Research paper

Exploring resistance pathways for first-generation NS3/4A protease inhibitors boceprevir and telaprevir using Bayesian network learning

Lize Cuypers [a,*], Pieter Libin [a,b], Yoeri Schrooten [a], Kristof Theys [a], Velia Chiara Di Maio [c], Valeria Cento [c], Maja M. Lunar [d], Frederik Nevens [e], Mario Poljak [d], Francesca Ceccherini-Silberstein [c], Ann Nowé [b], Kristel Van Laethem [a], Anne-Mieke Vandamme [a,f]

[a] KU Leuven, University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Herestraat 49, box 1040, 3000 Leuven, Belgium
[b] Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[c] Department of Experimental Medicine and Surgery, University of Rome "Tor Vergata", Rome, Italy
[d] Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
[e] University Hospitals Leuven, Department of Hepatology, Herestraat 49, 3000 Leuven, Belgium
[f] Center for Global Health and Tropical Medicine, Microbiology Unit, Institute for Hygiene and Tropical Medicine, University Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

ARTICLE INFO

ABSTRACT

Resistance-associated variants (RAVs) have been shown to influence treatment response to direct-acting antivirals (DAAs) and first generation NS3/4A protease inhibitors (PIs) in particular. Interpretation of hepatitis C virus (HCV) genotypic drug resistance remains a challenge, especially in patients who previously failed DAA therapy and need to be retreated with a second DAA based regimen. Bayesian network (BN) learning on HCV sequence data from patients treated with DAAs could provide insight in resistance pathways against PIs for HCV subtypes 1a and 1b, in a similar way as applied before for HIV. The publicly available 'Rega-BN' tool chain was developed to study associative analyses for various pathogens. Our first analysis, comparing sequences from PI-naïve and PI-experienced patients, determined that NS3 substitutions R155K and V36M arise with PI-exposure in HCV1a infected patients, and were defined as major and minor resistance-associated variants respectively. NS3 variant 174H was newly identified as potentially related to PI resistance. In a second analysis, NS3 sequences from PI-naïve patients who cleared the virus during PI therapy and from PI-naïve patients who failed PI therapy were compared, showing that NS3 baseline variant 67S predisposes to treatment-failure and variant 72I to treatment success. This approach has the potential to better characterize the role of more RAVs, if sufficient therapy annotated sequence data becomes available in curated public databases. In addition, polymorphisms present in baseline sequences that predispose patients to therapy failure can be identified using this approach.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The majority of hepatitis C virus (HCV) infected patients develops chronic hepatitis resulting in many people being at risk for liver-associated diseases (Bartosch et al., 2003). Worldwide, an estimated 80 million people are currently infected with HCV (Gower et al., 2014), about half of whom are not aware of their infection, hampering the desired scale-up of antiviral treatment. Since HCV manifests as seven genotypes (GTs) and >50 subtypes (Smith et al., 2014), developing a pan-genotypic drug proved to be challenging. However, concerted efforts led to the current generation of anti-HCV drugs with dramatically improved treatment success rates (Zeuzem et al., 2011; Cuypers et al., 2016b). As a result, highly efficient direct-acting antivirals (DAAs) were approved for HCV treatment starting from 2011, including NS3/4A protease inhibitors (PIs), NS5A inhibitors and NS5B polymerase inhibitors. Combination therapies currently available achieve sustained virological response (SVR) rates higher than 90–95% in all HCV genotypes (Welsch

et al., 2012; Foster et al., 2015; Zeuzem et al., 2015a). Unfortunately, for those that cannot clear their infection, viral drug resistance-associated variants (RAVs) are a threat, whether emerging during course of treatment, naturally occurring or acquired at infection (Chevaliez, 2011, Franco et al., 2014, Halfon and Locarnini, 2011, Romano et al., 2012, Wyles, 2013). Many uncertainties exist regarding the clinical impact of RAVs, making interpretation of HCV genotypic drug resistance a challenge. Some naturally occurring polymorphisms or treatment-associated substitutions have been shown to impact treatment response only in the context of specific HCV genotypes or subtypes (Barnard et al., 2013; Sarrazin et al., 2007; Sullivan et al., 2013; Susser et al., 2009). How polymorphisms are influencing therapy outcome depends also on the viral genetic sequence and drug context. Such knowledge is currently scarce, but urgently needed to establish proper HCV drug resistance interpretation algorithms that could help in the design of retreatment strategies. This is especially true in the case of NS5A RAVs, which tend to persist after failing therapy (Krishnan et al., 2015; Black et al., 2015; Dvory-Sobol et al., 2015).

The most abundant source of HCV genetic sequence data that can be used to investigate the relationship between amino acid variants and treatment success is for patients treated with the earliest DAAs, meaning PIs targeting NS3. The first 181 amino acids of the NS3 protein make up the enzyme serine protease that is responsible for cleaving the viral polyprotein, after activation by cofactor NS4A. Generations can be distinguished within the PI class based on binding properties with the catalytic triad of the NS3 protease (Salam and Akimitsu, 2013). The first-generation PIs are linear α-ketoamide compounds telaprevir and boceprevir, which covalently bind the enzyme and display only antiviral activity towards HCV genotype 1 (Raney et al., 2010). The second-generation PIs such as faldaprevir and simeprevir, bind the catalytic triad in a non-covalent manner, and show antiviral activity towards non-1 HCV genotypes too (De Luca et al., 2014). The most recent PIs feature a higher genetic barrier to resistance, as well as broader antiviral activity, e.g. paritaprevir and grazoprevir (Andreone et al., 2014; Cuypers et al., 2016b; Zeuzem et al., 2015a).

Associations between variants and treatment-exposure can be identified by the use of general linear regression analysis in a fast and easy way. However, basic statistics are not able to uncover in-depth multivariate associations and their dependencies, which is where graphical modeling techniques, such as Bayesian network (BN) learning, have been proven successful (Lee and Abbott, 2003). Each node in the network corresponds to a variable, and an arc between nodes encodes an unconditional dependence, which represents a direct influence between these variables. Bayesian network learning is a probabilistic model that describes statistical independencies between multiple variables (Pearl, 1998) by using a minimum number of arcs (Heckerman, 1999). During such a search, the benefit of adding an arc is evaluated against the additional cost of that arc (Deforche et al., 2006). In the context of HIV drug resistance (Deforche et al., 2006, 2007, 2008; Theys et al., 2010), resistance variants and epistatic interactions of resistance mutations and natural polymorphisms have been mapped, increasing the understanding of resistance pathways and prediction systems. Therefore, applying BN learning on HCV sequence data of patients treated with DAAs can increase the knowledge about epistatic interactions between amino acid variants, and their relation to exposure/response to treatment.

This proof-of-concept study will identify interactions between amino acid variants, drug-exposure and therapy response using a BN learning method that was implemented as a user-friendly tool chain. Therapy-annotated sequence data for patients treated with PIs were pooled, with the majority from the era when these PIs were used in triple therapy combinations with pegylated interferon-alpha (pegIFN-α) and ribavirin (RBV) (Zeuzem et al., 2011). Despite the fact that treatment regimens based on boceprevir or telaprevir are now contra-indicated for HCV due to their moderate antiviral activity, association with severe side effects (AASLD and EASL guidelines, Bacon et al., 2011,

Hézode et al., 2009) and low genetic barrier to resistance (Barnard et al., 2013; Sullivan et al., 2013), they still serve as a benchmark for the development of the newer generations of PIs. Mapping epistatic interactions between amino acids within NS3, and between amino acid variants and PI-exposure or therapy response will also be of value for other PIs, for which not enough virus sequence data is currently available.

2. Material and methods

2.1. Study population

NS3 protease sequences were gathered from Italy (Department of Experimental Medicine and Surgery of the University of Tor Vergata, Rome), Slovenia (University of Ljubljana) and Belgium (University Hospitals Leuven, Leuven). These three clinical databases were used in order to reduce potential bias from local epidemics. All the viral sequences were newly generated or obtained from samples taken from patients who were treated with a NS3/4A protease inhibitor based regimen. In total, 154 patients were included and for each patient one PI-naïve and/or one PI-experienced sequence was retained. The median age was 53 years (interquartile range (IQR): 45–63), and the majority of the patients were male (61%). The majority of patients were treated with first generation NS3/4A PIs telaprevir (52.8%) or boceprevir (31.6%), and rarely with second- or newer-generation PIs simeprevir, faldaprevir or paritaprevir (15.6%). All PIs were administered to the patients in the context of triple therapy with pegIFN-α and RBV, except for patients on triple DAA combination (paritaprevir boosted with ritonavir, combined with NS5A inhibitor ombitasvir and NS5B polymerase inhibitor dasabuvir).

HCV RNA viral load was quantified using commercial assays COBAS Ampliprep/TaqMan HCV Quantitative Test v2.0 (Roche Diagnostics, Pleasanton, USA) or Abbott Real Time HCV assay (Abbott Laboratories, Abbott Park, IL, USA). HCV genotype was determined by line probe assays (InnoLipa, Versant HCV Genotype 2.0 Assay, Siemens, Healthcare Diagnostic Inc., Tarrytown, NY, USA) according to the manufacturer's instructions. The UZ Leuven ethical committee approved this study (number ML9219 and ML10770). Samples from Italy were anonymized and research was not conducted within the context of a clinical trial, so this study was exempt from ethical approval (Italian law - art. 6 and art. 9, leg. 211/2003 and 196/2003). Similarly, in accordance with national legislation of the Republic of Slovenia, informed consent was not required for research using archival clinical samples if clinical samples are coded and tested anonymously.

2.2. Protease sequences

Viral sequences encoding the first 181 amino acids in the NS3 gene were determined by in house amplification and sequencing protocols for HCV1a and HCV1b separately (Cento et al., 2012). All sequences obtained in the context of this study, were submitted to Genbank (accession numbers KX852142-KX852220 and KX825921-KX825931). Viral sequences originating from Italian patients were published already (Cento et al., 2015a, 2015b).

Additional sequences from before start of therapy of PI-treated patients were selected from the public database GenBank (https://www.ncbi.nlm.nih.gov/), at least those annotated with HCV genotype, subsequent treatment regimen and therapy response, resulting in 319 additional sequences from patients that were still PI-naïve. The corresponding sequences of these patients after PI-experience, or any other PI-treated patients, were not available in Genbank. The sequence length of the 319 Genbank sequences was shorter than 181 amino acids, specifically 84 or 85 amino acids, resulting in the analysis of separate datasets that each focused on a different part of the NS3 protease gene.

2.3. Subtyping

HCV genotypes and subtypes were assigned using the Oxford HCV automated subtyping tool version 2 (Alcantara et al., 2009; De Oliveira et al., 2005), COMET (Struck et al., 2014) and by manual phylogenetic analysis using reference sequences from the Los Alamos HCV sequence database (Kuiken et al., 2008). Sequences were aligned against reference sequence H77 (NC_004102) using an *in-house* developed codon aware alignment tool (Cuypers et al., 2015). To improve the quality, alignments were manually edited in Seaview V4.0 (Gouy et al., 2010). Phylogenetic analysis was performed using a maximum-likelihood approach and GTR gamma as model of substitution with RAxML V8.0.20 (Stamatakis, 2014), generating 1000 bootstrap replicates to evaluate the tree robustness.

2.4. Selective pressure

Selective pressure was investigated using the fixed effects model (FEL) and the mixed effects model of evolution (MEME) methods implemented in HyPhy v2.2.1 (Murrell et al., 2012; Pond et al., 2005), which accommodates site-by-site variation. The latter also models variable dN/dS ratios across lineages at an individual site, starting from a maximum-likelihood inferred tree. The MEME method is more sensitive in detecting potential selective pressure occurring only along some branches, however only when a measurable proportion of lineages are experiencing non-synonymous evolution (Murrell et al., 2012). An amino acid position was considered to be under positive selective pressure if the ratio of non-synonymous and synonymous amino acid substitutions at that codon (dN/dS ratio) was >1 ($p < 0.05$). Amino acid positions with dN/dS <1, and p-value <0.05, were classified as positions under negative selective pressure. The number of sites under positive and negative selective pressure for sequences from PI-naïve, PI-experienced on-treatment and PI-experienced post-treatment patients, were compared for, using the two-tailed z-test.

2.5. Assignment of sequences according to treatment groups

Sequences were grouped based on exposure to a specific NS3/4A protease inhibitor (Table 1). All sequences allocated to the 'PI-experienced' group were from patients treated with boceprevir, telaprevir or faldaprevir since no sequences with simeprevir or paritaprevir exposure were available in our dataset. The sequence with a sample date closest to start of PI treatment was chosen for PI-naïve sequences. In cases where more than one PI-experienced sequence was available, the latest on-treatment sequence, or if not available, the first post-treatment sequence was selected. Sequences were excluded when the duration of HCV antiviral treatment during which the sample was taken, was <14 days (Cento et al., 2015a, 2015b). Table 2 summarizes more detailed information for all samples concerning the time of sampling compared to the start and end date of PI treatment. For the PI-naïve sequences, treatment regimen details are listed in Table 3, while the majority of strains from Genbank originated from patients who received telaprevir (83.4%) as antiviral treatment, and a minority that was treated with boceprevir (16.6%).

2.6. Known drug resistance-associated variants (RAVs)

RAVs were determined using the Geno2pheno [hcv] v2.0 tool (Kalaghatgi et al., 2016) and based on a study published by Lontok et al., 2015. The following NS3 amino acid positions were retained for analysis: 36, 41, 43, 54–56, 80, 87, 107, 117, 122, 132, 155, 156, 158, 168, 170, 174 and 175. While some of these variants may be naturally occurring variants, others arose under treatment and could be called resistance-associated substitutions (RASs) (Pawlotsky, 2016). We chose to use the term RAVs to cover both types of variants.

Table 1
PI-exposure of all 193 sequences from the 154 patients included in the study cohort. Listed per PI, samples were ordered based on sampling time (PI-naïve and PI-experienced, either on- or post-PI treatment).

| | Treatment-annotated sequences – 154 patients | | | |
| | PI-naïve | PI-experienced | | |
PI[a]	PI-naïve[b]	On-treatment	Post-treatment	Total
Boceprevir	36	5	20	61
Telaprevir	76	14	12	102
Simeprevir	3	1	0	4
Faldaprevir	12	0	4	16
Paritaprevir	10	0	0	10
Total	137	20	36	193

[a] All protease inhibitors were co-administrated with pegIFN-α + RBV, except for paritaprevir that was boosted with ritonavir and combined with ombitasvir and dasabuvir.
[b] For each PI naïve patient, the future PI treatment regimen and treatment response (SVR or failure) was known.

2.7. Identification of treatment-associated variants

Comparison of sequences from PI-naïve and PI-experienced patients was done cross-sectionally using sequence data of all 154 patients. A longitudinal approach was not feasible given the limited number of patients with both PI-naïve and PI-experienced sequences. This comparison was carried out using a tool-chain adapted from one developed by Deforche et al. (2006) (Libin, 2014). A Fisher's exact test was used to select amino acid changes that were significantly linked with PI-exposure (at significance level of 0.05), hereafter referred to as treatment-associated variants. These variants served as input for the BN learning procedure. In general, resistance-associated variants are expected to be associated to treatment, while treatment-associated variants do not always confer resistance (Theys et al., 2013).

2.8. Bayesian network learning

Treatment-associated variants (see Section 2.6) were complemented with all known resistance-associated variants (see Section 2.5), and served as input for the BN learning procedure. The first network focused on interactions between amino acid variants and PI-exposure comparing sequences from PI-naïve and PI-experienced patients. The goal of the second network was to explore natural RAVs in PI-naïve patients, potentially related with treatment outcome. The subsequent PI regimen and therapy response was known for all PI-naïve patients included in this network.

BN learning was performed using the B-course software (Myllymäki et al., 2002) using the simulated annealing algorithm and by maximizing the posterior probability of the model (Deforche et al., 2006). During such a search, the benefit of adding an arc is evaluated against the additional cost of that arc (Deforche et al., 2006). Dependencies were visualized in a directed acyclic graph and formed the qualitative component of the network (Theys et al., 2010). Each node corresponded to a variable, and an arc between nodes encoded an unconditional dependence, which represented a direct influence. Robustness of network features

Table 2
The selection of samples according to their timing with respect to the PI treatment period. For all samples, the total range, median and interquartile range (IQR) of the difference between the sampling dates and the start or end date of PI treatment, are summarized.

	PI-naïve - number of days before start of treatment	On-treatment – number of days on treatment	Post-treatment – number of days after stop of treatment
Total range	0–2431	15–274	5–369
Median	0	53	96
IQR	0–22	26–91	63–196

Table 3
PI treatment outcome of all 456 PI-naïve sequences.
Listed per subset of PI-naïve sequences, samples were ordered based on the specific PI they were starting treatment with, and the respective PI treatment outcome (PI-failure and PI treatment success). Three different subsets are described, depending on their coverage of the HCV protein NS3 (more details in the 'Results' Section 3.2): 1) coverage from NS3 amino acid positions 1 to 86; 2) a second subset of PI-naïve sequences with sequence information for NS3 positions 29 to 113; and 3) sequences covering the smallest part of the NS3 protease gene, from amino acid positions 29 to 85.

		PI-naïve sequences		
	PI[a]	Subset 1	Subset 2	Subset 3
PI-failure	Boceprevir	52	31	63
	Telaprevir	180	75	234
	Faldaprevir	9	9	9
PI treatment success	Boceprevir	16	27	27
	Telaprevir	52	107	107
	Simeprevir	3	3	3
	Faldaprevir	3	3	3
	Paritaprevir	10	10	10
Total		325	265	456

[a] All protease inhibitors were co-administrated with pegIFN-α + RBV, except for paritaprevir that was boosted with ritonavir and combined with ombitasvir and dasabuvir.

was assessed with a non-parametric bootstrap using 100 replicates (Friedman et al., 1999). Only arcs, but ignoring arc direction, with a bootstrap support over 65% were considered robust (Deforche et al., 2006).

In HIV research, a mutation that confers phenotypic resistance on its own and plays a key role in drug resistance is considered to be a major mutation (Shafer, 2002). A minor mutation increases resistance only in presence of a major mutation, or compensates for a possible fitness impact of other mutations, and therefore is selected only in the presence of other mutations. We used the same scheme for HCV RAVs. Semantic meaning of the BN with respect to drug resistance can be postulated for the presence of arcs in the network between amino acids and for the network structure around the drug node (Deforche et al., 2006). As a minor mutation interacts epistatically with a corresponding major mutation, the BN indicates this relationship by an arc between these mutations. The presence of a minor mutation is dependent on the presence of the corresponding major mutation, and thus not expected to be directly connected to the treatment node in the network. In contrast, a major mutation is connected to the treatment node. Other epistatic interactions can also be identified, which means that the association of an amino acid variant with treatment-experience can be dependent on the presence of the genetic background. Arc coloring reflected the estimated semantic meaning, in order to improve the interpretation of the graph.

A tool chain called 'Rega-BN' that includes all pre-processing steps, the identification of treatment-associated variants, and the Bayesian network procedure itself, was developed. This tool is freely available (www.github.com/rega-cev/rega-bn), and can be used for various pathogens.

3. Results

3.1. Investigation of RAVs associated with PI-exposure

3.1.1. Patients and sequences

For this part of the analysis, only the sequences from the 154 patients sampled in Belgium, Italy and Slovenia were included, most of which newly generated for this project. Subtyping was performed using the commercial InnoLipa assay and the results from the subtyping tools Oxford, COMET, and manual phylogenetic analysis with RAxML, were fully concordant. All sequences could be assigned to either HCV1a (33%) or HCV1b (67%) subtypes. Sequences were categorized according to the PI administered, with the majority of patients treated with either boceprevir or telaprevir, respectively 61 and 102 NS3 sequences. The remaining sequences included in the study cohort were

from patients treated with simeprevir (4), faldaprevir (16) and paritaprevir (10). In the boceprevir group, 59% of the sequences were classified as PI-naïve (36 sequences: 22% HCV1a – 78% HCV1b) and 41% (25 sequences: 28% HCV1a – 72% HCV1b) as PI-experienced. For telaprevir, a higher number of sequences was available: 76 PI-naïve (38% HCV1a – 62% HCV1b), and 26 PI-experienced sequences (50% HCV1a – 50% HCV1b).

3.1.2. Known resistance-associated variants

Fig. 1 shows the prevalence of previously reported resistance-associated variants, in naïve and treatment-experienced patients, treated with boceprevir or telaprevir, and selected for the study cohort. However, specific RAVs like V36A/G, T54C/G, V55A, R155G/I and A156V, known to confer drug resistance to boceprevir and telaprevir, were absent in our selected dataset.

3.1.3. Treatment-associated variants

The following amino acid variants were significantly associated with PI-exposure (selected by the Fisher's exact test as explained in Section 2.6): 20 NS3 amino acid positions with 22 different treatment-associated variants (7S, 35I, 36M, 40A, 42T, 48I/V, 56F, 61T, 62R, 64I, 66S, 71V, 72I, 86P, 89Q, 132V, 147A, 155K, 170I/V and 175L). All these variants, complemented with other known RAVs, were added as variables to map their dependency on PI-exposure using the Bayesian network approach.

3.1.4. Selective pressure analysis

Using the FEL and MEME method, the complete dataset and three partial datasets were analyzed separately: NS3 sequences from PI-naïve patients, sequences sampled on-treatment and sequences from post-treatment samples. Positive selective pressure was not detected in any of the datasets, by none of the two methods. A large difference was observed between the number of positions found to be under negative selective pressure, higher for the FEL compared to the MEME method. Using FEL, 91.1%, 87.3%, 77.9% and 70.0% of the amino acid positions were identified to be under negative selective pressure, for the entire dataset, PI-naïve sequences, on-treatment, and post-treatment data, respectively. Using the MEME method, a lower number of positions was found to be under negative selective pressure, for the PI-naïve dataset these were positions 6, 14, 18, 49, 60, 121 and 173 (3.9%), for the on-treatment datasets positions 7, 66, 95, 174 and 178 (2.8%), and for the post-treatment sequences 95, 122, 174 and 178 (2.2%). Since there was a partial overlap, this results in 13 sites (7.2%) for the complete dataset. Among these positions, two have been reported as resistance-associated: positions 122 and 174.

3.1.5. Interpreting the inferred Bayesian network of PI naïve and PI experienced patients

Using all 193 HCV1a and HCV1b sequences of the study cohort, NS3 variant R155K was identified as a major drug resistance mutation, given the robust unconditional dependency on PI-exposure (ePI) in the network (Fig. 2). Resistance-associated variant V36M appeared to be directly dependent on R155K, and could thus be considered a minor drug resistance mutation. Variant R155M interacted with variant V36L, with the latter influenced by background polymorphism R117C, with neither of these three variants being drug-related. Additionally, PI-exposure was directly related to variant 174H (<70% - not shown in the network). Although this variant is not known to result in drug resistance and this link was only supported by a low number of bootstrap replicates, this association could suggest 174H is a PI resistance mutation. In a separate HCV1a network of 40 PI-naïve and 20 PI-experienced sequences, the direct influence between major drug resistance mutation R155K and PI-exposure as well as variant V36M, was confirmed. This was not the case for HCV1b, where an interaction between variant V170I and PI-exposure was observed, however supported by a non-robust arc (networks not shown).
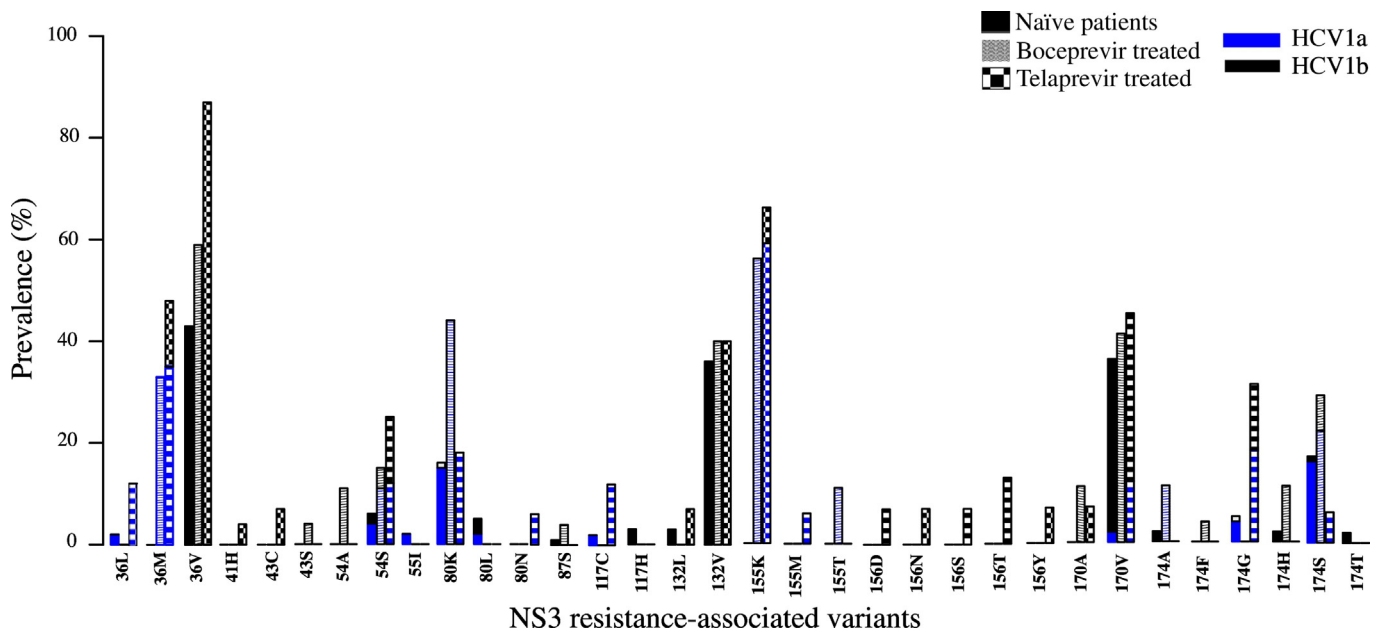
Fig. 1. Prevalence (%) of resistance-associated variants known to confer drug resistance to boceprevir and/or telaprevir. For each variant, the three bars represent the prevalence in naïve (left bar – full color), boceprevir (centered bar - striped) and telaprevir (right bar - blocked) treated patients. HCV1a and HCV1b sequences are indicated in blue and black, respectively (see inset legend). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Investigation of baseline variants predictive for therapy failure

3.2.1. Patients and sequences

The 137 PI-naïve sequences used in the first network together with the 319 additional PI-naïve sequences from Genbank resulted in a dataset of 456 sequences. Subtyping classified 44.7% of these sequences as HCV subtype 1a while the remaining 252 strains were classified as HCV1b.

3.2.2. Potential treatment-predictive variants

All NS3 sequences from PI-naïve patients who cleared HCV with PI treatment were compared to sequences from PI-naïve patients who failed PI therapy (Table 3), with the first subset of this analysis including 325 sequences covering NS3 amino acid positions 1 to 86. Respectively, 84 and 241 sequences from PI-naïve patients that cleared their virus and patients who failed PI treatment respectively, were compared. A set of 11 amino acids with 13 variants (35I, 40A, 42T, 48V, 56F, 61T, 62R, 64I, 66S, 71I/V, 72I/T) was identified to be significantly associated with PI treatment outcome (Fisher's exact test). For the second subset of 265 PI-naïve sequences with sequence information for NS3 positions 29 to 113 (respectively 150 and 115 strains of patients that successfully ended treatment vs patients that failed to clear HCV after PI therapy), 13 amino acids with 15 variants (35I, 40A, 42T, 48V, 61T, 62R, 64I, 66S, 67S, 71I/V, 72I/T, 86Q, 89Q) were significant and selected (Fisher's exact test). The last network included all sequences (respectively 150 vs 306 strains of PI-naïve patients that cleared HCV and failed PI treatment) but covered the smallest part of the NS3 protease (29–85), resulting in the selection of 11 amino acids with 13 variants (same selection as the first subset) as input for Bayesian network learning.

3.2.3. Interpreting the inferred Bayesian network of pre-therapy HCV sequences

Two Bayesian networks of respectively 325 and 456 PI-naïve NS3 sequences covering NS3 amino acids 1–86 and 29–85, both suggested that 72I is predictive of treatment success although the connecting arc had a low bootstrap support (<70% - Fig. 3A). No other amino acid variant was predictive of treatment failure or success. Polymorphism 72I was not only linked to PI therapy success, but also to the wild-type variants

64I and 66S, both by the presence of well-supported arcs. On NS3 residue 72, amino acid Threonine was also included in the network, however not directly associated with treatment outcome. A third partial BN of 265 PI-naïve NS3 sequences covering NS3 residues 29–113 identified variant 67S as a major predictor of PI treatment failure, supported by 100 bootstrap replicates (Fig. 3B).

4. Discussion

4.1. The value of Bayesian network learning

To our knowledge, this is the first study that implemented Bayesian network learning to assess drug resistance pathways for HCV. At least three other studies that apply this approach on HCV sequence data have been published, however these focused on the prediction of liver fibrosis progression (Lara et al., 2014), the prediction of treatment outcome based on early virological response and liver inflammation stage (Bijedic et al., 2012), or on the presence of nucleotide features (KayvanJoo et al., 2014). Bayesian network learning provides an intuitive framework that produces results that can be easily understood by clinicians and virologists. A tool chain that includes all pre-processing steps as well as the Bayesian network analysis, was developed to reduce the threshold for researchers with a biological background to perform such associative analyses (Libin, 2014), and is freely available (www. github.com/rega-cev/rega-bn). This tool chain is fully automated and has already been applied to HIV and HCV (Libin, 2014; Cuypers et al., 2016a), and is built in a generic way to accommodate the analysis of different pathogens.

4.2. Selection of patients and sequences

In order to study dependencies between amino acid variants and PI-exposure or treatment success/failure, HCV genetic sequences from infected patients sampled before, during and after treatment were collected retrospectively. The majority of patients in the participating hospitals were treated with a first-generation PI in the context of a clinical trial or other study, not the ideal sampling strategy to obtain samples before and during PI-exposure. In many cases, only a PI-naïve sample from before the start of the study was available as access to samples taken
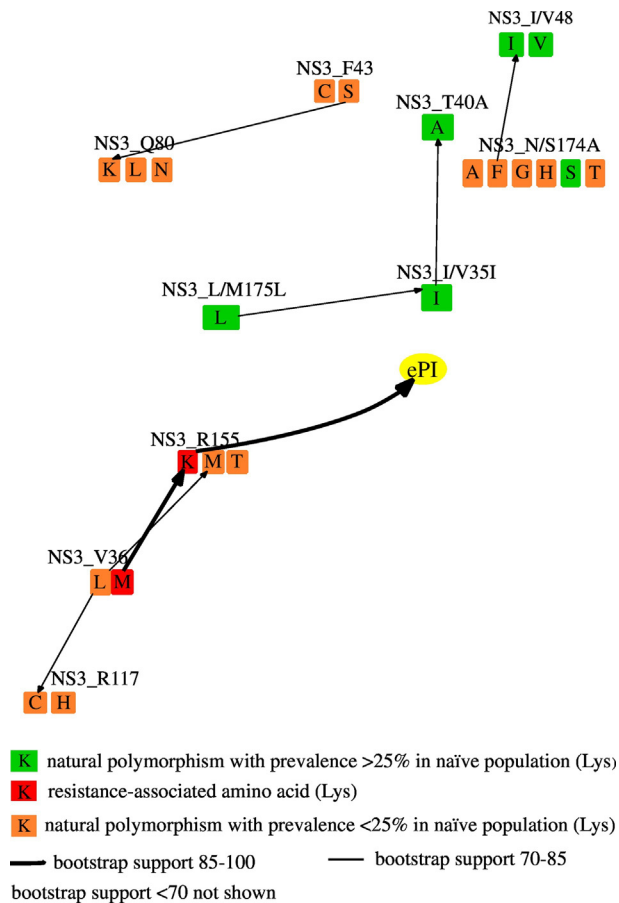
Fig. 2. Annotated Bayesian network showing dependencies between PI-associated variants and PI-exposure (ePI) in the HCV1a and HCV1b pooled dataset. An arc represents a direct dependency between the corresponding variables with thickness proportional to bootstrap support (85–100%: full line - 70–85%: full thin line – <70% no arc drawn). The color code of amino acid variants is defined based on their link with PI-exposure (ePI) and their prevalence in the PI-naïve population: natural polymorphisms with a prevalence >25% (green), natural polymorphisms with a prevalence <25% (orange), and resistance-associated variants that are directly linked to ePI or with one variant in between them and ePI (red). An antagonistic arc with a wild type was treated the same as a synergistic arc with mutations at this position. Arc direction has no causal meaning, but may indicate a non-additive multivariate effect. Only the part of the network around the treatment node is visualized to simplify the graph. Variables not connected with any arc were not shown in the network. For all remaining NS3 positions in the network, the wild-type variants in HCV1a (reference H77) and HCV1b (reference Con1) are listed before the amino acid position, separated by a slash (for example L/M175 meaning that Leucine is the wild type amino acid in HCV1a and Methionine in HCV1b). If both subtypes share the same wild-type, only one amino acid is listed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

during the course of the clinical trial was not allowed. As summarized in Table 2, the overall time range between sampling and starting PI treatment was broad, potentially introducing a bias in the analysis. Especially for post-treatment samples (median 96 days, but with a range up to 369 days after stop of treatment), sampling time was less appropriate. As such, treatment-associated variants may have been missed, since it has been reported that NS3 variants tend to disappear within one year after end of therapy (Barnard et al., 2013; Sullivan et al., 2013), and even shorter in the case of newer generation DAAs (Krishnan et al., 2015). For on-treatment sequences, several samples taken at early time points (<14 days) could not be included in the study since it has been shown that RAVs are only detected in plasma after two weeks when using population-based Sanger sequencing (Cento et al., 2015b). A second limitation was the small overall number of NS3 sequences available for the analysis. We were unable to extend the dataset with PI-experienced data from public databases since at the time of analysis,

no NS3 sequences annotated with detailed information regarding the HCV genotype and subtype, treatment regimen and treatment response were available. We strongly encourage the submission of viral sequence data annotated with clinical information. This will reduce bias in a similar analysis. Despite these limitations, this study, with 473 patients and 512 sequences annotated with treatment information, is the largest of its kind. This allowed us to draw some conclusions regarding particular amino acid variants, as well as demonstrate the value of our approach.

4.3. Resistance-associated variants in the dataset

In contrast to HIV, where several genotypic drug-resistance interpretation systems already exist (Vercauteren and Vandamme, 2006), few exist for HCV and the information required to design detailed interpretation algorithms for HCV is lacking. The most extensive genotypic interpretation system for HCV drug resistance variants is the Geno2pheno tool, based on in vitro and in vivo data (Barnard et al., 2013; Lontok et al., 2015; Sarrazin et al., 2007; Sullivan et al., 2013; Susser et al., 2009). This algorithm reports 50 variants to NS3/4A protease inhibitors, 38 of which are prevalent in our dataset, with only variants V36A/G/L, T54C/G, V55A, R155G/I/M and A156S/T/V, known to confer resistance to boceprevir or telaprevir, missing. The scarcity of data in investigations of HCV drug resistance compared to HIV, will not soon be overcome, since genotypic resistance testing is not routinely performed in case of virological failure or before start of treatment, with a few exceptions. One such exception is in the case of HCV1a cirrhotic patients, where the highly prevalent variant Q80K (Fig. 1, Cuypers et al., 2015, Sarrazin et al., 2015) needs to be monitored before starting the combination of simeprevir and sofosbuvir (Lawitz et al., 2015). Another exception is in cases where the treating physician of HCV1a infected patients wants to initiate treatment with grazoprevir and elbasvir. In these cases, genotyping can support the decision to extend treatment and to associate ribavirin when high-level resistance NS5A variants are present at baseline (AASLD and EASL guidelines 2015, and EASL guidelines 2016).

4.4. Absence of positive selective pressure

To detect selective pressure, both the FEL and MEME methods were used, while only results obtained with FEL were reported in previous studies (Cuypers et al., 2015 and Cuypers et al., 2016a, 2016c). It is known that the FEL method might score too many sites under negative selective pressure if some lineages have a different selective pressure than others. The MEME method is more sensitive in detecting potential positive selective pressure, for example from drugs, along some branches, provided that a measurable proportion of lineages experience non-synonymous evolution (Murrell et al., 2012). It is therefore not surprising that fewer sites were found to be consistently under negative selective pressure using MEME versus FEL, in the PI-naïve dataset here 3.9% versus 87.3%, in agreement with 32%–83% reported in previous studies of drug naïve sequences, by Cuypers et al. (2015) and (2016a, 2016c). When re-analyzing the larger datasets formerly used in Cuypers et al. (2015) (data not shown), the proportion of sites under negative selective pressure was also consistently lower using MEME versus FEL. Therefore, the large difference observed here might be mainly due to a few lineages with some positions not under negative selective pressure, in the context of the majority of lineages experiencing negative selective pressure for those sites, or alternatively, due to the lack of necessary divergence for MEME to adequately detect sites under negative selective pressure (Murrell et al., 2012). Also, in this dataset, none of the sites was detected to be under positive selective pressure, similar as in a previous larger dataset from DAA-naïve patients for which positively selected sites in NS3 were identified only within the helicase region and not in the protease region (Cuypers et al., 2015). These findings are in agreement with previous data suggesting that HCV mainly experiences neutral evolution (Cuypers et al., 2015, 2016c). This might explain the low number of amino acid variants
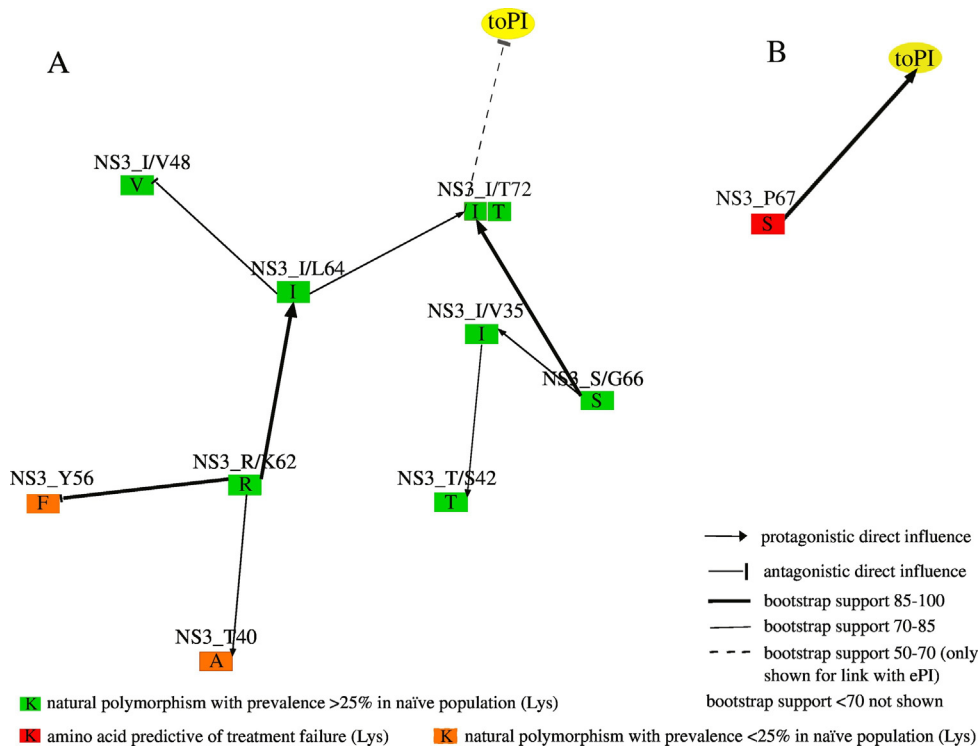
Fig. 3. Annotated Bayesian network predicting the link of variant 67S with PI treatment failure (toPI), based on NS3 sequences from PI-naïve patients. A) PI-naïve NS3 sequences covering the first 86 NS3 residues resulted into the same network as obtained by using PI-naïve NS3 sequences covering NS3 residues 29–85 (visualized only once). B) PI-naïve sequences covering NS3 residues 29–113 identified variant P67S as a major predictor of PI-treatment failure. The rest of this network is not visualized, as only dependencies between wild-type amino acid variants were identified. For more information, see the caption of Fig. 2.

identified to be associated with PI-exposure using Bayesian network learning, in comparison with what is generally seen in for example HIV drug resistance analyses (Deforche et al., 2006, 2007, 2008) for which a larger number of positions was identified to be under positive selective pressure (Snoeck et al., 2011).

4.5. Associations between NS3 drug-resistance variants and PI-exposure

Due to the limited availability of NS3 sequences, all sequences available in the study cohort were used for Bayesian network learning in a cross-sectional approach comparing drug-naïve and drug-experienced sequences. Treatment-associated variants were selected based on the results of the Fisher's exact test. The network identified an association between drug resistance mutation R155K (Bartels et al., 2008; Larrat et al., 2015; Lontok et al., 2015) and PI-exposure, suggesting this is a major mutation, while V36M is a minor drug resistance-associated variant, dependent also on the presence of R155K. These associations were confirmed in a separate HCV1a analysis, but no in the HCV1b analysis, probably due to R155K being less prevalent since two nucleotide substitutions (instead of one) are required in HCV1b patients (Lim et al., 2012). Cross-resistance between first- and second-generation PIs has been reported for this variant (Kieffer et al., 2012). All HCV1a patients in this dataset who harbored variant R155K during course of treatment, failed to clear HCV infection with a triple therapy based on boceprevir or telaprevir, and were classified as null-responders. Drug resistance-associated variants R155M and V36L interacted, and both are known to confer resistance towards telaprevir in HCV1a infected patients (Lontok et al., 2015). Additionally, variant V36L was identified to be dependent on polymorphisms on NS3 position 117 (Susser et al., 2012).

4.6. NS3 baseline variants predictive of treatment outcome

NS3 sequences of PI-naïve patients who cleared HCV were compared to sequences of PI-naïve patients who failed PI therapy. PI-naïve

sequences from Genbank, annotated with treatment outcome, were added to improve statistical power. Two networks covering NS3 protease residues 1 to 86 and 29 to 85 both suggested a predictive value of amino acid 72I for PI treatment success, although supported by <70% bootstrap replicates. Polymorphism 72T on the other hand was previously identified as a compensatory mutation in the NS3/4A protease region (Lopez-Labrador et al., 2008), however, in the Bayesian network it was not directly linked to treatment outcome. Amino acid variants located at position 42, 62, 64 and 66, were observed around regions with higher amino acid entropy (Cuypers et al., 2015; Lopez-Labrador et al., 2008). A third Bayesian network covering NS3 residues 29 to 113 showed variant 67S to be predictive of PI treatment failure, in agreement with previous findings that HCV1a infected patients carrying this variant failed triple therapy (Cento et al., 2014). Mutation P67S was formerly identified in replicons isolated from one single resistant clone (Trozzi et al., 2003). As already suggested in a former study (Cento et al., 2014), concomitance of variants P67S and N174G (variable not shown in the network) was observed.

5. Conclusion

A tool chain was developed to allow researchers to perform associative analyses using Bayesian network learning without extensive bioinformatics skills. However, limitations in the study design, lack of positive selective pressure and limited access to treatment-annotated sequence data, reduced the number of dependencies identified in this study, yet important associations were found. A direct influence between variant R155K and PI-exposure in HCV1a infected patients was observed, as well as an epistatic interaction between variants 155K and 36M. NS3 polymorphism 67S at baseline was associated with subsequent PI-failure, while baseline variant 72I seemed to be predictive of treatment success. Given the genetic changes in NS3 under treatment, and the absence of positive selective pressure, it seems that HCV mainly experiences neutral evolution, even during therapy. Bayesian network learning of large HCV sequence data from patients treated with DAAs

has the potential to provide insights on the specific role of RAVs. This is especially interesting in case of retreatment with NS5A inhibitors since NS5A RAVs tend to persist for a long time.

Funding

Acknowledgments

References

Alcantara, L.C.J., Cassol, S., Libin, P., et al., 2009. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res. 37:634–642. http://dx.doi.org/10.1093/nar/gkp455.

Andreone, P., Colombo, M.G., Enejosa, J.V., et al., 2014. ABT-450, ritonavir, ombitasvir, and dasabuvir achieves 97% and 100% sustained virologic response with or without ribavirin in treatment-experienced patients with HCV genotype 1b infection. Gastroenterology 147:359–365. http://dx.doi.org/10.1053/j.gastro.2014.04.045.

Bacon, B.R., Gordon, S.C., Lawitz, E., et al., HCV RESPOND-2 Investigators, 2011. Boceprevir for previously treated chronic HCV genotype 1 infection. N. Engl. J. Med. 364: 1207–1217. http://dx.doi.org/10.1056/NEJMoa1009482.

Barnard, R., Howe, J.A., Ogert, R.A., et al., 2013. Analysis of boceprevir resistance associated amino acid variants (RAVs) in two phase 3 boceprevir clinical studies. Virology 444: 329–336. http://dx.doi.org/10.1016/j.virol.2013.06.029.

Bartels, D.J., Zhou, Y., Zhang, E.Z., et al., 2008. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naïve subjects. J. Infect. Dis. 198:800–807. http://dx.doi.org/10.1086/591141.

Bartosch, B., Dubuisson, J., Cosset, F.L., 2003. Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. J. Exp. Med. 197:633–642. http://dx.doi.org/10.1084/jem.20021756.

Bijedic, N., Hamulic, I., Vukobrat-Bijedic, Z., Husic-Selimovic, A., 2012. Bayesian Network Model of HCV Therapy Response Prediction in FBiH. 2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics, September 20–22, 2012, Subotica, Serbia.

Black, S., Pak, I., Ingravallo, P., et al., 2015. Resistance Analysis of Virologic Failures in Hepatitis C Genotype 1 Infected Patients Treated With Grazoprevir/Elbasvir +/− Ribavirin: The C-Worthy Study. 50th EASL Vienna 2015, p. P0891.

Cento, V., Mirabelli, C., Salpini, R., et al., 2012. HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. PLoS One 7: e39652. http://dx.doi.org/10.1371/journal.pone.0039652.

Cento, V., Di Maio, V.C., Di Paolo, D., et al., 2014. HCV Genetic Background and Early Detection of Resistance Mutations Are Associated With Low Viral Decay and Viral Failure to First Generation Protease Inhibitors in Difficult-to-Treat patients. 12th European Workshop on HIV & Hepatitis, Barcelona, Spain, March 26–28, 2014. Abstract O_6.

Cento, V., Di Paolo, D., Di Carlo, D., 2015a. Hepatitis C virus RNA levels at week-2 of telaprevir/boceprevir administration are predictive of virological outcome. Dig. Liver Dis. 47:157–163. http://dx.doi.org/10.1016/j.dld.2014.11.010.

Cento, V., Tontodonati, M., Di Maio, V.C., et al., 2015b. Kinetics of hepatitis C virus RNA decay, quasispecies evolution and risk of virological failure during telaprevir-based triple therapy in clinical practice. Dig. Liver Dis. 47:233–241. http://dx.doi.org/10.1016/j.dld.2014.12.004.

Chevaliez, S., 2011. Antiviral activity of the new DAAs for the treatment of hepatitis C virus infection: virology and resistance. Clin. Res. Hepatol. Gastroenterol. 35: S46–S51. http://dx.doi.org/10.1016/S2210-7401.

Cuypers, L., Li, G., Libin, P., et al., 2015. In context of eradication of the hepatitis C virus: genetic diversity and selective pressure of HCV genotypes 1–6. Viruses. 7: 2039–5016. http://dx.doi.org/10.3390/v7092857.

Cuypers, L., Snoeck, J., Kerremans, L., et al., 2016a. HCV1b genome evolution under selective pressure of the cyclophilin inhibitor alisporivir during the DEB-025-HCV-2013 phase II clinical trial. Infect. Genet. Evol. 44:169–181. http://dx.doi.org/10.1016/j.meegid.2016.06.050.

Cuypers, L., Ceccherini-Silberstein, F., Van Laethem, K., et al., 2016b. Impact of HCV genotype on treatment regimens and drug resistance: a snapshot in time. Rev. Med. Virol. http://dx.doi.org/10.1002/rmv.1895.

Cuypers, L., Li, G., Neumann-Haefelin, C., et al., 2016c. Mapping the genomic diversity of HCV subtypes 1a and 1b: implications of structural and immunological constraints for vaccine and drug development. Virus Evol. 2. http://dx.doi.org/10.1093/ve/vew024.

De Luca, A., Bianco, C., Rossetti, B., 2014. Treatment of HCV infection with the novel NS3/4A protease inhibitors. Curr. Opin. Pharmacol. 18:9–17. http://dx.doi.org/10.1016/j.coph.2014.07.016.

De Oliveira, T., Deforche, K., Cassol, S., et al., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21:3797–3800. http://dx.doi.org/10.1093/bioinformatics/bti607.

Deforche, K., Silander, T., Camacho, R., et al., non-B workgroup, 2006. Analysis of HIV-1 pol sequences using Bayesian networks: implications for drug resistance. Bioinformatics 22:2975–2979. http://dx.doi.org/10.1093/bioinformatics/btl508.

Deforche, K., Camacho, R., Grossman, Z., et al., 2007. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. Infect. Genet. Evol. 7:382–390. http://dx.doi.org/10.1016/j.meegid.2006.09.004.

Deforche, K., Camacho, R.J., Grossman, Z., et al., non-B Workgroup, 2008. Bayesian network analyses of resistance pathways against efavirenz and nevirapine. AIDS 22: 2107–2115. http://dx.doi.org/10.1097/QAD.0b013e32830fe940.

Dvory-Sobol, H., Wyles, D., Ouyang, W., et al., 2015. Long-term persistence of HCV NS5A variants after treatment with NS5A inhibitor ledipasvir. 50th EASL Vienna 2015, p. O059.

Foster, G.R., Afdhal, N., Roberts, S.K., et al., ASTRAL-2 Investigators, ASTRAL-3 Investigators, 2015. Sofosbuvir and velpatasvir for HCV genotype 2 and 3 infection. N. Engl. J. Med. 373 (27):2608–2617. http://dx.doi.org/10.1056/NEJMoa1512612.

Franco, S., Tural, C., Nevot, M., et al., 2014. Detection of a sexually transmitted hepatitis C virus protease inhibitor-resistance variant in a human immunodeficiency virus-infected homosexual man. Gastroenterology 174:599–601. http://dx.doi.org/10.1053/j.gastro.2014.05.010.

Friedman, N., Goldszmidt, M., Wyner, A., 1999. Data Analysis with Bayesian Networks: A Bootstrap Approach. Proceedings of the Fifteenth conference of uncertainty in artificial intelligence. Morgan Kaufmann Publicers Inc., San Francisco.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27:221–224. http://dx.doi.org/10.1093/molbev/msp259.

Gower, E., Estes, C., Blach, S., et al., 2014. Global epidemiology and genotype distribution of the hepatitis C virus infection. J. Hepatol. 61:S45–S57. http://dx.doi.org/10.1016/j.hep.2014.07.027.

Halfon, P., Locarnini, S., 2011. Hepatits C virus resistance to protease inhibitors. J. Hepatol. 55:192–206. http://dx.doi.org/10.1016/j.jhep.2011.01.011.

Heckerman, D., 1999. A Tutorial on Learning with Bayesian Networks. In Learning in Graphical Models. MIT Press, pp. 301–354 ISBN: 0-262-60032-3.

Hézode, C., Forestier, N., Dusheiko, G., et al., PROVE2 Study Team, 2009. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N. Engl. J. Med. 360:1839–1850. http://dx.doi.org/10.1056/NEJMoa0807650.

Kalaghatgi, P., Sikorski, A.M., Knops, E., et al., 2016. Geno2pheno[HCV] - a web-based interpretation system to support hepatitis C treatment decisions in the era of direct-acting antiviral agents. PLoS One 11, e0155869. http://dx.doi.org/10.1371/journal.pone.0155869.

KayvanJoo, A.H., Ebrahimi, M., Haqshenas, G., 2014. Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms. BMC Res. Notes 7:565. http://dx.doi.org/10.1186/1756-0500-7-565.

Kieffer, T.L., De Meyer, S., Bartels, D.J., et al., 2012. Hepatitis C viral evolution in genotype 1 treatment-naïve and treatment-experienced patients receiving telaprevir-based therapy in clinical trials. PLoS One 7. http://dx.doi.org/10.1371/journal.pone.0034372.

Krishnan, P., Tripathi, R., Schnell, G., et al., 2015. Long-Term Follow-Up of Treatment-Emergent Resistance-Associated Variants in NS3, NS5A and NS5B With Paritaprevir/r, Ombitasvir- and Dasabuvir-based Regimens. 50th EASL Vienna 2015, p. O057.

Kuiken, C., Hraber, P., Thurmond, J., et al., 2008. The hepatitis C sequence database in Los Alamos. Nucleic Acids Res. 36, D512–D516.

Lara, J., López-Labrador, X.F., González-Candelas, F., et al., 2014. Computational models of liver fibrosis progression for hepatitis C virus chronic infection. BMC Bioinforma. 15: S5. http://dx.doi.org/10.1186/1471-2105-15-S8-S5.

Larrat, S., Vallet, S., David-Tchouda, S., et al., 2015. Naturally occurring resistance-associated variants of hepatitis C virus protease inhibitors in poor responders to pegylated interferon-ribavirin. J. Clin. Microbiol. 53:2195–2202. http://dx.doi.org/10.1128/JCM.03633-14.

Lawitz, E., Matusow, G., DeJesus, E., et al., 2015. A Phase 3, Open-Label, Single-Arm Study to Evaluate the Efficacy and Safety of 12 Weeks of Simeprevir (SMV) Plus Sofosbuvir (SOF) in Treatment-Naïve or –Experienced Patients With Chronic HCV Genotype 1 Infection and Cirrhosis: OPTIMIST-2. 50th EASL, Vienna, Austria, April 22-26, 2015 (Abstract LP04).

Lee, S.M., Abbott, P.A., 2003. Bayesian networks for knowledge discovery in large datasets: basic for nurse researchers. J. Biomed. Inform. 36:389–399. http://dx.doi.org/10.1016/j.jbi.2003.09.022.

Libin, P., 2014. Applying Graphical Modelling Techniques to Virological Data. Thesis for the Fulfilment of Master of Science in Applied Informatics, Vrije Universiteit Brussel, June 14.

Lim, S.R., Qin, X., Susser, S., et al., 2012. Virologic escape during danoprevir (ITMN-191/ RG7227) monotherapy is hepatitis C virus subtype dependent and associated with R155K substitution. Antimicrob. Agents Chemother. 56:271–279. http://dx.doi.org/ 10.1128/AAC.05636-11.

Lontok, E., Harrington, P., Howe, A., et al., 2015. Hepatitis C virus drug resistance-associated substitutions: state of the art summary. Hepatology 62:1623–1632. http://dx.doi. org/10.1002/hep.27934.

Lopez-Labrador, F.X., Moya, A., Gonzàlez-Candelas, F., 2008. Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. Antivir. Ther. 13, 481–494.

Murrell, B., Wertheim, J.O., Moola, S., et al., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, e1002764. http://dx.doi.org/10.1371/ journal.pgen.1002764.

Myllymäki, P., Silander, T., Tirri, H., Uronen, P., 2002. B-course: a web-based tutorial for Bayesian and causal data analysis. Int. J. Artif. Intell. Tools. 11:387–396. http://dx. doi.org/10.1142/S02182130020000940.

Pawlotsky, J.M., 2016. Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. Gastroenterology 151:70–86. http://dx.doi.org/10.1053/j.gastro. 2016.04.003.

Pearl, J., 1998. Graphical models for probabilistic and causal reasoning. In: Gabbay, D.M., Smets, P. (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision. Kluwer Academic Publishers, Dordrecht, pp. 367–389.

Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679. http://dx.doi.org/10.1093/bioinformatics/bti079.

Raney, K.D., Sharma, S.D., Moustafa, I.M., et al., 2010. Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target. J. Biol. Chem. 285:22725–22731. http://dx.doi.org/10.1074/jbc.R110.125294.

Romano, K.P., Ali, A., Aydin, C., et al., 2012. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. PLoS Pathog. 8, e1002832. http://dx.doi. org/10.1371/journal.ppat.1002832.

Salam, K.A., Akimitsu, N., 2013. Hepatitis C virus NS3 inhibitors: current and future perspectives. Biomed. Res. Int. http://dx.doi.org/10.1155/2013/467869.

Sarrazin, C., Kieffer, T.L., Bartels, D., et al., 2007. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. Gastroenterology 132, 1767–1777 (PMID: 17484874).

Sarrazin, C., Lathouwers, E., Peeters, M., et al., 2015. Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region. Antivir. Res. 116: 10–16. http://dx.doi.org/10.1016/j.antiviral.2015.01.003.

Shafer, R.W., 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. Clin. Microbiol. Rev. 15, 247–277.

Smith, D.B., Bukh, J., Kuiken, C., et al., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatol. 59:318–327. http://dx.doi.org/10.1002/hep.26744.

Snoeck, J., Fellay, J., Bartha, I., et al., 2011. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. Retrovirology 8:87. http://dx.doi.org/10.1186/1742-4690-8-87.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. http://dx.doi.org/10.1093/ bioinformatics/btu033.

Struck, D., Lawyer, G., Ternes, A.M., et al., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42, e144. http://dx.doi. org/10.1093/nar/gku739.

Sullivan, J.C., De Meyer, S., Bartels, D.J., et al., 2013. Evolution of treatment-emergent resistant variants in telaprevir phase 3 clinical trials. Clin. Infect. Dis. 57:221–229. http:// dx.doi.org/10.1093/cid/cit226.

Susser, S., Welsch, C., Wang, Y., et al., 2009. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. Hepatology 50:1709–1718.

Susser, S., Schelhorn, S.E., Lange, C.M., et al., 2012. Ultratiefe pyrosequenz-analyse (UDPS) von neu beschriebenen seltenen resistenzvarianten der hepatitis C virus NS3 protease bei patienten, die mit telaprevir oder boceprevir behandelt wurden. Z. Gastroenterol. 50:K052. http://dx.doi.org/10.1055/s-0032-1323987.

Theys, K., Deforche, K., Libin, P., et al., 2010. Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. J. Gen. Virol. 91:1898–1908. http://dx.doi.org/10.1099/vir.0.022657-0.

Theys, K., Abecasis, A.B., Vandamme, A.M., 2013. HIV-1 drug resistance: where do polymorphisms fit in? Future Microbiol 8:303–306. http://dx.doi.org/10.2217/fmb.13.10.

Trozzi, C., Bartholomew, L., Ceccacci, A., et al., 2003. In vitro selection and characterization of hepatitis C virus serine protease variants resistant to an active-site peptide inhibitor. J. Virol. 77:3669–3679. http://dx.doi.org/10.1128/JVI.77.6.3669-3679.2003.

Vercauteren, J., Vandamme, A.-M., 2006. Algorithms for the interpretation of HIV-1 genotypic drug resistance information. Antivir. Res. 71:335–342. http://dx.doi.org/10. 1016/j.antiviral.2006.05.003.

Welsch, C., Jesudian, A., Zeuzem, S., et al., 2012. New direct-acting antiviral agents for the treatment of hepatitis C virus infection and perspectives. Gut 61:36–46. http://dx.doi. org/10.1136/gutjnl-2012-302144.

Wyles, D.L., 2013. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J. Infect. Dis. 207 (Suppl. 1):S33–S39. http://dx.doi.org/10.1093/infdis/ jis761.

Zeuzem, S., Andreone, P., Pol, S., et al., 2011. Telaprevir for retreatment of HCV infection. N. Engl. J. Med. 364:2417–2428. http://dx.doi.org/10.1056/NEJMoa1013086.

Zeuzem, S., Ghalib, R., Reddy, K.R., et al., 2015a. Grazoprevir-Elbasvir combination therapy for treatment-naïve cirrhotic and noncirrhotic patients with chronic HCV genotype 1, 4 or 6 infection: a randomized trial. Ann. Intern. Med. 163:1–13. http://dx.doi.org/10. 7326/M15-0785.