

北大汇丰

PHBS FINANCIAL FRONTIER DIALOGUE

金融前沿对话

2020年第7期 总第54期



PHBS HFRI
北京大学汇丰金融研究院

主办单位：北京大学汇丰金融研究院
院长：海闻
执行院长：巴曙松
秘书长：本力
编辑：都闻心（执行） 曹明明 方培豪 朱伟豪

北京大学汇丰金融研究院简介

北京大学汇丰金融研究院（The HSBC Financial Research Institute at Peking University，缩写 HFRI）成立于 2008 年 12 月，研究院接受汇丰银行慈善基金会资助，致力于加强国内外著名高校、金融研究机构，以及知名金融学者之间的交流，构建开放的学术、政策交流平台，旨在提高中国金融理论与政策的研究水平，促进学术繁荣与发展，加强与政府决策部门的联系与合作，为政府决策提供参考意见，为国际金融体制改革和中国金融业的发展做出贡献。

北京大学汇丰金融研究院院长为北京大学校务委员会副主任、北京大学汇丰商学院院长海闻教授，执行院长为中国银行业协会首席经济学家、中国宏观经济学会副会长巴曙松教授。

基因组学与人工智能：应对传染病的挑战

【对话主持】

巴曙松（北京大学汇丰金融研究院执行院长、中国银行业协会首席经济学家、中国宏观经济学会副会长）

【特邀嘉宾】

周巍（生捷科技首席执行官）

我开始关注传染病是在埃默里大学做博士后的时候，当时我的科研项目是大肠癌症，但我有很多朋友在美国疾控中心（CDC）工作，他们主要研究传染病。我和这些传染病“病毒猎手们”在一起，学到了不少关于病毒的故事。2007 年我开始做传染病检测芯片，与中国 CDC 探讨过关于传染病芯片的合作项目，多年来我对传染病一直很感兴趣。从风投角度看，这个行业目前并不是很热门，但是我觉得这次新冠病毒疫情之后，投资界对这个行业的看法将有所改变。

一、病毒是一个小程序

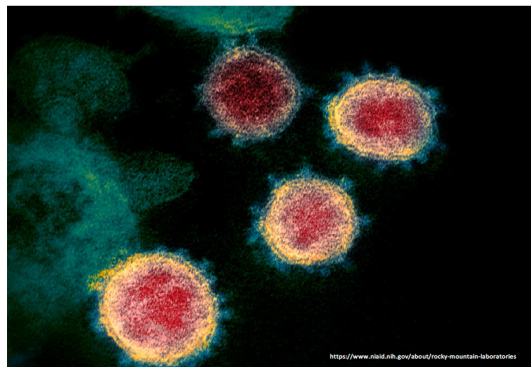


图 1 SARS-CoV-2 病毒图像

图 1 是新冠病毒的电子显微镜照片。病毒几乎是一个全生物信息的载体。病毒就像一个 U 盘，里面有芯片，外面有包装，还要有 USB 接口。病毒的主要零部件是 RNA 核糖核苷酸（存放程序），边缘上有包膜蛋白（包装），还有一个很重要的是刺突蛋白（接口）。病毒没有自己的 CPU，所以无法自己复制蛋白，需要把自己的信息输入到细胞中，让宿主细胞帮助其复制繁殖。在病毒完成复制之后，经过组装过程，再从细胞中钻出来。从图 2 可以看出，冠状病毒（黄色）从宿主细胞中钻出来。

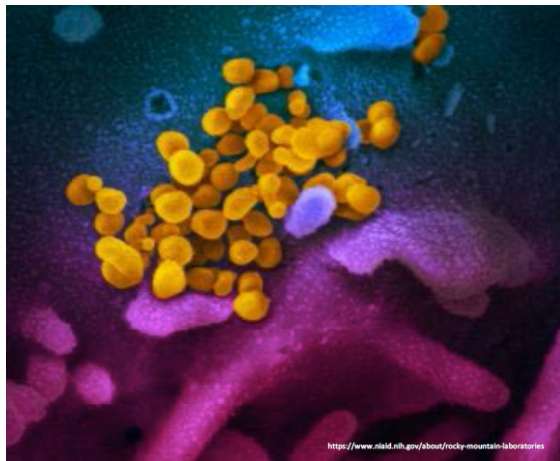


图 2 SARS-CoV-2 病毒（黄色）完成繁殖

基因组就是核糖核苷酸，由 4 个字母组成，字母组合的变化加上修饰，或者说 DNA 序列，编码了地球上几乎所有生物的多样性。图 3 是新冠肺炎病毒的 RNA 序列，代表了病毒的整个基因组信息。

随后加拿大研究人员宣布首次成功测序冠状病毒基因组。而此次新冠病毒从发现病例到基因组测序完成，整个过程速度非常快。2019年12月发现第一个病例，2020年1月5日全基因组序列公布并共享。目前病毒已经被分离，但是还没看到猴子实验（寄主实验），中国和美国CDC都在进行相关实验。

新冠病毒与其他冠状病毒的相似性很高，因此之前的研究结论都可以使用。比如，我们知道冠状病毒由哪些蛋白质组成，可以看到病毒零部件的组成情况，也可以通过病毒家谱掌握其进化过程。这次新冠病毒基因组公布的不同之处是，数据完全公开，公众都可以拿到，分析软件也是开源的。大家可以使用 blast 网站（<https://blast.ncbi.nlm.nih.gov/Blast.cgi>）进行病毒序列比较。可以发现，该病毒基因组和 SARS 病毒的相似性大概在 82.3% 左右。最近网上有很多关于新冠病毒来源的讨论，其中存在很多低质量的文章，没有通过科学家的同行评议就发表了。我找到了一篇比较值得信任的文章（见图 4），这些作者是很有名的病毒学家。这篇文章中有几个重要的结论，比如从基因组序列测算估计，该病毒是在 2019 年 11 月下旬到 12 月初这段时间内出现的。

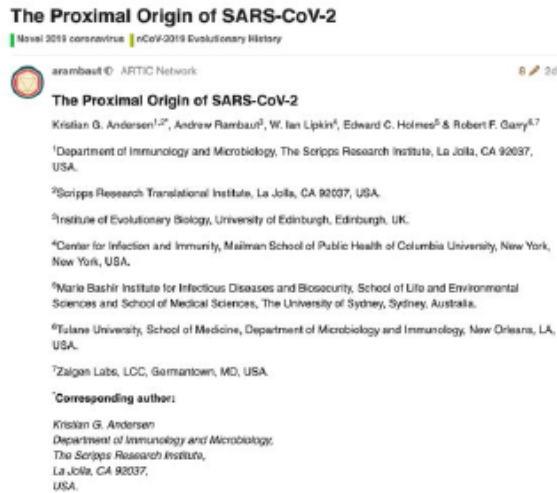


图 4 SARS-COV-2 病毒基因组分析

三、基因组学与人工智能及其应用

目前为止，对冠状病毒基因组的分析主要采用的是传统生物信息分析工具，但是人工智能的应用越来越重要。人工智能的主要应用领域是图像分析和语言分析。病毒基因组主要是字母排列组合，其实和语言非常类似。所以人工智能算法对分析基因序列非常有用。图像分析的人工智能工具，也已经应用于基因序列分析，这是因为核酸序列可以转换成图像，之后就可以用人工智能分析软件来进行生物数据分析。

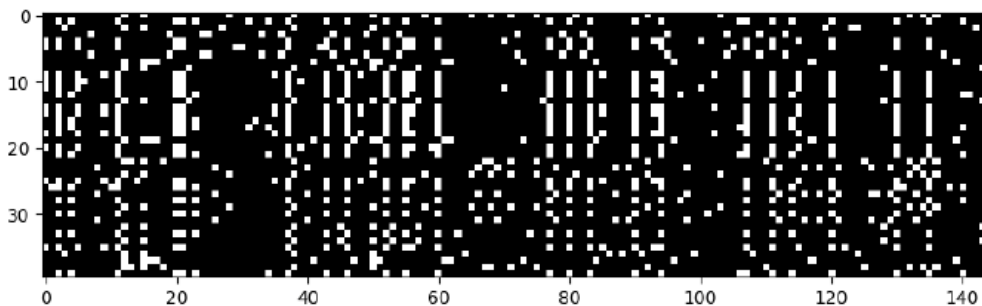


图 5 核酸序列转换成图像

我举两个人工智能在生物分析方面的应用。第一个例子是利用机器学习预测携带人畜共患病病毒的啮齿动物。第二个例子是利用机器学习方法从测序数据中发现新病毒。我们在检测人类全基因组时会测出一些细菌和病毒，科学家用机器学习的方法在人的基因组中找到了6000种新病毒。近年美国CDC举行了一个流感预测挑战赛，公众可以使用公开数据对当年流感发生的概率进行预测判断。

大家可能对基于基因组学的疫苗开发比较感兴趣。目前在开发的疫苗和基因组学有很大相关性。科学家可以在体外合成病毒蛋白，进而用来配置疫苗。在过去十几年中，很多公司在开发RNA疫苗。冠状病毒的主要零部件是RNA，RNA进入人的细胞中之后就开始繁殖复制。利用这个原理，如果将RNA中的一小块放在细胞中，人类细胞会合成抗原，从而起到疫苗的作用。

四、病毒监控和检测是关键

科学家正在通过人工智能搜索冠状病毒的潜在治疗方法。实际上抗病毒药物的开发需要很久的时间。虽然疫苗和抗病毒药非常重要，但是病毒的监控和检测是最关键的。人跟所有动物一样，人类的身体充满了病毒。人体约有40万亿细菌，尽管我们没有确切知道人体中有多少种病毒，但是在自然界中，病毒的总数通常是细菌的10倍左右，因此据估计，人体中约有380万亿病毒。

我们的基因组中也记录了与病毒相互作用的历史。转座子（Transposon）是整合入人类基因组的病毒留下的遗迹，人类基因组

中有多达 44% 的 DNA 是转座子, 其中最常见两种 Alu 和 LINE (L1) 就占人类基因组的 30% 以上。如果我们提高算法的敏感性, 几乎 80% 的人类基因组可能与病毒有关。相比之下, 人类基因组编码蛋白质只占 2%。人内源性逆转录病毒(HERVs)至少占人类基因组的 8%。

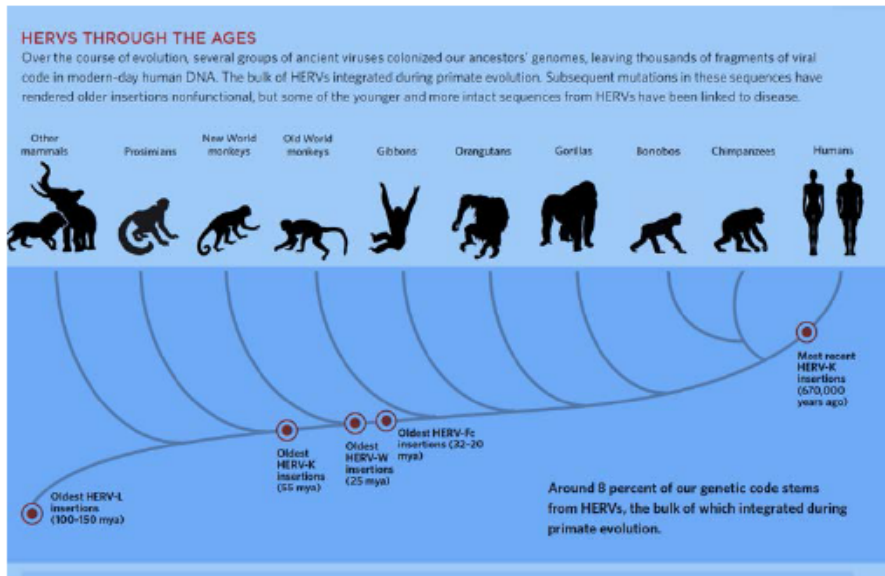


图 6 动物进化过程中病毒逐渐进入人类基因组

大多数动物都带有大量病毒, 其中蝙蝠尤为特殊。蝙蝠基因组的特殊设计使得其免疫系统可以容忍病毒。为了应对蝙蝠强大的免疫系统, 病毒必须进化出快速在细胞中传播的能力, 这可能也是许多来自蝙蝠的病毒具有高致病性的原因。

短期来看, 比较大的传染病爆发每隔几年就有一次, 但是实际上如果从人类发展历史来看, 有无穷多次病毒和疾病的发生。病毒也是可以进化的, 疫苗和药物无法彻底控制病毒。例如, 季节性流感已经存在了很长时间, 我们仍然没有 100% 有效的疫苗和药物。对于传染病的预防, 早期病毒的监控和识别非常重要, 可以给我们一些时间来

遏制病毒大流行，同时开发疫苗。

五、个性化的预防和治疗

关于新冠病毒核酸的临床诊断检测方法，现在主要有两种：一是基因测序，如果样品没有污染，测序结果是非常精确的。但是问题是速度慢、成本高、通量低，测序结果的生物信息学解释很复杂。二是核酸诊断，主要用的是实时逆转录 PCR，相对便宜，通量极高，速度快（30 分钟-2 小时，不包括样品制备）。但会受到许多因素影响，容易出现假阴性，很难检测到病毒突变，并且难以同时检测多种病原体。

那么为什么很难对普通人群进行病毒感染普检？假设一种检测方法有 5% 的假阳性，如果感染率为 1.5%，那么在 1000 人中有 15 人受到感染。大约 65 人将测试呈阳性，50 人为假阳性，测试阳性的人中只有 23% 确实是阳性，77% 的人为假阳性。从美国 CDC 流感快速检测的数据看，如果检测方法准确性在 98%，感染率在 2.5%，假阳性的概率为 44-61%，因此普查的效果会很差。

If Influenza Prevalence is...	And Specificity is...	Then PPV is...	False Pos. rate ¹ is...
VERY LOW (2.5%)	MODERATE (80%)	VERY LOW (6-12%)	VERY HIGH (88-94%)
VERY LOW (2.5%)	HIGH (98%)	LOW (39-56%)	HIGH (44-61%)
MODERATE (20%)	MODERATE (80%)	LOW (38-56%)	HIGH (44-62%)
MODERATE (20%)	HIGH (98%)	HIGH (86-93%)	LOW (7-14%)

<https://www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm>

图 7 美国 CDC 流感快速检测数据

从目前新冠病毒核酸临床诊断检测结果看，检出率只有 30-50%，

美国 CDC 的 RT-PCR 试剂盒也有问题。由于时间紧迫，很多试剂盒还没得到充分验证就被送到一线应用。随着套件设计和质量的提高以及操作人员的培训强化，假阴性率应会下降。

为什么相同的病毒会对人产生不同的结果？基因组的区别起到了很重要的作用。当年在猪流感出来时，欧洲科学家开始寻找一些基因变异，他们发现有一个基因，携带该基因的动物对猪流感非常敏感，会出现较严重的临床后果。禽流感出现时，香港科学家发现，如果将某两个基因型整合，可以区分并预判出较严重的病例。

Identification of *TMPRSS2* as a Susceptibility Gene for Severe 2009 Pandemic A(H1N1) Influenza and A(H7N9) Influenza

Zhongshan Cheng,^{1,4} Jie Zhou,^{1,2,3,4} Kelvin Kai-Wang To,^{1,2,3,4} Hin Chu,^{1,2,3} Cun Li,¹ Dong Wang,¹ Dong Yang,¹ Shufa Zheng,⁵ Ke Hao,¹ Yohan Bossé,⁶ Ma'en Obeidat,³ Corry-Anke Brandsma,¹⁰ You-Qiang Song,³ Yu Chen,⁶ Bo-Jian Zheng,^{1,2,3} Lanjuan Li,⁴ and Kwok-Yung Yuen,^{1,2,3,4}

¹Department of Microbiology, ²Research Centre of Infection and Immunology, ³State Key Laboratory of Emerging Infectious Diseases, ⁴Carol Yu Centre for Infection, and ⁵Department of Biochemistry, The University of Hong Kong, Pok Fu Lam; ⁶State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, First Affiliated Hospital, College of Medicine, Zhejiang University, China; ⁷Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York; ⁸Department of Molecular Medicine, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Laval University, and ⁹University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada; and ¹⁰Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, The Netherlands

The genetic predisposition to severe A(H1N1)2009 (A[H1N1]pdm09) influenza was evaluated in 409 patients, including 162 cases with severe infection and 247 controls with mild infection. We prioritized candidate var-

图 8 对禽流感敏感基因的检测识别研究

如果我们能够准确地检测到所有病原体，就会减少误诊的概率。现在采用 CT 作为辅助的诊断依据，是非常明智的做法。通过病患基因组的测序，可以对临床结果进行预测，有利于提供个性化的护理和治疗。

六、问答环节

Q1: 您怎么评价中国科学家在此次疫情中的表现？您提到目前核酸检测检出率仅为 30-50%，这个指标是确诊双阳的关键指标，这

个结果是否说明国内确诊人数也只有实际患病人数的 30-50%?

A1: 从文章发表来看，中国科学家做了大量的工作，他们表现非常优秀。关于假阳性问题，通过与一线医生交流了解到，目前假阳性比例还是比较高的，一线医生使用 CT 来辅助判断，是非常好的方法，检测准确率会逐步提升。

Q2: 因为 SNP（单核苷酸多态性）的不同，是否不同种族群体，对某些病毒更敏感、死亡率更高？

A2: 欧洲科学家对流感敏感性所做的研究，发现那个易感基因在白人中出现频率很低，但是在亚裔种群中很高。但是这个不是绝对的，可能针对另一个病毒，易感基因在白人种群中出现频率更高。

Q3: 目前这个病致死原因是不是与每个人的免疫力直接相关？

A3: 应该是有关系的。从流感来看，流感的敏感基因基本都和免疫系统有关系。

Q4: 这次 n-Coronavirus，因为研究疫苗的需要，在实验室合成的可能性是否存在？

A4: 根据知名病毒学家做的基因组分析，看不出该病毒是人工合成的样子。这种病毒是可以人工合成的，但是从生物信息上没看出任何迹象。

本文根据北京大学汇丰金融研究院执行院长巴曙松教授发起并主持的“全球市场与中国连线”第三百一十六期（2020年2月9日）内容整理而成，特邀嘉宾为生捷科技首席执行官周巍博士。

周巍，生物学博士，法律博士，基因组学芯片生产商生捷科技的首席执行官，他同时也是美国最大的眼科临床诊断实验室 MVL 的首席执行官。生捷科技生产超高密度 DNA 芯片用于基因分型、3D 测序和病原体检测。在创立生捷之前，他曾担任基因组学技术先驱 Affymetrix（现为赛默飞的一部分）分管知识产权及尖端技术的高级副总裁。周巍博士是斯坦福大学法学博士，埃默里大学医学院博士后，弗吉尼亚理工大学博士。他目前还担任 Fralin 生物医学研究所的顾问。

【免责声明】

“全球市场与中国连线”为中国与全球市场间内部专业高端金融交流平台。本期报告由巴曙松教授和郑子龙博士共同整理，未经嘉宾本人审阅，文中观点仅代表嘉宾个人观点，不代表任何机构的意见，也不构成投资建议。

本文版权为“全球市场与中国连线”会议秘书处所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复印、发表或引用本文的任何部分。



PHBS HFRI
北京大学汇丰金融研究院

