Analyzing the Effects of Pretreatment Diversity on HCV Drug Treatment Responsiveness Using Bayesian Partition methods

Yao Fu[2], **Gang Chen**[3], **Xuan Guo**[4], **Jing Zhang**[1,2,3,*], and **Yi Pan**[4,*]

[1]Department of Mathematics and Statistics, Georgia State University

[2]Program of Computational Biology and Bioinformatics, Yale University

[3]Department of Statistics, Yale University

[4]Department of Computer Science, Georgia State University, Atlanta, Georgia, USA

Abstract

Traditional therapies for Hepatitis C Virus (HCV) often yield unsatisfactory results. The reason for this may lie in the mechanism of drug resistance of the HCV virus. Despite doing a plain vanilla comparison between the treated and untreated groups, this paper takes a detour and investigates the drug resistance mechanism to interferon plus ribavirin combined therapy by comparing pretreatment sequence data between response and non-response patients in the NS5A region for genotype 1a HCV virus. We use Bayesian probabilistic models to detect single mutation or mutation combinations, and infer interaction structures between these mutations, to investigate the drug resistance combinations differences between those patients. We hope to decipher, at least partially, the reason behind the unsatisfactory results received from interferon plus ribavirin therapy.

Author Summary—HCV treatment results have been historically suboptimal[1–3]. HCV drug resistance, which further hinders the treatment effects, is caused by mutations of viral proteins that disrupt the drugs' binding but do not affect the viral survival. Due to the high rate and low fidelity of HCV replication, resistant strains quickly become dominant in a viral population under the selection pressure of a drug. M.J. Donlin et al indicate that pretreatment sequence diversity correlates with response effects[15]. We incorporate this idea and use a Bayesian approach to look into the pretreatment sequences diversity of HCV virus between response and non-response groups, under a combined treatment of interferon and ribavirin.

Introduction

As a single-strand RNA virus, Hepatitis C virus (HCV) has been categorized into at least six genotypes each with several subtypes. Spreading in different regions, the genotypes will have dissimilar response patterns to interferon-based therapy. In clinic, less than 20% of chronic patients show sustained response with interferon monotherapy treatment[1], while the therapy with the IFN and ribavirin combination showed significant improvement in

*Corresponding author: Jing Zhang, Program of Computational, Biology and Bioinformatics, Yale University, USA. jzhang47@gsu.edu, Yi Pan, Department of Computer, Science, Georgia State University, Atlanta, Georgia, USA. yipan@gsu.edu.

response rates[2,3]. An accurate interpretation of the mechanism behind the antiviral resistance to IFN therapy will be a key factor for developing better treatment strategies.

Many studies have found that variations in HCV sequences play a role in response to IFN-based therapies, especially for the variations in the NS5A region[4,5]. NS5A is a nonstructural protein which leads to the resistance of IFN treatment by blocking the function of an important mediator in IFN response, dsRNA dependent protein kinase (PKR)[6,7]. In 1995, an "interferon sensitivity determining region" (ISDR amino acid [aa] 2209–2248) defined by Enomoto et al. in NS5A, is enriched with mutations related to resistance to IFN[8]. This finding has been confirmed by several studies[9–11]. Moreover, some other study showed evidence that in PKR binding domain (PKRbd aa 2209–274) of NS5A mutations can hamper the viral replication[12]. However, conflicting results about these two regions are also found, which complicate the role of NS5A in response to IFN. In addition to ISDR and PKRbd, other domains in NS5A have also been concerned in resistance to virus, such as proline-rich region (PRR aa 2283–2327), AR1 (aa 2144–2185) and AR2 (aa 2221–2272) of nuclear localization signal (NLS), and the variable region 3 (V3 aa 2356–2379) in the C terminus[13].

This study will concentrate on NS5A region particular for HCV genotype 1a. In NS5A region there are 1344 base pairs, linking to 448 amino acids. The goal of this paper is to compare the pretreatment sequence patterns between those patients who respond positively to the treatment and those who don't, and then infer the possible mutation positions that may affect the treatment effects.

To better facilitate the discussion of our findings, we first introduce the inner structure of NS5A region. It constitutes the following regions: the membrane attachment region (aa 1–236); the carboxyl region (aa 237–448); and the regions within the carboxyl end, such as PKRbd (aa 237–302), variable region 4 (V4; aa 310–330), variable region 3 (V3; aa 381–409), the region between V4 and V3 (aa 331–380), and the downstream region of V3 (aa 410–448).

In this study, we employed Bayesian models, which are originally proposed by Zhang et al. for investigating mutation interactions of HIV caused by a certain drug treatment[14]. Based on the Bayesian variable partition (BVP) model, we first used Metropolis-Hastings algorithm on the data of the interferon treatment to sort out mutations associated with drug resistance, and then applied a recursive model selection (RMS) procedure on the selected mutation positions to infer the dependence structure with the interacting effects.

Methods

Here we first employed the Bayesian variable partition (BVP) model[15] to search for the mutation positions. After detecting the interaction mutation positions, we further applied the Recursive Model Selection[16] (RMS) on selected mutations by BVP to infer more detailed dependence structure among the interacting positions.

Bayesian Variable Partition Model

The response group and non-response group can be represented as two data matrices $A=[A_1, \ldots, A_m]$ (of dimension $N_A \times m$) and $B=[B_1, \ldots, B_m]$ (of dimension $N_B \times m$), respectively (each row is a sequence, each column is a position of protein HCV RT). Here N_A or N_A denotes the number of sequences in response or non-response group respectively, and m denotes the number of positions. The m positions can be partitioned into four sets: set 1 contains positions independent with each other sharing the same distribution in response and non-response groups; set 2 contains positions independent with each other but with different distributions in two groups; set 3 contains positions dependent with the same distribution in two groups; set 4 contains positions dependent with different distributions in two groups. These four sets are corresponding to the four hypotheses in the result section. Let $I=(I_1, \ldots, I_m)$ indicate the membership of the positions with $I_j=1,2,3$ and 4, respectively, and $A^{(1)}$ and $B^{(1)}$ denote the sequences in l^{th} set from two groups. Our goal is to infer the sets of positions with different distributions in two groups (that is $I_j = 2$ or 4).

Assume that there are c_j possible values (amino acid types) at position j, and let $\Theta_1=\{(\theta_{j1},\theta_{j2},\ldots,\theta_{jc_j}):I_j=1\}$ be the amino acid frequencies of each position in set 1 in both groups, thus, the likelihood of $(A^{(1)}, B^{(1)})$ is

$$P(A^{(1)}, B^{(1)}|\Theta_1)=\Pi_{j:I=1}\Pi_{k=1}^{c_j}\theta_{jk}^{n_{jk}} \quad (1)$$

where $\{n_{j1},\ldots,n_{jc_j}\}$ are number of sequences taking k^{th} value in $(A^{(1)}, B^{(1)})$. Assume a Dirichlet prior on Θ_1, that is, $\Theta_1 \sim Dirichlet(\alpha)$ where $\alpha(\alpha_1,\ldots, \alpha_{c_j})$ By integrating out Θ_1, we have the marginal probability:

$$P(A^{(1)}, B^{(1)}|I)=\Pi_{j:I=1}\left(\Pi_{k=1}^{c_j}\frac{\Gamma(n_{jk}+\alpha_k)}{\Gamma(\alpha_k)}\right)\frac{\Gamma(|\alpha|)}{\Gamma(N+|\alpha|)} \quad (2)$$

where $|\alpha|$ is the sum of all elements in α, and $N=N_A+N_B$.

Different to positions in set 1, two priors, Dirichlet (β^A) and Dirichlet (β^B), are used on the amino acid frequencies of each positions in set 2 in group A and B, respectively. By integrating out frequencies, we obtain

$$P(A^{(2)}|I)=\Pi_{j:1=2}\left(\Pi_{k=1}^{c_j}\frac{\Gamma(n_{jk}+\beta_k^A)}{\Gamma(\beta_k^A)}\right)\frac{\Gamma(|\beta^A|)}{\Gamma(N_A+|\beta^A|)} \quad (3)$$

$$P(B^{(2)}|I)=\Pi_{j:1=2}\left(\Pi_{k=1}^{c_j}\frac{\Gamma(n_{jk}+\beta_k^B)}{\Gamma(\beta_k^B)}\right)\frac{\Gamma(|\beta^B|)}{\Gamma(N_B+|\beta^B|)} \quad (4)$$

Positions in set 3 and 4 influence the resistance statuses through interactions. Thus, each amino acid combination over set 3 or 4 represents a potential mutation. Assume there are $c^{(3)}$ and $c^{(4)}$ possible value combinations for set 3 and 4, respectively, we use $Dirichlet(\gamma)$ prior on the combination frequencies in set 3 and use two priors, $Dirichlet$ (δ^A) and

Dirichlet(δ^B), on the combination frequencies in set 4 for response group and non-response group. By integrating out frequencies, we obtain

$$P(A^{(3)}, B^{(3)}|I) = \Pi_{j:I=3} \left(\Pi_{k=1}^{c^{(3)}} \frac{\Gamma(n_{jk}+\gamma_k)}{\Gamma(\gamma_k)} \right) \frac{\Gamma(|\gamma|)}{\Gamma(N+|\gamma|)} \quad (5)$$

$$P(A^{(4)}|I) = \Pi_{j:I=4} \left(\Pi_{k=1}^{c^{(4)}} \frac{\Gamma(n_{jk}+\delta_k^A)}{\Gamma(\delta_k^A)} \right) \frac{\Gamma(|\delta^A|)}{\Gamma(N_A+|\delta^A|)} \quad (6)$$

$$P(B^{(4)}|I) = \Pi_{j:I=4} \left(\Pi_{k=1}^{c^{(4)}} \frac{\Gamma(n_{jk}+\delta_k^B)}{\Gamma(\delta_k^B)} \right) \frac{\Gamma(|\delta^B|)}{\Gamma(N_B+|\delta^B|)} \quad (7)$$

Combining formulas from (1) to (7), we have the posterior distribution of I as

$$P(I|A,B) a P(A^{(1)},B^{(1)}|I)P(A^{(2)}|I)P(B^{(2)}|I)P(A^{(3)},B^{(3)}|I)P(A^{(4)}|I)P(B^{(4)}|I)P(I) \quad (8)$$

In this study, we assume most positions should be in set 1 or set 3 in prior (i.e. unassociated with drug resistance), $P(I_i=2)=P(I_i=4)=0.01$, and $P(I)=\Pi_{i=1}^{m}P(I_i)$. We further set the parameters for all Dirichlet priors to 0.5. A Markov chain Monte Carlo (MCMC) algorithm[17] can be designed to sample from this posterior distribution so as to infer which variables are associated with the treatment status. More details on BVP can be found in[18].

Recursive Model Selection

We applied the RMS procedure to infer the detailed dependence structure among the interacting positions generated by BVP. Our strategy is to recursively apply a model selection of two classes of cruder models, that is the chain-dependence model and the V-dependence model, until the data do not support more detailed models.

We say that a group of variables X_G follow a chain-dependence model if the index set G can be partitioned into three subsets U, V, and W such that X_U and X_W are independent given X_V, such as $X_U \rightarrow X_V \rightarrow X_W$. The joint distribution of a chain-dependence model is

$$P(X_G)=P(X_U)P(X_V|X_U)P(X_W|X_V) \quad (9)$$

We say that a group of variables X_G follow a V-dependence model if X_U and X_W are mutually independent, that is $X_U \rightarrow X_W \leftarrow X_V$. The joint distribution of a V-dependence model is

$$P(X_G)=P(X_U)P(X_W)P(X_V|X_U,X_W) \quad (10)$$

In these two models, only set W is allowed to be empty, in which case these models become the saturated model.

We use a model indicator $I^{CV} = (I_1^{CV}, I_2^{CV}, \ldots, I_L^{CV})$ to imply the membership of the L positions with $I_j^{cv} = 0$ for the chain-dependence model and $I_j^{cv} = 1$ for the V-dependence model. Let $\mathfrak{P}$ denote the set partition, the posterior distribution of I^{CV} and $\mathfrak{P}$ is

$$\mathbf{P}(\mathfrak{P}, \boldsymbol{I^{CV}} | \boldsymbol{Data}) \propto \boldsymbol{P}(\boldsymbol{Data} | \mathfrak{P}, \boldsymbol{I^{CV}}) \boldsymbol{P}(\mathfrak{P}) \boldsymbol{P}(\boldsymbol{I^{CV}}) \quad (11)$$

We set equal prior probability for I^{CV} and $\mathfrak{P}$ is An MCMC algorithm is designed to simulate from (11) and to find the optimal model type and variable partition. More details on RMS can be found in[13]. The procedure is applied recursively until only single-variable nodes are available.

BVP model and RMS procedure were utilized sequentially to the data of response and non-response patients. For the comparison, we applied BVP to 47 response datasets versus 29 non-response dataset, and contrived the difference between these two, recognizing that there exist different pretreatment patterns that we should account for differently. The detailed accession numbers for total 76 sequences are showed in Table 1, 2. More information about these sequences can be accessed in the reference[19–20].

Results

As M.J. Donlin et al. indicated, pretreatment sequence diversity correlates with response effects[15]. On observing this finding, our analysis is to test the following four proposed hypotheses. Hypothesis 1: the positions are independent with each other, and the probability distribution of the pretreatment sequences of response and non-response groups is the same; Hypothesis 2: the positions are independent with each other, and the pretreatment sequences of response and non-response groups have different probability distributions; Hypothesis 3: the positions are dependent, and the distribution of the pretreatment sequences of response and non-response groups is the same; Hypothesis 4: the positions are dependent, and the pretreatment sequences of response and non-response groups have different probability distributions.

We applied Bayesian Variable Partition (BVP) model and Recursive Model Selection (RMS) procedure to the pretreatment sequences of response (47 sequences) and non-response (29 sequences) samples, as described in detail in the methods section. We run the proposed method with multiple random restarts, and multiple chains will converge to different multiple local modes. So given different results, we consider all of them meaningful. We did not do any heuristic or subjective selection. Table 3 shows the results of the analysis, namely, positions which have 95% or more probability for us to infer one of our four hypotheses. As shown in Table 3, the results are not uniform from different Markov Chains. The limited sample size of our analysis may be the reason for this inconsistency. However, looking at the common positions from almost all the 20 chains still gives us a reliable idea of the mutation mechanism of positions 49, 349, and 199, 209, 242, 398 which have the highest frequencies among these 20 Markov chains. Positions 49 and 349 are statistically different in response and non-response patients and are independent of other positions. Positions 199, 209, 242, 398 are dependent and demonstrate significant difference

in response and non-response patients. Position 49 is in membrane attachment region; Position 349 is in the region between V3 and V4; Positions 199 and 209 are in membrane attachment region; Position 242 is in ISDR region; Position 398 is in V3 region. These positions may have some biological influence on drug resistance to IFN and ribivirin.

While analyzing single positions as above is helpful, a lot of positions are not mutating independently. [Figure 3] shows the interacting positions detected by BVP in response samples and [Figure 4] shows the interacting positions detected by BVP in non-response samples. Table 4 and Table 5 show the dependence structure inferred by RMS in detail, for the response and non-response group respectively. At position 285, we found that the frequency of amino acid E is 13.8% in non-response samples and 8.5% in the response samples. A more significant result was found at position 199, where the frequency of amino acid L decrease from 100% to 87.2%, from non-response samples to response samples. Similar yet less significant patterns were found at position 226, where the amino acid M decrease from 20.75 to 14.9%, from non-response samples to response samples. For dependent positions, we observed similar results, as shown jointly in Table 4 and Table 5. At positions 107, 226, 288, 410, 439, the amino acid combination EMIAE does not exist in response samples, which indicates that those positions combined may be a distinguishing factor for response and non-response patients. There are other non-existent combinations at those positions. For instance, KEIAG, TMVAG, TLIAE, are all combinations that only exist in non-response samples.

To show the prediction power of the reported mutations by our method, an SVM classifier was used to conduct a classification on the non-response and response sequences. The ROC curves and corresponding AUC values were showed in Figure 5. The SVM was implemented by libsvm[16], and the hyper- parameters were tuned to be optimal by the grid search. The kernel function applied was Radial basis function, which gave best results comparing to linear, polynomial, and sigmoid kernel functions. We chose four mutation positions combinations to illustrate the prediction power of our findings: (A) 44, 133, 226, 269, 276, 285, 288, 296, 304, 305, 382 (B) 24, 44, 107, 133, 135, 226, 304, 305, 422 (C) 23, 44, 61, 64, 71, 135, 245, 269, 276, 285, 291, 382, 388, 439, 445; (D) 107, 226, 388, 410, 439. From the ROC curves and AUC values, these mutation positions demonstrate considerable discrimination and prediction power with even small samples size (47 response sequences versus 29 non-response sequences).

Single positions significant under Fisher test also reveal differences between response and non-response samples in terms of the frequency of amino acid. At position 48, the frequency of amino acid R is 100% in non-response samples, while only 70.2% in response samples. At position 81, the frequency of amino acid R is about 15% higher in response samples. These results, combined with the more reliable evidence from Table 4 and Table 5, gave us a relatively complete picture of the differences between non-response and response samples.

Discussion

Utilizing the pioneering method proposed by Zhang et al[14], which employs Bayesian statistical modeling, we were able to detect and analyze, the complex interactions of mutations of the HCV protease and reverse transcriptase.

While this analysis helps present a relatively comprehensive picture of the different pretreatment structures of non-response and response patients, it admittedly omits many other factors that possibly influence HCV virus mutations. Despite all the possibilities that may emerge, this study has not only confirmed the original findings of HCV drug resistance but also demonstrated the long-puzzled selection pattern of HCV drug treatment effects. We are positive that the method and results presented here will make a stimulation of new and more accurate ways to decipher the myths behind drug resistance of HCV and other related diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

References

1. Poynard T, Leroy V, Cohard M, et al. Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: Effects of dose and duration. Hepatology. 1996; 24(4):778–789. [PubMed: 8855176]

2. Davis GL, Esteban-Mur R, Rustgi, et al. Interferon alfa-2b alone or in combination with ribavirin for the treatment of relapse of chronic hepatitis C. N Engl J Med. 1998; 339(21):1493–1499. [PubMed: 9819447]

3. McHutchison JG, Gordon SC, Schiff ER, et al. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. N Engl J Med. 1998; 339(21):1485–1492. [PubMed: 9819446]

4. Enomoto N, Sakuma I, Asahina Y, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med. 1996; 334(2):77–81. [PubMed: 8531962]

5. Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. J Gen Virol. 2004; 85(pt 9):2485–2502. [PubMed: 15302943]

6. Gale MJ, Korth MJ, Tang NM, et al. Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. Virology. 1997; 230(2):217–227. [PubMed: 9143277]

7. Gale M, Blakely CM, Kwieciszewski B, et al. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. Mol Cell Biol. 1998; 18(9): 5208–5218. [PubMed: 9710605]

8. Enomoto N, Sato C. Clinical relevance of hepatitis C virus quasispecies. J Viral Hepat. 1995; 2(6): 267–272. [PubMed: 8732171]

9. Enomoto N, Sakuma I, Asahina Y, et al. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J Clin Invest. 1995; 96(1):224–230. [PubMed: 7542279]

10. Gerotto M, Dal Pero F, Sullivan DG, et al. Evidence for sequence selection within the non-structural 5A gene of hepatitis C virus type 1b during unsuccessful treatment with interferon-α. J Viral Hepat. 1999; 6(5):367–372. [PubMed: 10607252]

11. Fukuda M, Chayama K, Tsubota A, et al. Predictive factors in eradicating hepatitis C virus using a relatively small dose of interferon. J Gastroenterol Hepatol. 1998; 13(4):412–418. [PubMed: 9641307]

12. Clemens MJ, Elia A. The double-stranded RNA-dependent protein kinase PKR: structure and function. J Interferon Cytokine Res. 1997; 17(9):503–524. [PubMed: 9335428]

13. Chung KM, Song OK, Jang SK. Hepatitis C virus nonstructural protein 5A contains potential transcriptional activator domains. Mol Cells. 1997; 7(5):661–667. [PubMed: 9387155]

14. Zhang J, Hou T, Wang W, et al. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. Proc Natl Acad Sci USA. 2010; 107(4):1321–1326. [PubMed: 20080674]

15. Donlin MJ, Cannon NA, Yao E, et al. Pretreatment sequence diversity differences in the full-length hepatitis C virus open reading frame correlate with early response to therapy. J Virol. 2007; 81(15):8211–8224. [PubMed: 17522222]

16. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST). 2011; 2(3):27.

17. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. Nature genetics. 2007; 39:1167–1173. [PubMed: 17721534]

18. Lian Z, Tian QN, Liu Y, et al. Detecting Hepatitis B Viral Amino Acid Sequence Mutations in Occult Hepatitis B Infections via Bayesian Partition Model. J Proteomics Bioinform. 2013; S6:005.

19. Nousbaum J, Polyak SJ, Ray SC, et al. Prospective characterization of full-length hepatitis C virus NS5A quasispecies during induction and combination antiviral therapy. J Virol. 2000; 74(19):9028–9038. [PubMed: 10982347]

20. Dal Pero F, Tang KH, Gerotto M, et al. Impact of NS5A sequences of hepatitis C virus genotype 1a on early viral kinetics during treatment with PEGinterferon-α2a plus ribavirin. J Infect Dis. 2007; 196(7):998–1005. [PubMed: 17763320]

21. Jain MK, Yuan HJ, Adams-Huet B, et al. Pegylated interferon and ribavirin promote early evolution of nonstructural 5A protein in individuals with hepatitis C who demonstrate a response to treatment. J infect Dis. 2009; 200(6):866–876. [PubMed: 19673650]
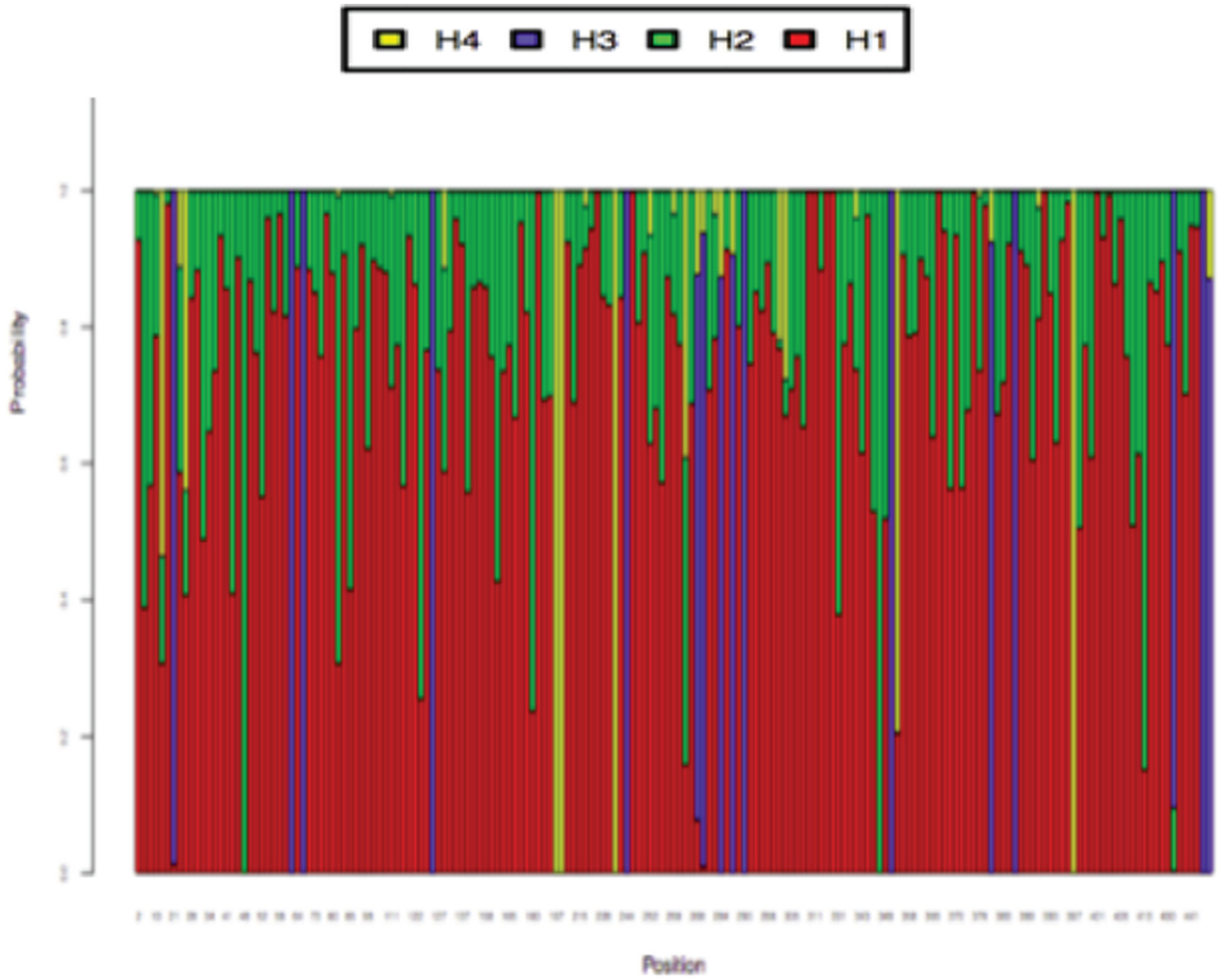
Figure 1.
Hypotheses distribution (H1, H2, H3 and H4) in a sample Markov Chain.
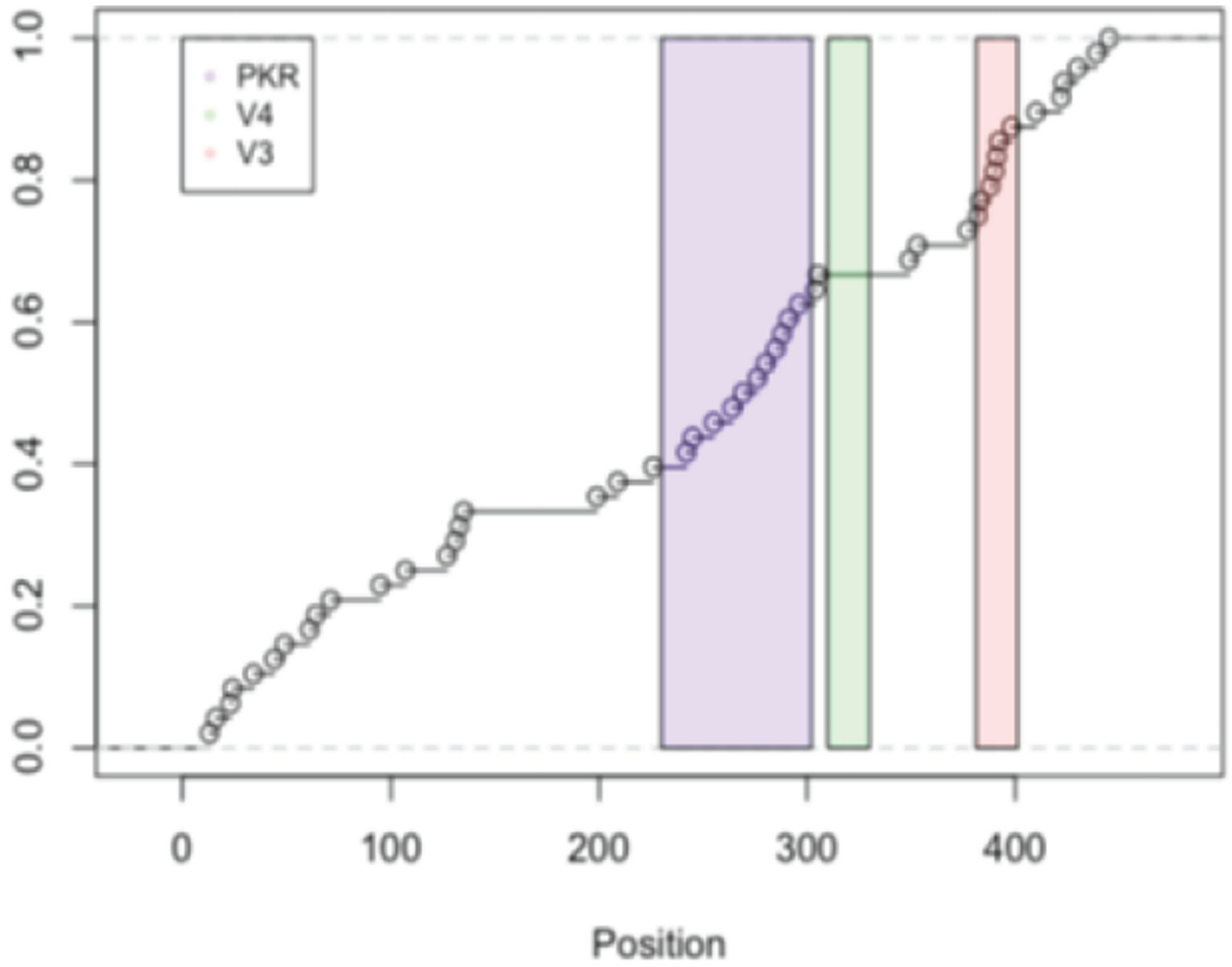
Figure 2.
Summary for Table 3
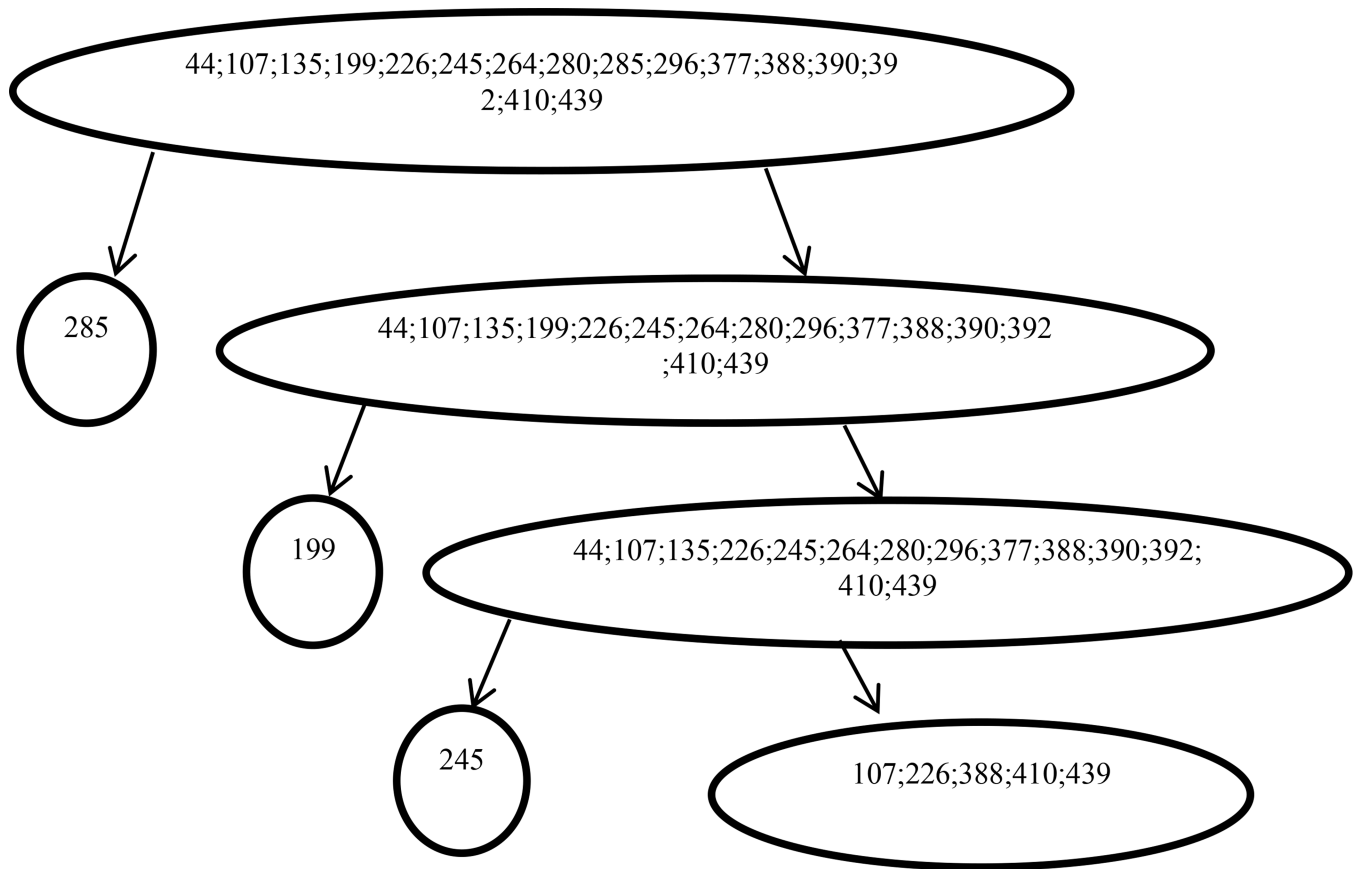
Figure 3.
Flowchart of detected mutation positions and position combinations in the pretreatment
sequence of patients who respond to the treatment.
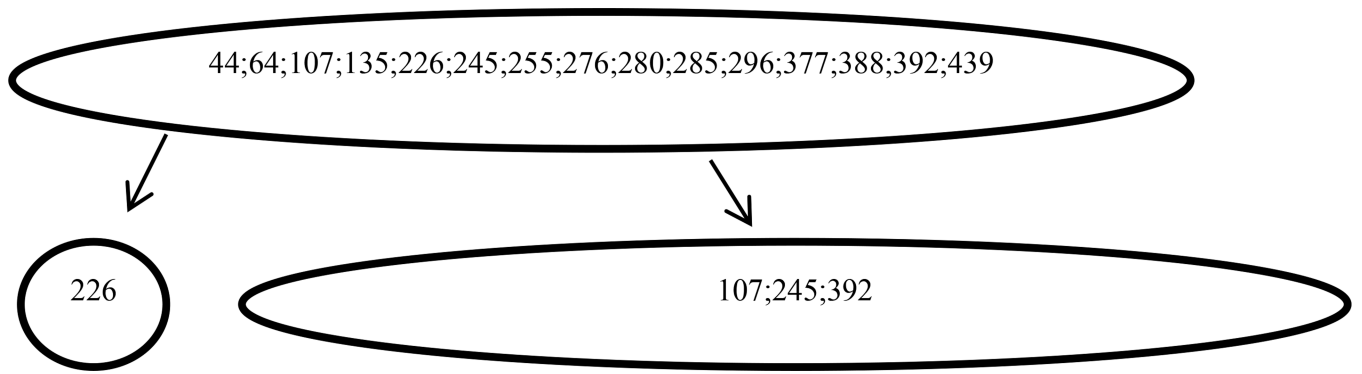
Figure 4.
Flowchart of detected mutation positions and position combinations in the pretreatment sequence of patients who don't respond to the treatment.
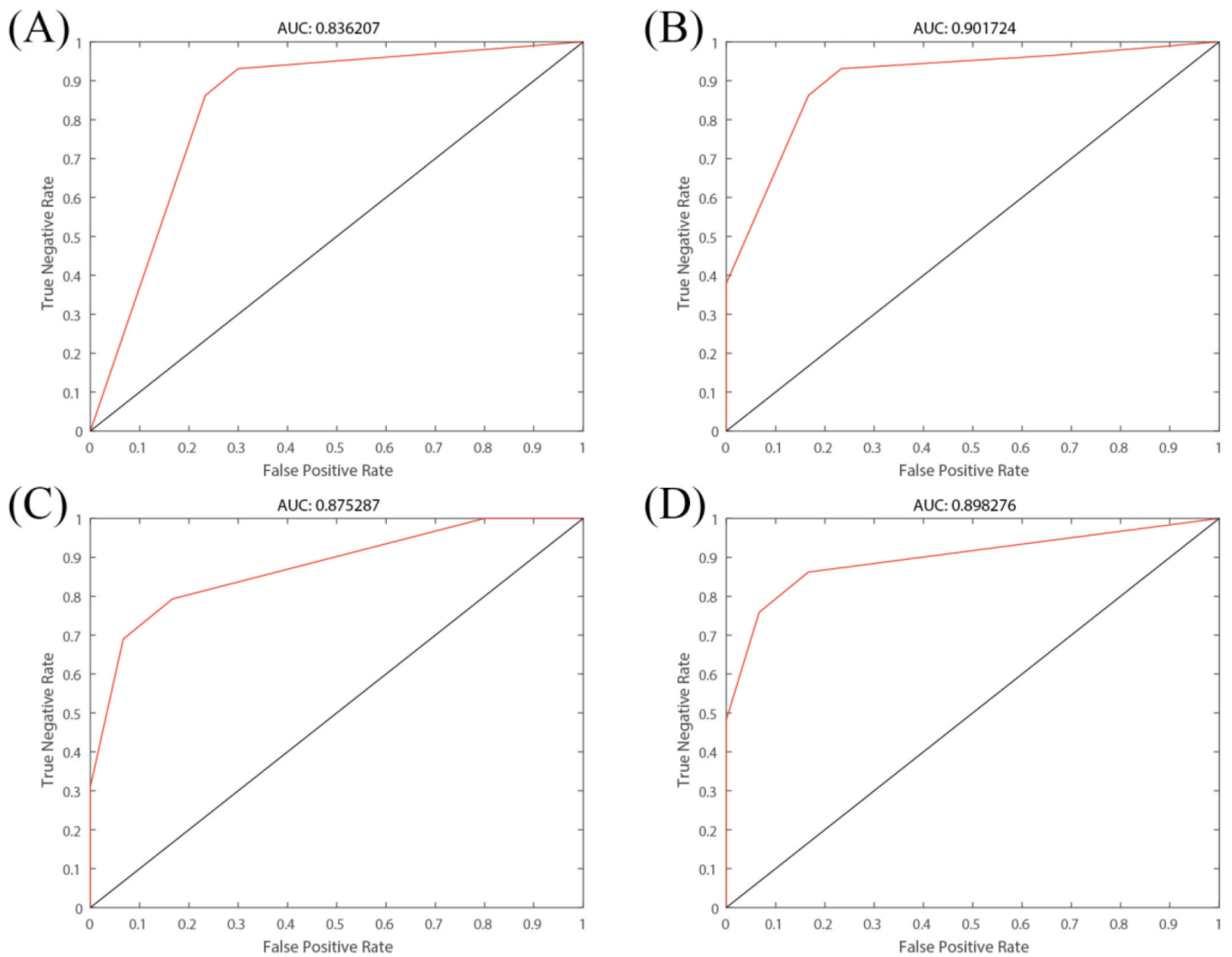
Figure 5.
ROC curve by using SVM with selected mutations as features. (A) Mutation positions: 44, 133, 226, 269, 276, 285, 288, 296, 304, 305, 382; (B) Mutation positions: 24, 44, 107, 133, 135, 226, 304, 305, 422; (C) Mutation positions: 23, 44, 61, 64, 71, 135, 245, 269, 276, 285, 291, 382, 388, 439, 445; (D) Mutation positions: 107, 226, 388, 410, 439.

Table 1

Accession numbers for 47 response sequences.

	Response Sequences					
	Accession number		Accession number		Accession number	
1	AF265047	17	AM600927	33	EF407419
2	AF265111	18	AM600936	34	EF407420
3	AF265129	19	AM600919	35	EF407421
4	AF265132	20	AM600923	36	EF407422
5	AF265138	21	AM600950	37	EF407423
6	AM600953	22	AM600914	38	EF407424
7	AM600938	23	AM600930	39	EF407425
8	AM600925	24	AM600944	40	EF407411
9	AM600932	25	AM600951	41	EF407412
10	AM600929	26	AM600942	42	FJ958369
11	AM600921	27	EF407413	43	FJ958414
12	AM600955	28	EF407414	44	FJ958465
13	AM600934	29	EF407415	45	FJ958543
14	AM600946	30	EF407416	46	FJ958850
15	AM600948	31	EF407417	47	FJ958939
16	AM600940	32	EF407418		

Table 2

Accession numbers for 29 non-response sequences

	Non-response Sequences					
	Accession Number		**Accession Number**		**Accession Number**	
1	EF407432	11	AF265105	21	EF407434
2	EF407437	12	AF265009	22	EF407435
3	EF407445	13	AF265028	23	EF407436
4	EF407427	14	EF407427	24	EF407437
5	EF407430	15	EF407428	25	EF407438
6	EF407436	16	EF407429	26	EF407439
7	AF265141	17	EF407430	27	EF407440
8	AF265135	18	EF407431	28	EF407441
9	AF265121	19	EF407432	29	EF407442
10	AF265117	20	EF407433		

Table 3

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

# Chain	H2 (P > 0.95)	H4 (P > 0.95)
1	49 349	23 64 71 245 269 276 288 291 382 388 445
2	49 349	13 390 391
3	49 349	34 242 377 398
4	Null	44 133 226 269 276 285 288 296 304 305 382
5	49 349	61 64 135 296 423 430
6	49 349	16 23 61 135 226 255 388
7	49 349	199 209 242 398
8	49 349	24 44 107 133 135 226 304 305 422
9	49 349	23 44 61 64 71 135 245 269 276 285 291 382 388 439 445
10	49 349	133 264 280 305 392
11	49	16 95 131 383 410 439
12	49 349	34 199 209 242 377 398
13	49 349	71 127 245 269 276 285 288 353 439 445
14	49 349	23 64 71 245 269 276 285 288 291 382 388 445
15	349	242 390 391 398
16	49 349	Null
17	49 349	34 95 135 255 377
18	49 349	199 209 242 398
19	49 349	133 269 276 285 288 304 305 382
20	49 349	199 209 242 398

Table 4

Detailed position interaction relations for positions for the pretreatment sequence of patients who respond to the treatment.

Positions	Amino Acids Frequency		
285	E	D	V
Non-response	13.80%	86.20%	0.00%
response	8.50%	89.40%	2.10%

Positions		
199	L	V
Non-response	100.00%	0.00%
response	87.20%	12.80%

Positions					
245	A	T	V	N	Y
Non-response	44.80%	48.30%	6.90%	0.00%	0.00%
response	29.80%	63.80%	2.10%	2.10%	2.10%

Positions	E+M+I+A+E	K+L+I+A+E	T+V+I+A+E	T+L+I+A+G	K+E+I+A+G	T+M+V+A+G	T+L+I+A+E	T+V+I+A+G	T+E+I+V+G	T+M+I+A+G
107/226/388/410/439										
Non-response	6.90%	6.90%	10.30%	6.90%	10.30%	6.90%	10.30%	6.90%	3.40%	3.40%
response	0.00%	2.10%	8.50%	8.50%	0.00%	0.00%	0.00%	19.10%	0.00%	0.00%

Positions	K+V+V+A+G	K+M+I+A+G	M+E+I+A+E	K+V+I+A+E	T+E+I+A+G	K+V+I+A+G	K+V+V+A+E	T+M+I+A+E	K+M+I+A+E	K+L+I+D+E
107/226/388/410/439										
Non-response	6.90%	3.40%	3.40%	3.40%	3.40%	3.40%	3.40%	0.00%	0.00%	0.00%
response	0.00%	0.00%	0.00%	0.00%	2.10%	0.00%	0.00%	2.10%	6.40%	2.10%

Positions	S+V+I+A+E	T+M+V+A+E	T+L+T+A+E	K+V+I+G+E	T+W+T+A+D	T+M+V+S+G	T+V+I+T+E	E+M+I+A+G	K+E+A+A+E	T+E+I+A+E
107/226/388/410/439										
Non-response	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
response	2.10%	2.10%	2.10%	2.10%	2.10%	2.10%	4.30%	2.10%	2.10%	2.10%

Positions	K+L+V+A+G	T+E+V+A+G
107/226/388/410/439		
Non-response	0.00%	0.00%
response	2.10%	2.10%

Table 5

Detailed position interaction relations for positions for the pretreatment sequence of patients who don't respond to the treatment.

Positions	Amino Acids Frequency									
226	M	L	V	E	W					
Non-response	21.70%	11.50%	49.30%	0.70%	0.70%					
response	24.00%	18.00%	38.00%	6.00%	0.00%					
107/245/392	E+A+N	K+A+N	T+A+N	T+T+D	K+T+N	T+T+N	T+V+S	T+V+D	M+A+D	K+T+S
Non-response	6.90%	17.20%	6.90%	6.90%	10.30%	27.60%	3.40%	3.40%	3.40%	3.40%
response	2.10%	8.50%	12.80%	17.00%	8.50%	34.00%	0.00%	0.00%	0.00%	2.10%
107/245/392	T+A+D	K+A+D	T+T+V	S+A+N	T+N+N	K+V+N	K+Y+N			
Non-response	3.40%	6.90%	0.00%	0.00%	0.00%	0.00%	0.00%			
response	2.10%	2.10%	2.10%	2.10%	2.10%	0.00%	2.10%			