution of the *t*-statistics for 500 SNPs from across the genome frequency matched to the *IFNL4* SNP rs12979860 (SNPs listed in Supplemental Table S2; Supplemental Data). The red line shows the fitted normal distribution to the data. The *t*-statistic is from a linear regression where the dependent variable is the MFED or UpA frequency and the independent variable is the SNP with a dominant genetic model (coded the same as *IFNL4* SNP rs12979860 CC vs. non-CC genotypes). The blue arrows indicate the *t*-statistic for the *IFNL4* SNP rs12979860 CC vs. non-CC genotypes.

tertiary structure prediction (e.g., Jabbari et al. 2018; El Fatmi et al. 2019; Singh et al. 2019). It is entirely possible, for example, that pseuoknot or kissing loop interactions may play further stabilizing roles in GORS-associated RNA structures. The perplexing spread of MFED values based only on secondary structure-based energy calculations within individual HCV subtypes (Fig. 1B) or genotypes (Supplemental Fig. S4; Supplemental Data) might conceivably be reconciled if additional tertiary elements involved in RNA structure formation could be incorporated in MFE calculations.

RNA structure plasticity

This exploratory study investigated the extent and conservation of RNA structure formation in different genotypes of HCV. In areas of the coding sequences where conserved pairings were predicted between genotypes, such as in the core and NS5B regions, these were fully consistent with those previously determined by nuclease mapping, SHAPE analysis and functional studies (Tuplin et al. 2004; You et al. 2004; McMullan et al. 2007; Diviney et al. 2008; Mauger et al. 2015; Pirakitikulr et al. 2016). The contour plots, however, revealed far more RNA structure that differed between genotypes throughout the E1–NS5B regions. Most RNA structures were indeed unique to individual genotypes of HCV (Fig. 2A) that otherwise shared only in their elevated MFED scores as a marker of RNA structure formation (Fig. 1B). These findings are consistent with the evident differences in RNA structures revealed by SHAPE mapping of genomes of genotypes 1a, 2a, and 3a (Mauger et al. 2015), and indeed visually apparent from

the match between contour plot representations of the published SHAPE structures and RNAFold predictions for these genotypes (Fig. 3A). HCV therefore demonstrates considerable plasticity in the nature of its RNA pairings between subtypes and genotypes in most of the genome. This suggests that it is simply folding rather than the actual topography and potential interactions with cellular or viral RNA structures or proteins that is functionally important.

The highly variable structures formed by different genotypes visualized in the contour plots (Fig. 2A) account for the previous difficulties in mapping and functionally characterizing RNA structures in HCV. Indeed, few of the RNA structures investigated functionally after their identification by SHAPE were conserved across genotypes and even fewer possessed obvious replication functions. For example, effects on virus replication kinetics were modest or absent in mutants with disrupted base pairings in SL783 (core gene), SL1412 (E1), SL6038 (NS4A), and SL8001 (NS5B) in the JFH genotype 2a strain (Pirakitikulr et al. 2016). Similarly, disruption of J750 (core) and J8640 (NS5B) showed no effect on the replication of Jc1, while less than one log reductions were observed on disruption of J7880 (NS5B) and J8880 (NS5B) (Mauger et al. 2015). Systematic, large-scale mutagenesis of 17 consecutive sequence segments showed mild to modest replication effects only in the two final segments spanning positions 8441–8767 and 8768–9087 upstream of the CRE (SL9266) and an absence of cell culture functional elements in the rest of the coding region (Chu et al. 2013). These effects are collectively different from the previously documented lethal phenotype associated with disruption

of the HCV CRE (SL9266 or NS5B 3.2) on replication that is conserved across all genotypes.

The diversity of predicted RNA structure elements between HCV variants (Fig. 2A; Supplemental Fig. S3; Supplemental Data) was mirrored by FMDV. Its genome similarly shows elevated MFED scores across its genome (Supplemental Fig. S4; Supplemental Data) despite the evident differences in RNA structures formed by different serotypes across its more variable structural gene region. Collectively, these findings challenge the prevailing paradigm of viral RNA structures being discrete elements, conserved in base-pairings with defined functions and being highly evolutionarily stable. GORS, based on the current and previous analyses, differs in all of these characteristics, being pervasive throughout the genome, variable in pairings, likely mediating a general conformational effect on interactions with the cell and being highly evolutionarily plastic. While it has been argued that many or most of RNA structures predicted in the coding region of Jc1 (genotype 2a) strain may play replication or regulatory roles (Pirakitikulr et al. 2016), it seems difficult to imagine how any form of equivalence in functional properties could be maintained in other genotypes with quite radically different RNA structural organizations.

The evolution of RNA structure

The timescale for the evolution of different HCV genotypes and subtypes is currently uncertain. However, its estimated nucleotide substitution rate of $5–10 \times 10^{-4}$ substitutions per site per year (Markov et al. 2009; Gray et al. 2011) indicates that diversification of subtypes within a genotype has occurred over many hundreds of years; for example, the various genotype 2 subtypes have been proposed to have diverged in Guinea–Bissau in 1470 (range 1414–1582) (Markov et al. 2009). The eight HCV genotypes may have originated proportionately earlier (Smith et al. 1997). Given the evident differences between contour plots for HCV 1a, 1b, 2a, and 3a (Fig. 3), HCV has evidently largely remodeled its RNA secondary structure throughout the coding region over this period. At the same time, HCV has presumably managed to preserve sequence order-dependent RNA structure in its evolutionary intermediates, given the universal presence of elevated MFED values in all HCV genotypes (Fig. 1A; Supplemental Fig. S4, Supplemental Data). On a more immediate evolutionary timescale, we have obtained evidence for the evolution of RNA structure over the course of infection. The emergence of systematic differences in MFED values between study subjects with different *IFNL4* genotypes (Fig. 7) indicates effects of interferon lambda expression on genome configurations of genotypes 1 and 3 that could only have arisen over the perhaps 20–40 yr course of infection within each individual.

A link between GORS and cellular responses to infection was previously inferred from the evident association between its presence and virus persistence (Simmonds et al. 2004; Davis et al. 2008). It has been proposed that RNA structure formation may assist in the evasion of immune recognition of genomic RNA by RNaseL and PKR during persistent infections (Mauger et al. 2015). RNaseL is activated by 2′–5′ oligoadenylate molecules (2–5A) produced by a range of oligoadenylate synthetases (OASs) and subsequently targets single stranded RNA sequences for cleavage at UpA and UpU dinucleotide sites (Wreschner et al. 1981; Han and Barton 2002). The mammalian OAS/RNaseL plays a major role in the control of HCV and other RNA virus infections (Han and Barton 2002; Silverman 2007; Chakrabarti et al. 2011; Kristiansen et al. 2011; Li et al. 2016) and may significantly influence the outcomes of HCV infections and their responsiveness to treatment. Differences between RNaseL sensitivity between HCV genotypes and associated differences in UpA frequencies (Fig. 8; Washenberger et al. 2007) and RNA structure formation (Fig. 1B) are not inconsistent with this association.

In unraveling these factors, it has been found by several groups that intrinsic levels of ISG expression are greater in genotype 1 infections than in genotype 2 or 3 (Chen et al. 2010; Holmes et al. 2015; d'Avigdor et al. 2019). The internal environment in genotype 1-infected hepatocytes may be consequently more hostile for HCV replication. Among the various antiviral pathways activated by IFN, RNaseL potently restricts RNA virus replication through cleavage of viral RNA genomic sequences at UpA and UpU dinucleotide sites. Significantly, HCV genotype 1 shows a markedly greater suppression of UpA frequencies than genotypes 2 and 3 and consequent alterations in codon usage (Han and Barton 2002; Washenberger et al. 2007) (also evident in Fig. 8); this has been proposed by the authors as a necessary adaptive change to minimize the antiviral activity of RNaseL. The greater degree of folding of RNA in genotype 1 (mean MFED values of 9.0%, compared to 8.4% in genotype 3 [Fig. 1B]) may similarly reflect an adaptive change to reduce the frequencies of UpA/UpU dinucleotides in single stranded RNA contexts that can be targeted by RNaseL.

The same pattern of linked variables is mirrored on a smaller scale between individuals with different *IFNL4* genotypes. The greater degree of RNA structure formation in those with CT and TT genotypes (Fig. 1B) and the reduced frequencies of UpA dinucleotides (Fig. 8; Ansari et al. 2019) represents a microcosm of the larger virus genotype effects. They may reflect the much shorter term outcomes of more effective virus control mediated through RNaseL and other ISGs consequent to IFNλ4 expression in the liver during the course of chronic infection (Dill et al. 2011). Extending the model further, the reported evidence for less divergence of HCV genotype 3 sequences in those

with CT or TT genotypes (Ansari et al. 2019) which was also reflected in genotypes 1a and 1b (Fig. 7) may reflect the greater selection strength operating on viruses in cells with enhanced innate immune control. The finding in the current study that diversity at synonymous sites is also more restricted in non-CC individuals indicates a selection pressure potentially operating at the RNA level rather than on the viral proteome as previously suggested (Ansari et al. 2019). The observation in the latter study for associations at synonymous positions within several codons with *IFNL4* SNP rs12979860 provides further evidence for selection at the RNA level. Its focus on genotype 3a, where the differences in synonymous variability between CC and non-CC individuals are much less apparent than in genotypes 1a and 1b (Fig. 7), potentially accounts for the differences in conclusions reached. RNA structure requirements and maintaining low frequencies of UpA dinucleotides may indeed place quite different selection pressures of HCV infection in different *IFNL4* backgrounds at both nonsynonymous and synonymous sites. More broadly, the cellular host response may therefore be a potent factor influencing the course of evolution of HCV within an infected individual.

To conclude, we have shown that RNA secondary structure in HCV genomes pervades whole genomes of HCV and other viruses that establish persistent infections. GORS not only shows substantial changes in response to the host environment mediated through the *IFNL4* polymorphism, but it has also been entirely remodeled in most of the genome over the longer period of its divergence into different subtypes and genotypes (Fig. 2). In contrast to discrete, conserved RNA structural elements used by RNA viruses for replication and translation functions, GORS is pervasive and profoundly adaptive over even relatively short evolutionary periods. While the current study findings support the previously suggested link between RNA folding and RNaseL susceptibility, the broader underlying reasons for virus genomes becoming structured in this way require considerable further investigation.

MATERIALS AND METHODS

Sequence data sets

HCV complete polyprotein sequences were obtained from previously recruited cohorts (BOSON, the Early Access Programme [EAP], STOP-HCV-1 and the UK cirrhosis study) (Foster et al. 2015; McLauchlan et al. 2017; Fawsitt et al. 2019). Patients from the BOSON cohort originated from Australia, Canada, New Zealand, United Kingdom and United States, while the remaining cohorts were of UK-only origin. The BOSON study was conducted in accordance with the International Conference on Harmonization Good Clinical Practice Guidelines and the Declaration of Helsinki (clinical trial registration number: NCT01962441). The EAP study conformed to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by the institution's human research committee. The EAP patients were enrolled by consent into the HCV Research UK registry. Ethics approval for HCV Research UK was given by NRES Committee East Midlands—Derby 1 (Research Ethics Committee reference 11/EM/0314). A subset of genotype 3a subjects with self-reported white ancestry infected with HCV genotype 3a for which we had obtained both host genome-wide SNP data was selected for genome wide association study. Duration of infections were not determined for the study subjects although the cohorts varied in their progression and disease severity, ranging from mild (STOP-HCV-1, BOSON) to severe (EAP, cirrhosis cohort).

Available coding complete genome sequences of other viruses (FMDV serotype 0 [FMDV-O], HPgV-1, HPeV-3, JEV, MNV, and EV-A71) were downloaded from Genbank in 2019; redundant sequences were removed by filtering out sequences showing <1% nucleotide sequence from others in the data set (accession numbers of the selected sequences listed in Supplemental Table S1; Supplemental Data).

RNA structure prediction

Minimum folding energies were predicted using the RNAFold.exe program in the RNAFold package, version 2.4.2 (Lorenz et al. 2011) using with default parameters (37°C, allow isolated base pairs, linear sequence, unlimited pairing span) using sequential 240 base sequence fragments incrementing by 12 bases between fragments. Contour plots were generated from ensemble RNA structure predictions (Wuchty et al. 1999) of sequential 1600 base fragments of the alignment incrementing by 400 bases between fragments using the program SubOpt.exe, with the following additional parameters—all suboptimal structures, maximum 50 analyzed. Pairing predictions supported by >50% of suboptimal structures were used in the consensus contour plot; predictions from sites without a majority prediction were discarded. Folding heterogeneity was expressed as the difference in folding depth between different sequences at each position in the alignment; low values therefore correspond to sites with concordant pairing predictions. Analyses using RNAFold.exe and SubOpt.exe were invoked through the programs Folding Energy Scan and StructureDist in the SSE package version 1.4 (Simmonds 2012) (http://www.virus-evolution.org/Downloads/Software/). StructureDist in version 1.4 of SSE has been extended by the calculation of folding depths used for contour plotting. Sequence distances were calculated in the SSE package.

Host genetics

Host genetics information was based on a host genome-wide genotyping platform (Affymetrix UK Biobank Chip) and other demographic data collected for previous studies (Ansari et al. 2017, 2019).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

Members of the STOP-HCV Consortium

J. Ball,[7] D. Bonsall,[8] D. Brainard,[9] G. Burgess,[10] J. Dillon,[11] G. Foster,[12] C. Gore,[13] N. Guha,[14] R. Halford,[15] K. Whitby,[16] C. Holmes,[17] A. Howe,[18] E. Hudson,[19] S. Hutchinson,[20] S. Khakoo,[21] P. Klenerman,[22] N. Martin,[23] B. Massetto,[24] T. Mbisa,[25] J. McHutchison,[26] J. McKeating,[27] A. Miners,[28] A. Murray,[29] P. Shaw,[30] C. Spencer,[31] E. Thomson,[32] P. Vickerman,[33] and N. Zitzmann[34]

ACKNOWLEDGMENTS

REFERENCES

Al Sunaidi M, Williams CH, Hughes PJ, Schnurr DP, Stanway G. 2007. Analysis of a new human parechovirus allows the definition of parechovirus types and the identification of RNA structural domains. *J Virol* **81:** 1013–1021. doi:10.1128/JVI.00584-06

[7]University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, United Kingdom
[8]University of Oxford, Peter Medawar Building for Pathogen Research, Oxford OX1 3SY, United Kingdom
[9]Gilead Sciences, Inc., Foster City, California 94404, USA
[10]Conatus Pharmaceuticals, Inc., San Diego, California 92127, USA
[11]University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, United Kingdom
[12]Queen Mary's University of London, London E1 4AT, United Kingdom
[13]Hepatitis C Trust, London SE1 3YD, United Kingdom
[14]University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, United Kingdom
[15]Hepatitis C Trust, London SE1 3YD, United Kingdom
[16]Gilead Sciences, Inc., Uxbridge, Middlesex UB11 1AF, United Kingdom
[17]University of Oxford, Oxford OX1 3LB, United Kingdom
[18]BC Centre for Excellence in HIV/AIDS, St Paul's Hospital, Vancouver, British Columbia, Canada V6Z 1Y6
[19]University of Oxford, Peter Medawar Building for Pathogen Research, Oxford OX1 3SY, United Kingdom
[20]Glasgow Caledonian University, Glasgow G4 0BA, Scotland, United Kingdom
[21]University of Southampton, Southampton SO17 1BJ, United Kingdom
[22]University of Oxford, Peter Medawar Building for Pathogen Research, Oxford OX1 3SY, United Kingdom
[23]UC San Diego, La Jolla, California 92093-0507, USA
[24]Gilead Sciences, Inc., Foster City, California 94404, USA
[25]Public Health England, London NW9 5EQ, United Kingdom
[26]Gilead Sciences, Inc., Foster City, California 94404, USA
[27]University of Birmingham, Centre for Human Virology, Edgbaston, Birmingham B15 2TT, United Kingdom
[28]London School of Hygiene and Tropical Medicine, London WC1H 9SH, United Kingdom
[29]OncImmune Limited, Clinical Sciences Building, Nottingham City Hospital, Nottingham NG5 1PB, United Kingdom
[30]Merck & Co., Inc., Kenilworth, New Jersey 07033, USA
[31]University of Oxford, Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, United Kingdom
[32]University of Glasgow, MRC-CVR, Glasgow G61 1QH, United Kingdom
[33]University of Bristol, Clifton BS8 2BN, United Kingdom
[34]University of Oxford, Oxford OX1 3QU, United Kingdom

Ansari MA, Pedergnana V, LCI C, Magri A, Von Delft A, Bonsall D, Chaturvedi N, Bartha I, Smith D, Nicholson G, et al. 2017. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet* **49:** 666–673. doi:10.1038/ng.3835

Ansari MA, Aranday-Cortes E, Ip CL, da Silva Filipe A, Lau SH, Bamford C, Bonsall D, Trebes A, Piazza P, Sreenu V, et al. 2019. Interferon lambda 4 impacts the genetic diversity of hepatitis C virus. *Elife* **8:** e42463. doi:10.7554/eLife.42463

Baechlein C, Fischer N, Grundhoff A, Alawi M, Indenbirken D, Postel A, Baron AL, Offinger J, Becker K, Beineke A, et al. 2015. Identification of a novel hepacivirus in domestic cattle from Germany. *J Virol* **89:** 7007–7015. doi:10.1128/JVI.00534-15

Baechlein C, Baron AL, Meyer D, Gorriz-Martin L, Pfankuche VM, Baumgartner W, Polywka S, Peine S, Fischer N, Rehage J, et al. 2019. Further characterization of bovine hepacivirus: antibody response, course of infection, and host tropism. *Transbound Emerg Dis* **66:** 195–206. doi:10.1111/tbed.12999

Beames B, Chavez D, Lanford RE. 2001. GB virus B as a model for hepatitis C virus. *ILAR J* **42:** 152–160. doi:10.1093/ilar.42.2.152

Bukh J, Apgar CL, Govindarajan S, Purcell RH. 2001. Host range studies of GB virus-B hepatitis agent, the closest relative of hepatitis C virus, in New World monkeys and chimpanzees. *J Med Virol* **65:** 694–697. doi:10.1002/jmv.2092

Chakrabarti A, Jha BK, Silverman RH. 2011. New insights into the role of RNase L in innate immunity. *J Interferon Cytokine Res* **31:** 49–57. doi:10.1089/jir.2010.0120

Chen L, Borozan I, Sun J, Guindi M, Fischer S, Feld J, Anand N, Heathcote J, Edwards AM, McGilvray ID. 2010. Cell-type specific gene expression signature in liver underlies response to interferon therapy in chronic hepatitis C infection. *Gastroenterology* **138:** 1123–1133.e1121-1123. doi:10.1053/j.gastro.2009.10.046

Chu D, Ren S, Hu S, Wang WG, Subramanian A, Contreras D, Kanagavel V, Chung E, Ko J, Amirtham Jacob Appadorai RS, et al. 2013. Systematic analysis of enhancer and critical *cis*-acting RNA elements in the protein-encoding region of the hepatitis C virus genome. *J Virol* **87:** 5678–5696. doi:10.1128/JVI.00840-12

Condy JB, Hedger RS, Hamblin C, Barnett IT. 1985. The duration of the foot-and-mouth disease virus carrier state in African buffalo (i) in the individual animal and (ii) in a free-living herd. *Comp ImmunolMicrobiolInfect Dis* **8:** 259–265. doi:10.1016/0147-9571(85)90004-9

Cortey M, Ferretti L, Perez-Martin E, Zhang F, de Klerk-Lorist LM, Scott K, Freimanis G, Seago J, Ribeca P, van Schalkwyk L, et al. 2019. Persistent infection of African buffalo (*Syncerus caffer*) with foot-and-mouth disease virus: limited viral evolution and no evidence of antibody neutralization escape. *J Virol* **93:** e00563-19. doi:10.1128/JVI.00563-19

Coyne KP, Dawson S, Radford AD, Cripps PJ, Porter CJ, McCracken CM, Gaskell RM. 2006. Long-term analysis of feline calicivirus prevalence and viral shedding patterns in naturally infected colonies of domestic cats. *Vet Microbiol* **118:** 12–25. doi:10.1016/j.vetmic.2006.06.026

d'Avigdor WMH, Budzinska MA, Lee M, Lam R, Kench J, Stapelberg M, McLennan SV, Farrell G, George J, McCaughan GW, et al. 2019. Virus genotype-dependent transcriptional alterations in lipid metabolism and inflammation pathways in the hepatitis C virus-infected liver. *Sci Rep* **9:** 10596. doi:10.1038/s41598-019-46664-0

Davis M, Sagan S, Pezacki J, Evans DJ, Simmonds P. 2008. Bioinformatic and physical characterisation of genome-scale ordered RNA structure (GORS) in mammalian RNA viruses. *J Virol* **82:** 11824–11836. doi:10.1128/JVI.01078-08

Dill MT, Duong FH, Vogt JE, Bibert S, Bochud PY, Terracciano L, Papassotiropoulos A, Roth V, Heim MH. 2011. Interferon-induced gene expression is a stronger predictor of treatment response than IL28B genotype in patients with hepatitis C. *Gastroenterology* **140:** 1021–1031. doi:10.1053/j.gastro.2010.11.039

Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, Evans DJ. 2008. A hepatitis C virus *cis*-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B. *J Virol* **82:** 9008–9022. doi:10.1128/JVI .02326-07

El Fatmi A, Bekri MA, Benhlima S. 2019. RNAknot: a new algorithm for RNA secondary structure prediction based on genetic algorithm and GRASP method. *J Bioinform Comput Biol* **17:** 1950031. doi:10.1142/S0219720019500318

Fawsitt CG, Vickerman P, Cooke G, Welton NJ. 2019. A cost-effectiveness analysis of shortened direct-acting antiviral treatment in genotype 1 noncirrhotic treatment-naive patients with chronic hepatitis C virus. *Value Health* **22:** 693–703. doi:10.1016/j.jval .2018.12.011

Forrester NL, Boag B, Moss SR, Turner SL, Trout RC, White PJ, Hudson PJ, Gould EA. 2003. Long-term survival of New Zealand rabbit haemorrhagic disease virus RNA in wild rabbits, revealed by RT-PCR and phylogenetic analysis. *J Gen Virol* **84:** 3079–3086. doi:10.1099/vir.0.19213-0

Foster GR, Pianko S, Brown A, Forton D, Nahass RG, George J, Barnes E, Brainard DM, Massetto B, Lin M, et al. 2015. Efficacy of sofosbuvir plus ribavirin with or without peginterferon-alfa in patients with hepatitis C virus genotype 3 infection and treatment-experienced patients with cirrhosis and hepatitis C virus genotype 2 infection. *Gastroenterology* **149:** 1462–1470. doi:10.1053/j .gastro.2015.07.043

Goodfellow IG, Kerrigan D, Evans DJ. 2003. Structure and function of the poliovirus *cis*-acting replication element (CRE). *RNA* **9:** 124–137. doi:10.1261/rna.2950603

Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol Biol* **11:** 131. doi:10.1186/1471-2148-11-131

Han JQ, Barton DJ. 2002. Activation and evasion of the antiviral 2′–5′ oligoadenylate synthetase/ribonuclease L pathway by hepatitis C virus mRNA. *RNA* **8:** 512–525. doi:10.1017/S135583820 2020617

Holmes JA, Congiu M, Bonanzinga S, Sandhu MK, Kia YH, Bell SJ, Nguyen T, Iser DM, Visvanathan K, Sievert W, et al. 2015. The relationships between IFNL4 genotype, intrahepatic interferon-stimulated gene expression and interferon treatment response differs in HCV-1 compared with HCV-3. *Aliment Pharmacol Ther* **42:** 296–306. doi:10.1111/apt.13263

Hoofnagle JH. 2002. Course and outcome of hepatitis C. *Hepatology* **36:** S21–S29. doi:10.1053/jhep.2002.36227

Hsu CC, Riley LK, Wills HM, Livingston RS. 2006. Persistent infection with and serologic cross-reactivity of three novel murine noroviruses. *Comp Med* **56:** 247–251.

Jabbari H, Wark I, Montemagno C, Will S. 2018. Knotty: efficient and accurate prediction of complex RNA pseudoknot structures. *Bioinformatics* **34:** 3849–3856. doi:10.1093/bioinformatics/ bty420

Kristiansen H, Gad HH, Eskildsen-Larsen S, Despres P, Hartmann R. 2011. The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *J Interferon Cytokine Res* **31:** 41–47. doi:10.1089/jir.2010.0107

Li Y, Banerjee S, Wang Y, Goldstein SA, Dong B, Gaughan C, Silverman RH, Weiss SR. 2016. Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. *Proc Natl Acad Sci* **113:** 2241–2246. doi:10.1073/pnas .1519657113

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6:** 26. doi:10.1186/1748-7188-6-26

Markov PV, Pepin J, Frost E, Deslandes S, Labbe AC, Pybus OG. 2009. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol* **90:** 2086–2096. doi:10.1099/vir.0 .011569-0

Mauger DM, Golden M, Yamane D, Williford S, Lemon SM, Martin DP, Weeks KM. 2015. Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc Natl Acad Sci* **112:** 3692–3697. doi:10 .1073/pnas.1416266112

McFadden N, Arias A, Dry I, Bailey D, Witteveldt J, Evans DJ, Goodfellow I, Simmonds P. 2013. Influence of genome-scale RNA structure disruption on the replication of murine norovirus-similar replication kinetics in cell culture but attenuation of viral fitness in vivo. *Nucleic Acids Res* **41:** 6316–6331. doi:10.1093/nar/ gkt334

McLauchlan J, Innes H, Dillon JF, Foster G, Holtham E, McDonald S, Wilkes B, Hutchinson SJ, Irving WL. 2017. Cohort profile: the hepatitis C virus (HCV) research UK clinical database and biobank. *Int J Epidemiol* **46:** 1391–1391h. doi:10.1093/ije/ dyw362

McMullan LK, Grakoui A, Evans MJ, Mihalik K, Puig M, Branch AD, Feinstone SM, Rice CM. 2007. Evidence for a functional RNA element in the hepatitis C virus core gene. *Proc Natl Acad Sci* **104:** 2879–2884. doi:10.1073/pnas.0611267104

Oldstone MB. 2009. Anatomy of viral persistence. *PLoS Pathog* **5:** e1000523. doi:10.1371/journal.ppat.1000523

Pfaender S, Cavalleri JM, Walter S, Doerrbecker J, Campana B, Brown RJ, Burbelo PD, Postel A, Hahn K, Anggakusuma, et al. 2015. Clinical course of infection and viral tissue tropism of hepatitis C virus-like nonprimate hepaciviruses in horses. *Hepatology* **61:** 447–459. doi:10.1002/hep.27440

Pirakitikulr N, Kohlway A, Lindenbach BD, Pyle AM. 2016. The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol Cell* **62:** 111–120. doi:10.1016/j.molcel.2016.01 .024

Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, Rice CM. 2009. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457:** 882–886. doi:10.1038/nature07684

Polaris Observatory HCV Collaborators. 2017. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. *Lancet Gastroenterol Hepatol* **2:** 161–176. doi:10 .1016/S2468-1253(16)30181-9

Ramsay JD, Evanoff R, Wilkinson TE Jr, Divers TJ, Knowles DP, Mealey RH. 2015. Experimental transmission of equine hepacivirus in horses as a model for hepatitis C virus. *Hepatology* **61:** 1533–1546. doi:10.1002/hep.27689

Randall RE, Griffin DE. 2017. Within host RNA virus persistence: mechanisms and consequences. *Curr Opin Virol* **23:** 35–42. doi:10 .1016/j.coviro.2017.03.001

Rieder E, Paul AV, Kim DW, van Boom JH, Wimmer E. 2000. Genetic and biochemical studies of poliovirus *cis*-acting replication element *cre* in relation to VPg uridylylation. *J Virol* **74:** 10371–10380. doi:10.1128/JVI.74.22.10371-10380.2000

Silverman RH. 2007. Viral encounters with 2′,5′-oligoadenylate synthetase and RNase L during the interferon antiviral response. *J Virol* **81:** 12720–12729. doi:10.1128/JVI.01471-07

Simmonds P. 2012. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes* **5:** 50. doi:10.1186/1756-0500-5-50

Simmonds P, Tuplin A, Evans DJ. 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded

RNA viruses: implications for virus evolution and host persistence. *RNA* **10:** 1337–1351. doi:10.1261/rna.7640104

Simmonds P, Kakasiliotis I, Chaudhry Y, Evans DJ, Goodfellow IG. 2008. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res* **36:** 2530–2546. doi:10.1093/nar/gkn096

Singh J, Hanson J, Paliwal K, Zhou Y. 2019. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun* **10:** 5407. doi:10.1038/s41467-019-13395-9

Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, Yap PL, Simmonds P. 1997. The origin of hepatitis C virus genotypes. *J Gen Virol* **78:** 321–328. doi:10.1099/0022-1317-78-2-321

Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, Bassendine M, Spengler U, Dore GJ, Powell E, et al. 2009. IL28B is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. *Nat Genet* **41:** 1100–1104. doi:10.1038/ng.447

Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, et al. 2009. Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. *Nat Genet* **41:** 1105–1109. doi:10.1038/ng.449

Thackray LB, Wobus CE, Chachu KA, Liu B, Alegre ER, Henderson KS, Kelley ST, Virgin HW. 2007. Murine noroviruses comprising a single genogroup exhibit biological diversity despite limited sequence divergence. *J Virol* **81:** 10460–10473. doi:10.1128/JVI.00783-07

Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, Kidd J, Kidd K, Khakoo SI, Alexander G, et al. 2009. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* **461:** 798–801. doi:10.1038/nature08463

Thomson GR, Vosloo W, Esterhuysen JJ, Bengis RG. 1992. Maintenance of foot and mouth disease viruses in buffalo (Syncerus caffer Sparrman, 1779) in southern Africa. *Rev Sci Tech* **11:** 1097–1107. doi:10.20506/rst.11.4.646

Trivedi S, Murthy S, Sharma H, Hartlage AS, Kumar A, Gadi S, Simmonds P, Chauhan LV, Scheel TKH, Billerbeck E, et al. 2017. Viral persistence, liver disease and host response in Hepatitis C-like virus rat model. *Hepatology* **68:** 435–448. doi:10.1002/hep.29494

Tuplin A, Evans DJ, Simmonds P. 2004. Detailed mapping of RNA secondary structures in core and NS5B coding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* **85:** 3037–3047. doi:10.1099/vir.0.80141-0

Washenberger CL, Han JQ, Kechris KJ, Jha BK, Silverman RH, Barton DJ. 2007. Hepatitis C virus RNA: dinucleotide frequencies and cleavage by RNase L. *Virus Res* **130:** 85–95. doi:10.1016/j.virusres.2007.05.020

Westbrook RH, Dusheiko G. 2014. Natural history of hepatitis C. *J Hepatol* **61:** S58–S68. doi:10.1016/j.jhep.2014.07.012

Wreschner DH, McCauley JW, Skehel JJ, Kerr IM. 1981. Interferon action–sequence specificity of the ppp(A2′p)$_n$A-dependent ribonuclease. *Nature* **289:** 414–417. doi:10.1038/289414a0

Wuchty S, Fontana W, Hofacker IL, Schuster P. 1999. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49:** 145–165. doi:10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G

You S, Stump DD, Branch AD, Rice CM. 2004. A *cis*-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J Virol* **78:** 1352–1366. doi:10.1128/JVI.78.3.1352-1366.2004