

# 语言模型竟能追踪病毒

## 可预测突变识别疫苗有效目标

■本报记者 唐凤

病毒可能比人们想象的更狡猾，它能不断伪装自己，逃避“疫苗猎人”的追捕。

一直以来，研制流感、艾滋病疫苗之所以如此困难，原因之一便是这些病毒的变异非常迅速。这使得它们可以通过一种被称为病毒逃逸的过程，避开特定疫苗产生的抗体。

美国麻省理工学院研究人员现在设计了一种计算病毒逃逸的新模型。该模型基于最初用来分析语言的模型，可以预测病毒表面蛋白的哪些部分更容易发生突变，从而使病毒能够逃逸，也可以识别出不太可能发生突变的部分，使它们成为新疫苗的靶标。

“病毒逃逸是个大问题。”麻省理工学院计算机科学与人工智能实验室教授 Bonnie Berger 说，“流感病毒表面蛋白和 HIV（艾滋病病毒）包膜表面蛋白引发的病毒逃逸是造成目前没有通用疫苗的主要原因。这两种疾病每年都会导致数十万人死亡。”

在 1 月 15 日发表于《科学》的一项研究中，Berger 及同事确定了流感、HIV 和新冠病毒疫苗的可能目标。研究人员还将该模型用于研究最近在英国和南非出现的新病毒新变种。研究人员说，尚未经过同行评审的相关分析显示，这些病毒的基因序列应该被进一步调查，以确定它们是否有可能逃脱现有疫苗的影响。

### 病毒也有语言

不同类型的病毒以不同速度发生基因突变，HIV 和流感是突变最快的病毒之一。

“HIV 和流感病毒突变得很快，这是它们生物学复制的结果。例如，HIV 和流感遗传物质的复制机制很容易出错，从而导致突变。”该研究通讯作者、麻省理工学院生物工程助理教授 Bryan Bryson 接受《中国科学报》采访时表示。

为了让这些突变促进病毒逃逸，病毒必须改变其表面蛋白的形状，这样抗体就不能再与它们结合。然而，这种蛋白质不会发生使其失去功能的变化。

Berger、Bryson 以及研究生 Brian Hie 等人，决定使用一种被称为语言模型的计算模型进行建模。这种模型来自自然语言处理（NLP）领域，最初被设计用来分析语言模式，特别是某些单词同时出现的频率，以预测哪些单词可以用来完成一个句子，比如要补全“萨莉在（ ）中吃了鸡蛋”，NLP 模型可能预测“早餐”或“午餐”。

“我们对 NLP 语言模型的最新进展感到兴奋，这些模型可以通过学习原始文本理解人



快速变异使一些病毒能够避开特定疫苗产生的抗体。

图片来源: MIT

类语言。于是，我们认为，由于病毒最丰富的数据是原始的病毒序列，我们也可以通过训练语言模型从病毒序列数据集中学习非常复杂的模式。”Bryson 说。

当这种模型应用于生物信息，如基因序列时，其语法类似于确定特定序列编码的蛋白质是否具有功能，语义类似于确定蛋白质是否能够呈现新的形状，帮助它逃避抗体。因此，使病毒能够逃逸的突变必须保持序列的语法性，但同时能以一种有用的方式改变蛋白质的结构。

### 用序列训练模型

“病毒想要逃离人类的免疫系统，又不想因突变而死亡或无法复制，换句话说，它既想保持健康，又想充分伪装自己，以防被人体免疫系统检测到。”Hie 说。

为模拟这一过程，研究人员训练了一个 NLP 模型分析基因序列中的模式，该模型可以预测具有新功能但仍遵循蛋白质结构生物

学规则的新序列。这样建模的一个显著优点是它只需要序列信息，这比获得蛋白质结构容易得多。

此外，该模型可以在相对少量的信息上进行训练——在这项研究中，研究人员使用了 60000 条 HIV 序列、45000 条流感序列和 4000 条冠状病毒序列。

“语言模型非常强大，因为它们可以学习这个复杂的分布结构，并从序列变化中获得一些对功能的洞见。”Hie 告诉记者，“我们在每个氨基酸位置都有大量的病毒序列数据，模型通过训练数据学习氨基酸特性。”

该模型一旦经过训练，研究人员便能使用它来预测冠状病毒刺突蛋白、HIV 包膜蛋白和流感血凝素（HA）蛋白的序列变化，这些蛋白或多或少可能产生逃逸突变。

“发现看似不相关的科学分支之间的联系，可能会发展出加速一个分支研究的新方法。该研究提供了一个此类联系的示例。作者发现了病毒与自然语言之间的相似之处，进而提出了一种识别突变的强大新方法，这种突变

可以使病毒通过中和抗体逃脱识别。”未参与该研究的美国国家医学图书馆的 Teresa M. Przytycka 等人在相关评论文章中指出。

### 知己知彼 阻断逃逸

对于流感，该模型揭示了最不可能发生突变和产生病毒逃逸的序列是在 HA 蛋白的茎部。这与最近研究一致，表明针对 HA 茎部的抗体几乎可以提供全面的保护，以对抗任何流感病毒株。

在对 HIV 的研究中，研究人员发现，该蛋白的 V1-V2 高变区域有许多可能的逃逸突变，这与之前的研究结果一致，他们还发现了逃逸概率较低的序列。

该模型对冠状病毒的分析表明，被称为 S2 亚基的刺突蛋白的一部分最不可能产生逃逸突变。但新冠病毒变异的速度仍是一个问题，因此目前部署的抗击新冠肺炎大流行的疫苗在多长时间内保持有效性尚不清楚。

“目前，对于新冠病毒，我们认为该模型可以迅速标记出与以前看到的病毒序列有本质区别的新序列，以便在实验室进行下一步测试。”Berger 告诉《中国科学报》，“你可以想象，模型能检查每一个新序列，而改变超过一定阈值的序列就需要在实验室中展开进一步研究。”

初步证据表明，这种病毒的变异速度不像流感或 HIV 那么快。然而，研究人员最近发现了在新加坡、南非和马来西亚出现的新突变，他们认为应该对潜在的病毒逃逸进行调查。

研究人员认为，目前面临的问题仍然是新冠病毒的变异速度有多快。

“该病毒种类繁多，控制其复制的生物机制因人而异，所以尽管它们有共同的特征，但仍需要对每种病毒进行专门研究，以了解其突变率。”Bryson 说，“我们最好的见解将来自于动物感染模型，其中完整的免疫反应是存在的，因此可以理解总的免疫压力是如何影响病毒突变率的。之后，我们需要对新冠病毒感染者的病毒进行测序，以识别这些病毒中存在的突变。”

此外，研究人员现在正与其他人合作，利用他们的模型确定癌症疫苗的可能目标，从而刺激人体自身免疫系统摧毁肿瘤。他们说，该模型还可以用来设计小分子药物，这种药物可能不会引发结核病等疾病的耐药性。

相关论文信息：

<https://doi.org/10.1126/science.abd7331>  
<https://doi.org/10.1126/science.abf6894>

# 世卫总干事：各国应从新冠疫情中汲取三个教训

据新华社电 世界卫生组织总干事谭德塞 1 月 18 日表示，新冠疫情持续时间已有一年，世卫组织所有会员国及世卫组织自身应从疫情中汲取三个教训，分别涉及对疫情的准备和应对、人与动物和地球的关系以及加强世卫组织建设等方面。

谭德塞在当天举行的世卫组织执委会会议上发言时表示，这三个教训适用于所有人，是未来“学习、变革、创新和成长所需要的”。

首先，在应对大流行的准备和应对方面，新冠疫情让世界上即使最富有、最强大的国家都感到惊讶和措手不及，这暴露出各国在应急准备方面普遍存在投入不足的问题。谭德塞认为，新冠大流行表明，以往根据专家对各国防疫准备状况的评估而推出的一些疫情应对工具已显不足。

其次，人们应该认识到，人类、动物和地球的健康

是紧密交织在一起的。由于近年来 70% 以上新发现的疾病与动物的传播有关，谭德塞敦促各国从根本上加强监测和管理人类、动物和生态系统之间的接触风险，以保护和促进人类健康。他同时呼吁解决影响人类、动物和地球之间关系的各种问题，包括森林砍伐、过度开垦、环境污染、气候变化等。

第三，世界需要一个强大的世卫组织。谭德塞表示，各会员国已确定世卫组织未来面临的最大障碍之一是可持续和可预测的筹资。为此，他要求世卫组织基金会在未来 3 年从新的渠道筹集 10 亿美元，将其中的 70% 至 80% 用于世卫组织自身，剩余资金用于其他公共卫生组织，重点是民间组织。

谭德塞表示，世界仍面临着前所未有的危险，但也有一个前所未有的机会，“让健康成为发展的脉搏，在此基础上建设一个更安全、更公平的世界”。（刘曲）

## 科学快讯

（选自 Science 杂志，2021 年 1 月 15 日出版）

### 植物根系通过限制乙烯扩散感知土壤压实度

用铁锹穿过压实的土壤很困难，植物的根在压实的土地上生长时似乎也有同样的问题。

作者发现不是物理阻力，而是信号通路抑制生长。挥发性植物激素乙烯会通过土壤扩散，但压实土壤会减少这种扩散，增加根部组织附近乙烯的浓度。过多乙烯引发的细胞信号级联阻止了根的生长。

因此，通过气体的扩散可读出土壤的压实度，以便为植物的根系生长寻找营养。

相关论文信息：

<https://doi.org/10.1126/science.abf3013>

### 重元素中 $\alpha$ 簇的形成

$\alpha$  衰变是铀等重元素的一种常见的放射性模式，它需要失去由两个质子和两个中子组成的粒子。尽管经过了一个多世纪的研究，这些  $\alpha$  粒子在稳定核和不稳定核中何时何地形成，仍然是一个悬而未解的问题。

作者用高能质子轰击了一系列稳定的锡同位素，并探测到喷出的  $\alpha$  粒子的丰度与质量数成反比。这一观察将  $\alpha$  粒子的积累与核表面的中子皮肤厚度联系起来，适用于从放射性衰变到量子动力学模型。

相关论文信息：

<https://doi.org/10.1126/science.abe4688>

### 一维偶极气体进入强相关预热态拓扑界

大多数非平衡的多粒子系统最终会达到热态。然而，一些一维量子系统并没有热化，尽管它们可能会进入长期的平衡态。

作者研究了这样一个由偶极原子组成的系统，该系统被限制在像二维阵列的细长管的光学势中。作者以规定的方式循环原子之间的接触相互作用，创造了越来越多的激发态非热量子多体态。

相关论文信息：

<https://doi.org/10.1126/science.abb4928>

### 羟基化作用让植物可化学防御害虫而不产生自毒作用

植物产生各种各样的分子来保护它们不受饥饿昆虫的侵害。作者分析了植物自我保护和自我毒害之间的平衡。

在野生烟草中，两种细胞色素 P450 酶在 17-羟基香叶酰芳樟醇二萜苷的生物合成途径中起作用，有助于防止有毒二萜衍生物在植物体内积累。

这些相同的二萜衍生物在昆虫食草动物摄入后形成，并通过抑制植物和昆虫鞘脂生物合成而引起毒性。

相关论文信息：

<https://doi.org/10.1126/science.abe4713>

（冯维维译）

## “小柯”秀

一个会写科学新闻的机器人

《细胞》

### 感染通过训练肠道菌群提高宿主抵抗力

美国国立卫生研究院 Yasmine Belkaid、Apollo Stacy 等研究人员合作发现，感染可通过训练肠道菌群提高宿主对于病原体的抵抗力。1 月 15 日，《细胞》在线发表了这一成果。

由于感染在整个进化过程中普遍存在，研究人员假设宿主能够采用各种策略通过感染引起的微生物群来获取自身益处，并更好地防御以后的感染。这一假设的推论是，定期接触病原体可能有益于微生物群产生最佳的抗菌能力。然而，在工业化社会中，人们通过改善的卫生条件来减少对某些种类病原体的接触，从而导致了微生物群落多样性和有益功能的急剧下降。与此相吻合的是，野生小鼠比实验室小鼠能够抵抗更多的感染，这表明具有感染史的小鼠表现出免疫力的增强，对

肠道菌群扰动也具有更高的抵抗力。

研究人员发现，有感染史的肠道菌群能够为宿主提供更强的抗感染力。这种长期的功能重塑与胆汁酸代谢改变有关，从而导致利用牛磺酸的生物类群扩增。值得注意的是，仅供应外源性牛磺酸足以诱导微生物群功能的这种改变并增强抵抗力。从机制上讲，牛磺酸可增强微生物群中硫化物的产生，硫化物是细胞呼吸的抑制剂，这是宿主被多种病原体入侵的关键。因此，硫化物的药物螯合干扰了微生物群的组成并促进了病原体的入侵。总之，这项工作表明，在感染之后，宿主可以利用牛磺酸作为营养物质来滋养和训练微生物群，从而增强其对随后感染的抵抗力。

相关论文信息：<https://doi.org/10.1016/j.cell.2020.12.011>

《自然》

### 伴侣蛋白介导的自噬可维持造血干细胞功能

日前出版的《自然》在线发表了美国阿尔伯特·爱因斯坦医学院 Ana Maria Cuervo 和 Britta Will 小组合作的研究成果，他们发现伴侣蛋白介导的自噬（CMA）维持了造血干细胞（HSC）的功能。

研究人员发现 CMA 参与维持成年小鼠 HSC 的功能。CMA 是一种溶酶体选择性降解蛋白质的形式。CMA 是干细胞蛋白质质量控制和 HSC 活化后脂肪酸代谢上调所必需的。研究发现，随着年龄的增长，HSC 中 CMA 的活性降低，并证明利用遗传手段或药物激活 CMA 可以恢复小鼠和人类 HSC 的功能。

总之，该发现为理解 CMA 在维持 HSC 质量控制、适当能量学和 HSC 长时功能中的作用提供了见解。该工作表明，CMA 可能是衰老或干细胞移植过程中增强 HSC 功能的潜在靶点。

据了解，大部分静息 HSC 的激活是血细胞终身产生的先决条件。这个过程需要对主要生物分子调控以满足 HSC 分裂和代谢需求。但是干细胞活化后调控细胞重编程以及其随后恢复静息的机制尚未得到充分表征。

相关论文信息：

<https://doi.org/10.1038/s41586-020-03129-z>

更多内容详见科学网小柯机器人频道：

<http://paper.sciencenet.cn/Alnews/>

## 科学线人

全球科技政策新闻与解析

### 美国任命首位部长级科学顾问



埃里克·兰德

图片来源: AP PHOTO

日前，美国当选总统拜登宣布选择一位研究政策专家担任总统科学顾问和白宫科学技术政策办公室（OSTP）主任。

这个熟悉的面孔是现年 63 岁的埃里克·兰德，他是由哈佛大学和麻省理工学院联合管理的博德研究所的主席和创始董事。作为一名数学家出身的分子生物学家，兰德在前总统奥巴马的领导下担任了 8 年的总统科技顾问委员会（PCAST）联席主席，并与后者的科学顾问 John Holdren 密切合作，同时与拜登进行了互动。

“兰德是一个极好的选择，他将成为了一位了不起的科学顾问。”Holdren 说，兰德是科学上的博学者，在政策上也是如此。在奥巴马当政期间，PCAST 发布了 39 份报告，每一份都有兰德的贡献，其中 6 份报告涵盖了以前的流行病和公共卫生危机。

他将是第一个兼任这两项职务的生物学家，且科学顾问这一职务首次提升为内阁级别，意味着兰德在内阁中占有了一席之地。

兰德具有很高的声望，曾领导人类基因组图谱绘制计划。

拜登要求兰德和他的团队回答有关科技未来的 5 个问题，并就如何应对公共卫生威胁、减轻气候变化影响、保持国家创新的世界领先地位、利用科学改善社会公平以及加强美国的研究事业，向政府提出建议。拜登还成立了一个团队应对新冠肺炎带来的公共卫生危机。

同样重要的是，拜登指出，为完成这些任务，兰德应该“广泛和透明地与美国不同科学领导层合作，让更广泛的美国公众参与其中”。这被认为是向拜登前任经常忽视或拒绝的科学家释放的“欢迎回家”的信息。

目前尚不清楚 OSTP 在兰德领导下的角色将如何演变。Holdren 认为，“兰德的工作是确保总统在任何政策问题背后的科学问题上获得最好的建议”。（王方）

### 《科学》系列期刊改变开放获取政策



《科学》及旗下期刊将有针对性地试用一项新的绿色开放获取政策。

图片来源: Loic Venance/AFP/Getty

为了向开放获取迈进，《科学》出版商将允许一些在其订阅期刊上发表文章的作者在相关条款下，公开分享他们已被接收的草稿，这意味着任何人都可以复制或转发这些作品。

这一变化确保了那些在 S 计划倡议下坚持开放获取政策的资助机构所资助的科学家，仍然可以在《科学》系列期刊上发表文章。大约有 24 家资助机构签署了 S 计划，该计划于 2021 年 1 月 1 日正式启动，当然各个机构的具体启动日期不同。

在过去的两个月里，为了响应 S 计划，许多订阅期刊已经引入了作者支付费用以使他们的论文开放获取的政策。但《科学》出版方、美国科学促进会（AAAS）却要避免这种情况，因为担心引入开放获取出版费用，可能会超出作者的经济能力。

而 AAAS 的新政策则允许由一些 S 计划机构资助的研究人员在论文发表后可以免费在网上发布论文的草稿。一些参与 S 计划的机构，如英国国家基金、英国研究与创新委员会等，还没有最终确定关于手稿共享的政策，所以该政策还不适用于它们。

之前，AAAS 已经允许这种由作者发起的即时共享（有时被称为绿色开放获取），但条款规定，手稿只能在个人或机构的网页上共享，不能转发。研究人员必须等上 6 个月，才能在文献服务检索系统等资料库中张贴手稿。这并没有使参与 S 计划的资方满意，他们认为作者应在论文发表时、在完全开放的许可下分享被接收的手稿。

2020 年 7 月，一些参与 S 计划的资助者甚至表示，将把作者保留公开分享已被接收手稿的权利作为资助的法律条件——不管期刊的出版协议怎么说。现在，AAAS 表示，接受这个“权利保留计划”机构资助的科学家将能够申请共享手稿开放许可。其他科学家则无法做到这一点。（鲁亦）