TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 06/10 pp484-490
Volume 20, Number 5, October 2015

Investigating Genotype 1a HCV Drug Resistance in NS5A Region via Bayesian Inference

Yao Fu, Gang Chen, Lizhi Fu, and Jing Zhang*

Abstract: Hepatitis C virus (HCV) treatment is on the cutting edge of medicine. Due to the high rate of mutations and low fidelity of HCV replication, resistant strains quickly become dominant in a viral population under the selection pressure of a drug. In this paper, we examined the drug resistance mechanism in the NS5A region of genotype 1a HCV virus by comparing the sequence data from interferon-ribavirin treated and untreated patients. To find the drug resistance difference, we used innovative Bayesian probability models to detect mutation combinations and inferred detailed interaction structures of these mutations. We aim to provide reference to drug design and mutation mechanism understanding through our work.

Key words: Bayesian model; Hepatitis C virus (HCV); drug resistance; NS5A

1 Introduction

Hepatitis C virus (HCV) is a single-strand RNA virus that was classified into at least six genotypes with several subtypes in each genotype. Different genotypes spread in different regions and have different response patterns to interferon-based therapy[1]. Clinically, interferon monotherapy treatment only leads to less than 20% sustained response of chronic patients, while IFN and ribavirin combined therapy has significantly improved response rates[2, 3]. Understanding the mechanism of antiviral resistance to IFN therapy will shed light on the design of better treatment strategies.

Variations, especially those in the NS5A region[4, 5] in HCV sequences, have been found to play a role in response to IFN-based therapies. NS5A is a nonstructural protein that hampers the function of dsRNA dependent protein kinase (PKR), an important mediator of IFN response, leading to resistance of IFN treatment[6, 7]. In 1995, Enomoto and Sato[8] defined an "interferon sensitivity determining region" (ISDR aa 2209–2248) in NS5A, implying that mutations in this region was related to resistance to IFN. Several studies have confirmed this finding[9–11]. Furthermore, another study found that mutations in PKR binding domain (PKRbd aa 2209–2274) of NS5A can block viral replication[12]. However, contradictory results are also obtained for these two regions, which makes the role of NS5A to IFN ambiguous. In addition, other domains in NS5A were also involved in HCV drug resistance, such as variable region 3 (V3 aa 2356–2379), a nuclear localization signal (NLS), AR1 (aa 2144–2185), AR2 (aa 2221–2272), and proline-rich region (PRR aa 2283–2327)[13].

This study will focus on NS5A region in HCV genotype 1a. NS5A region has 1344 base pairs, corresponding to 448 amino acids. The aim of this paper

• Yao Fu is currently working as a bioinformatics scientist in Bina Technologies, Roche, Redwood City, CA 94065, USA. This paper is a part of work when she was in Yale University, New Haven, CT 06520, USA. E-mail: yao.fu@yale.edu.
• Gang Chen currently works in the financial service industry. E-mail: g.chen@aya.yale.edu.
• Lizhi Fu is with Department of Mathematics & Statistics, Georgia State University, Atlanta, GA 30303, USA. E-mail: lfu1@student.gsu.edu.
• Jing Zhang is with the Department of Statistics, Yale University, New Haven, CT 06520, and Department of Mathematics & Statistics, Georgia State University, Atlanta, GA 30303, USA. E-mail: jzhang47@gsu.edu.
* To whom correspondence should be addressed.
Manuscript received: 2015-07-25; accepted: 2015-08-06

is to find possible amino acids that are correlated with drug resistance.

Zhang et al.[14] proposed an innovative method for investigating mutation interactions of HIV after certain drug treatment. This method has been used in detecting genome-wide associations on multiple diseases[15–29]. In this paper, we applied a Bayesian model equipped with Metropolis-Hastings algorithm on the data of the interferon-based drug treatment to select mutations associated with drug resistance. In addition, the Recursive Model Selection (RMS)[14] step is used to infer the dependence structure in the interacting mutation positions found by the Bayesian Variable Partition (BVP) model. This study confirms the findings of several previous studies and uncovers several interacting structures for these mutations.

2 Methods

2.1 BVP and RMS

Here, we first used the BVP model[14] to detect the mutation positions. After that, we applied RMS to infer the detailed dependence structure among the interacting positions.

2.2 Bayesian variable partition model

We will set up two data matrices $\boldsymbol{A} = [\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m]$ (dimension $n_A \times m$) and $\boldsymbol{B} = [\boldsymbol{B}_1, \cdots, \boldsymbol{B}_m]$ (dimension $n_B \times m$), where $\boldsymbol{A}$ and $\boldsymbol{B}$ denote the treated and untreated group, respectively (each row is one sequence, each column is one position of protein). Here n_A or n_B denotes the number of sequences in treated or untreated group and m denotes the number of positions of protein. For the distribution of the positions of these two groups, we propose the following four hypotheses:

- **H1 hypothesis**: The positions are independent of each other and treated group and untreated group have same probability distribution;
- **H2 hypothesis**: The positions are independent and treated group and untreated group have different probability distribution;
- **H3 hypothesis**: The positions are dependent and treated group and untreated group have same probability distribution;
- **H4 hypothesis**: The positions are dependent and treated group and untreated group have different probability distribution.

We are especially interested in positions in H2 and H4. Given that position i is from H2 hypothesis, assume that there are c_i possible values (amino acid) at position i. Suppose for every sequence in group $\boldsymbol{A}$, we have p_1 for the first value, p_2 for the second value, $\cdots$, p_{c_i} for the last value, and $\sum_{j=1}^{c_i} p_j = 1$. Then the likelihood for group $\boldsymbol{A}$'s data at position i is

$$P(\boldsymbol{A}_i | p_1, \cdots, p_{ci}, \mathrm{H2}) = \prod_{j=1}^{c_i} p_j^{n_j},$$

where n_j denotes the number of sequences who take the j-th value in $\boldsymbol{A}_i$. For every sequence in $\boldsymbol{B}$, we have p'_1 for the first value, p'_2 for the second value, $\cdots$, p'_{c_i} for the last value, and $\sum_{j=1}^{c_i} p'_j = 1$. Then the likelihood for group $\boldsymbol{B}$'s data at position i is

$$P(\boldsymbol{B}_i | p'_1, \cdots, p'_{c_i}, \mathrm{H2}) = \prod_{j=1}^{c_i} (p'_j)^{n'_j},$$

where n'_j denotes the number of individuals who take the j-th value in $\boldsymbol{B}_i$.

H2 means $p_j \neq p'_j$. However, we do not know these p_j and p'_j. So we assume they are random and use Dirichlet prior on them.

$$p \sim \mathrm{Dirichlet}(\alpha_1, \cdots, \alpha_{c_i}):$$

$$P(p_1, \cdots, p_{c_i} | \mathrm{H2}, \alpha_1, \cdots, \alpha_{c_i}) = \frac{1}{\boldsymbol{B}(\alpha)} \prod_{j=1}^{c_i} p_j^{\alpha_j - 1},$$

$$\text{where} \quad \boldsymbol{B}(\alpha) = \frac{\prod_{j=1}^{c_i} \Gamma(\alpha_j)}{\Gamma\left(\sum_{j=1}^{c_i} \alpha_j\right)}, \alpha = (\alpha_1, \cdots, \alpha_{c_i}),$$

$$\text{and} \quad \Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} \mathrm{d}t.$$

$$p' \sim \mathrm{Dirichlet}(\alpha'_1, \cdots, \alpha'_{c_i}):$$

$$P(p'_1, \cdots, p'_{c_i} | \mathrm{H2}, \alpha'_1, \cdots, \alpha'_{c_i}) = \frac{1}{\boldsymbol{B}(\alpha')} \prod_{j=1}^{c_i} (p'_j)^{\alpha'_j - 1},$$

$$\text{where} \quad \boldsymbol{B}(\alpha') = \frac{\prod_{j=1}^{c_i} \Gamma(\alpha'_j)}{\Gamma\left(\sum_{j=1}^{c_i} \alpha'_j\right)}, \alpha' = (\alpha'_1, \cdots, \alpha'_{c_i}),$$

$$\text{and} \quad \Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t} \mathrm{d}t.$$

So

$$P(\boldsymbol{A}_i, p_1, \cdots, p_{c_i} | \mathrm{H2}) =$$

$$\prod_{j=1}^{c_i} p_j^{n_j} \times \text{Dirichlet}(\alpha_1, \cdots, \alpha_{c_i}) =$$

$$\frac{1}{\boldsymbol{B}(\alpha)} \prod_{j=1}^{c_i} p_j^{n_j+\alpha_j-1};$$

$$P(\boldsymbol{B}_i, p'_1, \cdots, p'_{c_i}|\text{H2}) =$$

$$\prod_{j=1}^{c_i} (p'_j)^{n'_j} \times \text{Dirichlet}(\alpha'_1, \cdots, \alpha'_{c_i}) =$$

$$\frac{1}{\boldsymbol{B}(\alpha')} \prod_{j=1}^{c_i} (p'_j)^{n'_j+\alpha'_j-1}.$$

Integrating p and p', respectively, we have:

$$P(\boldsymbol{A}_i|\text{H2}) = \int_p P(\boldsymbol{A}_i, p_1, \cdots, p_{c_i}|\text{H2})\text{d}p =$$

$$\prod_{j=1}^{c_i} \frac{\Gamma(n_j+\alpha_j)}{\Gamma(\alpha_j)} \frac{\Gamma\left(\sum_{j=1}^{c_i} \alpha_j\right)}{\Gamma\left(\sum_{j=1}^{c_i} (n_j+\alpha_j)\right)},$$

$$P(\boldsymbol{B}_i|\text{H2}) =$$

$$\int_{p'} P(\boldsymbol{B}_i, p'_1, \cdots, p'_{c_i}|\text{H2})\text{d}p' =$$

$$\prod_{j=1}^{c_i} \frac{\Gamma(n'_j+\alpha'_j)}{\Gamma(\alpha'_j)} \frac{\Gamma\left(\sum_{j=1}^{c_i} \alpha'_j\right)}{\Gamma\left(\sum_{j=1}^{c_i} (n'_j+\alpha'_j)\right)}.$$

Then

$$P(\boldsymbol{A}_i, \boldsymbol{B}_i|\text{H2}) = P(\boldsymbol{A}_i|\text{H2}) \times P(\boldsymbol{B}_i|\text{H2}).$$

H1 means $p_j = p'_j$. So for H1 hypothesis, similarly, we have

$$P(\boldsymbol{A}_i, \boldsymbol{B}_i|\text{H1}) = \int_p P(\boldsymbol{A}_i, \boldsymbol{B}_i, p_1, \cdots, p_{c_i}|\text{H1})\text{d}p =$$

$$\int_p \frac{1}{\boldsymbol{B}(\alpha)} \prod_{j=1}^{c_i} p_j^{n_j+n'_j+\alpha_j-1} \text{d}p =$$

$$\prod_{j=1}^{c_i} \frac{\Gamma(n_j+n'_j+\alpha_j)}{\Gamma(\alpha_j)} \frac{\Gamma\left(\sum_{j=1}^{c_i} \alpha_j\right)}{\Gamma\left(\sum_{j=1}^{c_i} (n_j+n'_j+\alpha_j)\right)}.$$

For H4 hypothesis, assume that there are c possible value combinations of the dependent positions. Suppose for every sequence in group $\boldsymbol{A}$, we have p_1 for the first combination, p_2 for the second combination, $\cdots$, p_c for the last combination, and $\sum_{j=1}^{c} p_j = 1$. For everyone in group $\boldsymbol{B}$, we have p'_1 for the first combination, p'_2 for the second combination, $\cdots$, p'_c for the last combination, and $\sum_{j=1}^{c} p'_j = 1$. Then we can obtain:

$$P(\text{dependent positions in } \boldsymbol{A}|\text{H4}) =$$

$$\prod_{j=1}^{c} \frac{\Gamma(n_j+\alpha_j)}{\Gamma(\alpha_j)} \frac{\Gamma\left(\sum_{j=1}^{c} \alpha_j\right)}{\Gamma\left(\sum_{j=1}^{c} (n_j+\alpha_j)\right)},$$

$$P(\text{dependent positions in } \boldsymbol{B}|\text{H4}) =$$

$$\prod_{j=1}^{c} \frac{\Gamma(n'_j+\alpha'_j)}{\Gamma(\alpha'_j)} \frac{\Gamma\left(\sum_{j=1}^{c} \alpha'_j\right)}{\Gamma\left(\sum_{j=1}^{c} (n'_j+\alpha'_j)\right)},$$

where $n_j(n'_j)$ is the number of the j-th combination in $\boldsymbol{A}$ ($\boldsymbol{B}$), and

$$P(\text{dependent positions in } \boldsymbol{A}, \boldsymbol{B}|\text{H4}) =$$
$$P(\text{dependent positions in } \boldsymbol{A}) \times P(\text{dependent positions in } \boldsymbol{B}).$$

For H3 hypothesis, similarly, we have

$$P(\text{dependent positions in } \boldsymbol{A}, \boldsymbol{B}|\text{H3}) =$$

$$\prod_{j=1}^{c} \frac{\Gamma(n_j+n'_j+\alpha_j)}{\Gamma(\alpha_j)} \frac{\Gamma\left(\sum_{j=1}^{c} \alpha_j\right)}{\Gamma\left(\sum_{j=1}^{c} (n_j+n'_j+\alpha_j)\right)}.$$

We define an indicator vector $\boldsymbol{I} = [I_1, \cdots, I_m]$ to indicate the hypothesis of different positions belong to, where $I_i = 1$ means position i from H1, $I_i = 2$ means position i from H2, $I_i = 3$ means position i from H3, and $I_i = 4$ means position i from H4. Currently, we are interested in the posterior distribution of $\boldsymbol{I}$ given the data matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e., $P(\boldsymbol{I}|\boldsymbol{A}, \boldsymbol{B})$. According to the Bayes' theorem, we have

$$P(\boldsymbol{I}|\boldsymbol{A}, \boldsymbol{B}) = \frac{P(\boldsymbol{I})P(\boldsymbol{A}, \boldsymbol{B}|\boldsymbol{A})}{\sum_{\text{all possible } \boldsymbol{I}} P(\boldsymbol{I})P(\boldsymbol{A}, \boldsymbol{B}|\boldsymbol{I})}.$$

Therefore, we know that $P(\boldsymbol{I}|\boldsymbol{A},\boldsymbol{B}) \propto P(\boldsymbol{I})P(\boldsymbol{A},\boldsymbol{B}|\boldsymbol{I})$. Based on the four hypotheses, we have

$$P(\boldsymbol{A},\boldsymbol{B}|\boldsymbol{I}) = \prod_{I_i=1,2} P(\boldsymbol{A}_i,\boldsymbol{B}_i|I_i)\times$$
$$P(\text{dependent positions from H3})\times$$
$$P(\text{dependent positions from H4}).$$

In our study, we assume in prior most positions should be in H1 or H3 (i.e., unassociated with drug resistance),

$$P(I_i = 1) = P(I_i = 3) = 0.98,$$
$$P(I_i = 2) = P(I_i = 4) = 0.01,$$
$$P(\boldsymbol{I}) = \prod_{i=1}^{m} P(I_i).$$

2.3 Metropolis-Hastings algorithm

Then we designed a Metropolis-Hastings (M-H) algorithm (which is an algorithm for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution) to sample from the posterior distribution of $\boldsymbol{I}$. 10 independent Markov chains are stimulated with 20 000 iterations and the sampling starts from the 5000th iteration. The procedure of M-H is as follows:

(1) Initialize: randomly assign $\boldsymbol{I}$ a starting value $\boldsymbol{I}^{(t)}$ $(t = 0)$.

(2) Propose: randomly choose one $I_i^{(t)}$ and change it to other values with equal probabilities, the new $\boldsymbol{I}$ is $\boldsymbol{y}$.

(3) Evaluate: since the propose is symmetric, so the acceptance probability is $\alpha(\boldsymbol{I}^{(t)},\boldsymbol{y}) = \min\{1, P(\boldsymbol{I}=\boldsymbol{y}|\boldsymbol{A},\boldsymbol{B})/P(\boldsymbol{I}=\boldsymbol{I}^{(t)}|\boldsymbol{A},\boldsymbol{B})\}$.

(4) Move: generate u from Uniform(0,1) and set

$$\boldsymbol{I}^{(t+1)} = \begin{cases} \boldsymbol{y}, & \text{if } u \leqslant \alpha(\boldsymbol{I}^{(t)},\boldsymbol{y}); \\ \boldsymbol{I}^{(t)}, & \text{otherwise.} \end{cases}$$

(5) If $t \geqslant N$, stop; otherwise set $t = t + 1$ and go to Step (2) (Here N is the total number of iterations).

We applied BVP and RMS sequentially to the data of treated and untreated patients. For the treated and untreated comparison, we applied BVP to all treated datasets (67 sequences) versus untreated dataset (457 sequences), and devised the difference between these two, realizing that there exist different treated outcomes that we should account for differently.

3 Results

The structure of NS5A region is constituted by the membrane attachment region (aa 1–236), the carboxyl region (aa 237–448). The carboxyl region consists of PKRbd (aa 237–302), variable region 4 (V4; aa 310–330), variable region 3 (V3; aa 381–409), the region between V3 and V4 (aa 331–380), and V3 downstream region (aa 410–448).

Our analysis is based on the four hypotheses we mentioned above. Hypotheses 2 and 4 are most probably for most of the mutating position and our findings indeed verified this observation.

In order to study the mutation pattern differences between treated and untreated individuals, we used BVP and RMS to analyze the sequences from treated (67 sequences) and untreated (457 sequences) patients. Table 1 shows the results detailing the positions that have high enough probabilities (95% or more) for us to make inference regarding our hypotheses. As we can see from Table 1, the results from different Markov chains seem to vary, possibly because of the noise from local modes, where chains stuck. An interesting observation is that among positions from about 380 to 400, there are no differences between treated and untreated patients in this region. A sample Markov chain from the 20 chains we generated with different hypotheses distributions is available in Fig. 1. This region is corresponding to V3 in NS5A. From this information, we could conclude

Table 1 Positions whose posterior probabilities of H2 or H4 are larger than 0.95.

Chain	H2 (P>0.95)	H4 ($P > 0.95$)
1	null	56
2	null	343, 404
3	null	14, 101, 166, 313
4	null	160
5	null	null
6	null	160, 164
7	null	14, 24, 101, 166, 203, 313, 337
8	null	68, 161, 284, 444
9	null	68, 161, 444
10	null	78, 135, 296
11	null	null
12	null	78, 135, 296
13	null	227, 270, 279
14	null	175, 222, 270, 284, 381
15	null	14, 62, 166, 236, 313
16	null	23, 101, 131, 249, 295, 313, 385
17	null	260
18	null	14, 68, 101, 161, 166, 249, 257, 313, 384, 385
19	null	30, 56
20	null	270, 279, 284, 381

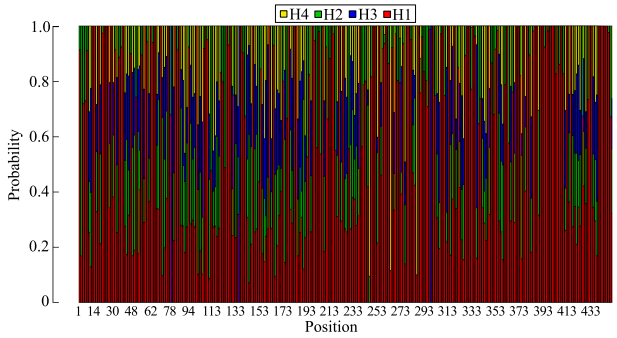

Fig. 1 Hypotheses distribution (H1, H2, H3, and H4) in a sample Markov Chain.

that IFN plus ribavirin therapy does not cause sequence changes in V3.

Aside from the interesting finding above, Fig. 2 and Table 2 show BVP-detected interacting positions and RMS-inferred dependence structure. In positions 78, 135, and 296, we found that amino acid combination KAI, which consisted 5.8% in the untreated sample, disappeared in the treated sample. Similar decreases were seen in some other combinations: RSI, KTL, and RAV. It seems that interferon-based treatment is very useful for the amino acids combinations mentioned above. It is also worth noting that some significant increase was observed. The combination KSV almost tripled its part from 0.6% to 1.5% after treatment, the combination KTI increased to 16.7% in the treated samples compared to 13.2% in untreated ones, and the combination RTV increased from 2.1% to 4.5%. Results from positions 343 and 404 exhibit similar trends with some combinations decrease in frequency significantly while some increase in frequency after treatment.

We also looked at single positions significant under Fisher test and noticed that even single positions revealed treatment effects in terms of amino acid frequency. At position 311, the ratio of amino acid S

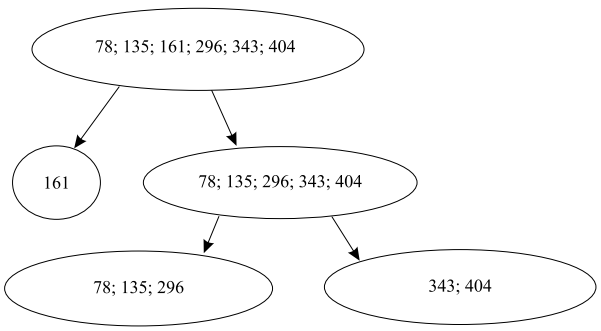

Fig. 2 Flowchart of detected mutation positions and position combinations.

Table 2 Detailed position interaction relations for positions 78, 135, 296, 343, 404, 311, and 358.

Amino acids combinations (in order of 78, 135, 196)	Percentage in untreated sequences (%)	Percentage in treated sequences (%)
KAV	19.9	22.7
KTI	13.2	16.7
KTV	6.1	4.5
RTI	48.5	47.0
RTV	2.1	4.5
RAI	2.1	3.0
KAI	5.8	0.0
RSI	0.3	0.0
KTL	0.3	0.0
RAV	0.6	0.0
RNI	0.3	0.0
KSV	0.6	1.5
Positions 343, 404		
Amino acids combinations (in order of 343, 404)	Percentage in untreated sequences (%)	Percentage in treated sequences (%)
PC	83.4	80.3
AP	0.3	0.0
PG	0.3	0.0
AH	3.4	3.0
PR	6.1	7.6
PS	2.1	3.0
PP	0.6	1.5
AR	1.2	1.5
AC	0.9	1.5
PH	1.2	0.0
PL	0.3	0.0
PQ	0.0	1.5
Position 311 (Fisher test significant)		
Amino acids	Percentage in untreated sequences (%)	Percentage in treated sequences (%)
P	61.3	66.7
Q	31.3	21.2
S	2.8	1.5
R	1.8	9.1
H	1.8	0.0
L	0.6	0.0
K	0.3	1.5
Position 358 (Fisher test significant)		
Amino acids	Percentage in untreated sequences (%)	Percentage in treated sequences (%)
K	99.4	0.6
R	95.5	4.5

decreased from 31.3% to 21.2% after treatment, and so did amino acids H and L, both disappeared after treatment. At position 358, the ratio of amino acid R increased almost three folds from 0.6% to 4.5% after treatment.

4 Discussion

Utilizing the innovative method proposed by Zhang et al.[14], which connected Bayesian statistical modeling with molecular dynamic simulations, we were able to investigate the complex interactions of drug resistance mutations in HCV proteases and reverse transcriptase.

This paper, however, is admittedly a baseline analysis of the treatment effects that ignores many other factors that possibly affect HCV virus mutations. For instance, different pre-treatment patterns are observed in patients data, which may explain some of the inconsistency exhibited in the result. Another issue lies in the differences of treatment effects, which may also affect the genome of sequence of HCV virus.

Despite all other possibilities may occur, this study has demonstrated the original findings of HCV drug resistance and showed a long puzzled selection pattern of HCV drug treatment effects. All the methods and results here will enlighten new and more accurate ways to uncover some new finds behind HCV drug resistance and other related diseases.

Acknowledgements

Dr. Jing Zhang was supported by start-up funding and Sesseel Award from Yale University. The computation was done with the help from the Yale University Biomedical High Performance Computing Center, which was supported by the NIH grant RR19895.

References

[1] T. Poynard, V. Leroy, M. Cohard, T. Thevenot, P. Mathurin, P. Opolon, and J. P. Zarski, Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: Effects of dose and duration, *Hepatology,* vol. 24, no. 4, pp. 778–789, 1996.

[2] G. L. Davis, R. Esteban-Mur, V. Rustgi, J. Hoefs, S. C. Gordon, C. Trepo, M. L. Shiffman, S. Zeuzem, A. Craxi, M. H. Ling, et al., Interferon alfa-2b alone or in combination with ribavirin for the treatment of relapse of chronic hepatitis C. International Hepatitis Interventional Therapy Group, *N. Engl. J. Med.,* vol. 339, no. 21, pp. 1493–1499, 1998.

[3] J. G. McHutchison, S. C. Gordon, E. R. Schiff, M. L. Shiffman, W. M. Lee, V. K. Rustgi, Z. D. Goodman, M. H. Ling, S. Cort, and J. K. Albrecht, Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group, *N. Engl. J. Med.,* vol. 339, no. 21, pp. 1485–1492, 1998.

[4] N. Enomoto, I. Sakuma, Y. Asahina, M. Kurosaki, T. Murakami, C. Yamamoto, Y. Ogura, N. Izumi, F. Marumo, and C. Sato, Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection, *N. Engl. J. Med.,* vol. 334, no. 2, pp. 77–81, 1996.

[5] A. Macdonald and M. Harris, Hepatitis C virus NS5A: Tales of a promiscuous protein, *J. Gen. Virol.,* vol. 85, no. Pt 9, pp. 2485–2502, 2004.

[6] M. J. Gale, Jr., M. J. Korth, N. M. Tang, S. L. Tan, D. A. Hopkins, T. E. Dever, S. J. Polyak, D. R. Gretch, and M. G. Katze, Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein, *Virology,* vol. 230, no. 2, pp. 217–227, 1997.

[7] M. Jr Gale, C. M. Blakely, B. Kwieciszewski, S. L. Tan, M. Dossett, N. M. Tang, M. J. Korth, S. J. Polyak, D. R. Gretch, and M. G. Katze, Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: Molecular mechanisms of kinase regulation, *Mol. Cell Biol.,* vol. 18, no. 9, pp. 5208–5218, 1998.

[8] N. Enomoto and C. Sato, Clinical relevance of hepatitis C virus quasispecies, *J. Viral. Hepat.,* vol. 2, no. 6, pp. 267–272, 1995.

[9] N. Enomoto, I. Sakuma, Y. Asahina, M. Kurosaki, T. Murakami, C. Yamamoto, N. Izumi, F. Marumo, and C. Sato, Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region, *J. Clin. Invest.,* vol. 96, no. 1, pp. 224–230, 1995.

[10] M. Gerotto, F. Dal Pero, D. G. Sullivan, L. Chemello, L. Cavalletto, S. J. Polyak, P. Pontisso, D. R. Gretch, and A. Alberti, Evidence for sequence selection within the non-structural 5A gene of hepatitis C virus type 1b during unsuccessful treatment with interferon-alpha, *J. Viral. Hepat.,* vol. 6, no. 5, pp. 367–372, 1999.

[11] M. Fukuda, K. Chayama, A. Tsubota, M. Kobayashi, M. Hashimoto, Y. Miyano, H. Koike, M. Kobayashi, I. Koida, Y. Arase, et al., Predictive factors in eradicating hepatitis C virus using a relatively small dose of interferon, *J. Gastroenterol Hepatol,* vol. 13, no. 4, pp. 412–418, 1998.

[12] M. J. Clemens and A. Elia, The double-stranded RNA-dependent protein kinase PKR: Structure and function, *J. Interferon Cytokine Res.,* vol. 17, no. 9, pp. 503–524, 1997.

[13] K. M. Chung, O. K. Song, and S. K. Jang, Hepatitis C virus nonstructural protein 5A contains potential transcriptional activator domains, *Mol. Cells,* vol. 7, no. 5, pp. 661–667, 1997.

[14] J. Zhang, T. Hou, W. Wang, and J. S. Liu, Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance, *Proc. Natl. Acad. Sci. USA,* vol. 107, no. 4, pp. 1321–1326, 2010.

[15] V. Svicher, C. Alteri, A. Artese, J. M. Zhang, G. Costa, F. Mercurio, R. D'Arrigo, S. Alcaro, G. Palu, M. Clementi, et al., Identification and structural characterization of novel genetic elements in the HIV-1 V3 loop regulating coreceptor usage, *Antivir. Ther.,* vol. 16, no. 7, pp. 1035–1045, 2011.

[16] B. Y. Zhang, J. Zhang, and J. S. Liu, Block-based Bayesian epistasis association mapping with application to Wtccc Type 1 diabetes data, *Ann. Appl. Stat.,* vol. 5, no. 3, pp. 2052–2077, 2011.

[17] V. Svicher, V. Cento, M. Bernassola, M. Neumann-Fraune, F. Van Hemert, M. Chen, R. Salpini, C. Liu, R. Longo, M. Visca, et al., Novel HBsAg markers tightly correlate with occult HBV infection and strongly affect HBsAg detection, *Antiviral Res.,* vol. 93, no. 1, pp. 86–93, 2012.

[18] M. Chen, V. Svicher, A. Artese, G. Costa, C. Alteri, F. Ortuso, L. Parrotta, Y. Liu, C. Liu, C. F. Perno, et al., Detecting and understanding genetic and structural features in HIV-1 B subtype V3 underlying HIV-1 co-receptor usage, *Bioinformatics,* vol. 29, no. 4, pp. 451–460, 2013.

[19] C. Mirabelli, M. Surdo, F. Van Hemert, Z. Lian, R. Salpini, V. Cento, M. F. Cortese, M. Aragri, M. Pollicita, C. Alteri, et al., Specific mutations in the C-terminus domain of HBV surface antigen significantly correlate with low level of serum HBV-DNA in patients with chronic HBV infection, *J. Infect.,* vol. 70, no. 3, pp. 288–298, 2015.

[20] I. Kozyryev and J. Zhang, Bayesian analysis of complex interacting mutations in HIV drug resistance and cross-resistance, *Adv. Exp. Med. Biol.,* vol. 827, pp. 367–383, 2015.

[21] Y. Fu, G. Chen, X. Guo, Y. Pan, and J. Zhang, Analyzing the effects of pretreatment diversity on HCV drug treatment responsiveness using Bayesian partition methods, *J. Bioinfo. Proteomics. Rev*., vol. 1, no. 1, pp. 1–6, 2015

[22] X. Guo, J. Zhang, Z. Cai, D. Du, and Y. Pan, DAM: A Bayesian method for detecting genome-wide associations on multiple diseases, *Lecture Notes in Computer Science,* vol. 9096, pp 96–107, 2015

[23] M. Surdo, M. F. Cortese, C. Mirabelli, R. Salpini, J. Zhang, F. Van Hemert, V. Cento, M. Pollicita, G. Gubertini, G. M. De Sanctis, et al., P691 key patterns of HBsAg mutations correlate with mechanisms underlying levels of serum HBVDNA, *Journal of Hepatology*, 2014, dio:10.1016/S0168.

[24] Z. C. Lian, Q. Tian, Y. Liu, and J. Zhang, Detecting hepatitis B viral amino acid sequence mutations in cccult hepatitis B infections via Bayesian partition model, *J. Proteomics Bioinform*., 2013. doi: 10.4172/jpb.S6-005.

[25] I. Kozyryev and J. Zhang, Bayesian exploration of multilocus interactions on the genome-wide scale, *American Journal of Bioinformatics*, vol. 1, no. 1, pp. 70–78, 2012. doi: 10.3844/ ajbsp.2012.70.78.

[26] J. Zhang, Z. Wu, C. Gao, and M. Zhang, High-order interactions in rheumatoid arthritis detected by Bayesian method using genome-wide association studies data, *American Medical Journal*, vol. 3, no. 1, pp.56–66, 2012. doi: 10.3844/amjsp.2012.56.66.

[27] J. Zhang, T. Hou, Y. Liu, G. Chen, X.Yang, J. Liu and W. Wang, Systematic investigation on interactions for HIV drug resistance and cross-resistance among protease inhibitors, *Journal of Proteome Science & Computational Biology*, 2012, doi: http://dx.doi.org/10.7243/2050-2273-1-2.

[28] I. Kozyryev and J. Zhang, Bayesian determination of disease associated differences in haplotype blocks, *Am. J. Bioinform.*, vol. 1, no. 1, pp. 20–29, 2012.

[29] J. Zhang, Q. Zhang, D. Lewis, and M. Q. Zhang, A Bayesian method for disentangling dependent structure of epistatic interaction, *Am. J. Biostat.*, vol. 2, no. 1, pp. 1–10, 2011.

Yao Fu got her PhD degree in computational biology and bioinformatics from Yale University in 2015. She is now working as a bioinformatics scientist in Roche. Her research interests are bioinformatics including next generation sequencing data processing and analysis, functional interpretation of human genetic variations, network topologies, statistical genetics, and cancer genomics.

Gang Chen graduated from Yale College with a BA in economics and mathematics in 2013. He is interested in Bayesian statistics and computational biology. He currently works in the financial service industry.

Jing Zhang got her PhD degree in statistics from Harvard University in 2009. She was an assistant professor in Department of Statistics, Yale University from 07/2010-06/2014. After that, she moved to Georgia State University as an assistant professor till now. Her research interest is bioinformatics. She is professional in Bayesian Analysis.

Lizhi Fu got his master degree from Zhejiang University in 2012. He is currently a graduate student in Dept. of Biology and Dept. of Mathematics and Statistics, Georgia State University. His research interests are bioinformatics, preclinical drug screen, and clinical trial phase I – III drug evaluation.