

新型冠状病毒人群易感性分析报告

背景

新型致病性冠状病毒（SARS-CoV-2）的出现与其在中国乃至世界范围内的迅速蔓延已构成了一场全球公共卫生危机。[现有研究](#)表明，SARS-CoV-2进入细胞，需要病毒刺突(S)蛋白与细胞受体血管紧张素转换酶2(ACE2)结合，以及宿主细胞蛋白酶(TMPRSS2)启动S蛋白。据此，我们对ACE2、TMPRSS2基因在人群中的突变情况和表达情况进行分析，希望能解释人群易感性差异，为保护高危人群提供理论证据支持。

方法

1. 基因信息收集
- 收集特定基因在核苷酸数据库(NCBI Genes, GeneCards)、蛋白质数据库(Uniprot, NCBI Protein)中的基本信息，并对可能的多个基因isoform序列进行多重比对分析。
2. 突变信息收集
- 收集特定基因在正常人群突变数据库(dbSNP, gnomad, CMDDB)、疾病人群数据库(GDC)中的基本突变信息。
3. 基因表达相关情况收集
- 于GTEx数据库中，分析特定基因在多种组织中的表达状况，并收集与其特定组织内表达相关的eQTL变异位点加以分析。

结果

ACE2基因

1. isoform序列多重比对
- 对NCBI数据库中5个ACE2的isoform序列进行多重比对，发现其在与冠状病毒spike蛋白结合的氨基酸残基位置上高度保守，尚不能认为不同isoform对冠状病毒的结合力有差异。

24	27	31	34	37	38	41	42	45	79	82	83	90	325	329	330	353	354	
L	T	T	Y	Q	E	Y	Q	V	L	T	Y	D	Q	E	N	K	G	civet ACE2
N	T	N	Q	E	D	Y	Q	L	T	S	F	T	Q	A	N	H	G	mouse ACE2
K	S	K	Q	E	D	Y	Q	L	I	N	F	N	P	T	N	H	G	rat ACE2
Q	T	K	H	E	D	Y	Q	L	L	M	Y	N	Q	E	N	K	G	human ACE2
N473	Y475	Y475	Y440	Y491	Y436	Y484	Y436	Y484	L472	L472	N473	T402	R426	R426	T486	G488	Y491	human SARS
		Y442	N479			T486	Y484				Y475					T487	G488	
						T487										Y491		

2. 影响病毒结合的ACE2突变位点分析

之前**已有研究**发现，位于人类ACE2蛋白序列31、41、82-84、353-357位置上的氨基酸残基对于冠状病毒的S蛋白结合十分重要。

1. 人群突变位点数据库

目前**已有研究**应用中国代谢解析计划(ChinaMAP)、千人基因组(1KGP)突变位点数据库对ACE2基因内存在的错义突变进行分析，发现其错义突变位点均位于ACE2蛋白胞外拓扑结构域内，但都**不改变关键位点上氨基酸结构**。

通过中国百万人基因组计划数据库(CMDB)查询ACE2相应突变位点，发现其**错义突变位点**均位于ACE2蛋白胞外拓扑结构域内，但都**不改变关键位点上氨基酸结构**。

查看NCBI dbSNP数据库中关于**ACE2的编码区SNP信息**，并筛选出3个与SARS冠状病毒S蛋白结合关键位点上的错义突变，但这3个错义突变在人群中的突变频率(AF)都很小，提示在不同人群中均缺乏对冠状病毒刺突蛋白结合天然抵抗的ACE2变异。

- 82AA: Met [M] -> Ile [I] [rs766996587](#)——AF:0.00001 (2/182758, GnomAD_exome)
- 84AA: Pro [P] -> Thr [T] [rs759134032](#)——AF:0.00001 (1/182792, GnomAD_exome)
- 355AA: Asp [D] -> Asn [N] [rs961360700](#)——AF:0.00001 (2/170310, GnomAD_exome)

查看**gnomAD**人群基因组突变频率数据库中关于ACE2的编码区SNP信息，该数据库网站暂时无法访问，待补充。

2. 疾病人群突变位点数据库

查看TCGA项目的GDC数据库中关于**ACE2的突变信息**。其存在错义、终止、移码突变共计120个，但都**不改变关键位点上氨基酸结构**。

3. 基因表达情况分析

1. 人体组织中表达状况

ACE2主要在睾丸、小肠末端回肠、内脏脂肪、肾脏、心脏、甲状腺组织之中高表达(TPM>5)，可能使上述组织细胞成为易受新冠病毒感染的对象，并引起肺炎外的一系列继发病状。

目前有**相应研究**应用单细胞RNA-seq测序数据，提出消化系统也是新冠病毒的潜在感染路径，进一步论证了上述推测。

2. 表达数量性状基因座(eQTLs)分析

为找到可能影响到关键组织(如肺部、肾脏、睾丸等)中ACE2表达的eQTLs，我们对GTEx数据库中**所有ACE2相关的eQTLs**进行分析，发现睾丸组织中存在2个使ACE2表达增加的eQTL突变(rs143695310, rs3053013)，且其在GnomAD数据库中的AF分别为0.0226、0.0218，提示有少部分携带该突变人群的男性生殖系统可能更易受新冠病毒侵袭受损，需要加以关注。

TMPRSS2基因

1. isoform序列多重比对

对NCBI数据库中2个TMPRSS2的isoform序列(529, 492AA)进行多重比对，发现其**在胞外的拓扑结构域高度保守**，亦即不同的isoform对其胞外酶解功能无影响。

2. 影响细胞蛋白酶活性的TMPRSS2突变位点分析

考虑到不同数据库间对于参照的(canonical)isoform选择并不一致，我们以NCBI和Uniprot数据库的选择为准，以较短的isoform(492AA, transmembrane protease serine 2 isoform 2)为基准进行分析。

依据Uniprot数据库中对于跨膜丝氨酸蛋白酶2(TMPRSS2)功能改变的相关突变位点信息，位于人类TMPRSS2蛋白序列255、441位置上的氨基酸残基，分别对酶的催化、非催化序列分裂以及催化活性起着重要作用。

1. 人群突变位点数据库

通过中国百万人基因组计划数据库(CMDB)查询TMPRSS2相应突变位点，发现其**错义突变位点**均位于跨膜丝氨酸蛋白酶2非催化链上，亦即**不改变重要功能位点上氨基酸结构**。

查看NCBI dbSNP数据库中关于**TMPRSS2的编码区SNP信息**，并筛选出2个与跨膜丝氨酸蛋白酶2催化能力改变相关的错义突变，但这2个错义突变在人群中的突变频率(AF)都很小，提示在不同人群中均缺乏使跨膜丝氨酸蛋白酶2催化能力减弱的TMPRSS2变异。

- 255AA: Arg [R] -> Ser [S] [rs769655195](#)——AF:0.00002 (3/125568, TOPMED)
- 441AA: Ser [S] -> Gly [G] [rs1292701415](#)——AF:0.00001 (1/125568, TOPMED)

查看**gnomAD**人群基因组突变频率数据库中关于ACE2的编码区SNP信息，该数据库网站暂时无法访问，待补充。

2. 疾病人群突变位点数据库

查看TCGA项目的GDC数据库中关于**TMPRSS2的突变信息**。其存在错义、终止、移码突变共计76个，但都**不改变重要功能位点上氨基酸结构**。

3. 基因表达情况分析

1. 人体组织中表达状况

TMPRSS2主要在前列腺、胃部、横结肠、胰腺、肺部、小肠末端回肠、唾液腺、肾脏、食道黏膜、甲状腺、肝脏等具有分泌功能的组织之中高表达(TPM>10)。结合对ACE2高表达基因的分析结果，我们认为，小肠末端回肠、肾脏组织中ACE2、TMPRSS2的高表达状况可能会使其极易被新冠病毒入侵并造成损伤，新冠病毒感染患者的消化系统、泌尿生殖系统可能会出现相应的并发症状。

事实上，目前对于新冠肺炎患者的**临床特征分析**显示，纳入的38名新冠病毒感染者中有1名患者出现了腹泻症状，并且13名ICU患者中有3名伴随急性肾损伤。而在另一项**单中心回顾性观察研究**中，52名重症患者大部分都伴随器官损伤，包括15例(29%)急性肾损伤和15例(29%)肝功能不全病例，这也在临床上初步证实了我们的结论。

2. 表达数量性状基因座(eQTLs)分析

为找到可能影响到关键组织(如肺部、肾脏、睾丸等)中TMPRSS2表达的eQTLs，我们对GTEx数据库中**所有TMPRSS2相关的eQTLs**进行分析，发现192个分别与肺部、睾丸组织相关的eQTLs，受所携带的不同eQTLs影响，不同个体在肺部或睾丸组织内的TMPRSS2表达情况可能存在差异(增加或降低)，进而可能对于病毒侵袭的反应程度不同，这一推测有待未来进一步研究论证。