Functional Evaluation of Causal Mutations Identified in Human Genetic Studies

by

Yi-Fan Lu

Department of Molecular Genetics and Microbiology
Duke University

Date:_______________________

Approved:

___________________________

David B. Goldstein, Co-Supervisor

___________________________

Douglas A. Marchuk, Co-Supervisor

___________________________

David M. Tobin

___________________________

Georgia D. Tomaras

___________________________

Yong-Hui Jiang

___________________________

Shelton S. Bradrick

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Molecular Genetics and Microbiology in the Graduate School
of Duke University

2016
ABSTRACT

Functional Evaluation of Causal Mutations Identified in Human Genetic Studies

by

Yi-Fan Lu

Department of Molecular Genetics and Microbiology
Duke University

Date:_______________________
Approved:

___________________________
David B. Goldstein, Co-Supervisor

___________________________
Douglas A. Marchuk, Co-Supervisor

___________________________
David M. Tobin

___________________________
Georgia D. Tomaras

___________________________
Yong-Hui Jiang

___________________________
Shelton S. Bradrick

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Molecular Genetics and Microbiology in the Graduate School of Duke University

2016
Abstract

Human genetics has been experiencing a wave of genetic discoveries thanks to the development of several technologies, such as genome-wide association studies (GWAS), whole-exome sequencing, and whole genome sequencing. Despite the massive genetic discoveries of new variants associated with human diseases, several key challenges emerge following the genetic discovery. GWAS is known to be good at identifying the locus associated with the patient phenotype. However, the actually causal variants responsible for the phenotype are often elusive. Another challenge in human genetics is that even the causal mutations are already known, the underlying biological effect might remain largely ambiguous. Functional evaluation plays a key role to solve these key challenges in human genetics both to identify causal variants responsible for the phenotype, and to further develop the biological insights from the disease-causing mutations.

We adopted various methods to characterize the effects of variants identified in human genetic studies, including patient genetic and phenotypic data, RNA chemistry, molecular biology, virology, and multi-electrode array and primary neuronal culture systems. Chapter 1 is a broader introduction for the motivation and challenges for functional evaluation in human genetic studies, and the background of several genetics
discoveries, such as hepatitis C treatment response, in which we performed functional characterization.

Chapter 2 focuses on the characterization of causal variants following the GWAS study for hepatitis C treatment response. We characterized a non-coding SNP (rs4803217) of \textit{IL28B (IFNL3)} in high linkage disequilibrium (LD) with the discovery SNP identified in the GWAS. In this chapter, we used inter-disciplinary approaches to characterize rs4803217 on RNA structure, disease association, and protein translation.

Chapter 3 describes another avenue of functional characterization following GWAS focusing on the novel transcripts and proteins identified near the \textit{IL28B (IFNL3)} locus. It has been recently speculated that this novel protein, which was named IFNL4, may affect the HCV treatment response and clearance. In this chapter, we used molecular biology, virology, and patient genetic and phenotypic data to further characterize and understand the biology of IFNL4. The efforts in chapter 2 and 3 provided new insights to the candidate causal variant(s) responsible for the GWAS for HCV treatment response, however, more evidence is still required to make claims for the exact causal roles of these variants for the GWAS association.

Chapter 4 aims to characterize a mutation already known to cause a disease (seizure) in a mouse model. We demonstrate the potential use of multi-electrode array (MEA) system for the functional characterization and drug testing on mutations found in neurological diseases, such as seizure. Functional characterization in neurological
diseases is relatively challenging and available systematic tools are relatively limited. This chapter shows an exploratory research and example to establish a system for the broader use for functional characterization and translational opportunities for mutations found in neurological diseases.

Overall, this dissertation spans a range of challenges of functional evaluations in human genetics. It is expected that the functional characterization to understand human mutations will become more central in human genetics, because there are still many biological questions remaining to be answered after the explosion of human genetic discoveries. The recent advance in several technologies, including genome editing and pluripotent stem cells, is also expected to make new tools available for functional studies in human diseases.
Contents

Abstract.............................................................................................................................................. iv

List of Tables ........................................................................................................................................... xii

List of Figures ......................................................................................................................................... xiii

Acknowledgements ............................................................................................................................. xv

1. Introduction .......................................................................................................................................... 1

  1.1 Functional evaluation in human genetics ..................................................................................... 1

    1.1.1 Human disease and genetic discovery .................................................................................. 1

    1.1.2 Motivation: moving from genetic association to biology ..................................................... 4

    1.1.3 Unknown causal variants ...................................................................................................... 4

    1.1.4 Investigate biology behind disease-causing mutations ......................................................... 5

    1.1.5 Other challenges in functional evaluation in human genetics .............................................. 6

  1.2 Hepatitis C treatment response ..................................................................................................... 8

    1.2.1 HCV biology ......................................................................................................................... 8

    1.2.2 IL28B (IFNL3) discovery for HCV treatment responses ...................................................... 9

    1.2.3 IL28B (IFNL3) and spontaneous clearance of HCV ............................................................ 11

    1.2.4 HCV prevalence and geography .......................................................................................... 12

    1.2.5 Implications .......................................................................................................................... 13

  1.3 Functional characterization in epilepsy and neurological diseases ............................................... 14

    1.3.1 Background .......................................................................................................................... 14

    1.3.2 Functional evaluation .......................................................................................................... 15
1.3.3 Approach and limitation ................................................................. 16

1.4 Thesis overview .......................................................................................... 17

1.4.1 Multidisciplinary approach to characterize candidate causal variants for hepatitis C treatment response .......................................................... 17

1.4.2 Multi-electrode array to characterize mutations in neurological diseases ...... 18

2. IFNL3 mRNA structure is remodeled by a functional noncoding polymorphism ...... 20

2.1 Introduction ...................................................................................................... 20

2.2 Materials and Methods .................................................................................. 22

2.2.1 Genotyping .................................................................................................. 22

2.2.2 Statistical analysis ......................................................................................... 22

2.2.3 Cell culture and transfection ........................................................................ 23

2.2.4 Polysome profiling ......................................................................................... 24

2.2.5 RNA SHAPE analysis ................................................................................... 25

2.2.6 SHAPE-MaP library construction ................................................................. 26

2.2.7 SHAPE-MaP data analysis .......................................................................... 27

2.3 Results ............................................................................................................. 29

2.3.1 Association analysis by a large cohort with chronic hepatitis C ................. 29

2.3.2 rs4803217 impacts IFNL3 mRNA translation efficiency ............................. 33

2.3.3 rs4803217 alters structural conformation of the IFNL3 3' UTR ................. 38

2.4 Discussion ....................................................................................................... 46

3. Functional characterization of interferon-λ4 encoded near the IL28B (IFNL3) locus for HCV treatment response ................................................................. 50
4.2.4 Tonic-clonic seizure testing.................................................................79
4.2.5 MEA recording and spike classification ...........................................80
4.2.6 MEA Data analysis ...........................................................................81
R package ....................................................................................................81
Network detection ......................................................................................81
Network burst detection ............................................................................82
Stable period selection ..............................................................................83
Permutation analysis ..................................................................................83
Burst detection ............................................................................................84
Permutation test ..........................................................................................85
Fisher’s Method ..........................................................................................86
4.3 Results .....................................................................................................87
4.3.1 Cortical neurons from Celf4 homozygous knockout (KO) mice showed elevated spike intensity in network bursts.........................................................87
4.3.2 Minimal difference in mean firing rate (MFR) was observed in Celf4 KO neurons .............................................................................................................91
4.3.3 Higher spike frequency and shorter inter-spike interval (ISI) in bursts in Celf4 KO cortical neurons ................................................................................91
4.3.4 Fluoxetine reduces tonic-clonic seizures in Celf4-deficient mice .............95
4.3.5 Fluoxetine rescues spike intensity in network burst phenotype of Celf4 KO neurons ........................................................................................................97
4.3.6 Additional effects of fluoxetine on MEA .............................................100
4.3.7 Carbamazepine did not rescue the spike intensity in network burst phenotype .............................................................................................................102
4.3.8 Carbamazepine in Celf4 deficient mice ................................................................. 104

4.4 Discussion.................................................................................................................. 105

5. Conclusions and future directions ............................................................................. 108

5.1 Molecular genetics and Hepatitis C ....................................................................... 108

5.2 Functional modeling and precision medicine in epilepsy ...................................... 109

5.3 Future directions...................................................................................................... 110

Appendix A: Additional information for IFNL3 mRNA structure ................................ 111

Appendix B: Additional information for IFNL4 functional characterization ............... 119

Appendix C: Additional information for Celf4 functional characterization on multi-electrode array .............................................................................................................. 121

References ...................................................................................................................... 126

Biography ......................................................................................................................... 137
List of Tables

Table 1: Association of rs4803217 with SVR in the IDEAL cohort................................. 31
Table 2: Association of rs4803217 with the pre-treatment viral load............................. 32
Table 3: Association of candidate causal variants with clinical phenotypes by a recessive genetic model............................................................... 59
Table 4: Association of candidate causal variants with clinical phenotypes by an additive genetic model............................................................... 60
Table 5: IFNL4 mRNA isoforms expressed in stimulated PHHs from two donors.......... 70
Table 6: MEA phenotype of Celf4 KO neurons............................................................. 90
Table 7: Combined permutation p value of each treatment comparison...................... 100
Table 8: Raw SHAPE reactivities.................................................................................. 111
Table 9: Raw data for IFNL4 mRNA isoforms expressed in stimulated PHHs from two donors............................................................... 119
Table 10: Intensity in network burst is insensitive to the bin size selection in the network burst detection algorithm............................................................... 121
Table 11: Raw and combined permutation p value of each treatment comparison........ 122
Table 12: Mouse raw data of fluoxetine dosage and seizure responses................. 123
List of Figures

Figure 1: Variant IFNL3 3′ UTRs differentially inhibit reporter gene expression. ........ 35

Figure 2: Polysome profiles of IFNL3 reporter mRNAs............................................. 37

Figure 3: SNP-induced structural changes in the IFNL3 3′ UTR. ................................. 42

Figure 4: Secondary structure models for the IFNL3 3′ UTR. ..................................... 43

Figure 5: Site-directed mutagenesis of the IFNL3 3′ UTR.......................................... 45

Figure 6: IFNL4 is anti-viral against infectious Flaviviridae...................................... 62

Figure 7: IFNL4 inhibits production of infectious HCV and induces STAT1 phosphorylation after over-expression................................................................. 64

Figure 8: IFNL4 is not functionally secreted.............................................................. 66

Figure 9: IFNL4 anti-viral activity requires the IFNL receptor...................................... 68

Figure 10: Expression of IFNL4 mRNA isoforms in A549 cells.................................... 69

Figure 11: Celf4 knockout neurons showed elevated spike intensity in network bursts, but moderate elevation of mean firing rate (MFR) on MEA. ................................. 89

Figure 12: Normalized Histograms of bursting features............................................. 93

Figure 13: Fluoxetine reduced tonic-clonic seizures in Celf4 deficient mice. ................. 96

Figure 14: The spike intensity in network burst phenotype of Celf4 KO neurons was rescued by fluoxetine treatment. ................................................................. 99

Figure 15: Carbamazepine (CBZ) did not rescue the spike intensity in network burst phenotype................................................................. 104

Figure 16: Alignment of primate IFNL2 and IFNL3 sequences................................. 113

Figure 17: Analysis of IFNL3 reporter constructs on HeLa, PH5CH8, and LH86 cells. 114

Figure 18: SHAPE-MaP deconvolution of multiple alleles...................................... 115
Figure 19: A549 cells inducibly express IFNL3 mRNA and are heterozygous at rs4803217.

Figure 20: Analysis of median Shannon entropy as a function of window size.

Figure 21: SHAPE-directed minimum free energy secondary structure for the IFNL3 “G” allele 3’ UTR.

Figure 22: Predicted structure of the C45G mutant stem loop.

Figure 23: The effects of IFNL3 or IFNL4 plasmids over a range of concentrations.

Figure 24: Representative raster plots of 60-second recordings in different developmental stages.

Figure 25: Selected MEA features of experiments performed in this study.

Figure 26: Celf4 heterozygous KO neurons showed the same trend of elevation in spike intensity in network bursts but the difference was insignificant determined by permutation.
Acknowledgements

I would like to thank my advisor, David Goldstein, who provided me
tremendous guidance and support during my graduate school career. I appreciate the
training and research opportunities at Center for Human Genome Variation (Duke
University Medical Center) and Institute for Genomic Medicine (Columbia University
Medical Center) for me to learn genetics and science. I learned tremendously from
David Goldstein for his insights in genetics and his leadership role to run the laboratory
and perform collaborative research.

I would like to thank Shelton Bradrick for being my thesis committee member
and scientific mentor, and his incredible support and advice on conducting research
during my graduate school career. I would like to thank Thomas Urban who gave me
tremendous help and support to conduct genetic research while I was at Center for
Human Genome Variation at Duke University.

I would like to thank and acknowledge the members in my thesis committee,
Douglas Marchuk, David Tobin, Georgia Tomaras, Yong-Hui Jiang and Shelton
Bradrick, for their support during my course of developing and conducting research for
my Ph.D. dissertation.

I would like to give special thanks to all scientists in Center for Human Genome
Variation/Institute for Genomic Medicine, Erin Heinzen, Liz Cirulli, Kimberly Pelak,
Elizabeth Ruzzo, Molly Cook, Anna Need, Jessica Maia, Xiaodi Yao, Mingfu Zhu, Slavé Petrovski, Nicole Walley, Brian Krueger, Kenneth Cronin, Curtis Gumbs, Patrick Shea, Sophie Colombo, Sahar Gelfman, Matthew Halvorsen, Diana Hall, Xiaolin Zhu, Ayal Gussow, Melodi McSweeney, Sarah Kleinstein, Sarah Dugger, Tiffany Tzeng, Gabi Griffin, Francesca La Carpia, Sergey Kisselev, Denise Collado, Rashmi Kanagaratnam, Samantha Shaltz, Quanli Wang, Daniel Fernandez, Joshua Bridgers, Nick Ren, Wayne Frankel, Michael Boland, and Matthew Harms for their help, comments, and supports.
1. Introduction

1.1 Functional evaluation in human genetics

1.1.1 Human disease and genetic discovery

Human genetics has long been studied both to understand how genetic variation influences risk of disease and infer aspects of human evolutionary history. We all differ at the level of our DNA sequence, and geneticists obsess over trying to understand the significance of this genetic diversity. Understanding human genetic diversity is essential to understanding the biology of our diseases of various kinds, from the genetically more simple to more complex, and how we respond to treatment at both the population and individual levels [2]. Indeed, improving our knowledge of human disease biology is the primary driver behind the largest and most systematic studies of human genetic diversity today.

The successful completion of the Human Genome Project in 2003 was the first in a series of large multinational public efforts that began to move the field of medical genetics away from purely descriptive documentation of patients’ physical features coupled with laborious one-by-one examination of a small subset of their genes for potentially pathogenic changes. For example, the International HapMap Project’s collection of millions of genotypes from four global populations was indispensable to

---

1 Portions of this chapter were modified from a published work that was co-authored by David Goldstein, Misha Angrist, and Gianpiero Cavalleri 1. Lu, Y.F., et al., Personalized medicine and human genetic diversity, Cold Spring Harb Perspect Med, 2014. 4(9): p. a008581.
the pursuit of hereditary changes in genes that contribute to disease by providing the platform for so called “genome-wide association studies” (GWAS). GWAS gave us the ability to efficiently and comprehensively assay genetic variants that are common in a population and identify those that appear more commonly in patients with a given disease than they do in controls without the disease. Such variants can sometimes provide clues to the genetic basis of human disease [3].

The basic framework used in these GWAS studies is to select key variants that inform about virtually all common variations in the human genome. These specially selected variants are often called tagging single-nucleotide polymorphisms (SNPs) because they are near perfect surrogates for variants not directly assayed. These variants could be tested easily by newly developed technologies using specially designed genotype chips. GWAS chips are also relatively inexpensive; one can now genotype a million variants for <$50 a sample. Applied to large studies involving thousands of disease (case) and nondisease (control) individuals, the GWAS approach provided the framework to associate specific genetic variants and their cognate genomic regions with diseases, even if the study design was not well suited to identifying the actual genetically causal variants (the reasons will be discussed later on in this chapter). The GWAS approach was successful in that it provided much needed momentum in the push to identify disease genes [3].
It was not until the development of novel DNA-sequencing techniques in the last few years that the study of rare variants became logistically and financially feasible [4]. The relative importance of rare and common variants in the traits that impose the greatest public health burden in the developed world remains unclear. One good example of common variant that influences treatment response and disease susceptibility is \textit{IL28B} (\textit{IL28B} was later renamed \textit{IFNL3}) and hepatitis C treatment response. The detailed background of this genetic discovery is described in Chapter 1.2.

Rare variants can also play a crucial role in human diseases. We know that sometimes cases of relatively common and certainly complex diseases can be caused by rare genetic changes of large effect. Among the best examples of the latter is neuropsychiatric disease, including conditions such as autism, epilepsy, and schizophrenia, in which rare, large genetic rearrangements (so-called “copy-number” variants) collectively account for a small but significant fraction of cases [5]. Another illustrative example of rare variants with large effects is epilepsy. In a recent report on two classical epileptic encephalopathies (infantile spasms and Lennox–Gastaut syndrome), researchers have discovered statistically significant enrichment of \textit{de novo} mutations, that is, new variants that arise in the germline of the patient’s parents, in specific gene sets. Some of these genes have significantly more \textit{de novo} mutations in the patient cohort than would be expected by chance. This finding demonstrates that \textit{de novo} mutations (occurring at one of several different genes) can have a strong influence on the risk of epilepsy [6].
1.1.2 Motivation: moving from genetic association to biology

These genetic tools make genetic discovery feasible through properly designed experiments, such as case-control, family, or trio studies. However, several issues arise when geneticists attempt to move from genetic association to the underlying biological basis of disease-causing mutations. Functional studies play a key role to answer these questions. First, functional evaluation is used, together with population genetics data, to solve the ambiguity of the causal mutations and then further understand the biological basis of the genetic association. Second, in another scenario where the disease-causing mutation is already known, we use functional evaluation to investigate the effect of the mutations to learn biological and therapeutic insights from the mutation.

1.1.3 Unknown causal variants

Although the current genome technologies are relatively good at identifying the mutations that are associated with the disease, the tools available, for example GWAS, might not able to inform the exact variant that is responsible for the phenotype. The SNPs identified in GWAS are unlikely to be responsible for the phenotype themselves because they are often distally located from known functional elements, such as promoter binding sites or open reading frames. Rather, these variants likely co-segregate with other polymorphisms that exert biological functions. Identifying the causal variants that exert biological functions is important to understand the underlying biology and insights for potential therapies.
This issue of unknown causal variants becomes a major challenge following the GWAS studies, because researchers who identify the association often have no idea what the next step is to investigate and interpret the biology. There are numerous possible SNPs that co-segregate (in high linkage disequilibrium) with the discovery SNP and it is difficult to interrogate their functions. Chapter 2 and 3 in this dissertation focus on the topic to investigate the causal variant(s) responsible for the GWAS association for the hepatitis C treatment response.

1.1.4 Investigate biology behind disease-causing mutations

Another challenge in human genetics is that even the causal mutations are already known, the underlying biology remains largely unclear. This fact poses another issue in which no effective therapeutic intervention could be tested because of the limited knowledge of the mutation. It is more similar to a situation that we know where the exact causal mutation is, but can do very little about it.

There are more examples like this in sequencing studies for rare mutations in which researchers are able to determine the causality of a mutation by well-designed statistical approaches. One example is the de novo mutations identified in epileptic encephalopathies (infantile spasms and Lennox–Gastaut syndrome) that several mutations are significantly enriched in the patient cohort than it is expected by chance [6]. However, despite the fact that we know the clear link between certain mutations
with the phenotype, not much is known for the pathological pathways or any candidate target for therapies.

This limitation might be cause by the insufficient platform to characterize such mutation in a cell-based or animal model. For example, mutations in ion channels genes found in neurological disease are relatively feasible to characterize because of the well-established patch-clamping techniques and knowledge of these channels. However, mutations in non-ion channel genes in the same disease often pose new challenges because researchers often have no idea where to start to investigate. Therefore, it would be extremely beneficial have more universal and diverse tools available to answer the questions of biological effects of these mutations.

The focus of functional evaluation in this type of situation is drastically different from the issue discussed in Chapter section 1.1.3 to find the causal variant, but is also an essential part in the human genetics. Chapter 4 describes a mouse seizure model that Celf4 haploinsufficiency is clearly responsible for the seizure phenotype in mice, and we performed further functional characterization and to test potential therapies.

1.1.5 Other challenges in functional evaluation in human genetics

Several other challenges, however, remain for the functional evaluation for human diseases. One example is finding a relevant experimental modeling system, either in vitro or in vivo. It would be extremely difficult to characterize a mutation if there were no prior established knowledge on the function of the protein or the relevant cell
or tissue type that has been established. The functional characterization of some type of
diseases is particular difficult, for example, neurological diseases, because of the reason
that the effort or techniques required to establish a relevant experimental system is
relatively challenging than most of other diseases. Most immortalized cell lines do not
retain the neuronal properties or express relevant genes. Animal models, in most cases
mouse models, generally took a substantial effort and time to create and work on. In
some instances, it would be particularly difficult to work on the animal models if the
mutation induces lethality and requires conditional genetic manipulation.

Overall, functional evaluation is emerging as a central challenge in human
genetics, both to identify the causal variants and to understand the biology of disease-
causing mutation. My graduate work focused on solving these central issues by a variety
of approaches, and spanned a range of major challenges in human genetics. The rest of
Chapter 1 covers the biological backgrounds and genetics of the diseases I worked on
during my graduate school career. Chapter 2 and Chapter 3 focus to characterize the
causal variants following a GWAS study. In Chapter 4, I was trying to characterize a
mutation with known causality to the phenotype and test compounds to revert the
phenotype both in vivo and in vitro.
1.2 Hepatitis C treatment response

1.2.1 HCV biology

HCV is a positive-strand RNA virus belonging to the family Flaviviridae. HCV transmission is mainly through blood-to-blood contact and chronic infection usually results in fibrosis, cirrhosis, liver carcinoma, and even liver failure. It is estimated that 170 million people are chronically infected by HCV worldwide, and it is the major cause for liver transplants in the United States. Because HCV has been a serious public health problem in the United States and worldwide, there have been efforts to develop treatments for chronic HCV infection. However, the treatment success rate has been unsatisfactory. PegIFN-α combined with ribavirin (RBV) therapy has been widely used to treat chronically infected HCV patients since 2002. The treatment success rate is moderate (from 20% to 70%) and is dependent on a patient’s ancestry. Treatment success is defined as reaching sustained virological response (SVR), when the blood viral load is suppressed below the detectable level for 24 wk after 48 wk of combination treatment [7]. In East Asian populations, the PegIFN-α plus RBV treatment for chronically infected HCV patients has been shown to reach 76% of the overall SVR rate, which is dramatically higher than the 56% SVR rate of European-Americans and 24% of African-Americans [8, 9]. Before the genetic discovery of IL28B (later renamed IFNL3), the reason for the differences observed among major ethnic groups was unclear, and race had been used as a profiling feature to predict HCV treatment response.
1.2.2 *IL28B (IFNL3)* discovery for HCV treatment responses

The GWAS performed by Ge and colleagues, as well as studies performed by two other groups, identified a SNP (rs12979860) on the IL28B locus associated with the response of PegIFN-α plus RBV therapy. This genetic variant (rs12979860) is a C-to-T substitution with C being the major allele in Europeans and East Asians. The relative risk for SVR (chance to reach treatment success) is around threefold higher in C/C than non-C/C patients (including C/T and T/T), and is statistically highly significant. Similar results were also found in several other studies; patients with the homozygous C/C genotype at *IL28B (IFNL3)* generally have a two to three times higher treatment success rate than patients with C/T or T/T genotypes [9-11]. European-American patients with C/C genotypes under different treatment regimens show ~80% SVR, compared with 30% and 40% SVR rates of C/T and T/T genotypes, respectively. In African-Americans, patients with the C/C genotype show ~50% of the SVR rate compared with <20% of the SVR rate for C/T and T/T patients. The overall effect of the IL28B polymorphism is, therefore, substantial in predicting HCV treatment response. In general, regardless of ethnicity, the C/C genotype has higher SVR rate than non-C/C genotypes (twofold higher in European-Americans and Hispanics, and threefold in African-Americans). This result suggests that C/C universally favors treatment success versus non-C/C, although in African-Americans, the same C/C genotype shows a lower SVR rate than in European-Americans (50% in African-Americans vs. 80% in European-Americans). The
factors that cause this success rate difference in C/C genotype among individuals of different ethnicities are still unclear. However, despite the clear association and great impact of IL28B on HCV therapy and clearance, little was known about the mechanism and the causal variant(s) underlying the GWAS association.

IL28B has received a great deal of attention since the GWAS discovery for its ability to predict the pretreatment drug response outcome, and the potential for its biological antiviral activities. Before the GWAS, the reason behind the link between ethnicity and drug responses was elusive, but we now clearly know that the IL28B allele frequencies show very different distributions across populations. Using random controls with unknown hepatitis C status, 90% of the East Asian population carried the IL28B C allele versus 70% in European-Americans. However, in the African-American population, the C allele has become the minor allele (smaller allele frequency) at 40%. Strikingly, according to the study performed by Ge and colleagues (2009), the C allele frequency showed linear correlation with the SVR rate in four distinct populations. This concordance strongly suggests that the difference observed in HCV treatment response can be mostly explained by the allele frequency distribution among populations. In a subsequent study by Thomas et al. (2009), 51 geographical subpopulations were examined for the IL28B (IFNL3) polymorphism [12]. The results were similar: the C allele frequency was highest in Asian populations, modest in European populations, and the
lowest in African ancestry populations. This result showed the *IL28B (IFNL3)* allele frequency distribution in higher resolution and corroborated the initial observations.

This correlation substantially explains the reason why different populations have significantly different treatment success rates. Up until now, the only gene for which there is strong evidence of an influence on HCV treatment response has been *IL28B*. An extensive search for other genetic factors that might contribute to HCV treatment response has been performed, but no statistically significant result for other genes that modify the effect of *IL28B (IFNL3)* has been found thus far.

The profile based on race to predict treatment success rate in the past is now proven to be overly simplified. It is actually the *IL28B (IFNL3)* genotype that plays a major role in determining treatment response, not ethnicity, and the differences observed among ethnicity can be explained merely by the allele frequency differences among geographic populations. HCV treatment response is a great example of how allele frequency can affect treatment outcomes among populations, and it seems highly likely that there will be other examples like this to be found in the future.

### 1.2.3 *IL28B (IFNL3)* and spontaneous clearance of HCV

Spontaneous clearance is the clearance of virus by the immune system without the administration of additional drugs. Based on studies of the natural history of HCV, 20%–30% of infected patients can spontaneously clear the virus, whereas the other 70%–80% become chronically infected and require drug therapy. The spontaneous clearance
rate was estimated to be 36% in patients of non-African ancestry and 9% in patients of African ancestry [13]. Soon after the discovery of genetic association with treatment response for HCV, *IL28B* again was shown to be associated with the spontaneous clearance of HCV. Thomas and colleagues examined the *IL28B (IFNL3)* polymorphism in six independent patient cohorts with the diagnosis of HCV infection. Patients were categorized as being chronically infected or having spontaneously cleared HCV by at least two blood tests separated by an interval of at least 6 months. Strikingly, the C allele of *IL28B* (rs12979860) also favors HCV clearance in these cohorts consisting of both European- and African-Americans. Individuals with the *IL28B C/C* genotype were, once again, two to three times more likely to clear the virus than the non-C/C patients. This result was similar to what had been observed in drug-induced HCV clearance. This finding suggests that *IL28B (IFNL3)* has a universal effect on HCV resolution in natural settings without the administration of drugs, an important biological clue.

**1.2.4 HCV prevalence and geography**

Because there is clear evidence of *IL28B (IFNL3)* association with both treatment-induced and spontaneous viral clearance, it would be intriguing to know whether *IL28B (IFNL3)* is also associated with the geographic distribution of HCV prevalence. However, the prevalence of HCV in major continents and *IL28B (IFNL3)* frequency do not seem to be highly correlated. Although in most African countries, the prevalence rates are >3% of the total population (>3% is considered high prevalence for HCV), many
East Asian countries (Asian populations have the highest rate of protective \textit{IL28B C} allele) also comprise the majority of HCV chronic infections worldwide. For example, there is a 3.2\% seroprevalence rate in China, which accounts for a major global HCV-infected population [14]. Many believe that country-specific features of the health care system itself may play a major role in determining the likelihood of HCV exposure. For example, the availability of safe injections dramatically decreases the chance of exposure. Nevertheless, because HCV was first discovered in 1989, it has been impossible to obtain actual data of global HCV prevalence before industrialization.

\textbf{1.2.5 Implications}

It was already well known that African ancestry individuals respond more poorly to HCV drug treatment than Caucasian and Asian individuals. In 2009, GWAS discovered a SNP (also known as rs12979860) in the \textit{IL28B (IFNL3)} locus that is highly associated with patient drug responses to medicines designed to treat HCV [9]. Allele frequencies of \textit{IL28B (IFNL3)} polymorphism were found to differ largely among these ethnic populations, and explain the differences of treatment success rate among those populations. \textit{IL28B (IFNL3)} encodes interferon-\(\lambda\)-3, which is an important cytokine for innate immunity and one of the first responders to the invasion of foreign pathogens. Some believe the allele frequency of \textit{IL28B} has been selected among different populations by one or more pathogen and, thus, evolved at different stages of human history. However, the exact natural selection pressure that causes the distinct pattern of
allele frequency is unclear. Overall, the discovery of \textit{IL28B} (\textit{IFNL3}) polymorphism illustrates that the frequency distribution of certain risk alleles is sufficient to affect the disease progression and drug responses.

Thanks to advances in tools for human genetic research, GWAS methods provide us with insight on common variants and infectious disease. In many carefully controlled clinical trials for HCV treatments, clear and consistent correlation between treatment success and the presence of the \textit{IL28B} (\textit{IFNL3}) polymorphism has been shown. The genetic discovery shatters the long-lasting myth that race plays a role in HCV clearance. In fact, most of the difference in SVR rate can be explained solely by the frequency differences of \textit{IL28B} (\textit{IFNL3}) alleles among populations. HCV infection, therefore, provides a great example of how common variation affects disease susceptibility and drug response.

\textbf{1.3 Functional characterization in epilepsy and neurological diseases}

\textbf{1.3.1 Background}

Epilepsy is one of the most common neurological diseases with a lifetime risk estimated at about three percent of the population [15]. Although genetics is thought to play a major role among the non-acquired epilepsies, the currently known human epilepsy genes account for a small fraction of epilepsy cases [16]. The available antiepileptic therapies still leave approximately 30% of patients with poorly controlled seizures [17, 18]. There is broad agreement that therapies targeted to the underlying
molecular causes of disease—identified through genetic interpretation of patient sequence data—might help address this significant unmet clinical need for seizure control [16]. However, despite the advances in gene discovery, the available experimental tools for screening new antiepileptic therapies remain limited and applicable to only a fraction of the genes identified to date. This has led to an increased interest in developing in vitro screening platforms that can reliably assess candidate therapies in genetic models of epilepsy [16].

1.3.2 Functional evaluation

Seizures are often characterized by hypersynchronous discharges that may occur at a specific region of the cortex and spread into contiguous areas of the brain. The cellular mechanism of seizure initiation is thought to be the network hypersynchronization and high frequency bursts consisting of increased density of action potentials, presumably due to an excitation/inhibition imbalance [19]. Sufficiently synchronized bursts may pass the threshold of surrounding inhibition and activate neighboring neurons leading to broader recruitment, network propagation and ultimately seizures. This phenotypic manifestation is similar among different types of clinical seizures, although the underlying molecular causes can be markedly different. While the in vitro manifestations of seizures are not fully understood, it is thought that both increased synchronicity of network firing and increased bursting are analogues of the in vivo phenotype [20].
1.3.3 Approach and limitation

Although the cellular or biochemical assays can be performed to understand the functional change induced by the mutations in neurological diseases, the interpretation of the result is often limited. First of all, the regular sources of cells – in most cases immortalized cell lines – do not have the neuronal characteristics or lack the expression of neuronal-related genes. This is a major disadvantage because the cellular or molecular results will then become difficult to interpret and connect to the neuronal phenotypes.

Electro-physiology using cells or tissue slides from mutant mouse model is often a feasible way. However, it is a very specialized technique and often requires years of training for individual scientists and substantial effort for a research laboratory to procure equipment and establish the know-how expertise. This fact precludes electro-physiology being adopted as a generalized platform for functional characterization. Successful characterization often requires the substantial involvement of a specialized laboratory with certain areas of expertise.

The third option for functional characterization is to create knock-in or knockout animal models. However, a conditional knockout system needs to be created, if the perturbation of the mutant gene is lethal. In general, although creating animal models to study neurological diseases has great advantages to thoroughly characterize the mutations, it also requires significant time and effort and it is difficult to quickly test for pathogenic mutations by this approach.
Overall, despite there are a number of available options for functional characterization, it will still be extremely beneficial to have more general platforms that can be used for characterization for neurological diseases. Chapter 4 covers a “proof of principle” example of using multi-electrode arrays (MEA) as a platform for functional characterization and drug testing in neurological diseases.

1.4 Thesis overview

The goal of my Ph.D. dissertation is to use functional characterization approaches to 1) identify causal variants responsible for the association in human genetic studies, and 2) investigate the biological effects of disease causing mutations. I aim to integrate these approaches and results to demonstrate a framework that utilizes various approaches to understanding the effect of mutations associated with human diseases.

1.4.1 Multidisciplinary approach to characterize candidate causal variants for hepatitis C treatment response

We demonstrate a multi-disciplinary approach involving genetics, molecular biology, and RNA chemistry that may be generalized to address GWAS associations with non-coding genetic variants that may alter RNA structure and post-transcriptional gene regulation. Nucleotide-resolution SHAPE probing revealed that the 3′ UTR of the rs4803217 G allele IFNL3 mRNA, which is associated with HCV clearance, forms a well-defined structure whereas that of the T allele mRNA is dynamic. The large alteration
induced by a single nucleotide change illustrates the extent to which non-coding genetic variants can have significant functional effects by impacting RNA structure.

Another avenue of research aimed to better understand the association between the recently described IFNL4 gene and control of HCV infection. We analyzed IFNL4 characteristics using molecular *in vitro* assays and primary hepatocyte cultures, and also performed genetic analyses in a well-defined cohort of chronic HCV patients. Some of our findings replicate recently published data, but we believe this is important considering that there is an element of controversy surrounding the relevance of IFNL4 to HCV clearance. Together, our data suggest that IFNL4 is an atypical interferon-λ whose expression may be maladaptive to control of HCV infection but its exact role *in vivo* requires more investigation.

**1.4.2 Multi-electrode array to characterize mutations in neurological diseases**

We show that multi-electrode array (MEA) monitoring of *in vitro* neuronal networks derived from a mouse epilepsy model shows significant differences in network behavior in comparison to the wildtype neuronal networks. Because this modeling paradigm can accommodate proteins that cannot be characterized in well-established modeling paradigms appropriate for ion channels, this result will be of significant interest given the rapid rate of discovery of epilepsy genes.

We also demonstrate that a compound that reduces seizure activity *in vivo* in the mouse model also reverts the *in vitro* MEA network phenotype of the mutant model
back toward the wildtype neuronal networks. This work suggests that \textit{in vitro} neuronal networks monitored by multi-electrode arrays (MEA) may be a general modeling platform appropriate for not only revealing the effects of pathogenic mutations, but useful for identifying compounds that revert those effects both \textit{in vivo} and \textit{in vitro}. 
2. IFNL3 mRNA structure is remodeled by a functional noncoding polymorphism

2.1 Introduction

Human genetic variation has a large influence on an individual’s susceptibility to infectious diseases. Studies on the hepatitis C virus (HCV), a major cause of liver disease [22], have revealed a striking example of how human genetics affects the outcome of infection. Genome-wide association studies performed on diverse patient populations have identified polymorphisms near the interferon-λ 3 (IFNL3; formerly IL28B) gene that predict the efficacy of interferon-based therapy for chronic infection. These polymorphisms are also strongly predictive of spontaneous HCV clearance during the acute phase of infection [9-12]. IFNL3 is a secreted cytokine that binds a specific cell surface receptor complex expressed on epithelial cells, leading to JAK-STAT signal transduction and expression of interferon-stimulated genes [23-25]. IFNLs have been implicated in control of viral infections of epithelial-derived tissues, such as gut, lung, and liver [26-29]. IFNLs inhibit replication of multiple viruses in vitro [30-32] and IFNL3 genotype has been linked to liver interferon-stimulated gene mRNA expression in chronic HCV patients and primary human hepatocytes [33-35]. The mechanisms by which IFNL3 genetic variants function to control HCV infection are not clear. Steady-

---

1 This chapter is part of a published work that was co-authored by Shelton Bradrick, David M. Mauger, Thomas Urban, David Goldstein, and Kevin Weeks 21. Lu, Y.F., et al., IFNL3 mRNA structure is remodeled by a functional non-coding polymorphism associated with hepatitis C virus clearance. Sci Rep, 2015. 5: p. 16037.
state levels of IFNL3 mRNA in liver tissue and peripheral blood mononuclear cells may be linked to HCV clearance because the protective alleles (rs12979860 C and rs8099917 T) are correlated with a modest (< 2-fold) increase in IFNL3 mRNA levels [10, 11, 36, 37]. However, other studies on different patient populations and cell types failed to detect correlations between differences in IFNL3 mRNA expression and genotype in liver tissue [33, 34] or plasmacytoid dendritic cells [38]. Thus, the propensity for HCV clearance appears to be related to differences beyond IFNL3 mRNA levels.

Although primary hepatocyte cultures have been shown to express IFNL3 upon HCV infection [26, 39], evidence is emerging to support a role for dendritic cell subtypes in the production of IFNL3 [40-42]. In addition, levels of IFNL3 protein were found to be strongly correlated with IFNL3 genotype in serum samples from HCV-infected patients [43, 44]. One study also suggested that serum IFNL3 levels are a stronger predictor of viral clearance than IFNL3 genotype [45]. These observations, taken together, suggest that IFNL3 abundance may be regulated by differential efficiency of mRNA translation as a function of IFNL3 genotype.

Here we report the genetic association and functional consequences of a SNP (rs4803217) located in the 3′ untranslated region (UTR) of IFNL3 mRNA. This SNP was previously reported to regulate IFNL3 expression by changing seed base pairing with HCV-induced microRNAs [46]. We genotyped a large number of patients with chronic hepatitis C from the IDEAL Cohort [47] to evaluate associations of this variant with
treatment phenotypes. Molecular analyses suggested that rs4803217 is functional in regulation of mRNA translation efficiency. RNA structure probing by SHAPE provided a physical explanation for these functional differences and revealed that IFNL3 mRNA structure is markedly altered by this common SNP. Together, these findings support a model in which IFNL3 mRNA translation efficiency, governed by allele-specific mRNA structures, that may be important for clearance of HCV infection.

2.2 Materials and Methods

2.2.1 Genotyping

DNA from the IDEAL cohort was used for genotyping. The genotyping of rs4803217 was performed by quantitative PCR (Bioline) using, forward primer 5´-ACCTG AGATT TTATT TATAA ATTAG CCACT TG(G/T)-3´, and reverse primer 5´-CTTTT CCTCA TGTGTT TATTT CAACA AGGAT TTC-3´. Data were clustered by Ct value for genotype calling. Rs368234815 was genotyped by the TaqMan genotyping assay (Life Technologies) according to the protocol recommended by the manufacturer, forward primer 5´-TGGGT CCTGT GCACG GTGAT-3´, reverse primer 5´-TCCCT CAGCG CCTTG GCA-3´, and probe 5´-CGCAG (AA/C)GGCC CCCCG G-3´.

2.2.2 Statistical analysis

Statistical analysis of phenotypic data was performed in STATA (StataCorp) and R (www.r-project.org). Sustained virological response (SVR) was used as the binary dependent variable, and genetic polymorphisms were considered independent variables.
in the logistic regression model (the assumption of logistic regression requires the dependent variable to be binary). Both discovery SNP and rs4803217 were used as independent variables in multiple logistic regression to distinguish its independent association after accounting for the effect of each other. Viral loads (IU/ml) were transformed by Box-Cox method and considered as the dependent variable in linear regression to meet the multivariate normality assumption. Sample size calculation was performed by the pwr package in R and the significance level was set at $p = 0.01$ under the multiple regression setting. All the $p$ values reported in this paper are based on the two-sided statistical tests.

2.2.3 Cell culture and transfection

PH5CH8 [48], LH86 [49], A549 and stable HeLa-FRT cells expressing TetR were cultured in high glucose DMEM supplemented with 10% heat-inactivated FBS and non-essential amino acids. Stable HeLa-FRT cells (a gift from Elena Dobrikova, Duke University) harboring RLuc reporter trans-genes were established as described38 by co-transfection of a flp recombinase plasmid (pOG44, Life Technologies) with RLUC reporter plasmids cloned into pcDNA5/FRT/TO (Life Technologies, see below). Stable cell lines were selected with blasticidin (2.5 $\mu$g/ml) and hygromycin (200 $\mu$g/ml). Analysis of RLuc levels (Renilla Luciferase Assay System, Promega) in HeLa cell lines was conducted in the absence or presence of tetracycline (1 $\mu$g/ml) induction. Stable reporter cell lines were independently derived and analyzed three times.
Transient transfections of PH5CH8, LH86, or HeLa cells were performed using the pcDNA5/FRT/TO reporter plasmids described above. Cells were plated onto 24-well plates at a density of 7.5 × 105 cells/well 24 hours prior to co-transfection with 50ng (LH86, HeLa) or 100ng (PH5CH8) of RLuc reporter plasmids and 20 ng of pGL3 firefly luciferase plasmid (Promega) using 0.5 μL Lipofectamine 2000 (Life Technologies). Cells were lysed 24 hours after transfection and luciferase measurements were made using by dual luciferase assay (Promega). All reporter plasmids were analyzed in triplicate and P values were calculated by unpaired t-test. A549 transfections for RT-PCR analysis of IFNL3 expression were performed with 7.5 × 105 cells plated 24 hours prior to transfection into 24-well plates. Cells were transfected with 800 ng pcDNA5/FRT/TO plasmid, poly I:C (Sigma), or purified HCV JFH1 strain [50] RNA produced by in vitro transcription with T7 RNA polymerase (Ambion). Four hours after transfection, total RNA samples were generated using Trizol reagent (Life Technologies) and used for standard RT-PCR detection of IFNL3 mRNA using primers designed to amplify the open reading frame.

2.2.4 Polysome profiling

HeLa cell lines were cultured on 10 cm dishes and processed essentially as described [51]. Briefly, cells were washed, scraped and pelleted in ice-cold PBS containing 200 μ M cycloheximide. Cell pellets were resuspended in 0.4 ml lysis buffer [400 mM KOAc, 25 mM K-HEPES (pH 7.5), 15 mM Mg(OAc)2, 1 mM DTT, 200 μ M
cycloheximide] and incubated on ice for 30 minutes before centrifugation for 10 minutes at 12,000 × g to remove insoluble material. Supernatants were layered onto 10 ml 15–50% sucrose density gradients containing 400 mM KOAc, 25 mM K-HEPES (pH 7.5) and 15 mM Mg(OAc)2 and then centrifuged for 3 hours at 35,000 rpm using an SW41 rotor. Gradients were fractionated (Isco) and 12 fractions were collected for RNA extraction using Trizol LS (Life Technologies). Polysome analysis was conducted with single replicate per cell line in three biological replicates conducted on different days.

2.2.5 RNA SHAPE analysis

Adherent A549 cells were plated in a 6-well culture dish and grown overnight to 90% confluency. IFNL3 mRNA was induced by transfecting cells (using Lipofectamine 2000) with 4 µg of poly I:C (Sigma) and incubated for 4 hours before washing with PBS (pH 7.2). For SHAPE modification of total deproteinized mRNA, RNA was purified using a modified Trizol (Invitrogen) extraction protocol. The cells were lysed with 1.0 mL of Trizol for 5 min at room temperature. The cell lysate was placed in a microfuge tube and 0.2 mL of chloroform was added followed 15 seconds of shaking vigorously by hand. The sample was incubated for 2 min at room temperature, and the sample was spun in the centrifuge at 12,000 × g for 15 min. The deproteinized total RNA in the aqueous layer was removed and immediately equilibrated into native folding buffer [50 mM HEPES (pH 8.0), 5 mM MgCl2, 200 mM potassium acetate] using a pre-equilibrated G-25 spin column (GE Health sciences). Deproteinized RNAs were incubated at 37 °C for
15 minutes in native folding buffer in order to allow folding state to come to equilibrium and then treated with 0.1 volume of DMSO, or 100 mM 1-methyl-7-nitroisatoic anhydride (1M7) in DMSO at 37 °C for 5 min. EDTA was added to 10 mM final concentration; reactions were chilled on ice; and the RNA was precipitated with isopropanol. For the parallel denaturing modification control, deproteinized total RNA was equilibrated into 1X denaturing buffer [50 mM HEPES (pH 8), 4 mM EDTA, 50% formamide] using a pre-equilibrated G-25 spin column; heated to 95 oC for 1 min; modified with 0.1 volume of 100 mM 1M7 in DMSO; chilled on ice for 2 min; and recovered by precipitation with isopropanol.

### 2.2.6 SHAPE-MaP library construction

DNA libraries for massively parallel sequencing were prepared as described [52]. Briefly, targeted SHAPE-MaP reverse transcription reactions (20 µL) contained 2 µM of IFNL3 RT primer (5′-GTCTT TTCCT CATTG TTTAT TTC-3′) in 1x SHAPE-MaP buffer [50 mM Tris-HCl (pH 8), 75 mM KCl, 6 mM MnCl2] with 1 µ L reverse transcriptase (Superscript II, Invitrogen). Reactions were incubated at 42 °C for 3 hours, and cDNA products were purified (G-25 microspin column; GE Lifesciences). Purified cDNAs were amplified for 20 cycles using Q5 DNA Polymerase (NEB) with Illumina adapted IFNL3 targeted primers (Forward 5′-GACTG GAGTT CAGAC GTGTG CTCTT CCGAT CNNNN NCCTCCACCA TTGGC TGC-3′, Reverse: 5′-CCCTA CACGA CGCTC TTCCG ATCTN NNNNN NNCTC ATTGT TTATT TCAAC AAGGA
TTTC-3′). The forward primer was designed with a 3′ mismatch to the IFNL2 mRNA to limit amplification of this similar transcript. PCR primers were designed to amplify sequences spanning different exons, so analyzed PCR product could only be generated from a spliced mRNA and not contaminating genomic DNA. PCR products were purified using a PureLink PCR purification kit (Invitrogen), and libraries were amplified by a second PCR reaction using Q5 DNA Polymerase (NEB) and Illumina TRUseq PCR primers. PCR libraries were quantified by fluorescence (Qubit fluorometer; Life Technologies), analyzed with a Bioanalyzer DNA kit (Agilent), and sequenced on a MiSeq instrument (Illumina).

### 2.2.7 SHAPE-MaP data analysis

SHAPE-MaP reactivities for each nucleotide within the IFNL3 3′ UTR were generated from the raw FASTQ files. Raw SHAPE reactivity data are shown in (Table 8 Appendix A). A custom analysis pipeline identified all of the sequencing reads from IFNL3 rs4803217 G, IFNL3 rs4803217 T, and IFNL2 mRNAs. IFNL2 reads were determined by analyzing six nucleotide positions that vary between IFNL2 and IFNL3. Reads with three or more nucleotide matches to IFNL2 (1.2% of total reads) were removed prior to RNA structure analysis. Reads from the two IFNL3 alleles were sorted into separate FASTQ files and then analyzed using ShapeMapper [52]. Shannon entropies and minimum-free energy genome secondary structure models were generated from SHAPE reactivities using a custom RNA folding pipeline [52]. Shannon
entropy values were calculated to quantify the well-determinedness of structural states for a given RNA region. The pipeline interfaces with RNAstructure v5.5 in order to computationally model RNA secondary structure. The RNA secondary structure models were diagramed using VARNA [53].
2.3 Results

2.3.1 Association analysis by a large cohort with chronic hepatitis C

The most strongly associated single nucleotide polymorphisms (SNPs) found in genome-wide association studies (GWAS) were rs12979860 in European-Americans and African-Americans, and rs8099917 in East Asians. These SNPs likely co-segregate with polymorphisms that directly affect biological functions related to IFNL expression or activity. To identify candidate functional variants in the IFNL3 gene region, we searched for variants in high linkage disequilibrium with rs12979860 (hereafter referred to as the ‘discovery SNP’) by analysis of available human genome sequence data (1000genomes.org). We identified a SNP (rs4803217) in the 3′ UTR of IFNL3 that is highly correlated (best tagging SNP according to 1000 Genomes Project) to the discovery SNP [54-57] (Figure 1A). The rs4803217 G allele is correlated with the protective C allele of the discovery SNP. This variant was an intriguing candidate as the source of viral infection-related phenotypes because 3′ UTRs often contain cis-acting RNA elements that control mRNA translation and decay. Rs4803217 is also flanked by AU-rich elements (AREs; Figure 1B), sequence motifs that have been linked to post-transcriptional gene regulation35. Therefore, we conducted a large-scale genetic association study examining rs4803217 and HCV clearance in a cohort of chronically-infected patients.
We genotyped and analyzed these variants in the IDEAL Cohort [47], a large sample of chronic HCV patients treated with pegylated IFN-α and ribavirin combination therapy (n = 792 European-American; n = 169 African-American). The sample sizes of both populations are sufficient to achieve greater than 0.99 in statistical power (1.00 in European-American and 0.992 in African-American under the significance level of 0.01). Linkage disequilibrium analysis in HCV patients revealed that rs4803217 is strongly associated with rs12979860 in European-American patients ($r^2 > 0.97$; Table 1), with somewhat lower linkage disequilibrium in African-Americans ($r^2 = 0.913$). We performed association tests of rs4803217 with clinical phenotypes using logistic regression. The association of rs4803217 with sustained virological response (defined as absence of detectable HCV RNA in serum at least 24 weeks after discontinuation of treatment) was extremely significant ($p = 2.48 \times 10^{-25}$) in the European-American population (Table 1). The discovery SNP has been previously associated with viral burden such that patients with the protective allele (C) exhibit a slightly higher baseline viral load [9]. We found that rs4803217 exhibited comparable association with pre-treatment baseline viral load by linear regression in both European- and African-Americans compared with the discovery SNP (Table 2).
Table 1: Association of rs4803217 with SVR in the IDEAL cohort.

<table>
<thead>
<tr>
<th>Discovery SNP in GWAS</th>
<th>r²*</th>
<th>OR**</th>
<th>P</th>
<th></th>
<th>r²*</th>
<th>OR**</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12979860-C</td>
<td>-</td>
<td>3.92</td>
<td>4.18×10⁻²⁵</td>
<td>-</td>
<td>-</td>
<td>2.10</td>
<td>0.006</td>
</tr>
<tr>
<td>Candidate functional variant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4803217-G</td>
<td>0.977</td>
<td>3.93</td>
<td>2.48×10⁻²⁵</td>
<td>0.219</td>
<td>0.913</td>
<td>1.77</td>
<td>0.030</td>
</tr>
<tr>
<td>rs368234815-TT</td>
<td>0.977</td>
<td>3.96</td>
<td>1.81×10⁻²⁵</td>
<td>0.118</td>
<td>0.708</td>
<td>2.13</td>
<td>0.004</td>
</tr>
</tbody>
</table>

| European-American N = 792     |       |      |         |           |         |      |         |
|                                | r²    | OR** | P       |           | r²      | OR** | P       |
| Comparison between functional SNPs |       |      |         |           |         |      |         |
| rs4803217-G (correcting for rs368234815-TT) | 0.965 | 1.69 | 0.393   |           | 0.701   | 0.78 | 0.629   |
| rs368234815-TT (correcting for rs4803217-G) | 2.37  | 0.158|         |           | 2.64    | 0.065|         |

Logistic regression or multiple logistic regression were performed by additive genetic model

* r² represents pairwise LD with GWAS discovery SNP (rs12979860)
** OR = odds ratio

Next, we performed a multiple logistic regression analysis to examine whether rs4803217 showed independent association with patient phenotypes in excess of that attributed to the discovery SNP. This analysis revealed that rs4803217 did not show significant association with SVR when we adjusted for rs12979860 genotype, in either European- or African-Americans (Table 1). Thus, although the IDEAL cohort is one of the largest and most data-rich clinical trial cohorts available for genetic studies of patient response to IFN-based therapy, the extremely high linkage disequilibrium between variants in this region precluded independent statistical association of rs4803217 with SVR in this sample size.

Recently, a functional dinucleotide SNP rs368234815 that can influence the expression of IFNL4 protein was discovered upstream of the IFNL3 locus and was implicated in the clearance of HCV [58]. (Detailed rs368234815 biology will be discussed in Chapter 3). In addition, the putatively favorable rs4803217 G allele was proposed to have an unexpected negative effect on the decrease of HCV RNA level after treatment in
the presence of the unfavorable rs368234815 ΔG allele [59], suggesting that rs4803217 may modify the effect of the rs368234815 variant associated with IFNL4. Here, we also performed the same multiple logistic regression or multiple linear regression approaches to assess whether rs4803217 or rs368234815 showed stronger association with patient phenotypes. We found that, despite the large sample size of the IDEAL cohort, this analysis did not reveal significant independent association of either SNP after correcting for the effect of each other (Table 1 and Table 2), although we note an interesting opposite trend of odds ratios (OR) for SVR after mutual adjustment (0.78 and 2.64 for rs4803217 G and rs368234815 TT, respectively) in African American patients (Table 1). Overall, rs4803217 demonstrated similar levels of association with patient phenotypes as did the discovery SNP or a nearby functional SNP rs368234815, suggesting that rs4803217 is an interesting candidate variant to further functionally investigate.

Table 2: Association of rs4803217 with the pre-treatment viral load.

<table>
<thead>
<tr>
<th>Discovery SNP in GWAS</th>
<th>coefficient</th>
<th>P (after correcting for rs12979860)</th>
<th>coefficient</th>
<th>P (after correcting for rs12979860)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12979860-C</td>
<td>12.89</td>
<td>7.81×10⁻⁷</td>
<td>20.68</td>
<td>5.41×10⁻⁶</td>
</tr>
<tr>
<td>Candidate functional variant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4803217-G</td>
<td>12.61</td>
<td>1.35×10⁻⁶</td>
<td>21.25</td>
<td>2.25×10⁻⁶</td>
</tr>
<tr>
<td>rs368234815-TT</td>
<td>13.03</td>
<td>5.77×10⁻⁷</td>
<td>22.20</td>
<td>7.57×10⁻⁷</td>
</tr>
</tbody>
</table>

Comparison between functional SNPs

<table>
<thead>
<tr>
<th>European-American N = 792</th>
<th>coefficient</th>
<th>P</th>
<th>coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4803217-G (correcting for rs368234815-TT)</td>
<td>-5.79</td>
<td>0.677</td>
<td>9.04</td>
<td>0.253</td>
</tr>
<tr>
<td>rs368234815-TT (correcting for rs4803217-G)</td>
<td>18.71</td>
<td>0.178</td>
<td>14.61</td>
<td>0.066</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>African-American N = 169</th>
<th>coefficient</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4803217-G</td>
<td>9.04</td>
<td>0.253</td>
</tr>
<tr>
<td>rs368234815-TT</td>
<td>14.61</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Association tests were performed by linear regression or multiple linear regression (additive genetic model)


### 2.3.2 rs4803217 impacts IFNL3 mRNA translation efficiency

Although rs4803217 did not independently predict SVR in patients, based on association evidence it is a good candidate to investigate. Furthermore, its presence in the IFNL3 mRNA 3′ UTR prompted us to test functionality. The IFNL3 3′ UTR is 176 nucleotides (nt) in length, is composed of 65% A and U nucleotides, and harbors several functional AREs [46], including two (ARE1 and ARE2) flanking rs4803217 (Fig. 1B). Comparison of UTRs from primate IFNL3 and IFNL2 genes indicates that the region around rs4803217 is highly conserved and that the T variant is the ancestral allele (Appendix A: Figure 16).

We first evaluated the effects of the variant IFNL3 3′ UTRs on steady-state mRNA and protein levels using a reporter assay. We took advantage of a stable HeLa cell line that harbors a genomic flippase recognition target site [60] to create isogenic cell lines with distinct Renilla luciferase (RLuc) reporter constructs controlled by a CMV promoter that is transcriptionally regulated by tetracycline repressor protein. This system allows characterization of gene expression in cells containing a single copy of the reporter gene. Three cell lines were established containing either RLuc alone or RLuc fused to each of the two IFNL3 3′ UTR variants (Figure 1C). The presence of either IFNL3 3′ UTR caused modest but statistically significant ($p = 0.028$ for G allele, $p = 0.046$ for T allele, unpaired t-test) reductions in the levels of IFNL3 chimeric mRNAs compared to RLuc alone (Figure 1D). Interestingly, levels of the RLuc protein were
reduced by a much larger amount, 4- and 7-fold for the rs4803217 G and T reporters, respectively, compared to the reporter lacking an IFNL3 3' UTR (Figure 1E). The differential effects on mRNA versus protein levels strongly suggest that the IFNL3 3' UTR regulates gene expression by repressing the efficiency of mRNA translation rather than mRNA abundance in HeLa cells. Importantly, the 3' UTR SNP affected the magnitude of this repression: the rs4803217 T allele produced approximately 50% less RLuc than did the corresponding G allele mRNA (Figure 1E; p = 0.0005, unpaired t-test). Induction of reporter mRNA expression levels by tetracycline treatment reduced RLuc differences between the cell lines, suggesting that reporter mRNA over-expression led to escape from endogenous regulatory factor(s) present in HeLa cells (Appendix A: Figure 17A,B). We also tested the IFNL3 reporter constructs in an immortalized human hepatocyte cell line (PH5CH8), a human hepatoma cell line (LH86), and HeLa cells by transient plasmid transfection (Appendix A: Figure 17C). In each cell line, the patterns of reporter expression mirrored that in the stable HeLa cell lines (Figure 1E).
Figure 1: Variant IFNL3 3′ UTRs differentially inhibit reporter gene expression.

(A) IFNL genes and genetic variants. (B) The IFNL3 mRNA 3′ UTR sequence. The three AREs (1–3) are indicated in red and the rs4803217 SNP is bracketed. (C) IFNL3 reporter constructs. Transcription of each reporter mRNA is directed by the constitutive CMV promotor. (D) Relative levels of each RLuc reporter mRNA. RNA levels were normalized to GAPDH; asterisk indicates that differences are significant at the p < 0.05 level. (E) RLuc protein levels. Protein levels were measured by luciferase assay and normalized to total protein. Triple asterisks indicate that differences are significant at the p < 0.001 level. All data are shown as mean values ± s.d.
We next directly tested the hypothesis that rs4803217 regulated efficiency of mRNA translation. We used polysome profiling to examine whether the variant IFNL3 reporter mRNAs were differentially associated with translating ribosomes in stable HeLa cell lines. The extent of association with multiple ribosomes is an indication of mRNA translation efficiency. IFNL3 variant reporter cell lines, grown to equal densities, were harvested, lysed and subjected to sucrose density gradient centrifugation to separate non-translating and ribosome-associated mRNA. Gradients were subsequently fractionated and RLuc reporter and GAPDH mRNA levels were measured by RT-qPCR. Ribosomal RNA sedimentation patterns from the IFNL3 two reporter cell lines were nearly identical (Figure 2A), indicating that bulk mRNA translation rates did not differ between cell lines. In contrast, the rs4803217 G and T IFNL3 reporter mRNAs showed distinct profiles across the gradient. Across three biological replicates, the G allele mRNA was more highly represented in denser gradient fractions, reflecting incorporation into polysomes, than the T allele mRNA (Fig. 2B,C). We quantified the areas underneath the lines for both G and T alleles and found that the allelic difference in area between polysome fractions (fractions 7–12) and non-polysome (fractions 1–6) fractions were significantly different (p = 0.01) (Figure 2C). This result indicates that the G allele mRNA was more abundant in polysome fractions compared to the T allele mRNA. Correspondingly, levels of non-translating mRNA (sedimenting between the 40S ribosomal subunit and the 80S monosome) were higher
Figure 2: Polysome profiles of IFNL3 reporter mRNAs.
(A) Top: Representative ribosome profiles. Locations of the 40S subunit, 80S monosome, and polysome peaks are indicated. Profiles were obtained using a sucrose density gradient, monitored by absorbance trace (254 nm). Bottom: Analysis of extracted total RNA samples in each gradient fraction for each IFNL3 reporter cell line. Ribosomal RNA species are indicated at right. (B) Relative levels of RLuc reporter mRNA normalized to GAPDH mRNA for each IFNL3 cell line as a function of ribosome gradient fraction. The relative levels (%) of each mRNA in each gradient fraction are indicated. (C) Areas underneath the lines in Figure 2B were quantified for both polysome fractions (fractions 7–12) and non-polysome fractions (fractions 1–6). The plot shows the differences in area between the G and T alleles (G minus T). Unpaired t-test (two-tailed) was used to calculate the P value from the three independent experiments (p = 0.01). Error bar represents the SEM.

for the rs4803217 T mRNA. We did note that a significant amount (15%) of the T-allele reporter mRNA was found in dense gradient fractions (10–12), suggesting that a portion of the T allele reporter mRNA escapes repression and is efficiently translated. Combined with the discrepant luciferase protein and reporter mRNA measurements shown in Figure 1, these data strongly suggest that the rs4803217 SNP regulates efficiency of IFNL3 reporter mRNA translation in HeLa cells.

2.3.3 rs4803217 alters structural conformation of the IFNL3 3′ UTR

There is extensive literature linking 3′ UTRs with translational control [see reference [61] for a recent review]. A growing body of evidence indicates that 3′ UTR structure regulates mRNA translation and decay by modulating interactions with trans-acting factors [62]. Moreover, genetic variants that alter mRNA structure and function, termed RiboSNitches, have been described [63]. We directly examined whether the rs4803217 SNP physically modulates the structure of endogenously expressed IFNL3
mRNA using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) [64]. We examined the structure of IFNL3 mRNA under non-denaturing conditions in the context of total purified cellular RNA. SHAPE analysis was performed on deproteinized RNA as this approach allows for direct assessment of mRNA structure independently of trans-acting proteins or RNAs. The structure of the IFNL3 mRNA was probed using the fast-reacting SHAPE reagent 1-methyl-7-nitroisatoic anhydride (1M7) and products were analyzed by massively parallel sequencing using the SHAPE-MaP strategy [52] (Appendix A: Figure 18). We used human A549 lung adenocarcinoma cells for these experiments because they express IFNL3 mRNA upon transfection with poly(I:C) (Appendix A: Figure 19A). Importantly, A549 cells are heterozygous at rs4803217 and levels of mRNA from each allele were directly correlated with the gene dosage (Appendix A: Figure 19B). Using the MaP approach, sequencing reads derived from IFNL3 were sorted by SNP identity at rs4803217; this made it possible to examine the structures of both alleles under identical conditions in the same experiment (Appendix A: Figure 18). SHAPE data were collected on a 266 nt long region of the IFNL3 mRNA containing the last 125 nt of the coding sequence (including the stop codon) and the 3' UTR, excluding the terminal 23 nt and the poly(A) tail.

SHAPE-MaP reactivities report a model-free measurement of the degree of RNA structure [52, 64]. The rs4803217 G allele had lower SHAPE reactivity, and thus more stable structure, near the region close to rs4803217 than did the T allele (Figure 3A). We
next calculated Shannon entropy values for each rs4803217 allele. Shannon entropies are derived from a SHAPE-directed partition function and report a measure of whether an RNA region is likely to form a single well-defined structure [52, 65, 66]. The median Shannon entropies across the IFNL3 \(^{-}\) UTR were much lower for the G allele, especially near the rs4803217 site, than for the T variant (Figure 3B and Appendix A: Figure 20).

We next performed a correlation analysis of SHAPE reactivities between the G and T alleles by calculating R2 across the mRNA regions to address the overall structural similarity between the two alleles (Figure 3C). The SHAPE reactivities in the open reading frame (ORF) are highly correlated between alleles, and moderately correlated at the very end of 3 \(^{-}\) UTR. In contrast, a very low or zero correlation was observed in the region of rs4803217. This indicates that a drastic local change in structure is induced by the SNP. In sum, the G allele 3 \(^{-}\) UTR adopts an overall well-defined, highly structured conformation; whereas, nucleotide-resolution SHAPE probing indicates that the T allele adopts multiple conformations. In addition, structural differences between the two 3 \(^{-}\) UTR alleles are centered on the SNP region.

SHAPE data can be used to develop accurate models for large, complexly structured RNAs with well-defined structures [67, 68]. SHAPE-directed RNA structure models for each variant over the 3 \(^{-}\) portion of the open reading frame and 3 \(^{-}\) UTR are summarized in arc plots that display predicted short- and long-range base pairing
Given its low Shannon entropy, the structure of the G allele could be modeled with a high degree of confidence, as evidenced by the preponderance of highly
Figure 3: SNP-induced structural changes in the IFNL3 3′ UTR.

(A) SHAPE reactivity profiles. Low median SHAPE reactivities correspond to highly structured regions in the RNA. Reactivities are shown as the median over 5-nt windows. “Nucleotide position” values indicate nt locations in the context of the total SHAPE data (B) Shannon entropies (medians over 5-nt windows). Peaks indicate regions with high Shannon entropies and that likely adopt multiple conformations. (C) Correlation of median SHAPE reactivities between the G and T alleles. R2 was calculated over 100-nt windows. (D) SHAPE-directed RNA secondary structure models for the 3′ UTR. Base pairs are shown as arcs. Arcs are colored by pairing probability.

probable helices (Figure 3D, bottom). In contrast, the 3′ UTR for the T allele is predicted to adopt multiple conformations, comprised of helices with base pairs of lower individual probability (Figure 3D, top). The combined SHAPE reactivity and entropy differences indicate that the T-allele mRNA has a more variable structure than the G-allele mRNA.

The SHAPE-directed RNA secondary structure model indicated that the rs4803217 G nucleotide forms a canonical base pair within a stable stem-loop (Figure 4A and Appendix A: Figure 21) that contains portions of ARE1 and ARE2. In contrast, the U variant is not predicted to stably adopt this stem-loop structure (Figure 4A). We calculated the free energy change of 3′ UTR RNA folding [69] for each allele and found that the rs4803217 G to U change alters relative free energy by approximately 7 kcal/mol. This difference in relative free energy is close to the largest possible increment achievable by altering a single base pair [70]. We calculated the expected free energy
changes for all possible nucleotide changes in the context of the IFNL3 3′ UTR (529 sequence variants) and in the context of the intact mRNA (2851 sequence variants). The

Figure 4: Secondary structure models for the IFNL3 3′ UTR.

(A) SHAPE-directed RNA secondary structure models of the IFNL3 3′ UTR G (top) and T (bottom) alleles. rs4803217 site is shown for each structure. Nucleotides are colored by SHAPE reactivity. Highly probable (>80%) helices for each allele are shown. Locations of AU-rich elements are indicated on each allele structure. (B) Calculated free energy change increments (∆∆G) for all possible nucleotides changes in the IFNL3 3′ UTR (left) or complete intact mRNA (right). Batch calculation was performed by mfold 3.0.
increase in folding free energy due to the change from G to U at rs4803217 ranked the highest among all possible substitutions in 3′ UTR, and the third among all possible substitutions in the intact mRNA, indicating that this position has a strong influence on RNA folding (Figure 4B). Direct SHAPE experimental interrogation and free energy increment calculations thus suggest that the rs4803217 SNP induces close to the largest possible change in RNA structure.

We performed mutagenesis to further test whether functional differences for rs4803217 reflect changes in RNA structure. Three mutant versions of the IFNL3 reporter construct were established that contained mutations at the rs4803217 position (nt 52 of the 3′ UTR) and/or at the nt position predicted to base-pair with rs4803217 (nt 45 of the 3′ UTR) (Figure 5A). Changing rs4803217 from G to C resulted in reduced expression, similar to the rs4803217 T reporter construct (Figure 5B). Restoring the predicted base pair in this mutant by introducing a G at nt 45 increased expression, consistent with RNA structure opposing repression. Interestingly, introduction of G at nt 45 (C45G) in the context rs4803217 G, which is predicted to interrupt base-pairing, had increased expression which was opposite of the effect on expression initially expected. However, this mutant resulted in a free energy of RNA folding similar to the rs4803217 G 3′ UTR (Figure 5B) by inducing a register shift in base pairing and ultimately preserved the stem loop associated with escape from repression [69] (Appendix A: Figure 22). For each of
the constructs analyzed, the expression levels were inversely correlated with folding free energy. Together, the mutagenesis data further reinforce the model that the T-allele has a more variable structure than the G-allele mRNA and that stable structure is associated with enhanced HCV clearance.

**Figure 5:** Site-directed mutagenesis of the IFNL3 3′ UTR.

(A) The rs4803217 G and T allele reporter constructs were compared to three mutants in stable HeLa cell lines. (B) Normalized relative light units (RLU) for each reporter
construct is shown above. Data are shown as mean values ± s.d. and folding free energies for the full 3′ UTR sequence are shown.

2.4 Discussion

Multiple genome-wide association studies have identified polymorphisms near the IFNL3 gene that predict efficacy of therapy and spontaneous HCV clearance [9-12]. Although rs4803217 is not independently associated with patient phenotypes in the cohort we analyzed, this SNP occurs in the 3′ UTR of IFNL3 and showed clear functional effects on reporter gene expression, suggesting a role for this variant in control of HCV infection.

Rs4803217 influences regulation of mRNA translation efficiency by the IFNL3 3′ UTR in HeLa cells, perhaps through altering functionality of AREs. Although cytokine 3′ UTR AREs are commonly associated with enhanced mRNA decay, they have also been implicated as translational regulators [71-74] and mRNA decay and translation are closely associated cytoplasmic processes [75, 76]. Due to gene regulation at the level of mRNA translation, the modest differences in IFNL3 mRNA levels, as a function of genotype, reported in some studies may underestimate actual differences in IFNL3 protein levels, as we observed here for the IFNL3 rs4803217 SNP (Figure 1). This is consistent with a recent report of ~7-fold differences in IFNL2/3 serum levels between HCV patients who are homozygous at the discovery SNP26. Consistent with this work, a recent independent study identified a distinction between CC versus non-CC genotype
(discovery SNP) for IFNL3 protein plasma levels [44]. However, we acknowledge that our results do not formally establish a causal relationship between rs4803217 and IFNL3 levels in patients.

To examine and define the physical basis for the mechanism of IFNL3 translational regulation, we used the SHAPE-MaP strategy, which made it possible to resolve structural conformations and differences for the rs4803217 G and U variant mRNAs in a single heterozygous cell line, in a single experiment. We discovered that the nucleotide at rs4803217 influenced the global structure of IFNL3 mRNA. The G allele was much more highly structured and the free energy difference between the G- and U-containing mRNAs was predicted to be close to the upper limit in structural stabilities achievable by a single-nucleotide sequence change.

Our work strongly suggests that rs4803217 is a RiboSNitch41 that alters gene expression by inducing changes in RNA structure. Recent genome-wide analyses suggest that RiboSNitches are widespread in the human transcriptome [77]. The experimental and computational strategies developed here can be applied to other candidate functional variants to advance our understanding of how post-transcriptional gene regulation is affected by non-coding genetic polymorphisms. These approaches should be broadly useful for identifying human genome sequence variants in non-coding regions that exert functional effects by altering RNA structure.
The observed change in protein expression is likely the result of differential binding of \textit{trans}-acting factors, induced by the differential folding of the 3' UTR variants. The SHAPE-derived secondary structures show that rs4803217 T reduces structure of a region overlapping ARE1 and 2 (Figure 4A), suggesting that unidentified \textit{trans}-acting ARE-binding proteins may access these elements more efficiently in the T allele mRNA. Recently, McFarland et al. reported that the IFNL3 mRNA is regulated by AU-rich elements and also targeted by muscle-specific micro(mi)RNAs (miR-208b and miR-499a-5p) whose expression is triggered by HCV replication. These authors found that rs4803217 was functional using IFNL3 reporter mRNAs. However, the SNP did not confer regulation when ARE1 or ARE2 were mutated, in agreement with our observation that rs4803217 alters structural conformation of regions encompassing ARE1 and ARE2. With respect to the mechanism involving action of miRNAs, the rs4803217 T to G allele would disrupt miRNA-mRNA seed base-pairing. Thus, the dramatic effects of rs4803217 on IFNL3 mRNA structure that we report here may regulate interactions with ARE-binding proteins and miRNAs. However, we note that, when the SNP position was mutated to a C, we observed a level of reporter expression similar to that of the T-allele (Figure 5B). The rs4803217 C mutation abolishes seed-region base-pairing to the proposed regulatory miRNAs, similar to the G-allele. This suggests that the regulation we characterized in HeLa cells, which may lack expression of the implicated miRNAs, is mediated primarily by effects of RNA structure on
interaction with \textit{trans}-acting factor(s) other than miR-208b or miR-499a-5P. Further study is needed to define the complement of relevant \textit{trans}-acting factors that may differentially regulate variant IFNL3 mRNA transcripts through repression of translation and/or induction of mRNA decay.

In summary, in this chapter, we demonstrate a general strategy involving genetics, molecular biology, and RNA chemistry for understanding GWAS associations with non-coding genetic variants that may alter RNA structure and post-transcriptional gene regulation. Nucleotide-resolution SHAPE probing revealed that the 3′ UTR of the rs4803217 G allele IFNL3 mRNA, which is associated with HCV clearance, forms a well-defined structure whereas that of the T allele mRNA is dynamic. The large alteration induced by a single nucleotide change illustrates the extent to which non-coding genetic variants can have significant functional effects by impacting RNA structure.
3. Functional characterization of interferon-λ4 encoded near the *IL28B (IFNL3)* locus for HCV treatment response

**3.1 Introduction**

Human genetic variation impacts susceptibility to infectious diseases. This is particularly true for hepatitis C virus (HCV), a member of the Flaviviridae and a major world-wide cause of liver disease [22]. Multiple genome-wide association studies (GWAS) have identified polymorphisms near the interferon-λ3 (*IFNL3*; formerly *IL28B*) gene that predict efficacy of pegylated IFN-α/ribavirin combination therapy for chronic infection and spontaneous HCV clearance during the acute phase of infection [9-12]. The most strongly associated single nucleotide polymorphisms (SNPs) found in GWAS (known as discovery SNPs) were rs12979860 in European-Americans and African-Americans, and rs8099917 in East Asians. The former is located upstream of *IFNL3* and confers >2-fold likelihood of sustained virological response (SVR) between favorable and non-favorable alleles. These SNPs are unlikely to be responsible for improved clearance themselves because they are distally located from known functional elements. Rather, these variants likely co-segregate with other polymorphisms that exert biological functions related to *IFNL* expression or activity.

---

1 This chapter is part of a published work. My collaborators in this study include Shelton Bradrick, David Goldstein, and Thomas Urban. 32. Lu, Y.F., et al., *Interferon-lambda4 is a cell-autonomous type III interferon associated with pre-treatment hepatitis C virus burden*. Virology, 2015. 476: p. 334-40.
*IFNL3* is a member of the type III IFNs, whose genes are clustered together on chromosome 19 [23, 24]. The IFNLs bind a distinct receptor complex that is expressed on epithelial cells [25]. Receptor binding initiates JAK-STAT signal transduction, leading to expression of interferon-stimulated genes that confer an anti-viral state. Notably, IFNLs inhibit HCV replication *in vitro* [30, 31] and *IFNL3* genotype has been linked to liver interferon-stimulated gene mRNA expression [33, 34].

Recently, RNA sequencing of primary human hepatocytes (PHHs) treated with polyinosinic:cytidylic acid (poly I:C, a mimic of dsRNA) revealed transcriptional activity upstream of the *IFNL3* gene [58]. Many of the identified transcript isoforms contain putative open reading frames (ORFs). One transcript encoding a 179 amino acid (p179) protein exhibited antiviral activity against a non-infectious HCV replicon when overexpressed and was named *IFNL4*. Of note, the rs12979860 discovery SNP lies within the first intron of *IFNL4* (Figure 6A). A dinucleotide variant in this region (rs368234815 TT/ΔG) causes a frame shift on these ORFs and may alter translation of *IFNL4* mRNA [58]. A direct role for *IFNL4* in HCV clearance is presently unclear, and whether the *IFNL4* protein is endogenously produced in human hepatocytes infected with HCV is uncertain due to lack of Western blot or proteomic data.

Here, we examined the expression of *IFNL4* transcripts in PHHs and human cell lines, characterized *IFNL4* secretion and requirement for the IFNL receptor, and performed genetic analysis of rs368234815 in a large well-characterized chronic HCV
cohort. While the exact role of *IFNL4* in control of HCV infection is still elusive, the molecular and genetic evidence suggest that rs368234815 is likely relevant to viral clearance in patients infected with HCV.

3.2 Materials and Methods

3.2.1 Genotyping and statistical analysis

DNA from the IDEAL cohort [47] was used for genotyping. Taqman genotyping design for rs368234815 is: forward primer 5´-TGGGTCTGTGCACCGGTGAT-3´, reverse primer 5´-TCCCTCAGCGCCTTGGCA-3´, and probe 5´-CGCAG(AA/C)GGCCCCCCCGG-3´. Taqman genotyping was performed according to the protocol suggested by the manufacturer (Life Technologies). Genotypes for rs12979860 were obtained from the Illumina Human610-Quad genotyping array as described previously [9]. Logistic regression and linear regression were performed in STATA (StataCorp) and R (www.r-project.org). Clinical phenotypes [SVR, rapid virological response (RVR), and pre-treatment viral load] were used as a dependent variable, and genetic polymorphisms (rs12979860 and rs368234815) were used as independent variables as indicated in the tables for both logistic and linear regression models. All P values and 95% confidence intervals (CI) reported were calculated from two-tailed distributions. In multiple logistic or linear regression analyses, both rs12979860 and rs368234815 were considered as independent variables to assess independent associations after accounting for the effect of each other.
3.2.2 PHH experiments and RT-PCR

Cryopreserved (lot #: Hu8122) or freshly isolated PHHs (lot #: Hu8171) were obtained from Life Technologies. Cells in 24-well plates were maintained in William's E medium supplemented with the Maintenance Supplement Pack, Serum-free (Life Technologies) 37 °C overnight for recovery. The next day, cells were transfected with 5 µg of poly I:C (Sigma) and 6.4 µl Lipofectamine 2000 per well as described previously (Park et al., 2012). Total RNA was extracted by RNeasy mini column (Qiagen) and treated with on-column DNase to eliminate genomic DNA carry-over. Reverse transcription with random primers (Life Technologies) was subsequently performed to obtain cDNA. Forward primer 5′-GTCCT GTGCA CGGTG ATC-3′ and reverse primer 5′-GATAA CTGGC AATAA ATTTA AACCG GG-3′ were used to generate pooled PCR products from cDNA transcribed from the IFNL4 genomic region. pCR®-Blunt II-TOPO® was used for blunt end insertion of PCR products. Bacterial colonies were randomly picked for sequencing to determine isoform and genotype.

For studies of endogenous IFNL4 mRNAs in A549 cells, confluent monolayers in 24-well dishes were transfected with 1.6 µg poly I:C and 4 µl of lipofectamine 2000 per well. Four hours post-transfection, total RNA was isolated with Trizol reagent and then subjected to reverse transcription using a gene specific reverse primer (5′-TCTTT GATAA CTGGC AATAA-3′). PCR was subsequently performed with this and a forward primer (5′-GTCCT GTGCA CGGTG ATC-3′).
3.2.3 Cell culture, transfection, and virus infections

Huh7.5 (Blight et al., 2002) and A549 cells were cultured in high glucose DMEM supplemented with 10% heat-inactivated FBS and non-essential amino acids. A2EN cells stably transduced with ISRE-luciferase [78] were kindly provided by Raphael Valdivia (Duke University) and grown in keratinocyte serum-free medium (Invitrogen) supplemented with 30 µg/mL recombinant EGF, 0.1 ng/mL bovine pituitary extract and 0.4 mM CaCl2. For evaluation of IFNL antiviral potency and STAT1 phosphorylation, Huh7.5 cells were transfected with the indicated CMV promoter-driven expression plasmids in 24-well plates using 0.8 µg DNA and 2 µl Lipofectamine 2000 per well unless otherwise indicated. The next day, cells were re-fed media and infected with either HCV JFH1 [79], DENV-2 New Guinea C, or YFV 17D. HCV infections were performed at a multiplicity of infection (MOI) of 0.1 whereas flavivirus infections were carried out using an MOI of 1. At 24 (flaviviruses) or 72 (HCV) hours after infection, cells were fixed in cold methanol and stained with either 4G2 anti-envelope (flaviviruses) or C7-50 anti-core (HCV) antibodies as described (Kato et al., 2006). Rates of infection were determined using a Cellomics ArrayScan high content imager. Infectious HCV in cell supernatants were determined by TCID50 assay [80]. All reported infection experiments were performed at least three separate times and each experiment had three replicates per condition. Representative experiments are shown in Figure 6, Figure 7, Figure 8 and Figure 9.
For evaluation of IFNL activity in cell supernatants, Huh7.5 were transfected as described above and re-fed with supplemented DMEM 4 h after transfection. 48 h later, cell supernatants were harvested, clarified by centrifugation, and used to treat naïve cells. Subsequently, cell lysates were isolated 6 h after treatment for STAT1 Western blotting, or infected with YFV 24 h after treatment and rates of virus infection determined 24 h post-infection as described above.

For knockdown of IFNLR1, siRNA duplexes were obtained from Qiagen and used to transfecHuh7.5 cells with RNAiMAX (Invitrogen) at a final concentration of 30 nM. Cells were re-fed with media 24 h later and transfected with IFNL expression plasmids as described above. Cells were then infected with YFV 24 h after plasmids transfection and rates of infection determined as described above.

3.2.4 Western blotting

Cells were lysed using RIPA buffer and subjected to SDS-PAGE and Western blotting. Antibodies to total STAT1 and phospho-STAT1(Tyr 701) were obtained from Cell Signaling Technology. Antibody to polypyrimidine tract-binding protein was kindly provided by M.A. Garcia-Blanco (Duke University). For analysis of cell culture supernatants by Western blot, proteins were concentrated by precipitation with trichloroacetic acid and resuspended in sample buffer prior to fractionation by SDS-PAGE.
3.2.5 Cloning, and RT-qPCR

IFNL4 and p131 cDNA expression constructs encoding Halo tags were kindly provided by L. Prokunina-Olsson (NCI). For generation of IFNL3 expression constructs, the IFNL3 ORF (AY184374.1) encoding a C-terminal Flag tag was synthesized (Integrated DNA Technologies) and inserted into pcDNA5/FRT/TO (Invitrogen) using KpnI and NotI restriction sites. A comparable IFNL4 expression construct was similarly established [58]. Knockdown of IFNLR1 mRNA expression was determined by RT-qPCR. Briefly, Huh7.5 cells transfected in triplicate with siRNA duplexes (see above) were harvested 48 h post-transfection for RNA isolation (RNeasy, Qiagen) and cDNA synthesis using the High Capacity cDNA synthesis kit (Applied Biosystems). Subsequently, qPCR was performed using Power SYBR green master mix (Applied Biosystems) to measure IFNLR1 transcript levels using forward primer 5′-GGATCTGAAGTATGAGGTGC-3′ and reverse primer 5′-GTAGATGGTCTGACTGA-3′. Levels of β-2-microglobulin mRNA were analyzed for calibration using forward primer 5′-TGTCTGGGTTTCATCATCCGA-3′ and reverse primer 5′-TCACACGGCGATGACATCT-3′.

3.3 Results

3.3.1 Genetic analysis between rs368234815 and rs12979860 (the discovery SNP)

To further investigate the association of IFNL4 with clinical phenotypes, we genotyped and analyzed the rs368234815 and rs12979860 variants in the IDEAL Cohort
a large well-characterized sample of patients chronically-infected with HCV (n=794 European-American; n=171 African-American). Linkage disequilibrium (LD) analysis revealed that rs368234815 is strongly associated with rs12979860 in European-American patients (r^2=0.977), with somewhat lower LD in African-American patients (r^2=0.713).

We performed association tests of both rs12979860 and rs368234815 with patient phenotypes (SVR, RVR, and pre-treatment viral load) using a logistic regression or linear regression model to evaluate the correlation of each polymorphism with clinical features. The rs12979860 CC genotype was previously associated with a 2 fold higher chance of achieving SVR compared to that of combined non-CC genotypes (CT and TT), and the CT genotype only showed slightly higher SVR than the TT homozygous patients [9], suggesting a mixture between additive and recessive genetic models. Therefore, we performed regression analyses with both recessive and additive models separately (Table 3 and Table 4). In the recessive model (rs12979860 CC versus non-CC, and rs368234815 TT/TT versus non-TT/TT), the associations of rs12979860 and rs368234815 with SVR were highly significant in both populations (Table 3 and Table 4). However, because genetic variants surrounding IFNL3 and IFNL4 are in high linkage disequilibrium (inherited together as a linkage disequilibrium block), most polymorphisms in this region exhibit strong association with these clinical phenotypes.

In order to question whether rs368234815 shows independent association when controlled for the discovery SNP (rs12979860), which would support a causal role for
rs368234815, we performed multiple logistic regression analysis. After correcting for the effect of rs12979860, the P value for rs368234815 remained significant in both populations for SVR (P=0.039 in European-Americans and 0.042 in African-Americans), suggesting stronger association of rs368234815 with SVR than the discovery SNP (Table 3). However, the P values of both populations were driven by a small number of patients that have disparate genotype between rs12979860 and rs368234815 and are also of borderline significance (cutoff=0.05). In the additive model, rs368234815 no longer showed statistical significance after correcting for rs12979860 (Table 4). Therefore, the genetic evidence for whether rs368234815 is directly causal with respect to SVR is still insufficient and merits further investigation.

We previously observed that viral load is significantly correlated with rs12979860 genotype such that patients with the C allele have higher pre-treatment viral load (~1.5-fold; P=1.21 ×10^{-10}) [9]. Since rs368234815 has lower LD (r^2=0.713) with the discovery SNP in African-Americans compared with Caucasians (r^2=0.977), there is a higher frequency of disparate genotypes and therefore greater resolution to identify independent effects of rs368234815 and rs12979860. We found that in African-American patients, both rs12979860 and rs368234815 are associated with pre-treatment viral load (P=1.00 ×10^{-3} for rs12979860 and 2.99×10^{-5} for rs368234815). After correcting for the discovery SNP (rs12979860) in the multiple linear regression analysis, rs368234815 showed independent association with pre-treatment viral load (P=0.012; Table 2). The
independent association of rs368234815 with pre-treatment viral load was also evident in multiple linear regression analysis using an additive genetic model (P=0.042 after correcting for rs12979860; Table 4). Taken together, these observations suggest that IFNL4 expression may influence pre-treatment viral load in HCV-infected patients.

**Table 3: Association of candidate causal variants with clinical phenotypes by a recessive genetic model.**

<table>
<thead>
<tr>
<th></th>
<th>European-American (N = 794)</th>
<th>African-American (N = 171)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logistic regression</td>
<td>Logistic regression</td>
</tr>
<tr>
<td></td>
<td>SVR RVR</td>
<td>SVR RVR</td>
</tr>
<tr>
<td></td>
<td>OR (95% CI) P</td>
<td>OR (95% CI) P</td>
</tr>
<tr>
<td>rs12979860 (CC vs non-CC)</td>
<td>7.30 (5.04-10.91) 5.11 x 10^{-6}</td>
<td>7.42 (5.24-10.50) 1.31 x 10^{-6}</td>
</tr>
<tr>
<td>rs368234815 (TT/TT vs non-TT/TT)</td>
<td>11.44 (9.13-115.46) 0.0049</td>
<td>10.21 (8.97-106.92) 0.0053</td>
</tr>
<tr>
<td>rs12979860(correcting for rs368234815)</td>
<td>0.05 (0.08-0.64) 0.71</td>
<td>0.77 (0.075-7.89) 0.825</td>
</tr>
<tr>
<td>rs368234815(correcting for rs12979860)</td>
<td>11.44 (9.13-115.46) 0.0049</td>
<td>10.21 (8.97-106.92) 0.0053</td>
</tr>
<tr>
<td></td>
<td>Multiple logistic regression</td>
<td>Multiple logistic regression</td>
</tr>
<tr>
<td>rs12979860 (CC vs non-CC)</td>
<td>5.70 (2.36-13.74) 1.08 x 10^{-4}</td>
<td>6.39 (1.17-32.73) 0.032</td>
</tr>
<tr>
<td>rs368234815 (TT/TT vs non-TT/TT)</td>
<td>8.04 (3.39-20.81) 1.79 x 10^{-4}</td>
<td>3.47 (0.60-20.28) 0.107</td>
</tr>
<tr>
<td>rs368234815 (correcting for rs12979860)</td>
<td>6.23 (1.66-36.48) 0.042</td>
<td>0.45 (0.03-5.37) 0.548</td>
</tr>
</tbody>
</table>

LD is pairwise linkage disequilibrium between rs12979860 and rs368234815.

* Odds ratio (95% confidence interval).

** In African-Americans, N = 157 in RVR.
Table 4: Association of candidate causal variants with clinical phenotypes by an additive genetic model.

**European-American (N = 794)**

<table>
<thead>
<tr>
<th>LD (r²) = 0.977</th>
<th>Logistic regression</th>
<th>Linear regression</th>
<th>SVR</th>
<th>RVR</th>
<th>Viral load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)¹</td>
<td>P</td>
<td>OR (95% CI)¹</td>
<td>P</td>
<td>coefficient (95% CI)</td>
</tr>
<tr>
<td>rs12979860</td>
<td>3.92 (3.03 - 5.07)</td>
<td>4.18×10⁻⁵</td>
<td>5.10 (3.35 - 7.76)</td>
<td>3.46×10⁻⁴</td>
<td>12.89 (7.81 - 17.98)</td>
</tr>
<tr>
<td>rs368234815</td>
<td>3.96 (3.06 - 5.13)</td>
<td>1.81×10⁻⁵</td>
<td>4.92 (3.25 - 7.45)</td>
<td>5.48×10⁻⁴</td>
<td>13.03 (7.96 - 18.10)</td>
</tr>
</tbody>
</table>

Multiple logistic regression  | Multiple linear regression |
| rs12979860 (correcting for rs368234815) | 1.19 (0.26 - 5.42) | 0.826 | 5.41 (0.65 - 44.83) | 0.118 | -0.89 (-34.22 - 32.43) | 0.958 |
| rs368234815 (correcting for rs12979860) | 3.35 (0.73 - 15.29) | 0.118 | 0.94 (0.12 - 7.57) | 0.955 | 13.91 (-19.32 - 47.13) | 0.412 |

**African-American (N = 171)**

<table>
<thead>
<tr>
<th>LD (r²) = 0.713</th>
<th>Logistic regression</th>
<th>Linear regression</th>
<th>SVR</th>
<th>RVR</th>
<th>Viral load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)¹</td>
<td>P</td>
<td>OR (95% CI)¹</td>
<td>P</td>
<td>coefficient (95% CI)</td>
</tr>
<tr>
<td>rs12979860</td>
<td>2.19 (1.29 - 3.72)</td>
<td>0.004</td>
<td>3.29 (0.92 - 11.75)</td>
<td>0.066</td>
<td>20.23 (11.65 - 28.82)</td>
</tr>
<tr>
<td>rs368234815</td>
<td>2.21 (1.32 - 3.71)</td>
<td>0.003</td>
<td>3.74 (1.07 - 13.01)</td>
<td>0.038</td>
<td>21.63 (13.20 - 30.06)</td>
</tr>
</tbody>
</table>

Multiple logistic regression  | Multiple linear regression |
| rs12979860 (correcting for rs368234815) | 1.44 (0.55 - 3.78) | 0.463 | 1.19 (0.12 - 11.50) | 0.881 | 6.34 (-9.53 - 22.22) | 0.431 |
| rs368234815 (correcting for rs12979860) | 1.64 (0.64 - 4.21) | 0.304 | 3.24 (0.34 - 30.73) | 0.306 | 16.32 (0.57 - 32.06) | 0.042 |

*Odds ratio (95% confidence interval)
**In African-Americans, N=157 in RVR
LD is pairwise linkage disequilibrium between rs12979860 and rs368234815
3.3.2 IFNL4 exhibits broad antiviral activity

The dinucleotide variant rs368234815 (TT/ΔG) near the discovery SNP (rs12979860) has been proposed to influence HCV clearance through the production of IFNL4 in subjects carrying the ΔG allele (Figure 6A). To expand on these findings, we investigated the effects of IFNL4 over-expression on replication of several infectious Flaviviridae species in Huh7.5 human hepatoma cells [81]: HCV (JFH-1 strain), dengue virus serotype 2 (DENV2; New Guinea C strain), and the yellow fever virus vaccine (YFV; 17D strain). Transfection of vector plasmid, a construct encoding an inactive
Figure 6: IFNL4 is anti-viral against infectious Flaviviridae.

(A) Relative positions of IFNL3 and IFNL4 mRNAs are indicated along with locations of genetic variants relevant to this study. Proteins produced by IFNL4 transcripts are indicated for the rs368234815 ΔG genotype. (B) Huh7.5 cells transiently transfected with the indicated expression plasmids were infected with HCV, DENV2, or YFV-17D and rates of infection were measured by quantitative immunofluorescence using antibodies against viral antigens. All data are shown as mean±S.D. (C) Representative immunofluorescence images that illustrate levels of virus infection after transfection with the indicated plasmids are shown.
isoform of IFNL4 (p131; Figure 6A), and an IFNL3 expression plasmid served as controls. Replication of each virus was similarly inhibited by expression of IFNL3 and IFNL4 compared to the vector alone and to inactive p131 as measured by immunofluorescence staining for viral antigen followed by quantitative imaging (Figure 6B and C). The effects of IFNL3 or IFNL4 plasmids were similar over a range of concentrations (Appendix B: Figure 23), suggesting that the two IFNLs have similar potencies when expressed transiently. Measurements of infectious HCV in supernatants of transfected cells revealed an approximate 10-fold reduction in infectious particles due to either IFNL3 or IFNL4 over expression (Figure 7A). In addition, STAT1 phosphorylation levels were similar in lysates from cells transfected with either IFNL expression construct (Figure 7B). These observations confirm the antiviral potency of IFNL4 when exogenously expressed by plasmid transfection.
Figure 7: IFNL4 inhibits production of infectious HCV and induces STAT1 phosphorylation after over-expression.

(A) Huh7.5 cells were transfected with the indicated plasmids as in Fig. 1 and infected with HCV. Supernatants were collected at 72 h post-infection and titered by TCID$_{50}$ assay. (B) Western blot analysis of STAT1, STAT1-PO$_4$ and polypyrimidine tract-binding protein (PTB) is shown from cells transfected with the indicated plasmids for 18 h. Quantification of STAT1-PO$_4$ signals normalized to PTB is shown below the STAT1-PO$_4$ blot.

3.3.3 IFNL4 anti-viral activity is cell autonomous

IFNs are typically secreted molecules and trigger paracrine and autocrine signal transduction pathways leading to expression of interferon-stimulated genes (ISGs) that produce an anti-viral phenotype [82, 83]. Recombinant IFNL4 was previously described
to lack biological activity for unknown reasons [58]. To test whether IFNL4 is functionally secreted from cells, Huh7.5 cells were transfected with expression plasmids as described in Figure 6 and supernatants were harvested 48 h post-transfection. Subsequently, naïve Huh7.5 cells were treated with supernatants and then infected with YFV to assay for anti-viral effects. The IFNL3-containing supernatant clearly antagonized YFV infection, as expected. In contrast, IFNL4 supernatant lacked detectable antiviral activity (Figure 8A). We also utilized a stable cell line harboring firefly luciferase under control of an interferon-sensitive response element (ISRE) to analyze supernatants from transfected cells. Whereas supernatant from cells transfected with the IFNL3 plasmid triggered signal transduction to the ISRE promoter at the lowest dose utilized, the supernatants from cells that express IFNL4 did not produce a response (Figure 8B). Similarly, IFNL3-containing supernatant strongly induced STAT1 phosphorylation in treated cells compared to supernatant from IFNL4-transfected cells (Figure 8C).

We further analyzed IFNL4 secretion by Western blot analysis. Huh7.5 cells were transfected with plasmids encoding either flag-tagged IFNL4, halo-tagged IFNL4, or halo-tagged p131. After 48 h, equivalent fractions of cell lysates and supernatants were subjected to SDS-PAGE and Western blotting using IFNL4 antibody. Each protein was abundantly expressed in cell lysates and migrated as multiple bands of different apparent molecular weight, possibly reflecting post-translational glycosylation [84].
contrast, each protein was barely detectable in supernatant samples, confirming that IFNL4 and p131 are inefficiently secreted. Together, these data suggest that IFNL4 is an atypical anti-viral factor that does not require secretion for activity in vitro.

Figure 8: IFNL4 is not functionally secreted.

(A) Huh7.5 cells were transfected with the indicated plasmids and supernatants harvested 48 h p.t. Naïve cells were infected with YFV-17D 24 h after treatment with the respective supernatants and levels of virus infection determined 24 h later by quantitative immunofluorescence. (B) Supernatants generated after plasmid transfection
were used to treat A2EN cells that stably harbor a luciferase transgene under control of an ISRE element. (C) Conditioned media from cells expressing the indicated proteins were used to treat naïve cells for 4 h and the same factors were analyzed by Western blot as in Figure 7. Quantification of STAT1-PO4 signals normalized to PTB is shown below the STAT1-PO4 blot.

3.3.4 The antiviral activity of IFNL4 requires IFNLR1 expression

To be classified as a type III interferon, IFNL4 should functionally require IFNLR1, the class II cytokine receptor that, along with IL10Rβ, transduces signals for all α-interferons [23]. To determine whether IFNL4 activity requires IFNLR1, we inhibited expression of IFNLR1 in Huh7.5 cells using three independent siRNAs and then expressed p131, IFNL3, or IFNL4 proteins in these cells. Knockdown of IFNLR1 mRNA levels was verified by RT-qPCR due to lack of reliable antibodies for the IFNLR1 protein (Appendix B: Figure 23B). Cells were then infected with YFV to assess anti-viral phenotypes. Transfection of cells with each of the three siRNAs negated the antiviral effects of both IFNL3 and IFNL4 (Figure 9). These results indicate that IFNLR1 expression is required for IFNL4 biological activity.
Non-silencing (NS) and three individual IFNLR1 siRNAs were introduced into Huh7.5 cells followed by transfection with the indicated expression plasmids. Subsequently, cells were infected with YFV-17D to evaluate cellular antiviral phenotype. All data show mean values±S.D. Asterisks (**) indicate P values of≤0.005 as calculated by unpaired Student’s t-test.

3.3.5 Endogenous IFNL4 mRNA isoforms are expressed in lung epithelial carcinoma cells

No known immortalized human cell lines are known to express IFNL4. We stimulated multiple cell lines (A549, Huh7, HepG2 and Huh7.5) by transfection with poly I:C and performed RT-PCR with pan-amplifying IFNL4 primers. Only A549 carcinoma cells expressed IFNL4 transcripts, each of which contained a novel intron located upstream of the start codon. Specifically, three transcripts were identified, the most abundant of which corresponded to mRNA isoforms predicted to express p107 and p131 proteins in the context of the rs368234815 ΔG genotype (Figure 10). These mRNA isoforms were also the most abundant transcripts identified in PHHs (see below). The third transcript contained novel splice sites and does not correspond to previously reported IFNL4 mRNAs. These observations suggest that A549 cells may be a tractable model for studies of IFNL4 alternative pre-mRNA splicing and transcriptional regulation.
Figure 10: Expression of IFNL4 mRNA isoforms in A549 cells.

Cells were either mock-treated or transfected with poly I:C and total RNA isolated 4 h post-transfection for RT-PCR detection of IFNL4. Identified transcripts and corresponding predicted proteins based on rs368234815 genotype are depicted on the right.

3.3.6 IFNL4 is weakly expressed in primary human hepatocytes

IFNL4 is predicted to be synthesized only by the rs368234815 ΔG allele through alternative splicing (Figure 6A). We evaluated the relative frequency of IFNL4 mRNA transcripts in primary human hepatocytes (PHHs) from two rs368234815 heterozygous donors stimulated by transfection of poly I:C. We sequenced individual clones derived from RT-PCR amplification using primers designed to hybridize to all known IFNL4 mRNA isoforms. Sanger sequence analysis of 105 individual clones revealed similar ratios of mRNA isoforms between the two donors (Appendix B: Figure 9).

Approximately 72% of transcripts were derived from the beneficial rs368234815 TT allele (Table 5). This was surprising, as most of these transcripts are predicted targets of nonsense mediated decay (NMD) due to the presence of an in-frame stop codon located upstream of exon junctions. Out of 46 and 59 clones analyzed for each PHH donor,
respectively, only 2 were identified in a single donor that encoded functional IFNL4.

This suggests that, in the context of rs368234815 heterozygosity, a large majority (~98%) of spliced mRNA transcripts arising from the IFNL4 locus do not produce functional IFNL4 protein.

**Table 5: IFNL4 mRNA isoforms expressed in stimulated PHHs from two donors.**

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Genotype</th>
<th>Amino acids</th>
<th>Count</th>
<th>Subject to NMD</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p131</td>
<td>TT</td>
<td>75</td>
<td>19</td>
<td>Yes</td>
<td>18.1</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>131</td>
<td>12</td>
<td>No</td>
<td>11.4</td>
</tr>
<tr>
<td>p179</td>
<td>TT</td>
<td>123</td>
<td>24</td>
<td>Yes</td>
<td>22.9</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>179</td>
<td>2</td>
<td>No</td>
<td>1.9</td>
</tr>
<tr>
<td>p170</td>
<td>TT</td>
<td>124</td>
<td>10</td>
<td>No</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>170</td>
<td>1</td>
<td>No</td>
<td>1.0</td>
</tr>
<tr>
<td>p107</td>
<td>TT</td>
<td>51</td>
<td>11</td>
<td>Yes</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>107</td>
<td>9</td>
<td>No</td>
<td>8.6</td>
</tr>
<tr>
<td>Others</td>
<td>TT</td>
<td>–</td>
<td>12</td>
<td>–</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>–</td>
<td>5</td>
<td>–</td>
<td>4.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>TT</td>
<td><strong>76</strong></td>
<td></td>
<td></td>
<td><strong>72</strong></td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td><strong>29</strong></td>
<td></td>
<td></td>
<td><strong>28</strong></td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td></td>
<td>105</td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>

**3.4 Discussion**

Multiple genome-wide association studies have identified polymorphisms near the *IFNL3* gene that predict HCV clearance, but the functional genetic variant(s) contributing to this effect phenotype have been elusive [9-12]. Here, we analyzed the genetic associations and functional effects of rs368234815. Our *in vitro* assays clearly indicated that IFNL4 exhibits antiviral potency against multiple members of the
Flaviviridae when over-expressed by plasmid transfection in cultured human hepatoma cells. IFNL4 potency also requires the cognate IFNLR1 that is utilized by other IFNL cytokines, and is therefore accurately classified as a member of the IFNL family. An important feature of IFNL4 appears to be its relatively inefficient secretion, suggesting that autocrine signaling through binding to IFNLR1 may occur in the context of endosomes. In agreement with our observations, Hamming et al. [84] recently reported that over-expressed IFNL4 is not efficiently secreted and requires IFNLR1 for anti-viral activity. Together, these findings establish that IFNL4 is a newly identified type III IFN that is functional if expressed to sufficient levels.

We observed that in rs368234815 heterozygous PHHs, IFNL4 mRNAs transcribed from the rs368234815 ΔG allele are not abundantly expressed, compared to TT allele mRNAs which are predicted to be degraded by NMD. This observation was unexpected, and suggests that endogenous IFNL4 mRNA isoforms may be poorly translated, because NMD requires cytoplasmic translation [85], and/or that genetic variation surrounding IFNL4 differentially regulates the transcription of IFNL4 isoforms based on genotype. Moreover, the observation that only ~2% of mRNA transcripts arising from IFNL4 loci in heterozygous primary cells encode the active IFNL4 protein may explain the difficulty of detecting endogenous IFNL4 by traditional Western blot procedures (data not shown). IFNL4 mRNA isoforms that would produce inactive proteins (p107 and p131) were relatively abundant in the analyzed PHH cultures and we observed a similar trend
in cultured A549 cells where only transcripts excluding the IFNL4-specific exon were detectable. These observations suggest that alternative splicing of IFNL4 pre-mRNA could significantly regulate production of IFNL4 protein and subsequent anti-viral responses. It will be of interest to investigate whether common genetic variants may influence IFNL4 splicing or transcription.

Our findings and those of other groups present an apparent paradox: IFNL4 exhibits clear anti-viral activity in vitro yet its potential for expression is correlated with poor outcome of HCV infection. While this is seemingly incongruous, we suggest that two features of IFNL4 may contribute to its potential maladaptive role in HCV control. First, IFNL4 does not appear to be strongly expressed in primary cells. Second, IFNL4 is not efficiently secreted in vitro and is predicted to act in an autocrine manner in the liver. Based on these observations, it is possible that HCV infection triggers weak hepatic IFNL4 expression that is ineffective at halting productive infection, yet induces a state of refractoriness to secreted IFNL3 and IFN-α. In agreement with this scenario, multiple studies have documented elevated basal ISG expression in livers of patients who respond poorly to IFN-based therapy [34, 35, 37, 86, 87] and it is possible that this is driven by expression of IFNL4 in patients with rs368234815 ΔG.

However, rs368234815 ΔG and rs4803217 T are in high LD (r2=0.966 in European-Americans, 0.701 in African-Americans). Thus, individuals with HCV-induced IFNL4 expression would be predicted to express less IFNL3 in hepatocytes due to post-
transcriptional regulation by rs4803217, potentially exacerbating the maladaptive effects of IFNL4 expression.

The association between rs368234815 and pre-treatment viral load in the IDEAL cohort is consistent with previously reported findings that rs368234815 is significantly better correlated with normalized viral load decrease during treatment than the rs12979860 discovery SNP [58]. These findings are also consistent with a recent report describing a positive correlation between liver IFNL4 mRNA and HCV RNA levels [88]. However, although the IDEAL cohort is one of the largest and best characterized HCV cohorts available, there is presently not enough statistical evidence for stronger association of rs368234815 with SVR than the discovery SNP in this cohort. Despite this, rs368234815 has been reported to be significantly more predictive of SVR than rs12979860 in a Swiss cohort of chronically-infected patients [89]. However, in the Swiss cohort, the reported LD between rs12979860 and rs368234815 is r2=0.91, which is lower than the LD reported in this study (r2=0.977). This discrepancy may be explained by differences in population structure between the two cohorts, although both cohorts are composed of self-reported Caucasians. The extremely high LD in the Caucasian population in the IDEAL cohort may also limit the statistical power to distinguish the independent association of the two variants.

In summary, we explored rs368234815 as a candidate causal variant underlying the GWAS association for HCV clearance using diverse experimental approaches.
However, a definitive causal role for rs368234815 in determining HCV clearance (SVR) is still uncertain and merits further investigation, including demonstration of endogenous IFNL4 protein by Western blot and/or proteomic methods.
4. Multi-electrode array models and functional characterization of seizures

4.1 Introduction

In this chapter, we evaluate the multi-electrode array (MEA) as a potential in vitro platform to functionally characterize neurological disease, such as seizure. We demonstrate that the MEA platform is able to recapitulate not only epilepsy-relevant phenotypes of a mouse model, but also the mouse response to drug therapy.

MEAs contain arrays of electrodes that detect electrical signals from electrically active cells. Cultured cells are seeded onto MEA plates that are embedded with an array of electrodes that detect action potentials for recording and downstream analyses. MEAs detect the change in extracellular field potentials that reflect the intrinsic action potentials of cultured neurons and thereby directly monitor neuronal network activities without damaging the cells. MEAs have been extensively used to assess neurotoxicity of compounds but are more recently being used to study the behavior of neurons carrying genetic mutations of interest [20, 90-94]. Because they are non-invasive, MEAs can monitor the whole course of neuronal network development in which cells form connectivity from early to mature stages without intracellular perturbation. The cellular

1 The results of this chapter were submitted for publication entitled “Multi-electrode array models support the use of fluoxetine to treat seizures caused by CELF4 deficiency” to be considered by Annals of Clinical and Translational Neurology. Collaborators in this work include: Sahar Gelfman, Quanli Wang, Xiaodi Yao, Slavé Petrovski, Tiffany J. Tzeng, K. Melodi McSweeney, Erin L. Heinzen, Shelton S. Bradrick, Andrew S. Allen, Yong-hui Jiang, Steven Petrou, Wayne N. Frankel, and David B. Goldstein.
mechanism of seizure initiation is thought to be the network hypersynchronization and high frequency bursts consisting of increased density of action potentials, presumably due to an excitation/inhibition imbalance [19]. While the \textit{in vitro} manifestations of seizures are not fully understood, it is thought that both increased synchronicity of network firing and increased bursting are analogues of the \textit{in vivo} phenotype [20].

\textit{CELF4} is one of six mammalian CELF (CUGBP and ETR3-like Factor) RNA-binding proteins involved in multiple steps in RNA processing, including alternative splicing, RNA stability, and translational regulation. \textit{CELF4} encodes 486 amino acids and, like the other CELF proteins, binds RNA through RNA recognition motif (RRM) domains. In the adult mouse, \textit{CELF4} is most highly expressed in cerebral cortex and hippocampus and appears restricted to excitatory neurons. \textit{Celf4} deficiency is known to cause neurological phenotypes in mice including, most prominently convulsive seizures [95, 96]. In the context of Del(18)(q12.2q21.1) syndrome, \textit{CELF4} haploinsufficiency is also associated with seizures in human, although of about 20 patients described to date only one microdeletion is limited to \textit{CELF4} [97-99]. Mouse \textit{CELF4} was previously shown to bind the 3’ untranslated region (UTR) of target mRNAs, including those highly enriched for synaptic function [100]. The mRNA and protein levels of several \textit{CELF4} targets are known to be decreased at the steady-state level as a result of \textit{Celf4} deficiency. In particular, the protein level of serotonin receptor 2c (HTR2C) is reduced by nearly 60% in \textit{Celf4} knockout (KO) mice compared to the wildtype (WT) [95]. Although the
physiological causes of epilepsy in Celf4 deficient mice are likely to be complex and due to multiple dysregulated mRNAs, because fluoxetine is known to inhibit serotonin re-uptake, it was previously tested in Celf4 knockout mice and shown to lead to a 50% decrease in spike-wave discharges associated with absence-like non-convulsive seizures [95, 101, 102].

Although epilepsy phenotypes attributed specifically to Celf4 deficiency are clear in mouse models, beyond Del(18)(q12.2q21.1) causality in human disease has not yet been established. Nevertheless, it seems plausible that CELF4 mutations might play a more common role in human epilepsy than has so far been recognized.

Here, we use Celf4 knockout mice for “proof of concept” of whether the MEA platform can recapitulate both excitability phenotypes and the in vivo mouse response to fluoxetine. We have constructed a pipeline for the analysis of the MEA multi-dimensional output (representing time, frequency and channels) and used both published and novel algorithms to test the results [103]. Using these tools the neuronal network of cortical neurons from Celf4 KO mice demonstrated a clear pattern of denser bursts compared to the usually sparser bursting features of the WT. In the synchronized bursting (network bursts), elevated spike intensity in network bursts was observed in the Celf4 KO neurons. We also characterized the change in activity after fluoxetine treatment and observed that fluoxetine rescued the spike intensity in network bursts in the KO neurons when compared to the WT level. These results demonstrate a successful
characterization of a seizure phenotype on the MEA platform and support MEA as a screening platform for candidate epilepsy therapies.

4.2 Materials and Methods

4.2.1 Mouse

Celf4 knockout mice on either the 129S1/SvImJ (congenic) or the C57BL/6J coisogenic inbred strain were used for studies [96]. All mice were housed, bred with approval of the respective Institutional Animal Care and Use Committee (IACUC) at Duke University or at The Jackson Laboratory. The animal studies were conducted based on the United States Public Health Service’s Policy on Humane Care and Use of Laboratory Animals.

4.2.2 Disassociation and culturing of primary mouse cortical neurons

Primary mouse cortical neurons were obtained from the new born P0-1 mice from the same litter. 48-well MEA plates were coated with 0.5 % PEI and laminin (20 ug/mL) prior to the seeding of disassociated neurons. Dissection and disassociation protocols were adopted and optimized from the published work of cortical culture on MEA 5. Cerebral Cortex was removed in ice-cold Hank’s Balanced Salt Solution (HBSS) buffer, and mechanically disassociated by scissors. The tissue was subsequently subject to trypsin (8 minutes) followed by DNase (6 minutes) treatment in 37 Celsius bead bath. Trypsin was neutralized by MEM (10% FBS, glucose, 5 mM HEPES, and Penicillin Streptomycin) and centrifuged at 200 rcf for 5 mins followed by trituration by flame-
polished glass Pasteur pipette. 150,000 cells were seeded on top of each well covered by electrodes. Plating was done in a mixed pattern to control for confounding between spatial effects and genotype on the plates. MEM (10% FBS, glucose, 5 mM HEPES, and Penicillin Streptomycin) was replaced by Neurobasal-A media (Life Technologies) supplied with B-27, 5mM HEPES, and Penicillin Streptomycin after three hour of incubation. Arabinofuranosyl cytidine (ara-C) (5 µM) was added to the cultured neuron at DIV3 and washed away at DIV5. 45% of media was changed every two days starting from DIV12.

4.2.3 Drug treatments on cultured neurons

Fluoxetine (Sigma-Aldrich; catalogue number F132) and were dissolved in dimethyl sulfoxide (DMSO) as 50 mM stock, and carbamazepine were dissolved in DMSO as 0.5 M stock. Fluoxetine or carbamazepine was administered to neurons at the indicated concentrations. At the days of media change, 40% of old media was removed and fresh Neurobasal-A media containing the same concentration of fluoxetine or carbamazepine was added to the wells.

4.2.4 Tonic-clonic seizure testing

Mice that had two handling associated seizures were qualified to enter the study. Mice were then trained to consume frozen peanut butter pellets ("Jif Creamy") for three to five consecutive days, delivered on a spatula through the cage lid until they ate the pellet within 5 minutes. Once trained, dosages of fluoxetine with various combinations
were used in this study on both maintenance days (days between test days) and the test days (Appendix C: Table 12). On test days, mice were tested for seizures approximately one hour following consumption, based on the known time of peak effect for fluoxetine. Four separate tests were performed with a frequency of one test every three days. Average daily fluoxetine intake was calculated for each mouse and categorized the cohort into three groups for data analysis (average daily fluoxetine dosage 0 mg/kg, 15-18 mg/kg, and 22.5-24.6 mg/kg).

For CBZ \textit{in vivo} experiments, CBZ (25 mg/kg) was given to trained mice daily in frozen peanut butter pellets [104-106]. The same seizure testing procedure described above was used, except on test days the mice were tested 30 minutes following consumption, based on the shorter known time of peak effect for CBZ.

\textbf{4.2.5 MEA recording and spike classification}

The Maestro 48-well plates (Axion Biosystems, Atlanta, GA), which contains 768 electrodes on this plate, were recorded at a rate of 12,500 Hz with AxIS 1.9 software daily for 15 minutes on a 37\textdegree C heat block in a closed chamber filled with Carbogen (5\% carbon dioxide). The threshold of spike detection was set as an adaptive threshold at 7 Standard Deviation (SD). The spike train was detected based on a moving window where deviations of at least 7-fold from the standard deviation of the overall electrical measurements were classified as a spike.
4.2.6 MEA Data analysis

R package

The R package sjemea developed and maintained by Stephen J. Eglen (http://www.damtp.cam.ac.uk/user/sje30/r/) was used to extract MEA features at the electrode level for spikes, bursts and network spikes from Axion BioSystems recordings [103]. We then develop code and algorithms to compute and aggregate corresponding features at the well level and perform analysis using the R environment.

Network detection

The algorithm developed by Eytan and colleagues was used to detect the network spikes within each well on a plate 19. A network spike is detected for a well if there are at least N active electrodes with each having at least E spike(s) detected within a common time bin of T milliseconds. The parameters N and T are chosen following Eytan’s suggestion but alternative values were also used to perform sensitivity analyses. The parameter E introduced here serves as a tuning parameter to deal with scenarios where one electrode might be recording multiple neurons in a densely populated cell culture. Setting E to 1 defaults our network spike detection to Eytan’s algorithm.

Following the network spike detection, other features such as duration of each network spike is computed using R package sjemea and spikes within each network spike are identified and tallied to generate network spike related features. For example, the
feature ‘spikes in network spike’ for each well is defined as the total number of spikes that are within the duration of all identified network spikes.

**Network burst detection**

Network burst patterns at the well level were also investigated using a method that considers bursts at the electrode level as well as the synchronization of bursts across a well. Firstly, spike time within spike trains from all electrodes was normalized through binning. Since the spike detection algorithm sets a default minimum distance between any two spikes to greater than 2.16ms, we used a bin size of 2ms to guarantee that at most one spike is called within each bin. Secondly, a Gaussian Filter with a given window size (5 separate window sizes were used) was applied to smooth the binned spike train from each electrode. The smoothed signal was then further standardized to have a maximum signal value of 1. All smoothed signals at the electrode level were then further added together and smoothed again using the same Gaussian filter. The final result from this step is a smoothed signal at a given window size that measures the overall synchronization of all electrodes in a well, with larger values indicating higher level of synchronized bursting activities. Thirdly, the Otsu global thresholding method is applied to the well level signal from the previous step to automatically detect the burst intervals [107]. This method was chosen for its simplicity and parameter free nature, although other methods, such as adaptive thresholding, can be utilized. Finally, based on the network burst intervals from Otsu thresholding, we collected all burst
features at the well level, such as network burst frequencies, inter network burst intervals etc.

**Stable period selection**

To test several DIVs from the same experiment, a time window is selected where MFR of WT wells had a minimal fluctuation between DIVs. The selection of stable period was dependent on the developmental curve and slightly different among experiments, but the stable period generally lasted 3-6 days of consecutive recording (Table 6). This selection of DIV windows was used for all statistical tests (except drug treatment) in this report including network synchronization, MFR, network bursting, and burst distributions.

**Permutation analysis**

In the network synchronization analysis, the Mann–Whitney U test and permutation of a stabilized window of days *in vitro* (DIVs) was used. The purpose was to determine whether the mean value of wells corresponding to the actual genotype labels are more significant than randomly shuffled well labels generated by the permutation. In order to avoid batch-to-batch variation, we only compared the KO and WT wells from the same experiment from littersmates and then used Fisher’s method to combine the permutation p values from these independently performed experiments.
Burst detection

Bursts were detected using the Max Interval burst detection algorithm (Neuroexplorer software, Nex Technologies, Littleton, MA) implemented in sjemea. The same set of parameters was used in all recordings and the values were chosen based on the literature and experimentation [108, 109]. More specifically, the parameters were set at a burst consisting of at least 5 spikes and lasts at least 50 milliseconds and that the maximum duration between two spikes within a burst to be 50 milliseconds. Adjacent bursts were further merged if the duration between them is less than 0.8 second. Five burst features were evaluated, inter burst interval (IBI), inter spike interval (ISI) within burst, burst duration, number of spikes in a burst and spike frequency within a burst. For each feature, a normalized histogram with values between zero and one was calculated to adjust for variability between different electrodes. The data from all electrodes, per genotype, were combined to accommodate for electrodes with variable levels of activity. The normalized histograms were calculated for both genotypes by averaging all the electrodes belonging to a genotype. At this point, the distributions for each of the five bursting features were derived based on the complete WT and separate the Celf4 KO recordings. The difference between the two distributions was tested by calculating the earth mover’s distance (EMD) from the R emdist package [110]. Distance between distributions was then tested using a permutation test to evaluate the strength of the difference between the original distributions and 10,000 random sets.
Feature distributions of all participating electrodes in all selected DIVs are combined to one dataset per genotype, on which the EMD and permutation tests are performed. Fisher's method is then used to combine the permutation results from all six experiments [111, 112].

The above method allows the observation of differences between genotypes along full recordings, while still accounting for variability of the activity between channels and wells that could arise from the different types of neurons in a cortical culture or the distance of a recorded neuron from an electrode.

**Permutation test**

Permutation was performed in R (www.r-project.org). To generate each permutation dataset, labels (genotypes \ treatments) of the individual wells were randomly shuffled but observations within each well were kept intact. This preserves correlations between time points within wells while breaking any relationship with genotype and subsequent outcome. This basic permutation scheme was used in two different analyses:

1. For burst distributions analyses:

   For each well we generate a normalized histogram of the values of each bursting feature (duration, frequency, etc) over a specific time range. For example, normalized histogram of spike frequency is generated over the possible range of 0-300 spikes/second. For each of 10,000 permuted datasets we compute an average of these
normalized histograms within each genotype group. These distributions are then compared for each of the burst features using both Earth Mover’s Distance (EMD), using R package emdist (http://www.rforge.net/emd) and maximum distance (MD). Similar values (EMD and MD) are also computed for the original (unpermuted) data. A permutation p-value for EMD was computed as the proportion of permuted data EMD values that are equal to or greater than the original EMD value from the unpermuted dataset. A permutation p-value for MD was defined similarly.

2. For all features other than burst distributions:

For each of 10,000 permuted datasets the Mann–Whitney (MW) test was performed comparing distributions in outcome variables between genotypes and its p-value was recorded. A permutation p-value for the MW test was computed as the proportion of permuted data MW p-values that were less than or equal to the MW p-value form the original unpermuted dataset.

**Fisher’s Method**

Fisher method was used to combine p values of independently performed experiments (biological replicates). R scripts (www.r-project.org) were developed to allow the input of multiple p values to generate a combined p value by the Fisher’s method previously described in the literature [111, 112].
4.3 Results

4.3.1 Cortical neurons from Celf4 homozygous knockout (KO) mice showed elevated spike intensity in network bursts

We hypothesized that cultured neuronal networks monitored by MEAs would reveal, compared with WT, excitability phenotypes caused by the epileptogenic variant. Celf4 KO newborn mice were dissected and disassociated progenitor neurons were obtained from the same litter for each of the three genotypes: homozygous KO (Celf4−/−), heterozygous KO (or Celf4wt/−), and the WT (Celf4wt/wt) genotypes. Disassociated neurons from cerebral cortex of both genotypes were plated in equal density (150,000 cells per well or 4.7K cells/mm2) on 48-well MEA plates and recorded by the Axion Maestro system (Axion Biosystems). Spike trains were then subjected to comprehensive analysis for detecting network synchronization (network spike and network burst) and bursting features (see Methods).

Both WT and KO neurons showed increased synchronization along the course of development, suggesting the maturation of the network through formation of connectivity and neuronal circuits as previously described by Eytan and colleagues [103]. A raster plot of increased synchrony during development is shown in Appendix C: Figure 24. In order to quantify network synchronization and statistically identify the effect induced by Celf4 deficiency, we adopted the algorithm developed by Eytan and colleagues from the R package sjemea to detect correlated firing among electrodes in the same well as a “network spike” [103]. Additionally, a novel algorithm was developed to
capture the synchronized bursting phenotype (network burst) (Figure 11A). A permutation method was adopted to assess the statistical significance between genotypes of the same plate, and Fisher’s method was used to determine the combined significance of independently performed experiments (see 4.2 Materials and Methods).

Among the network burst and network spike features, spike intensity in network bursts was identified to be the feature that Celf4 KO neurons consistently showed significant elevation compared to the WT neurons across six independent experiments (combined permutation p value $1.58 \times 10^{-9}$) (Figure 11B, Table 6 and Figure 25). Since each well may contain different numbers of active electrodes (AE) (defined as electrode with greater than 5 spikes per minute), which can lead to higher overall spike intensity per well because of the higher number of AE, this variability was controlled by dividing by the number of AE in the network burst analysis. After controlling for number of AE, the intensity in network burst feature remained highly significant (combined permutation p value $2.29 \times 10^{-8}$) (Figure 11C and Table 6). We also confirmed that this significant distinction between KO and WT neurons was insensitive to the change of bin size, which is the time frame in millisecond (ms) used to calculate the accumulative spike intensity, for network burst detection algorithm (bin size ranging from 5 ms to 100 ms) (Appendix C: Table 10). A representative plot for results generated by different bin sizes is shown in Figure 11A. These results demonstrated spike intensity in network burst is a major MEA phenotype.
**Figure 11:** *Celf4* knockout neurons showed elevated spike intensity in network bursts, but moderate elevation of mean firing rate (MFR) on MEA.

(A) A representative raster plot and network burst detection of cortical neurons dissected from newborn mice that were plated on 48-well MEA plate. The red lines represent the global threshold for network burst detection. (B) *Celf4* KO neurons showed elevated spike intensity in network burst. (C) Spike intensity in network burst was normalized to the number of active electrodes (AE) to control variation in the number of AE among wells. (D) The mean firing rate (MFR) of *Celf4* KO and WT neurons across DIVs. The number of replicates (wells) in each genotype in Figure 11A-C is n = 15-16. The mean value of each well is used to calculate the overall mean ± SEM for each genotype. Permutation p values for each experiment were shown in Table 6.
of Celf4 neurons and supports MEA as a tool to characterize epileptogenic mutations through the analysis on in vitro network excitability.

Since Celf4 heterozygous mice also showed limbic and tonic-clonic seizures in vivo but with a later onset than homozygotes [95], we also investigated whether the heterozygous neurons showed similar alterations in network synchronization by the same statistical method described above. However, no significant elevation of spike intensity in network burst was detected compared to the WT neurons, although heterozygous neurons also showed a trend of elevation in spike intensity in network burst but with smaller magnitudes of difference (Appendix C: Figure 26). Therefore, we focused drug testing experiments on homozygous KO neurons where a significant MEA phenotype was found.

Table 6: MEA phenotype of Celf4 KO neurons.

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>Statistical test window (based on stability of MFR)</th>
<th>Spike intensity in network burst (permutation p value)</th>
<th>Burst distribution (EMD permutation p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIV16–18</td>
<td>DIV16–19</td>
<td>DIV18–23</td>
</tr>
<tr>
<td>10325</td>
<td>0.0001</td>
<td>0.0017</td>
<td>0.00008</td>
</tr>
<tr>
<td>103134</td>
<td>0.0001</td>
<td>0.1005</td>
<td>0.0094</td>
</tr>
<tr>
<td>103210</td>
<td>0.034</td>
<td>0.0035</td>
<td>0.0037</td>
</tr>
<tr>
<td>103231</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0017</td>
</tr>
<tr>
<td>103446</td>
<td>0.0008</td>
<td>0.0010</td>
<td>0.17</td>
</tr>
<tr>
<td>10343</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Permutation was performed 10,000 times to determine the permutation p value.
*Fisher’s method was used to combine p values of independently performed experiments.
**Spike intensity is divided by the number of active electrode (AE) to control for the variation in number of AE in each well.
4.3.2 Minimal difference in mean firing rate (MFR) was observed in Celf4 KO neurons

Elevated mean firing rate (MFR) is a commonly reported MEA feature when investigating epileptogenic or neurological mutations in neurons cultured on MEA [20, 91, 92]. Here we adopted the sjemea R package to analyze the mean firing rate (MFR) differences between Celf4 KO and WT neurons. The average values of each active electrode of a full recording were calculated for each Celf4 KO and WT well. Overall, Celf4 KO neurons showed a slight trend of elevation of MFR by active electrodes compared to WT neurons, but the combined permutation p value from six independent experiments shows only a modest significance (combined p value by Fisher’s method is 0.042) (Figure 11D and Figure 25). The magnitude of MFR elevation for the Celf4 KO neurons is also much smaller than what was previously reported in Tsc1 and Syn1 seizure models assessed using the MEA [20, 91]. This result indicates that MFR is not a major MEA phenotype for Celf4 KO neurons.

4.3.3 Higher spike frequency and shorter inter-spike interval (ISI) in bursts in Celf4 KO cortical neurons

We next examined the bursting patterns from the networks of cortical neurons. An analysis pipeline was developed (see 4.2 Materials and Methods) and the following five burst features were analyzed and compared between KO and WT neurons: inter burst interval (IBI), inter spike interval within bursts (ISI), number of spikes in bursts, duration of bursts, and spike frequency within bursts. The earth mover’s distance (EMD)
method was adopted to test the difference in the two distributions and a permutation method is used to assess whether the genotype effect is greater than random assignment of well labels (see 4.2 Material and Methods). Figure 12A shows a representative bursting calling result by the algorithm.

Strikingly, in all six experiments, the distribution of spike frequency in *Celf4* KO bursts was elevated and shifted to higher frequency (Hz) (Figure 12B, Table 6, and Figure 25). To account for high/low firing electrodes in each well, the values of each feature (spike frequency, ISI, etc.) are normalized to one. When we tested by the permutation of EMD values to determine the significance of the difference between WT and KO distributions (see 4.2 Material and Methods), six experiments support a very significant difference and the combined Fisher’s p value is $2.13 \times 10^{-8}$. 
Figure 12: Normalized Histograms of bursting features.

(A) A representative raster plot demonstrating the results of burst calling by the maximum interval method. Red line represents the length of a burst and the number in blue color is the number of spikes in a burst. (B) A normalized histogram of the frequency of spikes within bursts. (C) A representative normalized histogram of inter-spike interval (ISI) within bursts and (D) the normalized histogram of number of spikes within bursts. The permutation p values and the normalized histograms of spike frequency and ISI within bursts of all experiments were shown in Table 6 and Figure 25.
This result of higher spike frequency within bursts in Celf4 KO neurons is not only highly reproducible among experiments, but also corroborated with the network burst results previously described in this report. The burst detection algorithm adopted a maximum distance algorithm and was independent of the network burst detection framework. Both algorithms, with distinct methods, revealed similar results on the spike frequency feature in network bursting or bursting in Celf4 KO neurons.

Next, we found the inter-spike interval (ISI) values within bursts also differed between WT and the Celf4 KO neurons (Figure 12C and Table 6). A significantly shorter interval was observed between spikes in Celf4 KO (combined Fisher’s p value of all six experiments is $4.52 \times 10^{-8}$), which also corresponded with the higher frequency of spiking in bursts and network bursts that were observed in the Celf4 KO neurons.

A consistent trend of a higher number of spikes in Celf4 KO bursts was observed compared to the WT but with a less significant combined permutation p value ($8.17 \times 10^{-4}$) (Figure 12D and Table 6). Celf4 KO networks also had a tendency to produce bursts with longer durations compared to WT neurons, but this difference in the histogram distribution was not present in all the experiments that were performed. For inter-burst interval (IBI) values, there was not a consistent difference between WT and Celf4 KO neurons (Table 6). The overall effect on number of spikes in bursts, IBI, and duration of bursts are much weaker and less consistent among experiments and therefore were not considered as dominant MEA phenotypes for Celf4 KO neurons.
4.3.4 Fluoxetine reduces tonic-clonic seizures in Celf4-deficient mice

CELF4 was previously reported to regulate the expression of a subset of genes both in mRNA and protein level [95, 100]. One of the CELF4 binding targets is the mRNA that encodes serotonin receptor 2c (HTR2C), and the KO of CELF4 has been shown to decrease HTR2C protein expression by 60% in mouse hippocampus [95]. Fluoxetine, a clinically available anti-depressant, is a compound that inhibits the re-uptake of serotonin in the synaptic junctions, and is thought to increase the local concentration of serotonin within the synapse. Although the roles of serotonin neuromodulation in the Celf4 seizure disorder is not established, this motivated a previous trial of fluoxetine as a potential therapy, whereby it was shown to decrease the number of spike-wave discharges of Celf4 mutants by 50% [95]. While the spike-wave discharges – the electrophysiological hallmark of an absence seizure – is a relatively minor seizure type in Celf4 mutants, the effect of fluoxetine on the more major seizure type – generalized tonic-clonic seizures – was not previously examined.

The poor survival of Celf4+/− homozygotes to adulthood [95] makes it a challenge to do a drug study in vivo with sufficient sample sizes. However, recurrent tonic-clonic seizures are readily observed in over 70% of Celf4 heterozygous adults older than 90 days, following routine handling such as weekly cage changes [96]. To determine whether fluoxetine could decrease the incidence of handling-associated tonic-clonic seizures, an assay was developed whereby fluoxetine was administered daily orally and
non-invasively (without handling) in frozen peanut butter pellets (see 4.2 Materials and Methods) [113], and mice then were tested for handling-associated tonic-clonic seizures every three days for four tests total. The results demonstrated that fluoxetine had a significant impact on the reduction of seizure frequency of handling associated tonic-clonic seizures in Celf4 deficiency mice ($p = 0.001$ by repeated measure ANOVA on the raw response data), with continued treatment in the high dose range effectively eliminating handling-associated seizures (Figure 13). In sum, this in vivo fluoxetine treatment study further elaborated the effect of fluoxetine previously reported by showing the tonic-clonic seizure reductions in Celf4 deficient mice.

![Figure 13: Fluoxetine reduced tonic-clonic seizures in Celf4 deficient mice.](image)

*Celf4* heterozygous KO mice from the age of three months old and having exhibited at least two handling seizures during cage changes prior to the study, entered the study and were given pellet of fluoxetine or vehicle as controls ($N = 8$ in vehicle control; $N = 8$
in 15-18 mg/kg and N = 10 in 22.5-24.6 mg/kg group). Four tests were performed with a frequency of one test per every three days. Repeated measures two-way ANOVA (RM-ANOVA) was performed on the raw response data to determine the statistical significance of the effect of dosage and time (four seizure tests on the same animal were considered repeated measures). Dosage (0, 15-18, or 22.5-24.6 mg/kg) and time (test one to four) were used as independent variables to test the difference in the seizure frequency. The p value for the dosage effect of RM-ANOVA is 0.001. The dosage effects of fluoxetine-treated groups compared to the 0 mg/kg controls were both statistically significant (p = 0.0074 for 15-18 mg/kg group and p < 0.0001 for 22.5-24.6 mg/kg group). Error bar represents SEM. Seizure frequencies were normalized to the highest value within each dosage group.

4.3.5 Fluoxetine rescues spike intensity in network burst phenotype of Celf4 KO neurons

To evaluate whether MEA can also be used as a tool to screen for compounds with potential therapeutic benefits, we investigated whether any existing compounds are capable of reverting the MEA phenotypes described above toward the level of WT neurons. Since fluoxetine was documented to reduce SWD in Celf4 deficient mice and was further confirmed to reduce tonic-clonic seizures in vivo, these facts motivated a trial of fluoxetine in the MEA platform to determine whether fluoxetine could rescue the network bursting and bursting MEA phenotypes of Celf4 KO neurons.

It was previously documented in the literature that fluoxetine can reduce the net current in vitro by 50% at around 5 µM concentration (IC50 = 5.54 µM based on the dose response curve data) [114]. Because of the rationale that an appropriate concentration should have a substantial effect on neurons while avoid overdosing that will completely silence the field potentials, a trial of 5 µM of fluoxetine or vehicle (DMSO) was
administered to the *Celf4* KO neurons on the MEA. We examined whether fluoxetine decreased the level of the major network synchronization phenotype, spike intensity in network burst. Strikingly, the spike intensity in network bursts decreased significantly following the fluoxetine administration (Figure 14A and 14B). This significant reduction was apparent whether number of active electrodes (AE) was controlled for or not. Combined permutation p value for KO versus fluoxetine-treated KO neurons was $p = 9.2 \times 10^{-3}$, suggesting a significant reduction after fluoxetine treatment compared to the mock treated neurons. We further asked the question whether fluoxetine treatment could revert the *Celf4* KO neurons to the WT levels for the spike intensity in network burst phenotype. The original combined permutation p value between WT and KO from three independent experiments was $5.3 \times 10^{-3}$, but became insignificant ($p = 0.96$) when comparing WT to the fluoxetine-treated KO neurons (Table 7). This result suggests *Celf4* KO neurons treated by fluoxetine no longer demonstrated the network phenotype (elevated spike intensity in network burst) and supports the rescue effect induced by fluoxetine.
Figure 14: The spike intensity in network burst phenotype of Celf4 KO neurons was rescued by fluoxetine treatment.

(A) The spike intensity in network burst in Celf4 KO after fluoxetine (5 μM) treatment no longer elevated compared to the wildtype. (B) Spike intensity in network burst was normalized to the number of active electrodes of each well. (C) The mean firing rate was decreased by around 50% after cortical neurons were treated by fluoxetine. (D) A representative plot of normalized histograms of the inter-burst interval (IBI). Y axis represents bursts distribution while X axis represents inter-burst time in seconds. (E) A representative plot of normalized histograms showing the duration of burst decreased after fluoxetine treatment. In Figure 3A-C, data are presented as mean ± SEM (n = 6-7 in each treatment group). Permutation p values for all replicate experiments after fluoxetine treatment were shown in Table 7.
Table 7: Combined permutation p value of each treatment comparison.

<table>
<thead>
<tr>
<th></th>
<th>Fluoxetine (FLX)</th>
<th>Carbamazepine (CBZ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Combined</td>
<td>Combined</td>
</tr>
<tr>
<td></td>
<td>permutation p</td>
<td>permutation p</td>
</tr>
<tr>
<td></td>
<td>value by Fisher's method</td>
<td>value by Fisher's method</td>
</tr>
<tr>
<td>Network burst (permutation p value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spike intensity in network burst (per well)</td>
<td>0.0053</td>
<td>6.08×10\textsuperscript{-7}</td>
</tr>
<tr>
<td>Spike intensity in network burst (by number AE)*</td>
<td>0.0065</td>
<td>1.51×10\textsuperscript{-6}</td>
</tr>
<tr>
<td>Burst distribution (EMD permutation p value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spike frequency in burst</td>
<td>2.32×10\textsuperscript{-7}</td>
<td>2.13×10\textsuperscript{-6}</td>
</tr>
<tr>
<td>Inter-spike interval (ISI)</td>
<td>6.13×10\textsuperscript{-5}</td>
<td>6.13×10\textsuperscript{-5}</td>
</tr>
<tr>
<td>Number of spikes in burst</td>
<td>0.0453</td>
<td>0.0453</td>
</tr>
<tr>
<td>Inter-burst interval (IBI)</td>
<td>0.0252</td>
<td>0.0252</td>
</tr>
<tr>
<td>Duration of burst</td>
<td>0.0016</td>
<td>0.0168</td>
</tr>
</tbody>
</table>

* Spike intensity is divided by the number of active electrode (AE) to control for the variation in number of AE in each well.

4.3.6 Additional effects of fluoxetine on MEA

In addition, fluoxetine also decreased the MFR by 50%, when compared to vehicle (DMSO) alone, which is consistent with the known function of fluoxetine to reduce IA (the integral of the outward current) in cultured neurons (Figure 14C) [114].
The distributions of bursting features were also examined for the experiments treated with fluoxetine. However, fluoxetine did not completely rescue spike frequency, ISI, and the number of spikes in a burst. The combined permutation p value is still significant when comparing WT and fluoxetine-treated KO histograms (p = 1.21×10^{-4} for spike frequency in burst, and p = 1.04×10^{-4} for ISI in burst) (Table 7). However, the original combined permutation p values for mock treated KO and WT neurons was 2.32×10^{-7} for spike intensity in burst, and the fluoxetine treatment on KO neurons increased the combined permutation p value by three orders of magnitude (1.21×10^{-4}), suggesting although fluoxetine did not completely revert the burst histogram distribution to WT, the effect of fluoxetine still shifted KO neurons slightly toward the WT distribution (without fluoxetine p value = 2.32×10^{-7}, and with fluoxetine p value = 1.21×10^{-4}).

In addition to the investigation of the main MEA phenotype of Celf4 KO neurons, we also identified additional effects on the MEA induced by the fluoxetine treatment that were absent from either WT or KO mice. Fluoxetine-treated neurons exhibited a distinct pattern of IBIs that centered around three to five seconds (Figure 14D). The neurons after treatment did not show the high abundance of IBIs in 0.5-2 seconds as previously observed for the wildtype and Celf4 KO. In addition, Celf4 KO neurons treated with fluoxetine showed a significant reduction in the burst duration (combined permutation p value of 3.29×10^{-7}). This increased IBI and decreased burst duration
reflects shorter bursts and the longer quiescent period between bursts and created a unique signature of fluoxetine treatment on the MEA platform.

4.3.7 Carbamazepine did not rescue the spike intensity in network burst phenotype

Since fluoxetine treatment also reduced MFR in the Celf4 KO neurons, the reduction of MFR could also influence other MEA features, especially spike frequency-related features. Here we tested another drug, carbamazepine (CBZ) on Celf4 neurons and showed that the network phenotype (spike intensity in network burst) was not rescued by CBZ while CBZ did reduce the MFR by the same level.

Carbamazepine is one of the first choice anti-epileptic drugs commonly prescribed for patients with focal epilepsy. The mechanism of action of carbamazepine is thought to be through the binding to voltage-gated sodium channels and stabilizes the channel opening to reduce the excitability [115]. The dose response data were obtained from the published literatures and 50 µM was used for the purpose to decrease MFR by 50% [116-118]. After CBZ administration on the MEA, a 50% decrease in mean firing rate (MFR) in the CBZ treated wells was observed (Figure 15C). However, contrary to the fluoxetine treatment results, the spike intensity in network burst did not drop and
Figure 15: Carbamazepine (CBZ) did not rescue the spike intensity in network burst phenotype.

(A) The spike intensity in network burst in Celf4 KO after CBZ (50 µM) treatment was still elevated compared to the wildtype. (B) Spike intensity in network burst was normalized to the number of active electrodes of each well. (C) CBZ decreased mean firing rate (MFR) by around 50%. N=12 in each genotype. (D) A representative normalized histograms of the duration of burst after CBZ treatment. Data are presented as mean± SEM in Figure 4A-C. Permutation p values for all replicate experiments after CBZ treatment were shown in Table 2. (E) N = 5 in vehicle and N = 5 in CBZ 25 mg/kg group. Repeated measures two-way ANOVA (RM-ANOVA) was performed on the raw response data to determine the statistical significance of the effect of dosage and time. The p value for the dosage effect of RM-ANOVA is 0.39 (not significant). Error bar represents SEM. Seizure frequencies were normalized to the highest value within each dosage group.

remained increased for the KO neurons after CBZ treatment (Figure 15A-B). The combined permutation p values are still significant in KO neurons with or without CBZ treatment compared to the WT (p = 6.08×10⁻⁷ without CBZ treatment and p = 7.16×10⁻⁷ with CBZ treatment) (Table 7). For burst distributions, CBZ also did not affect most of the burst distributions, except for creating a unique distribution in burst durations (Figure 15D and Table 7). In summary, despite both compounds decreased the mean firing rate by a similar level, we did not observe the same effect of carbamazepine as fluoxetine in rescuing the spike intensity in network bursts.

4.3.8 Carbamazepine in Celf4 deficient mice

Although there was no previous evidence in the literature that CBZ reduces tonic-clonic seizures in Celf4 deficiency mice, we also performed a test to determine the \textit{in vivo} efficacy of CBZ on Celf4 deficiency-induced tonic-clonic seizures. Mice were
orally fed CBZ daily and assessed for handling-associated tonic-clonic seizures for four tests, similar to the fluoxetine study described above (see 4.2 Materials and Methods). The result shows that CBZ does not have any protective effect compared to the control group (vehicle only) of the same batch of mice tested at the same time. The mean value of seizure frequency is higher in the CBZ treated group compared to the control group, and the difference between treatments is insignificant ($p = 0.39$ by repeated measure ANOVA) (Figure 15E).

4.4 Discussion

In this chapter, we demonstrate the effectiveness of adopting MEA as a platform to model seizure phenotype and drug treatment in vitro. We utilized the primary cortical neurons from the Celf4 KO mouse model and identified excitability phenotypes in the Celf4 knockout neurons on MEA, which can be rescued by fluoxetine, but not carbamazepine. Carbamazepine provided a useful comparison: it did reduce the mean firing rate (MFR) but unlike fluoxetine did not alter the more epilepsy-relevant spike intensity in network burst phenotype. This result further supports the interpretation that the fluoxetine rescue results cannot simply be explained by the MFR reduction, which would likely be achieved by many standard antiepileptic compounds. These results not only successfully paralleled the seizure phenotype observed in the Celf4 KO mouse itself; it also recapitulated the response of the mouse model to fluoxetine. This
recapitulation suggests that compounds found to have efficacy in *in vitro* phenotypes monitored by MEAs may identify compounds with efficacy *in vivo*.

Although our current MEA throughput is not yet high enough to perform high throughput screening on diverse libraries, candidate testing of targeted compounds is certainly feasible through the current 48-well format, as used in this study. This suggests that it is possible to design experiments aimed at optimizing treatment for genetic epilepsies by selecting compounds with only the most promising *in vitro* effect.

Moreover, new formats for MEAs are likely to allow substantial increases in screening capacity. This advance in technology presents a great opportunity for the development of personalized medicine in the future for epilepsies and other neurological diseases [1, 16, 119, 120].

Despite the successful parallels between *in vitro* and *in vivo* model in this study, this approach may also have limitations. Our current protocol utilizes primary cortical neurons, and thus may not be capable of capturing the mechanisms that originate from cell types of other brain regions, such as hippocampus or thalamus. However, tissues such as hippocampal culture yield fewer cells that can limit the number of replicates available to statistically distinguish the phenotype or assess the drug rescue effect. Still, as long as a hippocampal neuron gene, for example, is expressed in cortical neurons, when separated from its *in vivo* context, it may well elicit network dysfunction in a manner that reflects the genetic lesion in the native network.
Celf4 deficiency was known to decrease the protein expression of serotonin receptor 2c (Htr2c) in mice. The presumed mechanism of fluoxetine action is to block the serotonin re-uptake [101, 102]. While fluoxetine has a significant tonic-clonic seizure mitigating effect on Celf4 deficiency in vivo as observed through the MEA system, the role of serotonin neuromodulation in this complex seizure disorder is not yet understood. We note that CELF4 is proposed to regulate the translation of hundreds of target mRNA through binding to 3’ untranslated regions (UTR). Among these targets is Nav1.6 which is known as the major initiator of forward-propagating action potentials of excitatory neurons [100, 121].

Although fluoxetine was not intended to treat or alleviate epilepsy symptoms, several reports documented the use of fluoxetine in epilepsy patients and suggest some therapeutic effect [122, 123]. However, concerns were also raised for the use of fluoxetine to treat seizures because of exacerbation in several patients and also the provoking of seizures as a side effect from high dose fluoxetine [124, 125]. Based on these results, we do not suggest fluoxetine as a potential therapy for epilepsy, nor do we know the extent to which the effects are specific to the Celf4 model. Rather, we offer the recapitulation of excitability phenotypes in the MEA as an example of the potential power and specificity of the MEA model.
5. Conclusions and future directions

5.1 Molecular genetics and Hepatitis C

Although many works have been dedicated to understand the molecular genetics of hepatitis C clearance, many questions still remain despite the substantial progress in this field. Part of the reason is that the lack of comprehensive understanding of in vivo mechanism of chronic HCV infection and its interaction with the immune system. It has been believed that hepatocytes play a major role in IFN secretion during the virus encounter. However, more and more evidence suggested that certain immune cell types might be responsible of the major IFN secretion during infection. It is crucial to ask the question of the effect of genetic variants in the most relevant cell types. However, in vivo mechanistic studies for hepatitis C is somehow difficult because of the suitable animal model and availability of patient samples. This also reflected again the fact the functional evaluation for mutations in human genetic studies relies largely upon the basic understanding of physiological and biochemical knowledge of the pathological pathway. It is generally difficult for geneticists to jump into an unknown field and figure out the effect of the mutations without substantial prior knowledge of the protein or disease. This fact should also be kept in mind when deciding to prioritize the mutations for functional characterization.

The functional characterization for HCV treatment response also pointed out the difficulty after GWAS to identify causal variants given the fact that many candidate
causal variants may exist with high LD in the nearby genomic region. It requires substantial efforts to gather a large size of patient cohort in order to distinguish the association signals for the causal variant, and also many independent functional works to ask the question of causality for each candidate causal variants. This fact also points out the importance of establishing a prioritizing method for functional evaluation on candidate variants.

5.2 **Functional modeling and precision medicine in epilepsy**

Chapter 4 demonstrates a successful example of recapitulation of seizure phenotypes and drug response *in vitro* on MEA. The results demonstrated in Chapter 4 may also be promising to functionally validate mutations identified through sequencing studies. Although here we still utilize cortical neurons from a genetically modified mouse model, similar approaches, such as transient knockdown or overexpression of epilepsy-causing genes can, in theory, also serve to model seizures “in a dish.” Current molecular tools, such as promoter-driven small hairpin RNA (shRNA) or open reading frame (ORF) constructs packaged in lentivirus, are certainly suitable to achieve such goals. While there are various benefits of creating and testing animal models and deriving cultures from them, these alternative *in vitro* approaches provide the opportunity to quickly screen candidate epilepsy mutations discovered from genome sequencing. This concept has been proven successful on MEA by modulating the levels
of gene expression by lentiviral transduction on WT neurons in a pilot study (McSweeney et al., manuscript under review).

We think that our study suggests a very promising future for MEA to become a platform to study epilepsy and to perform efficient in vitro drug screening. We anticipate the development of this MEA model to facilitate the genetic findings and targeted therapies and eventually moving forward to benefit patients with epilepsy.

5.3 Future directions

The recent advance in genome editing tools (CRISPR/CAS9 system) is expected to accelerate the functional characterization for human mutations because of the drastically lowered cost for engineering mutations into cells or an organism. The importance of functional evaluation in human genetics is expected to keep increasing, and hopefully through the combination with novel technology, such as genome editing, we will have more tools available to characterize the effect of mutations identified in human genetic studies.
# Appendix A: Additional information for IFNL3 mRNA structure

## Table 8: Raw SHAPE reactivities

| Raw SHAPE reactivities | sequence | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 |
Figure 16: Alignment of primate IFNL2 and IFNL3 sequences.

Location of the rs4803217 SNP is indicated. AREs are underlined in the consensus sequence.
Figure 17: Analysis of IFNL3 reporter constructs on HeLa, PH5CH8, and LH86 cells.

(A) Stable HeLa cell lines described in Figure 1 were induced to express the indicated reporter mRNAs by treatment with tetracycline and cell lysates were analyzed for luciferase activity at 2 and 3 hours after tetracycline treatment. Data from untreated cells is shown as the “0” hour time point. Data are shown as mean values ± s.d. (B) Same as in panel (A) showing only the untreated cells. (C) PH5CH8, LH86, and HeLa cells were transiently transfected with the indicated reporter plasmids as described in the Materials and Methods. For each cell line, the RLuc value was set to equal 1. P values were calculated by unpaired t-test (** p = 0.018, * p = 0.021, * p = 0.030).
Figure 18: SHAPE-MaP deconvolution of multiple alleles.

Total cellular RNA was isolated from poly(I:C)-transfected A549 cells and treated with 1M7 reagent after RNA purification. Modified or DMSO-treated (control) RNA samples were subjected to IFNL3-targeted SHAPE-MaP. Resulting sequencing reads were sorted for nucleotide identity at the rs4803217 locus and used to generate structural profiles for each individual allele.
Figure 19: A549 cells inducibly express IFNL3 mRNA and are heterozygous at rs4803217.

(A) RT-PCR analysis of IFNL3 mRNA expression in A549 cells. Cells were either untreated (M), or transfected with vector plasmid (DNA), poly(I:C) (pI:C), or synthetic HCV genomic RNA from the JFH1 strain (HCV). Total RNA was harvested 4 hours after transfection and analyzed by RT-PCR and agarose gel electrophoresis. (B) Analysis of A549 DNA and RNA amounts for the IFNL3 3’ UTR rs4803217 SNP. DNA and RNA samples were analyzed by qPCR and RT-qPCR, respectively, using an assay that discriminates rs4803217 G versus T/U alleles. A549 cells are triploid at IFNL3 with two alleles for rs4803217 T and one for G. mRNA levels reflect the gene dosage.

Figure 20: Analysis of median Shannon entropy as a function of window size.

(A) 3 nt, (B) 11 nt, or (C) 21 nt. For reference, Figure 3 uses a window size of 5 nts.
Figure 21: SHAPE-directed minimum free energy secondary structure for the IFNL3 “G” allele 3’ UTR.

The structure includes 11 nt of coding sequence including the stop codon (in red). Location of rs4803217 is indicated by an arrow.
Figure 22: Predicted structure of the C45G mutant stem loop.

The C45G mutation is indicated in red and location of rs4803217 is shown. Note the predicted re-folding of the stem loop due in the C45G mutant.
## Appendix B: Additional information for IFNL4 functional characterization

Table 9: Raw data for IFNL4 mRNA isoforms expressed in stimulated PHHs from two donors.

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Genotype</th>
<th>Amino acids</th>
<th>Count</th>
<th>Subject to NMD</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>p131</td>
<td>TT</td>
<td>75</td>
<td>10</td>
<td>Yes</td>
<td>21.7%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>131</td>
<td>3</td>
<td>No</td>
<td>6.5%</td>
</tr>
<tr>
<td>p179</td>
<td>TT</td>
<td>123</td>
<td>11</td>
<td>Yes</td>
<td>23.9%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>179</td>
<td>2</td>
<td>No</td>
<td>4.3%</td>
</tr>
<tr>
<td>p170</td>
<td>TT</td>
<td>124</td>
<td>1</td>
<td>No</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>179</td>
<td>2</td>
<td>No</td>
<td>2.2%</td>
</tr>
<tr>
<td>p107</td>
<td>TT</td>
<td>51</td>
<td>5</td>
<td>Yes</td>
<td>10.9%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>107</td>
<td>5</td>
<td>No</td>
<td>10.9%</td>
</tr>
<tr>
<td>Others</td>
<td>TT</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>10.9%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>6.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>TT</td>
<td></td>
<td><strong>32</strong></td>
<td></td>
<td><strong>70%</strong></td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td></td>
<td><strong>14</strong></td>
<td></td>
<td><strong>30%</strong></td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td></td>
<td></td>
<td><strong>46</strong></td>
<td></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Genotype</th>
<th>Amino acids</th>
<th>Count</th>
<th>Subject to NMD</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>p131</td>
<td>TT</td>
<td>75</td>
<td>9</td>
<td>Yes</td>
<td>15.3%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>131</td>
<td>9</td>
<td>No</td>
<td>15.3%</td>
</tr>
<tr>
<td>p179</td>
<td>TT</td>
<td>123</td>
<td>13</td>
<td>Yes</td>
<td>22.0%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>179</td>
<td>0</td>
<td>No</td>
<td>0.0%</td>
</tr>
<tr>
<td>p170</td>
<td>TT</td>
<td>124</td>
<td>9</td>
<td>No</td>
<td>15.3%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>170</td>
<td>0</td>
<td>No</td>
<td>0.0%</td>
</tr>
<tr>
<td>p107</td>
<td>TT</td>
<td>51</td>
<td>6</td>
<td>Yes</td>
<td>10.2%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>107</td>
<td>4</td>
<td>No</td>
<td>6.8%</td>
</tr>
<tr>
<td>Others</td>
<td>TT</td>
<td>-</td>
<td>7</td>
<td>-</td>
<td>11.9%</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>TT</td>
<td></td>
<td><strong>44</strong></td>
<td></td>
<td><strong>75%</strong></td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td></td>
<td><strong>15</strong></td>
<td></td>
<td><strong>25%</strong></td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td></td>
<td></td>
<td><strong>59</strong></td>
<td></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>
Figure 23: The effects of IFNL3 or IFNL4 plasmids over a range of concentrations.

(A) The indicated amounts of plasmid expression constructs for IFNL3 and IFNL4 were transfected into Huh7.5 cells and infected 24 h later with YFV-17D. After 24 h of infection, cells were stained for YFV E2 glycoprotein and analyzed by high content imaging to determine rates of infection. (B) Huh7.5 cells were transfected with control siRNA or three separate siRNA duplexes targeting the IFNLR1 mRNA. After 48 h, cells were collected for RNA isolation and RT-qPCR analysis of IFNLR1 transcript levels. Levels of β2 microglobulin mRNA were used for calibration and values for non-silencing siRNA were normalized to 1.
Appendix C: Additional information for Celf4 functional characterization on multi-electrode array

Table 10: Intensity in network burst is insensitive to the bin size selection in the network burst detection algorithm.

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>10325</th>
<th>103134</th>
<th>103210</th>
<th>103231</th>
<th>103446</th>
<th>103433</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical test window (based on stability of MFR)</td>
<td>DIV16–18</td>
<td>DIV16–19</td>
<td>DIV16–19</td>
<td>DIV18–23</td>
<td>DIV18–23</td>
<td>DIV15–18</td>
</tr>
<tr>
<td>Intensity in network burst (permutation p value)</td>
<td>0.0001</td>
<td>0.0021</td>
<td>0.0019</td>
<td>0.0125</td>
<td>0.0019</td>
<td>0.0375</td>
</tr>
<tr>
<td>bin size = 5 ms (per well)</td>
<td>1.95×10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 5 ms (by nAE**)</td>
<td>2.59×10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 10 ms (per well)</td>
<td>0.0001</td>
<td>0.0007</td>
<td>0.0104</td>
<td>0.0211</td>
<td>0.0011</td>
<td>0.0040</td>
</tr>
<tr>
<td>bin size = 10 ms (by nAE)</td>
<td>3.94×10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 20 ms (per well)</td>
<td>0.0001</td>
<td>0.0017</td>
<td>0.0086</td>
<td>0.0198</td>
<td>0.0023</td>
<td>0.0066</td>
</tr>
<tr>
<td>bin size = 20 ms (by nAE)</td>
<td>1.02×10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 30 ms (per well)</td>
<td>0.0001</td>
<td>0.0017</td>
<td>0.0088</td>
<td>0.0363</td>
<td>0.0018</td>
<td>0.0414</td>
</tr>
<tr>
<td>bin size = 30 ms (by nAE)</td>
<td>1.58×10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 40 ms (per well)</td>
<td>0.0001</td>
<td>0.0019</td>
<td>0.0082</td>
<td>0.0495</td>
<td>0.0001</td>
<td>0.0074</td>
</tr>
<tr>
<td>bin size = 40 ms (by nAE)</td>
<td>2.29×10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 50 ms (per well)</td>
<td>0.0001</td>
<td>0.0016</td>
<td>0.0086</td>
<td>0.0484</td>
<td>0.0001</td>
<td>0.0085</td>
</tr>
<tr>
<td>bin size = 50 ms (by nAE)</td>
<td>2.84×10^-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 100 ms (per well)</td>
<td>0.0001</td>
<td>0.0018</td>
<td>0.0129</td>
<td>0.1028</td>
<td>0.0001</td>
<td>0.3432</td>
</tr>
<tr>
<td>bin size = 100 ms (by nAE)</td>
<td>4.18×10^-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Fisher’s method was used to combine p values of independently performed experiments

**Spike intensity is divided by the number of active electrode (AE) to control for the variation in number of AE in each well
### Table 11: Raw and combined permutation p value of each treatment comparison.

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>Fluoxetine (FLX)</th>
<th>Carbamazepine (CBZ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT vs KO</td>
<td>WT vs KO</td>
</tr>
<tr>
<td></td>
<td>KO vs FLX</td>
<td>KO vs CBZ</td>
</tr>
<tr>
<td></td>
<td>KO vs FLX</td>
<td>KO vs CBZ</td>
</tr>
<tr>
<td>Spike Intensity in network burst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 30 ms (per well)</td>
<td>0.0088</td>
<td>0.0587</td>
</tr>
<tr>
<td>bin size = 30 ms (by nAE**)</td>
<td>0.0194</td>
<td>0.0703</td>
</tr>
<tr>
<td>Burst distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spike frequency in burst</td>
<td>0.0035</td>
<td>0.0001</td>
</tr>
<tr>
<td>Inter-spike interval (ISI)</td>
<td>0.0038</td>
<td>0.0001</td>
</tr>
<tr>
<td>Number of spikes in burst</td>
<td>0.034</td>
<td>0.3955</td>
</tr>
<tr>
<td>Inter-burst interval (IBI)</td>
<td>0.017</td>
<td>0.5647</td>
</tr>
<tr>
<td>Duration of burst</td>
<td>0.031</td>
<td>0.0529</td>
</tr>
<tr>
<td>Spike Intensity in network burst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin size = 30 ms (per well)</td>
<td>0.8485</td>
<td>0.6783</td>
</tr>
<tr>
<td>bin size = 30 ms (by nAE)</td>
<td>0.7296</td>
<td>0.3858</td>
</tr>
<tr>
<td>Burst distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spike frequency in burst</td>
<td>0.0097</td>
<td>0.0005</td>
</tr>
<tr>
<td>Inter-spike interval (ISI)</td>
<td>0.0099</td>
<td>0.0001</td>
</tr>
<tr>
<td>Number of spikes in burst</td>
<td>0.52</td>
<td>0.0611</td>
</tr>
<tr>
<td>Inter-burst interval (IBI)</td>
<td>0.017</td>
<td>0.0001</td>
</tr>
<tr>
<td>Duration of burst</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

*Fisher's method was used to combine p values of independently performed experiments.**Spike intensity is divided by the number of active electrode (AE) to control for the variation in number of AE in each well.
Table 12: Mouse raw data of fluoxetine dosage and seizure responses.

<table>
<thead>
<tr>
<th>Mouse ID</th>
<th>Genotype</th>
<th>Between test day dose</th>
<th>Test day dose</th>
<th>Average daily dose bin</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>48694</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>44291</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48678</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48684</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48691</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48701</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>49282</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>49304</td>
<td>Celf4+/+</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>49305</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>49296</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49300</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49288</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49301</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>49306</td>
<td>Celf4+/+</td>
<td>15</td>
<td>15</td>
<td>15-18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>48689</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>15</td>
<td>15-18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48934</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>15</td>
<td>15-18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48639</td>
<td>Celf4+/+</td>
<td>20</td>
<td>30</td>
<td>22.5-24.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48700</td>
<td>Celf4+/+</td>
<td>20</td>
<td>30</td>
<td>22.5-24.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47633</td>
<td>Celf4+/+</td>
<td>20</td>
<td>30</td>
<td>22.5-24.6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>49305</td>
<td>Celf4+/+</td>
<td>20</td>
<td>30</td>
<td>22.5-24.6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48695</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>15</td>
<td>22.5-24.6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48944</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>22.5</td>
<td>22.5-24.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>48698</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>22.5</td>
<td>22.5-24.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>48688</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>30</td>
<td>22.5-24.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48690</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>30</td>
<td>22.5-24.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48947</td>
<td>Celf4+/+</td>
<td>22.5</td>
<td>30</td>
<td>22.5-24.6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 24: Representative raster plots of 60-second recordings in different developmental stages.
Raster plots of active electrodes from DIV7, DIV12, DIV17, and DIV22 are shown. X axis represents a time frame of 60-second recordings. Increasing synchrony among electrodes was observed along the development.

Figure 25: Selected MEA features of experiments performed in this study.

MFR by the number of active electrodes, spike intensity in network burst, histogram of spike frequency in bursts, and histogram of ISI in bursts are shown for six independently performed experiments. P values for each experiment are shown in Table 1. Error bar represents mean ± SEM.
Figure 26: Celf4 heterozygous KO neurons showed the same trend of elevation in spike intensity in network bursts but the difference was insignificant determined by permutation.

Neurons from heterozygous and WT newborn mice of the same litter were plated and cultured on 48-well MEA plate. The data were collected and analyzed the same way presented in the Result and Experimental Procedures sections. Error bar represents mean ± SEM.
References


Biography

Yi-Fan Lu was born on April 7th, 1987 in Taipei, Taiwan. He graduated from Nation Taiwan University (NTU) with a major in Life Sciences and came to Duke University through the Molecular Genetics and Microbiology (MGM) program.

During his undergraduate career, he spent most of his time in Dr. Shau-Chi Chi laboratory studying virology in aquatic organisms. He also visited North Carolina State University (NCSU) in Dr. Frederick Fuller laboratory for lentivirus research. After rotations during his first year in MGM at Duke, he joined the Goldstein lab in the summer of 2011, focusing on the genetic and functional analysis of hepatitis C treatment responses, and functional analyses of mutations for neurological diseases.

Publications


Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 347, 1436-1441. 2015


8. Wu YC, **Lu YF**, Chi SC. Anti-viral mechanism of barramundi Mx against betanodavirus involves the inhibition of viral RNA synthesis through the interference of RdRp. Fish & shellfish immunology 28, 467-475. 2010

Scholarships and Awards

  - Ministry of Education, Taiwan

- **Government Scholarship for Studying Abroad**, 2012–2013
  - Ministry of Education, Taiwan

- **American Association for the Advancement of Science (AAAS)**
  - Program for Excellence in Science

- **Chancellor’s Scholarship**, 2010–2012
  - Duke University