This is a pre- or post-print of an article published in
Irving, W.L., Rupp, D., McClure, C.P., Than, L.M.,
Titman, A., Ball, J.K., Steinmann, E., Bartenschlager,
R., Pietschmann, T., Brown, R.J.P.
Development of a high-throughput pyrosequencing assay for
monitoring temporal evolution and resistance associated
variant emergence in the Hepatitis C virus protease
coding-region
(2014) Antiviral Research, 110 (1), pp. 52-59.

Development of a high-throughput pyrosequencing assay for monitoring temporal evolution and resistance associated variant emergence in the Hepatitis C virus protease coding-region

William L. Irving^{a*}, Daniel Rupp^{b, c*}, C. Patrick McClure^a, Lwin Mar Than^a, Andrew Titman^a, Jonathan K. Ball^a,
Eike Steinmann^d, Ralf Bartenschlager^{b, c}, Thomas Pietschmann^d and Richard J. P. Brown^{d#}

^aThe School of Life Sciences and the NIHR Nottingham Digestive Diseases Biomedical Research Unit, The University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK

^bDepartment of Infectious Diseases, Molecular Virology, Heidelberg University, 69120 Heidelberg, Germany and ^cGerman Center for Infection Research, Heidelberg University

^dDivision of Experimental Virology, TWINCORE, Centre for Experimental and Clinical Infection Research; a joint venture between the Medical School Hannover (MHH) and the Helmholtz Centre for Infection Research (HZI), Feodor-Lynen-Str. 7, 30625 Hannover, Germany

*These authors contributed equally

Corresponding author:

Richard J. P. Brown

Tel. +49 511 22 00 27 133

Fax. +49 511 22 00 27 137

e-mail: richard.brown@twincore.de

ABSTRACT

A new generation of drugs targeting the non-structural (NS) proteins of the Hepatitis C virus (HCV) will substantially increase treatment success rates, reducing global infections. Amongst the NS proteins, the NS3 protease represents an important drug target, responsible for liberation of mature NS proteins from the nascent HCV polyprotein and suppression of host innate immunity. Despite this, the evolutionary stability of the genomic locus encoding the NS3 protease is poorly characterized in chronic HCV infection. To address this shortfall, we developed a high-throughput amplicon pyrosequencing protocol and utilised it to monitor NS3 protease coding-sequence evolution for over a decade in two patients. Although patient-specific evolutionary trends were apparent, the protease amino acid population consensus remained stable with a massive excess of synonymous mutations observed, confirming this locus is under strong purifying selection during chronic infection within individual patients. No evidence for continuous immune escape was detected. Additionally, both patients failed protease inhibitor (PI) therapy and protease sequence diversity pre- and post-therapy were also assessed. No baseline resistance associated variants (RAVs) contributed to treatment failure. Significant reductions in viral diversity were observed post-PI therapy, indicating a population bottleneck occurred. The genetic vestiges of this bottleneck were still detectable 18 months after therapy discontinuation. Although significant enrichment of the Q80L mutation was observed in one patient, genetic and phenotypic data reveal no detectable RAV persistence post-therapy failure. Together this investigation provides a sensitive and reproducible high-throughput framework to interrogate viral sequence diversity at high-resolution, with potential applications for routine monitoring of treatment regimens. This study also reveals novel insights into the evolutionary processes that shape NS3 sequence divergence in both chronic HCV infection and post PI-therapy failure.

1. INTRODUCTION

HCV is an enveloped virus and possesses a ~9.6 kb +sense RNA genome containing an open-reading frame which is translated as a single polyprotein. Cleavage by host and viral proteases into 10 mature viral proteins then occurs. The non-structural proteins are required for viral RNA (vRNA) replication and virion assembly. HCV infection is the leading indication for liver transplantation in developed nations and constitutes a major healthcare and economic burden worldwide (Gravitz, 2011). Approximately 75-85% of infections progress to chronicity, predisposing carriers to an increased risk of developing cirrhosis and hepatocellular carcinoma (Hoofnagle, 2002).

Until recently HCV treatment options were limited in their efficacy and did not target the virus directly, promoting viral clearance by boosting host immune responses. Genotype 1 viruses are the most globally prevalent and the predominant circulating strain in developed nations (Gravitz, 2011). The paucity of response rates in genotype 1 infections to pegylated interferon- α and ribavirin (PEG-IFN α /RBV) therapy (Welsch et al., 2012) has driven the development of alternative therapeutic options: small molecule inhibitors targeted at specific steps in the viral life-cycle and known as Direct Acting Antivirals (DAAs).

In combination with PEG-IFN α /RBV, the first generation of licensed protease inhibitors (PIs) have substantially increased treatment response rates in genotype 1 infected patients (Salam and Akimitsu, 2013; Welsch et al., 2012). However, RAV emergence to DAAs is well documented (Halfon and Locarnini, 2011; Sarrazin and Zeuzem, 2010). Consequently, the success of DAA treatments may be compromised by baseline polymorphisms which may confer resistance to specific DAAs (Kuntzen et al., 2008), in addition to breakthrough DAA-induced resistant viruses. Genetic surveillance of RAVs at the inpatient level is thus essential to facilitate targeted healthcare delivery and effective management of antiviral regimens. To this end, we developed a high-throughput amplicon pyrosequencing strategy, which we utilised to monitor the evolutionary stability of the viral NS3 protease coding-region in two chronically infected patients over a 10 year period. In addition, both patients failed to achieve viral clearance after therapy with NS3 PI Faldaprevir

(Llinas-Brunet et al., 2010) in combination with PEG-IFN α /RBV. Consequently, characterised RAV frequencies pre- and post-therapy were examined, to determine whether undetectable baseline RAVs contributed to treatment failure, and assess the extent of long-term RAV persistence post-DAA therapy failure. Treatment-naïve and post-treatment failure protease coding-regions were engineered into sub-genomic replicons (SGRs) for phenotyping. Together this pilot study sheds new light on HCV NS3 protease evolution over the course of chronic infection and assesses long-term RAV persistence at unprecedented resolution.

2. MATERIALS AND METHODS

2.1. Amplification of HCV NS3 protease domain

Extraction of viral RNA (vRNA) from patient sera was performed using a QIAamp MinElute Viral Spin kit (Qiagen) according to the manufacturer's protocol. Both patients provided informed consent. Synthesis of cDNA via gene specific primer reverse transcription (GSP-RT) was performed using Superscript III Reverse transcriptase (Life Technologies) according to the manufacturer's instructions, with minor modifications. 8 μ l of extracted vRNA was used for GSP-RT in 20 μ l volumes using 2pmol primer (5'-TCYGCTTGGTCMAGGAC-3') (Imhof and Simmonds, 2010). Reverse transcription reactions were incubated at 50°C for 1 hour followed by 42°C for 1 hour.

Nested PCR was performed using previously described 1a-specific primer pairs (Imhof and Simmonds, 2010). First round amplifications were performed in 50 μ l volumes containing 5 μ l of viral cDNA, 0.2 μ M of each 1st round primer, 1.5 U Platinum *Taq* High Fidelity polymerase (Invitrogen), 1x High Fidelity polymerase buffer and 2 mM MgSO₄. Cycling parameters comprised initial denaturation at 94°C for 2 minutes followed by 35 cycles of 94°C for 15 seconds, 50°C for 30 seconds and 68°C for 2 minutes, with a final extension step at 68°C for 10 minutes. Second round amplifications were performed using identical conditions to the 1st round, except 5 μ l of 1st round product was used as template and 2nd round primers replaced 1st round primers. Cycling parameters for 2nd round amplification were identical to the 1st round with minor modifications: the cycle number was increased to 45, the annealing temperature was increased to 55°C and

the annealing time reduced to 1.5 minutes. Amplicons were cleaned-up using QIAquick PCR purification columns (Qiagen) and sequenced using BigDye™ v3.1 (Applied Biosystems).

2.2. Roche 454 pyrosequencing

GS FLX Titanium adaptor sequences were incorporated onto the 5' and 3' ends of each amplicon (0.5-1 µg) via integration PCR, using one custom barcoded megaprimer pair per amplicon. Tagged amplicons were subsequently pooled and run on an agarose gel, prior to gel extraction. The amplicon pool was used to generate a DNA library which was bidirectionally sequenced using the Roche 454 GS-FLX system. Sequence data were analyzed using CLC genomics workbench v4.9.

2.3. 5'NCR Quantitative PCR

Quantitative PCR was performed on a Stratagene MX4000 instrument in 20 µl volumes containing 0.2 µM forward (5' - GCGCAACCGGTGAGTACA -3') and reverse (5' - ACTCGCAAGCACCTATCAG -3') primers, 0.5 µl viral cDNA and 10 µl of 2× SYBR Green JumpStart *Taq* ReadyMix (Sigma) reagent with additional ROX reference dye as per the manufacturer's instructions. Cycling conditions comprised initial denaturation at 95°C for 2 minutes, followed by 50 cycles of 95°C for 30 seconds, 60°C for 1 minute, 72°C for 30 seconds, with a further denaturation for 1 minute at 95°C followed by a temperature gradient from 55°C to 95°C at 30 seconds/°C to determine amplicon melt peaks.

2.4. Subgenomic replicon luciferase reporter assays

The Q80L mutation, enriched post-Faldeprevir therapy in Patient B, was engineered into H77 (subtype 1a) (Esser-Nobis et al., 2013) and Con1 (subtype 1b) subgenomic replicons (SGRs) via overlapping mutagenesis PCR (Friebe et al., 2005). Additionally, entire NS3 protease sequences derived from patient B pre-treatment (Q80) and post- treatment (L80) viruses were engineered into a Con1 SGR backbone (Qi et al., 2009). *In vitro* transcription and electroporation of subgenomic HCV RNAs, in addition to luciferase assays, were performed as previously described (Koutsoudakis et al., 2006; Lohmann et al., 1999).

2.5. Titration of Boceprevir and Telaprevir

Boceprevir and Telaprevir were dissolved in dimethylsulfoxide (DMSO) at a stock concentration of 10mM and diluted to the indicated concentrations for titration with an adjusted DMSO concentration of 0.05%

(v/v). Titrations were started 4h post electroporation and cells were incubated for 68h prior to Firefly luciferase assays.

3. RESULTS

3.1. Temporal sampling of chronic HCV infection

HCV+ serum samples were retrospectively obtained from two patients enrolled in the therapy arm of the SILEN-C2 trial, whereby treatment-experienced patients received PEG-IFN α /RBV therapy supplemented with the PI Faldaprevir. (Sulkowski et al., 2013b). Both patients were male, chronically infected for over 10-years with HCV subtype 1a viruses and non-responders to previous PEG-IFN α /RBV treatment. Preceding trial commencement, both patients were determined to have advanced fibrosis (Ishak stage 5) via liver biopsy. Patient A received therapy for 4 weeks prior to discontinuation due to adverse side-effects. Patient B received therapy for 24-weeks before discontinuation. Both patients were determined to be HCV RNA+ at therapy discontinuation. Seven temporally-spaced pre-SILEN-C2 samples, in addition to one post-SILEN-C2 sample were investigated for each patient (Figure 1A).

3.2. Evaluation of 454 amplicon sequencing

Pyrosequencing was performed using the Roche 454 platform. Per amplicon, average nucleotide depth ranged between 3,278-7,159, average read length ranged between 421-451 bases, average read count ranged between 4,246-9,286 reads, and total nucleotides sequenced ranged between 1,847,196 and 4,058,738 bases. Thus, 454 run statistics were comparable for all sequenced amplicons (Fig 1B).

Quantification of viral genomic cDNA input into each 454 amplicon was assessed via 5'UTR qPCR (Fig 1C). These data reveal 454 run statistics from amplicons derived from high-titre cDNAs are indistinguishable from those derived from low-titre cDNAs, which differed from each other by as much as three orders of magnitude (Figs 1B and 1C). Amplicons derived from low copy number cDNAs are marked with an asterisk in Figure 1C and were omitted from subsequent analyses.

3.3. Quantification of NS3 diversity via pyrosequencing

Temporal sequence evolution of NS3 protease domain over the course of chronic HCV infection was tracked via pyrosequencing (Figures 2 and 3). Mutational spectra were determined at the 1% level across the entire protease domain: In both patients prior to therapy, at all sampling points, an excess of synonymous mutations over non-synonymous mutations is apparent, relative to the master sequence. The population amino acid consensus remains stable for nearly a decade in both patients, although fluctuations in position and frequency of sites exhibiting silent polymorphism are observed. Over the course of infection, patient-specific trends are observed with greater diversity seen in patient B, when compared to patient A. In patient A, between the first (Figure 2A) and last sampling before therapy (Figure 2C), a significant increase ($p<0.05$) in frequencies of SNVs is observed. Contrastingly, a quantitative (although non-significant) reduction in frequencies of SNVs is observed in patient B between the first (Figure 3A) and last sampling before therapy (Figure 3F).

Post-Faldeprevir therapy, a significant reduction in the total amount of sites exhibiting polymorphism was observed in both patients, when compared to the pre-therapy population (Figure 2. Patient A: $A_6 = 57 > A_8 = 32$. $p<0.05$. Figure 3. Patient B: $B_7 = 63 > B_8 = 20$: $p<0.001$). The vestiges of this bottleneck, a significant reduction in NS3 diversity, are still apparent almost two years after therapy cessation. Patient-specific trends are observed in the post-therapy population. While the reduction in viral diversity is significant in patient A (Figure 2D), the observed bottleneck is less extreme than that observed for patient B (Figure 3G). This may be reflective of the differential durations of therapy received by each patient (Figure 1A). In patient B, a total of five silent mutations achieved fixation in the post-therapy sample. A single non-synonymous mutation (A239T) encoding a Q80L amino acid replacement was enriched post-therapy (95%: 7667/8104 reads) (Figure 3G). A239T was also detected in the pre-therapy sample via 454-pyrosequencing (4%: 213/5512 reads) although below conventional sequencing detection levels (Figure 3F).

Excluding Q80L, no detectable non-synonymous polymorphism $>1\%$ was apparent at any of the eight major residues reported as conferring resistance to a range of PIs (Bartenschlager et al., 2013; Halfon and

Locarnini, 2011; Sarrazin and Zeuzem, 2010) in either pre- or post-therapy populations. Consequently, levels of non-synonymous sequence polymorphism were determined at high resolution (>0.01%) at these critical residues (Figure 4). For control purposes, non-synonymous variation was also determined at critical active site residues which are essential for protease activity. Low frequency non-synonymous variation was detected at residues associated with DAA resistance. However, excluding Q80L, variability at resistance associated residues was comparable in pre- and post-therapy samples, and similar to observed variability at active site residues (Figure 4). Thus, this variation likely represents either stochastic errors introduced by the viral RdRp, or background PCR error, rather than low-level RAV persistence. Unfortunately we did not have access to on-treatment serum samples to investigate which (if any) mutations contributed to treatment failure. Minor enrichment of V170I was observed in the post-PI therapy population of patient B. However, this mutation is reported to circulate naturally and is not a described resistance mutation.

3.4. Phenotyping of the Q80L mutation

Independent of genetic background, L80 does not exert any deleterious effect on viral replication when compared to the Q80 wild-type (Figure 5A-C). Unfortunately, we were unable to access the compound Faldeprevir to test whether the L80 mutation conferred resistance in our phenotyping assay. However, importantly, L80 does not confer cross-resistance to Boceprevir (Figure 5D-E) or Telaprevir (Figure 5G-H) in a subtype 1b or subtype 1a SGR background. In addition, post-treatment B8 NS3 protease coding-regions containing Q80L were not demonstrably more sensitive to Boceprevir (Figure 5F) or Telaprevir (Figure 5I) than pre-treatment B7 NS3 protease coding-regions containing Q80, indicating future therapeutic options including additional PIs are unlikely to be compromised for patient B.

4. DISCUSSION

Using 454-pyrosequencing to interrogate NS3 intrahost polymorphism, we observe patient-specific trends in synonymous variation over the course of chronic HCV infection, while the stability of the population amino acid consensus is maintained. Our intrahost analysis confirms non-synonymous variation in NS3 is generally selectively deleterious and removed from the population via purifying selection, as evidenced by the massive excess of synonymous variation. The genetic data presented herein provides compelling evidence for

extreme structural/function constraints restricting fixation of deleterious mutations in the NS3 protease domain: the majority of non-synonymous variation represents deleterious mutational space. Indeed, despite the inherent high mutation rate of the viral RdRp, the restrictions on the mutability of the NS3 protease domain locus, in addition to other NS proteins, have been successfully exploited therapeutically as evidenced by the successful development of a range of DAA inhibitors (Bartenschlager et al., 2013). These data also indicate the NS3 protease is not subject to constantly evolving host immune targeting over the course of chronic infection, as there are no genetic signatures indicating viral escape.

While incorporation of DAAs into antiviral regimens will significantly improve treatment success rates, the minority who fail therapy due to the emergence of resistant strains require further monitoring and investigation. In this study, we also endeavoured to investigate patterns of protease coding-region evolution post-Faldeprevir therapy failure, to assess levels of RAV persistence and thus inform future therapeutic options. Pyrosequencing revealed the genetic signatures of a population bottleneck, which presumably occurred during therapy. This reduction in viral diversity is still detectable approximately 18 months after the cessation of therapy. Additionally, no compelling evidence for RAV persistence is detected in virus derived from sera, with levels of non-synonymous variation at described resistance sites comparable to that observed for catalytic triad residues and thus likely to represent inherent assay background error.

Faldeprevir treatment failures are generally associated with mutations at protease amino acid residues 155 and 168, with R155K RAVs frequently associated with genotype 1a treatment failures and D168V RAVs commonly observed in genotype 1b treatment failures (Berger et al., 2013; Sulkowski et al., 2013a). Neither of these mutations could be detected in the pre- or post-therapy samples from Patient A or B. In patient B, the Q80L mutation is detectable via pyrosequencing at low levels pre-Faldeprevir therapy and becomes dominant post-therapy. Interestingly, Q80L mutations do not render replicons resistant to a range of PIs (Lenz et al., 2010) and are not implicated in Faldeprevir resistance *in vitro* (Lagace et al., 2012) or *in vivo* (Berger et al., 2013). Our data demonstrate the Q80L mutation is selectively neutral with respect to viral

replication and does not confer cross-resistance to Boceprevir or Teleprevir. It is therefore likely that L80 has reached virtual fixation post-therapy due to the population bottleneck which occurs during treatment. Together these data indicate that PI therapy can sometimes facilitate fixation of stable non-synonymous neutral mutations in the viral population. While RAV persistence in the liver remains a possibility, this is unlikely as HCV is a +sense RNA virus which completes its replication cycle outside the nucleus. Together, the stable reduction in inpatient viral diversity coupled with no obvious cross-resistance or RAV persistence in circulating virus indicates that future therapeutic options for Faldeprevir treatment failures are unlikely to be comprised long term. However, baseline Q80K mutations have been reported to have significant impact on the efficacy of the PI Simeprevir (Report, 2013). While Q80L does not confer resistance to Faldeprevir, Boceprevir or Teleprevir, its effect on the efficacy of Simeprevir requires further investigation.

The mechanisms underlying treatment failure in the two studied patients remain unresolved, although duration of treatment, previous non-response to IFN α /RBV and host genetics at the interferon lambda (*IFNL*) locus are likely to be critical determinants. Patient A was withdrawn from the SILEN-C2 trial after four weeks due to adverse side-effects, which likely explains the non-response to Faldeprevir/IFN α /RBV. While patient B completed 24-weeks of therapy, a previous history of non-response to IFN α /RBV significantly reduces sustained virological response (SVR) rates to Faldeprevir/IFN α /RBV (Sulkowski et al., 2013b) when compared to treatment-naïve SVR rates (Sulkowski et al., 2013a). Underlying this, SNPs at the *IFNL3* (*IL28B*) locus and indels producing the novel *IFNL4* gene profoundly influence the outcome of infection and response to IFN-based therapies (Ge et al., 2009; Sheahan et al., 2014; Suppiah et al., 2009; Tanaka et al., 2009; Thomas et al., 2009). Thus, host genetics are expected to play an important role in determining the outcome of any HCV IFN-containing therapies.

Many studies which utilise NGS to investigate viral intrahost complexity fail to quantitate viral input prior to amplicon sequencing. Importantly, in this study we employed GSP-RT of the NS3 protease coding-region, to enable accurate quantification of viral cDNA input to each amplicon via 5'NCR qPCR. GSP-RT avoids off-target

reverse transcription of host mRNA and inflated viral load estimation due to circulating 5' subgenome contamination (Shimizu et al., 2006), both artefacts associated with reverse transcription via random hexamers. The data generated indicate quantification of viral genomic input prior to any NGS strategy is essential to avoid making erroneous claims about viral population diversity and identify where significant re-sampling is likely to occur due to low genomic copy number input.

In this study, 454-pyrosequencing was utilised to interrogate HCV protease coding-region evolution due to its compatibility with amplicon sequencing (Radford et al., 2012), although one can envisage adapting the workflow to any of the currently available NGS platforms, in addition to alternative regions of the viral genome targeted by DAAs. Our study clearly demonstrates that, if appropriately controlled, 454-pyrosequencing is a highly sensitive tool to monitor viral population diversity and emergence of potentially resistant strains. Indeed, NGS methods for interrogating viral diversity allow viral populations to be scanned at unprecedented depth, many orders of magnitude greater than clonal sequencing. Together this study demonstrates the applicability of our methods to monitoring and management of therapeutic regimens, in addition to studies of viral evolution and pre-clinical evaluation of RAV emergence to antiviral compounds. The data generated also reveals the signatures of strong purifying selection in chronic HCV infection, no evidence for continuous immune escape, genetic bottlenecking and fixation of neutral mutations occurring post PI-therapy and no evidence for long-term stability of RAVs.

5. ACKNOWLEDGEMENTS

RB and WI were funded by a University of Nottingham Faculty Research Strategy Group (FRSG) pump-priming award. TP was supported by the iMed initiative of the Helmholtz Association of German Research Centres. ES was supported by DFG (STE 1954/1-1). DR was supported by the MD/PhD-Program of the Hartmut Hoffmann-Berling International Graduate School of Molecular and Cellular Biology. We thank Steven Ryder and Susan Slininger for invaluable assistance concerning acquisition of patient information and clinical trial details.

REFERENCES

- Bartenschlager, R., Lohmann, V., Penin, F., 2013. The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nat Rev Microbiol* 11, 482-496.
- Berger, K.L., Lagace, L., Triki, I., Cartier, M., Marquis, M., Lawetz, C., Bethell, R., Scherer, J., Kukulj, G., 2013. Viral resistance in hepatitis C virus genotype 1-infected patients receiving the NS3 protease inhibitor Faldaprevir (BI 201335) in a phase 1b multiple-rising-dose study. *Antimicrob Agents Chemother* 57, 4928-4936.
- Esser-Nobis, K., Romero-Brey, I., Ganten, T.M., Gouttenoire, J., Harak, C., Klein, R., Schemmer, P., Binder, M., Schnitzler, P., Moradpour, D., Bartenschlager, R., Polyak, S.J., Stremmel, W., Penin, F., Eisenbach, C., Lohmann, V., 2013. Analysis of hepatitis C virus resistance to silibinin in vitro and in vivo points to a novel mechanism involving nonstructural protein 4B. *Hepatology* 57, 953-963.
- Friebe, P., Boudet, J., Simorre, J.P., Bartenschlager, R., 2005. Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. *Journal of virology* 79, 380-392.
- Ge, D., Fellay, J., Thompson, A.J., Simon, J.S., Shianna, K.V., Urban, T.J., Heinzen, E.L., Qiu, P., Bertelsen, A.H., Muir, A.J., Sulkowski, M., McHutchison, J.G., Goldstein, D.B., 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461, 399-401.
- Gravitz, L., 2011. Introduction: a smouldering public-health crisis. *Nature* 474, S2-4.
- Halfon, P., Locarnini, S., 2011. Hepatitis C virus resistance to protease inhibitors. *J Hepatol* 55, 192-206.
- Hoofnagle, J.H., 2002. Course and outcome of hepatitis C. *Hepatology* 36, S21-S29.
- Imhof, I., Simmonds, P., 2010. Development of an intergenotypic hepatitis C virus (HCV) cell culture method to assess antiviral susceptibilities and resistance development of HCV NS3 protease genes from HCV genotypes 1 to 6. *Journal of virology* 84, 4597-4610.
- Koutsoudakis, G., Kaul, A., Steinmann, E., Kallis, S., Lohmann, V., Pietschmann, T., Bartenschlager, R., 2006. Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. *Journal of virology* 80, 5308-5320.
- Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A.M., Young, S.K., Lee, B., Heckerman, D., Carlson, J., Reyor, L.L., Kleyman, M., McMahon, C.M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A.D., Lauer, G.M., Rosen, H.R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A.Y., Xing, Y., Schneidewind, A., Madey, M.A., Fleckenstein, J.F., Park, V.M., Galagan, J.E., Nusbaum, C., Walker, B.D., Lake-Bakaar, G.V., Daar, E.S., Jacobson, I.M., Gomperts, E.D., Edlin, B.R., Donfield, S.M., Chung, R.T., Talal, A.H., Marion, T., Birren, B.W., Henn, M.R., Allen, T.M., 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. *Hepatology* 48, 1769-1778.

Lagace, L., White, P.W., Bousquet, C., Dansereau, N., Do, F., Llinas-Brunet, M., Marquis, M., Massariol, M.J., Maurice, R., Spickler, C., Thibeault, D., Triki, I., Zhao, S., Kukulj, G., 2012. In vitro resistance profile of the hepatitis C virus NS3 protease inhibitor BI 201335. *Antimicrob Agents Chemother* 56, 569-572.

Lenz, O., Verbinen, T., Lin, T.I., Vijgen, L., Cummings, M.D., Lindberg, J., Berke, J.M., Dehertogh, P., Fransen, E., Scholliers, A., Vermeiren, K., Ivens, T., Raboisson, P., Edlund, M., Storm, S., Vrang, L., de Kock, H., Fanning, G.C., Simmen, K.A., 2010. In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435. *Antimicrob Agents Chemother* 54, 1878-1887.

Llinas-Brunet, M., Bailey, M.D., Goudreau, N., Bhardwaj, P.K., Bordeleau, J., Bos, M., Bousquet, Y., Cordingley, M.G., Duan, J., Forgione, P., Garneau, M., Ghiro, E., Gorys, V., Goulet, S., Halmos, T., Kawai, S.H., Naud, J., Poupart, M.A., White, P.W., 2010. Discovery of a potent and selective noncovalent linear inhibitor of the hepatitis C virus NS3 protease (BI 201335). *J Med Chem* 53, 6466-6476.

Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L., Bartenschlager, R., 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* 285, 110-113.

Qi, X., Bae, A., Liu, S., Yang, H., Sun, S.C., Harris, J., Delaney, W., Miller, M., Mo, H., 2009. Development of a replicon-based phenotypic assay for assessing the drug susceptibilities of HCV NS3 protease genes from clinical isolates. *Antiviral Res* 81, 166-173.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. *J Gen Virol* 93, 1853-1868.

Report, F., 2013. FDA ANTIVIRAL DRUGS ADVISORY COMMITTEE MEETING OCTOBER 24, 2013 BACKGROUND PACKAGE FOR NDA 205123 SIMEPREVIR (TMC435).

Salam, K.A., Akimitsu, N., 2013. Hepatitis C virus NS3 inhibitors: current and future perspectives. *Biomed Res Int* 2013, 467869.

Sarrazin, C., Zeuzem, S., 2010. Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* 138, 447-462.

Sheahan, T., Imanaka, N., Marukian, S., Dorner, M., Liu, P., Ploss, A., Rice, C.M., 2014. Interferon lambda alleles predict innate antiviral immune responses and hepatitis C virus permissiveness. *Cell host & microbe* 15, 190-202.

Shimizu, Y.K., Hijikata, M., Oshima, M., Shimizu, K., Yoshikura, H., 2006. Detection of a 5' end subgenome of hepatitis C virus terminating at nucleotide 384 in patients' plasma and liver tissues. *Journal of viral hepatitis* 13, 746-755.

Sulkowski, M.S., Asselah, T., Lalezari, J., Ferenci, P., Fainboim, H., Leggett, B., Bessone, F., Mauss, S., Heo, J., Datsenko, Y., Stern, J.O., Kukulj, G., Scherer, J., Nehmiz, G., Steinmann, G.G., Bocher, W.O., 2013a. Faldaprevir combined with pegylated interferon alfa-2a and ribavirin in treatment-naïve patients with chronic genotype 1 HCV: SILEN-C1 trial. *Hepatology* 57, 2143-2154.

Sulkowski, M.S., Bourliere, M., Bronowicki, J.P., Asselah, T., Pawlotsky, J.M., Shafran, S.D., Pol, S., Mauss, S., Larrey, D., Datsenko, Y., Stern, J.O., Kukolj, G., Scherer, J., Nehmiz, G., Steinmann, G.G., Bocher, W.O., 2013b. Faldaprevir combined with peginterferon alfa-2a and ribavirin in chronic hepatitis C virus genotype-1 patients with prior nonresponse: SILEN-C2 trial. *Hepatology* 57, 2155-2163.

Suppiah, V., Moldovan, M., Ahlenstiel, G., Berg, T., Weltman, M., Abate, M.L., Bassendine, M., Spengler, U., Dore, G.J., Powell, E., Riordan, S., Sheridan, D., Smedile, A., Fragomeli, V., Muller, T., Bahlo, M., Stewart, G.J., Booth, D.R., George, J., 2009. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nature genetics* 41, 1100-1104.

Tanaka, Y., Nishida, N., Sugiyama, M., Kurosaki, M., Matsuura, K., Sakamoto, N., Nakagawa, M., Korenaga, M., Hino, K., Hige, S., Ito, Y., Mita, E., Tanaka, E., Mochida, S., Murawaki, Y., Honda, M., Sakai, A., Hiasa, Y., Nishiguchi, S., Koike, A., Sakaida, I., Imamura, M., Ito, K., Yano, K., Masaki, N., Sugauchi, F., Izumi, N., Tokunaga, K., Mizokami, M., 2009. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nature genetics* 41, 1105-1109.

Thomas, D.L., Thio, C.L., Martin, M.P., Qi, Y., Ge, D., O'Huigin, C., Kidd, J., Kidd, K., Khakoo, S.I., Alexander, G., Goedert, J.J., Kirk, G.D., Donfield, S.M., Rosen, H.R., Tobler, L.H., Busch, M.P., McHutchison, J.G., Goldstein, D.B., Carrington, M., 2009. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 461, 798-801.

Welsch, C., Jesudian, A., Zeuzem, S., Jacobson, I., 2012. New direct-acting antiviral agents for the treatment of hepatitis C virus infection and perspectives. *Gut* 61 Suppl 1, i36-46.

Figure 1. Patient sampling and 454-pyrosequencing run statistics. (A) HCV+ sera sampling and therapy timeline for patient A (left panel) and patient B (right panel). Black closed triangles represent serum sampling time-points, grey shaded bars represent IFN α /RBV therapy duration and black shaded bars represent Faldaprevir/IFN α /RBV therapy duration. (B) 454-pyrosequencing run statistics for patient A (left panel) and patient B (right panel). Mean read length (black bars), total read count (dark grey bars), mean coverage (light grey bars) and total nucleotides sequenced (white bars) are plotted for each amplicon. (C) Viral genomic cDNA input into each 454 amplicon for patient A (left panel) and patient B (right panel). Amplicons generated from low copy cDNAs are marked with a * and were excluded from the subsequent analysis.

Figure 2. Tracking of NS3 protease coding-region evolution in patient A via NGS. Nucleotide depth and mutation plots for the NS3 protease domain from Patient A are ordered chronologically down the page, with the earliest sampling at the top (A) and the final post-therapy sampling at the bottom (D). For each plot, all mutations present at greater than 1% are plotted relative to the master sequence, the consensus sequence derived from the first sampling-point. For individual plots, the x-axis represents nucleotide position in the NS3 protease domain. The y-axis represents nucleotide coverage across NS3 protease domain, with synonymous (d_s) (black bars) and non-synonymous (d_N) (blue bars) SNV frequencies in the viral population relative to the master sequence highlighted. Nucleotide coverage at each position is represented by continuous background plot (sky blue). Closed triangles (black) positioned below the x-axis the represent control non-synonymous SNV (E176G) which was incorporated into 50% of the 2nd round antisense primers. Statistical significance of differences in numbers of SNVs were assessed via χ^2 tests. Bracket linking panel A with panel C indicates differences in SNVs numbers over the course of chronic infection with associated p -value (* = $p < 0.05$). Bracket linking panel C with panel D indicates differences in SNVs numbers when comparing pre- and post-PI therapy populations with associated p -value (* = $p < 0.05$).

Figure 3. Tracking of NS3 protease coding-region evolution in patient B via NGS. Nucleotide depth and mutation plots for the NS3 protease domain from Patient B are ordered chronologically down the page, with the earliest sampling at the top (A) and the final post-therapy sampling at the bottom (G). Nucleotide depth and mutation plots are presented as previously in Figure 3. Chromatographs derived from parallel Sanger sequencing of amplicons are positioned below pre-therapy (F) and and post-therapy (G) mutation plots and represent NS3 nucleotide positions 238, 239 and 240 encoding amino acid position 80. Open triangles located above the post-therapy mutation plot (G) represent d_s (black) or d_N (blue) mutations that have achieved fixation in the viral population. Statistical significance of differences in numbers of SNVs were assessed via χ^2 tests. Bracket linking panel A with panel F indicates differences in SNVs numbers over the course of chronic infection (NS = non-significant). Bracket linking panel F with panel G indicates differences in SNVs numbers when comparing pre- and post-PI therapy populations with associated p -value (***) = $p < 0.001$).

Figure 4. Detectable polymorphism at critical residues. Nucleotide depth and mutation plots at 0.01% for critical residues in the NS3 protease domain associated with PI-resistance. Boxed residues H57, D81 and S139 represent control catalytic triad residues critical for protease activity. The x-axis represents nucleotide depth across the NS3 protease, with nucleotide coverage (black bars) and non-synonymous variation at critical residues (d_N) (blue bars) presented. Dotted lines represent mean variation detectable at 1%. Pre- and post-PI therapy plots for patient A (A-B respectively) are presented in the two left hand panels while pre- and post-PI therapy plots for patient B (C-D respectively) are presented in the two right hand panels.

Figure 5. Replication and resistance profiles of Q80 (WT) and L80 (mutant) SGRs. Replication dynamics of Q80 and L80 in the context of Con1 (A) and H77 (B) SGR backgrounds. Replication dynamics of entire patient B protease domains (subtype 1a) from pre-PI (Q80) and post-PI (L80) samples (C) in a Con1 SGR background (subtype 1b). RLU: relative light units. Sensitivity to Boceprevir of Q80 and L80 in Con1 (D) and H77 (E) SGR backgrounds. As a positive control the T54A mutation, conferring resistance against Boceprevir and

Telaprevir, was introduced into Con1 SGR via overlapping mutagenesis PCR. Sensitivity to Boceprevir of entire patient B protease domains from pre-PI (Q80) and post-PI (L80) samples (F) in a Con1 SGR background. Sensitivity to Teleprevir of Q80 and L80 in Con1 (D) and H77 (E) SGR backgrounds. Sensitivity to Telaprevir of entire patient B protease domains from pre-PI (Q80) and post-PI (L80) samples (F) in a Con1 SGR background. Data presented are the means from three independent experiments. Error bars represent standard deviation around the mean. A hybrid replicon (Hybrid), carrying the H77 NS3 protease domain in a Con1 backbone was generated as reference construct (Qi et al., 2009). A Δ NS3 protease 1b SGR was created as a replication negative control.