
病毒会“逃跑”“语言”露马脚

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/12425.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

病毒会“逃跑”“语言”露马脚。



快速的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

病毒可能比人们想象得更狡猾，它能不断伪装自己，逃避疫苗猎人的追捕。

一直以来，研制对抗流感和艾滋病病毒（HIV）等病毒的有效疫苗之所以如此困难，原因之一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良好靶标。

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病毒包膜表面蛋白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日发表在《科学》上的一项研究中，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冠病毒新变种。研究人员说，尚未经过同行评审的相关分析发现，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的速度发生基因突变，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，这是它们复制生物学的结果。例如，HIV和流感遗传物质复制的机制容易出错，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在接受《中国科学报》记者采访时表示。

为了让这些突变促进病毒逃逸，它们必须帮助病毒改变其表面蛋白质的形状，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对这些标准进行建模。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。然后，这些模型就可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，这些模型可以通过训练原始文本来理解人类语言。于是，我们认为，由于病毒最丰富的数据只是原始的病毒序列，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，语法类似于确定特定序列编码的蛋白质是否具有功能的规则，语义意义类似于蛋白质是否能够呈现新的形状，帮助它逃避抗体。因此，使病毒能够逃脱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

如果病毒想要逃离人类的免疫系统，又不想让自己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，以便不会被人体免疫系统检测到。Hie说。

为了模拟这一过程，研究人员训练了一个NLP模型来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。这种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6万条HIV序列、4.5万条流感序列和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共现和共变的这些特性。

一旦该模型被训练，研究人员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，可能会发展出来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。作者们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而逃脱识别。未参与该研究的美国国家医学图书馆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。这与最近的研究一致，研究表明，针对HA茎部的抗体可以提供几乎全面的保护，以对抗任何流感毒株。

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将在多长时间内保持有效尚不清楚。

目前，对于新冠病毒，我们认为我们的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，以便在实验室进行进一步测试。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进行进一步研究。

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，我们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，人们仍需要对每种病毒进行专门研究，以了解它们的突变率。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，因此我们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，以识别感染这些人的病毒中存在的突变。

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，它还可以用于设计小分子药物，这种药物可能不太可能引发结核病等疾病的耐药性。（来源：中国科学报唐凤）

相关论文信息：<https://doi.org/10.1126/science.abd7331>

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

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：Bryan Bryson 来源：《科学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发