

Published online ahead of print on 12 May 2008 as DOI 10.1099/vir.0.2008/002469-0

**Positive selection at the receptor-binding site of haemagglutinin H5 in viral sequences derived from human tissues**

Alita Kongchanagul, Ornpreeya Suptawiwat, Mongkol Uiprasertkul, Pilaipan Puthavathana and Prasert Auewarakul

Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

**Correspondence**

Prasert Auewarakul

sipaw@mahidol.ac.th

Highly pathogenic H5N1 avian influenza virus has spread through at least 45 countries in three continents. Despite the ability to infect and cause severe disease in humans, the virus cannot transmit efficiently from human to human. The lack of efficient transmission indicates the incompleteness of the adaptation of the avian virus to the new host species. The required mutations for the complete adaptation and the emergence of a potential pandemic virus are likely to originate and be selected within infected human tissues. Differential receptor preference plays an important role in the species-tropism of avian influenza. We have analysed quasispecies of sequences covering the receptor-binding domain of the haemagglutinin gene of H5N1 viruses derived from fatal human cases. We employed a likelihood ratio test to identify positive-selection sites within the quasispecies. Nine of seventeen positive-selection sites identified in our analyses were found to be located within or flanking the receptor-binding domain. Some of these mutations are known to alter receptor-binding specificity. This suggests that our approach could be used to screen for mutations with significant functional impact. Our data provide new candidate mutations for the viral adaptation to a human host, and a new approach to search for new genetic markers of potential pandemic viruses.

## INTRODUCTION

The high replication error of RNA viruses, such as influenza virus, results in a mixed viral population with many variants, referred to as quasispecies (Eigen, 1996). Although, most sequence variations are neutral and offer no competitive advantage, the quasispecies provides multiple variants, which can be readily selected if there is a change in selective pressure. Transmission of the virus to a new host species provides new selective pressures and can result in the expansion of the 'best fit' minor variants for adaptation to the new environment.

In general, human and avian influenza A viruses differ in their recognition of host cell receptors. Human influenza viruses preferentially recognize  $\alpha$ -2,6-linked sialic acid, while avian influenza viruses recognize  $\alpha$ -2,3-linked sialic acid (Rogers & Paulson, 1983). However, the highly pathogenic avian influenza viruses subtype H5N1 can transmit directly from avian species to humans (Subbarao *et al.*, 1998; Tran *et al.*, 2004). Even though the H5N1 viruses can infect and cause severe disease in humans, they do not bind the  $\alpha$ -2,6-linked sialic acid receptor with high affinity (Ha *et al.*, 2001; Stevens *et al.*, 2006). This property is believed to be one of the major factors that prevent the H5N1 virus from transmitting efficiently amongst humans and causing a pandemic. Amino acid substitutions in the haemagglutinin (HA) gene can lead to the altered receptor-binding preference of the virus from  $\alpha$ -2,3-linked to  $\alpha$ -2,6-linked sialic acid (Auewarakul *et al.*, 2007; Stevens *et al.*, 2006; Yamada *et al.*, 2006). This would enable avian H5N1 viruses to recognize human-type host cell receptors and could potentially enable the virus to transmit efficiently within the human population and cause a catastrophic pandemic. Therefore, it is extremely important to monitor the viral changes that may lead to the emergence of pandemic viruses.

The earliest possible detection of selected mutants is by looking at sequences within viral quasispecies before they expand and become dominant virus. In order to detect mutants that might have altered phenotypes we studied the viral sequences at the level of quasispecies. We conducted a study where the viral sequence was directly amplified, cloned and sequenced from a nasopharyngeal aspirate or tissue specimens. The specimens were obtained from fatal human cases in Thailand.

Selection at the protein level can be measured by  $\omega$  ( $d_N/d_S$ ), in which  $d_N$ =non-synonymous substitution rate (non-synonymous changes per non-synonymous site) and  $d_S$ =synonymous substitution rate (synonymous changes per synonymous site). If amino acid changes provide better fitness, the mutations will be fixed at a higher rate than synonymous mutations. This results in  $d_N > d_S$  and  $\omega > 1$ . Originally,  $\omega$  was calculated as an

average for the whole gene, which does not allow sensitive detection of individual amino acid residue under positive selection. Subsequently codon-based models that allow the  $\omega$  ratio to vary amongst sites were developed (Nielsen & Yang, 1998; Yang & Nielsen, 2000). These models describe  $\omega$  ratio distribution amongst sites: M0 assumes one constant  $\omega$  for all sites; M3 classifies sites into discrete classes with different  $\omega$ ; M7 allows  $\omega$  to vary according to a distribution that represents negative or neutral selection; M8 adds on top of M7 a discrete  $\omega$  class for sites with positive selection ( $\omega > 1$ ). Likelihood ratio test is used as a statistical test of goodness-of-fit to compare the two models and test whether the more complex model, e.g. M8 or M3, fits the dataset significantly better than the simpler model, e.g. M7 or M0. An empirical Bayes approach is then used to calculate the posterior probability that each site is from a particular site class, and sites with high posterior probabilities coming from the class with  $\omega > 1$  are inferred to be under positive selection (Yang *et al.*, 2005).

## **METHODS**

### **Patient.**

The first patient (patient A) was a non-autopsy case, previously reported (Auewarakul *et al.*, 2007). Patient A was a 5-year-old boy who had progressive viral pneumonia that led to respiratory failure and death by 12 days after the onset of illness. In December 2005, he developed a fever, stomach ache, nausea and vomiting and was admitted to the hospital 10 days later when he developed dyspnea and a chest radiography showed patchy infiltration at the right middle lobe. The symptom of dyspnea worsened to respiratory failure and the pulmonary infiltration spread to both lungs on the following day. The diagnosis of avian influenza was suspected on day 12. The use of the antiviral drug oseltamivir was started and the patient expired on the same day. The patient was not known to have had direct contact with any sick or dying birds, but he played in the yard where the birds were often present. Nasopharyngeal aspirate was collected on day 12 after the onset of illness.

Autopsy was conducted in the second (patient B) and third patients (patient C). Patient B, previously reported (Uiprasertkul *et al.*, 2007), was a 48-year-old man who had progressive viral pneumonia in October 2005. He had a fever, cough, runny nose, myalgia and chest pain at the onset of illness. Dyspnea developed on day 2 of the illness and a chest radiograph showed interstitial infiltrations at the right upper and left middle lung fields and a mass-like infiltration at the right middle lung field. The diagnosis of avian influenza was suspected on day 4 of the illness after a history of direct contact with dying chickens was revealed. Respiratory secretions were then sent to national laboratories and they were confirmed positive for influenza (H5N1) virus. The patient died on day 6 of the illness.

Patient C, previously reported (Uiprasertkul *et al.*, 2005), was a 6-year-old boy who had progressive viral pneumonia in January 2004. He was initially treated with multiple broad-spectrum anti-microbial agents. Virological diagnosis of H5N1 infection was made on day 7 of the illness. After oseltamivir became available in Thailand, he was treated on day 15 of his illness with this agent until he died. He was also treated with methylprednisolone on day 15 until death and with granulocyte colony-stimulating factor for leukopenia from day 5 to 10 of the illness. The patient died on day 17 of the illness.

The use of the patients' specimens was approved by the Ethics Committee of the Faculty of Medicine Siriraj Hospital.

### **Viral RNA, cloning and quasispecies analysis.**

For patient A, total RNA from the nasopharyngeal specimen was extracted according to the manufacturer's protocol (QIAmp RNA mini kit; Qiagen). For patients B and C, total

RNAs were extracted by using Trizol from paraffin-embedded blocks of lung and intestine tissue samples and then purified using Qiagen RNaseasy kit according to the manufacturer's instructions.

A fragment of the HA gene covering the receptor-binding site (nt 413–905) was amplified from RNA extracted from the specimen by using the high fidelity enzyme *Pfu* (Promega) and the primers HHAf2 (5'-GGTCCAGTCATGAAGCCTCA-3') and HA-H5r12 (5'-TTTATCGCCCCATTGGAGT-3'). The PCR product was cloned into pGEM T-Easy. One hundred clones of each sample were picked up and sequenced.

The selective pressures acting on the receptor region were estimated by using the CODEML program in the PAML package. We used models M7 and M8, where M7 contains 10  $\omega$  categories to describe  $\omega$  amongst sites, all constrained to be  $<1$ ; M8 differs from M7 only in that it estimates  $\omega$  for an extra class of sites (p10) at which  $\omega$  can be  $>1$  (Yang, 1997). Models were compared using a likelihood ratio test and the Bayes Empirical Bayes (BEB) method was used for a posteriori estimation of individual codons under positive selection (Yang *et al.*, 2005).

## RESULTS

We have analysed the HA sequences from either each individual patient or each organ (lung or intestine) of the patient. During a phylogenetic analysis (data not shown), the consensus wild-type sequence of each patient was found to be similar to one another and closely related to other clade 1 sequences from Thailand. All together 17 positive-selection sites were identified (Table 1) within this sequenced region spanning 143 aa residues (128–270). At these positions, the sequence majority in all the samples contains similar amino acids as in the consensus sequence of all human H5N1 viruses available in the GenBank database, except for positions 133 and 138 where mutant amino acids were found in higher frequencies. Total numbers of synonymous and non-synonymous substitutions, as well as lists of non-synonymous substitutions in all the RNA samples are shown in Table 2. Some of these mutations with low frequencies were not picked up as positive-selection site by the BEB analysis. Amongst the positive-selection sites, six sites were repeatedly found when each patient was analysed individually. These sites were 133, 138, 161, 186, 222 and 227 (H3 numbering system). Of these all but the 133 and 161 positions are in the known receptor-binding domain. The receptor-binding site of HA at the tip of HA1 globular domain is composed of three secondary structure elements: the 190 helix (residues 190–198), the 130 loop (residues 135–138) and the 220 loop (residues 221–228), forming the sides of each site; and the base made up of the conserved residues Tyr<sup>98</sup>, Trp<sup>153</sup>, His<sup>183</sup> and Tyr<sup>195</sup> (Skehel & Wiley, 2000). Although, the positions 133 and 230 are not in the receptor-binding domain, they flank the 130 and 220 loops and they showed positive selection in our analyses, suggesting that mutations at these sites might also contribute to the receptor-binding adaptation. A138V, N186K and S227N mutations were previously reported to confer  $\alpha$ -2,6-linked sialic acid binding to H5N1 virus (Auewarakul *et al.*, 2007; Gambaryan *et al.*, 2006; Shinya *et al.*, 2005; Yamada *et al.*, 2006). The mutant sequences in our analyses are similar to these mutations at positions 138 and 227, whereas our mutation at position 186 is N186D. Amongst the 17 positive-selection sites that were identified in our study, eight sites are not known to be related to the receptor-binding domain. But, these sites all showed only low mutation frequencies. In other words, all strong positive-selection sites with high mutation frequencies are related to the receptor-binding domain.

In patients A, B and C, we found seven, five and ten positive-selection sites, respectively. Of these five, three and five sites, respectively, are in or they flank the receptor-binding domain (Table 1). The higher frequency of positive-selection sites in patient C may be related to the fact that this patient died on day 17 of the illness, while the samples from the

other two patients were collected earlier in the course of the illness. The longer duration of infection provided a longer period under the selective pressure and might cause the virus to gain more adaptation. In patient C, there were markedly more positive-selection sites that are not related to the receptor-binding domain, including the *N*-linked glycosylation site at position 158. This suggests that other selective pressures, such as the immune response, might be involved in late phases of the disease.

When each type of tissue from patients B and C was analysed, four and seven positive-selection sites were identified in the lung and intestine, of which two and four sites are related to the receptor-binding domain, respectively. (Table 1). There are some differences in the positive-selection sites from different tissues. In particular, the S227N mutation was found in high frequency (36.7%) in the intestine of patient C and in low frequency (3%) in the intestine of patient B, while it was absent in the lung tissue from both cases. This suggested a strong selection for this mutation and a compartmentalization of the viral population within the patients.

Most of the identified positive-selection sites concentrated in the N-terminal 2/3 part of the sequences. Sixteen sites were identified in the region covering the receptor-binding domain from residue 128 to 230, whereas only two sites were identified in the rest of the sequences (residue 231–270).

Mutations that have been reported to alter receptor-binding specificity of H5 are A138V, N186K, Q196R, S227N, Q226L and G228S (Auewarakul *et al.*, 2007; Gambaryan *et al.*, 2006; Shinya *et al.*, 2005; Stevens *et al.*, 2006; Yamada *et al.*, 2006). Our analyses provide new candidate mutations that may affect the receptor-binding specificity. Whether these mutations have functional consequences is under investigation.

## DISCUSSION

Minimal positive selection in the C-terminal region outside the receptor-binding domain suggests that the observed positive-selection sites were not fortuitous. Although, most of the positive-selection sites were within or very close to the receptor-binding domain, there were also other sites not related to the receptor-binding pocket. It is not clear what drove positive selection at these sites. It is possible that a mutation outside the receptor-binding pocket may induce conformational changes that indirectly affect the pocket. On the other hand, among the eight positive-selection sites outside the receptor-binding pocket, five sites are on the surface of HA (129, 158, 161, 163 and 168) and only three sites are hidden inside the molecule (180, 150 and 152). Although, these five exposed sites are not within known antigenic epitopes (Kaverin *et al.*, 2002, 2007), it is possible that they constitute an unknown epitope.

The observed differences in the viral quasispecies in lung and intestine suggest that there might be a compartmentalization of viral infection and that the selective pressure might be different among different tissues. Human lung has been shown to contain  $\alpha$ -2,3-linked sialic acid in alveolar epithelial cells, whereas human intestinal epithelium lacks this receptor. In human intestine the  $\alpha$ -2,3-linked sialic acid receptor was identified only on neurons (Yao *et al.*, 2008). It is not clear whether the difference in the sialic acid receptor distribution between the two tissues contributed to the different selection of viral sequences.

Our data demonstrate adaptation of the receptor-binding domain of H5N1 virus in infected human tissues. If allowed to be transmitted further to other human hosts, the mutants would be likely to be selected further and expand, and eventually cause emergence of a potential pandemic virus. Understanding the adaptation is therefore of utmost importance. Adaptation of avian influenza virus to human host involves multiple mechanisms. However, the receptor usage preference is likely to be a major step in the adaptation process. Although, there have been reports on mutations that altered receptor-binding specificity of H5, those reported mutation only conferred partial switching from  $\alpha$ -2,3-linkage tropism to dual tropism (Auewarakul *et al.*, 2007; Gambaryan *et al.*, 2006; Stevens *et al.*, 2006; Yamada *et al.*, 2006). It is likely that these mutations are not sufficient and full switching to  $\alpha$ -2,6-linkage tropism is probably needed for an efficient transmission in the human population. Such mutation that can cause a complete switching of H5 is not known. The fact that our analyses could pick up mutations those that are known to change the receptor-binding property of H5N1 viruses indicates that they can be used to screen and search for mutations with significant functional effects. Our analyses offer an approach to

find candidate mutations, which should be studied further for determining functionality. Finding mutations with pandemic potential before the actual emergence of such viruses will provide genetic markers for vigilant monitoring, which will hopefully help us to avoid the pandemic.

## **ACKNOWLEDGEMENTS**

This work was supported by a research grant from the National Center for Genetic Engineering and Biotechnology of Thailand.

## REFERENCES

- Auewarakul, P., Suptawiwat, O., Kongchanagul, A., Sangma, C., Suzuki, Y., Ungchusak, K., Louisirirotchanakul, S., Lerdsamran, H., Pooruk, P. & other authors (2007).** An avian influenza H5N1 virus that binds to a human-type receptor. *J Virol* **81**, 9950–9955. [Medline](#)
- Eigen, M. (1996).** On the nature of virus quasispecies. *Trends Microbiol* **4**, 216–218. [Medline](#)
- Gambaryan, A., Tuzikov, A., Pazynina, G., Bovin, N., Balish, A. & Klimov, A. (2006).** Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology* **344**, 432–438. [Medline](#)
- Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. (2001).** X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc Natl Acad Sci U S A* **98**, 11181–11186. [Medline](#)
- Kaverin, N. V., Rudneva, I. A., Ilyushina, N. A., Varich, N. L., Lipatov, A. S., Smirnov, Y. A., Govorkova, E. A., Gitelman, A. K., Lvov, D. K. & Webster, R. G. (2002).** Structure of antigenic sites on the haemagglutinin molecule of H5 avian influenza virus and phenotypic variation of escape mutants. *J Gen Virol* **83**, 2497–2505. [Medline](#)
- Kaverin, N. V., Rudneva, I. A., Govorkova, E. A., Timofeeva, T. A., Shilov, A. A., Kochergin-Nikitsky, K. S., Krylov, P. S. & Webster, R. G. (2007).** Epitope mapping of the hemagglutinin molecule of a highly pathogenic H5N1 influenza virus by using monoclonal antibodies. *J Virol* **81**, 12911–12917. [Medline](#)
- Nielsen, R. & Yang, Z. (1998).** Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936. [Medline](#)
- Rogers, G. N. & Paulson, J. C. (1983).** Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **127**, 361–373. [Medline](#)
- Shinya, K., Hatta, M., Yamada, S., Takada, A., Watanabe, S., Halfmann, P., Horimoto, T., Neumann, G., Kim, J. H. & other authors (2005).** Characterization of a human H5N1 influenza A virus isolated in 2003. *J Virol* **79**, 9926–9932. [Medline](#)
- Skehel, J. J. & Wiley, D. C. (2000).** Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69**, 531–569. [Medline](#)
- Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C. & Wilson, I. A. (2006).** Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410. [Medline](#)

- Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C. & other authors (1998).** Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**, 393–396. [Medline](#)
- Tran, T. H., Nguyen, T. L., Nguyen, T. D., Luong, T. S., Pham, P. M., Nguyen, V. C., Pham, T. S., Vo, C. D., Le, T. Q. & other authors (2004).** Avian influenza A (H5N1) in 10 patients in Vietnam. *N Engl J Med* **350**, 1179–1188. [Medline](#)
- Uiprasertkul, M., Puthavathana, P., Sangsiriwut, K., Pooruk, P., Srisook, K., Peiris, M., Nicholls, J. M., Chokephaibulkit, K., Vanprapar, N. & Auewarakul, P. (2005).** Influenza A H5N1 replication sites in humans. *Emerg Infect Dis* **11**, 1036–1041. [Medline](#)
- Uiprasertkul, M., Kitphati, R., Puthavathana, P., Kriwong, R., Kongchanagul, A., Ungchusak, K., Angkasekwinai, S., Chokephaibulkit, K., Srisook, K. & other authors (2007).** Apoptosis and pathogenesis of avian influenza A (H5N1) virus in humans. *Emerg Infect Dis* **13**, 708–712. [Medline](#)
- Yamada, S., Suzuki, Y., Suzuki, T., Le, M. Q., Nidom, C. A., Sakai-Tagawa, Y., Muramoto, Y., Ito, M., Kiso, M. & other authors (2006).** Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382. [Medline](#)
- Yang, Z. (1997).** PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556. [Medline](#)
- Yang, Z. & Nielsen, R. (2000).** Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**, 32–43. [Medline](#)
- Yang, Z., Wong, W. S. & Nielsen, R. (2005).** Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**, 1107–1118. [Medline](#)
- Yao, L., Korteweg, C., Hsueh, W. & Gu, J. (2008).** Avian influenza receptor expression in H5N1-infected and noninfected human tissues. *FASEB J* **22**, 733–740. [Medline](#)

**Table 1.** Positive-selection sites on the HA gene from residue 128 to 270

Sources of sequences, wild-type and mutant amino acids, positions (H3 numbering), frequencies of mutant amino acids and site-specific  $\omega$  ( $d_N/d_S$ ) $\pm$ SEM are shown ( $\omega > 1$  indicates positive selection). The wild-type amino acids are from the consensus sequence of all human H5N1 viruses available in the GenBank database. Bold-typed residues are related to the receptor-binding domain. The residues that have been shown to carry receptor preference determinant for H5N1 are underlined.

Sample	Patient characteristic	Wild-type	Site	Mutant	Frequency (%)	$\omega$ ( $\pm$ SEM)
Patient A (nasopharyngeal secretion)	5-year-old boy, diagnosed in December 2005, specimen collected on day 12 after onset of illness	S/L	<b>133</b>	<b>V</b>	59	9.78 $\pm$ 0.94
		<u><b>A</b></u>	<u><b>138</b></u>	<u><b>V</b></u>	<u>58</u>	<u>9.78<math>\pm</math>0.94</u>
		Y	161	H	7	8.85 $\pm$ 2.89
		<u><b>N</b></u>	<u><b>186</b></u>	<u><b>D</b></u>	<u>3</u>	<u>6.86<math>\pm</math>4.29</u>
		N	<b>224</b>	<b>S</b>	3	9.76 $\pm$ 1.03
		<u><b>S</b></u>	<u><b>227</b></u>	<u><b>N</b></u>	<u>2</u>	<u>9.54<math>\pm</math>1.72</u>
Patient B (lung and intestinal tissues)	48-year-old man, diagnosed in October 2005, died on day 6 after onset of illness	N	250	S	2	6.90 $\pm$ 4.28
		T	163	A/T	6	7.77 $\pm$ 1.47
		<b>L</b>	<b>194</b>	<b>P</b>	6	7.84 $\pm$ 1.35
		<b>K</b>	<b>222</b>	<b>E</b>	6	5.67 $\pm$ 3.03
		<b>M</b>	<b>230</b>	<b>T</b>	5	4.26 $\pm$ 3.14
		I	252	V	7	7.83 $\pm$ 1.36

Patient C (lung and intestinal tissues)	6-year-old boy, diagnosed in January 2004, died on day 17 after onset of illness	H	129	Q	3	7.76±1.43
		<b>S/L</b>	<b>133</b>	<b>V/F</b>	13/1	7.89±1.12
		<b>A</b>	<b>138</b>	<b>V</b>	<u>39</u>	<u>7.88±1.14</u>
		N	158	S/D	1/1	6.97±2.48
		Y	161	C/H/R	1/2/1	7.89±1.12
		Y	168	C	1	4.93±3.53
		W	180	R	1	7.28±2.24
		<b>N</b>	<b>186</b>	<b>D</b>	<u>19</u>	<u>7.89±1.12</u>
		<b>K</b>	<b>222</b>	<b>E</b>	45	7.89±1.12
		<b>S</b>	<b>227</b>	<b>N</b>	<u>18</u>	<u>7.89±1.12</u>
Lung (pooled sequences from patients B and C)		<b>W</b>	<b>153</b>	<b>R</b>	1	8.62±2.03
		Y	161	H/R	3/1	8.78±1.69
		T	163	A	6	5.57±3.81
		<b>L</b>	<b>194</b>	<b>P</b>	6	8.89±1.43
Intestine (pooled sequences from patients B and C)		<b>S/L</b>	<b>133</b>	<b>V/F</b>	11/1	6.53±1.17
		Y	161	C/R	1	6.30±1.59
		W	180	R	1	4.85±2.80
		<b>N</b>	<b>186</b>	<b>D</b>	<u>23</u>	<u>5.53±2.27</u>
		<b>K</b>	<b>222</b>	<b>E</b>	34	6.53±1.17
		<b>S</b>	<b>227</b>	<b>N</b>	<u>23</u>	<u>5.36±2.37</u>

**Table 2.** Numbers of all synonymous and non-synonymous substitutions and a list of non-synonymous substitutions in each dataset

RNA sample	No. synonymous substitution	No. non-synonymous substitution	List of non-synonymous substitution (frequency)
Patient A nasopharyngeal aspirate	17	138	S128R (1), L133V (54), S136G (1), A138V (53), A138E (1), Q142R (1), K157R (3), Y161H (7), N172D (1), N186D (3), E190D (1), L194H (1), N224S (3), N224D (2), S227N (2), S227G (1), N240S (1), N250S (2)
Patient B lung	98	94	S128G (1), A131V (1), A131T (3), Y141C (1), K144R (1), V152A (1), W153R (1), I155V (1), N158D (1), Y161H (1), T163A (11), I164V (1), N169S (1), V178A (2), D187G (1), D187N (2), E190G (1), K193R (18), K193S (3), K193E (1), L194P (13), N197D (1), P198L (1), T200A (1), I202V (1), N210S (1), I217V (1), K222E (2), G228E (2), M230T (10), F233S (1), F223L (1), N244D (1), F245S (1), I252V (3), E255G (1)
Patient B intestine	35	62	E134V (2), S136G (2), R149K (2), S167G (1), S167C (1), Y168* (1), N172S (1), H183R (1), H184R (1), N186D (2), R193K (11), L194P (2), Y195* (1), Q196R (2), P198S (1), T199A (1), V204A (1), T206A (2), K222E (7), N224H (1), G225R (1), S227N (2), M230T (1), F233S (1), P239S (1), N244S (1), I252V (8), I268V (2), K270E (2)

Patient C lung	20	145	S128N (1), H129Q (1), L133V (6), S136G (4), A138V (55), S146P (1), W153* (1), I155T (1), N158D (1), Y161H (3), Y161R (1), Y168C (2), N169H (1), I182V (1), N186D (1), T199I (2), T206A (1), T208A (1), T219I (1), E222K (56), R229G (1), S247N (2), F251S (2)
Patient C intestine	31	189	S127A (1), H129Q (3), A131D (1), L133V (19), L133F (1), S136G (1), A138V (20), S146P (1), W153* (1), N158S (3), Y161C (1), Y161R (1), L177S (1), W180R (2), I182T (1), N186D (37), K193R (1), Y201* (1), N210D (1), E222K (51), S227N (36), I252V (4), K262R (1)

\*, Stop codon.