UCLA
UCLA Electronic Theses and Dissertations

Title
High-Resolution Functional Profiling Hepatitis C Virus Genome

Permalink
https://escholarship.org/uc/item/86z9j5qz

Author
QI, HANGFEI

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

High-Resolution Functional Profiling Hepatitis C Virus Genome

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular and Medical Pharmacology

by

Hangfei Qi

2012

ABSTRACT OF THE DISSERTATION

High-Resolution Functional Profiling Hepatitis C Virus Genome

by

Hangfei Qi

Doctor of Philosophy in Molecular and Medical Pharmacology

University of California, Los Angeles, 2012

Professor Ren Sun, Chair

Hepatitis C virus (HCV) is a major cause of human liver diseases and the mortality associated with chronic infection has been increasing due to lack of effective treatments. Thus, better understanding the functional domains in the virus genome and their contributions to viral-host interactions will shed light on novel drug development.

In an attempt to systematically map the anti-interferon (IFN) functional domains in HCV genome, we established a genetic profiling platform by combining high-throughput mutagenesis and next-generation sequencing. An IFN-α screen performed with a 15-nucleotide random insertion library identified four regions with increased IFN-sensitivity when mutated. Further analyzing the N-terminus of core in a secondary screen with saturation mutagenesis revealed that phenylalanine 24 is essential for inhibiting STAT1 phosphorylation thereby blocking IFN signaling transduction. The genomic IFN-α screen suggested that p7 is a novel immune evasion protein. To interrogate the mechanism that governs p7 counteracting the IFN response, a cDNA

library screen of liver-specific interferon-stimulated genes was conducted, and individual effect of 107 genes on p7 mutant virus replication was determined. The screen showed that virus with p7 knockout is more sensitive to IFI6-16 over-expression. Further analysis demonstrated a physical interaction between p7 and IFI6-16, suggesting p7 may target IFI6-16 to actively suppress innate immune response.

The same concept of systematic profiling approach was applied to determine the residues interacting with Daclatasvir, which was identified as an effective NS5A inhibitor. Quantitatively examining the fitness of a saturation mutant library within domain IA of NS5A upon the drug treatment uncovered potential drug-interaction residues. Epistatic interactions among these residues are correlated with genotype-specific differences in drug-sensitivity. Remarkably, the fitness score of all possible substitutions and their drug-sensitivity allow for systematical mapping of the genetic barriers and prediction of evolutionary paths for potentially emerging resistances.

Taken together, we have profiled the HCV genome to define the essential residues for evading host immune responses or mediating drug interactions. We suggest that the genetic profiling platform described in this thesis can be generally applied in interrogating virus-host interactions and chemical-target interactions, which will provide comprehensive knowledge on new therapeutic strategies to overcome persistent HCV infection.

The dissertation of Hangfei Qi is approved.

Sam Chow

Asim Dasgupta

Huiying Li

Ren Sun, Committee Chair

University of California, Los Angeles

2012

献给

我亲爱的丈夫和我们的父母

This dissertation is dedicated

to

My beloved husband, Qingshan Wei, and our parents,

for their continuous support, encouragement and love.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION ii

DEDICATIONS. v

TABLE OF CONTENTS vi

LIST OF FIGURES ix

LIST OF TABLES xi

ACKNOWLEDGEMENTS xii

VITA xv

PUBLICATIONS AND PRESENTATIONS xv

CHAPTER 1. INTRODUCTION 1

1.1 Hepatitis C virus 2

1.2 HCV experimental systems 3

1.3 HCV life cycle 5

1.4 Success of HCV infection (evasion of innate antiviral responses by HCV) 6

1.5 Drug development for antiviral therapy 8

1.6 Bibliography 11

CHAPTER 2. A GENETIC PLATFORM FOR SYSTEMATIC MAPPING OF ANTI-INTERFERON FUNCTIONS IN THE HEPATITIS C VIRUS GENOME 17

2.1 Abstract 18

2.2 Introduction 19

2.3 Results 20

2.3.1 High-resolution profiling of the HCV genome. 20

2.3.2 Constructing the tagged viruses based on the profiling data. 25

2.3.3 IFN-α selection of the mutant library revealed four IFN-hypersensitive domains. 25

2.3.4 Validation of the phenotype identified by IFN-α screen. 27

2.3.5 Profiling the critical residues of the core N-terminus. 29

2.3.6 Residue phenylalanine 24 is critical for inhibition of the IFN responsive signaling pathway. 31
2.4 Discussion 32
2.5 Materials and methods 37
2.6 Acknowledgements 44
2.7 Supplementary information 45
2.8 Bibliography 54
CHAPTER 3. P7 IS A NOVEL IMMUNE EVASION PROTEIN OF HCV 58
3.1 Abstract 59
3.2 Introduction 60
3.3 Results 61
3.3.1 Mutations in p7 confer hypersensitivity of the virus to IFN-α treatment. 61
3.3.2 Liver-specific ISG library screen identifies genes that specifically inhibit p7 mutant virus replication. 62
3.3.3 P7 specifically interacts with IFI6-16. 65
3.4 Discussion 68
3.5 Materials and Methods 72
3.6 Bibliography 76
CHAPTER 4. SYSTEMATICALLY EXPLORING DISTINCT FITNESS LANDSCAPES OF AN HCV PROTEIN UNDER DRUG TREATMENT 80
4.1 Abstract 81
4.2 Introduction 82
4.3 Results 84
4.3.1 Fitness landscape mapping of NS5A domain 1A during HCV viral replication. 84
4.3.2 Fitness profile reveals residues encoding critical functions. 89
4.3.3 Daclatasvir drives alteration of fitness landscape. 93

4.3.4 Selection of mutant library in Daclatasvir reveals epistatic interactions among the critical drug associated residues. 94
4.3.5 Functional profiling identifies potential emergence of drug-resistant mutations. 97
4.4 Discussion 100
4.5 Materials and Methods 104
4.6 Bibliography 107
CHAPTER 5. CONCLUSIONS AND PERSPECTIVES 110

LIST OF FIGURES

Figure 2- 1. A schematic diagram of HCV genome profiling method. 21

Figure 2- 2. Frequency of each mutant indicated by different insertion sites, in the input and selected libraries. 23

Figure 2- 3. Verification of the profiling method by constructing epitope tagged mutant viruses. 26

Figure 2- 4. The genomic screen of mutant virus library with IFN-α treatment and validation of individual mutant viruses. 28

Figure 2- 5. Identification of the critical residues in the N-terminal core region for counteracting IFN-α response by saturation mutagenesis screen and validation. 30

Figure 2- 6. Phenylalanine 24 is critical for blocking STAT1 phosphorylation induced by IFN-α. 33

SI Figure 2- 1. 15nt genomic mutagenesis screen. 45

SI Figure 2- 2. IFN selection of the 15nt mutant library. 47

SI Figure 2- 3. Validation for the IFN screen with genomic mutagenesis library. 48

SI Figure 2- 4. Amino acid 17-28 is important for inhibiting IFN signaling pathway. 49

SI Figure 2- 5. Saturation mutagenesis of N-terminus of core. 50

SI Figure 2- 6. The fitness profile of mutant in the libraries. 51

SI Figure 2- 7. The frequency of mutation F24Q, F24W, K23N and K23R in the IFN screen. ...52

SI Figure 2- 8. Comparison of genetic profiling platform with traditional genetics strategies. 53

Figure 3- 1. Mutations in p7 confer hypersensitivity to IFN treatment. 63

Figure 3- 2. Liver-specific ISG library screen for identifying ISGs that are regulated by p7. 67

Figure 3- 3. P7 interacts with IFI6-16. 69

Figure 4- 1. Schematic picture showing the construction of saturation mutant library in a sub-domain of NS5A of HCV. 86

Figure 4- 2. The fitness landscape of amino acids 18-103 in NS5A in virus replication. 88

Figure 4- 3. Functional profiling of NS5AD1A reveals the structure-function interactions of the domain. 90
Figure 4- 4. I52 is an essential solvent-exposed hydrophobic residue critical for virus assembly 92
Figure 4- 5. Functional profiling under drug treatment reveals residues mediating drug-protein interaction and the epistatic interactions among them. 95
Figure 4- 6. Epistatic interactions between critical drug-associated residues and prediction of drug-resistant mutant emergence in patients. 98

LIST OF TABLES

Table 4- 1. Summary of input mutant library properties..... 87

ACKNOWLEDGEMENTS

As I move into this section of the dissertation, it marks the end of my graduate school. It has been over six years since the first time I visited UCLA when I was still a junior undergraduate student. Now when I turn around and look over these years at UCLA, too many individuals have made profound impacts on my life, to whom I feel deeply grateful.

First and foremost I would like to express my sincerest gratitude to my PhD supervisor and mentor, Dr Ren Sun, who completely transformed my life. It would never have been possible without the opportunity of working on an undergraduate thesis in his lab that he offered many years ago. I appreciated the substantial amount of guidance, support, encouragement and more importantly, the patience that he has with me when I had to learn everything from scratch. I am truly grateful for the freedom in pursuit of scientific interests and the potential career developments that he has provided. His enthusiasm and passion towards science and life has motivated and guided me through the PhD study at UCLA. Many of the work experiences with him during the past years are lifetime treasures.

I also want to sincerely thank my committee member, Dr. Sam Chow, with whom I did my first rotation and got the first taste of virology. Dr. Asim Dasgupta, who is also our collaborator, has offered many research resources and provided sound advice on my research projects. I also want to thank Dr. Huiying Li for her guidance and suggestions on my oral qualification exam, which I always remember: be precise.

There are so many people who have been involved in the creation of this work. In particular, I wish to express special acknowledgements to Dr. Anders Olson, for his great patience and tremendous input in the last project of my thesis. Without the high standard that he set for the project, the chapter four would never have been possibly done. I would like to thank Dr. Vaithilingaraja Arumugaswami, Roland Remenyi, who taught me about hepatitis c virus from the first day I joined this lab, for all of their guidance, instruction, and discussion on the projects. Thanks go to Nicholas, for his instant assistance in the large-scale data analysis and extremely

quick response whenever I have a request on data analysis. For my talented and hard-working undergraduate students Virginia Chu and Shawna Truong, thanks for their contributions to the projects and great company in the lab for the past two years. I would like to thank all of the past and current Sun Lab members for making such a collaborative and extremely friendly working environment. I am grateful to meet each of them, and I am truly proud of being part of this lovely family.

I want to thank all my collaborators inside and outside of UCLA. Thanks to Dr. Zugen Chen for kindly providing the facilities and the tutorials on the next-generation sequencing and helping work out the enrichment conditions. Thanks to Dr. Weidong Zhong, who generously provided the compound for the study of my last project. Thanks to Daniel Su, Dr. Shu-Hwa Chen and Dr. Chung-Yen Lin for doing the bioinformatics analyses. Thanks to Dr. Ruian Ke and Dr. James Lloyd-Smith for their help on the mathematical simulation.

Lastly, I would like to express my sincerest thanks to my dear friends for their constant support and invaluable friendship, which greatly enriched my graduate school life at UCLA and helped me mature through this process. I finally want to dedicate this dissertation to my parents and my beloved husband, who have always been there, loving, caring and supportive throughout my life. Without them, I would not have been what I am today.

Chapter two is a version the following manuscript: Qi H, Su SY, Chen Z, Wu NC, Arumugaswami V, Truong S, Chu V, C. Olson CA, Remenyi R, Young A, Wu TT, Chen SH, Lin CY, and Sun R. A genetic platform for systematical mapping of anti-interferon functions in the Hepatitis C virus genome.

Chapter three is a version of the following manuscript: Qi H, Chu V, Wu NC, Truong S, Chan E, Sun R. P7 is a novel immune evasion protein of HCV.

Chapter four is a version of the following manuscript: Qi H, Olson CA, Wu NC, Chu V, Truong S, Chen Z, Zhong W, Su SY, W TT, Chen SH, Sun R. Exploring distinct fitness landscapes of an HCV protein under anti-viral drug treatment.

VITA

2007	B.S. Polymer Science and Engineering Zhejiang University, China
2009	Advancement to Candidacy for Ph. D. Molecular and Medical Pharmacology University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

1. Qi H, Su SY, Chen Z, Arumugaswami V, Truong S, Chu V, Remenyi R, Young A, Chen SH, Lin CY, and Sun R. High-resolution Profiling HCV Genome by Combining Random Insertion Mutagenesis with Next-generation Sequencing. Oral presentation. 18th International Symposium on Hepatitis C Virus and Related Viruses. Seattle, Washington, September 2011.

2. Qi H, Su SY, Chen Z, Wu NC, Arumugaswami V, Truong S, Chu V, C. Olson CA, Remenyi R, Young A, Wu TT, Chen SH, Lin CY, and Sun R. A genetic platform for systematical mapping of anti-interferon functions in the Hepatitis C virus genome. *Manuscript in submission*.

3. Qi H, Chu V, Wu NC, Truong S, Chan E, Sun R. P7 is a novel immune evasion protein of HCV. *Manuscript in preparation*.

4. Qi H, Olson CA, Wu NC, Chu V, Truong S, Chen Z, Zhong W, Su SY, W TT, Chen SH, Sun R. Exploring distinct fitness landscapes of an HCV protein under anti-viral drug treatment. *Manuscript in preparation.*

5. Young A., Wu NC, Chen SH, Lu IHs, Lin CY, Chin RG, Al-Mawsawi LQ, Olson CA, Qi H, Gong D, Nelson SF, Liu Y, Wu TT, Sun R. Single nucleotide-resolution functional profiling of the Influenza A virus genome. *Manuscript submitted*.

6. Olson CA, Adams JD, Takahashi TT, Qi H, Howell SM, Wu TT, Roberts RW, Sun R, Soh HT. Rapid mRNA Display Selection of an IL-6 Inhibitor Using Continuous Flow Magnetic Separation. *Angew Chem Int Ed Engl*. 2011 Jul 14.

7. Kim S, Kim YC, Qi H, Su K, Morrison SL, Chow SA. Efficient Identification of Human Immunodeficiency Virus type I Mutants Resistant to a Protease Inhibitor Using a Random Mutant Library. *Antimicrobial Agents and Chemotherapy*. 2011 Nov; 55(11): 5090-8.

8. Weng D, Qi H, Wu TT, Yan M, Sun R, Lu Y... Visible light powered self-disinfecting coatings for influenza viruses. *Nanoscale* 2012 Apr 28; 4(9): 2870-4.

CHAPTER 1

INTRODUCTION

1.1 Hepatitis C virus

Hepatitis C virus (HCV) is an enveloped single positive-strand RNA virus encoding a polyprotein of around 3000 amino acids [1, 2]. The existence of the virus was first postulated as a nonA, nonB (NANB) infectious agent, which was recognized as a major cause of transfusion-acquired hepatitis in the 1970s [3-5]. However, the virus was not identified until 1989, when a group of scientists led by Michael Houghton cloned the cDNAs, the expression of which was immunoreactive with serum antibodies from patient with chronic NANB hepatitis [6]. This discovery facilitated subsequent characterization of the virus, and led to the establishment of diagnostic assays to screen blood supplies, which was one of the primary viral transmission routes [1]. Nonetheless, an estimated 130-170 million people worldwide are HCV-positive, and chronic infection remains a significant health concern [2].

Most acute HCV infections are asymptomatic, and about 15% to 30% of them are resolved spontaneously, while the other 70% to 85% of patients develop persistent infection [7-9]. Since the liver is the major site of HCV replication, patients with chronic infection are at high risk for developing severe liver diseases, including liver injury, progressive fibrosis, cirrhosis and hepatocellular carcinoma [10, 11].

The HCV genome is composed of two un-translated regions (5'UTR, 3'UTR), three structural proteins (core, E1, E2) and seven nonstructural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) [12]. The linear viral RNA genome, which is complexed with multiple copies of nucleocapsid protein (core), is housed in an enveloped viron composed with a lipid bilayer and two envelope glycoproteins (E1 and E2) [1]. The virus particles are estimated to range from 30 nm to 80 nm in diameter, based on filtration [13, 14]. Recent electron microscopic imaging of virus particles produced in tissue culture revealed the spherical virons with 50nm in diameter [1].

1.2 HCV experimental systems

After the landmark report of HCV identification, the major limitation has been the inability to efficiently culture the virus in tissue culture and lack of small animal models. It was not until 1996, when researchers realized that the original HCV cDNA clones missed a highly conserved 3'UTR that the complete HCV genome sequence was reconstructed and the derived RNA transcripts was shown to produce infectious virus in chimpanzee [15-18]. In the absence of a permissive tissue culture system, this infectious system has been taken as a useful tool to study virus evolution and immune response to viral infection *in vivo* [19]. These studies along with bioinformatics and biochemical analysis have defined the structure of HCV genome, polyprotein-processing mechanisms, protein topology, and some protein functions. However, the fact that the transcript viral RNA failed to replicate after transfection into cultured cells delayed the HCV research and hampered high-throughput chemical screens for antiviral drug development.

In 1999, Lohmann et al established the first subgenomic RNA replication system, by expressing the HCV nonstructural genes of genotype 1b (Con-1) via an internal ribosome entry site (IRES) from encephalomyocarditis virus (EMCV) with the selectable neomycin resistance gene in a human hepatoma cell line (Huh-7) [20]. Following the transfection of RNA, HCV nonstructural proteins were expressed in G418-resistant cells and persistent HCV RNA replication was established. This system was then further optimized by Blight et al., who isolated a subclone of HCV replicon cell line (Huh-7.5) [21] with enhanced permissiveness for HCV RNA replication due to the defect in the retinoic-acid inducible gene I (RIG-I) of the cell [22]. Sequencing of viral RNA revealed several cell culture-adaptive mutations in nonstructural protein regions that enhance RNA replication [23]. Further studies subsequently constructed genotype 1a [24], 2a replicon systems [25-27] in different cell lines, including 293, HeLa, HepG, and a mouse hepatoma line, suggesting that RNA replication is not restricted to hepatocytes or human cells. Intriguingly, genotype 2a replicon shows robust replication efficiency without

further acquiring adaptive mutations [25-27]. These replicon systems provided an important tool for studying virus-host interactions and screening for antiviral compounds. Unfortunately, incorporation of adaptive mutations in the full-length genomes in genotype 1a and 1b still yielded no infectious virons, which hampered the studies at certain steps of virus life cycle [24, 28-30].

To overcome this challenge, researchers developed model systems that allowed for examining the roles of HCV glycoproteins in virus entry and identifying cellular receptors. Retrovirus pseudoparticles displaying HCV-glycoproteins (known as HCVpp) were produced by co-transfecting: 1) the Gag-Pol proteins of retroviruses, 2) a retroviral genome harboring a reporter gene for virus entry detection such as green fluorescent protein or luciferase protein, and 3) the unmodified HCV envelope glycroproteins (E1 and E2) into 293T cells [31, 32]. The entry of HCVpp was shown to be a glycoprotein-dependent event and was easily detected by measuring reporter gene activity. This surrogate expression system allowed for demonstration of critical cellular receptors required for virus entry, but the inability to generate infectious virus particles in cell culture remained a big challenge. Examination of the adaptive mutations in NS5A, which were associated with reduced level of NS5A phosphorylation for efficient RNA replication [33, 34], led to the hypothesis that NS5A hypophosphorylation was essential for genome replication and that hyperphosphorylation was a negative regulator of this process, but critical for late viral life cycle events, such as assembly and packaging. In agreement with this idea, studies showed that incorporation of adaptive mutations failed to produce infectious particles although genome replication is robust in tissue culture [24, 28, 30, 35]. Furthermore, the adaptive mutations attenuated the infectivity of viruses in studies with chimpanzee [29]. All of the above results suggested that genotype 2a strain would be capable of producing infectious virus in cell culture due to the fact that it does not require additional adaptation for RNA replication. As predicted, a full-length genotype 2a strain, called JFH-1, derived from a Japanese patient, was able to produce infectious HCV viral particles (HCVcc) in Huh-7 cell lines

[36], and higher titer was obtained by more passages and using Huh-7.5 and its derivative cell line (Huh-7.5.1) cells [37]. A further improvement was done by the Rice group, who generated a chimeric genome using the structural region from another genotype 2a isolate, J6 and the nonstructural region from JFH1 to produce highly infectious viruses [38]. The viruses collected in the supernatant were shown to be infectious in naïve Huh-7 cells [38, 39], chimpanzees [40], and uPA-SCID mice carrying humanized liver grafts [40]. Afterwards, JFH-1 based chimera with genotype 1a were also established using the same idea with much lower titers [41]. The experimental systems have been widely used to facilitate HCV studies in understanding the virus life cycle, exploring virus-host interactions, and identifying novel therapeutic targets [12].

1.3 HCV life cycle

With the established HCV experimental systems, the virus life cycle events, including entry, RNA replication, assembly and packaging has been extensively examined [12]. Studies showed that HCV infection is initiated through binding of viral glycoproteins with several cell-surface receptors, including CD81, scavenger receptor class B, type I (SR-BI), claudin-1 (CLND1), and occludin [42-44]. Upon binding to the entry factors, the virus undergoes receptor-mediated endocytosis triggered by the low pH of the endosomal pathway. The nucleocapsid protein is uncoated to release the RNA genome into the cytoplasm. The viral genome is utilized as a template for translating polyprotein, which is further processed into 10 viral proteins by cellular signal peptidase, signal peptide peptidase and its own protease [1]. Following the structural protein region is a small membrane associated protein called p7. It is a putative ion channel protein and was shown to be critical for in *vivo* infection in chimpanzee [45]. The NS2 protein, which is the first nonstructural protein after p7, is not required for virus genome replication [20]. However, it harbors viral protease activity required for HCV protein maturity and plays important roles in virus assembly and packaging [46]. The nonstructural proteins, including NS3, NS4A, NS4B, NS5A and NS5B, form a replication complex with the positive-strand RNA genome in a

virus-derived ER-like membrane structure termed the membranous web to initiate genome replication through synthesis [47]. A negative-strand RNA is synthesized and serves as a template for multiple rounds of nascent, positive-strand RNA synthesis, which results in asymmetric accumulation of approximately ten positive-strands for each negative-sense genome. The newly synthesized RNA is packaged into nucleocapsid proteins and assembled with HCV glycoproteins to form a viron. The virion is matured and secreted through ER and secretary pathway to release from the infected cells, which completes the life cycle [12].

1.4 Success of HCV infection (evasion of innate antiviral responses by HCV)

Upon viral infection, host has developed various mechanisms to control virus replication. Innate immune responses are known as the first line of defense against pathogen infection [48]. They are usually activated via host recognition of molecule patterns expressed by the pathogens, which triggers signaling pathways to initiate an innate immune response to provide an immediate protection against infection [48, 49]. A common downstream output of these pathways is the secretion of interferons (IFNs), including type I IFNs (IFN-β and over a dozen of IFN-α subtypes), type II IFN (IFN-γ), and type III IFN (IFN-λ, also known as IL-29, IL-28A, and IL-28B). Type I and type III IFN further induces the expression of different but possibly overlapped classes of genes to establish antiviral states in cells [50, 51]. Recently, sequencing of IFN-λ3 (IL-28B) revealed that polymorphisms of the gene correlate with the outcome of HCV infection and the response to IFN-α treatment [52, 53].

Type-I IFN (IFN-α) has been clinically used to control HCV replication with an approximate 50% success rate in patients [2]. It's known that IFN responses can defend mammalian hosts from virus infection by blocking viral genome replication, degrading viral genome, inhibiting viral protein synthesis, and affecting viral protein post-translational processes to control virus replication at different steps [54]. Upon the recognition of viral infection, type I IFN expression is induced through two major independent pathways of host defense triggered

by dsRNA, namely retinoic acid-inducible gene I (RIG-I) [55] and Toll-like recepter (TLR3) signaling pathways [56]. The IFNs are secreted and bound to their receptors on the targeted cell surface. This results in the activation of the Jak/STAT pathway, where the signal transducer and activator of transcription (STAT) proteins are phosphorylated, dimerized and associated with IRF-9. The complex translocates to the nucleus and binds to the IFN-stimulated response elements (ISRE) within the promoter region of IFN-stimulated genes (ISGs), consequently inducing the expression of multiple antiviral effectors.

However, viruses have evolved to utilize different strategies to evade multiple layers of immune response to support efficient viral replication, which could minimize the antiviral efficacy of IFN-α therapy [57]. Several mechanisms have been proposed to explain the lack of response of HCV infection to IFN-α therapy [58]. One example is HCV protease NS3/4A blocking dsRNA-induced IFN production by cleaving the C-terminal region of IPS-1/Cardif [59], resulting in its subcellular redistribution to prevent downstream activation of IRF-3 and induction of IFN-β [60]. Similarly, NS3/4A protease can mediate proteolysis of TLR3 adaptor protein, TRIF, and attenuate the induction of IFN-β [61]. NS3/4A also inhibits activation of IRF-3 by interfering with the interaction between IRF-3 and TBK1 [62]. The HCV core protein has been shown to impede Jak-STAT signaling either through direct interaction with STAT1 [63] or through induction of SOCS3 expression, which is associated with poor treatment outcome in patients according to the clinical studies. Furthermore, studies also indicate that E2 [64] and NS5A [65, 66] interfere with dsRNA dependent protein kinase R (PKR) and abolish its antiviral activity. Interestingly, another group revealed that HCV hijacks the function of PKR, thereby differentially down-regulates the synthesis of antiviral proteins [67].

Although HCV infection evokes a broad range of innate immune responses, the virus still manages to survive in the host by evading recognition, blocking signaling transduction, or even completely counteracting the antiviral function of the immune effectors. It seems that during evolution, the virus has successfully adapted itself to its human host by evolving mechanisms to

modulate IFN induction, limit the expression of ISGs, and impair the antiviral activities of the interferon effectors, thereby allowing HCV to successfully evade the multilayered antiviral actions of the host response, contributeing to its persistence and resistance to therapy. Fully understanding the viral interface to host immunity is of terrific interest and importance for developing immunomodulators to better control viral replication.

1.5 Drug development for antiviral therapy

HCV infection has emerged as a major cause of human hepatocellular carcinoma (HCC), a major type of liver malignancy, with an estimated number of over 500,000 cases diagnosed every year. According to the International Agency for Research on Cancer, liver cancer is the fifth most common cancer in men and seventh in women [68]. Because of its poor prognosis, HCC is the third leading cause of cancer deaths. Epidemiological studies have shown that about 75% of HCC cases are strongly associated with chronic viral infection, and HCV is known to associate with about 47% HCC cases in developed countries [69]. Studies aiming to evaluate the effect of HCV elimination on the incidence of HCC indicate that more effective therapies to eradicate the virus will also reduce the risk of progression to hepatocellular carcinoma in a population with an estimated 17-fold risk increase compared to healthy individuals [70-72]. However, HCV infection, as pointed out by Laurent Gravitz, has been “Long overshadowed by HIV, the hepatitis C virus is starting to take its toll. And the heat is on to find and treat those affected” [73]. There are about 80% of HIV-positive patients in the United States that are aware of their viral infection, but only 30% of awareness in HCV-infected patients and approximately 5% worldwide, according to the data from Centers for Disease Control and Prevention. In addition, neither prophylactic nor therapeutic vaccine is available for HCV infection. Before the recent approval of the two protease inhibitors [74], the standard treatment has been combination therapy with interferon (IFN-α) and ribavirin which only offers a limited response

rate [75] but with severe side effects. Therefore, there is a pressing need for developing novel IFN-free antiviral therapeutics with strong potency and high specificity.

With the broadening knowledge of HCV replication and function of viral proteins, numerous attempts have been made to target the enzyme proteins, e.g. protease NS3/4A and the RNA-dependent RNA polymerase NS5B as direct-acting antiviral treatments [76]. Two protease inhibitors (Telaprevir and Boceprevir) have already been approved and show significantly improved antiviral function when combined with IFN in therapy [74]. Genotype 1 is the most prevalent virus in the US and Europe and is least responsive to the standard treatment (PEG-IFN/ribavirin). Luckily, the responsive rates have been significantly increased (from 20-30% to 70-80%) with the combination therapy of the two new protease inhibitors and IFN-α [77].

With the recent establishment of HCV replicon and HCVcc systems as reviewed [12], reverse pharmacology has been widely adopted to screen for small molecules that inhibit viral replication with unknown mechanisms. Daclatasvir, which is also known as BMS790052, is a new class of antiviral drug discovered by *Bristol-Myers Squibb (BMS).* Daclatasvir effectively inhibits the replication of HCV by blocking the function of NS5A at pM concentrations [78]. The drug is currently in phase III clinical trial and giving promising preliminary results without clearly defined working mechanism.

Each step of the viral cycle can potentially be targeted for drug development. Many viral replication inhibitors that are identified either through rational drug design or large-scale screen, are in (pre) clinical development [76].

However, in theory, the high error rate of NS5B polymerase, which can be up to 10^{-3} base/genome site/year, in combination with high turnover rate and viral load (replication ranging between 10^{10} to 10^{12} virons per day and the half life of 2 to 3 hours) [79], creates an extremely large and divergent genetic background, from which resistant variants can arise and dominate the population under a drug-selective pressure. Studies showed that there are preexisting drug-resistance mutations in the population, which then become dominant and generate resistance to

the compounds quickly [80]. This is in agreement with clinical observation of drug-resistance mutations emerging within days of therapy initiation with protease inhibitors, suggesting resistance preexists *in vivo* as well [81]. Thus, it is of interest and importance to design antiviral drugs against host regulators that are critical for virus replication, which potentially have much higher genetic barrier to resistance than that of the virus. As described earlier in this chapter, CD81 is one of the four essential host factors for HCV entry. Neutralizing antibodies against CD81 have been shown to block HCV infection *in vitro* and *in vivo* [82] Another recently identified host factor (phosphatidyldinositol-4-kinaseIIIα, PI4KIIIα) [83-86] required for HCV replication is now another host target for the treatment of HCV infection entering clinical trials [87]. Cyclophilin A is known as a cellular cis–trans-prolyl isomerase required for HCV RNA replication and assembly [88, 89]. Disruption of this interaction by cyclophilin A (cyclosporine A) inhibitor leads to abrogation of virus replication [88], which suggests cyclosporine A as a good candidate for antiviral therapy.

Better understanding of the virus has led to the abundance of candidate anti-HCV drugs that are under development. However, how to make a wise choice of drug for better efficacy with reduced side effects remains a challenge.

1.6 Bibliography

1. Lindenbach, B.D., H.J. Thiel, and C.M. Rice, *Flaviviridae: The viruses and Their Repliation*, in *Fields of Virology*, D.M. knipe and P.M. Howley, Editors. 2007, Lippincott Williams & Wilkins: Philadelphia. p. 1101.

2. Lemon, S.M., et al., *Hepatitis C Virus*, in *Fields of Virology*, D.M. knipe and P.M. Howley, Editors. 2007, Lippincott Williams & Wilkins: Philadelphia. p. 1253.

3. Feinstone, S.M., et al., *Transfusion-Associated Hepatitis Not Due to Viral Hepatitis Type A or B.* New England Journal of Medicine, 1975. **292**(15): p. 767-770.

4. Alter, H., et al., *TRANSMISSIBLE AGENT IN NON-A, NON-B HEPATITIS.* The Lancet, 1978. **311**(8062): p. 459-463.

5. Hollinger, F.B., et al., *Non-A, Non-B Hepatitis Transmission in Chimpanzees: A Project of the Transfusion-Transmitted Viruses Study Group.* Intervirology, 1978. **10**(1): p. 60-68.

6. Choo, Q.L., et al., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome.* Science, 1989. **244**(4902): p. 359-62.

7. Lehmann, M., et al., *High rate of spontaneous clearance of acute hepatitis C virus genotype 3 infection.* Journal of Medical Virology, 2004. **73**(3): p. 387-391.

8. Kamal, S.M., et al., *Peginterferon Alfa-2b Therapy in Acute Hepatitis C: Impact of Onset of Therapy on Sustained Virologic Response.* Gastroenterology, 2006. **130**(3): p. 632-638.

9. Thomas Dl, A.J.R.R.M. and et al., *The natural history of hepatitis c virus infection: Host, viral, and environmental factors.* JAMA: The Journal of the American Medical Association, 2000. **284**(4): p. 450-456.

10. El-Serag, H.B., *Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma.* Gastroenterology, 2012. **142**(6): p. 1264-1273.e1.

11. Lavanchy, D., *The global burden of hepatitis C.* Liver International, 2009. **29**: p. 74-81.

12. Tellinghuisen, T.L., et al., *Studying Hepatitis C Virus: Making the Best of a Bad Virus.* Journal of Virology, 2007. **81**(17): p. 8853-8867.

13. Yuasa, T., et al., *The Particle Size Of Hepatitis C Virus Estimated By Filtration Through Microporous Regenerated Cellulose Fibre.* Journal of General Virology, 1991. **72**(8): p. 2021-2024.

14. He, L.-F., et al., *Determining the Size of Non-A, Non-B Hepatitis Virus by Filtration.* Journal of Infectious Diseases, 1987. **156**(4): p. 636-640.

15. Kolykhalov, A.A., et al., *Transmission of Hepatitis C by Intrahepatic Inoculation with Transcribed RNA*. Science, 1997. **277**(5325): p. 570-574.

16. Yanagi, M., et al., *Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a,Äâchimpanzee.* Proceedings of the National Academy of Sciences, 1997. **94**(16): p. 8738-8743.

17. Kolykhalov, A.A., S.M. Feinstone, and C.M. Rice, *Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA.* J Virol, 1996. **70**(6): p. 3363-71.

18. Kolykhalov, A.A., et al., *Hepatitis C Virus-Encoded Enzymatic Activities and Conserved RNA Elements in the 3,Ä⩽ Nontranslated Region Are Essential for Virus Replication In Vivo.* Journal of Virology, 2000. **74**(4): p. 2046-2051.

19. Bukh, J., *A critical role for the chimpanzee model in the study of hepatitis C.* Hepatology, 2004. **39**(6): p. 1469-1475.

20. Lohmann, V., et al., *Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line.* Science, 1999. **285**(5424): p. 110-113.

21. Blight, K.J., J.A. McKeating, and C.M. Rice, *Highly Permissive Cell Lines for Subgenomic and Genomic Hepatitis C Virus RNA Replication.* Journal of Virology, 2002. **76**(24): p. 13001-13014.

22. Sumpter, R., Jr., et al., *Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I.* J Virol, 2005. **79**(5): p. 2689-99.

23. Blight, K.J., A.A. Kolykhalov, and C.M. Rice, *Efficient Initiation of HCV RNA Replication in Cell Culture.* Science, 2000. **290**(5498): p. 1972-1974.

24. Blight, K.J., et al., *Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture.* J Virol, 2003. **77**(5): p. 3181-90.

25. Date, T., et al., *Genotype 2a hepatitis C virus subgenomic replicon can replicate in HepG2 and IMY-N9 cells.* J Biol Chem, 2004. **279**(21): p. 22371-6.

26. Kato, T., et al., *Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon.* Gastroenterology, 2003. **125**(6): p. 1808-17.

27. Kato, T., et al., *Nonhepatic cell lines HeLa and 293 support efficient replication of the hepatitis C virus genotype 2a subgenomic replicon.* J Virol, 2005. **79**(1): p. 592-6.

28. Ikeda, M., et al., *Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells.* J Virol, 2002. **76**(6): p. 2997-3006.

29. Bukh, J., et al., *Mutations that permit efficient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpanzees.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14416-21.

30. Pietschmann, T., et al., *Persistent and transient replication of full-length hepatitis C virus genomes in cell culture.* J Virol, 2002. **76**(8): p. 4008-21.

31. Bartosch, B., J. Dubuisson, and F.L. Cosset, *Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes.* J Exp Med, 2003. **197**(5): p. 633-42.

32. Hsu, M., et al., *Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7271-6.

33. Evans, M.J., C.M. Rice, and S.P. Goff, *Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication.* Proc Natl Acad Sci U S A, 2004. **101**(35): p. 13038-43.

34. Appel, N., T. Pietschmann, and R. Bartenschlager, *Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain.* J Virol, 2005. **79**(5): p. 3187-94.

35. Yi, M. and S.M. Lemon, *Adaptive mutations producing efficient replication of genotype 1a hepatitis C virus RNA in normal Huh7 cells.* J Virol, 2004. **78**(15): p. 7904-15.

36. Wakita, T., et al., *Production of infectious hepatitis C virus in tissue culture from a cloned viral genome.* Nat Med, 2005. **11**(7): p. 791-6.

37. Zhong, J., et al., *Robust hepatitis C virus infection in vitro.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9294-9.

38. Lindenbach, B.D., et al., *Complete replication of hepatitis C virus in cell culture.* Science, 2005. **309**(5734): p. 623-6.

39. Pietschmann, T., et al., *Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras.* Proc Natl Acad Sci U S A, 2006. **103**(19): p. 7408-13.

40. Lindenbach, B.D., et al., *Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro.* Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3805-9.

41. Yi, M., et al., *Production of infectious genotype 1a hepatitis C virus (Hutchinson strain) in cultured human hepatoma cells.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2310-5.

42. von Hahn, T. and C.M. Rice, *Hepatitis C Virus Entry.* Journal of Biological Chemistry, 2008. **283**(7): p. 3689-3693.

43. Liu, S., et al., *Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection.* J Virol, 2009. **83**(4): p. 2011-4.

44. Ploss, A., et al., *Human occludin is a hepatitis C virus entry factor required for infection of mouse cells.* Nature, 2009. **457**(7231): p. 882-6.

45. Sakai, A., et al., *The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences.* Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11646-51.

46. Grakoui, A., et al., *A second hepatitis C virus-encoded proteinase.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10583-7.

47. Egger, D., et al., *Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex.* J Virol, 2002. **76**(12): p. 5974-84.

48. Kawai, T. and S. Akira, *The roles of TLRs, RLRs and NLRs in pathogen recognition.* Int Immunol, 2009. **21**(4): p. 317-37.

49. Bowie, A.G. and L. Unterholzner, *Viral evasion and subversion of pattern-recognition receptor signalling.* Nat Rev Immunol, 2008. **8**(12): p. 911-922.

50. Marcello, T., et al., *Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics.* Gastroenterology, 2006. **131**(6): p. 1887-98.

51. Thomas, E., et al., *HCV infection induces a unique hepatic innate immune response associated with robust production of type III interferons.* Gastroenterology, 2012. **142**(4): p. 978-88.

52. Ge, D., et al., *Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance.* Nature, 2009. **461**(7262): p. 399-401.

53. Tanaka, Y., et al., *Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C.* Nat Genet, 2009. **41**(10): p. 1105-9.

54. Katze, M.G., Y. He, and M. Gale, Jr., *Viruses and interferon: a fight for supremacy.* Nat Rev Immunol, 2002. **2**(9): p. 675-87.

55. Saito, T., et al., *Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA.* Nature, 2008. **454**(7203): p. 523-7.

56. Katze, M.G., et al., *Innate immune modulation by RNA viruses: emerging insights from functional genomics.* Nat Rev Immunol, 2008. **8**(8): p. 644-654.

57. Gale, M., Jr. and E.M. Foy, *Evasion of intracellular host defence by hepatitis C virus.* Nature, 2005. **436**(7053): p. 939-45.

58. Lemon, S.M., *Induction and Evasion of Innate Antiviral Responses by Hepatitis C Virus.* Journal of Biological Chemistry, 2010. **285**(30): p. 22741-22747.

59. Meylan, E., et al., *Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus.* Nature, 2005. **437**(7062): p. 1167-72.

60. Loo, Y.-M., et al., *Viral and therapeutic control of IFN-beta promoter stimulator 1 during hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2006. **103**(15): p. 6001-6006.

61. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(8): p. 2992-2997.

62. Otsuka, M., et al., *Interaction between the HCV NS3 protein and the host TBK1 protein leads to inhibition of cellular antiviral responses.* Hepatology, 2005. **41**(5): p. 1004-1012.

63. Lin, W., et al., *Hepatitis C Virus Core Protein Blocks Interferon Signaling by Interaction with the STAT1 SH2 Domain.* Journal of Virology, 2006. **80**(18): p. 9226-9235.

64. Taylor, D.R., et al., *Inhibition of the Interferon- Inducible Protein Kinase PKR by HCV E2 Protein.* Science, 1999. **285**(5424): p. 107-110.

65. Gale Jr, M.J., et al., *Evidence That Hepatitis C Virus Resistance to Interferon Is Mediated through Repression of the PKR Protein Kinase by the Nonstructural 5A Protein.* Virology, 1997. **230**(2): p. 217-227.

66. Gale, M., et al., *Control of PKR Protein Kinase by Hepatitis C Virus Nonstructural 5A Protein: Molecular Mechanisms of Kinase Regulation.* Molecular and Cellular Biology, 1998. **18**(9): p. 5208-5218.

67. Garaigorta, U. and F.V. Chisari, *Hepatitis C Virus Blocks Interferon Effector Function by Inducing Protein Kinase R Phosphorylation.* Cell Host & Microbe, 2009. **6**(6): p. 513-522.

68. IARC, *Liver Cancer Incidence, Mortality and Prevalence Worldwide in 2008.* GLOBOCAN 2008 (International Agency for Research on Cancer), 2008.

69. Yang, J.D. and L.R. Roberts, *Hepatocellular carcinoma: A global view.* Nat Rev Gastroenterol Hepatol, 2010. **7**(8): p. 448-58.

70. Yoshida, H., et al., *Benefit of interferon therapy in hepatocellular carcinoma prevention for individual patients with chronic hepatitis C.* Gut, 2004. **53**(3): p. 425-430.

71. George, S.L., et al., *Clinical, virologic, histologic, and biochemical outcomes after successful HCV therapy: A 5-year follow-up of 150 patients.* Hepatology, 2009. **49**(3): p. 729-738.

72. Bartosch, B., et al., *Hepatitis C virus-induced hepatocarcinogenesis.* J Hepatol, 2009. **51**(4): p. 810-20.

73. Gravitz, L., *Introduction: A smouldering public-health crisis.* Nature, 2011. **474**(7350): p. S2-S4.

74. Enserink, M., *First Specific Drugs Raise Hopes for Hepatitis C.* Science, 2011. **332**(6026): p. 159-160.

75. Feld, J.J. and J.H. Hoofnagle, *Mechanism of action of interferon and ribavirin in treatment of hepatitis C.* Nature, 2005. **436**(7053): p. 967-972.

76. De Clercq, E., *The race for interferon-free HCV therapies: a snapshot by the spring of 2012.* Rev Med Virol, 2012.

77. Schlutter, J., *Therapeutics: New drugs hit the target.* Nature, 2011. **474**(7350): p. S5-S7.

78. Gao, M., et al., *Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect.* Nature, 2010. **465**(7294): p. 96-100.

79. Neumann, A.U., et al., *Hepatitis C Viral Dynamics in Vivo and the Antiviral Efficacy of Interferon-Œ± Therapy.* Science, 1998. **282**(5386): p. 103-107.

80. Robinson, M., et al., *Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10290-5.

81. Thompson, A.J., S.A. Locarnini, and M.R. Beard, *Resistance to anti-HCV protease inhibitors.* Curr Opin Virol, 2011. **1**(6): p. 599-606.

82. Meuleman, P., et al., *Anti-CD81 antibodies can prevent a hepatitis C virus infection in vivo.* Hepatology, 2008. **48**(6): p. 1761-8.

83. Berger, K.L., et al., *Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication.* Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7577-82.

84. Borawski, J., et al., *Class III phosphatidylinositol 4-kinase alpha and beta are novel host factor regulators of hepatitis C virus replication.* J Virol, 2009. **83**(19): p. 10058-74.

85. Tai, A.W., et al., *A functional genomic screen identifies cellular cofactors of hepatitis C virus replication.* Cell Host Microbe, 2009. **5**(3): p. 298-307.

86. Trotard, M., et al., *Kinases required in hepatitis C virus entry and replication highlighted by small interference RNA screening.* FASEB J, 2009. **23**(11): p. 3780-9.

87. DeRose, P., P.E. Thorpe, and D.E. Gerber, *Development of bavituximab, a vascular targeting agent with immune-modulating properties, for lung cancer treatment.* Immunotherapy, 2011. **3**(8): p. 933-44.

88. Ciesek, S., et al., *Cyclosporine A inhibits hepatitis C virus nonstructural protein 2 through cyclophilin A.* Hepatology, 2009. **50**(5): p. 1638-45.

89. Yang, F., et al., *Cyclophilin A is an essential cofactor for hepatitis C virus infection and the principal mediator of cyclosporine resistance in vitro.* J Virol, 2008. **82**(11): p. 5269-78.

90. Chatterji, U., et al., *HCV resistance to cyclosporin A does not correlate with a resistance of the NS5A-cyclophilin A interaction to cyclophilin inhibitors.* J Hepatol, 2010. **53**(1): p. 50-6.

CHAPTER 2

A GENETIC PLATFORM FOR SYSTEMATIC MAPPING OF ANTI-INTERFERON FUNCTIONS IN THE HEPATITIS C VIRUS GENOME

2.1 Abstract

To study the mechanisms by which HCV antagonizes the interferon (IFN) response, we have developed a high-throughput, quantitative genome-scale profiling platform that enables systematic mapping of HCV sequences critical for anti-IFN function at single amino acid resolution. In our two-tiered approach, a full genome screen for regions sensitive to IFN-α treatment when mutated was followed by an in-depth screen to characterize all possible amino acid changes in a selected region. First, a genomic IFN-α screen performed with a 15-nucleotide (nt) insertion library of mutant HCV identified 75 locations in the virus genome, which showed increased sensitivity to IFN-α treatment when mutated. The mutations conferring altered IFN-sensitivity are clustered in N-terminal core-, N-terminal p7-, and domain II & III of NS5A-coding regions. Second, we analyzed one of the regions using a novel saturation mutagenesis approach, where each codon in the N-terminus of core protein was substituted with all 19 other possible amino acids. This revealed that amino acids 17-27 are critical for counteracting IFN response. In particular, phenylalanine 24 is an important residue that is required to impede the phosphorylation of STAT1 thereby blocking IFN signaling. This profiling method can provide a basic and generalizable approach for viral genetics, enabling systematical mapping of protein functions at single amino acid resolution.

2.2 Introduction

Hepatitis C virus (HCV), with an estimated 170 million people persistently infected worldwide, has emerged as a major cause of human liver diseases, including chronic hepatitis, cirrhosis, and hepatocellular carcinoma (HCC) [1, 2]. However, neither prophylactic nor therapeutic vaccines are available for HCV infection. IFN-α remains an important component of the standard treatment, but has significant side effects, and offers only a limited response rate [3].

HCV is an enveloped positive-strand RNA virus encoding a polyprotein of around 3000 amino acids. The genome is comprised of two untranslated regions (5'UTR, 3'UTR), three structural proteins (core, E1, E2) and seven nonstructural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) [4]. Due to the limited genome space, many of the viral proteins have evolved multiple functions required for viral survival within the host. For example, in addition to the roles in viral replication [5], core, E2, NS3/4A and NS5A also encode immune evasion functions [6].

Virus-host interactions, such as the virus-IFN response, are very complex and likely involve multiple elements with various mechanisms [3, 7, 8]. The IFN response defends mammalian hosts from virus infection by controlling viral replication at multiple steps [8]. Sensing of viral infection triggers type I IFN expression, and secreted IFNs then bind to their receptors on the targeted cell surface. This results in activation of the Jak/STAT pathway, where STAT proteins are phosphorylated, followed by dimerization and association with IRF-9. The complex translocates to the nucleus and binds to IFN-stimulated response elements (ISRE) within the promoter region of IFN-stimulated genes (ISGs), inducing the expression of multiple antiviral effectors. However, viruses have evolved to circumvent, via different strategies, multiple layers of the IFN response, which likely dampens the antiviral efficacy of IFN-α therapy [7]. Several viral mechanisms have been proposed to explain the lack of response of HCV infection to IFN-α therapy [6], including modulation of IFN induction, reduction of ISGs expression, and impairment of the antiviral activities of IFN effectors [9-16]. However, there are still some gaps of our understanding of the anti-IFN functions encoded in the HCV genome.

To interrogate such complex interactions and provide a comprehensive, high-resolution and unbiased profile of the anti-IFN functional sequences in the virus genome, we established a platform by combining two mutagenesis strategies-a whole genome-scale random insertion mutagenesis [17, 18] and a synthetic saturation mutagenesis-with next-generation sequencing technology. For the whole genome approach, a highly complex library of HCV mutants was constructed by randomly inserting 15 nucleotides into the virus genome (Figure 2-1A). The library was passaged in cell culture in the presence or absence of IFN-α. Quantitative genetic profiling by next-generation sequencing allowed us to determine the relative fitness of all the mutants under these two conditions. We identified four regions on the virus genome that increased sensitivity to IFN-α treatment when mutated: N-terminal region of core, N-terminus of p7, Domain II and III in NS5A and 3'UTR. To pinpoint the residues in the N-terminus of core protein critical for its anti-IFN function, we conducted a secondary screen focused on core residues 2-28. Using synthetic oligonucleotides, we randomly mutated each residue into all 19 other possible amino acids, which resulted in the parallel analysis of 540 amino acid variants with 864 codon variations. The fitness of each variant with/without treatment of IFN-α was also characterized by next-generation sequencing. We found that amino acids at position 17-27 are critical for counteracting the IFN response, and that phenylalanine 24 is an essential residue for inhibition of STAT1 phosphorylation thereby blocking IFN signaling transduction.

2.3 Results

2.3.1 High-resolution profiling of the HCV genome.

To systematically profile the HCV genome in an unbiased manner, we constructed a mutant library by *in vitro* Mu transposon-mediated random insertion mutagenesis on a plasmid carrying the HCV genome (pFNX-HCV, a genome that we chemically synthesized based on the

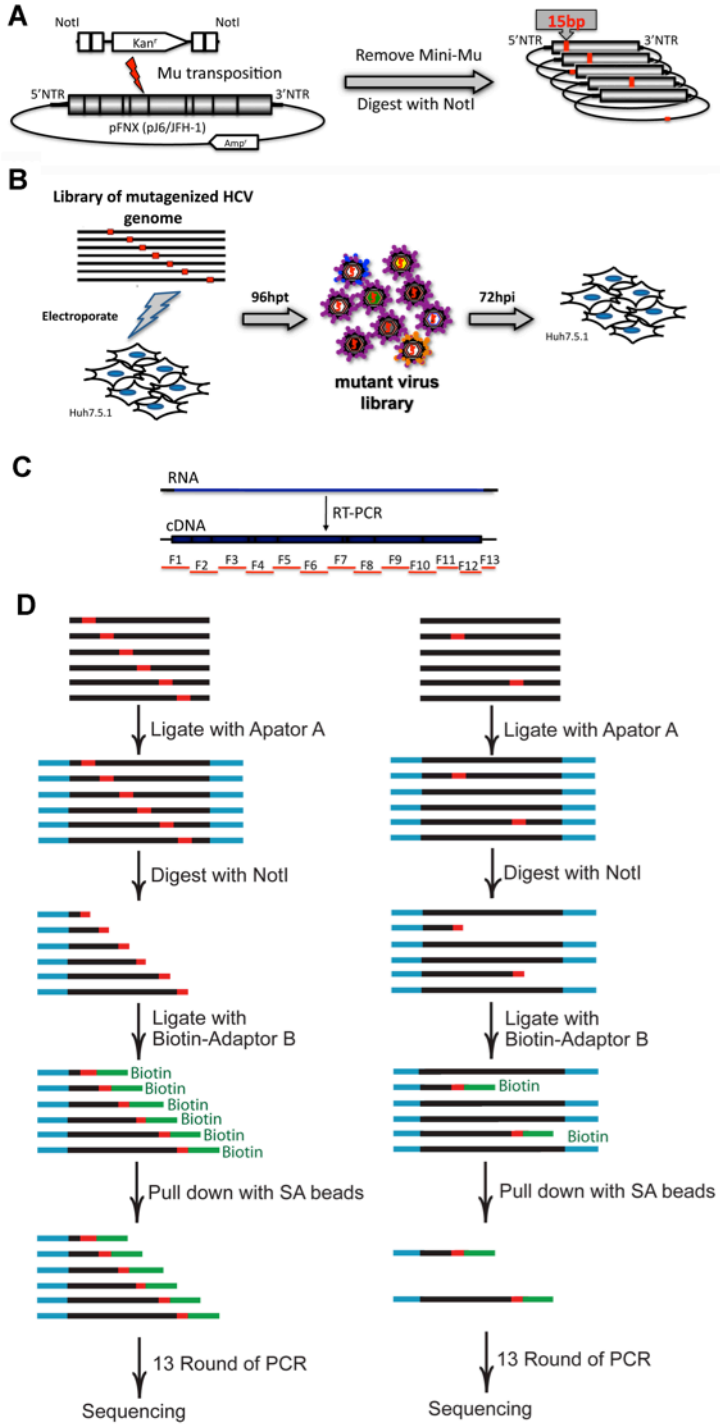


Figure 2- 1. A schematic diagram of HCV genome profiling method. (A) The mutant library was constructed by *in vitro* Mu transposon mediated random insertion mutagenesis on the plasmid carrying the HCV genome (pFNX-HCV). (B) The mutant plasmid library was linearized and transcribed into RNA *in vitro*, followed by electroporation into Huh-7.5.1 cells to reconstitute the mutant virus library. The virus library underwent two rounds of selection in Huh-7.5.1 cells. (C) The input RNA and total RNA from transfected or infected cells were isolated and reverse transcribed with

random hexamer. The cDNA then was used as PCR template for amplifying 13 fragments spanning the entire virus genome. (D) The purified PCR products were fragmented and ligated with sequencing adaptor A. After digestion with NotI enzyme, the library samples were ligated with the biotinylated sequencing adapter B via the NotI overhang. The biotinylated ligation products were then purified with SA beads and further enriched with PCR for adaptor-modified fragments for sequencing.

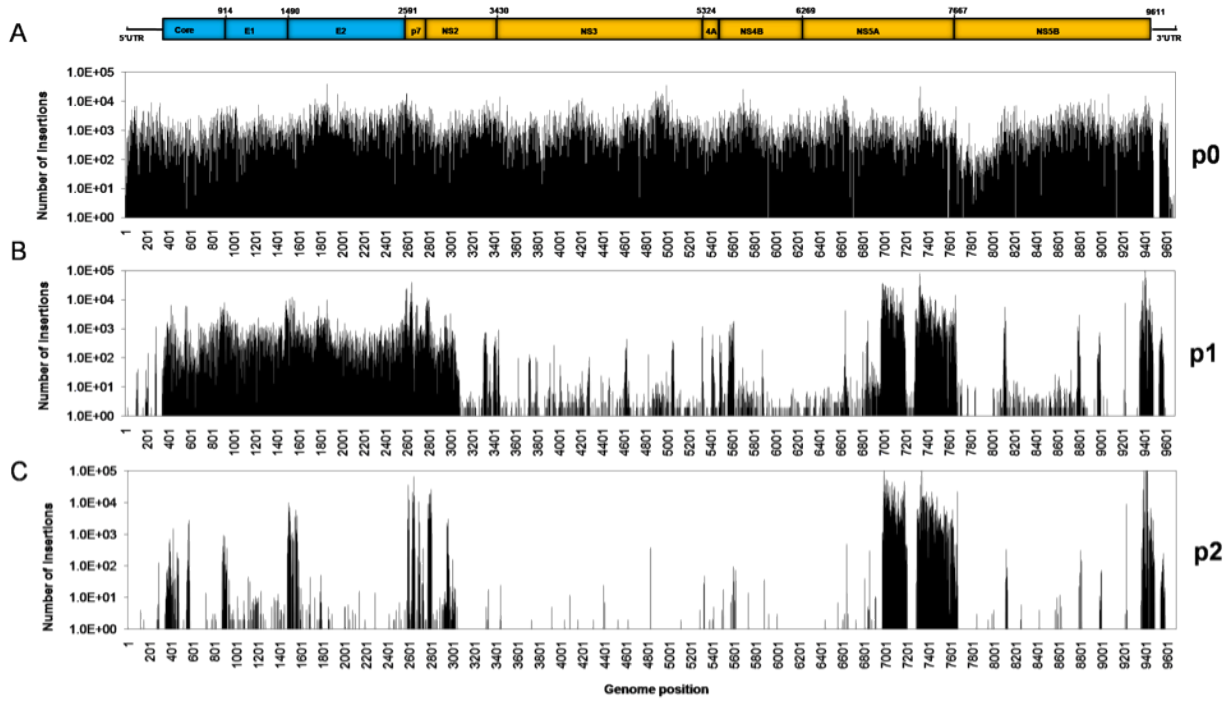


Figure 2- 2. Frequency of each mutant indicated by different insertion sites, in the input and selected libraries. The total RNA isolated from the transfected cells at 96 hours post transfection (96hpt) was analyzed as pool 1 (p1) and RNA from the infected cells at the end as pool 2 (p2). The in vitro transcribed RNA mutant library served as the input library pool 0 (p0). It is calculated from the sequencing data and normalized to account for differences in the total number of sequencing reads obtained from each library sample. X-axis represents the insertion position. Y-axis shows the number of each insertion detected by sequencing. The histogram panels show the frequency of mutants in p0 (A), p1 (B), and p2 (C). The schematic picture above the histogram shows the FNX-HCV virus genome composition. Blue is from J6 strain and yellow is from JFH1 strain.

chimeric genotype 2a clone, J6/JFH1) (Figure 2-1A). After removing the coding sequence in the transposon, a 15-nt insertion consisting of a NotI site and a 5-nt duplication from the targeted virus sequence remained, and randomly distributed throughout the virus genome. After reconstitution of the virus library, we passaged it in Huh-7.5.1 cells for two rounds at MOI 0.2 (Figure 2-1B). We then determined the fitness of individual mutants by analyzing the frequency of each mutant virus, indicated by different insertion sites, in various populations (p0, p1 and p2) with next-generation sequencing technology after enriching for the insertion sites harboring the NotI recognition sequence (Figure 2-1D). Our data showed that the insertion sites were uniformly distributed throughout the entire virus genome in the input library, except for the polyA region (Figure 2-2A). In contrast, mutants isolated from transfected cells at 96hpt (p1) showed that several regions, especially the non-structural proteins, did not tolerate insertions (Figure 2-2B), indicating their importance for viral genome replication, which is consistent with the known role in formation of replication complex [5]. After two rounds of infection in the cells (p2), we observed the relative loss of mutants in structural proteins (Figure 2-2C), suggesting that these domains are important for virus packaging and entry but not required for genome replication. Out of 8347 insertions identified in 9517 nucleotides of the genome, 600 were tolerated, 696 were attenuated, and 7051 were lethal for virus replication (SI Figure 2-1). It is noteworthy that mutations in the envelope glycoproteins, E1 and E2 are well tolerated for genome replication, which is in agreement with their function for virus entry [19]. However, the majority of the mutants are lethal for a complete virus life cycle, with the exception of a few residues in C-terminus of E1 (379-386) and N-terminus of E2 (384-390, 401-408). This is remarkably consistent with the definition of the region as the hypervariable region 1 (HVR1) [20, 21].

2.3.2 Constructing the tagged viruses based on the profiling data.

Profiling of the entire virus genome with 15-nt random insertions also provided a comprehensive overview of tolerated regions in the genome, facilitating the construction of epitope-tagged viruses. We constructed 5 viruses expressing either HA-, 6×His-, or Flag-tag epitopes within four different viral proteins (Core, E2, p7 and NS5A) (Figure 2-3A). Genome replication of these mutant viruses was not significantly affected by the insertion of the tags compared with wild type (WT) and a RNA polymerase-null mutant negative control virus (NS5B/GNN, with the catalytic residues GDD mutated to be GNN [22]) at 4, 48, and 96hpt (Figure 2-3B). Viral particle production at 48, 72 and 96hpt was comparable with WT, with the exception of core-HA tagged virus (Figure 2-3C). After a few rounds of adaptive passaging in cell culture, higher titer (10^4ffu/ml) was attainable. We also constructed two p7-tagged viruses without severely attenuating viral replication. P7 protein expression was visualized by immunofluorescence microscopy using anti-HA antibody in infected cells, which were confirmed by anti-E2 antibody (CBH5) [23] (Figure 2-3D). Previous efforts to construct growth-proficient viruses tagging majority of viral proteins for high-resolution microscopy have been challenging; an immediate benefit of our profiling data is information enabling the generation of reagents facilitating studies of HCV, by enabling tracking each protein and studying protein-protein interactions.

2.3.3 IFN-α selection of the mutant library revealed four IFN-hypersensitive domains.

After demonstrating the robustness of the genome-wide mutagenesis-profiling platform, we then utilized the approach to systematically identify the domains in the virus genome essential for circumventing the antiviral state induced by type I IFN. The library of mutant viruses was selected for two rounds under IFN-α treatment at 1U/ml (IC50, SI Figure 2-2A) (Figure 2-4A), and the fitness of each mutant was determined by next-generation sequencing. The effect of

A

Tagged protein /Insert location	Inserted amino acid sequence	Inserted tag sequence
Core/562/HA	$_{74}K$- CGRSK -S_{75}	$_{74}K$- YPYDVPDYAG -S_{75}
E2/1489/Flag	$_{383}A$- CGRNA -R_{384}	$_{383}A$- DYKDDDDK -R_{384}
p7/2644/HA	$_{768}G$- CGRNG -F_{769}	$_{768}G$- YPYDVPDYAG -F_{769}
p7/2644/His	$_{768}G$- CGRNG -F_{769}	$_{768}G$- HHHHHHG -F_{769}
NS5A/7147/Flag	$_{2269}S$- CGRTS -I_{2270}	$_{2269}S$- DYKDDDDK -I_{2270}

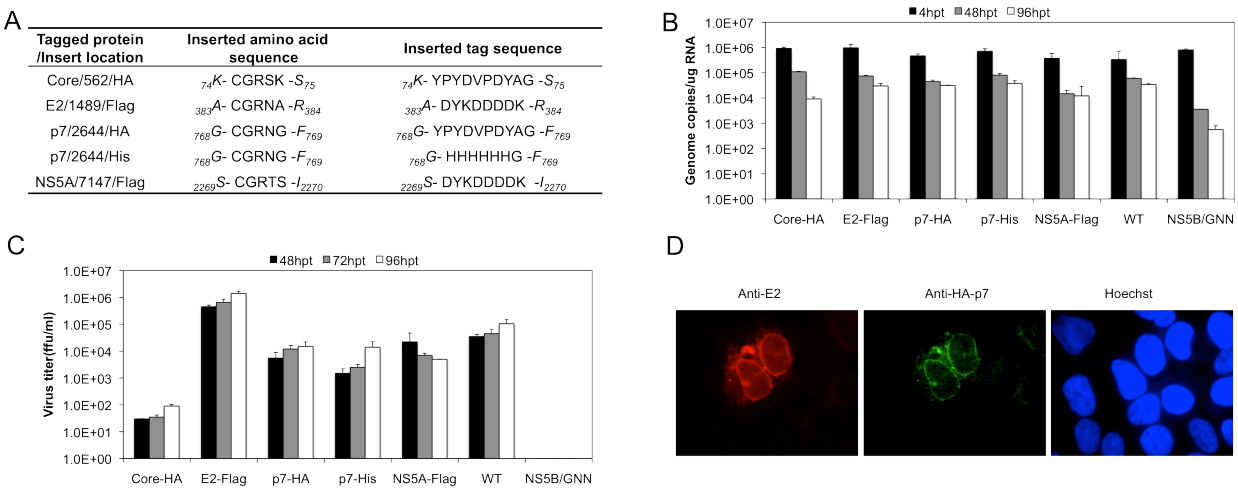


Figure 2- 3. Verification of the profiling method by constructing epitope tagged mutant viruses. (A) The epitope tags were inserted at the positions where 15-nt insertions were tolerated in the profiling screen. The sequences of 15-nt (5 amino acids) from the library and sequences of epitope tags are shown in the table. (B) The genome replication of the mutant viruses measured at 4hpt, 48hpt and 96hpt. (C) The production of epitope tagged mutant viruses particles measured at 48hpt, 72hpt and 96hpt. (D) The immunofluorescent staining of HA-tagged p7 (Green, secondary anti-mouse Alex488) can be detected in infected cells identified by anti-E2 antibody (CBH5) (Red, secondary anti-human Alex555). Hoechst 33342 (Blue) is used to stain the nuclei.

IFN-α on each mutant was evaluated by calculating the ratio of mutant virus abundance in IFN-α treated library (p2+IFN) to control (p2-IFN). The p-value was also determined for each mutant virus. We were particularly interested in IFN-α hypersensitive mutants based on the assumption that these viruses were inhibited due to the loss of their anti-IFN activity. With a cutoff for IFN-α hypersensitivity of ratio<0.5 and p-value<10-5, we found that mutations conferring increased IFN-sensitivity were clustered in four regions in the genome: N-terminus of core protein, N-terminus of p7, Domain II and III in NS5A protein and 3'UTR (Figure 2-4B).

2.3.4 Validation of the phenotype identified by IFN-α screen.

To verify the screen results, we constructed eight putative IFN-sensitive and one WT-like mutant virus (insertion at 7351) by inserting 15-nt at the positions identified in the screen. The nucleotide/amino acid sequences inserted in the virus genome are shown in SI Figure 2-3B. Infectious virus production of the mutants at 48 and 96hpt indicated that viral genome replication was not significantly affected by these insertions (Figure 2-4C). More importantly, consistent with our screen data, their replication was inhibited by IFN-α treatment quantitatively more than the WT-like control virus (Figure 2-4D). The phenotype was also observed by assaying the replication of viral genome by q-PCR (SI Figure 2-3B). Furthermore, we found that insertions in NS5A domain II and domain III were located within regions previously identified as interferon sensitivity-determining regions: (ISDR) [24, 25], PKR-binding domain (PKR-BD) [14, 15], variable region 3 (V3) and interferon/ribavirin resistance-determining region (IRRDR) [26-28] (SI Figure 2-3A). Collectively, these data demonstrate the reliability of the IFN screen results and more generally the validity and utility of our profiling platform.

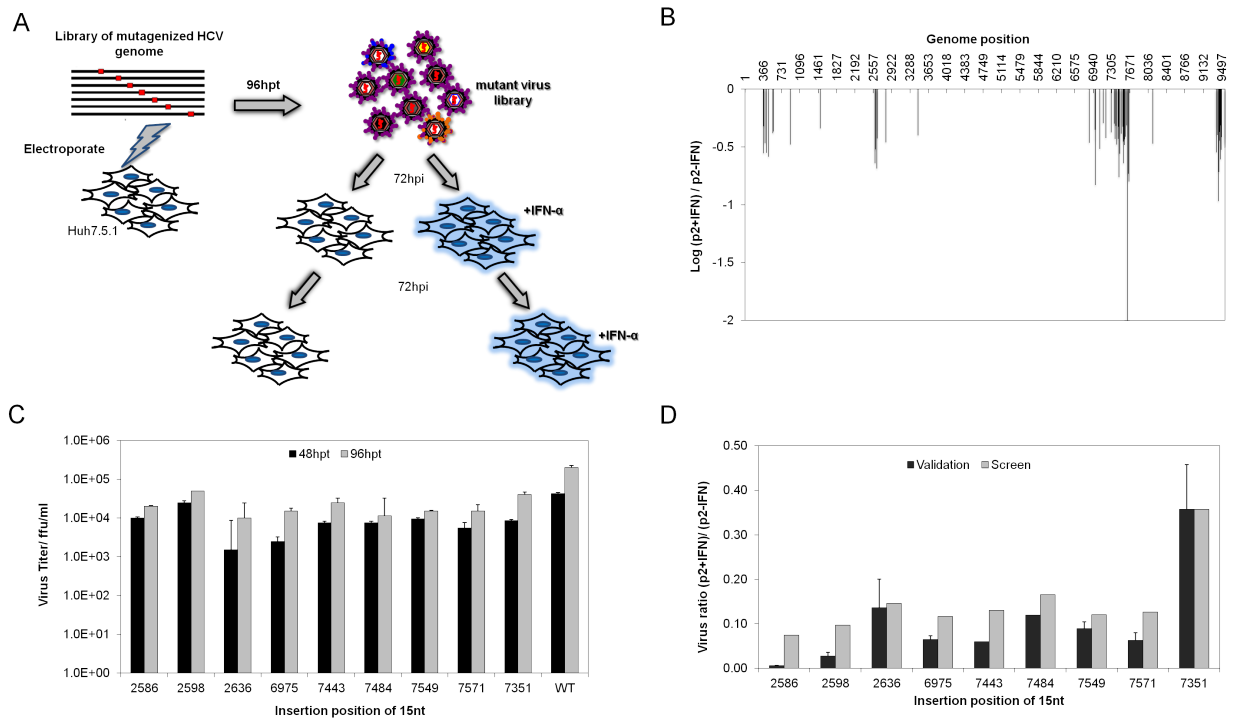


Figure 2- 4. The genomic screen of mutant virus library with IFN-α treatment and validation of individual mutant viruses. (A) The schematic diagram of selection to identify viral sequences critical for counteracting IFN responses. The library of mutant viruses was subjected to infect Huh-7.5.1 cells in the presence or absence of IFN-α treatment at the IC50 concentration (1U/ml, SI Fig. 2A). At 72 hours post infection, the supernatant was collected to infect naive cells for a second round selection (p2). (B) The IFN-sensitive mutations are clustered at 4 regions on the virus genome: N-terminus of core, N-terminus of p7, NS5A domain II and III and 3'UTR. X-axis indicates the positions of 15-nt insertion on the genome. Y-axis shows the ratio of mutant frequency with IFN-α treatment to without IFN-α treatment. (C) Eight putative IFN-sensitive mutants and one WT-like mutant (insertion at 7351) were constructed individually to characterize their sensitivity to IFN-α. The infectious virus particle production was measured at 48hpt (black) and 96hpt (grey). (D) The replication of the eight mutants in IFN-α treatment (black), and compared with the screen data (grey). Y-axis is the ratio of virus production in IFN-α treatment to control.

2.3.5 Profiling the critical residues of the core N-terminus.

To confirm the genomic IFN screen results and further map the critical residues in N-terminal region of core protein at single amino acid level, we constructed another mutant library where each residue at positions 2-28 was replaced by 3 continuous random nucleotides, which allowed for every possible amino acid substitution (Figure 2-5B). The mutant virus library was reconstituted, collected as pool 1 (p1) and selected for two rounds in cell culture in the presence (p2+IFN) or absence of IFN-α treatment (p2-IFN). The viral genome was extracted from the supernatant after each round of selection and used as template for reverse transcription followed by amplification of the N-terminal core region (amino acids 2-28) for sequencing. The mutant RNA library used for virus library reconstitution was assayed as pool 0 (p0). The fraction of WT virus in all of the pools provided an internal benchmark to determine the relative fitness of each mutant. Relative frequency was measured by calculating the ratio of mutant to WT percentage in each pool. The fitness of each mutant was evaluated as fold change of relative frequency compared with input library (p0) and presented in the form of heat maps (SI Figure 2-6). The fitness profile suggests that mutations at positions 2 through 16 are well tolerated, as expected considering the lack of structure in this region. Interestingly, positions 18-19, 24 and 26-28 are more sensitive to dissimilar nonsense mutations, which corresponds to a region of ordered secondary structure in nuclear magnetic resonance (NMR) experiments [29].

To evaluate the IFN-sensitivity of each mutant, we compared the relative frequency of each mutant with and without IFN-α treatment as a ratio relative to WT fraction (Figure 2-5C). The ratio represents the difference in IFN-sensitivity between mutants and WT. A ratio of 1 means that the mutant virus has the same IFN-sensitivity as WT, and a ratio smaller than 1 suggests stronger IFN-sensitivity than WT. Consistent with the whole genome profiling data, mutations in amino acids 17-27 confer IFN-sensitivity to the mutants, presumably due to the loss of anti-IFN function. Mutations within positions 2-16 had no effect. Notably, variations in IFN-sensitivity of mutants between and within positions demonstrate IFN-sensitivity does not

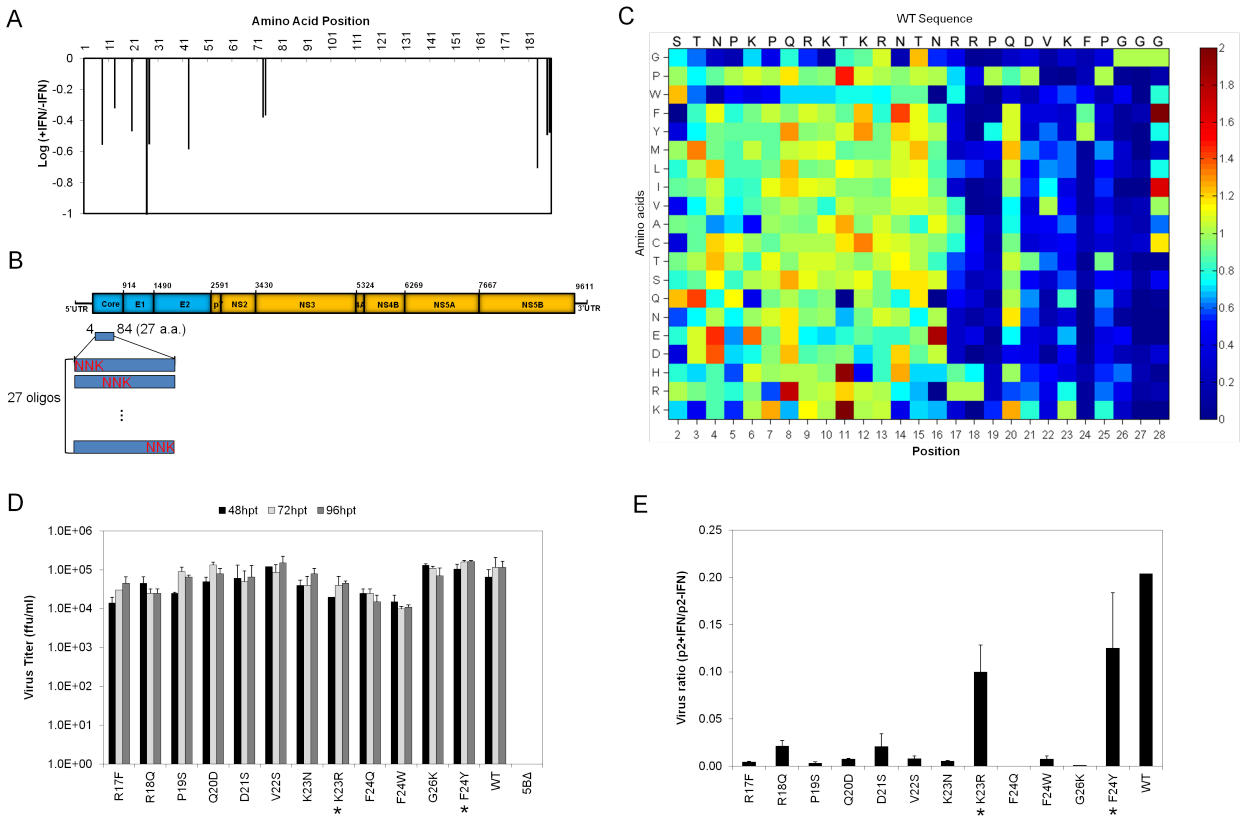


Figure 2- 5. Identification of the critical residues in the N-terminal core region for counteracting IFN-α response by saturation mutagenesis screen and validation. (A) The enlarged view of IFN screen results in the core protein region. (B) A schematic diagram of constructing saturation mutant library in region 2-28 of core protein. (C) The profile of IFN-sensitivity of each mutant virus relative to WT presented with a heat map. The WT sequence of each position is listed above the figure. X-axis is the position in the region. Y-axis is the list of all possible amino acid mutations. (D) Individual mutant viruses were constructed and the production of each mutant in the supernatant was assayed at 48hpt, 72hpt and 96hpt. NS5B/GNN is the replication defective control virus with catalytic residues GDD mutated to GNN. (E) The replication of the mutant viruses is inhibited significantly by IFN-α, compared with WT and the control viruses: K23R, F24Y. *: WT-like control viruses.

simply reflect replication fitness (SI Figure 2-7). Validation with individual mutant viruses further confirmed the screen. Mutations R17F, R18Q, P19S, Q20D, D21S, V22S, K23N, F24Q, F24W and G26K were constructed individually. Mutations K23R and F24Y, which did not cause IFN-hypersensitivity in the screen, were constructed as controls. Replication of the mutant viruses (R17F etc.) in Huh-7.5.1 cells was not significantly affected (Figure 2-5D), while the mutations caused increased sensitivity of these viruses to IFN-α treatment, compared with WT and control mutations K23R and F24Y (Figure 2-5E). This saturation mutagenesis analysis mapped the critical residues for counteracting IFN response at single amino acid level, which validates and complements the genome-scale screen.

2.3.6 Residue phenylalanine 24 is critical for inhibition of the IFN responsive signaling pathway.

With saturation mutagenesis analysis, we noticed that mutations at position 24 caused the mutant viruses to be hypersensitive to IFN treatment. To further understand how this residue facilitates the inhibition of the IFN signaling pathway, we constructed two mutant core proteins (F24Q and F24W) as well as the control construct (F24Y). Since STAT1 phosphorylation is one of the critical steps in the IFN signaling pathway, the effect of these proteins on counteracting the IFN response was assessed by examining STAT1 phosphorylation induced by IFN-α. We found that mutations F24Q and F24W were sufficient to abolish the inhibitory effect of core protein on STAT1 phosphorylation, but not F24Y (Figure 2-6A). Similarly, to evaluate the function of other residues, we constructed another three constructs, including a 4-amino-acid mutation construct (4a.a., mutations at position 21, 22, 23 and 25), 5a.a. (mutations at position 21, 22, 23, 24 and 25) and 8 a.a. (mutations at position 17, 19, 20, 21, 22, 23, 24 and 25), and examined their inhibitory effect on STAT1 phosphorylation. Intriguingly, the 4 a.a. variant retained some of the inhibitory function on the IFN signaling pathway, while an additional

mutation at position 24 completely abolished the inhibitory function of core protein (Figure 2-6B), suggesting that F24 plays a critical role in regulating STAT1 phosphorylation.

NMR structural analysis of amino acids 2-45 revealed the formation of a helix-loop-helix motif [29], and F24 locates at the last residue of the first α-helix (PDB entry 1CWX). A recent mutagenesis study of the region indicated that G33 is a conserved residue in core protein, presumably due to its steric interaction with neighboring amino acids (including F24) to maintain the overall conformation [30]. Since both F and Y contain a bulky aromatic side chain, we surmised that the motif maintained by this aromatic residue is critical for executing the anti-IFN function of core protein. As a result, loss of the aromatic side chain by substitution with Q or W, leads to hypersensitivity of the virus to IFN. Although W is also a bulky residue, it may be too large to be tolerated. As a consequence, viruses with mutation F24Q or F24W are more sensitive to IFN treatment compared with F24Y or WT (Figure 2-5E). This observation suggests an essential role for F24 in the anti-IFN function of core protein.

2.4 Discussion

Next-generation sequencing technologies have made significant impact on studies in various biological systems [31-34]. In this study, by combining whole genome-scale mutagenesis and regional saturation mutagenesis, we have developed a two-tiered, high-throughput, quantitative profiling platform to provide a comprehensive and unbiased insight into the domains critical for virus replication. The method is also sensitive enough to determine the role of individual residues in virus-host interactions.

To elucidate the regions that are essential for virus replication, we had previously conducted a study to map the JFH-1 genome [17]. However, the study was hindered by three factors: (1) The detection method, which is a capillary genotyping system, gives 1 or 2 nucleotides shift, causing uncertainty about insertion locations, reading frames, as well as

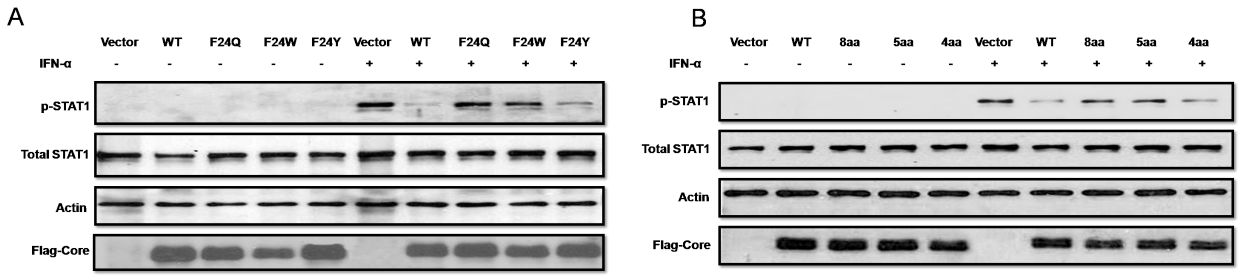


Figure 2- 6. Phenylalanine 24 is critical for blocking STAT1 phosphorylation induced by IFN-α. (A) 293T cells with WT or F24 mutant core proteins (F24Q, F24W, and F24Y) overexpressed were treated with 5U/ml of IFN-α for 30min, and phosphorylation on Tyr701 of STAT1 was detected by westernblot. The empty vector was included as control. (B) Phosphorylation of STAT1 in cells with WT or mutant core proteins (8a.a, 5a.a., and 4a.a.) overexpressed were induced by 5U/ml of IFN-α for 30min, and phosphorylation on Tyr701 was detected by westernblot.

inserted amino acid identities. (2) The input library started with 2399 mutant viruses, which leaves substantial region of virus genome uncharacterized. (3) The selection was performed by passing the virus RNA in transfected cells, due to the limited replication capacity of JFH-1 strain. This limited selection condition may result in missing the information of functional domains for some replications steps, e.g. egress and entry. To overcome the limitations, we adapted several approaches. First, to compensate for low mutation density, we constructed a high complex library of mutant HCV by inserting 15-nt at almost every base pair position in the virus genome. Next, we carried out the selection with multiple rounds of infection, allowing for investigating in a complete virus life cycle. Finally, by taking advantage of next-generation sequencing technology, we quantitatively characterized the library at high-resolution level. With the whole genome-scale mutagenesis, we are able to characterize 8347 different mutant viruses indicated by 8347 different insertion sites on the virus genome (9686bp), which is about 87% of the virus genome in a single experiment. The most tolerated region is within NS5A domain II and III, and the least tolerated the protein is NS3/4A (SI Figure 2-1C). Similar strategy has also been taken to map the temperature-sensitive mutants in Venezuelan equine encephalitis virus (VEEV) genome [18], which demonstrates the robustness of the whole genome profiling platform.

Utilization of next-generation sequencing allows us to precisely determine the insertion location and the phenotype of each mutant, allowing us to construct different epitope tagged viruses by inserting tags (HA/His/Flag) at the tolerated locations of viral proteins (core, E2, p7, NS5A). These mutant viruses will allow for tracking of the tagged proteins and can facilitate studies of virus-host interactions during infection. Thus, it has always been an interest for researchers to tag viral proteins and observe their localization during infection. However, successful tagging of proteins without affecting virus viability has remained a challenge [35]. It is worth pointing out that the positions previously identified to be tolerated for Flag-tag in E2 [36, 37] and GFP in NS5A [38] were also detected in our screen. Several attempts have been made

previously to tag p7 protein by fusing the tags to the protein either in an expressing vector [39-41] or in a replicon system [42]. Notably, none of these studies was able to generate viable viruses. In this study, with the precise determination of regions tolerated for insertions, we rationally constructed replication competent p7 tagged viruses, allowing for further exploration of the function of p7 in the context of viral replication.

The identification of domains critical for HCV in establishing resistance to IFN treatment is important for understanding the pathogenesis of HCV. Moreover, this could lead to novel antiviral treatment strategies to overcome viral resistance and benefit vaccine development efforts. Selection of the mutant virus library in the presence of IFN-α revealed four different regions as putative domains for the anti-IFN function. Intriguingly, the IFN-α sensitive mutations in NS5A were located at regions within ISDR, PKR-BD, V3 and IRRDR (SI Figure 2-3A), which is remarkably consistent with clinical studies [25-28, 43, 44]. P7, another protein identified as involved in IFN signaling in our screen, is an ion channel protein, comprised of two hydrophobic α-helices and a basic loop. In addition to its function in ion channel formation, p7 has also been implicated in mediating efficient assembly, release and production of infectious virions from host cells [45]. Here we found that mutations in p7, especially in the N-terminus of p7, cause IFN-hypersensitivity without affecting viral replication. Moreover, this phenotype can be rescued by overexpressing WT p7 protein in the infected cells. However, p7 does not diminish the ISRE activity induced by IFN-α, which suggests that p7 counteracts with the IFN response through regulating further downstream components. An *in vivo* study of experimentally infected chimpanzees demonstrated that HCV infection strongly induces the expression of ISGs in the liver [46]. HCV persists in the liver despite the induction of an antiviral state, suggesting that the virus encodes mechanisms to counteract the anti-viral functions executed by ISGs. This led us to hypothesize that p7 inhibits the function of the ISG(s) to facilitate robust viral replication despite the induction of an antiviral state by IFN.

Overexpression of core has been shown to interfere with the Jak-STAT signaling pathway through direct interaction with STAT1, blocking STAT1 phosphorylation [12]. Consistent with previous results, our whole genome profiling data suggests that N-terminal region of core protein is indispensable for its anti-IFN function. Moreover, saturation mutagenesis has identified the important role of amino acids 17-27 in regulating the IFN signaling pathway. These two mutagenesis methods together provide functional profiling images at different breadths and depths. They complement each other, by narrowing down the critical region(s) with entire genome profiling, and then characterizing in more detail the residues with saturation mutagenesis analysis.

This profiling method combines standard forward and reverse genetics, offering the possibility of identifying protein domains that are important for any particular selectable phenotype without previous knowledge of the protein (SI Figure 2-8). With saturation mutagenesis followed by deep sequencing, we demonstrated an extensive fitness landscape as well as functional profiling of amino acids 2-28 in core region, which is approximately 1% of the genome. The same approach has been successfully adopted to study the drug-protein interactions at another region, which comprises for 3% of the virus genome. By extending this mutagenesis approach to the entire genome (a library consisting of 102400 mutants), one would be able to functionally profile the whole virus genome and produce a fitness landscape with a heat map at the single amino acid level. In the present study, we focused on mapping sequences essential for HCV counteracting host immune response. The method is also generally applicable to other viruses and other cellular factors to provide fine-resolution functional maps. We anticipate that it will be a standard platform, which can be used in drug development to map drug-protein interactions, identify escape mutants and elucidate the mechanisms of newly developed drugs. Ultimately, a complete and densely mapped virus

genome reveals stretches with high genetic barriers to mutate, which can help to identify novel immune epitopes as targets for vaccine development.

2.5 Materials and methods

Cell culture, viruses and plasmids

The Huh-7.5.1 cell line was kindly provided by Dr. Francis Chisari from the Scripps Research Institute, La Jolla. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% of fetal bovine serum (FBS), 10mM non-essential amino acids (Invitrogen, Carlsbad, USA), 10mM HEPES, penicillin (100 units/ml), streptomycin (100 mg/ml), and 2mM L-glutamine at 37°C with 5% CO_2. 293T cells were grown in DMEM with 10% FBS, 100 units/ml of penicillin, and 100mg/ml of streptomycin.

pFNX-HCV is the plasmid we synthesized based on the chimeric sequence of J6/JFH1 virus. We introduced 7 nucleotide substitutions, resulting in synonymous mutations to the genome. To generate the 15-nt insertion library, the NotI site was also mutated. The construct and sequence are available upon request.

Wild-type core was cloned into an entry vector (pENTR, Invitrogen) by PCR amplification from pFNX-HCV, and recombined by an LR reaction into a modified destination vector resulting in a 3xFLAG epitope tag at the N-terminus of the protein. Individual mutant viruses with 15-nt insertion or epitope tags were constructed by site-directed PCR mutagenesis. Primer sequences for each 15-nt insertion or tagged virus are available upon request.

Construction and selection of 15-nt insertion mutant library

The plasmid carrying FNX-HCV was subjected to mu transposon mutagenesis and followed with selection on the plates with Kan and Amp for bacteria transformed with transposon-inserted plasmids. The plasmids were isolated and digested with NotI enzyme to remove the transposon

fragment and followed with ligation, which left 15-nt randomly inserted across the virus genome. We obtained a mutant plasmid library isolated from more than one million individual bacterial colonies, which is 100-fold greater than the number of all possible insertions. The plasmid library was used as a template to synthesize the mutant RNA library by *in vitro* transcription. We then reconstituted the virus library by electroporating the RNA into Huh-7.5.1 cells. The titer of the mutant library was $8.4x10^4$ focus forming units (ffu)/ml.

To identify the domains critical for viral replication, we passaged the library with 10 million of ffu virus particles in Huh-7.5.1 cells for two rounds at MOI 0.2 and isolated the total RNA from the infected cells at the end as pool 2 (p2) to evaluate the fitness of each mutant. The *in vitro* transcribed RNA mutant library served as the input library pool 0 (p0). The total RNA of transfected cells was collected as pool 1 (p1) at 96 hours post transfection (96hpt).

To systematically map the anti-IFN functional domains, the mutant virus library was subjected to infect naïve Huh-7.5.1 cells (36 million) or cells pretreated with 1U/ml of IFN-α 18h (36 million) before infection at M.O.I 0.2. The supernatant was collected at 72 hpi and passed on to infect naïve or IFN-α pretreated cells. The total RNA from infected cells was collected to evaluate the fitness of the mutant viruses.

Enrichment of insertion sites for next-generation sequencing

Total RNA was extracted from the libraries of transfected and infected cells by Trizol (Invitrogen) and reverse transcribed into cDNA with Superscript III (Invitrogen). The cDNA then was used as template for PCR to amplify 13 overlapping fragments covering the entire virus genome as described before[17]. The amplicons were mixed in an equimolar ratio to a final amount of 5ug. The mixtures were randomly fragmented by sonication and the ends were repaired with T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase to produce blunt, double stranded DNA fragments. The fragments were ligated to one of the adaptor, A before subjected to digesting with NotI enzyme (NEB). The digested fragments were ligated to the other biotin-

modified adaptor, B through NotI overhangs: GGCC with E. coli ligase at 16°C for overnight. The ligated products were purified by streptavidin-conjugated magnetic beads (Dynabeads® M-270 Streptavidin, Invitrogen), and enriched by PCR for the adapter-modified DNA fragments. Only the fragments with insertion sites could be amplified and sequenced, which allowed us to quantitatively and precisely determine the mutants in the libraries. The modified adaptor sequences (The barcodes used for distinguishing different libraries are underlined.):

A adaptor Top:

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT;

A adaptor Bot:

5'GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTT G;

B adaptors:

Not1-Top: 5'bio-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACT

Not 1-Bot: 5'GGCCAGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Not 2-Top: 5'bio-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGA

Not 2-Bot: 5'GGCCTCAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Not 3-Top: 5'bio-ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTG

Not 3-Bot: 5'GGCCCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Not 4-Top: 5'bio-ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAC

Not 4-Bot: 5'GGCCGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Immunofluorescent assay

At 48h post transfection, Huh-7.5.1 cells in 48-well plates were fixed with 4% paraformaldehyde for overnight. After 3 times wash with PBS, the cells were permeablized 0.05% Triton-x 100 for 10min and with blocked with 10% FBS, 3% BSA. The cells were incubated with human anti-E2 antibody (CBH5) and mouse anti-HA antibody (Sigma) at 4°C for overnight. Secondary

antibodies, goat anti-mouse IgG conjugated to Alex 488 (Invitrogen) and goat anti-human IgG conjugated to Alex 555 (Invitrogen) were added and incubated for 1 hr at room temperature. Cells were washed for 3 times with PBS and nuclei were stained with Hoechst 33342 (Invitorgen) before imaging.

Measuring virus titer

The virus titer was assayed by measuring the foci-forming unit of infectious virus particles per ml of supernatant. Culture supernatant taken from different time points, e.g. 48hpt, 72hpt and 96hpt were diluted in a 10-fold serial manner before applying onto the Huh-7.5.1 cells seeded in 96-well plates. 72hpt, the cells were fixed with 100% methanol and immunostained for core antigen with mouse monoclonal anti-core primary antibody C7-50 (Abcam, Cambridge, USA), followed by goat anti-mouse-Alex 488 secondary antibody. Virus titer was determined by counting the number of core antigen-positive foci at highest dilutions.

Quantitatively detection of viral genome with Q-PCR

Total RNA was isolated from transfected or infected cells by Trizol (Invitrogen) following the instruction of the manufacture. 500ng of RNA was used for reverse transcription using q-Script cDNA Synthesis Kit (Quantas). Primers used for HCV-specific qPCR was conducted by using primer pair: 5'-AGAGCCATAGTGGTCTGCG-3' and 5'-CTTTCGCAACCCAACGCTAC-3'; Actin: 5’-ACCTTCTACAATGAGCTGCG-3′ and 5′-CCTGGATAGCAACGTACATGG.

Construction of single amino acid mutant (saturation mutant) library

27 oligos with desired amino acid codon substituted with 3 continuous random nucleotides (NNK) were synthesized from IDT. The oligos contained a BtsI recognition site on each end, which allows generating “stick ends” matching the ends of the cassettes. The cassettes were established by amplifying the fragments (from pFNX-HCV) franking the region desired to be mutated with primers containing BtsI recognition site, and digested with BtsI enzyme (NEB) to

produce the “stick ends” matching the ones on the oligos. The oligos and the cassettes were ligated with T4 DNA ligase (Invitrogen) overnight at 16°C and purified with PCR columns (Invitrogen). The ligated product was subcloned into the pFNX-HCV vector via EcoRI and BsiWI restriction enzymes and transformed. Total approximately 50,000 colonies were collected for the library.

Selection of the saturation mutant library under IFN-α treatment

The mutant virus library (12ml) was subjected to infect naïve Huh-7.5.1 cells (4million) or cells pretreated with 1U/ml of IFN-α 18h before infection at M.O.I 0.2. The supernatant was collected at 72hpi and passed on to infect naïve or IFN-α pretreated cells. After two rounds of selection, the mutated region were amplified with PCR and processed with standard sample preparation protocol for HiSeq 2000 sequencing.

Enrichment of insertion sites for next-generation sequencing

Total RNA was extracted from the libraries of transfected and infected cells by Trizol (Invitrogen) and reverse transcribed into cDNA with Superscript III (Invitrogen). The cDNA then was used as template for PCR to amplify 13 overlapping fragments covering the entire virus genome as described before [17]. The amplicons were mixed in an equimolar ratio to a final amount of 5ug. The mixtures were randomly fragmented by sonication and the ends were repaired with T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase to produce blunt, double stranded DNA fragments. The fragments were ligated to one of the adaptor, A before subjected to digesting with NotI enzyme (NEB). The digested fragments were ligated to the other biotin-modified adaptor, B through NotI overhangs: GGCC with E. coli ligase at 16°C for overnight. The ligated products were purified by streptavidin-conjugated magnetic beads (Dynabeads® M-270 Streptavidin, Invitrogen), and enriched by PCR for the adapter-modified DNA fragments. Only the fragments with insertion sites could be amplified and sequenced, which allowed us to quantitatively and precisely determine the mutants in the libraries.

Determine 15 nt insertion sites from Illumina reads.

We mapped the HCV fragment from the reads to identify the base (N_5) prior to the 15nt sequence ($N_1N_2N_3N_4$**N_5**TGCGGCCGCA$N_1N_2N_3N_4N_5$, insertion is highlighted with underline) as the insertion site. Briefly, raw data was first processed to eliminate bad quality reads by SeqTrim [47]. In this study, different populations are linked with distinct barcodes; a successfully enriched HCV sequence read would be started with an 11-mer segment: 5'-DDDDggccgca $N_1N_2N_3N_4N_5$ . . . -3'. DDDD are the 4-mer barcode sequence and N is duplicated HCV sequence from the insertion sites. We marked the insertion site by the following rules: (1) for HCV sequence in sense direction, the fifth base after NotI site (N_5); (2) for HCV sequence in antisense direction, the first base after NotI site (N_1) is the complementary base of the insertion site. Binomial exact test was employed to calculate the p-value using a null hypothesis of 0.125.

Construction of saturation mutant library

27 oligos with desired amino acid codon substituted with 3 continuous random nucleotides (NNK) were synthesized from IDT. The oligos contained a BtsI recognition site on each end, which allows generating “stick ends” matching the ends of the cassettes. The cassettes were established by amplifying the fragments (from pFNX-HCV) franking the region desired to be mutated with primers containing BtsI recognition site, and digested with BtsI enzyme (NEB) to produce the “stick ends” matching the ones on the oligos. The oligos and the cassettes were ligated with T4 DNA ligase (Invitrogen) overnight at 16°C and purified with PCR columns (Invitrogen). The ligated product was subcloned into the pFNX-HCV vector via EcoRI and BsiWI restriction enzymes and transformed. Total approximately 50,000 colonies were collected for the library. A further detailed description of the materials and methods is provided in SI Materials and methods.

The full dataset, including whole genome mutagenesis, whole genome IFN screen, and saturation mutagenesis in the N-terminal core region will be available at http://eln.iis.sinica.edu.tw/hcv.

Sequence alignment for saturation mutagenesis library screen

Burrow-Wheeler Aligner was used to map the pair-end read by allowing 5 mismatches [48]. Sequencing error was corrected by reads pairing. SAMtools and BamTools were employed for sequence analysis [49, 50]. Custom Python script was created for the other downstream data analysis.

The full dataset, including whole genome mutagenesis, whole genome IFN screen, and saturation mutagenesis in the N-terminal core region will be available at http://eln.iis.sinica.edu.tw/hcv.

Westernblot for detection of STAT1 phosphorylation

293T cells were seeded at $5x10^4$ per well in 48-well plate. Wild type or mutant core constructs or empty vector was transfected into 293T cells with Lipofectamine 2000 (Invitrogen). 24 hours post transfection, the cells were treated with 2ug/ml of puromycin to select for transfected cells. 48 hours post treatment, the cells were treated with 5U/ml of IFN-α for 30min to induce the phosphorylation on STAT1. The other same set of cells was left untreated as control. Cell lysates were resolved by a 10% SDS-PAGE gel and transferred to a nitrocellulose membrane. The membranes were blocked (5% skim milk, 0.2% Tween-20 in PBS) and probed for phosphorylation of STAT1 on Tyr 701 (Cell signaling). The membrane was subsequently stripped to demonstrate equal loading of total STAT1 (Cell signaling) and beta-actin (Sigma). The expression of FLAG-core proteins was also detected with M2 mouse anti-FLAG antibody (Sigma).

2.6 Acknowledgements

We thank Dr. Francis Chisari (the Scripps Research Institute, La Jolla) for kindly providing the Huh-7.5.1 cell line. We are grateful to S. Foung for provision of the human anti-E2 antibody CBH5. We also thank Dr. Asim Dasgupta, Dr. Samuel French (UCLA) and Yong-Hoon Kim for their comments and suggestions on the manuscript. This work was supported by the following grants: National Natural Science Foundation of China (NSFC) 81172314, Margaret Early Medical Research Trust, and National Institute of Health AI078133 (RS), P30CA016042 (Jonson Comprehensive Cancer Center) and P30AI028697 (UCLA AIDS Institute/CFAR).

2.7 Supplementary information

A

Sample	pools
Input (RNA library)	p0
96hpt	p1
144hpi (round 2 of infection)	p2

B

	p0	p1	p2 (w/o IFN)
Number of mutant	8347	3880	1116
Total reads	6,415,885	3,048,022	4,537714
Average coverage/mutant	780	785	4066
Median reads/mutant	239	70	134

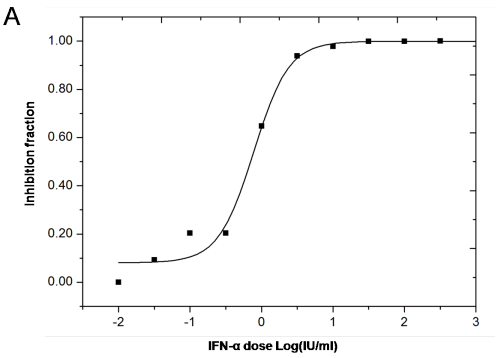
C

		Length	Lethal	Attenuated	Tolerated	Total insert	Lethal	Attenuated/Deleterious	Tolerated
1-340	5'UTR	340	287	9	0	296	96.96%	3.04%	0.00%
341-913	Core	573	373	107	17	497	75.05%	21.53%	3.42%
914-1489	E1	576	403	93	11	507	79.49%	18.34%	2.17%
1490-2590	E2	1101	807	92	20	919	87.81%	10.01%	2.18%
2591-2779	p7	189	107	51	16	174	61.49%	29.31%	9.20%
2780-3430	NS2	651	481	73	30	584	82.36%	12.50%	5.14%
3431-5323	NS3	1893	1621	35	1	1657	97.83%	2.11%	0.06%
5324-5485	NS4A	162	136	6	0	142	95.77%	4.23%	0.00%
5486-6268	NS4B	783	670	26	1	697	96.13%	3.73%	0.14%
6269-7666	NS5A	1398	672	70	462	1204	55.81%	5.81%	38.37%
7667-9442	NS5B	1776	1433	84	27	1544	92.81%	5.44%	1.75%
9443-9686	3'UTR	244	61	50	15	126	48.41%	39.68%	11.90%

SI Figure 2- 1. 15nt genomic mutagenesis screen. (A) Genomic mutagenesis screen libraries. To identify the domains critical for viral replication, we passaged the library with 10 million of ffu virus particles in Huh-7.5.1 cells for two rounds at MOI 0.2 and isolated the total RNA from the infected cells at the end as pool 2 (p2) to evaluate the fitness of each mutant. The *in vitro* transcribed RNA mutant library served as the input library pool 0 (p0). The total RNA of transfected cells was collected as pool 1 (p1) at 96 hours post transfection (96hpt). (B) Number of mutants detected in each library, total reads obtained from each library and the coverage for each mutant. The fitness of individual mutants was determined by analyzing the frequency of each mutant virus, indicated by different insertion sites, in various populations (p0, p1 and p2) with next-generation sequencing technology (Genome Analyzer IIx from Illumina). Our data showed that there were 8347 insertion sites corresponding to 8347 different mutant viruses in the input library, with an average coverage of 780 for each mutant, which was statistically significant. As expected, the insertion sites were uniformly distributed throughout the entire virus genome, except for the polyA region. In contrast, mutants isolated from transfected cells at 96hpt (p1) showed that several regions, especially the non-structural proteins, did not tolerate insertions, indicating their importance for viral genome replication, which is consistent with the known role in formation of replication complex. After two rounds of infection in the cells (p2), we observed the relative loss of mutants with insertions in structural proteins, suggesting that these domains are important for virus packaging and entry but not required for genome replication. (C) The distribution of mutations (the percentage and

number of insertions) in various regions of HCV genome and their phenotypes. To define a phenotype for each insertion mutant, the ratio of sequencing reads in p2 and p0 was calculated and presented in log scale (base 10). After the two rounds of selection, variants with frequencies maintained relative to input (positive log-scaled ratio) were considered as “tolerated”, while reduction by 0 to 3 logs in fitness as “attenuated”, and loss of variants (-3 logs or lower) as “lethal”.

A



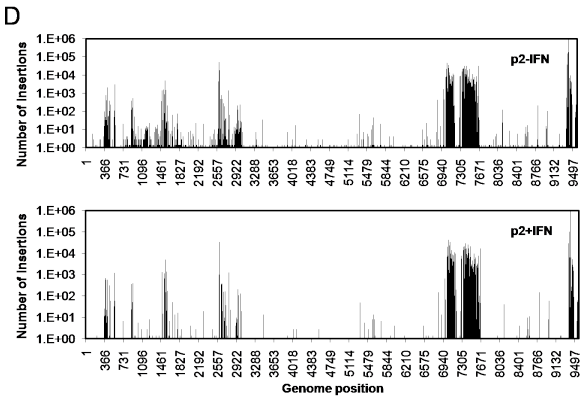
B

	p2 FN	P2+IFN
Number of mutant	1116	877
Total reads	4,537714	4,681573
Average coverage	4066	5338
Mean reads/mutant	134	386

C

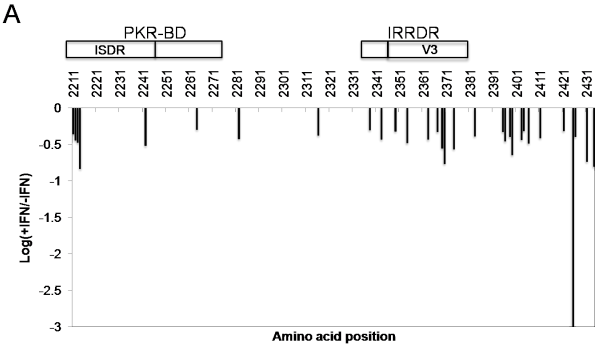
	Input (RNA)	96hpt	2nd round selection	
			w/o IFN	w/IFN
Lethal	/	4336	7231	7407
Tolerated/Attenuated	/	3880	1116	877
Total	8347	8347	8347	8347

D



SI Figure 2- 2. IFN selection of the 15nt mutant library. (A) IFN dose responsive curve of WT HCV virus. (B) Number of mutants detected in IFN selected and non-selected libraries, total reads obtained from each library and the coverage for each mutant. (C) The number of neutral, attenuated and lethal mutants detected in different libraries. (D) Frequency of each mutant indicated by different insertion sites, in the IFN selected and control libraries. X-axis represents the insertion position. Y-axis shows the number of each insertion detected by sequencing.

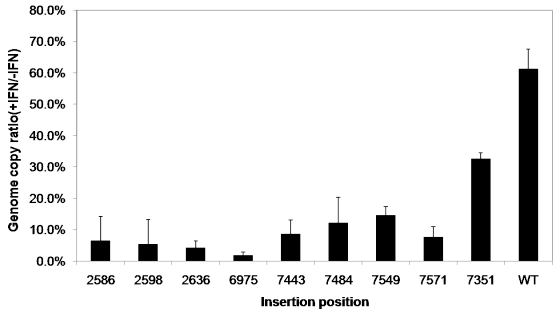
A



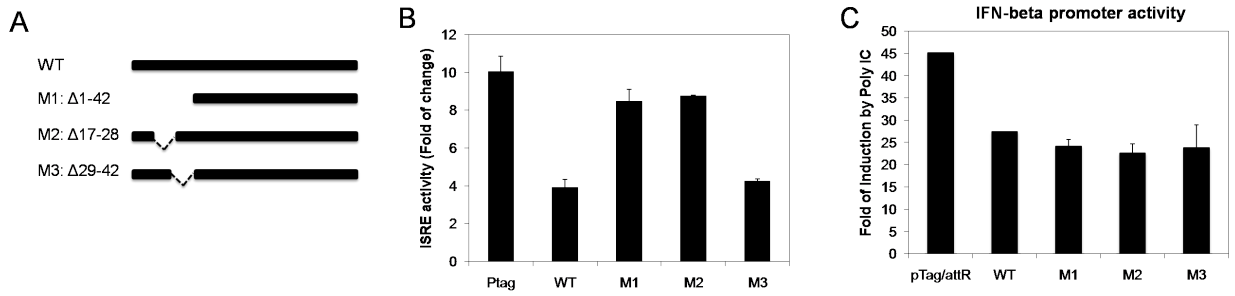
B

Mutant /Insert location	Inserted nucleotide sequence	Inserted amino acid sequence
2586	TGCGGCCGCA GCCGA	$_{748}A$- DAAAA -E_{749}
2598	TGCGGCCGCA CTAGA	$_{752}L$- DAAAL -E_{753}
2636	TGCGGCCGCA TAGCT	$_{765}S$- LRPHS -C_{766}
6975	TGCGGCCGCA TCAGC	$_{2211}S$- AAAAS -A_{2212}
7443	TGCGGCCGCA GCGGG	$_{2367}A$- GAAAA -G_{2368}
7459	TGCGGCCGCA CCGGC	$_{2373}G$- CGRTG -G_{2374}
7484	TGCGGCCGCA GCCGG	$_{2381}P$- VRPQP -A_{2382}
7549	TGCGGCCGCA CGGAC	$_{2403}D$- CGRTD -L_{2404}
7571	TGCGGCCGCA AGAGC	$_{2410}E$- LRPQE -L_{2411}
7351	TGCGGCCGCA GCGAG	$_{2337}E$- CGRSE -S_{2338}

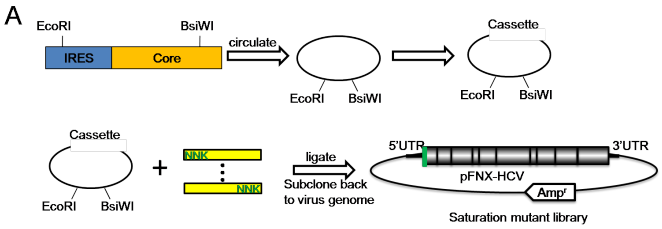
C



SI Figure 2- 3. Validation for the IFN screen with genomic mutagenesis library. (A) The IFN sensitive mutants are located in the regions, which are known as IFN sensitivity determine regions in clinical studies. (B) The sequence of individual mutant viruses constructed for validation. (C) The genome replication of individual mutants and WT with treatment of IFN. Data is presented as the ratio of genome copy number detected with IFN treatment to without IFN treatment.

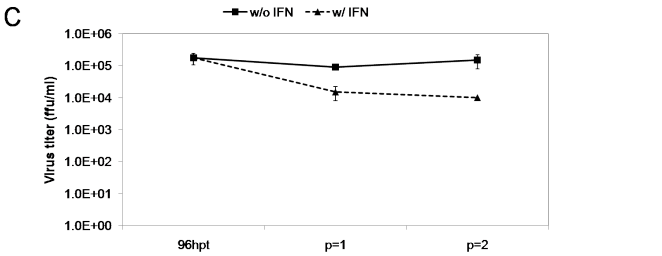


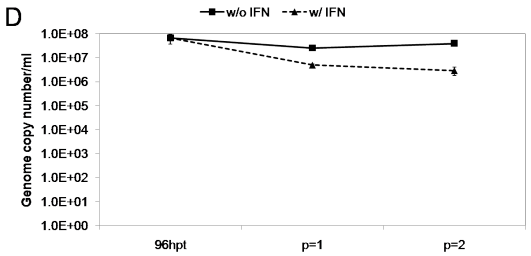
SI Figure 2- 4. Amino acid 17-28 is important for inhibiting IFN signaling pathway. To confirm that the N-terminal region of core protein counteracts the antiviral response (Fig. 5A), we cloned the WT and a series of N-terminally deleted HCV core constructs and tested whether overexpression of these proteins affects the IFN responsive signaling pathway by measuring the ISRE promoter activity. (A) Schematic picture showing the constructs of wt and mutants of core. (B) IRES report assay data showed that the full-length core protein inhibits the promoter activity by 60%, while deletion of amino acid 17-28 is sufficient to abrogate the inhibitory function of core on IRES promoter. (C) The IFN-beta promoter assay data showing that WT core and mutants are not differentially regulating the activity of the IFN-beta promoter.



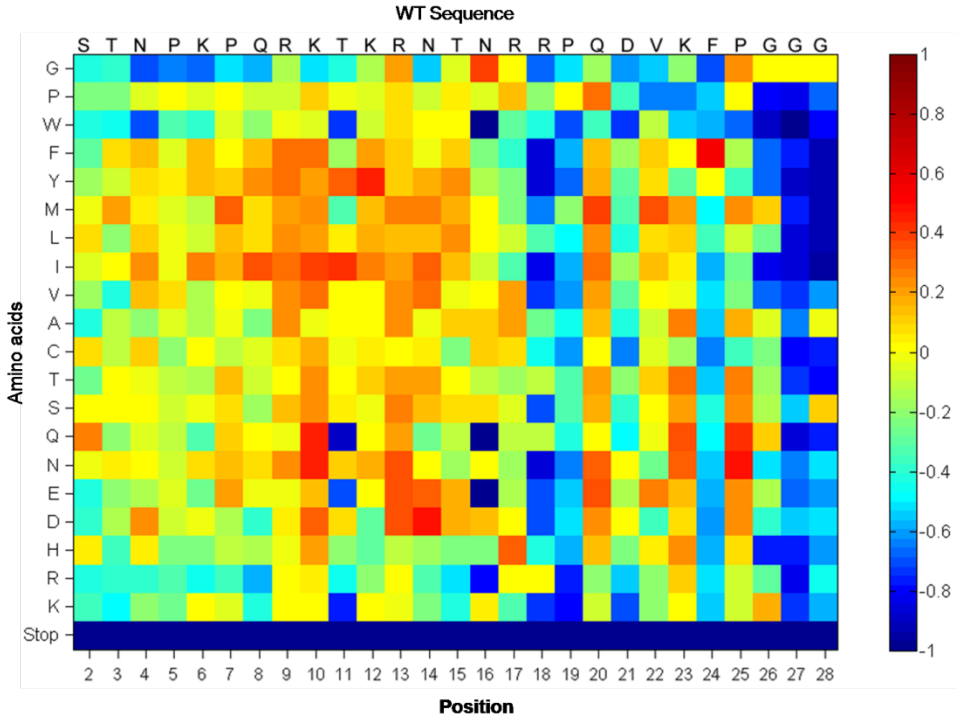
B

Library	Sequence Reads	Coverage
Plasmid	8,138,978	9420
RNA	4,574,166	5294
96hpt	8,523,693	9865
p2-IFN	4,243,896	4912
p2+IFN	6,067,784	7023

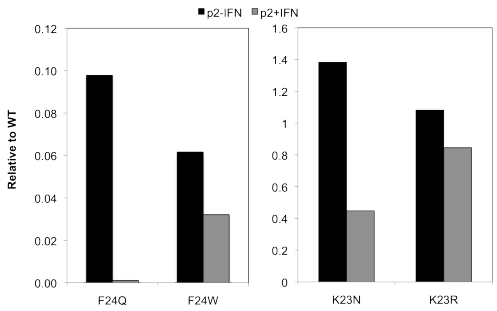




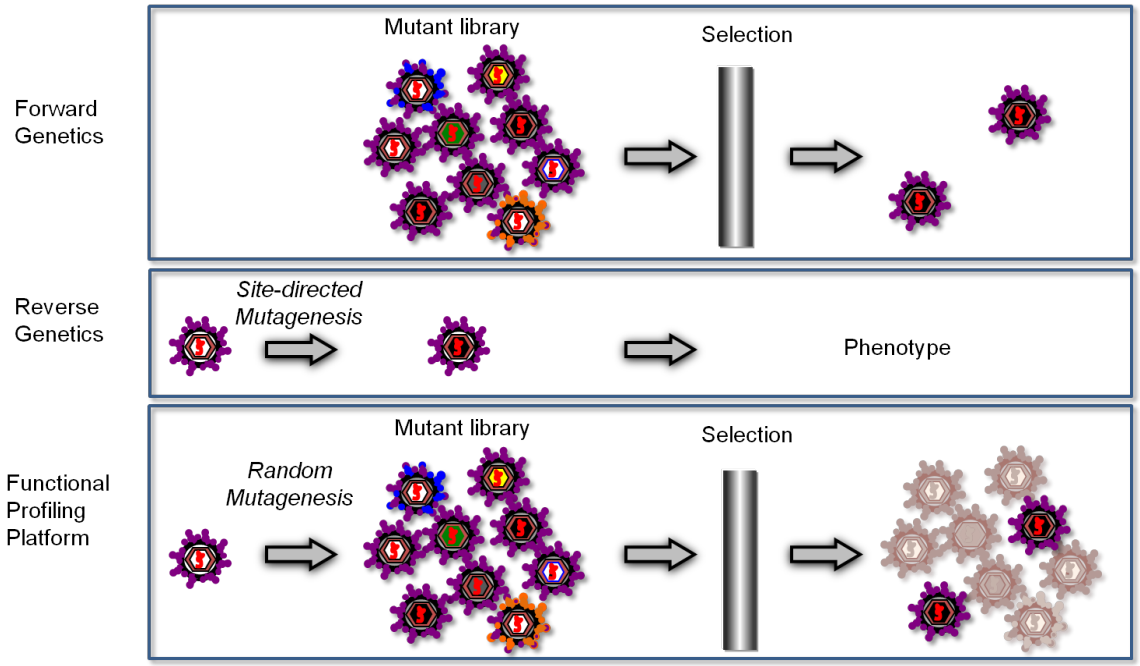
SI Figure 2- 5. Saturation mutagenesis of N-terminus of core. (A) Mutagenesis strategy of constructing saturated mutant library in fragment AA_2-AA_{28} of core protein. (B) The sequence reads obtained from each library and coverage for each mutant. Kinetics of mutant libraries selection in Huh-7.5.1 cells. The virus titer (C) and genome copy (D) number assayed for each library showing that the mutant library was completely covered.



SI Figure 2- 6. The fitness profile of mutant in the libraries. Heat map presenting the fitness of each mutant without IFN treatment, calculated by the selection coefficients.



SI Figure 2-7. The frequency of mutation F24Q, F24W, K23N and K23R in the IFN screen. Mutations F24Q and F24W confer the IFN sensitivity of the mutant viruses at different level. F24Q is more sensitive to IFN than F24W, despite the fact that F24W is less fit than F24Q, which suggests that the IFN sensitivity is not tracing the fitness of the virus. K23N is more sensitive to IFN than K23R, while K23N is also more fit than K23R, which shows that there is no strong correlation between the fitness of mutations and their IFN-sensitivity.



SI Figure 2- 7. Comparison of genetic profiling platform with traditional genetics strategies.

2.8 Bibliography

1. Lauer, G.M. and B.D. Walker, *Hepatitis C virus infection.* N Engl J Med, 2001. **345**(1): p. 41-52.

2. Yang, J.D. and L.R. Roberts, *Hepatocellular carcinoma: A global view.* Nat Rev Gastroenterol Hepatol, 2010. **7**(8): p. 448-58.

3. Feld, J.J. and J.H. Hoofnagle, *Mechanism of action of interferon and ribavirin in treatment of hepatitis C.* Nature, 2005. **436**(7053): p. 967-972.

4. Lindenbach, B.D. and C.M. Rice, *Unravelling hepatitis C virus replication from genome to function.* Nature, 2005. **436**(7053): p. 933-8.

5. Tellinghuisen, T.L., et al., *Studying Hepatitis C Virus: Making the Best of a Bad Virus.* Journal of Virology, 2007. **81**(17): p. 8853-8867.

6. Lemon, S.M., *Induction and Evasion of Innate Antiviral Responses by Hepatitis C Virus.* Journal of Biological Chemistry, 2010. **285**(30): p. 22741-22747.

7. Gale, M., Jr. and E.M. Foy, *Evasion of intracellular host defence by hepatitis C virus.* Nature, 2005. **436**(7053): p. 939-45.

8. Katze, M.G., Y. He, and M. Gale, Jr., *Viruses and interferon: a fight for supremacy.* Nat Rev Immunol, 2002. **2**(9): p. 675-87.

9. Meylan, E., et al., *Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus.* Nature, 2005. **437**(7062): p. 1167-72.

10. Loo, Y.-M., et al., *Viral and therapeutic control of IFN-beta promoter stimulator 1 during hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2006. **103**(15): p. 6001-6006.

11. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(8): p. 2992-2997.

12. Lin, W., et al., *Hepatitis C Virus Core Protein Blocks Interferon Signaling by Interaction with the STAT1 SH2 Domain.* Journal of Virology, 2006. **80**(18): p. 9226-9235.

13. Taylor, D.R., et al., *Inhibition of the Interferon- Inducible Protein Kinase PKR by HCV E2 Protein.* Science, 1999. **285**(5424): p. 107-110.

14. Gale Jr, M.J., et al., *Evidence That Hepatitis C Virus Resistance to Interferon Is Mediated through Repression of the PKR Protein Kinase by the Nonstructural 5A Protein.* Virology, 1997. **230**(2): p. 217-227.

15. Gale, M., et al., *Control of PKR Protein Kinase by Hepatitis C Virus Nonstructural 5A Protein: Molecular Mechanisms of Kinase Regulation.* Molecular and Cellular Biology, 1998. **18**(9): p. 5208-5218.

16. Garaigorta, U. and F.V. Chisari, *Hepatitis C Virus Blocks Interferon Effector Function by Inducing Protein Kinase R Phosphorylation.* Cell Host & Microbe, 2009. **6**(6): p. 513-522.

17. Arumugaswami, V., et al., *High-Resolution Functional Profiling of Hepatitis C Virus Genome.* PLoS Pathog, 2008. **4**(10): p. e1000182.

18. Beitzel, B.F., et al., *High-Resolution Functional Mapping of the Venezuelan Equine Encephalitis Virus Genome by Insertional Mutagenesis and Massively Parallel Sequencing.* PLoS Pathog, 2010. **6**(10): p. e1001146.

19. von Hahn, T. and C.M. Rice, *Hepatitis C virus entry.* J Biol Chem, 2008. **283**(7): p. 3689-93.

20. Hijikata, M., et al., *Hypervariable regions in the putative glycoprotein of hepatitis C virus.* Biochemical and Biophysical Research Communications, 1991. **175**(1): p. 220-228.

21. Weiner, A.J., et al., *Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins.* Virology, 1991. **180**(2): p. 842-848.

22. Kopp, M., et al., *Genetic Analysis of the Carboxy-Terminal Region of the Hepatitis C Virus Core Protein.* Journal of Virology, 2010. **84**(4): p. 1666-1673.

23. Hadlock, K.G., et al., *Human Monoclonal Antibodies That Inhibit Binding of Hepatitis C Virus E2 Protein to CD81 and Recognize Conserved Conformational Epitopes.* Journal of Virology, 2000. **74**(22): p. 10407-10416.

24. Enomoto, N., et al., *Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region.* The Journal of Clinical Investigation, 1995. **96**(1): p. 224-230.

25. Enomoto, N., et al., *Mutations in the Nonstructural Protein 5a Gene and Response to Interferon in Patients with Chronic Hepatitis C Virus 1b Infection.* New England Journal of Medicine, 1996. **334**(2): p. 77-82.

26. Duverlie, G., et al., *Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity.* Journal of General Virology, 1998. **79**(6): p. 1373-81.

27. Nousbaum, J.-B., et al., *Prospective Characterization of Full-Length Hepatitis C Virus NS5A Quasispecies during Induction and Combination Antiviral Therapy.* Journal of Virology, 2000. **74**(19): p. 9028-9038.

28. El-Shamy, A., et al., *Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy.* Hepatology, 2008. **48**(1): p. 38-47.

29. Ménez, R., et al., *Crystal Structure of a Hydrophobic Immunodominant Antigenic Site on Hepatitis C Virus Core Protein Complexed to Monoclonal Antibody 19D9D6.* The Journal of Immunology, 2003. **170**(4): p. 1917-1924.

30. Angus, A.G.N., et al., *Conserved Glycine 33 Residue in Flexible Domain I of Hepatitis C Virus Core Protein Is Critical for Virus Infectivity.* Journal of Virology, 2012. **86**(2): p. 679-690.

31. Kinde, I., et al., *Detection and quantification of rare mutations with massively parallel sequencing.* Proceedings of the National Academy of Sciences, 2011. **108**(23): p. 9530-9535.

32. Escobar-Gutierrez, A., et al., *Identification of Hepatitis C Virus Transmission Using a Next-Generation Sequencing Approach.* Journal of Clinical Microbiology, 2012. **50**(4): p. 1461-1463.

33. Davey, J.W., et al., *Genome-wide genetic marker discovery and genotyping using next-generation sequencing.* Nat Rev Genet, 2011. **12**(7): p. 499-510.

34. MacLean, D., J.D.G. Jones, and D.J. Studholme, *Application of 'next-generation' sequencing technologies to microbial genetics.* Nat Rev Micro, 2009. **7**(4): p. 287-296.

35. Takahashi, H., et al., *Biological properties of purified recombinant HCV particles with an epitope-tagged envelope.* Biochemical and Biophysical Research Communications, 2010. **395**(4): p. 565-571.

36. Merz, A., et al., *Biochemical and Morphological Properties of Hepatitis C Virus Particles and Determination of Their Lipidome.* Journal of Biological Chemistry, 2011. **286**(4): p. 3018-3032.

37. Prentoe, J. and J. Bukh, *Hepatitis C virus expressing flag-tagged envelope protein 2 has unaltered infectivity and density, is specifically neutralized by flag antibodies and can be purified by affinity chromatography.* Virology, 2011. **409**(2): p. 148-155.

38. Moradpour, D., et al., *Insertion of Green Fluorescent Protein into Nonstructural Protein 5A Allows Direct Visualization of Functional Hepatitis C Virus Replication Complexes.* Journal of Virology, 2004. **78**(14): p. 7400-7409.

39. Griffin, S.D.C., et al., *A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria.* Journal of General Virology, 2004. **85**(2): p. 451-461.

40. Griffin, S., et al., *Signal Peptide Cleavage and Internal Targeting Signals Direct the Hepatitis C Virus p7 Protein to Distinct Intracellular Membranes.* Journal of Virology, 2005. **79**(24): p. 15525-15536.

41. Carrere-KremerMontpellier-Pala, S., et al., *Subcellular Localization and Topology of the p7 Polypeptide of Hepatitis C Virus.* Journal of Virology, 2002. **76**(8): p. 3720-3730.

42. Haqshenas, G., et al., *Hepatitis C virus p7 protein is localized in the endoplasmic reticulum when it is encoded by a replication-competent genome.* Journal of General Virology, 2007. **88**(1): p. 134-142.

43. Munoz de Rueda, P., et al., *Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of Hepatitis C Virus Genotype 1 and Their Relationships to Pegylated Interferon-Ribavirin Treatment Responses.* Journal of Virology, 2008. **82**(13): p. 6644-6653.

44. Bouzgarrou, N., et al., *NS5AISDR-V3 region genetic variability of Tunisian HCV-1b strains: Correlation with the response to the combined interferon/Ribavirin therapy.* Journal of Medical Virology, 2009. **81**(12): p. 2021-2028.

45. Khaliq, S., S. Jahan, and S. Hassan, *Hepatitis C virus p7: molecular function and importance in hepatitis C virus life cycle and potential antiviral target.* Liver International, 2011. **31**(5): p. 606-617.

46. Su, A.I., et al., *Genomic analysis of the host response to hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2002. **99**(24): p. 15669-15674.

47. Falgueras, J., et al., *SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read.* BMC Bioinformatics, 2010. **11**(1): p. 38.

48. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-1760.

49. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

50. Barnett, D.W., et al., *BamTools: a C++ API and toolkit for analyzing and managing BAM files.* Bioinformatics, 2011. **27**(12): p. 1691-2.

CHAPTER 3

P7 IS A NOVEL IMMUNE EVASION PROTEIN OF HCV

3.1 Abstract

It has been well recognized that HCV encodes mechanisms to evade the multilayered antiviral actions of host defense system. Great progresses have been made in elucidating the strategies that HCV utilizes to down-regulate IFN production, impede IFN signaling transduction, and impair IFN stimulated gene expression. However, mechanisms governing viral proteins counteracting the antiviral functions of downstream IFN effectors were still limited, probably due to the lack of an experimental system to identify such interactions. With a high-resolution mutagenesis profiling approach, we found that mutations in p7 region conferred high levels of IFN-sensitivity, which can be alleviated by expression of wild type p7 protein. It suggests that p7 is a novel immune evasion protein of HCV. In an attempt to examine the mechanism of p7 in mediating immune response, we carried out a liver specific ISG library screen and identified 16 ISGs that significantly inhibits replication of p7 mutant viruses. Further mechanistic study revealed that p7 specifically interacts with IFI6-16, suggesting that p7 impedes the innate immunity through direct interaction with IFI6-16.

3.2 Introduction

Hepatitis C virus (HCV), with an estimated 170 million people persistently infected worldwide, has emerged as a major cause of human liver diseases, including chronic hepatitis, cirrhosis, and hepatocellular carcinoma (HCC) [1, 2]. However, neither prophylactic nor therapeutic vaccines are available for controlling HCV infection. IFN-α remains an important component of the standard treatment, but has significant side effects and offers only a limited response rate [3].

HCV is an enveloped positive-strand RNA virus encoding a polyprotein of around 3000 amino acids. The genome is comprised of two untranslated regions (5'UTR, 3'UTR), three structural proteins (core, E1, E2) and seven nonstructural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) [4]. Due to the limited genome space, many of the viral proteins have evolved multiple functions required for survival within the host. For example, in addition to the roles in viral replication [5], core, E2, NS3/4A and NS5A also encode immune evasion functions [6].

Virus-host interactions, such as the virus-IFN response, are very complex and likely involve multiple elements with various mechanisms [3, 7, 8]. The IFN response defends mammalian hosts from virus infection by controlling viral replication at multiple steps [8]. Sensing of viral infection triggers type I IFN expression, and secreted IFNs then bind to their receptors on the targeted cell surface. This results in activation of the JAK/STAT pathway, where STAT proteins are phosphorylated, followed by dimerization and association with IRF-9. The complex translocates to the nucleus and binds to IFN-stimulated response elements (ISRE) within the promoter region of IFN-stimulated genes (ISGs), inducing the expression of multiple antiviral effectors. However, viruses have evolved to circumvent, via different strategies, multiple layers of the IFN response, which likely dampens the antiviral efficacy of IFN-α therapy [7]. Previous studies uncovered several viral mechanisms to explain the lack of response of HCV infection to IFN-α therapy [6], through inhibiting the production of type I interferon and suppression of JAK/STAT signaling events to avoid the induction of an IFN mediated antiviral

state [9-16]. Great efforts were also devoted to interrogate the interactions of HCV and downstream IFN effectors and led to identification of several ISGs with inhibitory activity on HCV replication [6, 17-23]. In a recent study, a comprehensive library of human ISGs was cloned and over-expressed individually to test their ability in controlling the replication of several human viruses [24]. A subset of ISGs were found to inhibit HCV replication at different levels, and a larger group of ISGs were ineffective when over-expressed in viral infected cells due to unknown mechanisms.

In vivo study of experimentally infected chimpanzees has demonstrated that HCV infection strongly induces the expression of ISGs in the liver [25]. The ISG induction has also been observed in patients upon viral infection [26]. The fact that HCV persists in the liver despite the apparent induction of an antiviral state raises the possibility that the virus encodes mechanisms to counteract the anti-viral functions executed by ISGs. However, the current cDNA screens experimental setup is not optimal for identification of such interactions. To systematically explore the anti-IFN functional residues of HCV, we carried out a genomic mutagenesis screen of HCV in IFN-α selection, which resulted in identification of p7 as a novel immune evasion viral protein. A liver-specific ISG library [22, 23, 27, 28] screen on WT and p7 mutant virus replication revealed that IFI6-16 preferentially inhibits replication of p7 mutants but does not affect WT. Further examination on molecular interaction between the two proteins showed that p7 interacts and co-localizes with IFI6-16, suggesting that p7 antagonizes the antiviral responses of IFN through interacting with IFI6-16.

3.3 Results

3.3.1 Mutations in p7 confer hypersensitivity of the virus to IFN-α treatment.

To systematically explore the immune evasion functions encoded in HCV genome, we carried out an IFN-α screen with a comprehensive HCV mutant library with 15-nt randomly inserted in

the virus genome. Selection in 1U/ml of IFN-α treatment revealed that some mutations conferring hypersensitivity to IFN-α treatment clustered in amino-terminal (N-terminal) region of p7, presumably due to their loss of anti-IFN functions (Figure 3-1A). This screen result implies for the first time that p7 carries an immune evasion function, disruption of which causes significant inhibition by IFN treatment. This phenotype was validated with individually constructed p7 mutant viruses carrying 15-nt insertions at positions 2598 and 2636 (Figure 3-1B). An insertion at position 7351 in NS5A region was adopted as a negative control. To further examine the specificity of the IFN-α inhibition on p7 mutants, we tested whether overexpression of WT p7 protein can alleviate the inhibitory effect of IFN-α on p7 mutants. A Huh-7.5.1 cell line constitutively expressing p7 protein (Cp7) was established. A cell line harboring proteins core and E1 (CE1) served as a control (Figure 3-1C). We found that replication of p7 mutant viruses was inhibited by nearly 2 logs with IFN-α treatment (Figure 3-1D), while the defective replication phenotype was rescued to a level comparable to the control mutant (Figure 3-1E). The same rescuing effect was observed on the viral genome replication in Cp7 cells (Figure 3-1F,G), suggesting that p7 suppresses the antiviral effect activated by IFN-α treatment, and mutations in p7, as a consequence, lead to loss of the function and hypersensitivity to IFN-α.

3.3.2 Liver-specific ISG library screen identifies genes that specifically inhibit p7 mutant virus replication.

After confirming the regulatory function of p7 on the IFN antiviral effects, we then examined whether the IFN-sensitive response element (ISRE) promoter activity was affected upon p7 protein expression. HEK293T cells were transfected with plasmids carrying WT or mutant p7 along with a luciferase reporter under the control of ISRE, which is responsive to IFN-α induction. The transfected cells were subsequently stimulated with 50U/ml of IFN-α to induce the activation of ISRE promoter. The resultant luciferase activities were monitored after 20-hour treatment, and fold induction was calculated in comparison to untreated sample as the readout

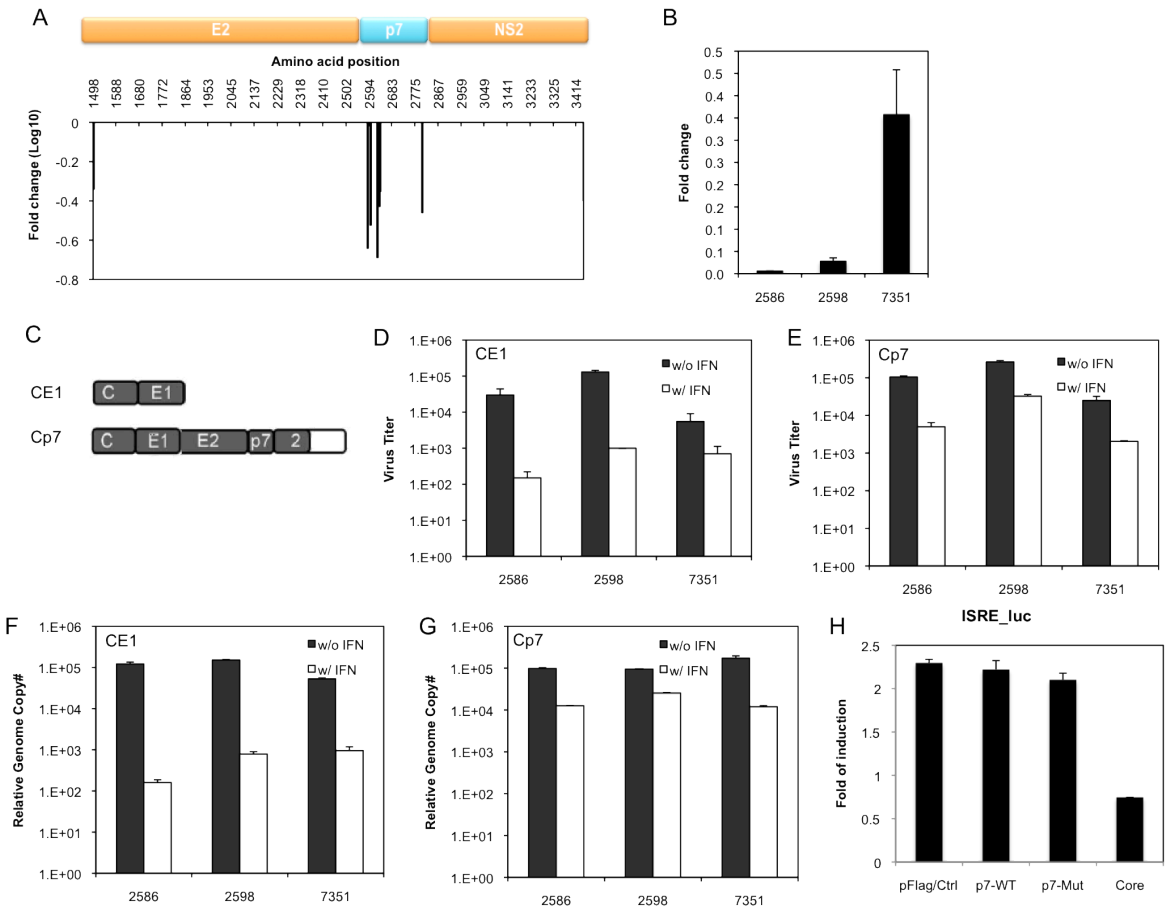


Figure 3- 1. Mutations in p7 confer hypersensitivity to IFN treatment. (A) 15-nt insertion mutagenesis profiling screen shows that mutations causing higher IFN-sensitivity are clustered in N-terminus of p7. (B) Individual mutant virus with 15-nt insertion in p7 confirmed the screen results. 2586, 2598 are the two p7 mutant viruses, and 7351 is a control mutant virus with 15-nt insertion in NS5A. The Huh-7.5.1 cells were pretreated with 1U/ml of IFN-α for 18 hours (hrs) before infection with the mutant viruses as indicated. At 72 hrs post infection, the virus production in the supernatant was measured and fold of changed caused by IFN treatment was calculated. (C) Two Huh-7.5.1 cell lines constitutively expressing CE1 or Cp7. (D) Infectious virus production in the supernatant of CE1 cell line at presence or absence of IFN-α treatment. (E) Infectious virus production in the supernatant of Cp7 cell line at presence or absence of IFN-α treatment. (F) Genome replication of mutant viruses in CE1 cell line measured by RT-q-PCR. (G) Genome replication of mutant viruses in Cp7 cell line measured by RT-q-PCT. (H) Effect of p7 overexpression on ISRE promoter activation upon IFN-α stimuli. HEK293T were transfected with ISRE-driving luciferase reporter plasmids along with the indicated plasmids respectively, and 24 hrs later, cells were left untreated or induced with 50U/ml of IFN-α for 20 hrs before luciferase assays were performed. Fold of activation was calculated in comparison to untreated control.

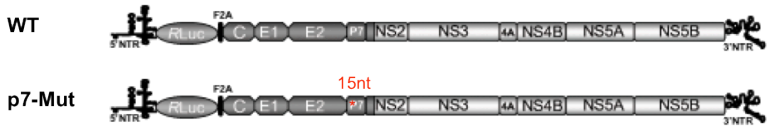
of ISRE activation. As shown in figure 3-1H, neither WT nor mutant p7 protein inhibits the activation of ISRE promoter induced by IFN-α, suggesting that p7 functions downstream of ISRE promoter.

To further interrogate the molecular basis of p7 counteracting innate immune response, we conducted a liver-specific ISG screen to search for ISGs that preferentially inhibit replication of p7 mutants. The assumption is that knockout of immune evasion function encoded in p7 restores the antiviral effect of the ISG, which was otherwise obstructed by WT p7 protein. We analyzed the published microarray data from IFN-treated liver hepatoma cells or fetal liver cells, and compiled a list of 107 ISGs that are expressed in liver cells upon IFN induction. We chose Huh-7.5.1 cell for the screen due to its defect in endogenous IFN production to give a clean background. Co-transfecting ISG constructs and a puromycin resistant vector allowed for selecting ISG-delivered Huh-7.5.1 cells, which were then challenged with WT or p7 mutant monocistronic Renilla luciferase HCV reporter viruses (Figure 3-2A). Viral replication in ISG-transfected cells was evaluated at 72 hrs post infection by measuring Renilla luciferase activity and normalized to that of RFP-transfected cells (Figure 3-2B). The antiviral effect of each ISG on p7 mutant and WT viral replication was compared and the ratio was taken. In line with other ISG screen done earlier in Huh-7.5 cells, a constrict range of ISGs presented antiviral activity, which is likely attributed to the ability of HCV to overcome their antiviral inhibition. As a result of high similarity in genome construction of p7 mutant and WT virus, we found that the ratio of their replication in most of the genes are close to 1 (Figure 3-2C, D). Interestingly, 16 genes were identified to inhibit the mutant virus to a greater extent than WT significantly ($p<0.05$). 12 out of 16 genes were further validated to consistently show stronger inhibition on the mutant virus replication (Figure 3-2E), suggesting a genetic interaction between p7 and the ISGs.

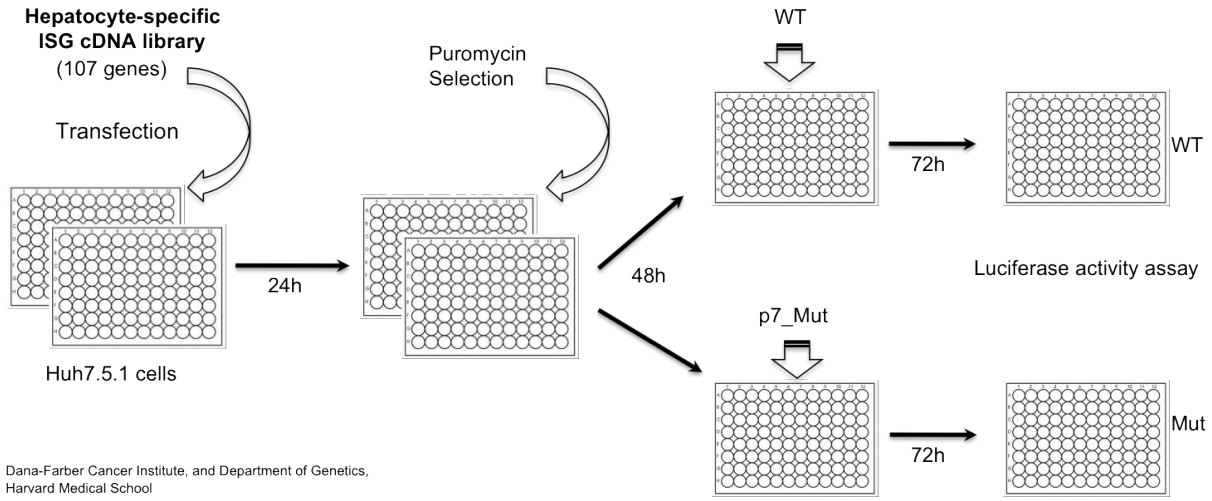
3.3.3 P7 specifically interacts with IFI6-16.

To further dissect the mechanism of genetic interactions between p7 and ISGs, we attempted to

A

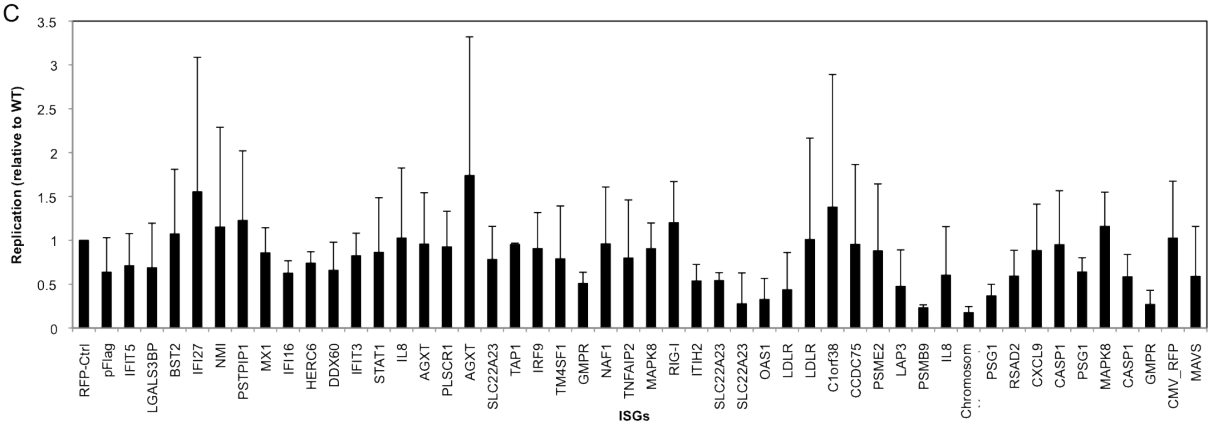


B

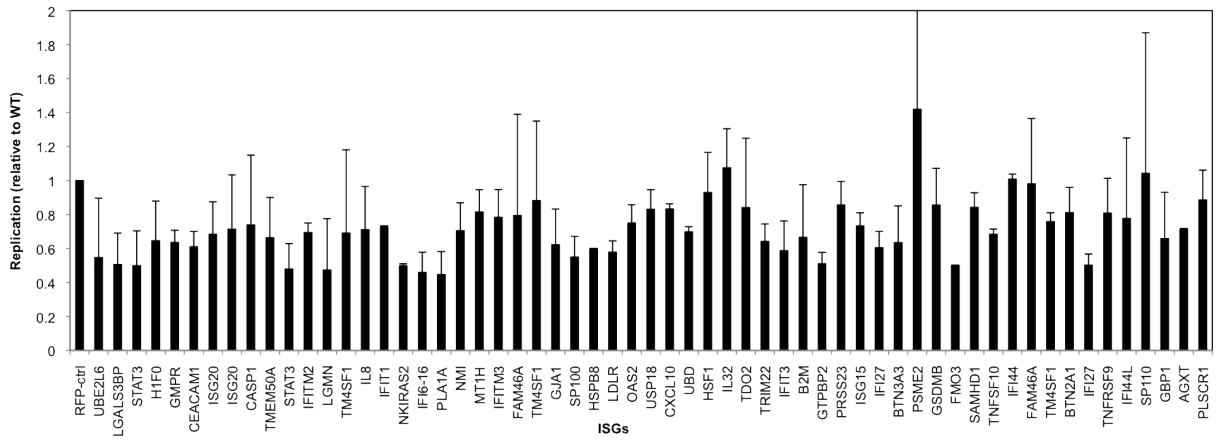


Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School

C



D



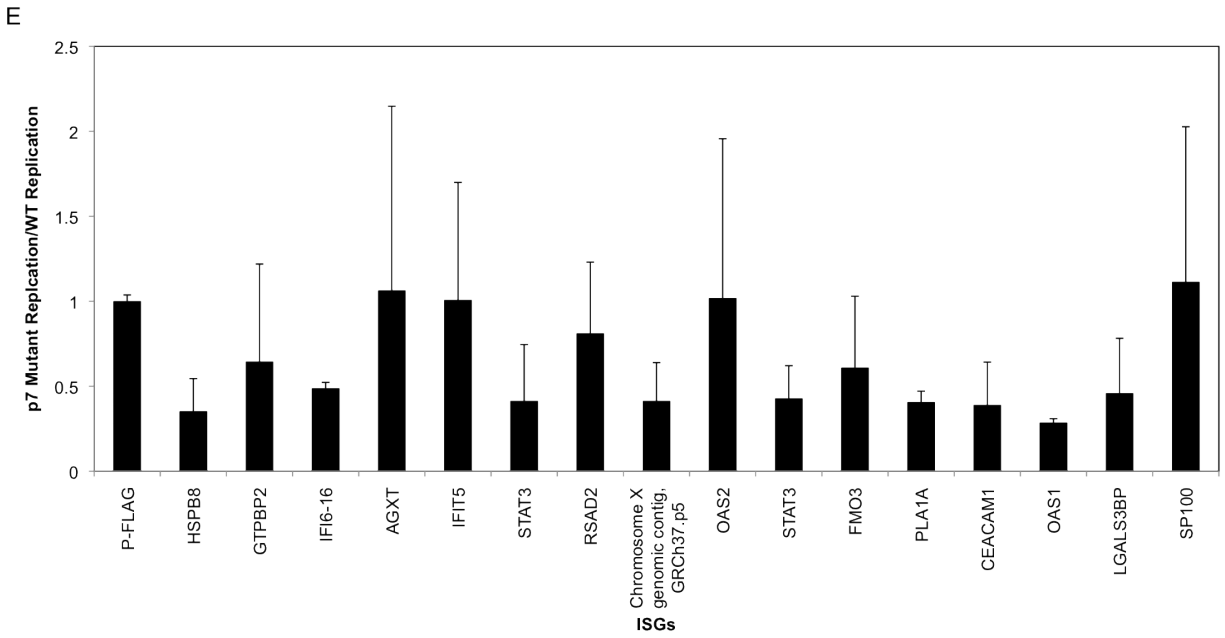


Figure 3- 2. Liver-specific ISG library screen for identifying ISGs that are regulated by p7. (A) Schematic of genome constructions of WT and p7 mutant (p7-Mut) Renilla reporter viruses. (B) Overall scheme of the liver-specific ISG library screen. ISG expression plasmid (107 total) was co-transfected with a puromycin-resistant plasmid into Huh-7.5.1 cells individually. Cells were selected with 2mg/ml of puromycin at 24 hrs post transfection. At 72 hrs post transfection (48 hrs post puromycin treatment), cells were changed with fresh growth medium and infected with WT or p7-Mut Renilla reporter viruses at an MOI of 0.1. At 72 hrs post infection, luciferase activities of infected Huh-7.5.1 cells were measured as the readout of viral replication. (C) and (D) Effects of ISGs, the control gene (RFP) and empty vector on p7-Mut virus replication in relative to that on WT control. The luciferase activity of p7-Mut in ISGs was normalized to that of in control gene (RFP) and compared with the normalized reading of WT. A relative activity equals to 1 suggests that the effect of ISG on p7-Mut equals that on WT virus. A bigger than 1 relative activity implies a weaker inhibition of ISG on p7-Mut, while a smaller than 1 ratio means a stronger inhibition of ISG on p7-Mut. (E) 16 top hits with p-value<0.05 were picked and validated in triplicates.

detect whether p7 and these ISGs are physically associated. Flag-tagged ISGs were individually co-transfected with HA-tagged p7 construct into Huh-7.5.1 cells. Co-immunoprecipitation results showed that IFI6-16 could interact strongly with p7. P7 and IFI6-16 were constructed in mammalian expression vectors as fusions to the epitope tags HA (pcDNA5) and Flag (pTAG). The interaction can be detected both in Flag-tagged IFI6-16 immunoprecipitated complex (Figure 3-3A) or HA-tagged p7 complex (Figure 3-3B).

The physical interaction between p7 and IFI6-16 was also confirmed by co-localization assay. Huh-7.5.1 cells were transfected with plasmids encoding p7 and IFI6-16 proteins, and subsequently fixed and probed with anti-Flag or anti-HA to detect the localization of IFI6-16 and p7. Confocal microscopy images show a perfect co-localization between these two proteins, verifying their physical interaction (Figure 3-3C).

3.4 Discussion

Host innate immune response has been described as the first defending line for eliminating invading viral pathogens [29]. However, since more than 50 years ago, when the antiviral effects of type I IFN were first described, it has also been recognized that viruses encode mechanisms to evade from the multiple layers of antiviral actions [7, 8, 30]. Great progresses have been made in elucidating the strategies that HCV utilizes to down-regulate IFN production, impede IFN signaling transduction, and impair IFN stimulated gene expression. However, details of mechanisms governing viral proteins counteracting the antiviral functions of downstream IFN effectors were still limited, probably due to the lack of an experimental system to identify such interactions. With a high-resolution mutagenesis profiling approach, we found that mutations in p7 region

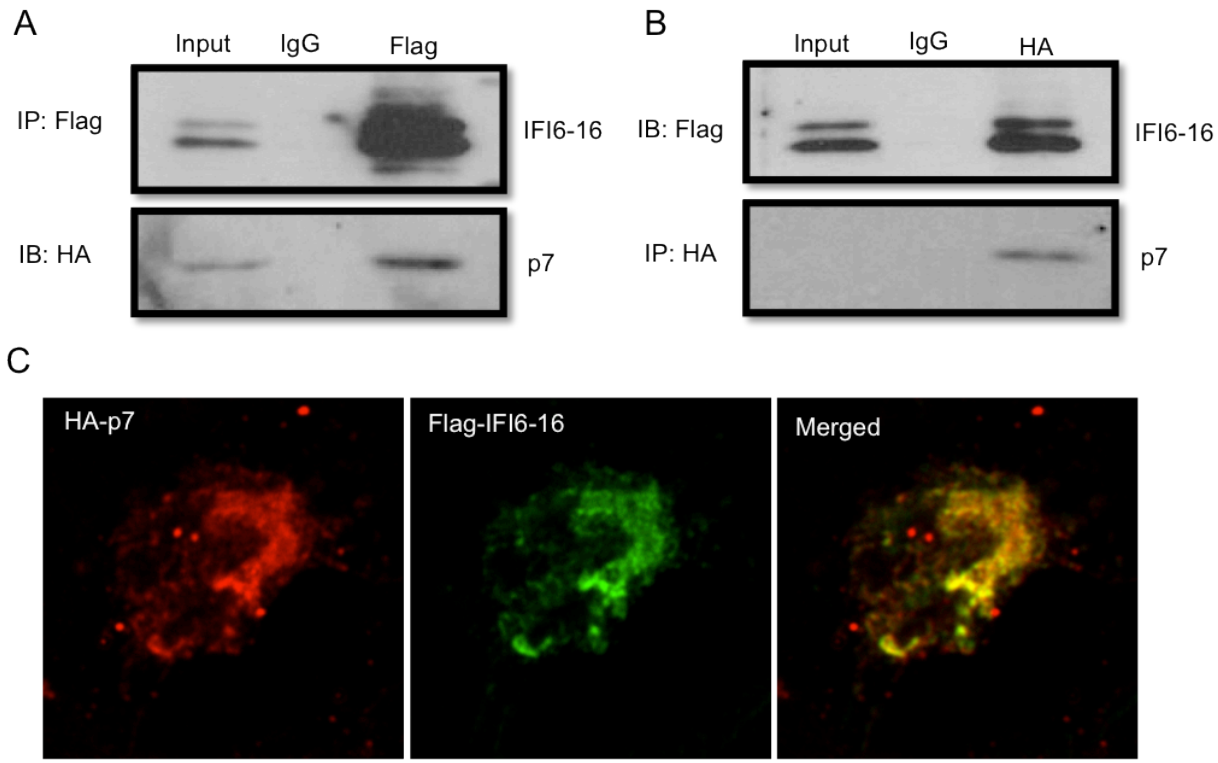


Figure 3- 3. P7 interacts with IFI6-16. (A) (B) Detecting the interaction of p7 and IFI6-16 by co-immunoprecipitation. HEK293T cells were transfected with FLAG-IFI6-16 and HA-p7. At 32 hrs post transfection, cell lysates were immunoprecipitated with anti-Flag, anti-HA, or non-specific anti-mouse IgG antibodies and subjected to western blotting with anti-Flag and anti-HA antibodies. (C) P7 and IFI6-16 are co-localized. Huh-7.5.1 cells were co-transfected with the indicated plasmids and immunofluorescence analysis was performed with the indicated antibodies.

conferred high levels of IFN-sensitivity, and this phenotype can be rescued by overexpression of wild type p7 protein in the infected cells, which implied p7 as a novel immune evasion protein of HCV. However, it is evident that p7 does not diminish the ISRE activity induced by IFN-α, which suggests that p7 might interfere with the antiviral functions of downstream of IFN effectors. A liver-specific ISG library screen identified specific genetic and molecular interaction between p7 and an ISG, IFI6-16.

P7 is known as an ion channel protein, comprised of two hydrophobic α-helices, and a basic loop located in the cytoplasm. Viral ion channel activity has been implicated in mediating virion entry, assembly, morphogenesis and secretion from host cells [31]. Here we found that mutations in p7, especially in the signaling peptide or N-terminus of p7 cause interferon sensitivity to the mutants without affecting the virus fitness.

Previous studies suggest that HCV proteins antagonize the innate immune response through inhibiting the production of type I interferon and suppression of JAK/STAT signaling events to avoid the induction of an IFN mediated antiviral state [32]. However, *in vivo* study of experimentally infected chimpanzees or human patient biopsies samples has demonstrated that HCV infection strongly induces the expression of ISGs in the liver [33, 34]. HCV persists in the liver despite the apparent induction of an antiviral state, raising the possibility that the virus encode mechanisms to counteract the anti-viral functions executed by ISGs. All these studies lead us to the hypothesis that p7 could be inhibiting the function of ISG(s) to facilitate robust viral replication despite the induction of anti-viral stage by IFN.

IFI6-16, also known as G1P3, was first identified as an interferon stimulated gene, whose mRNA was highly inducible in multiple cell lines upon type I interferon stimulation [35-38]. The expression of IFI6-16 is responsive to viral infections, including vesicular stomatitis virus (VSV), HCV, and cytomegalovirus [39]. It can also be induced by poly(I):poly(C) treatment and other immune regulators, namely lipopolysaccharide (LPS) and TNF-related apoptosis-induced

ligand (TRAIL) [39, 40]. Despite the early identification of IFI6-16 as an ISG and implication that it mediates innate immunity, the antiviral mechanism of the protein still remains obscure and inconclusive. Early study attempting to evaluate the antiviral function of IFI6-16 shows that introduction of IFI6-16 in a knockout cell line (HT1080_IFI6$^{-/-}$) does not affect the replication of encephalomyocarditis virus (EMCV), semliki forest virus or cocal virus, suggesting that IFI6-16 is not required to control these viral replication [41]. In contrast, IFI6-16 was identified as a negative regulator that markedly inhibited the replication of yellow fever virus [24]. The expression of the gene was also found to suppress respiratory syncytial virus replication and was down-regulated by the virus [42]. The antiapoptotic function of IFI6-16 was observed in immune cells to delay the apoptotic activity and maintain the immune response [43]. The effect of IFI6-16 on HCV replication, however, seems a little bit “controversial”. In the replicon cells harboring HCV subgenomic RNA, overexpression of IFI6-16 inhibited HCV replication and expression of the virus proteins, and knockdown increased the level of viral RNA replication. Moreover, IFI6-16 did not activate the IFN activation pathway, suggesting that it functions directly against viral replication without going through the IFN activity, which may induce and amplify antiviral actions [17, 18]. However, a comprehensive ISG cDNA screen using infection system suggests that IFI6-16 shows moderate or no significant suppression on HCV replication in either Huh-7 or Huh-7.5 cell lines [24]. In agreement with the screen results, HCV persists in the chimpanzee livers regardless of the up-regulation of IFI6-16, suggesting that the virus overcomes the antiviral functions of IFI6-16 [44]. Besides, comparison of ISG expression patterns among the IFN treatment responsive and nonresponsive patients concluded that the high level of endogenous IFI6-16 expression may hinder the response to therapy and cause ineffectiveness in clearing the infection [45]. Further studies also show that the high basal level of IFI6-16 in pretreated patients is an indicative marker of non-responsiveness in chronic hepatitis C patients [46].

The study we present here suggests that p7 functions as an immune evasion protein, most likely by counteracting the antiviral function of IFI6-16. This observation also explains the “controversial” observation that expression of IFI6-16 protein presents a substantial greater level of antiviral effect in HCV replicon system than that in infectious system, and provides an insight into mechanistic prediction on IFN nonresponsive outcome for some patients. The interaction could not have been identified without the genomic scale mutagenesis study. Collectively, these multilayered systematic approaches offer a comprehensive insight into HCV and host interactions, which will provide basis for understanding mechanisms of IFN resistance and developing new anti-viral therapies.

3.5 Materials and Methods

Cell culture, viruses and plasmids

The Huh-7.5.1 cell line was kindly provided by Dr. Francis Chisari from the Scripps Research Institute, La Jolla. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% of fetal bovine serum (FBS), 10mM non-essential amino acids (Invitrogen, Carlsbad, USA), 10mM HEPES, penicillin (100 units/ml), streptomycin (100 mg/ml), and 2mM L-glutamine at 37°C with 5% CO2. 293T cells were grown in DMEM with 10% FBS, 100 units/ml of penicillin, and 100mg/ml of streptomycin.

The pENTR-ISG library was a gift kindly provided by Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School. The ISGs were originally in the pENTR vector (Invitrogen), and recombined by an LR reaction into a modified destination vector resulting in a 3xFLAG epitope tag at the N-terminus of the protein. This process inserted 6 additional amino acids to the N-terminus of the protein (from the multiple cloning site) after the 3×FLAG tag and before the core ATG start site.

pFNX-HCV is the plasmid we synthesized based on the chimeric sequence of J6/JFH1 virus. We introduced 7 nucleotide substitutions, resulting in synonymous mutations to the

genome. The construct and sequence are available upon request. The monocistronic reporter virus was constructed as described before [47]. The p7 mutant viruses were cloned by introducing 15-nt insertion into desired positions accordingly. The primer sequences for mutant virus construction are listed below:

15nt_2586_F: GCCGATGCGGCCGCAGCCGAAGCAGCACTAGAG;

15nt_2586_R: TCGGCTGCGGCCGCATCGGCCTGGCCCAACAAG;

15nt_2598_F: CTAGATGCGGCCGCACTAGAGAAGCTGGTCATCTTGC;

15nt_2598_R: TCTAGTGCGGCCGCATCTAGTGCTGCTTCGGCCT;

15nt_2636_F: TAGCTTGCGGCCGCATAGCTGCAATGGCTTCCT;

15nt_2636_R: AGCTATGCGGCCGCAAGCTAGCTGCGCTCGCAG;

15nt_7351_F: GGGTCTGAGCGAGTGCGGCCGCAGCGAGAGCACCAT;

15nt_7351_R: ATGGTGCTCTCGCTGCGGCCGCACTCGCTCAGACCC;

The firefly luciferase reporter driven by the interferon-stimulated response element (ISRE_firefly-luciferase) and the renilla luciferase reporter driven by the promoter of housekeeping gene phosphoglycerate kinase (PGK_renilla-luciferase) were kind gifts from Dr. Genhong Cheng and Dr. Lily Wu, respectively (UCLA). Wild type and mutant p7 was cloned into the pcDNA5 vector (Invitrogen) by PCR amplification from pFNX24 with HA coding sequences, resulting in an HA epitope tag at the C-terminus of the protein. Primers are available upon request.

In vitro transcription

The virus genome containing plasmids were linearized with XbaI(NEB) and treated with Mungbean nuclease (NEB) for 30min at 30°C. The linearized DNA was then used as template for in vitro transcription to synthesize RNA virus genome with T7 polymerase following the protocol of the manufacture (Promega).

Reporter analysis

$5x10^4$ of HEK293T cells were seeded at 48-well plate prior to transfection. The cells were co-transfected with 20ng of ISRE_firefly-luciferase, 5 ng of PGK_renilla-luciferase, and 350 ng of viral protein expression plasmids or vector control using lipofectamine 2000 (Invitrogen). At 24 hours post transfection (hpt), the cells were treated with 50U/ml of IFN-α. At 20 hours post stimulation, firefly and renilla luciferase activity were measured by dual-luciferase reporter assay system (Promega Dual Luciferase Assay System). The same set of transfected cells was left untreated as control. Fold of activation was calculated by comparing the normalized firefly luciferase activity in the IFN-α treated sample with non-treated sample. To evaluate the replication efficiency of reporter viruses, the infected cells were lysed and the renilla luciferase activity was measured by renilla luciferase reporter assay kit (Promega Renilla Luciferase Assay System).

Immunoprecipitation and immunoblotting.

For co-immunoprecipitation assay to detect the interaction of IFI6-16 and p7, HEK293T cells were co-transfected with both constructs, washed once with PBS and then lysed on ice with lysis buffer (50 mM Tris pH 7.5, 1% NP-40, 0.25% sodium deoxycholate, 1% Tritonx-100, 150 mM NaCl, 5mM EDTA) supplemented with 1 mM PMSF, 1 mM Na_3VO_4, and 1 mM NaF. The cell lysates were then incubated with a specific antibody (2 ug antibody per sample) or control IgG conjugated Sepharose protein G beads (GE) overnight at 4°C. The immunoprecipitates were washed three times with lysis buffer then boiled and analyzed by western blot. The antibodies used in this study were as follows: mouse anti-FLAG (Sigma, F3165), mouse anti-HA (Sigma, H3663).

Immunofluorescent assay

At 48h post transfection, Huh-7.5.1 cells cultured on coverslips in 24-well plates were fixed with 4% paraformaldehyde for overnight. After 3 times wash with PBS, the cells were permeablized 0.05% Triton-x 100 for 10min and with blocked with 10% FBS, 3% BSA. The cells were incubated with human anti-HA antibody (Sigma, H3663) and rabbit anti-FLAG antibody at 4°C for overnight. Secondary antibodies, goat anti-rabbit IgG conjugated to Alex 488 (Invitrogen) and goat anti-mouse IgG conjugated to Alex 593 (Invitrogen) were added and incubated for 1 hr at room temperature. Cells were washed for 3 times with PBS and nuclei were stained with Hoechst 33342 (Invitrogen). The coverslips with cells were then removed from the plate and mounted to glasses with anti-fade reagent before imaging with confocal microscopy.

Measuring virus titer

The virus titer was assayed by measuring the focus-forming unit of infectious virus particles per ml of supernatant. Culture supernatant taken from different time points, e.g. 48hpt, 72hpt and 96hpt were diluted in a 10-fold serial manner before applying onto the Huh-7.5.1 cells seeded in 96-well plates. 72hpt, the cells were fixed with 100% methanol and immunostained for core antigen with mouse monoclonal anti-core primary antibody C7-50 (Abcam, Cambridge, USA), followed by goat anti-mouse-Alex 488 secondary antibody. Virus titer was determined by counting the number of core antigen-positive foci at highest dilutions.

Quantitatively detection of viral genome with q-PCR

Total RNA was isolated from transfected or infected cells by Trizol (Invitrogen) following the instruction of the manufacture. 500ng of RNA was used for reverse transcription using q-Script cDNA Synthesis Kit (Quantas). HCV-specific qPCR was conducted by using primer pair: 5'-AGAGCCATAGTGGTCTGCG-3' and 5'-CTTTCGCAACCCAACGCTAC-3'; Actin: 5'-ACCTTCTACAATGAGCTGCG-3' and 5'-CCTGGATAGCAACGTACATGG.

3.6 Bibliography

1. Lauer, G.M. and B.D. Walker, *Hepatitis C virus infection.* N Engl J Med, 2001. **345**(1): p. 41-52.

2. Yang, J.D. and L.R. Roberts, *Hepatocellular carcinoma: A global view.* Nat Rev Gastroenterol Hepatol, 2010. **7**(8): p. 448-58.

3. Feld, J.J. and J.H. Hoofnagle, *Mechanism of action of interferon and ribavirin in treatment of hepatitis C.* Nature, 2005. **436**(7053): p. 967-972.

4. Lindenbach, B.D. and C.M. Rice, *Unravelling hepatitis C virus replication from genome to function.* Nature, 2005. **436**(7053): p. 933-8.

5. Tellinghuisen, T.L., et al., *Studying Hepatitis C Virus: Making the Best of a Bad Virus.* Journal of Virology, 2007. **81**(17): p. 8853-8867.

6. Lemon, S.M., *Induction and Evasion of Innate Antiviral Responses by Hepatitis C Virus.* Journal of Biological Chemistry, 2010. **285**(30): p. 22741-22747.

7. Gale, M., Jr. and E.M. Foy, *Evasion of intracellular host defence by hepatitis C virus.* Nature, 2005. **436**(7053): p. 939-45.

8. Katze, M.G., Y. He, and M. Gale, Jr., *Viruses and interferon: a fight for supremacy.* Nat Rev Immunol, 2002. **2**(9): p. 675-87.

9. Meylan, E., et al., *Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus.* Nature, 2005. **437**(7062): p. 1167-72.

10. Loo, Y.-M., et al., *Viral and therapeutic control of IFN-beta promoter stimulator 1 during hepatitis C virus infection.* Proceedings of the National Academy of Sciences, 2006. **103**(15): p. 6001-6006.

11. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(8): p. 2992-2997.

12. Lin, W., et al., *Hepatitis C Virus Core Protein Blocks Interferon Signaling by Interaction with the STAT1 SH2 Domain.* Journal of Virology, 2006. **80**(18): p. 9226-9235.

13. Taylor, D.R., et al., *Inhibition of the Interferon- Inducible Protein Kinase PKR by HCV E2 Protein.* Science, 1999. **285**(5424): p. 107-110.

14. Gale Jr, M.J., et al., *Evidence That Hepatitis C Virus Resistance to Interferon Is Mediated through Repression of the PKR Protein Kinase by the Nonstructural 5A Protein.* Virology, 1997. **230**(2): p. 217-227.

15. Gale, M., et al., *Control of PKR Protein Kinase by Hepatitis C Virus Nonstructural 5A Protein: Molecular Mechanisms of Kinase Regulation.* Molecular and Cellular Biology, 1998. **18**(9): p. 5208-5218.

16. Garaigorta, U. and F.V. Chisari, *Hepatitis C Virus Blocks Interferon Effector Function by Inducing Protein Kinase R Phosphorylation.* Cell Host & Microbe, 2009. **6**(6): p. 513-522.

17. Itsui, Y., et al., *Expressional screening of interferon-stimulated genes for antiviral activity against hepatitis C virus replication.* Journal of Viral Hepatitis, 2006. **13**(10): p. 690-700.

18. Itsui, Y., et al., *Antiviral effects of the interferon-induced protein guanylate binding protein 1 and its interaction with the hepatitis C virus NS5B protein.* Hepatology, 2009. **50**(6): p. 1727-1737.

19. Zhu, H., et al., *Gene expression associated with interferon alfa antiviral activity in an HCV replicon cell line.* Hepatology, 2003. **37**(5): p. 1180-1188.

20. Wilkins, C., et al., *IFITM1 is a tight junction protein that inhibits hepatitis C virus entry.* Hepatology, 2012: p. n/a-n/a.

21. Randall, G., et al., *Silencing of USP18 Potentiates the Antiviral Activity of Interferon Against Hepatitis C Virus Infection.* Gastroenterology, 2006. **131**(5): p. 1584-1591.

22. Helbig, K.J., et al., *Analysis of ISG expression in chronic hepatitis C identifies viperin as a potential antiviral effector.* Hepatology, 2005. **42**(3): p. 702-710.

23. Jiang, D., et al., *Identification of Three Interferon-Inducible Cellular Enzymes That Inhibit the Replication of Hepatitis C Virus.* Journal of Virology, 2008. **82**(4): p. 1665-1678.

24. Schoggins, J.W., et al., *A diverse range of gene products are effectors of the type I interferon antiviral response.* Nature, 2011. **472**(7344): p. 481-485.

25. Su, A.I., et al., *Genomic analysis of the host response to hepatitis C virus infection.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15669-15674.

26. Biieche, I., et al., *Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection.* Virology, 2005. **332**(1): p. 130-144.

27. Bieche, I., et al., *Molecular profiling of early stage liver fibrosis in patients with chronic hepatitis C virus infection.* Virology, 2005. **332**(1): p. 130-144.

28. Geiss, G.K., et al., *Gene Expression Profiling of the Cellular Transcriptional Network Regulated by Alpha/Beta Interferon and Its Partial Attenuation by the Hepatitis C Virus Nonstructural 5A Protein.* Journal of Virology, 2003. **77**(11): p. 6367-6375.

29. Sadler, A.J. and B.R.G. Williams, *Interferon-inducible antiviral effectors.* Nat Rev Immunol, 2008. **8**(7): p. 559-568.

30. Katze, M.G., et al., *Innate immune modulation by RNA viruses: emerging insights from functional genomics.* Nat Rev Immunol, 2008. **8**(8): p. 644-654.

31. Khaliq, S., S. Jahan, and S. Hassan, *Hepatitis C virus p7: molecular function and importance in hepatitis C virus life cycle and potential antiviral target.* Liver International, 2011. **31**(5): p. 606-617.

32. Bowie, A.G. and L. Unterholzner, *Viral evasion and subversion of pattern-recognition receptor signalling.* Nat Rev Immunol, 2008. **8**(12): p. 911-922.

33. Bigger, C.B., et al., *Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees.* J Virol, 2004. **78**(24): p. 13779-92.

34. Lau, D.T.Y., et al., *Interferon regulatory factor-3 activation, hepatic interferon-stimulated gene expression, and immune cell infiltration in hepatitis C virus patients.* Hepatology, 2008. **47**(3): p. 799-809.

35. Friedman, R.L., et al., *Transcriptional and posttranscriptional regulation of interferon-induced gene expression in human cells.* Cell, 1984. **38**(3): p. 745-755.

36. Kelly, J.M., et al., *Differential regulation of interferon-induced mRNAs and c-myc mRNA by α- and γ-interferons.* European Journal of Biochemistry, 1985. **153**(2): p. 367-371.

37. Revel, M. and J. Chebath, *Interferon-activated genes.* Trends in Biochemical Sciences, 1986. **11**(4): p. 166-170.

38. Kusari, J., et al., *Expression of interferon-inducible genes in RD-114 cells.* Journal of Virology, 1987. **61**(5): p. 1524-1531.

39. Martensen, P.M. and J. Justesen, *Small ISGs coming forward.* Journal of Interferon & Cytokine Research, 2004. **24**(1): p. 19.

40. Cheriyath, V., D.W. Leaman, and E.C. Borden, *Emerging roles of FAM14 family members (G1P3/ISG 6–16 and ISG12/IFI27) in innate immunity and cancer.* Journal of Interferon & Cytokine Research, 2011. **31**(1): p. 9.

41. Porter, A.C.G. and J.E. Itzhaki, *Gene targeting in human somatic cells.* European Journal of Biochemistry, 1993. **218**(2): p. 273-281.

42. Zhao, D., et al., *Inhibition of G1P3 expression found in the differential display study on respiratory syncytial virus infection.* Virology Journal, 2008. **5**(1): p. 114.

43. Zhang, X., et al., *Reciprocal expression of TRAIL and CD95L in Th1 and Th2 cells: role of apoptosis in T helper subset differentiation.* Cell Death Differ, 2003. **10**(2): p. 203-210.

44. Kato, T., et al., *Interferon-inducible gene expression in chimpanzee liver infected with hepatitis c virus.* Virology, 1992. **190**(2): p. 856-860.

45. Sarasin-Filipowicz, M., et al., *Interferon signaling and treatment outcome in chronic hepatitis C.* Proceedings of the National Academy of Sciences, 2008. **105**(19): p. 7034-7039.

46. Sixtos-Alonso, M.S., et al., *IFN-stimulated Gene Expression Is a Useful Potential Molecular Marker of Response to Antiviral Treatment with Peg-IFNŒ± 2b and Ribavirin*

in Patients with Hepatitis C Virus Genotype 1. Archives of medical research, 2011. **42**(1): p. 28-33.

47. Arumugaswami, V., et al., *High-Resolution Functional Profiling of Hepatitis C Virus Genome.* PLoS Pathog, 2008. **4**(10): p. e1000182.

CHAPTER 4

SYSTEMATICALLY EXPLORING DISTINCT FITNESS LANDSCAPES OF AN HCV PROTEIN UNDER DRUG TREATMENT

4.1 Abstract

Hepatitis C virus (HCV) infection continues to cause a significant global disease burden with an estimated number of 170 million people persistently infected. Contentious efforts have been put into development of effective IFN-free drugs against viral infection ever since the establishment of HCV experimental systems. Daclatasvir (BMS-790052) has been identified as an extremely effective and potent HCV inhibitor, which is implied to interact with domain 1A of NS5A protein. Despite identification of resistant mutations during natural adaptive selection under drug treatment, the detailed working mechanism and full spectrum of potential escape mutations remain elusive. Here we utilized next-generation sequencing to quantitatively examine the replication fitness of each mutant virus in a saturation mutagenesis library within domain IA of NS5A, which yielded a fitness landscape of the region at single amino acid resolution. This led to identification of functional residues such as mediating RNA binding, involving protein-protein interactions, and regulating structural stability. Furthermore, selection under drug treatment revealed a panel of residues that influence the drug-sensitivity of the virus. Epistatic interactions were observed among these residues, and this epistasis explains the HCV genotype-specific difference in drug-sensitivity. Mathematical simulation predicted the potential emergence of resistant variants at different levels of medication adherences and postulated Y93W as a two-nucleotide-change mutation conferring strongest resistance. The comprehensive maps of fitness and drug-sensitivity for all possible single amino acid substitutions allowed for exploring the possible evolutionary paths of Y93W, suggesting the emergence of resistance is likely through Y93C. We suggest that the platform described here should generally be applicable to studying any drug-protein interactions to elucidate the working mechanisms of newly developed drugs.

4.2 Introduction

Hepatitis C virus is a positive single stranded RNA virus. The 9.6kb genome contains 2 untranslated regions (5'UTR, 3'UTR), and encodes 3 structural and 7 nonstructural proteins [1]. Persistent infection of HCV has been estimated in about 170 million people worldwide, and puts the patients at huge risks of developing severe liver diseases, including chronic hepatitis, cirrhosis, and hepatocellular carcinoma [2]. Traditional therapeutic treatment, the combination of interferon-α and ribavirin, only achieves limited viral response, yet causes severe side effects [3]. Tremendous efforts made to seek drugs with more effectiveness and less unintended effects lead to the approval of two protease inhibitors [4, 5]. However, the efficacy is highly dependent on the genotype and hindered by the arising resistant mutations [6, 7]. There are still unmet demands for development of novel drugs for effective treatment of HCV infection. The establishment of HCV infectious [8-10] and replicon cell system [11, 12] has facilitated the high-throughput screening of small-molecule inhibitors, and aided identification of many antiviral compounds [13].

Daclatasvir (BMS-790052) is identified as a potent antiviral agent, which blocks viral RNA replication by interfering with the function of nonstructural protein 5A (NS5A) [14, 15]. NS5A is known as a multifunctional viral protein, indispensable for viral genome replication, packaging, and innate immune evasion [16]. Its essential role during viral replication and lack of human homologous protein implies it to be a promising therapeutic target. Studies discovered that Daclatasvir possess a broad antiviral activity, with a medium effective concentration (EC_{50}) and cytotoxic concentration at a pico molar (pM) and micro molar (uM) range respectively, yielding a big potential therapeutic window [15]. However, the working mechanism and the complete resistant profile of the compound still need to be clarified. *In vitro* adaptation experiments, done by passing wild type (WT) HCV clones in the presence of the drug, uncovered the emergence of resistant mutations within domain IA of NS5A, suggesting a direct interaction between the drug and this domain [14, 15, 17, 18]. Clinical trial studies confirmed

correlation of resistant variants *in vivo* with those observed *in vitro* [15, 19]. Interestingly, viral breakthrough was detected in patients as early as 8 hours (hrs) post treatment, suggesting the preexistence of these minor quasispecies in the untreated population [6, 19]. Unfortunately, the study only monitored the major resistant strains after a short period of treatment (14 days), which may not reflect the entire resistant profile. The final outcome of resistance may be biased toward the prevalence of escape quasispecies at baseline, which could be different among individuals.

To systematically assess the drug-resistance profile, determine the fitness landscape of all possible variants, and provide insights into drug-protein interactions, we established a high-throughput genetic profiling approach by combining saturation mutagenesis and next-generation sequencing technology. Examining the fitness of all possible single amino acid substitution at each position in domain IA of NS5A allowed for identification of several functional residues, helping defining structure-function interactions. Differences in the fitness landscape caused by drug treatment revealed a full spectrum of the drug-sensitivity profile of NS5A domain IA in JFH1 background. Variants with high levels of resistance to the drug were identified and their evolutionary paths could be mapped. Positions with altered drug-sensitivity (either increased or decreased) were identified as putative drug-association residues, including 24 and 56, which were identified for the first time. The epistasis within variants at these positions showed correlation with the EC_{50} values of different genotypes, which explains the genotype-specific resistance conferred by the same amino acid substitution. Mathematical simulation revealed different drug resistant profiles at different levels of medication adherences and predicted the emergence of mutations conferring strong resistance. Mutation Y93W, which is the top candidate of drug resistance, requires two-nucleotide change. The genetic barriers and possible evolutionary paths were mapped with the fitness and drug-sensitivity profiles. The systematic analyses of these epistatic interactions can provide insights into the structure and function of the

protein, which is useful in expanding our knowledge of drug-protein interactions and helpful in designing next-generation drugs.

4.3 Results

4.3.1 Fitness landscape mapping of NS5A domain 1A during HCV viral replication.

To systematically explore the vitality of residues for viral replication, we have recently developed a high-throughput profiling system by integrating saturation mutagenesis and next-generation sequencing technology, enabling us to gain insight into protein structure-functions relationships at single amino acid resolution. Saturation mutations were introduced into domain IA of NS5A (amino acids 18-103) by substituting each codon with 3 continuous random nucleotides, allowing for changes of all possible amino acids at each position (Figure 4-1A, see materials and methods for detail). To reduce the possibility of introducing mutations during oligo synthesis, we divided the region into five segments to be mutated. The resultant five mutant libraries were kept separated throughout the screen to reduce sequencing capacity requirement. The DNA libraries, referred as pool 0 (p0) were isolated from more than 21,000 bacteria colonies to cover the complexity for approximately 50 times (Table 4-1A). After *in vitro* transcription, the RNA libraries were reconstituted into viral libraries (p1), which were subsequently subject to selection in hepatocyte cell line (Huh-7.5.1) for 5 rounds (p2_control through p6_control) of infection at low MOI (Table 4-1B). To examine the fitness of mutant viruses with the drug Daclatasvir (BMS_790052) and better understand the mechanisms of drug function, the libraries went through selection in the presence of the drug in parallel for 5 rounds (Figure 4-1B, p2_drug through p6_drug) with gradually increased concentration. After each round, change in frequency of each mutant virus, which reflects the effect of mutation on the fitness, was determined by next-generation sequencing (HiSeq 2000 from Illumina). The sequencing errors were reduced by using Illumina paired-end sequencing to a level that is much lower than the mutation rate, which allowing for the calling of mutants with high confidence [20].

Of the 23.9 million sequence reads that passed the quality filtration, each mutant virus was sequenced on average for about 1200 times, which yielded high statistic significance for quantification. Examination of the input library (p0) revealed 2.9 million reads corresponding to 1634 unique variants in the pool. After each round of selection, the frequency of every mutant and wild-type virus was monitored by deep sequencing. The WT fraction served as an internal benchmark to determine the relative frequency of each mutant. In brief, we normalized the frequency of each mutant to WT fraction within each pool, and calculated the frequency change from p0 to cancel out the difference caused by uneven input. The fitness score (W) of each mutant was given as an average of frequency change between successive rounds, and the selection coefficient (s) representing the difference in fitness between a mutant and WT was determined. Thereby, an s equals to 0 indicates the same fitness as WT virus, a negative s suggests a reduced fitness than WT, and a positive s denotes an increased fitness than WT. A comprehensive heat map was adopted to present the fitness landscape of the region for each variant as shown in Figure 4-2A.

To validate the profiling data, we reconstructed 16 mutant viruses with a wide spectrum of fitness at various positions based on a monocistronic *Renilla* luciferase HCV reporter virus, FNX24_RLuc for individual characterization. Mutations at same position yet with distinct fitness were also included. The mutant reporter virus defective in RNA polymerase activity (NS5B_GNN) served as a negative control and WT as positive. The phenotype of each mutant showed consistency with that of the screen (Figure 4-2B). The selection coefficients measured in screen and validation show strong correlation at high confidence (Figure 4-2C), demonstrating the accuracy of fitness measurements with the profiling method.

In agreement with the critical functions of NS5A required for viral replication, our data show that stop codons are the most unfavorable substitutions at any position of the region (Figure 4-3A) and silent mutations are tolerated, which suggests the effectiveness of the selective pressure and reliable means of monitoring the frequency change during selection.

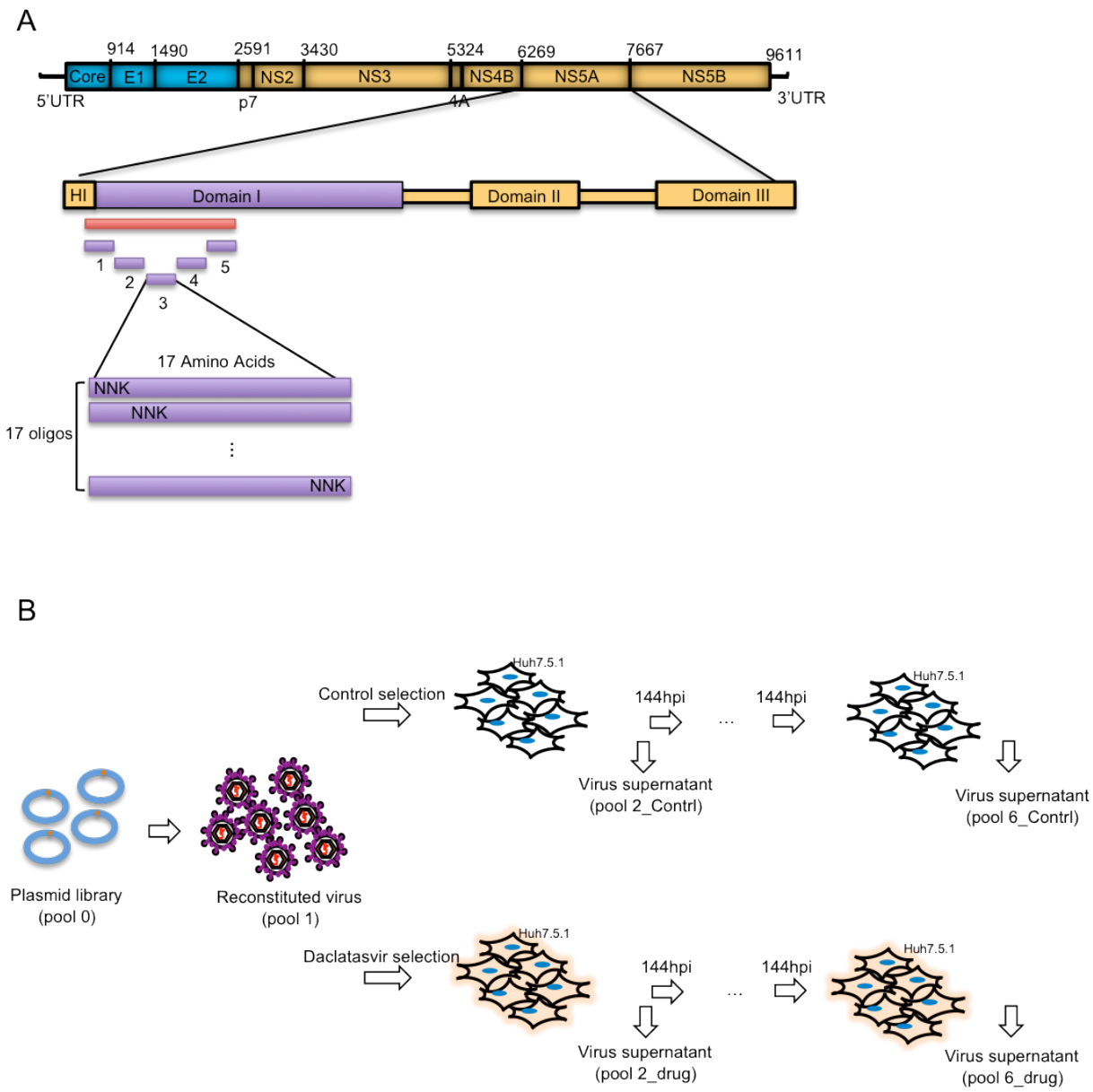


Figure 4- 1. Schematic picture showing the construction of saturation mutant library in a sub-domain of NS5A of HCV. (A) The area to be mutated was divided into 5 small regions, and each of them was composed of 17-18 amino acids. For each library, 17 (or 18) oligos, each of which contained one random codon at different position were synthesized and incorporated into the WT background of HCV. Every library covered all possible mutations for approximately 50 times. (B) The resultant viral library was then selected at the presence or absence of BMS-790052 selection for 5 rounds with gradually increased concentration of the drug (20pM, 20pM, 30pM, 50pM, and 100pM).

A

Segments	# of residues	Possible # of mutants	Colonies	Coverage
Region1	17a.a	544	26,477	49
Region2	17a.a	544	29,000	53
Region3	17a.a	544	21,460	40
Region4	17a.a	544	27,550	51
Region5	18a.a	576	24,360	42

B

Segments	Virus titer	MOI
Region1	$3X10^4$ ffuI/ml	0.15
Region2	$3.5X10^4$ ffu/ml	0.17
Region3	$2X10^4$ ffu/ml	0.1
Region4	$3X10^4$ ffu/ml	0.15
Region5	$5X10^4$ ffu/ml	0.25

Table 4- 1. Summary of input mutant library properties. (A) The size of each input library. Counting of bacterial colonies gave complexity and estimated coverage (average number of times each possible mutation was present in the library) (B) Virus titer and MOIs used for each segment in the first round of infection. To maintain the coverage, we introduced 12 ml of infectious viruses with various titer, which was at least 10 times the complexity of the plasmid library.

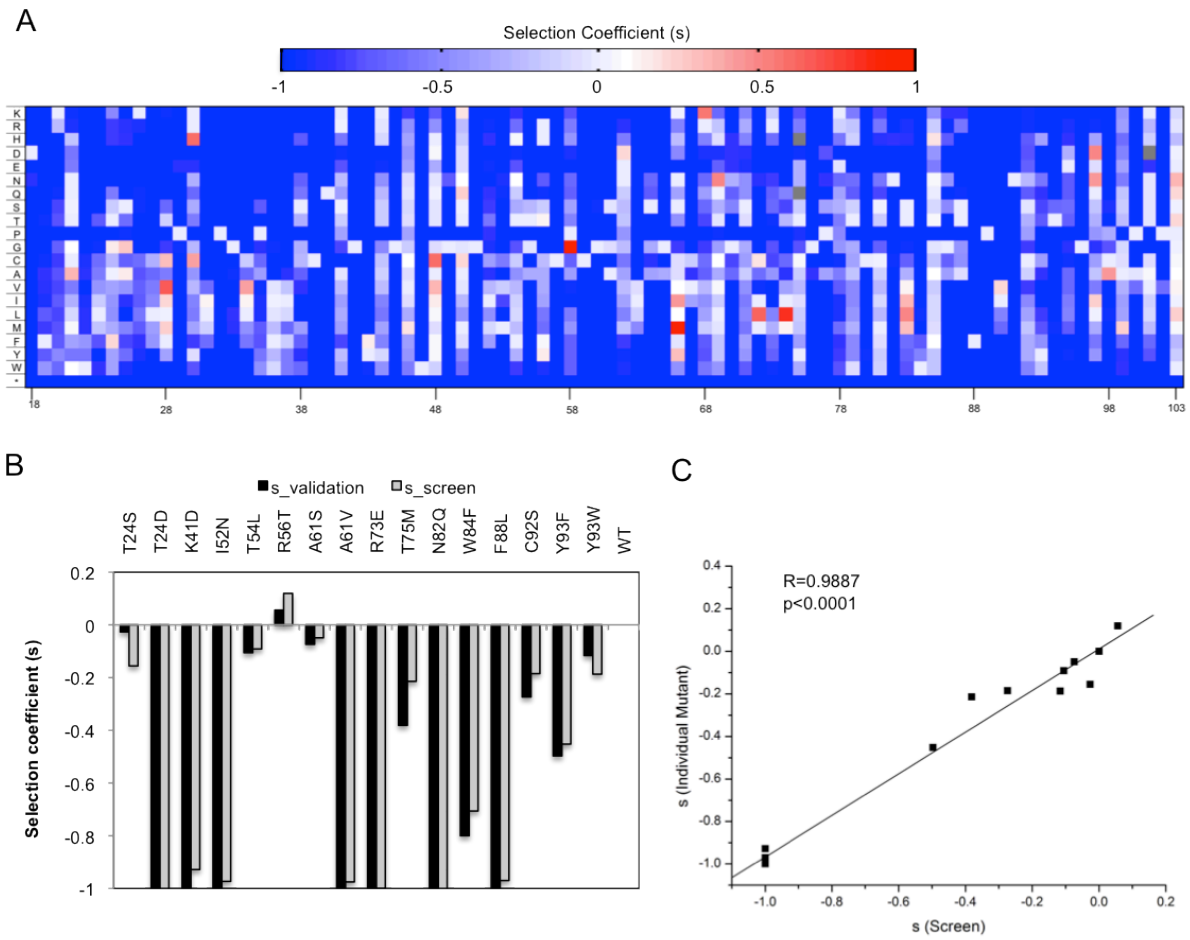


Figure 4- 2. The fitness landscape of amino acids 18-103 in NS5A in virus replication. (A) Heatmap showing the profile of fitness represented as selection coefficient (s) for each variant during viral replication in the tissue culture. Color indicates the fitness of each mutant calculated by selection coefficient in relative to WT. Red suggests a positive s and blue stands for a negative s. s=0 means the same fitness as WT virus; s<0 suggests a reduced fitness than WT and s>0 is increased fitness than WT. (B) Validation of the fitness measurements in the screen. Individual mutant virus containing: T24S, T24D, K41D, I52N, A61S, A61V, R73E, T75M, N82Q, W84F, F88L, C92S, Y93W were reconstructed based on a monocistronic *Renilla* luciferase HCV reporter virus, FNX24-RLuc and recovered by electroporating the viral RNA genome into Huh-7.5.1 cells. To estimate the selection coefficient of each mutant virus, the supernatant was collected and subjected to infect naïve Huh-7.5.1 cells for 2 passages. The replication of the viruses in each round was assayed by measuring the Renilla luciferase activity in the infected cells. (C) The selection coefficients of individual mutants in validation experiment correlate strongly (R=0.9887) and significantly (p<0.0001) with s in screen, suggesting that the fitness massively determined by the profiling platform is accurate.

4.3.2 Fitness profile reveals residues encoding critical functions.

Crystal structure analyses of the NS5A protein suggests the existence of a zinc-coordination motif comprised of four cysteines (Cys 39, Cys 57, Cys 59, and Cys 80), which have been postulated for their roles in maintenance of NS5A structural stability[21-23]. Indeed, the profiling data show that all of the substitutions are completely lethal at these four positions, and in contrast, Cys 92 tolerates most of the amino acids to certain levels, except for positively charged ones (Arg and Lys). This observation confirmed the presence of zinc-binding motif, formed by the four cysteines, excluding Cys 92. Other noticeable residues with significant sensitivity to mutations are at position 88 and 89, which are essential buried residues on the segment that passes through the core of the domain. All of the amino acid substitutions were lethal, indicating that proper volume of side chain is essential for retaining structural stability of the protein.

NS5A is previously known as a membrane associated protein [24-26], and the localization is mediated through the interaction of its N-terminal amphipathic helix membrane anchor with some unknown cellular membrane proteins [27]. It is clearly shown in our data that the residues (Phe19, Trp22) at the hydrophobic face exclusively tolerate for aromatic amino acids substitution, presumably facilitating membrane embedding, while residues (Asp18, Lys21) at the solvent exposed/polar side only allow mutations with same charges, which is perfectly in line with the NMR structural analyses [28, 29]. Moreover, the polar side of the amphipathic helix exhibits an asymmetric distribution of charged residues, with positively- and negatively- charged residues on either border of the polar face, and the opposite charged residues are remarkably unfavorable substitutions. Aspartic acid mutation at position 24, which lies in the basic line of the polar side, results in complete loss of fitness.

We found that position Lys41 (Arg41 in genotype 1b) and Lys44 exhibit a strict and consistent requirement for WT-like fitness: no acidic residues for both positions. These two residues together with the same sites from the other dimeric monomer are in a perfect

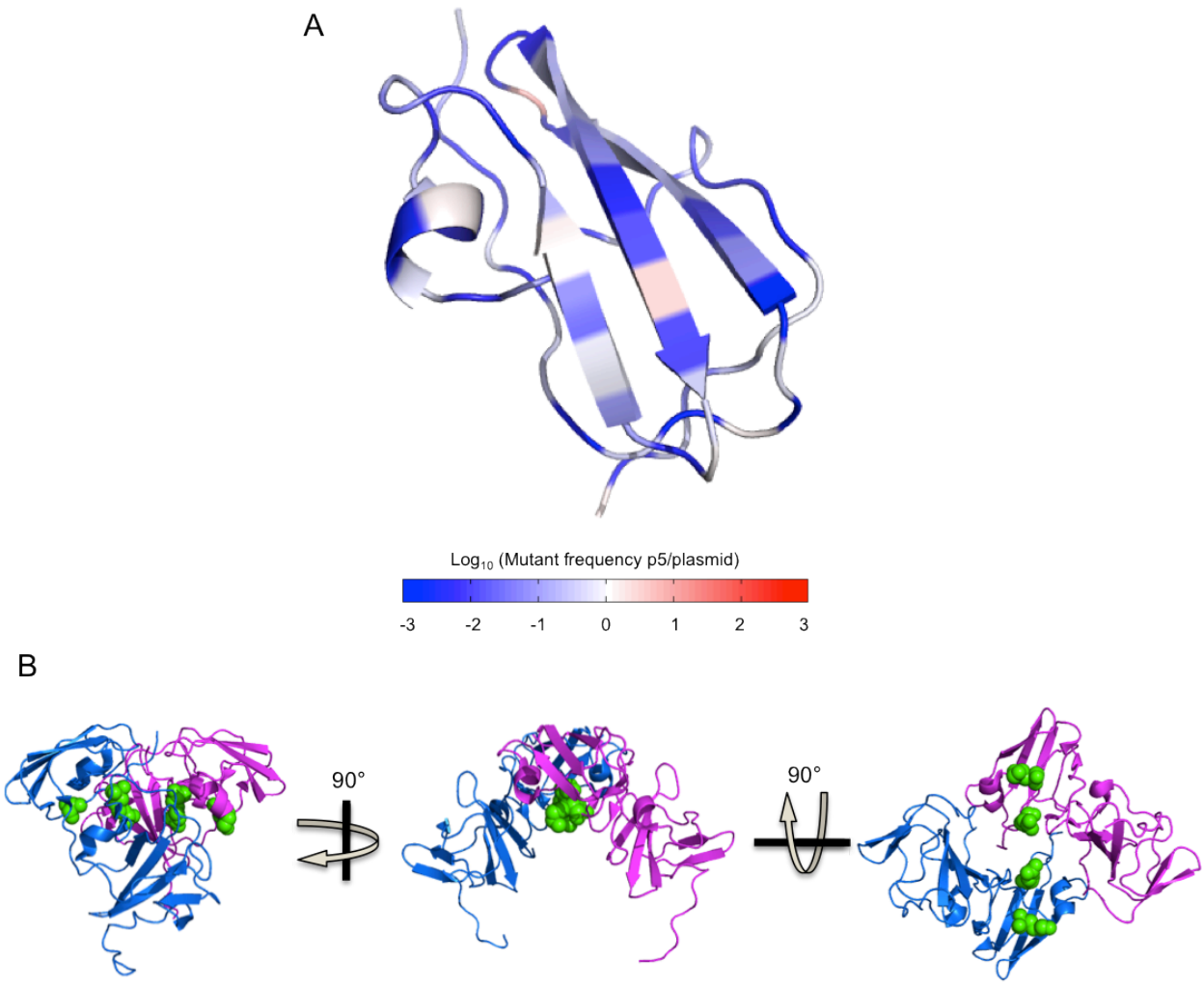


Figure 4- 3. Functional profiling of NS5AD1A reveals the structure-function interactions of the domain. (A) The color-coded structure of domain NS5AD1A shows the essentialness of each residue. The fold change of mutations (log_{10}) in pool 5 at each position was projected onto the structure with a blue-white-red color map in PyMOL. Blue color indicates a decreased frequency of mutations, and red color suggests an increased frequency of mutations. (B) The putative RNA-binding residues are aligned perfectly in the dimer structure. Ribbon diagrams of three rotations of the domain I dimer with residue 41 and 44 highlighted in green spheres.

alignment (Figure 4-3B), and form a basic groove, which has been speculated as an ideal RNA-binding pocket[22]. Thus this preferred physical property of the two residues observed in our profiling provides strong genetic evidence supporting the hypothesis. In fact, further examination suggested that the mutation of K41D resulted in defect of RNA replication.

Another interesting residue worth noting is at position 52, which is a solvent exposed hydrophobic residue on 3D structure [22] (Figure 4-4A). The profiling data suggest that exposed Ile52 strongly favors hydrophobic amino acids with a narrow range of geometry (Met is tolerated, but Val, Luc, Phe, and Cys cause attenuation), and this preference is confirmed with the lethal phenotype of a polar yet similar sized Asn substitution at this position (Figure 4-2B). This preference of tolerated physical property implies an essential role of Ile52 in intermolecular interaction, critical for HCV replication. Since NS5A is an indispensable member of viral replication complex, we then investigated the effect of I52N on RNA replication. The RNA of I52N, along with WT and replication deficient mutant NS5B_GNN were electroporated into Huh-7.5.1 cells and the Renilla luciferase activities in the transfected cells were measured as an indication of genome replication at different time points following transfection. As expected, the luciferase activity accumulates as the viral RNA gets replicated at later time points, and knockout of RNA polymerase enzyme active site results in abolishment of luciferase activity accumulation. Interestingly, mutation at position 52 has no influence on replication efficiency, implying an unimpaired RNA replication function (Figure 4-4C). However, the infectious viral production of the mutant is completely abrogated (Figure 4-4B), propounding the hypothesis that the mutation impairs the viral replication after genome replication, probably at the assembly step.

To further explore the mechanism of replication defect caused by I52N, we then analyzed the subcellular localization of the NS5A proteins in transfected Huh-7.5.1 cells by confocal microscopy. As shown in figure 4-4D, WT NS5A is found to form a reticular network pattern, which involves the nuclear membrane, surrounds the nucleus, and extends through the

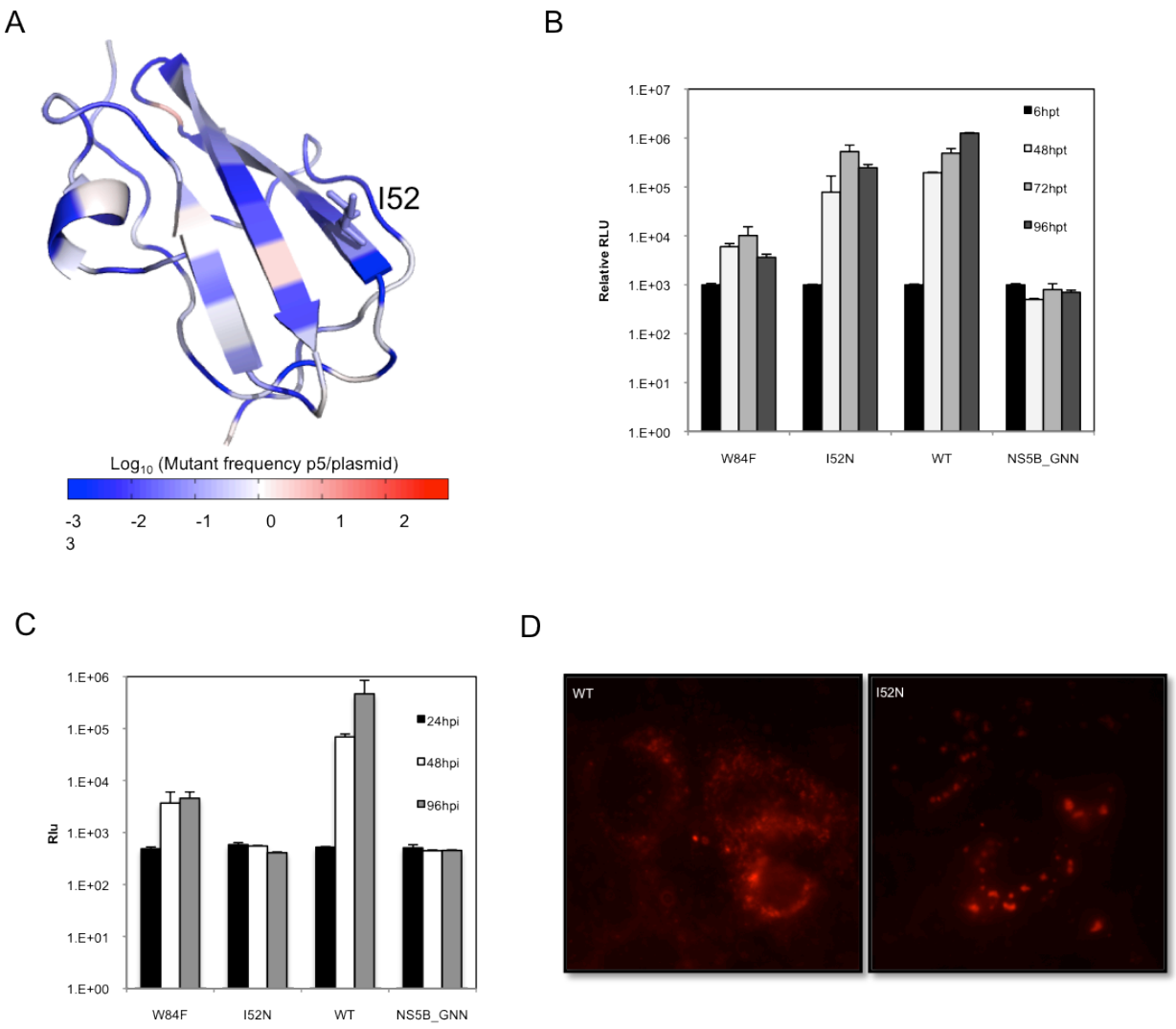


Figure 4- 4. I52 is an essential solvent-exposed hydrophobic residue critical for virus assembly. (A) I52, a solvent-exposed residue on the first β sheet of the domain, only tolerates for hydrophobic amino acids substitutions. (B) Analysis of mutant viral genome replication at 6h, 48h, 72h and 96h post transfection in Huh-7.5.1 cells. (C) Analysis of mutant viral infectivity in the supernatant of transfected cells at 24h, 48h and 96h post infection. (D) Immunostaining of NS5A protein in transfected Huh-7.5.1 cells at 72hpt.

cytoplasm. By contrast, the staining pattern of I52N is distinct from that of WT NS5A, and no nuclear membrane association staining is observed. Instead, the mutant NS5A tends to accumulate in the punctuated clusters, which are presumably the locations of replication complex. Consequently, these data suggest that through interacting with viral or cellular proteins, Ile at position 52 mediates NS5A subcellular localization, which is critical for infectious virus production. Mutations with any polar or bigger hydrophobic residues lead to disruption of the intermolecular interaction required for proper protein localization, thereby restrain infectious virus production.

4.3.3 Daclatasvir drives alteration of fitness landscape.

To interrogate the drug-protein interactions, which subsequently will shed light on mechanistic prediction and facilitating development of next-generation drug, we selected the mutant virus libraries under Daclatasvir treatment for 5 rounds with increased drug concentrations. The fitness of each mutant in 20pM drug selection (Wdrug) was determined as the frequency change in the consecutive rounds and the change in fitness score (Wdrug / Wcontrol) was calculated as presented with the heatmap shown in figure 4-5A. The color quantifies the fold change of fitness for each variant as a result of 20pM drug treatment. The mutations resulting in loss of fitness in control condition are colored in grey. As shown in the colored map, the mutations at positions 28, 31, 38, 92, and 93 are noticeably enriched upon drug selection, variants at positions 21, 56, and 68 are significantly diminished, and the frequency change at positions 24, 30, 62, and 75 is highly dependent on the property of the substituted amino acids. It is tenable to suppose that these positions are the critical residues that mediate the protein-drug interaction, and the dramatic alteration in fitness landscape pattern reflects the gain or loss of function in drug-interaction.

4.3.4 Selection of mutant library in Daclatasvir reveals epistatic interactions among the critical drug associated residues.

It is reported that various HCV genotypes show different sensitivity to Daclatasvir treatment with unclear mechanisms [15, 30]. Intriguingly, alignment of the critical residue sequences from different HCV isolates elucidates the genetic interactions among each other. It is interesting that one genotype 2a strain (J6) is 140-fold more resistant to the drug than its close subtype (JFH1) [30], and examination of the color-coded drug-sensitivity map (Figure 4-5B) explains that the difference is a result of the substitution of Luc to Met at position 31. Similarly, the difference in sensitivity for the two genotype 1a strains is likely due to the contribution of amino acids at position 30. Genotype 1b consistently shows the strongest susceptibility among all of the HCV genotypes, and mutations only yield low levels of resistance to the drug. Thr56 seems to be the major contributor in this case, and mutation R56T has been validated to confer over 5-fold higher drug-sensitivity to the J6/JFH1 virus, a level comparable to genotype 1b. This residue also plays a critical role in high drug-susceptibility of genotype 4b, another genotype highly sensitive to the drug treatment. Mutation Y93H has caught lots of attention for its ability of causing high resistance to the drug in replicon system as well as in patients [17, 19]. Moreover, introduction of this substitution yielded a significantly higher resistance (100-fold to 1000-fold) in the background of 1a, 1b, 2a, 3a and 4a genotypes than their respective original recombinants. In contrast, substitution of T93H in genotype 5a and 6a causes much lower drug-resistance (5-fold), corresponding to the fact that Thr93 is initially a more resistant residue compared to Tyr93, according to the heatmap. Surprisingly, NS5A in genotype 7a, which carries His at position 93 naturally, remains sensitive to the drug. Examination of the 7a sequence reveals a Ser30 that could cause severe attenuation in drug treatment according to the drug-sensitivity profile map. Therefore, the whole spectrum of drug-sensitivity profile correlates with the EC_{50} determined experimentally (Figure 4-5C), and suggests the existence of epistatic interactions among the critical drug associated residues.

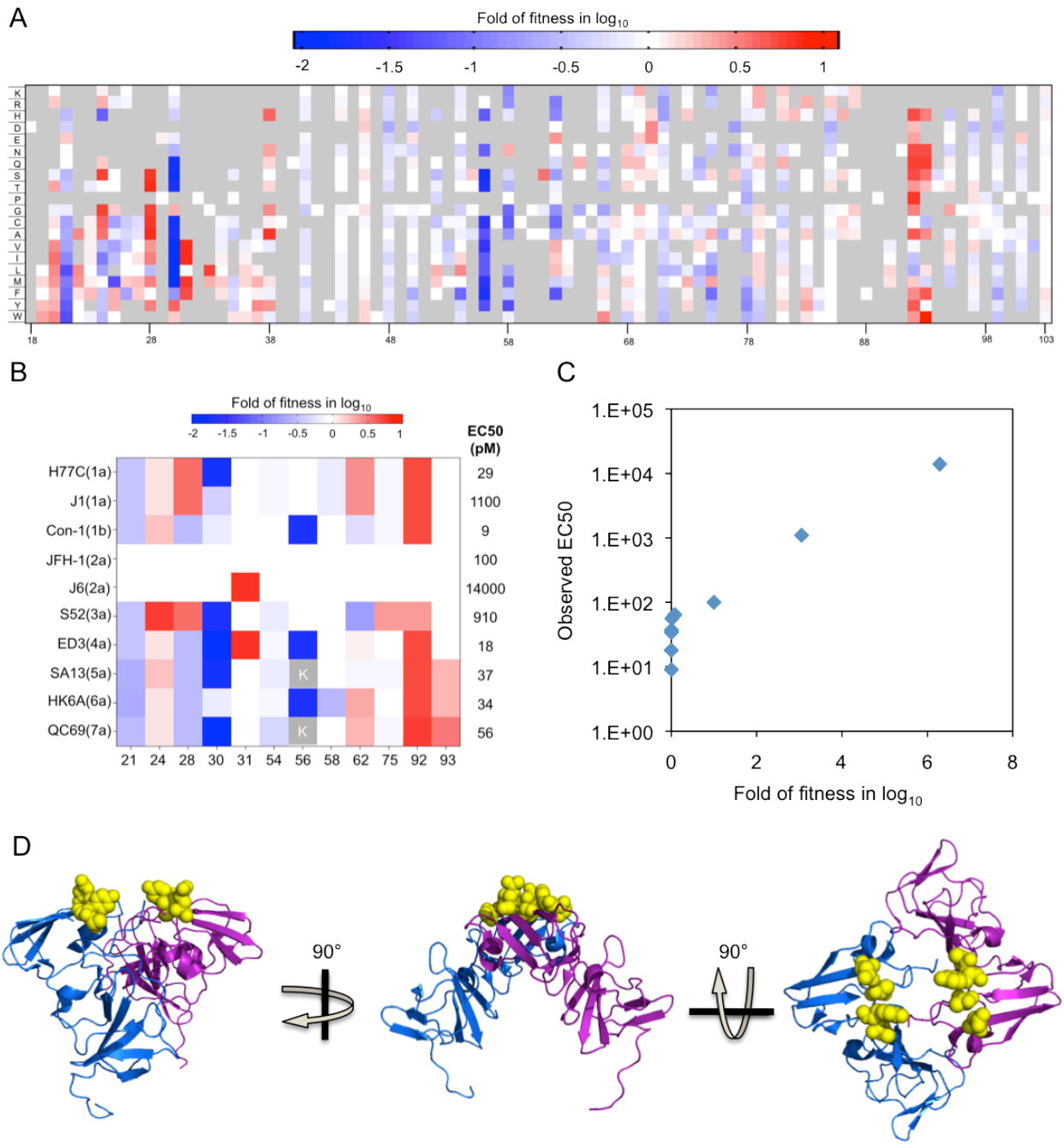


Figure 4- 5. Functional profiling under drug treatment reveals residues mediating drug-protein interaction and the epistatic interactions among them. (A) A heat map shows a fitness landscape shift induced by drug selection (20pM), which is indicated as the ratio of fitness score in drug treatment over the non-drug treated control for each mutant (Wdrug / Wcontrol). The mutants that are lethal in control are shown in grey. (B) 12 candidate residues critical for drug-sensitivity are aligned among 7 genotypes. The fitness fold change of each variant in drug treatment is colored as in figure 3. (C) The fold change of fitness is correlated with EC50 data determined previously.

(D) The drug-associated residues locate on the surface of the protein. Ribbon diagrams of three rotations of the domain I dimer with residue 52,54, 92 and 93 highlighted in yellow spheres.

4.3.5 Functional profiling identifies potential emergence of drug-resistant mutations.

The whole spectrum of fitness landscape under drug treatment offers a complete chart showing the behavior of drug-sensitivity for each possible mutation, and also predicts super strong drug-resistant mutants possibly emerging during treatment. To validate the prediction accuracy, we constructed 10 mutations with a wide range of drug-sensitivity as shown in figure 4-5A. The WT virus was included as a comparison. EC_{50} for each virus was assayed by examining the replication under treatment of a wide range of drug concentration (Figure 4-6A). As suggested by the heat map, the mutation Y93W confers the strongest resistance with a 4780-fold increase of EC_{50} from WT, while mutation R56T results in higher drug-sensitivity with a 5-fold decrease of EC_{50} from WT. The EC_{50} was found to be strongly correlated to the fitness score under drug treatment relative to control (Wdrug / Wcontrol) in the screen according to the regression analyses, demonstrating the reliability of the screen results (Figure 4-6B). The regressed relationship also allows for prediction of the EC_{50} for the rest viable mutants based on the ratio of W, which reveals 105 mutations (out of 677 non-lethal mutants) with at least 2-fold higher EC_{50} than WT. With a given assumption that a patient takes 60mg of Daclatasvir a day, mathematical simulation propounds the emergence of Y93W as a major and strong drug-resistant mutant in patients (Figure 4-6C). This model also predicts that with different levels of medication adherence (60% adherence or 80% adherence), L31I and F28C are only one nucleotide change from the WT sequences, and will become strong resistant strains that are difficult to eradicate. Moreover, these mutations, especially L31I and F28C, cause minor or no obvious fitness loss to the virus *in vitro*, suggesting a low trade-off of function and structure stability, which results in low genetic barriers to resistance.

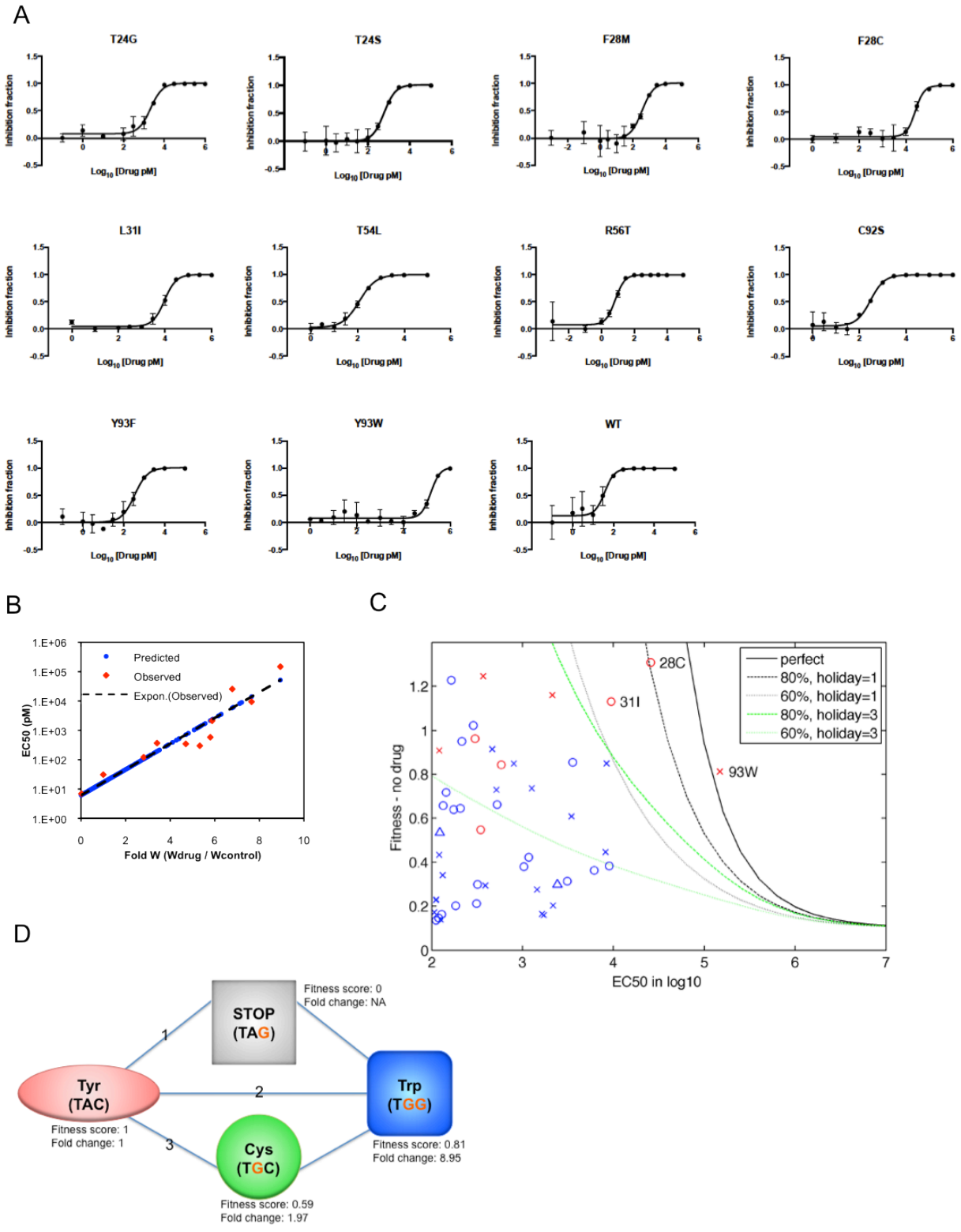


Figure 4- 6. Epistatic interactions between critical drug-associated residues and prediction of drug-resistant mutant emergence in patients. (A) Validation of the drug screen results. Mutant viruses with a variant of drug

resistance or sensitivity were reconstructed and their sensitivity to drug treatment was monitored individually under a serial concentration of drug treatment. The EC_{50} is calculated for each variant. (B) EC_{50} of the drug is estimated for each viable mutant virus by regression analysis on relationship of EC_{50} and fitness fold change (Wdrug / Wcontrol). (C) A mathematical model prediction of mutant virus emergence in different scenarios of drug adherence. (D) Evolutionary paths prediction of Y93W.

4.4 Discussion

NS5A is a multifunctional protein essential for several stages of HCV viral replication. The establishment of *in vitro* replication systems facilitated high-throughput screens for small molecules that inhibit viral replication with unknown mechanisms. Daclatasvir, which is also known as BMS790052, is a new class of antiviral drug discovered by *Bristol-Myers Squibb (BMS).* Daclatasvir effectively inhibits the replication of HCV by targeting NS5A at pM concentrations [15]. The drug is currently in phase III clinical trial and delivering promising preliminary results without a clearly defined working mechanism [13]. To systematically interrogate the protein structure function relationship, we mutagenized the domain IA of NS5A with saturation point mutagenesis, and quantitatively measured the fitness of each variant during its natural infection life cycle to explore the requirements of each position for maintaining viral replication. Application of the drug selection condition provided an insight into the critical drug-associated residues and their epistatic interactions, which will facilitate designing of second-generation drug with higher genetic barriers to resistance.

NS5A is an integral membrane protein with an amphipathic α-helix at its N-terminus (amino acids 1-25), followed by an unstructured linker region (amino acids 26-35) and three functional domains. The N-terminal domain (domain I, amino acids 36-198) is the only one that is structured. Projection of the effect caused by the amino acid substitution at each position onto the structure of NS5A provides an opportunity for identifying structural stability or functional specific requirements of the domain. In line with the structural property, the zinc-associated cysteines are extremely essential and do not tolerate any substitutions. Mutagenesis study with the amphipathic α-helical region reveals a hydrophobic face, that is only tolerant of hydrophobic voluminous amino acid mutations for its membrane association function, and a hydrophilic face with negative or positive charged residues aligned on either side of the edge, which potentially serves as an interaction surface. NS5A was previously reported as an RNA-binding protein associated with the 3'UTR region of HCV, although the specific binding sites remain elusive

[31]. In agreement with the structural prediction [22], our fitness landscape profile found two conserved non-acidic residues (amino acids 41, 44), providing genetic evidence of the existence of the RNA-binding motif.

Ile52 is a solvent-exposed hydrophobic residue, with strong preference on physical property of substitutions. Replacing Ile with Asn, as expected, leads to a defect in infectious virus production. Further examination identifies the lack of impairment of the mutation on genome replication, but disturbance of NS5A localization pattern. Given the fact that NS5A is actively involved in virus assembly, we suspect that the mutation obscures the interaction of NS5A with viral or cellular partner, leading to its deficient in viral particle packaging.

Tremendous efforts have been put into studies for identification and characterization of drug resistant mutations, since the discovery of the effective NS5A inhibitor. The mode of drug action on NS5A still remains elusive. Selection of the saturation mutant library allows for quantitative determination of drug-sensitivity of all possible variants in the domain IA of NS5A and identification of mutations conferring strong resistance, e.g. Y93W, F28C, which should be carefully monitored. Mutations at position 24 (T24G, T24S), which were identified for the first time, also show a significant elevated level of drug resistance *in vitro*. The previously reported resistant mutations have been recaptured in our screen with similar levels of resistance, except for some with differences in EC_{50} values presumably caused by the genotype backgrounds [14, 15, 17, 18]. More importantly, thoroughly exploring the fitness and drug-sensitivity of all possible variants with saturation mutagenesis allows for picturing and prediction of the possible evolutionary routes for emergence of drug-resistant mutations. Some resistant mutations, like Y93W, require two-nucleotide change from the WT sequence, and therefore is less likely to be observed with natural adaptation method starting with single clone. This mutation, however, is only one nucleotide different from Cys, which has already been observed at 10-15% frequency in patients only after a 4-day treatment with the drug [19]. The arising Y93C in patients despite the deleterious fitness (W=0.59) and weak drug resistance (fold of fitness=1.97) compared with

WT demonstrates its advantageous fitness under the selection pressure. Thus, it is tenable to expect the emergence of the mutation Y93W, which requires only one more nucleotide substitution (T to G), and confers substantially stronger resistance (fold of fitness=8.95, EC_{50}=148nM, 4773 folds of resistance) with less fitness loss (W=0.81) (Figure 4-6D). The mathematical simulation predicts that if emerges, this mutation will obscure the effectiveness of the drug and develop resistance in spite of good medication adherence (Figure 4-6C). Development of the escape mutation should be aware of along the treatment, and therefore combinational therapy is recommended to prevent the development and dissemination of resistant strains throughout or among infected individuals.

The drug screen also identified a spectrum of residues with altered drug sensitivity, including both resistance and hypersensitivity, conferred by the mutations. Sequence alignment of the seven genotypes reveals the epistatic interactions of the critical residues. The epistasis explains the low drug resistance in genotype 7a, naturally carrying a His at position 93, which causes a high level of resistance upon introduction into other genotype backgrounds [30]. Besides, position 56 is identified as a critical residue, mediating drug-binding affinity. Introduction of Thr at 56 causes substantial drug sensitivity, to which the high level of drug susceptibility of genotype 1b is likely attributed [32]. The variations seen at these positions and their differences in drug sensitivity are attributed to the wide range of EC_{50} observed in different genotypes [30, 32, 33].

Daclatasvir (BMS-790052) has been implicated in to interactions with domain 1A of NS5A, which results in the absence of NS5A from replication complex and blockage in viral genome replication [34]. Structural analyses of NS5A previously has yielded simultaneous existence of two different arrangements of domain I homodimer with non-overlapped association interfaces on the opposite side of each monomer [22, 23]. Interestingly, in either form of dimer structure, the drug sensitivity-determining residues locate away from the dimerization interfaces, which supports previous results showing that the drug does not interfere

with the dimerization of NS5A [34]. Instead, in one of the dimers, these residues localize on the surface of domain 1A and the physically adjacent linker region (amino acids 26-35) connecting the N-terminal amphipathic helix and the core of domain I (Figure 4-5D). It raises the possibility of them forming a putative drug binding pocket and preferentially stabilizing one dimer organization than the other. The two possible dimerization interfaces on the opposite side of the monomer leads to the prediction of superhelical array organization where the monomer is polymerized with alternative interfaces [23]. We hypothesize that the drug induces configuration shift towards one of the dimerization arrangements, the accumulation of which then stimulates the oligomerization. As a result, it depletes the NS5A required in replication complex.

Although this hypothesis is supported by the extremely low working concentration even in replicon cell lines that homogenously harbor highly active replication complex, it does not exclude the possibility that the drug directly competes with cellular or viral factors for NS5A binding, and as a consequence, abolishes membranous web formation. In fact, studies show that NS5A interacts with many host factors to hijack their cellular functions for facilitating viral replication at different stages. PI4KIIIα is a phosphatidylinositol 4-kinase identified as a host kinase targeted and relocated to HCV replication complex by NS5A [35-39]. Co-immunoprecipitation of NS5A deletion mutants and PI4KIIIα mapped the interaction to domain I [37, 38]. There is evidence demonstrating that the drug causes reduced interaction between NS5A and PI4KIIIα, and therefore hampers the viral genome replication. However, more direct evidences will be needed to elucidate the real mechanisms.

Overall, profiling the NS5A domain IA by combining saturation mutagenesis and next-generation sequencing allows thorough mapping of the fitness of each variant, and provides genetic evidences for functional residues essential for virus replication, including putative RNA binding sites, buried residues for structural stability, and solvent-exposed hydrophobic residue indispensable for virus assembly. Moreover, a distinct fitness landscape determined with 20pM of drug selection discloses the drug sensitivity-determining residues and their epistatic

interactions. It also presents a complete spectrum of single amino acid mutations resulting in high levels of drug resistance. Although it is an artificial mutagenesis process, the fitness landscapes with and without drug treatment provide a reference of genetic barriers for every possible change during viral evolution. This information may facilitate development of second-generation drugs with higher barriers. Furthermore, complementation of attenuated mutants in future studies will provide additional information on genetic linkage between amphipathic helix, flexible linker region, and the rest of domain, which further benefits mechanistic predictions. We anticipate that this systematic profiling approach will be generally applicable to studying any drug-protein interactions, identifying escape mutants, and elucidating the mechanisms for novel drug development.

4.5 Materials and Methods

Cell culture, viruses and plasmids

The Huh-7.5.1 cell line was kindly provided by Dr. Francis Chisari from the Scripps Research Institute, La Jolla. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 10% of fetal bovine serum (FBS), 10mM non-essential amino acids (Invitrogen, Carlsbad, USA), 10mM HEPES, penicillin (100 units/ml), streptomycin (100 mg/ml), and 2mM L-glutamine at 37°C with 5% CO_2. 293T cells were grown in DMEM with 10% FBS, 100 units/ml of penicillin, and 100mg/ml of streptomycin.

pFNX-HCV is the plasmid we synthesized based on the chimeric sequence of J6/JFH1 virus. We introduced 7 nucleotide substitutions, resulting in synonymous mutations to the genome. The construct and sequence are available upon request.

Construction of saturation mutant library in sub-domain IA of NS5A of HCV. The area to be mutated (86 amino acids) was divided into 5 small regions, and each of them is composed of

17-18 amino acids. For each region, 17 (or 18) oligos, each of which contains one random codon (NNK) at the desired position were synthesized from IDT. This mutagenesis results in all possible amino acid substitutions at a given position, which will facilitate exploring the function of each possible variant. The oligos contained a BtsI recognition site on each end, which allows for generating “stick ends” matching the ends of the cassettes. The cassettes were established by amplifying the fragments (from pFNX-HCV) franking the region desired to be mutated with primers containing BtsI recognition site, and digested with BtsI enzyme (NEB) to produce the “stick ends” matching the ones on the oligos. The oligos and the cassettes were ligated with T4 DNA ligase (Invitrogen) overnight at 16°C and purified with PCR columns (Invitrogen). The ligated product was subcloned into the pFNX-HCV vector via BamHI and RsrII restriction enzymes and transformed. Total approximately 50,000 colonies were collected for the library. Every library covers all possible mutations for approximately 70 folds.

Selection of the saturation mutant library under drug treatment

The mutant virus library (12ml) was subjected to infect naïve Huh-7.5.1 cells (4million) at M.O.I around 0.2 with or without BMS790052 treatment. The supernatant was collected at 144hpi and passed on to infect naïve cells for the second round of selection. The concentration of the drug for each round of selection is indicated in Fig 1C. After five rounds of selection, the viral genome was recovered from the supernatant, and the mutated regions were amplified with PCR and processed following the standard sample preparation protocol for HiSeq 2000 sequencing. Each library was tagged molecular barcodes, unique 7-bp sequences, allows for the identification and study of relative fitness levels in each selection pool.

Sequence alignment for saturation mutagenesis library screen.

Burrow-Wheeler Aligner was used to map the pair-end read by allowing 5 mismatches. Sequencing error was corrected by reads pairing. SAMtools and BamTools were employed for sequence analyses. Custom Python script was created for the other downstream data analyses.

Immunofluorescent assay

At 48h post transfection, Huh-7.5.1 cells in 24-well plates were fixed with 4% paraformaldehyde for overnight. After 3 times wash with PBS, the cells were permeablized 0.05% Triton-x 100 for 10min and with blocked with 10% FBS, 3% BSA. The cells were incubated with mouse anti-NS5A antibody at 4°C for overnight. Secondary antibodies, goat anti-mouse IgG conjugated to Alex 593 (Invitrogen) was added and incubated for 1 hr at room temperature. Cells were washed for 3 times with PBS and nuclei was stained with Hoechst 33342 (Invitrogen) before imaging.

Measuring virus titer

The virus titer was assayed by measuring the foci-forming unit of infectious virus particles per ml of supernatant. Culture supernatant taken from different time points, e.g. 48hpt, 72hpt and 96hpt were diluted in a 10-fold serial manner before applying onto the Huh-7.5.1 cells seeded in 96-well plates. 72hpt, the cells were fixed with 100% methanol and immunostained for core antigen with mouse monoclonal anti-core primary antibody C7-50 (Abcam, Cambridge, USA), followed by goat anti-mouse-Alex 488 secondary antibody. Virus titer was determined by counting the number of core antigen-positive foci at highest dilutions.

Quantitatively detection of viral genome with Q-PCR

Total RNA was isolated from transfected or infected cells by Trizol (Invitrogen) following the instruction of the manufacture. 500ng of RNA was used for reverse transcription using q-Script cDNA Synthesis Kit (Quantas). Primers used for HCV-specific qPCR was conducted by using primer pair: 5'-AGAGCCATAGTGGTCTGCG-3' and 5'-CTTTCGCAACCCAACGCTAC-3'; Actin: 5’-ACCTTCTACAATGAGCTGCG-3′ and 5′-CCTGGATAGCAACGTACATGG.

4.6 Bibliography

1. Appel, N., et al., *From structure to function: new insights into hepatitis C virus RNA replication.* J Biol Chem, 2006. **281**(15): p. 9833-6.

2. Lemon, S.M., et al., *Hepatitis C Virus*, in *Fields of Virology*, D.M. knipe and P.M. Howley, Editors. 2007, Lippincott Williams & Wilkins: Philadelphia. p. 1253.

3. Feld, J.J. and J.H. Hoofnagle, *Mechanism of action of interferon and ribavirin in treatment of hepatitis C.* Nature, 2005. **436**(7053): p. 967-972.

4. Enserink, M., *First Specific Drugs Raise Hopes for Hepatitis C.* Science, 2011. **332**(6026): p. 159-160.

5. Schlutter, J., *Therapeutics: New drugs hit the target.* Nature, 2011. **474**(7350): p. S5-S7.

6. Robinson, M., et al., *Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10290-5.

7. Thompson, A.J., S.A. Locarnini, and M.R. Beard, *Resistance to anti-HCV protease inhibitors.* Curr Opin Virol, 2011. **1**(6): p. 599-606.

8. Wakita, T., et al., *Production of infectious hepatitis C virus in tissue culture from a cloned viral genome.* Nat Med, 2005. **11**(7): p. 791-6.

9. Lindenbach, B.D., et al., *Complete replication of hepatitis C virus in cell culture.* Science, 2005. **309**(5734): p. 623-6.

10. Zhong, J., et al., *Robust hepatitis C virus infection in vitro.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9294-9.

11. Lohmann, V., et al., *Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line.* Science, 1999. **285**(5424): p. 110-113.

12. Blight, K.J., A.A. Kolykhalov, and C.M. Rice, *Efficient Initiation of HCV RNA Replication in Cell Culture.* Science, 2000. **290**(5498): p. 1972-1974.

13. De Clercq, E., *The race for interferon-free HCV therapies: a snapshot by the spring of 2012.* Rev Med Virol, 2012.

14. Lemm, J.A., et al., *Identification of Hepatitis C Virus NS5A Inhibitors.* Journal of Virology, 2010. **84**(1): p. 482-491.

15. Gao, M., et al., *Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect.* Nature, 2010. **465**(7294): p. 96-100.

16. Tellinghuisen, T.L., et al., *Studying hepatitis C virus: making the best of a bad virus.* J Virol, 2007. **81**(17): p. 8853-67.

17. Fridell, R.A., et al., *Resistance Analysis of the Hepatitis C Virus NS5A Inhibitor BMS-790052 in an In Vitro Replicon System.* Antimicrobial Agents and Chemotherapy, 2010. **54**(9): p. 3641-3650.

18. Wang, C., et al., *Hepatitis C Virus RNA Elimination and Development of Resistance in Replicon Cells Treated with BMS-790052.* Antimicrobial Agents and Chemotherapy, 2012. **56**(3): p. 1350-1358.

19. Fridell, R.A., et al., *Genotypic and phenotypic analysis of variants resistant to hepatitis C virus nonstructural protein 5A replication complex inhibitor BMS-790052 in Humans: In Vitro and In Vivo Correlations.* Hepatology, 2011. **54**(6): p. 1924-1935.

20. Fowler, D.M., et al., *High-resolution mapping of protein sequence-function relationships.* Nat Meth, 2010. **7**(9): p. 741-746.

21. Tellinghuisen, T.L., et al., *The NS5A Protein of Hepatitis C Virus Is a Zinc Metalloprotein.* Journal of Biological Chemistry, 2004. **279**(47): p. 48576-48587.

22. Tellinghuisen, T.L., J. Marcotrigiano, and C.M. Rice, *Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase.* Nature, 2005. **435**(7040): p. 374-379.

23. Love, R.A., et al., *Crystal Structure of a Novel Dimeric Form of NS5A Domain I Protein from Hepatitis C Virus.* Journal of Virology, 2009. **83**(9): p. 4395-4403.

24. Hijikata, M., et al., *Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus.* Proceedings of the National Academy of Sciences, 1993. **90**(22): p. 10773-10777.

25. Song, J., et al., *Nuclear localization and intramolecular cleavage of N-terminally deleted NS5A protein of hepatitis C virus.* Virus Research, 2000. **69**(2): p. 109-117.

26. Elazar, M., et al., *Amphipathic Helix-Dependent Localization of NS5A Mediates Hepatitis C Virus RNA Replication.* Journal of Virology, 2003. **77**(10): p. 6055-6061.

27. Cho, N.-J., et al., *Binding Dynamics of Hepatitis C Virus' NS5A Amphipathic Peptide to Cell and Model Membranes.* Journal of Virology, 2007. **81**(12): p. 6682-6689.

28. Brass, V., et al., *An Amino-terminal Amphipathic α-Helix Mediates Membrane Association of the Hepatitis C Virus Nonstructural Protein 5A.* Journal of Biological Chemistry, 2002. **277**(10): p. 8130-8139.

29. Penin, F.o., et al., *Structure and Function of the Membrane Anchor Domain of Hepatitis C Virus Nonstructural Protein 5A.* Journal of Biological Chemistry, 2004. **279**(39): p. 40835-40843.

30. Scheel, T.K.H., et al., *Recombinant HCV Variants With NS5A From Genotypes 1,7 Have Different Sensitivities to an NS5A Inhibitor but Not Interferon-α.* Gastroenterology, 2011. **140**(3): p. 1032-1042.e6.

31. Huang, L., et al., *Hepatitis C Virus Nonstructural Protein 5A (NS5A) Is an RNA-binding Protein.* Journal of Biological Chemistry, 2005. **280**(43): p. 36417-36428.

32. Okamoto, Y., et al., *Development of recombinant hepatitis C virus with NS5A from strains of genotypes 1 and 2.* Biochem Biophys Res Commun, 2011. **410**(3): p. 404-9.

33. Wang, C., et al., *In vitro activity of BMS-790052 on hepatitis C virus genotype 4 NS5A.* Antimicrob Agents Chemother, 2012. **56**(3): p. 1588-90.

34. Lee, C., et al., *The hepatitis C virus NS5A inhibitor (BMS-790052) alters the subcellular localization of the NS5A non-structural viral protein.* Virology, 2011. **414**(1): p. 10-8.

35. Berger, K.L., et al., *Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication.* Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7577-82.

36. Berger, K.L., et al., *Hepatitis C virus stimulates the phosphatidylinositol 4-kinase III alpha-dependent phosphatidylinositol 4-phosphate production that is essential for its replication.* J Virol, 2011. **85**(17): p. 8870-83.

37. Lim, Y.S. and S.B. Hwang, *Hepatitis C virus NS5A protein interacts with phosphatidylinositol 4-kinase type IIIalpha and regulates viral propagation.* J Biol Chem, 2011. **286**(13): p. 11290-8.

38. Reiss, S., et al., *Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment.* Cell Host Microbe, 2011. **9**(1): p. 32-45.

39. Tai, A.W. and S. Salloum, *The role of the phosphatidylinositol 4-kinase PI4KA in hepatitis C virus-induced host membrane rearrangement.* PLoS One, 2011. **6**(10): p. e26300.

CHAPTER 5

CONCLUSIONS AND PERSPECTIVES

Hepatitis C virus has been well recognized as a major cause of human liver diseases. Persistent infection of HCV leads to severe damages in livers and puts patients at huge risks of developing hepatitis, cirrhosis, and hepatocellular carcinoma, which is an indication of liver transplantation. However, the residual viruses in the body quickly infect and occupy the newly transplanted liver and leads to failure of the treatment. Studies aiming to evaluate the effect of HCV elimination on the hepatocellular carcinoma incidence indicate that successful eradication of HCV can reduce the risk of liver failure and liver cancer development. However, due to the asymptomatic disease progression and limited treatment strategies, the mortality associated with HCV chronic infection in the United States is as much as that of HIV. Therapeutic vaccine is not available for HCV treatment, and the standard treatment approach has been the combination therapy with IFN-α and ribavirin, which offers limited response rate but serious side effects. The recent approval of the two protease inhibitors, designed specifically targeting protease of HCV, has improved the treatment efficacy to about 70% in the most prevalent genotype as a supplement of standard interferon plus ribavirin therapy. However, the efficacy is highly dependent on the genotype and hindered by the arising resistant mutations. Given the fact that the virus replicates so fast and error-prone, escape mutants are expected according to the lesson that we learned from HIV battle. Therefore, there is a pressing need for understanding the mechanism(s) by which HCV develops resistance to the IFN treatment, meanwhile developing novel and more effective antiviral treatments.

It’s known that IFN response can defend mammalian host from virus infection by blocking viral genome replication, degrading viral genome, inhibiting viral protein synthesis and affecting viral protein post translational process to control virus replication at different steps. However, viruses have evolved to utilize different strategies to evade from multiple layers of immune response. Identification of any domain(s) critical for HCV to establish resistance to IFN treatment will offer novel antiviral treatment strategies to overcome viral resistance and will shed light on vaccine development.

The approval of the two protease inhibitors recently has increased the viral responsive rate in genotype 1b patients to about 70%. However, a lesson that we should learn from HIV treatment is that emergence of drug resistant variants always and ultimately results in failure of sustained viral response. Ever since the development of HCV experimental systems, tremendous strides have been made in large-scale chemical screens to search for novel antiviral compounds. These efforts have led to the identification of viral inhibitors with promising efficacy, yet with indefinite working mechanisms. One of the NS5A inhibitors, also known as Daclatasvir (BMS-790052), is currently in phase III of drug development and shows strong efficacy and potency in preliminary clinical trials. Despite arising of resistant mutations during natural adaptive selection under drug treatment, detailed working mechanism and the full spectrum of potential resistant mutations remain elusive.

In this thesis work, by combining high-resolution genome-scale mutagenesis and next-generation sequencing technology, we have established a systematic genetic profiling platform, which aims to map the functional residues on HCV genome in an unbiased manner. In particular, this approach enables to provide a landscape of anti-IFN functional domains and an insight into residues mediating drug-protein interactions.

Chapter two describes a high-throughput, genome-scale genetic profiling platform, which assists to systematically profile the anti-IFN functional domains in HCV genome and provide insights into mechanisms that govern IFN evasion of the virus. A highly complex library of mutant HCV was constructed by randomly inserting 15-nucleotides (nt) at almost every base pair position in the virus genome, and passaged in the presence or absence of IFN-α. Solexa sequencing was performed to quantitatively and accurately determine the frequency of each mutant in the populations, which led us to identify four regions, including N-terminal core-, N-terminal p7-, domain II & III of NS5A-coding regions, and 3'UTR showing increased sensitivity to IFN-α treatment when mutated. Further mapping of the critical residues in N-terminus of core protein by utilizing saturation mutant library revealed that phenylalanine 24 was one of the

important residues for the anti-IFN function of core protein. In particular, phenylalanine 24 is required to impede the phosphorylation of STAT1 thereby blocking IFN signaling. This profiling method provides a basic and generalizable approach for viral genetics, enabling systematical mapping of protein functions at single amino acid resolution. In future studies, more detailed mechanism can be explored to study why the residue F24 is critical in regulating the phosphorylation of STAT1. Co-immunoprecipitation or co-localization analysis of core and STAT1 can be carried out during the course of viral infection to investigate whether the regulation is through a direct interaction.

Chapter three reports a novel immune evasion viral protein, p7, which is previously known as an ion channel protein with unclear function for viral replication. The genomic profiling of HCV genome in IFN treatment described in chapter two suggests that mutations in p7 protein results in increased IFN-sensitivity of the mutants. This hypersensitivity can be rescued with over-expressed WT p7 protein, suggesting that p7 encodes functions essential for evading IFN responses. ISG cDNA screen identified that over-expression of IFI6 significantly inhibits the replication of p7 mutant virus in comparison to that of WT. Interestingly, p7 physically interacts with IFI6 based on sub-cellular localization and co-IP analyses. These genetic and molecular interactions between p7 and IFI6 propounds the hypothesis that p7 exhibits the anti-IFN function through obscuring the antiviral function of IFI6. Although IFI6 is originally identified as a gene that is upregulated upon stimulation of interferon, the exact role of controlling virus replication remains elusive. In future studies, RNA interference of endogenous IFI6 can be used to confirm the antiviral functions especially on p7 mutant virus. More definitive experiments such as IFA and EM can be performed to determine the sub-cellular localization of IFI6 during WT or p7 mutant HCV infection. Yeast two hybrid or protein mass spectrometry experiments can be carried out to identify its associated proteins for further mechanistic studies.

Chapter IV presents a comprehensive functional analysis of a saturation mutant virus library under different selection conditions. The fitness landscape of protein NS5A domain IA in

tissue culture revealed functional residues essential for protein structure stability or functions required for viral replication. Selection under drug treatment identified a panel of residues that influence the drug-sensitivity of the virus. Epistatic interactions observed among these residues explain the HCV genotype-specific difference in drug-sensitivity. These data also allow for systematical mapping of the genetic barriers for all possible mutations and prediction of evolutionary paths for potential emerging escape variants.

Taken together, we have functionally profiled the HCV genome at genomic scale and single amino acid resolution, and defined the essential residues for viral replication, as well as their functions in evading host immune responses. With the same concept of genetic profiling, we systematically characterized the residues in mediating drug interactions, determined the genetic barriers to resistance and predicted the potential emergence of escape mutants. We suggest that the genetic profiling platform described in this thesis can be generally applied in interrogating virus-host interactions and chemical-target interactions, which will provide comprehensive knowledge on new therapeutic strategies to overcome persistent HCV infection and resistance.