ACADEMIA JOURNAL OF BIOLOGY 2021, 43(1): 119–128
DOI: 10.15625/2615-9023/15829

POLYMORPHISM OF THE *TMPRSS2* GENE RELATING TO COVID-19 SUSCEPTIBILITY IN VIETNAMESE POPULATION

Nguyen Dang Ton[1,2], Vu Phuong Nhung[1], Duong Thu Trang[1], Nguyen Thi Thanh Hoa[1], Nguyen Hoai Nam[1], Hoang Thi Thuan[2], Ho Anh Son[3], Nguyen Thi Thanh Hai[4,5], Pham Ngoc Thach[4], Nong Van Hai[1,2], Nguyen Hai Ha[1,2,*]

[1]Institute of Genome Research, VAST, Vietnam
[2]Graduate University of Science and Technology, VAST, Vietnam
[3]Vietnam Military Medical University, Ha Noi, Vietnam
[4]National Hospital for Tropical Diseases, Ha Noi, Vietnam
[5]Hanoi Medical University, Ha Noi, Vietnam

Received 12 December 2020, accepted 16 March 2021

ABSTRACT

Recently, a contagious lung disease named coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), rapidly spread worldwide and has many serious consequences for human health. Human genetic polymorphisms may contribute to the variation of incidence, mortality as well as severity of COVID-19. To date, this factor in the Vietnamese population remains unknown. A cellular protease termed transmembrane protease serine 2 (TMPRSS2) was found to play a vital role in the entry of SARS-CoV-2 into host cells. In this study, we investigated polymorphisms in the *TMPRSS2* gene from 270 whole exome sequencing data of Vietnamese peoples. We also employed bioinformatics tools including SIFT, Polyphen-2, and PROVEAN to predict the possible function of missense variants. A total of 34 *TMPRSS2* variants were identified, of which, 29 were in non-coding regions and 14 were in coding regions. Variants found in exons included seven synonymous and seven non-synonymous point mutations, one of which was novel mutation (c.A1336C/p.R446R). Mutation c.G589A/p.V197M (rs12329760) possesses the highest frequency and was predicted to have the ability to damage protein by SIFT and Polyphen-2. In addition, the damaging possibility was also found in c.T244G/p.Y82D and c.C896T/p.A299V variants. This study contributes to the understanding of Vietnamese genetic variation databases relating to susceptibility to COVID-19.

Keywords: SARS-CoV-2, human *TMPRSS2*, genetic polymorphism.

Citation: Nguyen Dang Ton, Vu Phuong Nhung, Duong Thu Trang, Nguyen Thi Thanh Hoa, Nguyen Hoai Nam, Hoang Thi Thuan, Ho Anh Son, Nguyen Thi Thanh Hai, Pham Ngoc Thach, Nong Van Hai, Nguyen Hai Ha, 2021. Polymorphism of the *TMPRSS2* gene relating to COVID-19 susceptibility in Vietnamese population. *Academia Journal of Biology*, 43(1): 119–128. https://doi.org/10.15625/2615-9023/15829

**Corresponding author email*: nguyenhaiha@igr.ac.vn

INTRODUCTION

Late December 2019, a novel beta-coronavirus designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Gorbalenya et al., 2020) was detected to be the etiologic agent of a new viral pneumonia emerging in Wuhan city, Hubei Province, China (Zhu et al., 2020; Wang et al., 2020). Consequently, the transmission occurred from person to person (Chan et al., 2020; Dong et al., 2020) and this lung disease, then termed severe coronavirus disease 2019 (COVID-19) by World Health Organization (WHO), rapidly spread worldwide. As of 4 December 2020, there were approximately 63.9 million confirmed cases and 1.4 million deaths recorded during COVID-19 epidemic (WHO, 2020). SARS-CoV-2 entry is driven by binding of the viral spike (S) proteins to the peptidase angiotensin-converting enzyme 2 (ACE2), spike cleavage mediated by Furin, and S protein priming by a cellular protease named transmembrane protease serine 2 (TMPRSS2) (Li et al., 2003; Kuhn et al., 2004; Hoffmann et al., 2020; Walls et al., 2020; Wu et al., 2020). SARS-CoV-2 infection therefore depends on the expression of ACE2, TMPRSS2 and Furin.

Incidence and mortality of COVID-19 pandemic vary worldwide based on multiple factors such as age, race, ethnicity and health condition. For example, diabetes, hypertension and cardiovascular disease were indicated as being risk factors for severity and mortality in people infected with SARS-CoV-2 (De Almeida-Pititto et al., 2020). An assessment of the influence of COVID-19 among American Indian and Alaska Native in 23 States conducted by Hatcher and his colleages highlighted that the incidence rate for this population was 3.5 times higher than that for non-Hispanic white persons (Hatcher et al., 2020). Furthermore, another evaluation of COVID-19 impact variation across States of America and New York City also showed disparities in death rates associated with race, ethnicity and age (Gross et al., 2020). In these States, both the Black and Latinx experienced a significant higher risk of mortality than the White population.

Apart from environmental factors, health conditions and methods for prevention, genetic susceptibility may be a reasonable explanation for not only race- and ethnicity-stratified COVID-19 incidence and mortality, but also disease severity. Indeed, polymorphisms in genes responsible for SARS-CoV-2 entry such as *ACE2* and *TMPRSS2* play a causative role in variable transmissibility and disease progression among patients (Hou et al., 2020; Baughn et al., 2020). However, the genetic polymorphisms of these genes in the Vietnamese population have not been reported to date. Since the *ACE2* gene showed a low level of polymorphism in recent reports (Torre-Fuentes et al., 2020), the current study concentrated on the higher polymorphic *TMPRSS2*, to enhance understanding of this gene in the Vietnamese population and its association with virus diseases in further studies. In summary, we identified all genetic variants of *TMPRSS2* in 270 whole exome sequencing (WES) data of Vietnamese people. Furthermore, we used bioinformatics tools including SIFT, Polyphen-2 and PROVEAN to predict the possible functions of missense variants.

MATERIALS AND METHODS

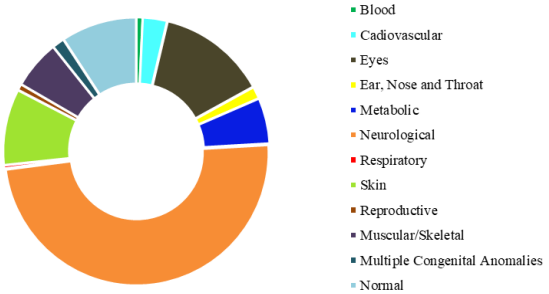


Figure 1. Distribution of subjects based on health condition

Our study was conducted using available WES data of 270 Vietnamese subjects. The data were collected and stored by the Institute of Genome Research from 2018 to 2020. This group comprises healthy people and patients

with diseases related to blood, cardiovascular system, eyes, ear, nose and throat, metabolism, neurological system, respiratory system, skin, reproductive system, muscular/skeletal and multiple congenital anomalies. Patients with Parkinson's disease of neurological group made up the highest proportion of subjects in this study.

Whole exome sequencing

For WES, library construction was performed by Sure Select V6-Post using the manufacturer's protocol (Agilent Technologies, Santa Clara, California, USA). The sequencing library is prepared by random fragmentation of DNA, following by 5' and 3' adapter ligation. Fragments ligated to adapter were subsequently amplified by PCR and gel purified. Enriched library was quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). DNA fragments distribution was confirmed by 2100 Bioanalyzers using High sensitivity DNA chip (Agilent Technologies, Santa Clara, California, USA) with an expected size range from 200 bp to 400 bp. Paired-end sequencing was conducted on the NovaSeq platform (Illumina, San Diego, California, USA) following the manufacturer's instructions. The mean exome coverage was more than 100X and each target base had at least 20X coverage.

WES data analysis

BWA (ver 0.7.10) was used to align raw sequence to the UCSC Human Reference Genome hg19. Single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels) were detected by GATK (ver 3.6) and SAMtools variant caller (ver 1.8). Removing duplicate sequence reads was performed by Markduplicates in the Picard package. A self-designed tool was used to collect profiles of all TMPRSS2 variants detected from WES data of 270 subjects.

Sorting intolerant from tolerant (SIFT)

SIFT is a popular bioinformatics tool used to predict the ability of amino acid substitutions as well as non-synonymous polymorphisms to affect the function of proteins based on sequence homology and physical properties of amino acids (https://sift.bii.a-star.edu.sg/). SIFT output is based on the scores that range from 0.0 to 1.0. The substitutions with scores less than 0.05 and with scores greater than or equal to 0.05 are considered to be damaging and tolerated, respectively.

Polymorphism phenotyping v2 (Polyphen-2)

Polyphen-2 is a bioinformatics tool that predicts whether amino acid substitutions have impacts on protein structure and function (http://genetics.bwh.harvard.edu/pph2/).

Polyphen-2 result depends on the scores ranging from 0.0 to 1.0. Variants with scores in the range from 0.0 to 0.15, from 0.15 to 1.0 and from 0.85 to 1.0 are predicted to be benign, possibly damaging and damaging, respectively.

Protein variation effect analyzer (PROVEAN)

PROVEAN is a software tool that predicts the possible effects of protein sequence variations, including amino acid substitutions and indels on the biological function of proteins (http://provean.jcvi.org/index.php). PROVEAN score threshold is set at -2.5. Variants with scores equal to or below -2.5 and with scores above -2.5 are predicted to be deleterious and neutral, respectively.

RESULTS

Genetic variants of *TMPRSS2* in the coding region

All coding variants of *TMPRSS2* detected in 270 Vietnamese subjects are shown in table 1. Variant identification relied on the nucleotide reference sequence NM_001135099. Totally, 14 variants were found in the coding region of *TMPRSS2*, comprising 7 synonymous SNPs: c.C192G, c.C222T, c.A336G, c.C888T, c.T879C, c.G1203A, c.A1336C and 7 non-synonymous SNPs: c.G23T, c.T209C, c.T244G, c.G589A, c.G634A, c.C896T, c.G1568A. Out of 14 variants, one was novel mutation (c.A1336C, p.R446R) which has not been reported in all public databases. Variants with the highest

frequencies are c.G589A (0.36296), c.C888T (0.26296) and c.A336G (0.20185) (Table 1).

As indicated in Figure 2, synonymous mutations were found mainly in Trypsin and Scavenger receptor cysteine-rich (SRCR)-like domain (p.I293I, p.G296G, p.V401V, p.R446R) and domains of unknown function (p.P64P, p.Y74Y, p.T112T). Non-synonymous variants are located in Trypsin domain (A299V, R523Q), SRCR-like domain (V197M, D212N) and domains of unknown function (G8V, V70A, Y82D).

Table 1. Variants of *TMPRSS2* in the coding region.

Exon	Location on chromosome	Reference SNP ID	Nucleotide change (*)	Amino acid change (*)	Frequency (**)	Frequency in this study	Number of subjects	
							Het	Hom
Exon 1	chr21: 42879909	rs75603675	c.G23T	p.G8V	0.24381	0.02407	13	0
Exon 3	chr 21: 42866440	rs141685390	c.C192G	p.P64P	0.00060	0.00370	2	0
	chr 21: 42866410	rs187460831	c.C222T	p.Y74Y	0.00020	0.00185	1	0
	chr 21: 42866423	rs201093031	c.T209C	p.V70A	0.00060	0.00370	2	0
	chr 21: 42866388	rs201679623	c.T244G	p.Y82D	0.00100	0.01296	7	0
	chr 21: 42866296	rs3787950	c.A336G	p.T112T	0.16294	0.20185	91	9
Exon 6	chr 21: 42852497	rs12329760	c.G589A	p.V197M	0.26138	0.36296	124	36
	chr 21: 42852452	rs748571451	c.G634A	p.D212N	Unknown	0.00741	4	0
Exon 9	chr 21: 42845366	rs150445636	c.C896T	p.A299V	0.00040	0.00185	1	0
	chr 21: 42845374	rs2298659	c.C888T	p.G296G	0.20947	0.26296	102	20
	chr 21: 42845383	rs17854725	c.T879C	p.I293I	0.36621	0.11111	52	4
Exon 11	chr 21: 42842654	rs140121827	c.G1203A	p.V401V	0.00060	0.00185	1	0
Exon 12	chr 21: 42840412	novel	c.A1336C	p.R446R	Unknown	0.00185	1	0
Exon 13	chr 21: 42839671	rs572530227	c.G1568A	p.R523Q	0.00020	0.00185	1	0

Notes: (*) NM_001135099; (**) Frequency in "The 1000 genomes project 2015"; Hom: Homozygous; Het: Heterozygous.

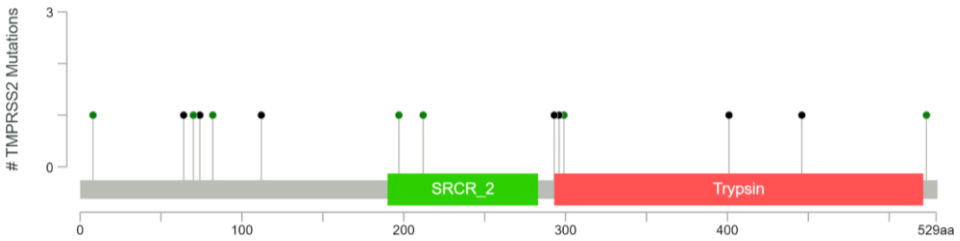
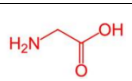
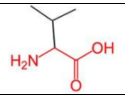
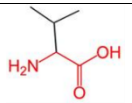
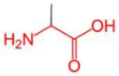
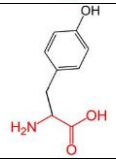
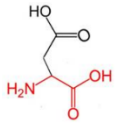
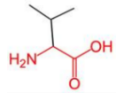
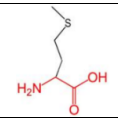
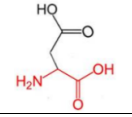
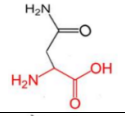
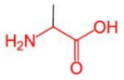
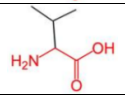
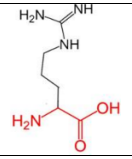
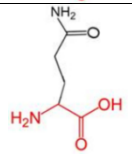


Figure 2. Distribution of *TMPRSS2* mutations mapped by cBioPortal for Cancer Genomics. Black: synonymous mutation, green: non-synonymous mutation

Schematic structures and properties of amino acid substitutions are shown in Table 2. The hydrophobicity-value are based on the hydrophobicity scale of Kyte and Doolittle (Kyte & Doolittle, 1982). The greater the values, the more hydrophobic the amino acids. Except for p.G8V and p.A299V that pertain to more hydrophobic mutant residues, most variants result in equal or less hydrophobic mutant amino acids in comparison to the wild-type one. The charge of each amino acid is characterized by the isoelectric point (pI), which is the pH at which the molecule is electrically neutral. There is no change in molecule's charge of 4 out of 7 amino acid replacements, including p.G8V, p.V70A, p.V197M and p.A299V. On the other hand, p.Y82D, p.D212N and p.R523Q lead to neutral-to-negative, negative-to-neutral and positive-to-neutral amino acid substitutions, respectively. The differences in molecular size are found in 6 substitutions. Of which, three mutants (p.G8V, p.V197M, p.A299V) are bigger and three mutants (p.V70A, p.Y82D, p.R523Q) are smaller than the wild-type residues.

Table 2. Schematic structures and properties of amino acid substitutions.

Amino acid substitution	Wild-type residue			Mutant residue			Size
	Schematic structures	Hydrophobicity-value (Kyte & Doolittle, 1982)	Charge (according to pI)	Schematic structures	Hydrophobicity-value (Kyte & Doolittle, 1982)	Charge (according to pI)	
p.G8V		-0.4	Neutral		4.2	Neutral	Mutant is bigger
p.V70A		4.2	Neutral		1.8	Neutral	Mutant is smaller
p.Y82D		-1.3	Neutral		-3.5	Negative	Mutant is smaller
p.V197M		4.2	Neutral		1.9	Neutral	Mutant is bigger
p.D212N		-3.5	Negative		-3.5	Neutral	Sizes are almost the same
p.A299V		1.8	Neutral		4.2	Neutral	Mutant is bigger
p.R523Q		-4.5	Positive		-3.5	Neutral	Mutant is smaller

Functional prediction of variants using SIFT, Polyphen-2 and Mutation Taster

SIFT, Polyphen-2 and Mutation Taster were employed to evaluate possible effects of non-synonymous mutations in our study (Table 3). These three predictions might pave the way to investigate the possibility and extent of the missense variants to damage the protein. It is noteworthy that 3 mutations were predicted to exert deleterious effects on protein function by 2 out of 3 models. In particular, damaging abilities of p.Y82D and p.A299V were shown by Polyphen-2 and PROVEAN, while that of p.V197M was reported by SIFT and Polyphen-2. The remains were benign, neutral, possibly damaging or tolerated.

Table 3. Prediction of functional effects of *TMPRSS2* variants.

Variants	Functional domain of protein	Function prediction		
		SIFT	Polyphen-2	PROVEAN
p.G8V	-	T	P	N
p.V70A	-	T	B	N
p.Y82D	-	T	D	D
p.V197M	SRCR-like domain	D	D	T
p.D212N	SRCR-like domain	T	B	N
p.A299V	Serine proteases, trypsin domain	T	D	D
p.R523Q	Serine proteases, trypsin domain	T	B	N

Notes: B: Benign; D: Damaging; N: Neutral; P: Possibly damaging; T: Tolerated, "-": unknown.

The results of Polyphen-2 prediction of p.V197M variant were shown in Figure 3. In Polyphen-2 prediction models, a combination of two pairs of datasets, including HumDiv and HumVar was used to test. HumDiv trained Polyphen-2 model prediction pointed out that it is probably damaging with a score of 0.997. Similarly, Humvar prediction also denoted that it is probably damaging with a score of 0.937. Hence, this result leads to the ability that c.G589A mutation might alter the function of protein.

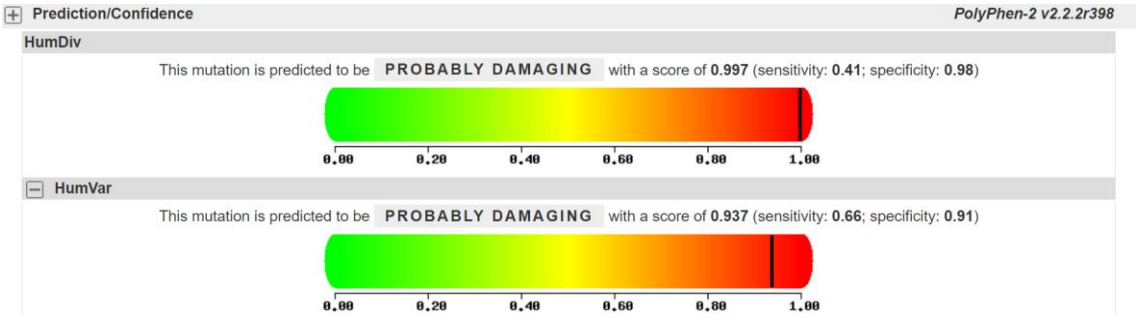


Figure 3. Polyphen-2 prediction of amino acid substitution caused by c.G589A/p.V197M (rs12329760)

Genetic variants of *TMPRSS2* in the non-coding regions

Our study reveals a total of 29 variants in non-coding regions (Table 4). Two point mutations were identified in 5' untranslated region (UTR5), including c.-25G>A (rs890983082) and a novel variant c.-35C>T. We found 26 variants in introns, of which one relates to the deletion of an A nucleotide (rs542502939), one arises from the insertion of a sequence of nucleotides (rs112132031), one involves the deletion of two nucleotides (rs140530035), and the

others occur through substitution. The frequency of each variant was calculated as shown in table 1. Variant rs140530035 shows the highest frequency (0.62222), followed by rs422471 (0.29444) and rs386416 (0.29259). The remaining mutations (rs141230106) were detected in 3' untranslated region (UTR3).

Table 4. Variants of *TMPRSS2* in the non-coding regions

Region	Location on chromosome	Reference SNP ID	Nucleotide change (*)	Frequency (**)	Frequency in this study	Number of subjects	
						Het	Hom
UTR5	chr 21: 42879956	rs890983082	c.-25G>A	-	0.00185	1	0
UTR5	chr 21: 42879966	unknown	c.-35C>T	-	0.00185	1	0
intron	chr 21: 42838094	rs542502939	delA	-	0.00185	1	0
intron	chr 21: 42838104	rs118028230	G>C	0.01877	0.12222	54	6
intron	chr 21: 42840172	rs743542	G>A	0.14936	0.00370	2	0
intron	chr 21: 42840187	rs190618812	G>A	-	0.00185	1	0
intron	chr 21: 42840580	rs462321	T>C	0.57808	0.01111	6	0
intron	chr 21: 42840595	rs462326	G>C	0.57308	0.00926	5	0
intron	chr 21: 42842713	rs112132031	insCTAAGG CCTCGGG	0.82987	0.22963	38	43
intron	chr21: 42842771	rs28524972	C>G	0.28714	0.01111	6	0
intron	chr 21: 42843960	rs464431	A>G	0.87380	0.01667	5	2
intron	chr 21: 42845633	rs2298660	C>T	0.25859	0.08889	32	8
intron	chr 21: 42845638	rs55964536	C>T	0.24241	0.00185	1	0
intron	chr 21: 42845642	rs2298661	C>A	0.26777	0.12222	42	12
intron	chr 21: 42848561-42848562	rs140530035	delAG	0.81330	0.62222	136	100
intron	chr 21: 42850911	rs9974995	C>T	0.26098	0.01111	6	0
intron	chr 21: 42850966	rs73372193	A>C	0.05611	0.01111	6	0
intron	chr 21: 42850977	rs9974933	A>G	0.26158	0.22963	108	8
intron	chr 21: 42851006	rs9975014	A>G	0.26238	0.23704	110	9
intron	chr 21: 42852591	rs378501	A>G	-	0.00370	2	0
intron	chr 21: 42860307	rs422471	C>T	0.55471	0.29444	129	15
intron	chr 21: 42860485	rs386416	G>C	0.55511	0.29259	128	15
intron	chr 21: 42860494	rs3819138	C>G	0.06849	0.05556	26	2
intron	chr 21: 42860593	rs3787947	C>T	0.30711	0.00556	3	0
intron	chr 21: 42861332	rs429442	C>T	0.27955	0.01111	6	0
intron	chr 21: 42869927	rs181592444	C>T	0.00220	0.00185	1	0
intron	chr 21: 42879724	rs8126497	G>A	0.10104	0.02222	12	0
intron	chr 21: 42879835	rs951118351	C>A	-	0.00370	2	0
UTR3	chr 21: 42838004	rs141230106	c.*65G>A	0.0012	0.00370	2	0

Notes: (*) NM_001135099; (**) Frequency in "The 1000 genomes project 2015"; Hom: Homozygous; Het: Heterozygous; "-": unknown.

DISCUSSION

Recently, Hou et al. (2020) highlighted that genetic variation in *ACE2* or *TMPRSS2* gene could personalize the treatment of people suffering from COVID-19. However, it might be difficult to directly disrupt ACE2 receptor in therapies because ACE2 serves as an essential regulator of heart function (Crackower et al., 2002). The absence of ACE2 activity in mice resulted in severe cardiac defects and an increase in angiotensin II levels. Meanwhile, *TMPRSS2* mutant mice were normal and displayed no significant

abnormalities in organ histology and function (Kim et al., 2006). Consequently, *TMPRSS2* protease could possibly be a promising candidate for development of treatments for COVID-19. Recently, a comparative study also reported that *TMPRSS2* gene might be a potential factor that guides the severity of the COVID-19 pandemic in the Italian population, which presented higher death rates and stratification in severity among sexes than East Asians (Asselta et al., 2020). Simultaneously, they found no strong evidence for the contribution of ACE2 in the severity of COVID-19. *TMPRSS2* gene therefore was chosen for this polymorphic investigation.

Protease-dependent SARS-CoV-2 infection as well as pathogenesis relies on endosomal cysteine proteases cathepsin B and L (CatB/L) and *TMPRSS2* protease (Hoffmann et al., 2020; Simmons et al., 2005; Shulla et al., 2011). The endosomal cysteine proteases-mediated entry requires acidic pH. The inhibitors of endosomal acidification, for example, NH_4Cl and bafilomycin A1, were demonstrated to be only effective for viral entry via CatB/L pathway. Interestingly, these drugs might exert a negligible effect on patients with *TMPRSS2* in wild-type form while CatB/L protease is dispensable. As a result, in the treatment of COVID-19, drugs against endosomal acidification might be potential for patients carrying missense *TMPRSS2* variants rather than patients with wild-type phenotype.

In the current study, nearly half of all subjects possess p.V197M (rs12329760) variant that was proposed to be damaging by bioinformatics tools. *TMPRSS2* rs12329760 variant was remarkably related to genomic rearrangements pertaining to TMPRSS2, which leads to the risk of prostate cancer (FitzGerald et al., 2008). Moreover, this group of researchers also showed rs12329760 mutation in another reference sequence (p.V160M) to be highly conserved among mammals. This missense variant is located in an exonic splicing enhancer srp40 site. Additionally, A allele is proposed to interrupt this exonic splicing enhancer, resulting in the possibility of exon skipping or protein malformation.

In conclusion, our study identified a total of 43 variants of the *TMPRSS2* gene in 270 Vietnamese's WES data. Variants found in coding regions included seven synonymous and seven non-synonymous point mutations, of which one was a novel mutation (c.A1336C/p.R446R). Variant c.G589A/p.V197M (rs12329760) possesses the highest frequency and was predicted to have the ability to damage protein by SIFT and Polyphen-2.

REFERENCES

Asselta R., Paraboschi E. M., Mantovani A., Duga S., 2020. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging (Albany. NY).*, 12(11): 10087–10098.

Baughn L. B., Sharma N., Elhaik E., Sekulic A., Bryce A. H., Fonseca R., 2020. Targeting *TMPRSS2* in SARS-CoV-2 Infection. *Mayo Clin. Proc.*, 95(9): 1989–1999.

Chan J. F.-W., Yuan S., Kok K.-H., To K. K.-W., Chu H., Yang J., Xing F., Liu J., Yip C. C.-Y., Poon R. W.-S., Tsoi H.-W., Lo S. K.-F., Chan K.-H., Poon V. K.-M., Chan W.-M., Ip J. D., Cai J.-P., Cheng V. C.-C., Chen H., Hui C. K.-M., Yuen K.-Y., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*, 395(10223): 514–523.

Crackower M. A., Sarao R., Oliveira-dos-Santos A. J., Da Costa J., Zhang L., 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature*, 417(6891): 822–828.

De Almeida-Pititto B., Dualib P. M., Zajdenverg L., Dantas J. R., De Souza F. D., Rodacki M., Bertoluci M. C., 2020. Severity and mortality of COVID 19 in patients with diabetes, hypertension and

cardiovascular disease: A meta-analysis. *Diabetol. Metab. Syndr.*, 12(1): 75.

Dong Y., Dong Y., Mo X., Hu Y., Qi X., Jiang F., Jiang Z., Jiang Z., Tong S., Tong S., Tong S., 2020. Epidemiology of COVID-19 among children in China. *Pediatrics*. https://doi.org/10.1542/peds. 2020-0702

FitzGerald L. M., Agalliu I., Johnson K., Miller M. A., Kwon E. M., Hurtado-Coll A., Fazli L., Rajput A. B., Gleave M. E., Cox M. E., Ostrander E. A., Stanford J. L., Huntsman D. G., 2008. Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: Results from a population-based study of prostate cancer. *BMC Cancer*, 8. https://doi.org/10.1186/1471-2407-8-230

Gorbalenya A., Baker S., Baric R., de Groot R., Drosten C., Gulyaeva A., Haagmans B., Lauber C., Leontovich A., Neuman B., Penzar D., Perlman S., Poon L., Samborskiy D., Sidorov I., Sola I., Ziebuhr J., 2020. Severe acute respiratory syndrome-related coronavirus : The species and its viruses – a statement of the Coronavirus Study Group. *Nat. Microbiol.*, https://doi.org/10.1101/2020. 02.07.937862

Gross C. P., Essien U. R., Pasha S., Gross J. R., Wang S. yi, Nunez-Smith M., 2020. Racial and Ethnic Disparities in Population-Level Covid-19 Mortality. *J. Gen. Intern. Med.*, 35(10): 3097–3099.

Hatcher S. M., Agnew-Brune C., Anderson M., Zambrano L. D., Rose C. E., Jim M. A., Baugher A., Liu G. S., Patel S. V., Evans M. E., Pindyck T., Dubray C. L., Rainey J. J., Chen J., Sadowski C., Winglee K., Penman-Aguilar A., Dixit A., Claw E., Parshall C., Provost E., Ayala A., Gonzalez G., Ritchey J., Davis J., Warren-Mears V., Joshi S., Weiser T., Echo-Hawk A., Dominguez A., Poel A., Duke C., Ransby I., Apostolou A., McCollum J., 2020. COVID-19 Among American Indian and Alaska Native Persons - 23 States, January 31–July 3, 2020. *MMWR. Morb. Mortal. Wkly. Rep.*, 69(34): 1166–1169.

Hoffmann M., Kleine-Weber H., Schroeder S., Krüger N., Herrler T., Erichsen S., Schiergens T. S., Herrler G., Wu N. H., Nitsche A., Müller M. A., Drosten C., Pöhlmann S., 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2): 271−280.e8.

Hou Y., Zhao J., Martin W., Kallianpur A., Chung M. K., Jehi L., Sharifi N., Erzurum S., Eng C., Cheng F., 2020. New insights into genetic susceptibility of COVID-19: An ACE2 and TMPRSS2 polymorphism analysis. *BMC Med.*, 18(1).

Kim T. S., Heinlein C., Hackman R. C., Nelson P. S., 2006. Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease. *Mol. Cell. Biol.*, 26(3): 965–975.

Kuhn J. H., Li W., Choe H., Farzan M., 2004., November Angiotensin-converting enzyme 2: A functional receptor for SARS coronavirus. *Cell. Mol. Life Sci.* https://doi.org/10.1007/s00018-004-4242-5

Kyte J., Doolittle R. F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157(1): 105–132.

Li W., Moore M. J., Vasllieva N., Sui J., Wong S. K., Berne M. A., Somasundaran M., Sullivan J. L., Luzuriaga K., Greeneugh T. C., Choe H., Farzan M., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965): 450–454.

Shulla A., Heald-Sargent T., Subramanya G., Zhao J., Perlman S., Gallagher T., 2011. A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol.*, 85(2): 873–882.

Simmons G., Gosalia D. N., Rennekamp A. J., Reeves J. D., Diamond S. L., Bates P., 2005. Inhibitors of cathepsin L prevent severe acute respiratory syndrome

coronavirus entry. *Proc. Natl. Acad. Sci. U. S. A.*, 102(33): 11876–11881.

Torre-Fuentes L., Matías-Guiu J., Hernández-Lorenzo L., Montero-Escribano P., Pytel V., Porta-Etessam J., Gómez-Pinedo U., Matías-Guiu J. A., 2020. ACE2, TMPRSS2, and Furin variants and SARS-CoV-2 infection in Madrid, Spain. *J. Med. Virol.*, 93(2): 863–869.

Walls A. C., Park Y. J., Tortorici M. A., Wall A., McGuire A. T., Veesler D., 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2): 281−292.e6.

Wang C., Horby P. W., Hayden F. G., Gao G. F., 2020. A novel coronavirus outbreak of global health concern. *Lancet*., https://doi.org/10.1016/S0140-6736(20)30185−9

WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020 (accessed 4 December 2020)

Wu C., Zheng M., Yang Y., Gu X., Yang K., Li M., Liu Y., Zhang Q., Zhang P., Wang Y., Wang Q., Xu Y., Zhou Y., Zhang Y., Chen L., Li H., 2020. Furin: A Potential Therapeutic Target for COVID-19. *IScience*, 23(10): 101642.

Zhu N., Zhang D., Wang W., Li X., Yang B., Song J., Zhao X., Huang B., Shi W., Lu R., Niu P., Zhan F., Ma X., Wang D., Xu W., Wu G., Gao G. F., Tan W., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.*, 382(8): 727–733.