

## 新型冠状病毒基因组变异与诊断

熊子军<sup>1\*</sup>, 张喆<sup>2\*</sup>, 王雅琦<sup>3</sup>, 朱德芹<sup>1</sup>, 郭佑民<sup>4</sup>

(1.西安交通大学, 陕西西安 710061; 2.珠海市人民医院, 广东珠海 519000; 3.苏州锐讯  
生物科技有限公司 江苏苏州 215000; 4.西安交通大学第一附属医院, 陕西西安 710061)

**摘要:** 快速准确的分子诊断是控制传染病和疫情爆发的关键工具。新型冠状病毒(简称 2019-nCoV 或 SARS-CoV-2)感染导致的肺炎患者数量迅速增加并引发了疫情蔓延, 2019-nCoV 基因组序列、核酸检测的诊断试剂和方法相继报道, 高深度基因组测序和 RT-PCR 是目前检测 2019-nCoV 的重要方法。但是核酸检测阴性而影像学诊断阳性的新型冠状病毒肺炎(简称 COVID-19 或 NCP)案例逐步增加, 给临床诊治带来了另一个巨大挑战, 因此, 提高核酸分子诊断的效能和灵敏度是亟待解决的问题。目前, 2019-nCoV 的起源、动物和人的传播途径、分子诊断及临床治疗、疫情爆发科学预警等已经成为研究热点。本文从遗传学、基因组学、病毒学和演化生物学的角度, 对 2019-nCoV 基因组变异与分子诊断的近期研究进展论述分析, 通过比较 2019-nCoV 与 SARS-CoV 基因组以及感染不同患者的 2019-nCoV 基因组变异, 为精准快速的分子诊断、抗病毒药物的靶点筛选提供依据, 也为 2019-nCoV 的追踪溯源和临床防治提供新思路。

**关键词:** 新冠病毒; 序列变异; 毒株; 分子分型

中图分类号: N3                      文献标志码: A

收稿日期: 2020-02-07                      修回日期: 2020-02-19

熊子军和张喆为同等贡献作者

通讯作者: 郭佑民, 教授, [cjr.guoyoumin@vip.163.com](mailto:cjr.guoyoumin@vip.163.com)

## Genomic variations of SARS-COV-2 and detection

XIONG Zi-jun<sup>1\*</sup>, ZHANG Zhe<sup>2\*</sup>, WANG Ya-qi<sup>3</sup>, ZHU De-qin<sup>1</sup>, GUO You-min<sup>4</sup>

1. Xi'an Jiaotong University, Xi'an 710061; 2. Zhuhai People's Hospital/Zhuhai Hospital  
Affiliated to Jinan University, Zhuhai 519000; 3. Suzhou Ruixun Biotech Co. Limited, Suzhou  
215000; 4. The First Affiliated Hospital, Xi'an Jiaotong University, Xi'an 710061, China)

**ABSTRACT:** Rapid and accurate detection of contagium virus is a key tool for controlling the outbreak. As the number of patients infected with 2019 novel coronavirus (2019-nCoV or

SARS-CoV-2) pneumonia is increasing and the epidemic is spreading, many hospitals, laboratories and pharmaceutical companies have developed diagnostic reagents that can detect SARS-CoV-2. Currently, genome sequencing or real-time PCR is the approach to detect SARS-CoV-2. However, there have been reported the clinical patients with pneumonia of negative results by SARS-CoV-2 nucleic acid detection but positive on the radiological CT scan. . It brings new challenges to the clinical practitioners since it is unclear for the specificity of CT to diagnose the 2019-nCoV pneumonia (NCP). Therefore, how to improve the diagnostic efficiency and the sensitivity of molecular detection remains to be solved. To analyze the variations in viral genomic sequence may be helpful in guiding the prevention and treatment of diseases infected by SARS-CoV-2. At present, controversy still exists over the source of SARS-CoV-2 even though the probable origin is bat in nature. This article summarizes the recent findings of SARS-CoV-2 from genetic, virological and evolutionary biological perspectives. By comparing the genomic variations among SARS-CoV-2 infected patients, we hope the findings can be used in the viral detection and potential antiviral therapy. Also, it will be great if we can trace the spread of the virus from one person to another, which will be a much effective way to predict and control the spread from the virus-carrier without symptoms at the incubation period to the others.

**KEY WORDS:** 2019-nCoV; sequence variation; virus strain; molecular typing

2019年12月以来,湖北省武汉市陆续发现了新型冠状病毒(以下简称2019-nCoV)感染导致新型冠状病毒肺炎(以下简称COVID-19)的患者,随之在全国和境外也相继出现此类病例<sup>[1, 2]</sup>。COVID-19已纳入《中华人民共和国传染病防治法》规定的乙类传染病,按照甲类传染病管理。传染源主要来源于2019-nCoV感染者,无症状携带者也是传染源。临床数据表明,2019-nCoV感染危重患者多见于老年人、糖尿病、高血压等有基础疾病的人群<sup>[3]</sup>。COVID-19流行病学特点表现为人群的普遍易感性。潜伏期为1~14天,大多为3~7天。临床主要症状为发热、乏力和干咳,危重患者与SARS感染相似,出现急性呼吸窘迫综合征、脓毒症休克和DIC等,预后较差<sup>[3, 4]</sup>。COVID-19已引起全球高度关注,很多国家已采取积极措施预防疫情的感染,避免疫情进一步蔓延<sup>[5, 6]</sup>。从分子水平对传染病的进行快速准确的诊断,了解病原的致病机制,是控制疫情爆发的关键工具<sup>[7]</sup>。目前,2019-nCoV基因组序列已经绘制完成(NCBI BioProject: PRJNA485481)。结合流行病学史、临床表现、胸部CT和核酸检测确诊COVID-19,但2019-nCoV精准快速诊断方法仍需要优化,亟需开发抗

2019-nCoV 的药物和疫苗<sup>[8]</sup>。本文从遗传学、基因组学、病毒学和演化生物学的角度,对 2019-nCoV 基因组变异与分子诊断的近期研究进展进行论述分析,通过比较 2019-nCoV 与 SARS 基因组以及感染不同患者的 2019-nCoV 基因组变异,为精准快速分子诊断、抗病毒药物靶点筛选提供依据,也为 2019-nCoV 追踪溯源、核酸分子诊断方案设计、临床防治提供新思路。

## 1 2019-nCoV 溯源、变异与进化基础

LLOYD 等<sup>[9]</sup>研究表明,携带未知病毒的蝙蝠、灵长类动物以及其他哺乳动物与人类频繁接触的地区,可能是新型传染源高发的地区,如巴西的寨卡病毒和西非的埃博拉病毒导致的疫情<sup>[10]</sup>。人类将不同的野生动物聚集在非自然环境中,如圈养大量野生动物,其携带的病毒等微生物可能产生致病的基因变异,通过排泄物、分泌物等方式传递给人类,从而导致人类疾病的发生。

冠状病毒于 1937 年从鸡分离出来,病毒颗粒的直径为 60~200nm,平均直径为 100nm,呈球形或椭圆形,具有多形性。病毒有包膜,包膜上存在棘突,整个病毒像日冕,不同的冠状病毒的棘突有明显的差异。人类冠状病毒(human coronavirus, HCoV)感染引起临床症状轻微、致病性弱。目前已发现 6 种(HCoV-229E、HCoV-OC43、HCoV-NL63、HCoV-HKU1、SARS-CoV 和 MERS-CoV)可以感染人类的冠状病毒,2019-nCoV 是第 7 种<sup>[11]</sup>。

2019-nCoV 在系统分类上属冠状病毒科(coronaviridae)冠状病毒 $\beta$ 属(coronavirus),本文收集的 2019-nCoV 全基因组数据来源于 NCBI 和新型冠状病毒信息库(<https://bigd.big.ac.cn/ncov>),相关项目编号为 PRJCA002165<sup>[5,11]</sup>。

通过上述基因组数据,本研究采用最大似然法构建进化树。结果显示,2019-nCoV 与 SARS 和蝙蝠冠状病毒亲缘关系最为接近(图 1)<sup>[12]</sup>。2019-nCoV 与蝙蝠冠状病毒、SARS 拥有共同的祖先序列,全基因组序列上有重要的相似性。2019-nCoV 基因组全序列与蝙蝠冠状病毒基因组全序列同源性达 87.99%;2019-nCoV 基因组全序列与 SARS 基因组全序列同源性为 79.5%<sup>[13]</sup>。

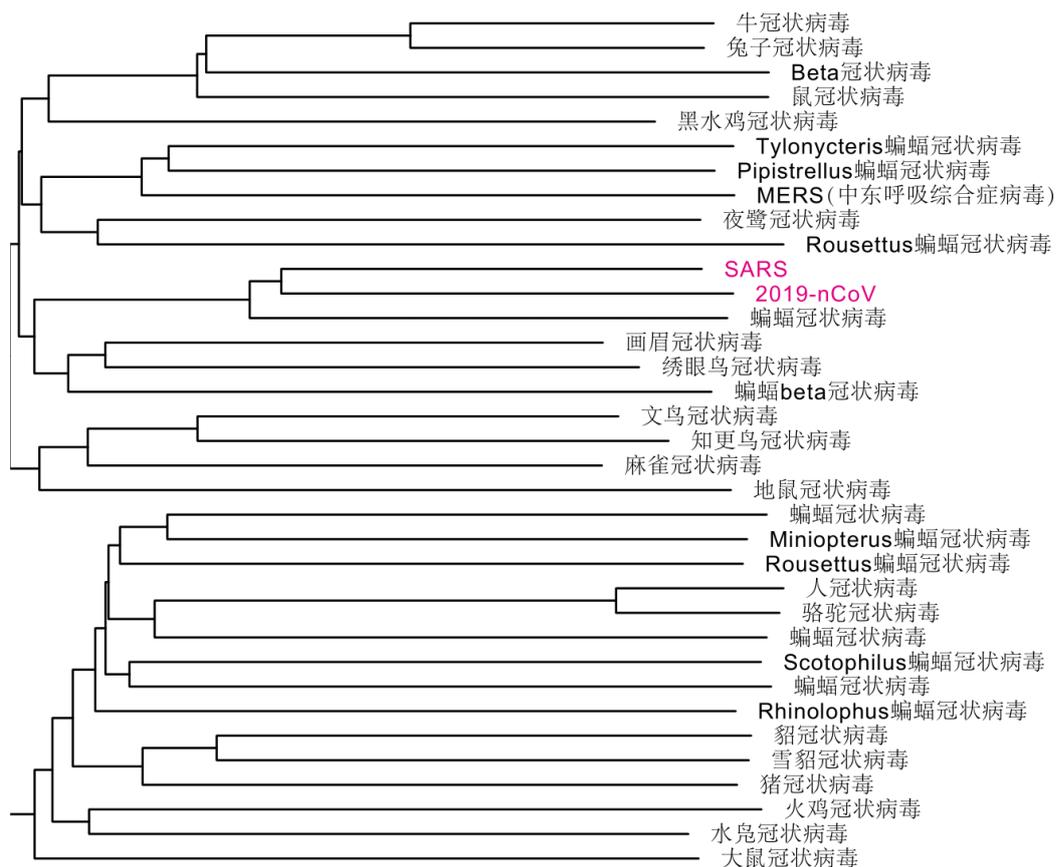


图 1 2019-nCoV 与其它冠状病毒的进化树（最大似然法）

Fig. 1 Phylogenetic tree of 2019-nCoV and other coronavirus (Maximum Likelihood Estimate)

## 2 2019-nCoV 的基因组结构、功能与致病亚型

NCBI 发布了 2019-nCoV 全基因组序列信息，长度为 29.9kb。本研究根据上述结果和 GUAN 等<sup>[14]</sup>发布的 SARS 基因组数据绘制了基因组结构比较图（图 2），基因组结构和注释显示，2019-nCoV 有 11 个功能基因（表 1），其中 3 个同源于 SARS，7 个同源于蝙蝠冠状病毒，1 个功能未知。

表 1. 2019-nCoV 蛋白编码基因注释信息

Tab. 1 Protein coding genes on 2019-nCoV

基因名字	序列长度	基因注释描述	同源序列
orf1ab	13218	Replicase polyprotein 1a	SARS
S	3822	Spike glycoprotein	SARS

ORF3a	828	Protein 3	Bat coronavirus
ORF4	228	Envelope small membrane protein	SARS
ORF5	669	Membrane protein	Bat coronavirus
ORF6	186	Non-structural protein 6	Bat coronavirus
ORF7a	366	Protein 7a	Bat coronavirus
ORF7b	132	Non-structural protein 7b	Bat coronavirus
ORF8	366	Non-structural protein 8	Bat coronavirus
ORF9	1260	Nucleoprotein	Bat coronavirus
ORF10	117	Unknown	NA

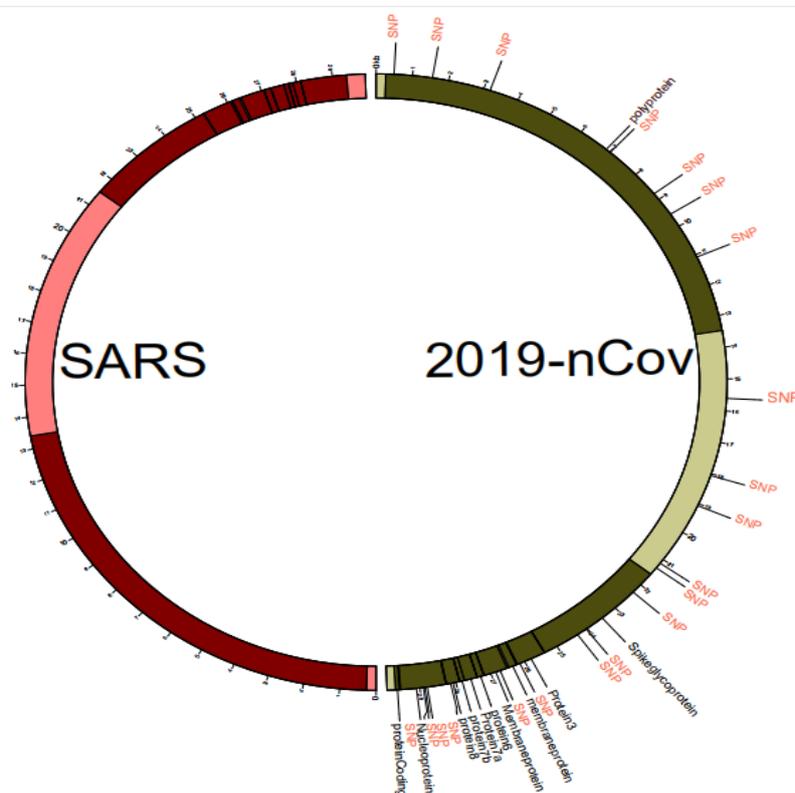


图 2 2019-nCoV 全基因组结构图、SNP 位点、蛋白编码基因在基因组分布

Fig. 2 Genomic structure, SNP sites and genes of 2019-nCoV

本文利用国家生物信息中心（CNGB）/国家基因组科学数据中心（NGDC）发布的 2019-nCoV 数据进一步分析了其基因组变异和临床表型有关个体致病亚型，发现 2019-nCoV 基因组存在毒株差异，表现为蛋白编码基因区域含 17 个单核苷酸多态性位点 SNP，可能引起致病基因变异。非蛋白编码区域含 5 个 SNP。图 3 显示了这些多态性 SNP 位点的基因组



图 4 2019-nCoV 的 ORF3a 和 ORF8 基因的蛋白序列变异

Fig. 4 Protein sequence variations of ORF3a and ORF8 in 2019-nCoVs

MN988713 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
MN994467 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
MN997409 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
MN985325 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
CNA0007335 Wuhan	-----TTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
NMDC60013002-07 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
MT007544 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
MN994468 Wuhan	ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTGTAGATCT
	*****
MN988713 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
MN994467 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
MN997409 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
MN985325 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
CNA0007335 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
NMDC60013002-07 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
MT007544 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTA-----T
MN994468 Wuhan	GGGAGGACTTGAAAGAGCCACCACATTTTCACCGAGGCCACGCGGAGTACGATCGAGTGT
	***** *

图 5 2019-nCoV 不同毒株序列变化

Fig. 5 Sequence variation of 2019-nCoVs

病毒的遗传变异，是两个不同的病毒株通过抗原转移或者基因重排，形成新的单倍型，病毒 RNA 的 SNP 无论是在病毒自身序列还是在其靶序列都可能存在。研究不同病毒 RNA 和靶序列的多态性，可以更好地达到干扰或过表达病毒 RNA 的效果，同时还可以有助于了解疾病的发生机制。基于发表的临床基因组数据<sup>[2, 3]</sup>，研究 2019-nCoV 的 SNP 的差异显示可能形成了新的单倍型（图 3）这样就可以用单倍型把不同病人区分，累计单倍型越多，越有意义，就可以分子分型临床表型，结合临床筛选单倍型，指导用药。因此，基于不同的临床表型，通过高通量快速准确的数字 PCR 技术和全基因组分型技术，得到更多的亚型的序列多态性数据，建立与疾病亚型对应的亚型，这将对于指导临床治疗具有很大价值。

### 3 2019-nCoV 核酸分子诊断进展

目前临床上出现了一些胸部 CT 检测阳性，而 2019-nCoV 核酸检测阴性的病例情况，考

考虑到影像学对于 NCP 诊断的特异性尚无数据,如何提高此类患者的诊断准确性和时效性,给一线的医疗人员带来了巨大挑战。按照国家卫生健康委员会《新型冠状病毒感染的肺炎诊疗方案(试行第五版)》要求,定义相关病例如下。①疑似病例:有流行病学史,且符合临床表现中任意 2 条。无明确流行病学史的,符合临床表现中的 3 条;②临床诊断病例:疑似病例具有肺炎影像学特征者;③确诊病例:临床诊断病例或疑似病例,具备以下病原学证据之一者:一是呼吸道标本或血液标本实时荧光 RT-PCR 检测 2019-nCoV 核酸阳性;二是呼吸道标本或血液标本病毒基因测序,与已知的 2019-nCoV 高度同源。

因此,临床实践证实,2019-nCoV 核酸检测必须结合临床、影像学、流行病学资料进行综合分析,才能确诊 COVID-19 患者;同时需要开发更灵敏和更特异的检测技术和靶向的治疗方法,以满足临床上的防控需求。目前可检测病毒的技术或者方法有定量 PCR、多重 PCR、反转录等,主要针对病毒正链单股 RNA、抗体等<sup>[15,16]</sup>。多个实验室和公司开发了不同的 2019-nCoV 的诊断试剂。但目前对不同单位试剂盒的性能尚无第三方评价数据,随着检测结果的假阴性报道<sup>[17]</sup>,评估试剂盒的灵敏度、特异度以及准确性已迫在眉睫。

病毒 RNA 序列的多态性研究,旨在全基因组序列中设计引物对,扩增出基因组序列变异特征,识别包括 2019-nCoV、SARS-CoV 和 MERS-CoV 等。第三代 Oxford Nanopore 纳米孔测序技术可以直接捕获 RNA 揭示 2019-nCoV 基因组的变异,不需要 RNA 的逆转录和扩增,可以直接对冠状病毒基因组进行 RNA 测序,从而检测病毒基因组中的 SNP 位点。因此,通过增加 COVID-19 患者样品内 RNA 亚型诊断的可行性,可以为基于亚型的病毒确认和毒株变异分析提供可能。

目前,微流控数字 PCR 将微升水平的 PCR 反应体系分割成数万至数十万个纳升级甚至是皮升为单位的微小反应单元,每个反应单元中不含靶标 DNA 或者含有 1 个至多个靶标 DNA,对各独立反应单元进行 PCR 扩增,检测最终各反应单元内是否有目标扩增产物,进而通过泊松分布即可计算出初始样本中靶标 DNA 的绝对拷贝数量,整个过程耗时一般为 45~90 分钟。数字 PCR 可检测出 0.01% 的突变率,甚至可以检出 0.001% 以下的突变率。相比而言,RT-PCR 或者被称作定量 PCR (qPCR) 反应的检测极限在 10<sup>1</sup> 个靶标数量级上,无法检测单分子靶标。当待测靶标丰度很低(1%~0.1%)时,qPCR 同样难以从高背景信号的前提下分辨出目标信息。而新一代测序(next generation sequencing, NGS)可检测 1%~5% 的低丰度突变,更低的突变丰度则需要依赖数字 PCR。理论上数字 PCR 可检测到单分子级别的靶标 DNA,是目前所有核酸检测方法中灵敏度最高的技术之一。因此,基于数字 PCR 可快速检出痕量病毒,为有效筛选病毒携带者提供可能。

## 小结

目前,已有大量的基于全基因组测序的和临床防治的 2019-nCoV 研究<sup>[2, 11, 13]</sup>,但仍缺乏精准而又快速的检测技术和针对 2019-nCoV 的靶向抗病毒药物和预防疫苗。2011 年至 2018 年,世界卫生组织追踪了 172 个国家的 1483 起传染病疫情,数据预测表明仍存在未知的新型疾病出现的可能性。因此,从遗传学、基因组学、病毒学和演化生物学的角度认识 2019-nCoV,对病毒传染源进行追踪溯源,基于病毒基因组信息了解新型冠状病毒在动物、人类和环境中传播过程,将会为控制传染源和疫情预警提供最有力证据。结合病毒演化机制、致病基因变异、临床表型、医学影像、分子病理学,将实验室检测的基因型和 COVID-19 患者临床表型统一,建立 2019-nCoV 基于临床表型的毒株亚型和诊断标准,对于指导临床有效防治该疾病具有非常重要的价值和意义。

致谢:感谢李生斌教授在文章设计、基因组分析、论文撰写给予指导,感谢常辽博士在基因组变异和临床表型数据分析、绘图等工作。

## 参考文献:

- [1] LI Q, GUAN X, WU P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia[J]. The New England Journal of Medicine, 2020, DOI:10.1056/NEJMoa2001316.
- [2] HUANG C, WANG Y, LI X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China[J]. The Lancet, 2020, DOI:10.1016/S0140-6736(20)30183-5.
- [3] CHEN N, ZHOU M, DONG X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study[J]. The Lancet, 2020, 395(10223): 507-513.
- [4] JIANG S, XIA S, YING T, et al. A novel coronavirus (2019-nCoV) causing pneumonia-associated respiratory syndrome[J]. Cellular & Molecular Immunology, 2020, DOI:10.1038/s41423-020-0372-4.
- [5] LI X, WANG W, ZHAO X, et al. Transmission dynamics and evolutionary history of 2019-nCoV[J]. Journal of Medical Virology, 2020, DOI:10.1002/jmv.25701.
- [6] HUI DS et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health--the latest 2019 novel coronavirus outbreak in Wuhan, China[J]. International Journal of

- Infectious Diseases, 2020, 2020(91): 264-266.
- [7] WANG FS, ZHANG C. What to do next to control the 2019-nCoV epidemic?[J]. The Lancet, 2020, 395(10222): 391-393.
- [8] WU JT, LENUG K, LEUNG GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study[J]. The Lancet, 2020, DOI:10.1016/S0140-6736(20)30260-9.
- [9] LLOYD JO et al. Superspreading and the effect of individual variation on disease emergence[J]. Nature, 2005, 2005(438): 355-359.
- [10] STEPHEN KG et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak[J]. Science, 2014, 345(6202): 1369-1372.
- [11] WU F, ZHAO S, YU B, et al. A new coronavirus associated with human respiratory disease in China[J]. Nature, 2020, DOI:10.1038/s41586-020-2008-3.
- [12] LI W, SHI Z, YU M, et al. . Bats are natural reservoirs of SARS-like coronaviruses[J]. Science 2005, 2005(310): 676-679.
- [13] ZHOU P, YANG X, WANG XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J]. Nature, 2020, DOI:10.1038/s41586-020-2012-7.
- [14] GUAN Y, ZHENG BJ, HE YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China[J]. Science 2003, 2003(302): 276-278.
- [15] CPRMAN VM et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR[J]. Eurosurveillance, 2020, 25(3): 23.
- [16] WANG M, CAO R, ZHANG L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro[J]. Cell research, 2020, DOI:10.1038/s41422-020-0282-0.
- [17] 莫茜, 秦炜, 傅启华, 关明. 正确认识新冠病毒核酸检测的影响因素[J]. 中华检验医学杂志, 2020, 43(00): E002-E002.
- MO X, QIN W, FU QH, et al. Understanding the influence factors in viral nucleic acid test of 2019 novel coronavirus (2019-nCoV)[J]. Chin J Lab Med, 2020, 43(00): E002-E002.

编辑 卓选鹏