

Immunopeptidomics for next-generation
bacterial vaccine development
免疫肽组学在下一代细菌疫苗开发中的应用

jhChen-陈家辉

2025 年 11 月 16 日

本文档主要介绍Immunopeptidomics for next-generation bacterial vaccine development（doi:10.1016/j.tim.2021.04.010）的相关内容。分享文献阅读中的心得体会。

抗菌素耐药性已成为日益严峻的全球性威胁，亟需可替代失效抗生素的疗法。本综述概述了针对胞内细菌病原体的免疫肽组学研究，并探讨了推进下一代疫苗研发面临的未来方向与挑战

1 术语解释

1.1 免疫肽组学 (Immunopeptidomics)

免疫肽组学 (Immunopeptidomics) 是一种专门研究和鉴定生物体中主要组织相容性复合体 (MHC) 所呈现的多肽 (肽段) 的组学技术。

免疫肽组学主要关注MHC分子 (包括人类的HLA分子) 在细胞表面呈现的所有肽段, 这些肽段的来源、序列、丰度、功能特性以及肽段-MHC复合物与T细胞免疫应答的关系都属于免疫肽组学的研究范畴。

1.1.1 MHC复合体

MHC (Major Histocompatibility Complex, 主要组织相容性复合体) 是一组编码细胞表面蛋白质的基因家族, 这些蛋白质在免疫系统中起着关键作用, 负责向T细胞呈递抗原肽段。MHC主要分为MHC I类分子 (MHC Class I) 和MHC II类分子 (MHC Class II)。

MHC I类分子 (MHC Class I) 由 α 链和 β 2-微球蛋白组成, 为跨膜蛋白, 表达在几乎所有有核细胞表面。主要负责呈递内源性抗原 (胞内合成的蛋白质), 包括细胞内自身蛋白、病毒蛋白、肿瘤抗原。CD8 T细胞识别MHC I类分子-肽段复合物。呈递途径为: 胞质内的细胞内抗原经蛋白酶体消化后, 可选择性通过内质网氨肽酶1/2 (ERAP1/2) 等蛋白酶进行进一步修剪, 形成约8-12个氨基酸长度的肽段。这些肽段通过抗原加工相关转运蛋白 (TAP) 进入内质网, 与MHC I类分子结合后, 整个复合物经高尔基体转运至细胞表面, 供CD8 T细胞监视并清除病原体感染细胞或恶性细胞。

MHC II类分子 (MHC Class II) 由 α 链和 β 链组成, 主要表达在抗原呈递细胞 (APC) 上如树突状细胞, 巨噬细胞, B细胞, 活化的T细胞。主要负责呈递外源性抗原如细菌蛋白、病毒颗粒、过敏原。CD4 T细胞识别MHC II类分子-肽段复合物。呈递途径为: 外源性抗原通过内吞作用进入细胞, 在酸性环境中被蛋白酶降解为肽段, 肽段与MHC II类分子结合, 随后转运至细胞表面。

在人类中, MHC被称为HLA (Human Leukocyte Antigen, 人类白细胞抗原)。

1.1.2 核细胞

核细胞是指含有细胞核的真核细胞。

1.2 细菌病原体 (Bacterial pathogen)

细菌病原体是指能够引起人类、动物或植物疾病的细菌。它们通过各种机制侵入宿主、破坏组织或干扰正常生理功能。

1.2.1 胞外细菌病原体 (Extracellular Bacterial Pathogens)

在宿主细胞外生存和繁殖，主要存于血液、组织间隙或体液中。通过分泌毒素、酶或其他因子来破坏宿主组织并引发免疫反应。

致病机制如通过分泌外毒素直接损伤宿主细胞；产生内毒素（如脂多糖）引起炎症反应；形成生物膜抵抗宿主防御

免疫应答包括产生抗体由体液免疫主导，中和毒素、调理吞噬作用，补体系统激活。胞外细菌疫苗主要诱导抗体产生。

经典例子包括金黄色葡萄球菌 (*Staphylococcus aureus*)、链球菌 (*Streptococcus* spp.) 和大肠杆菌 (*Escherichia coli*)。

1.2.2 胞内细菌病原体 (Intracellular Bacterial Pathogens)

侵入宿主细胞内并在细胞内生存和繁殖。它们能够避开宿主的体液免疫防御机制，利用宿主细胞资源进行增殖。主要针对巨噬细胞等吞噬细胞，通过抑制溶酶体与吞噬体的融合来避免被杀死。

致病机制如逃避抗体和补体系统的攻击，长期潜伏在宿主细胞内；操纵宿主细胞信号通路以促进自身存活和复制，干扰宿主细胞的正常功能。

免疫应答由细胞免疫主导 (T细胞应答)，*CD8* T细胞识别感染细胞，*Th1*型*CD4* T细胞激活巨噬细胞。胞内细菌疫苗需要激活细胞免疫，特别是T细胞应答。

典型例子包括结核分枝杆菌 (*Mycobacterium tuberculosis*)、沙门氏菌 (*Salmonella* spp.) 和李斯特菌 (*Listeria monocytogenes*)。

1.3 表位

表位（又称抗原决定簇，Antigen Determinant）是抗原分子上能够被免疫系统（抗体或T细胞受体）特异性识别和结合的最小结构单位。

1.3.1 B细胞表位（B-cell Epitope）

可被抗体识别和结合的位点，通常是抗原表面的三维结构，可以是蛋白质、多糖、脂质等。可分为线性表位（Linear Epitope）和构象表位（Conformational Epitope）。线性表位由连续的氨基酸序列组成，通常为5-20个氨基酸长度，在蛋白质变性后仍能被识别。构象表位由不连续的氨基酸在空间上折叠形成，依赖于蛋白质的三维结构，蛋白质变性后失去免疫原性。

1.3.2 T细胞表位（T-cell Epitope）

被T细胞受体识别的肽段，必须与MHC分子结合后呈递，通常是线性的短肽序列。分为MHC I类表位和MHC II类表位。MHC I类表位通常为8-11个氨基酸，主要被CD8 T细胞识别，来源于内源性蛋白（胞内合成的蛋白质）。MHC II类表位通常为13-25个氨基酸，主要被CD4 T细胞识别，来源于外源性蛋白（胞外摄入的蛋白质）。

2 背景

随着抗生素在20世纪上半叶的问世，许多细菌性疾病对人类造成的毁灭性主导地位得以终结。然而，抗生素的（过度）使用加速了抗生素耐药性 antimicrobial resistance（AMR）的出现。当前AMR形式严峻的致病菌中包含多种胞内细菌病原体。疫苗被认为是解决AMR威胁的最有效手段之一，因为它们可以预防性的阻断感染，从而减少对抗生素的需求。并且抗菌疫苗能持续对目标病原体保持有效性。

目前已经有多种疫苗成功应用如针对破伤风、白喉和百日咳的联合疫苗，以及b型流感嗜血杆菌（Hib）疫苗、肺炎球菌疫苗和脑膜炎球菌疫苗等等。上述疫苗都是针对于胞外细菌病原体设计的。仅有结核病和伤寒疫苗针对胞内病原体，由于胞内病原体的复杂免疫逃逸机制，现有疫苗的保护效果有限。因此，迫切需要开发针对胞内细菌病原体的有效疫苗。

对于胞内病原体，有效疫苗的研发长期受限于抗原认知不足以及多数疫苗平台无法激发强烈细胞毒性免疫应答的缺陷。新一代疫苗技术——包括病毒载体疫苗、DNA疫苗和mRNA疫苗——现已能克服后一障碍，这些疫苗既可诱导体液免疫又能激发细胞毒性免疫反应，且具备较短的研发周期，这三种技术均通过向宿主细胞递送编码实际抗原的遗传信息，实现胞内抗原合成，从而增强细胞毒性免疫应答。

于是开发新型疫苗的问题就转向了应选择何种细菌抗原编码至这些疫苗平台。相较于病毒，致病菌的基因组更大更复杂，通常每个细菌拥有数千个基因，这使得从病原体中筛选合适的抗原变得极具挑战性。图 1显示了免疫表位数据库（IEDB, www.iedb.org）中胞内细菌病原体覆盖范围的局限性。图中显示了胞内细菌病原体的蛋白质抗原数量占其各自病原体总蛋白质组的比例，条形图的黑色部分表示MHC I类分子呈递的抗原，灰色部分则描绘所有非MHC I类分子呈递的抗原。柱状图上方的数字代表2021年2月22日从免疫表位数据库（IEDB）获取的各病原体抗原总数。可以发现占比最高的结核分枝杆菌（Mtb）超过20%的蛋白质编码基因已被鉴定为抗原，而大多数的其他胞内细菌病原体抗原覆盖率均低于5%。并且只有其中的MHC I类可能引起细胞毒性免疫应答。即便免疫优势抗原的数量远远小于整个蛋白质组，但很容易推测出仍然存在大量未知的潜在抗原未被鉴定。

由里诺·拉普利在2000年开创了反向疫苗学技术，即利用完整病原体基因

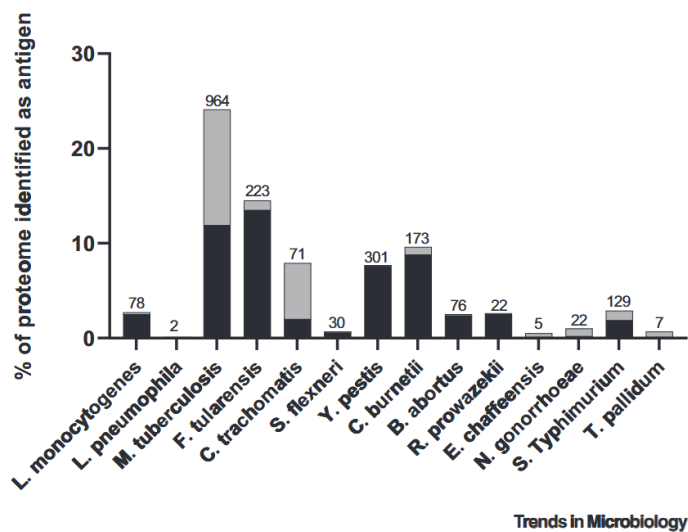


图 1: 胞内细菌病原体的抗原覆盖范围

组推断蛋白质序列，以鉴定分泌出的表面暴露蛋白作为合适的疫苗候选物。候选抗原在大肠杆菌或其他体系中表达后，再通过动物模型评估其免疫原性和保护效果。这种方法已成功应用于多种胞外细菌病原体疫苗的开发。基于这样的初始反向疫苗平台，结合近年来发展的包括B细胞受体库深度测序、基于结构的抗原设计及基于质谱的蛋白质组学等新兴技术，已经开发并推行了反向疫苗2.0平台。

3 免疫肽组学

3.1 免疫肽组学概述

目前能够直接检测抗原肽或蛋白质的蛋白质组学方法很丰富，比如免疫捕获质谱变体技术，电免疫沉淀法，多重亲和蛋白谱分析技术等等。这些方法均依赖于抗体分泌，因此仅能检测体液免疫应答的靶标。然而免疫肽组学可实现不依赖于抗体的抗原检测。

免疫组学旨在检测通过主要组织相容性复合体（MHC）呈递在细胞表面的抗原肽段，这些肽段通常被称为免疫肽、MHC相关肽或MHC配体。这些免疫肽是长度为8-25个氨基酸的抗原片段，在细胞内被装载到MHC复合体上，随后转运至细胞表面供T细胞识别。MHC分子分为I类和II类，分别介导胞质抗原与液泡/胞外抗原的呈递，并与功能特性各异的T细胞群体结合。虽然MHC II类分子主要局限于专业抗原呈递细胞（APC）表达，但MHC I类分子存在于所有有核细胞表面，通过与CD8细胞毒性T细胞结合启动受损细胞的死亡程序。本文将主要聚焦于讨论该机制。

3.2 免疫肽组学检测方法与技术

目前的MHC呈递的肽段的肽段通常通过质谱检测来实现，基于LC-MS/MS的HLA结合肽洗脱检测技术为测量抗原肽的首选方法。样品检测前首先需要经过分离和纯化的过程，目前的分离方法主要有两类：

温和酸洗脱法(MAE)，直接从完整细胞表面快速简易地获取免疫肽，但存在污染物肽共洗脱及MHC I/II类肽不可分离的缺陷；

免疫亲和纯化法(IP)，通过特异性抗体从细胞裂解液中纯化MHC分子-肽段复合物，随后洗脱肽段。需大量特异性抗体且样品制备时间较长，但能实现更优的免疫肽富集效果和更高的鉴定数量，且能够分离MHC I类和II类肽段。因此因此绝大多数免疫肽组学研究均采用IP方法（图2）。

细胞/组织通过温和去垢剂结合剪切应力裂解后，通过澄清液以及无菌过滤除去残留碎片/病原体，含有完整MHC-肽复合物的澄清裂解液与免疫亲和柱共同孵育。MHC特异性抗体通常通过共价交联固定于蛋白A/G包被磁珠，以避免抗体共洗脱造成的样本污染。经多次清洗后，采用酸性洗脱液

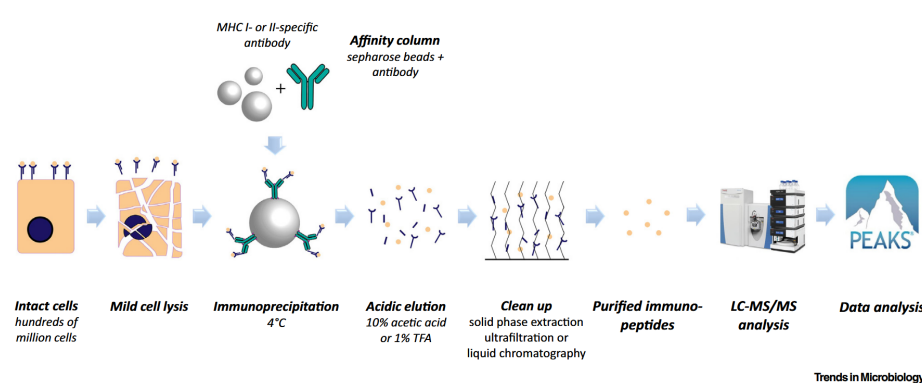


图 2: 基于免疫肽组学的免疫肽组分析流程

将MHC-肽复合物从免疫亲和柱上解离。洗脱液需通过液相色谱、超滤或C18固相萃取进一步纯化以去除MHC分子，从而实现短链亲水性免疫肽与长链疏水性MHC分子的分离。针对纯化后免疫肽的液相色谱-串联质谱分析，必须采用高灵敏度、高分辨率仪器及定制化参数以确保分析成功。所得原始数据通过如MaxQuant、PEAKS等软件进行数据库搜索和肽段鉴定。

由于免疫肽在细胞内的丰度极低，故进行检测时需要极大的初始样本量——通常需要数亿细胞或1克组织。因此必须最大限度减少样本处理步骤和表面暴露时间。

3.3 细菌免疫肽组学

免疫肽组学已广泛应用于肿瘤相关抗原的发现，细菌感染模型在免疫肽组学领域研究不足，但随着抗菌素耐药性的出现，对于细菌感染模型的研究正在逐渐增多。

首例细菌感染细胞的免疫肽组学研究由Hunt实验室于2002年报道，该研究从结核分枝杆菌H37Ra株系中鉴定出抗原肽段。 2017年，麦克默特里及其团队报道了H37Rv结核分枝杆菌株感染后呈现在非经典低多态性MHC-I类HLA-E分子上的免疫肽组。近期在Bettencourt等人针对BCG感染的人THP1巨噬细胞开展的研究中发现，多抗原的联合递送可使肺部和脾脏的细菌载量下降近十倍，这表明免疫肽组学鉴定的抗原组合尤其具有开发新型疫苗配方的重要潜力。

2008年，Robert Brunham实验室报道的研究表明感染鼠衣原体的小鼠骨髓来源树突状细胞（BMDCs）。在后续研究中，作者揭示了共有4种对人间皮细胞粘附至关重要的Pmps及MOMP（均为高丰度外膜蛋白）被识别，表明蛋白丰度、细胞定位及表达动力学对病原体蛋白递呈与抗原性起主导作用。

除衣原体外，Brunham实验室还将其免疫肽组学技术流程应用于鼠伤寒沙门氏菌感染的小鼠骨髓源性树突状细胞，共鉴定出1891个MHC I类肽和617个MHC II类肽。

2018年，Graham等人报道了一种名为BOTA（细菌源性T细胞抗原）的全基因组MHC II类限制性免疫优势表位预测算法。研究者采用同标标记LC-MS/MS技术，从表达H2-IA的单核细胞增生李斯特菌感染小鼠骨髓源性树突状细胞中鉴定了MHC II类结合肽段。BOTA整合了蛋白质定位、跨膜结构和结构域分布信息，在预测已鉴定的MHC II类李斯特菌免疫肽方面优于NetMHCIIpan。将BOTA与NetMHCIIpan联用分析炎症性肠病公开数据时，其预测效能也显著优于单独使用NetMHCIIpan。

最近，Gaur等人通过感染的人血源未成熟树突状细胞（iDCs）绘制了土拉弗朗西斯菌呈递的MHC I类肽段图谱。研究者报告了来自十个不同抗原的弗朗西斯菌源肽序列，并通过计算机模拟验证了其结果——针对三个最佳评分表位计算出高HLA结合亲和力与抗原性评分。

3.4 抗原鉴定到疫苗开发

免疫肽组学技术的持续改进将有助于未来更深入地挖掘细菌抗原库。肽段剪接在之前提到的所有研究中均被忽略。由于T细胞能够识别细菌剪接肽，对此类肽段的考量将产生更完整的抗原谱。类似地，其他翻译后修饰（如甲酰化或棕榈酰化）也可能存在于细菌免疫肽上，甚至宿主细胞来源的免疫肽在感染过程中可能呈现修饰变化。细菌感染还会干扰宿主细胞的抗原呈递过程，未来免疫肽组学研究应综合考虑此类免疫调节过程及翻译后修饰。尽管免疫肽组学通常需要大量起始材料，但最新技术突破已实现将样本量缩减至约 110^7 个细胞，并正在持续推进更小样本量的研究。

基于IP的方法，减少材料需求的方法包括，选择正确的细胞类型，找到正确的感染比例，以及通过免疫刺激如使用IFN γ 增加MHC表达。同样可以

联合多组学分析，比如转录组学来预测高表达抗原，可以通过AI技术来预测哪些蛋白序列更容易被蛋白酶体降解成适合MHC呈递的肽段以及对应的切割区域，以及分析预测对应细菌基因的表达模式。利用高灵敏度的检测技术也是较好的方法，结合AI可以从从复杂背景中提取低丰度免疫肽信号，识别免疫肽段的特征碎片模式从而实现高通量的数据筛选处理。

随着免疫肽组学应用于越来越多的细菌感染模型，如何从这些研究中筛选抗原以编码至有效的新一代疫苗成为关键问题。在开展体内试验之前，对已发现表位进行鉴定验证与抗原性确认是至关重要的首要步骤。抗癌疫苗通常编码具有肿瘤特异性和患者（HLA）特异性的突变表位，而抗菌疫苗则普遍采用完整蛋白抗原而非表位，以确保实现不依赖HLA的群体广谱保护。目前并未建立免疫肽组学来源抗原作为细菌疫苗候选物的优先选择指南，但通过整合前文研究，本文总结出了若干趋势和原则（如图3）。

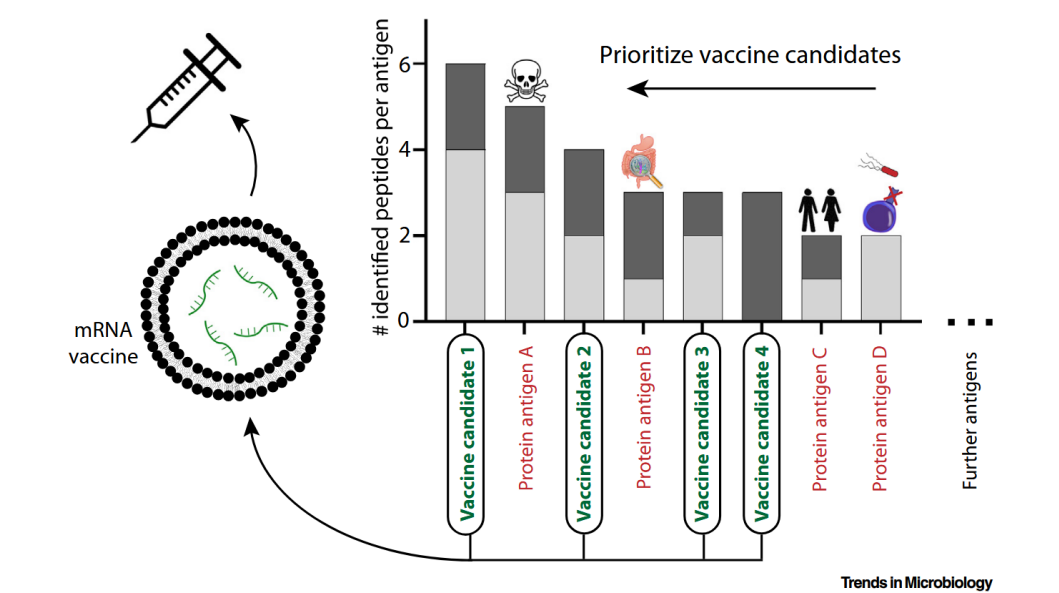


图 3: 基于免疫肽组学研究的疫苗候选优先化策略示意图

条形图呈现细菌感染细胞模型的假设性免疫肽组学实验结果，该筛选在与感染相关的两种宿主细胞类型浅灰色：上皮细胞；深灰色：抗原呈递细胞）中完成。纵轴显示两种筛选中各细菌蛋白抗原所对应的已鉴定肽表位数量。绿色标记抗原因在双重筛选中释放最多表位而被视为优质疫苗候选，这类抗原通常来源于细菌外周结构并与宿主细胞内环境密切接触。红色标记抗

原因存在显著毒性（抗原A）、与微生物组（抗原B）或宿主蛋白（抗原C）存在序列同源性、或T细胞反应性较低（抗原D）等问题，不被视为合适候选。将四种最优候选抗原编码于单一mRNA疫苗制剂中，既可引发广泛免疫应答，又能避免病原体因单一抗原变异引发的免疫逃逸。

释放最多表位的蛋白质可能代表免疫优势抗原。未来需要更灵敏的免疫肽组学筛查（每次识别数百而非数十个细菌表位）来进一步证实表位丰度与抗原免疫优势之间的潜在关联。其次，能引发强免疫应答的抗原通常源自细菌膜或周质区域，这与这些表面蛋白与宿主细胞内环境的密切接触相一致。第三，联合递送多个免疫肽组学发现的抗原似乎比单一抗原能提供更好保护，既可为不同人群提供稳定保护，又可避免单一抗原偏移。还应综合考虑疫苗抗原筛选的通用原则与工具，包括优先选择临床菌株中保守的抗原、排除有害抗原（如细菌毒素）以及避免与宿主和微生物组相似的抗原以防止自身免疫反应。

通过MHC-I类免疫肽组学鉴定的抗原可被轻松编码至针对胞质病原体的疫苗中——这类病原体（如单核细胞增生李斯特菌）通常需要强烈的细胞毒性CD8 T细胞反应以清除感染。事实上，被编码的抗原信息将在宿主细胞胞质内翻译成细菌蛋白，直接为MHC-I类分子呈递做好准备。对于非胞质型空泡病原体（如衣原体或沙门氏菌），体液免疫和CD4 T细胞反应通常更为关键，此时可采用分泌信号标签对MHC-II类鉴定的抗原进行编码，使分泌的细菌抗原能被专职抗原呈递细胞摄取并进行MHC-II类呈递。但需注意避免分泌相关糖基化可能引发的免疫耐受。值得注意的是，多数病原体的保护性免疫需要细胞免疫与体液免疫共同作用。传统全病原体疫苗或亚单位疫苗通常引发强抗体反应，而核酸疫苗现可诱导更强的细胞毒性反应。对于像结核分枝杆菌（Mtb）这类病原体，近年研究已明确细胞毒性反应的重要作用，MHC-I类免疫肽组学或有助于设计更有效的核酸疫苗。

4 总结和展望

免疫肽组学已成熟为一种强大的抗原发现技术。目前该技术仅应用于少数胞内细菌病原体，特别是分枝杆菌和衣原体，但随着抗菌素耐药性的加剧，预计未来数年将有更多研究揭示该领域的关键问题，如“基于细菌感

染模型的免疫肽组学筛选是否包含如何选择细菌抗原作为疫苗候选物的信息?”，“我们能否利用已识别表位的丰度及其在不同筛选中的重复出现频率来优先选择其母体抗原?”，“哪种宿主细胞类型、组织或器官最适合用于抗原筛选”等。已发现的抗原可直接编码于核酸疫苗中，构成一个多功能且强大的疫苗开发平台。通过进一步提高免疫肽组学研究的灵敏度，预计将增加抗原鉴定数量，并强化与病毒载体和mRNA疫苗平台的黄金联盟，从而推动开发有效的新型抗菌疫苗。