·综 述·

深度突变扫描技术在蛋白研究中的应用

李怡凡，王怡，张凯丽，李帅*

天津医科大学肿瘤医院乳腺病理研究室 国家恶性肿瘤临床医学研究中心 天津市恶性肿瘤临床医学研究中心 天津市“肿瘤防治”重点实验室 乳腺癌防治教育部重点实验室，天津 300060

李怡凡, 王怡, 张凯丽, 李帅. 深度突变扫描技术在蛋白研究中的应用[J]. 生物工程学报, 2023, 39(9): 3710-3723.
LI Yifan, WANG Yi, ZHANG Kaili, LI Shuai. Application of deep mutational scanning technology in protein research[J]. Chinese Journal of Biotechnology, 2023, 39(9): 3710-3723.

摘 要：作为细胞结构与功能的中心参与者，蛋白质一直是生命科学研究的中心主题。分析蛋白质序列变异对其结构、功能的影响，是研究蛋白的重要手段之一。近年一种称为深度突变扫描(deep mutational scanning, DMS)的技术被广泛应用于蛋白研究领域，其通过高丰度 DNA 文库在蛋白特定区域平行引入成千上万种突变，经筛选后，利用高通量测序为每一种突变打分，从而揭示序列与功能之间的相关性。深度突变扫描以其高通量、快速简易、节省人工等特点，已经成为蛋白质功能研究以及蛋白工程改造的一种重要方法，目前已在蛋白进化、抗体改造、致病突变鉴定等蛋白研究的多个领域广泛应用。本综述简要概括了深度突变扫描技术的原理，重点介绍了其在哺乳动物细胞中的应用，同时分析了目前的技术瓶颈，旨在为相关研究提供参考。

关键词：深度突变扫描；文库构建；高通量测序；哺乳动物细胞

Application of deep mutational scanning technology in protein research

LI Yifan, WANG Yi, ZHANG Kaili, LI Shuai*

Key Laboratory of Breast Cancer Prevention and Therapy, Ministry of Education, Tianjin Key Laboratory of Cancer Prevention and Therapy, Tianjin’s Clinical Research Center for Cancer, Department of Breast Cancer Pathology and Research Laboratory, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin 300060, China

Abstract: As central players in cellular structure and function, proteins have long been central themes in life science research. Analyzing the impact of protein sequence variation on its

资助项目：国家自然科学基金(31870860)；天津市医学重点学科(专科)建设项目(TJYXZDXK-012A)
This work was supported by the National Natural Science Foundation of China (31870860) and the Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-012A).
*Corresponding author. E-mail: shuaili@tmu.edu.cn
Received: 2022-12-30; Accepted: 2023-05-22; Published online: 2023-05-31

structure and function is one of the important means to study proteins. In recent years, a technology called deep mutational scanning (DMS) has been widely used in the field of protein research. It introduces thousands of mutations in parallel in specific regions of proteins through high-abundance DNA libraries. After screening, high-throughput sequencing is employed to score each mutation, revealing sequence-function correlations. Due to its high-throughput, fast and easy, and labor-saving features, DMS has become an important method for protein function research and protein engineering. This review briefly summarizes the principle of DMS technology, highlighting its applications in mammalian cells. Moreover, this review analyzes the current technical bottlenecks, aiming to facilitate relevant research.

Keywords: deep mutational scanning; library construction; high-throughput sequencing; mammalian cells

蛋白氨基酸序列变异可影响蛋白的结构与功能。对蛋白氨基酸序列进行突变扫描可帮助了解蛋白序列(sequence)与功能(function)的相关性，但如何进一步深度理解蛋白质序列中每个氨基酸和功能之间的关系仍是一个棘手的问题。近年随着基因合成、基因编辑、高通量测序(high-throughput sequencing, HTS)等技术的出现，解析蛋白序列——功能相关性的新研究方法不断出现，其中以深度突变扫描最具代表性。深度突变扫描(deep mutational scanning, DMS)又称饱和突变筛选(saturated mutagenesis screen)，是一种高通量地研究蛋白序列——功能相关性的实验方法，以高效和相对低的成本大规模地量化遗传变异的影响[1-2]。自 2010 年后，Fowler 等[3]、Ernst 等[4]和 Hietpas 等[5]发表的几篇代表性论文开启了深度突变扫描技术，自此此项技术得到快速发展。本文将深入介绍深度突变扫描技术，并着重探讨以哺乳动物细胞为平台的深度突变扫描应用。

1 DMS 实验流程

开展 DMS 实验首先要创建一个靶蛋白的突变文库，以展示其结构及功能多样性；接着建立合适的选择系统对蛋白文库进行筛选；最后通过高通量测序分析不同序列变异在功能筛选前后出现频率的变化，从而将蛋白序列与功能相关联。下面将简要介绍 DMS 的实验流程，包括突变文库的构建、选择系统的建立、高通量测序及数据分析[6] (图 1)。

1.1 突变文库的构建

构建饱和突变文库是 DMS 实验的起点。根据研究目的不同，可构建以下 2 种突变文库：(1) 单位点文库(single-site library)，对蛋白全长或某个区域的所有氨基酸进行逐一突变，由野生型突变为另外 19 种氨基酸，文库复杂度为 N*19 (N 代表待突变氨基酸位点数量)；(2) 多位点文库(multi-site library)，同时对多个氨基酸位点进行饱和突变，此类文库复杂度为 20^{N}[7-8] (图 2)。在突变文库构建、引入、表达过程中，研究者力求覆盖所有的序列可能性，使之得以呈现并被筛选，这是 DMS 实验能否成功的关键。突变文库变异序列的引入通常先从引物(寡核苷酸)合成开始，后续通过 PCR 过程将变异序列引入质粒或 PCR 产物中用于后续文库表达。突变引物也可以通过 PFunkel 等[9]方法直接引入到质粒载体中，形成质粒突变文库。此外，含变异序列的寡核苷酸还可以作为同源重组模板被引入到基因组(见 3.1)。

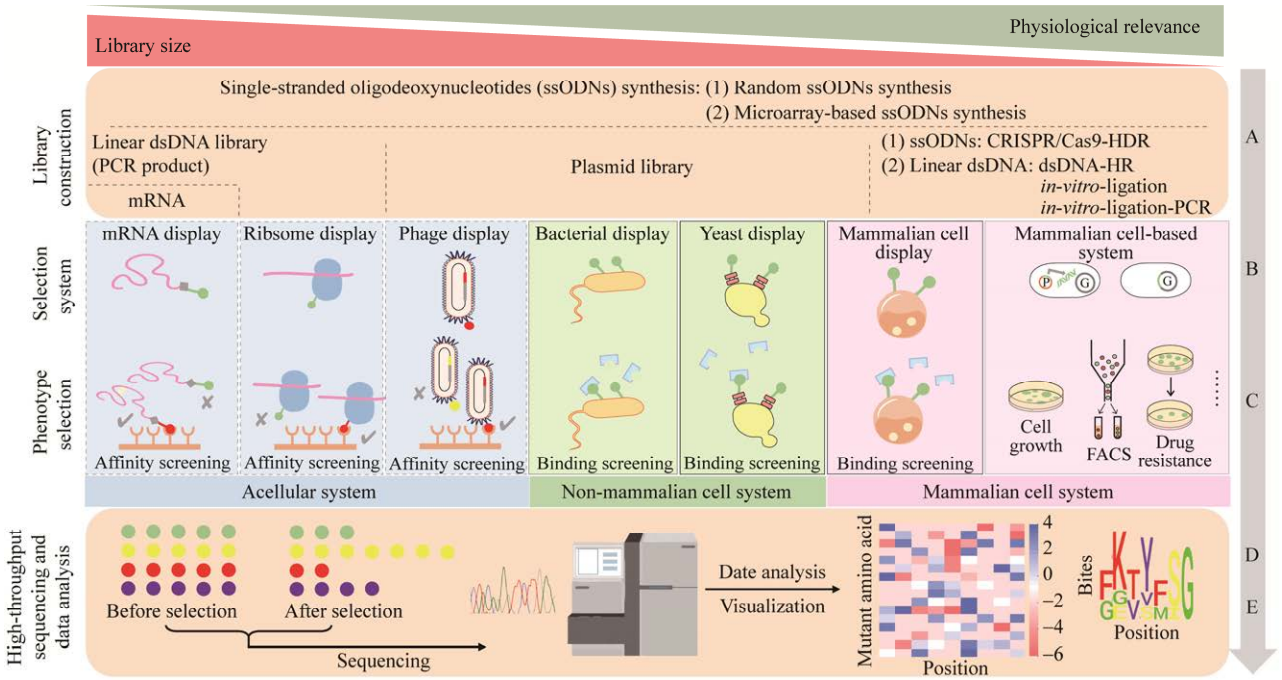

图 1　深度突变扫描实验流程　A：突变文库构建. B：突变文库引入不同的选择系统. C：表型筛选. D：高通量测序. E：数据分析及可视化

Figure 1　DMS experiment workflow. A: Mutant library construction. B: Introduction of mutational libraries into different selection systems. C: Phenotype selection. D: High-throughput sequencing. E: Data analysis and visualization.

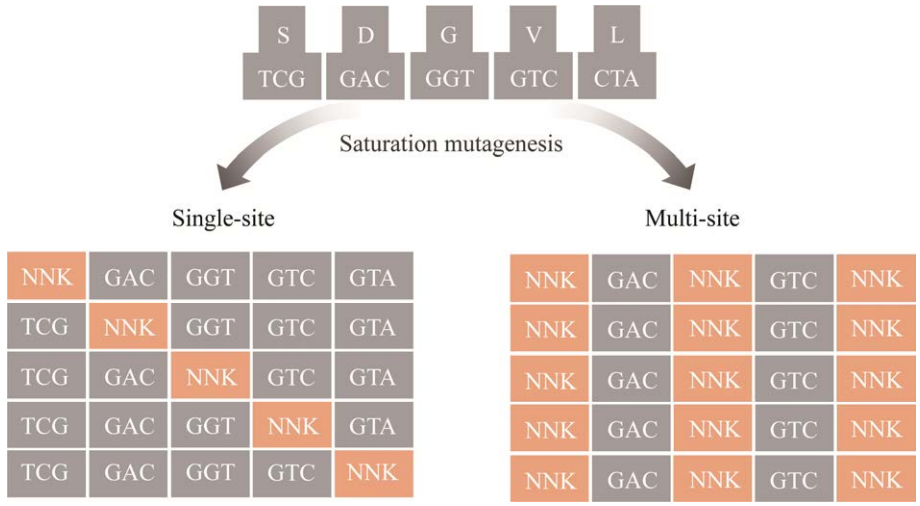

图 2　深度突变扫描文库设计　单位点文库对蛋白全长或某个区域的所有氨基酸进行 NNK 简并碱基逐一替换. 多位点文库同时对多个氨基酸位点进行饱和 NNK 简并碱基替换

Figure 2　DMS library design. Single-site library is generated by the one-by-one substitution of all individual amino acids with NNK degenerate codons. Multi-site library is generated by simultaneous substitution of multiple amino acids with NNK degenerate codons.

在单位点突变文库构建过程中可通过在对应氨基酸位点引入 NNK (或 NNS、NNB)简并碱基实现所有 20 种氨基酸变异的表达，每一个氨基酸位点需要设计一条引物[10]。多位点文库则通过直接在突变位点连续引入多个 NNK 简并碱基，通常通过合成单一引物完成。利用 NNK 简并碱基引入变异序列，引物设计简单，合成成本低廉，可覆盖所有 20 种氨基酸。但存在以下缺点：(1) 不同氨基酸代表性存在偏差，例如在 NNK 简并性下亮氨酸(leucine, Leu)出现的概率为 9.375% (3/32)，而酪氨酸(tyrosine, Tyr)的概率为 3.125% (1/32)，两者相差 3 倍；(2) 会出现终止密码子，且出现频率会随着 NNK 数目的增加而累积，当文库含有 20 个 NNK 简并碱基时，有约一半的文库含有终止密码子。这导致有效文库的比例下降。近年出现的三联核苷(trimer phosphoramidites)引物合成工艺部分解决了以上问题，但合成成本较为昂贵[11]。

此外，高通量引物合成技术近年来被大量应用于 DMS 领域[12-14]。芯片合成技术即基于微阵列的寡核苷酸合成技术，采用了多通道合成方式。利用该技术可在微阵列芯片上一次性合成成千上万种不同的寡核苷酸序列[13]。基于芯片的引物合成可实现突变文库的高度定制化，生成更精准的文库，既可以引入所有 20 种氨基酸序列，也可以实现固定氨基酸组合的引入[15]。

1.2 选择系统的建立及突变文库的引入

目前 DMS 实验所涉及的选择系统大致可以分为三类：(1) 非细胞系统(acellular system)；(2) 非哺乳动物细胞系统(non-mammalian cell system)；(3) 哺乳动物细胞系统(mammalian cell system)[16]。3 种选择系统在库容、生理病理相关性方面有各自的优缺点。

1.2.1 非细胞系统

常用的非细胞系统包括噬菌体展示[4,17]、mRNA 展示[18-20]、核糖体展示[21]等系统。非细胞系统的主要优势在于其文库规模较大，但局限是多进行基于亲和力表型进行筛选，难以开展基于功能的筛选。噬菌体展示系统是最常用的非细胞系统，其文库丰度可达 10^9[22]。Fowler 等[3]通过噬菌体展示技术对人类 Yes 相关蛋白 65 (human Yes associated protein 65, hYAP65) WW 结构域突变文库进行了展示，利用同源肽配体进行结合筛选，对筛选前后的突变文库进行 Illumina 测序，绘制出 WW 结构域的序列——功能地形图，该结果与已知的 WW 结构域特征相一致。此外，mRNA 展示系统、核糖体展示系统可通过体外转录翻译展示突变文库，文库丰度可达到更高水平(>10^{13})。

1.2.2 非哺乳动物细胞系统

非哺乳动物的细胞系统包括细菌、酵母系统，突变文库可以质粒形式转化进入细胞，文库规模受限于转化效率(细菌约 10^9，酵母约 10^7)[23-24]。细菌、酵母系统常基于展示技术、细胞生长、报告基因表达等进行选择。Kim 等[25]通过酵母细胞融合蛋白报告系统对蛋白降解子(degron) Deg1 进行 DMS，将来自酵母细胞 Matα2 蛋白的 Deg1 变异序列与酵母 Leu2 蛋白融合生成 Deg1-Leu2 蛋白，将编码融合蛋白的质粒转入酵母中。Deg1 中增加蛋白稳定性的突变将导致融合蛋白的累积，从而增加亮氨酸的产生，促进营养缺陷型酵母(–Leu)的生长。经条件培养基选择后会富集含有可增强 Deg1 稳定性的突变。最终通过对选择前后的变异序列进行高通量测序，完成了对 Deg1 序列 3 万多种变异的稳定性评分。

1.2.3 哺乳动物细胞系统

基于哺乳动物细胞是开展疾病相关蛋白 DMS 研究的最佳系统。在下文做详细介绍(见 3)。

1.3 高通量测序及数据分析

高通量测序技术的快速发展及应用是 DMS 技术得以建立的基础[26]。常用的测序方法有 Illumina 测序、Nanopore 测序、PacBio 公司的单分子实时测序(single molecule real-time sequencing)等[27-28]。Illumina 测序由于其成本、技术成熟度、市场占有率等优势成为 DMS 研究中最常用到的测序平台。

DMS 数据分析是将氨基酸序列变异与功能相关联的重要步骤。最简单的分析是将筛选后群体的频率与筛选前群体的频率进行比较，得到每种氨基酸变异的富集比。但通过这样简单处理所得到的数据可能存在一定偏差，需要利用生物信息分析工具进行进一步的精确分析。

常用的 DMS 数据分析工具有：Enrich[29]、dms_tools[30]、DiMSum[31]等，它们可将原始测序数据转化为突变功能评分，并生成可视化数据。Enrich 是一个分析 DMS 数据的统计框架，适用于多数常见的实验设计，能读取来自任何测序平台的 FASTQ 格式数据，识别蛋白突变信息并使用含有重叠序列的双端(paired-end)测序读数进行纠错。在此基础上 Fowler 等[32]开发了 Enrich2，增加了更复杂的统计分析，适用于基于细胞生长和结合筛选的 DMS 数据分析。dms_tools 可以在给定单个选择压力下推断每个序列位点对氨基酸的偏好或者评估这些偏好在不同选择压力下的变化程度，推断突变的影响。DiMSum 以 R/Bioconda 软件包的形式免费提供给使用者，用于从原始测序文件中获得可靠的突变适应度评分和误差评估，并在分析中采取补救措施。DiMSum 创新点在于使用了可解释的误差模型(interpretable error model)，能提供更为准确的误差分析。目前，Cenik 等[33]新开发出一种基于饱和诱变实验的软件工具包——satmut_utils (saturation mutagenesis utilities)，可用于对 DMS 数据进行分析，相比于 Enrich2、dms_tools、DiMSum 等软件需要特定的输入要求和实验设计，satmut_utils 设计更为灵活，可扩展到多种实验设计。

2 基于非哺乳动物细胞平台的 DMS

基于非哺乳动物细胞平台的 DMS 利用非细胞系统或非哺乳动物细胞系统开展对蛋白结构、进化的相关研究，文库丰度较高。

2.1 蛋白质结构预测

蛋白结构测定通常依赖于 X 射线晶体衍射、核磁共振及冷冻电镜等技术平台。DMS 的出现为研究人员深入了解并预测蛋白结构提供了新的辅助工具[34-35]。Adkar 等[36]建立了细菌毒素蛋白 CcdB 的单位点突变(single-site mutant)文库，通过细菌存活实验及高通量测序分析了这些变异对蛋白毒性的影响。DMS 结果表明 CcdB 中每个突变的功能评分(rank score)与已知结构中该残基距蛋白表面的距离相关。仅仅依据 DMS 数据，在不需要表达、纯化蛋白的情况下实现了对 CcdB 结构模型的准确预测。此外，DMS 数据也可用来确定蛋白质的三维结构。这种蛋白质三维结构的确定依赖于突变之间遗传相互作用即上位性的量化以及突变位点间的非独立相互作用。已有相关研究通过 DMS 产生的突变遗传相互作用数据来预测蛋白质或其结构域中残基对之间的结构接触，进而模拟出蛋白质或其结构域的二级结构或三级结构[37-38]。例如，Olson 等[20]分析了含 56 个氨基酸的蛋白 G 结构域 B1 (protein G domain B1, GB1)结构域的所有单一和几乎所有成对突变，通过与免疫球蛋白 G 片段可结晶(immunoglobulin G-fragment crystallizable, IgG-FC)的结合对选择前后的突变体进行 DMS，成功预测了残基的空间距离，

绘制出该结构域的二级和三级结构。

DMS 还可以揭示一些与疾病相关的固有无序蛋白(intrinsically disordered protein, IDP)的结构特征，例如神经退行性疾病相关 TDP-43 蛋白[39]、阿尔兹海默病相关 β-淀粉样蛋白(amyloid β-protein, Aβ)[40]，为绘制蛋白质结构图谱提供了低成本实验策略[41]。对于具有多种构象的固有无序蛋白来说，可能只有少数构象与功能相关。Newberry 等[42]在酵母细胞中对固有无序蛋白 α 突触核蛋白(α-synuclein)的 2 600 个突变进行了生长表型筛选，DMS 结果分析表明 α 突触核蛋白的活性构象是一个长且不间断的两亲性螺旋。通过筛选在某种特定应激环境下的蛋白突变体，DMS 提供了一种简便的方法来分析蛋白构象如何响应细胞环境变化的。

2.2 蛋白进化研究

突变是进化的核心，产生有益遗传突变的能力是生物进化的关键[43]。适应度地形(fitness landscape)是适应度在所有可能的序列空间中的功能分布，是多种进化理论的基础。适应度地形可以描述蛋白基因型与适应度的相关性，阐明进化轨迹。了解突变适应度分布是理解进化动力学、遗传变异、有害突变累积的基础，但其绘制需要对大量蛋白突变体适应度进行量化，这是一项艰巨的任务[44]。

DMS 的出现从一定程度上解决了这一难题，利用该技术可通过人为模拟的小规模自然选择实现对数万个突变的适应度进行量化，目前已经成功构建出多种蛋白的适应度地形模型[4,45]。TEM-1 β-内酰胺酶是研究蛋白质突变进化和适应度效应的常用模型，已有多个实验室分析了该酶基因突变对其进化的影响。TEM-1 可水解 β-内酰胺类抗生素(如氨苄西林)，导致对青霉素类抗生素的耐药性。细菌在这些抗生素存在环境下生存和繁殖的能力取决于细胞内 TEM-1 的活性。Firnberg[46]和 Stiffler[47]课题组分别利用 DMS 对 TEM-1 β-内酰胺酶基因的突变适应度效应进行了测定，提供了突变和抗生素耐药性的适应度地形图，可帮助预测该基因在不同选择条件下进化的潜力和局限性。Stiffler 课题组的 DMS 结果表明，相对于选择强度(抗生素浓度)，波动的选择条件更可能选择出具有更高酶活性的 TEM-1。

2.3 病毒研究

DMS 也是研究病毒变异的重要工具[48-49]。新冠病毒(SARS-CoV-2)通过其刺突糖蛋白的受体结合结构域(receptor binding domain, RBD)与细胞受体血管紧张素转化酶 2 (angiotensin converting enzyme 2, ACE2)特异性结合完成对细胞的感染[50]。2020 年美国华盛顿大学 Jesse D. Bloom 课题组对 RBD 结构域 331 至 531 位共计 201 个氨基酸位点进行了突变扫描，逐个地将每个氨基酸位点突变成另外 19 种非野生型氨基酸，构建了一个包含 3 819 种(19×201)序列变异的突变文库，将文库引入酵母细胞表达变异 RBD，使其展示在细胞表面[51]。将经荧光标记的 ACE2 蛋白分子与文库酵母细胞孵育结合，利用流式分选收集表达高亲和力 RBD 变异的酵母细胞。最后通过高通量测序分析不同变异的富集程度，从而获得每一种变异与 ACE2 的亲和力指数，绘制出 RBD 结构域突变热图。该课题组将结果以交互模式呈现在网址 https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS，为新毒株感染力预测、疫苗设计、抗体药物设计提供了重要信息。

3 基于哺乳动物细胞平台的 DMS

哺乳动物细胞是开展 DMS 实验的重要平台，在哺乳动物细胞中蛋白(尤其是跨膜蛋白)可保持其天然构象，受糖基化、磷酸化等化学

修饰与调控，可介导信号转导并产生下游生化效应，产生疾病相关表型。

3.1 哺乳动物细胞 DMS 突变文库的构建

在哺乳动物细胞内进行 DMS 关键在于如何将文库高效引入哺乳动物细胞中，目前有两种常见的方法。第一种方法基于质粒，将突变文库克隆到质粒载体中，而后将质粒文库引入哺乳动物细胞系(如 HEK293T 细胞)中实现文库的瞬时或稳定整合表达[52-53]。在质粒文库制备过程中，历经细菌转化、质粒扩增、细胞转染等步骤，文库规模不断缩小，仅可达 10^5。第二种方法利用 CRISPR/Cas9 系统将变异文库引入哺乳动物细胞系，如人 HEK293T 细胞[54]、HAP1 细胞[55]或小鼠 Ba/F3 细胞[56]等[57]。CRISPR/Cas9 系统中 gRNA 引导 Cas9 蛋白靶向基因组位点并产生双链断裂，随后利用同源定向修复(homology-directed repair, HDR)原理将含有文库变异序列的供体模板引入断裂位点，实现哺乳动物基因组特定区域的饱和编辑[58]。CRISPR/Cas9-HDR 方法尚存在一定的局限性：(1) 需要优化 CRISPR 系统的靶向效率[59]；(2) 整合细胞和非整合细胞难以分辨；(3) 受当前单链寡核苷酸(single-stranded oligodeoxynucleotides, ssODNs)合成长度的限制，只能在有限的区域引入突变(<200 bp)。

本课题组近年致力于基于哺乳动物细胞系(以易转染 HEK293T 细胞为主)的高丰度文库构建策略的开发，并积极尝试将构建策略应用于基于哺乳动物细胞平台的 DMS[60-61]。课题组相继开发了 dsDNA-HR、*in-vitro*-ligation 两种瞬时纳米抗体文库构建策略，高通量测序结果显示两种方案皆可在 HEK293T 细胞中表达上百万种不同的纳米抗体序列[61]。同时，课题组新开发了一种 *in-vitro*-ligation-PCR 策略，结合 piggyBac 系统，实现了突变文库在哺乳动物细胞的稳定整合。进一步通过绿色荧光蛋白发色团(chromophore) DMS 实验证实了此策略的可行性，为未来基于哺乳动物细胞的 DMS 实验提供了操作简单且丰度更高的文库构建方案(图 3)。

3.2 以哺乳动物细胞为平台的 DMS 应用

3.2.1 抗体改造

抗体以高亲和力和高特异性等特性成为近年增长最快的药物种类，用于从癌症到自身免疫性疾病、病毒感染等多种重大疾病的治疗[8]。抗体 DMS 的一个重要方向是通过对抗体互补决定区(complementarity determining region, CDR)进行突变，对抗体亲和力进行优化[59]。哺乳动物细胞抗体优化平台具有其独有的优势，即抗体可被糖基化，与其最终的药物形式更为接近。Reddy 等[62]使用含简并密码子的 ssODNs 突变文库，通过 CRISPR/Cas9 介导的 HDR 机制将文库引入曲妥珠单抗的 CDRH3 区域，利用 FACS 对文库表达细胞进行筛选，鉴定出了结合力更强的抗体序列。Powers 等[53]对抗表皮生长因子受体(anti-epidermal growth factor receptor, anti-EGFR)抗体 225 的人源化版本进行了 DMS，首先利用简并密码子生成单氨基酸替换文库并克隆到载体中，并将载体转染至哺乳动物细胞 293c18 中建立了含有 1 000 多种突变的抗体展示文库，利用 FACS 筛选表达高亲和力抗体的细胞，通过高通量测序最终鉴定了 67 个可增强亲和力的点突变。

3.2.2 致病性突变鉴定

蛋白序列突变可导致疾病的发生[63]。2015 年美国医学遗传学与基因组学会根据突变的潜在影响将突变分为 5 类，包括致病性(pathogenic)、可能致病性(likely pathogenic)、意义未确定(uncertain significance)、可能良性(likely benign)和良性(benign)[64]。目前大多数基因错义突变(missense mutation)都被归类为意义未确定突变(variants of uncertain significance, VUS)，是具有

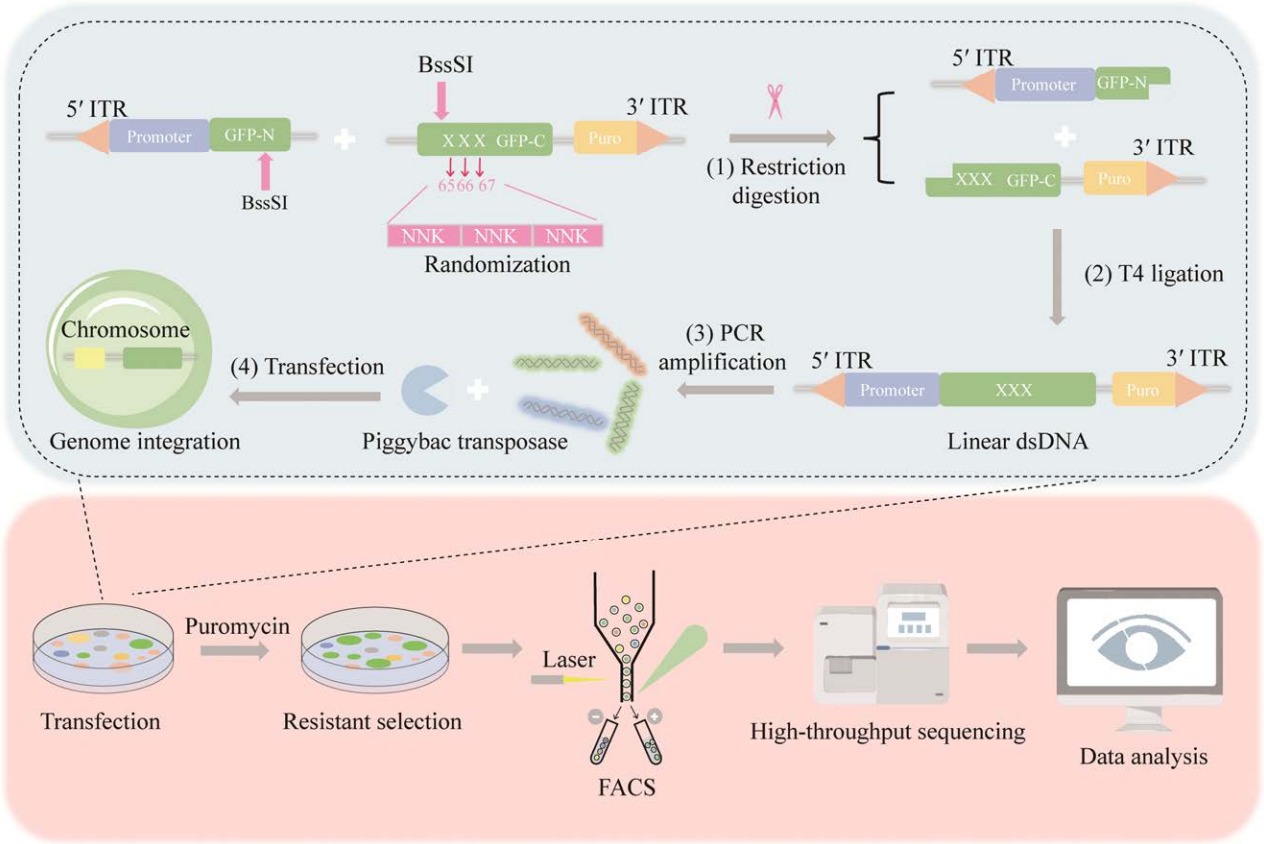

图 3　绿色荧光蛋白发色团深度突变扫描实验　通过 PCR 将 NNK 简并密码子引入绿色荧光蛋白发色团对应位点，利用 *in-vitro*-ligation-PCR 策略即在体外进行文库的酶切、连接和 PCR 扩增，并结合 piggyBac 系统，实现了突变文库在 HEK293T 细胞的稳定整合. 利用流式分选获得可产生绿色荧光的细胞，最后通过高通量测序分析所富集的发色团编码序列

Figure 3　GFP chromophore DMS experiment. The NNK degenerate codons were introduced into the corresponding site of the GFP chromophore by PCR, and the *in-vitro*-ligation-PCR strategy was used to construct mutational library through restrict digestion, ligation and PCR amplification, and by employing the piggyBac system, the mutant library was integrated into HEK293T cells genome. Green fluorescent cells were obtained by flow cytometry, and the enriched chromophore coding sequences were analyzed by high-throughput sequencing.

未知病理影响的基因突变[65]。这些不确定性给基因突变注释和疾病诊断带来重大挑战，DMS 可以对错义变异导致的疾病相关功能后果进行前瞻性评估[66-67]。乳腺癌易感基因 1 (breast cancer susceptibility gene 1, BRCA1)是重要的抑癌基因，BRCA1 胚系突变将显著提高女性乳腺癌、卵巢癌的发病风险[68-69]。约 10%的女性携带 BRCA1 错义突变，但目前仍有许多突变被列为“意义未明”，给患癌风险评估带来了巨大挑战。2018 年 Jay Shendure 课题组对位于 BRCA1 基因 13 个外显子上的近 4 000 种单核苷酸变异(single nucleotide variants, SNVs)进行了 DMS，对每一种 SNV 根据其对细胞同源重组修复能力的影响进行功能评分[70]。该研究共鉴定出约

400 个不影响 BRCA1 功能的 SNVs，以及约 300 个可影响 BRCA1 基因表达及功能的 SNVs。这些功能评分与已知的致病性变异或良性变异的临床评估相一致，表明 DMS 在 BRCA1 大规模 SNVs 功能注解上的可靠性。

PPARG 为过氧化物酶体增殖物激活受体 γ (peroxisome proliferator-activated receptor γ, PPARγ)编码基因，与孟德尔脂肪营养不良(Mendelian lipodystrophy)以及 2 型糖尿病患病风险增加相关。Majithia 等[71]在 THP1 细胞中引入了 PPARG 深度突变文库，根据 PPARγ 下游靶点 CD36 的表达水平进行流式细胞分选，分析了所有可能突变对 PPARG 蛋白功能的影响，他们对 55 个新错义突变进行了分类，并鉴定出了 6 个新致病性突变。

3.2.3 耐药突变鉴定

耐药性的产生是许多癌症治疗策略失败的重要原因[72]。DMS 可以分析靶蛋白不同突变对靶向药物的适应度，鉴定耐药性突变，为患者带来更为精准的诊疗信息[73]。Bolon 等[56]运用 CRISPR-Cas9 系统在白血病相关癌基因 BCR-ABL1 中引入突变，利用 DMS 方法系统地研究了不同突变对小鼠 Ba/F3 细胞增殖及对酪氨酸激酶抑制剂敏感性的影响。他们鉴定出了临床上所报道的所有 BCR-ABL1 病理性突变，获得了与临床流行率高度相关的结果。

BRAF 基因突变出现在多种肿瘤中，在黑色素瘤中突变频率特别高，高频致癌突变常发生在激酶结构域(kinase domain)。维罗非尼(vemurafenib)是 BRAF V600E 的选择性抑制剂，对黑色素瘤有显著抑制作用，可提高患者的生存率。然而，BRAF V600E 突变患者易出现维罗非尼继发性耐药。Wagenaar 等[74]利用 DMS 鉴定出 BRAF V600E 基因的一个继发性突变 L505H 可导致维罗非尼耐药。BRAF V600E/L505H 双突变可导致更高的激酶活性(较 V600E 更高)，并对不作用于 L505 位点的 BRAF 抑制剂较为敏感。

4 DMS 与机器学习

DMS 实验虽然实现了对序列——表型相关性的高通量量化分析，但在很多情况下受限于实验通量 DMS 尚不足以覆盖所有序列可能性。如何利用有限的 DMS 数据对蛋白序列——表型相关性进行全景式预测、描述？机器学习的深度介入与迭代是必由之路[75-76]。机器学习通过收集高质量的训练数据构建能够预测基因型——表型关系的模型，为蛋白质工程提供了一种强大的计算方法。近年机器学习已在抗体优化[77]、酶工程设计[78]、蛋白进化[79]等蛋白质工程领域得到了广泛应用。例如，Mason 等[62]利用哺乳动物细胞展示技术对曲妥珠单抗(trastuzumab)的 CDRH3 区域进行了 DMS，并将 DMS 数据库用于训练神经网络(neural network)。他们利用训练后的神经网络从包含近 1 亿种抗体变异序列的文库中预测了近千种候选抗体分子，并随机选取其中的 30 种变异抗体分子进行了后续验证，结果表明所有的 30 种抗体均可特异性结合 HER2 分子。此外，Taft 等[80]通过将 DMS 和机器学习相结合研究 SARS-CoV-2 RBD 结构域的变异规律以及这些变异介导免疫逃逸的可能性，开发出深度突变学习(deep mutational learning, DML)工具，为新毒株感染力预测、疫苗设计、抗体药物设计提供了重要预测工具。表 1 列举了上文中出现的典型 DMS 实验。

5 总结与展望

目前 DMS 技术已广泛应用于蛋白功能研究和工程改造等多个领域，有助于深入认识氨

表 1　深度突变扫描相关蛋白研究

Table 1　DMS for protein research

Protein/Gene	Selection system	Selection methodology	Variant space	References
hYAP65 WW domain	Phage display	Peptide ligand binding	Whole WW domain, >600 000 mutations	[3]
Protein G domain B1 (GB1)	mRNA display	Binding affinity screening	Whole GB1 domain	[20]
Protein degradation signal Deg1	Yeast	Growth rate screening	Deg1 residues 3–34, >30 000 mutations	[25]
CcdB	Bacterial	Cell toxicity screening	Whole gene, >1 200 single-site mutations	[36]
α-synuclein	Yeast	Growth phenotypes screening	2 660 single point mutations	[42]
TEM-1 β-lactamase	Bacterial	Antibiotic resistance	Whole gene, 4 997 mutations	[47]
SARS-CoV-2 spike glycoprotein	Yeast display	ACE2 binding affinity screening	Receptor binding domain, 201 positions	[51]
Anti-EGFR antibody	Mammalian cell display	FACS screening based on antigen affinity	CDRs, 1 060-point mutations	[53]
BCR-ABL1	Mammalian cell	Cells growth effect screening	311–319 positions	[56]
Therapeutic antibody trastuzumab	Mammalian cell display	FACS screening based on antigen affinity	About 1×10^4 variants	[62]
BRCA1	Mammalian cell	Homology-directed DNA repair function screening	13 exons	[70]
PPARG	Mammalian cell	FACS based on the expression level of CD36	Whole gene, 9 595 mutations	[71]
BRAFV600E	Mammalian cell	Drug resistance	77 amino acids surrounding binding site	[74]

基酸序列变异如何影响蛋白功能，协助设计出更好的疫苗、改造抗体药物、鉴定致病性突变等。该技术以其高通量、低成本、节省人力等优势推动了肿瘤、病毒、代谢性疾病等医学研究领域的发展。但目前并不是所有蛋白都可以开展 DMS 研究，需要根据蛋白已知功能设计有效的筛选体系。对于一些功能未知的蛋白，设计合理的高通量功能筛选方法是较为困难的。对于一些需要在组织、器官水平才能表现出功能的蛋白，目前还鲜有相关 DMS 研究。此外，如何合理设计突变文库，如何将突变文库高效地引入哺乳动物细胞，如何设计有效的筛选方案，如何利用机器学习协助处理并理解 DMS 数据，如何更好地将实验室 DMS 数据与临床数据相关联，这些都将是未来 DMS 研究应当重视的问题。

DMS 技术还处在快速发展阶段，随着高通量测序技术的发展、生物信息工具的开发以及人工智能的引入，该技术未来将有更为广泛的应用前景。同时，随着数据库共享以及数据可视化技术的发展，将为领域外研究人员读取、理解、应用 DMS 数据提供更多的便利，加速其研究创新。

REFERENCES

[1] KINNEY JB, MCCANDLISH DM. Massively parallel assays and quantitative sequence-function relationships[J]. Annual Review of Genomics and Human Genetics, 2019, 20: 99-127.

[2] WEIHJ, LI XH. Deep mutational scanning: a versatile tool in systematically mapping genotypes to phenotypes[J]. Frontiers in Genetics, 2023, 14:

1087267.
[3] FOWLER DM, ARAYA CL, FLEISHMAN SJ, KELLOGG EH, STEPHANY JJ, BAKER D, FIELDS S. High-resolution mapping of protein sequence-function relationships[J]. Nature Methods, 2010, 7(9): 741-746.
[4] ERNST A, GFELLER D, KAN ZY, SESHAGIRI S, KIM PM, BADER GD, SIDHU SS. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing[J]. Molecular BioSystems, 2010, 6(10): 1782.
[5] HIETPAS RT, JENSEN JD, BOLON DNA. Experimental illumination of a fitness landscape[J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(19): 7896-7901.
[6] WEILE J, ROTH FP. Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas[J]. Human Genetics, 2018, 137(9): 665-678.
[7] FOWLER DM, STEPHANY JJ, FIELDS S. Measuring the activity of protein variants on a large scale using deep mutational scanning[J]. Nature Protocols, 2014, 9(9): 2267-2284.
[8] HANNING KR, MINOT M, WARRENDER AK, KELTON W, REDDY ST. Deep mutational scanning for therapeutic antibody engineering[J]. Trends in Pharmacological Sciences, 2022, 43(2): 123-135.
[9] FIRNBERG E, OSTERMEIER M. PFunkel: efficient, expansive, user-defined mutagenesis[J]. PLoS One, 2012, 7(12): e52031.
[10] KILLE S, ACEVEDO-ROCHA CG, PARRA LP, ZHANG ZG, OPPERMAN DJ, REETZ MT, ACEVEDO JP. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis[J]. ACS Synthetic Biology, 2013, 2(2): 83-92.
[11] SUCHSLAND R, APPEL B, VIRTA P, MÜLLER S. Synthesis of fully protected trinucleotide building blocks on a disulphide-linked soluble support[J]. RSC Advances, 2021, 11(7): 3892-3896.
[12] PLESA C, SIDORE AM, LUBOCK NB, ZHANG D, KOSURI S. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes[J]. Science, 2018, 359(6373): 343-347.
[13] HUGHES RA, ELLINGTON AD. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology[J]. Cold Spring Harbor Perspectives in Biology, 2017, 9(1): a023812.
[14] KUIPER BP, PRINS RC, BILLERBECK S. Oligo pools as an affordable source of synthetic DNA for cost-effective library construction in protein-and metabolic pathway engineering[J]. ChemBioChem, 2022, 23(7): e202100507.
[15] KOSURI S, CHURCH GM. Large-scale *de novo* DNA synthesis: technologies and applications[J]. Nature Methods, 2014, 11(5): 499-507.
[16] FOWLER DM, FIELDS S. Deep mutational scanning: a new style of protein science[J]. Nature Methods, 2014, 11(8): 801-807.
[17] GARRETT ME, ITELL HL, CRAWFORD KHD, BASOM R, BLOOM JD, OVERBAUGH J. Phage-DMS: a comprehensive method for fine mapping of antibody epitopes[J]. iScience, 2020, 23(10): 101622.
[18] NEWTON MS, CABEZAS-PERUSSE Y, TONG CL, SEELIG B. *In vitro* selection of peptides and proteins—advantages of mRNA display[J]. ACS Synthetic Biology, 2020, 9(2): 181-190.
[19] HUANG YC, WIEDMANN MM, SUGA H. RNA display methods for the discovery of bioactive macrocycles[J]. Chemical Reviews, 2019, 119(17): 10360-10391.
[20] OLSON CA, WU NC, SUN R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain[J]. Current Biology, 2014, 24(22): 2643-2651.
[21] FUJINO Y, FUJITA R, WADA K, FUJISHIGE K, KANAMORI T, HUNT L, SHIMIZU Y, UEDA T. Robust *in vitro* affinity maturation strategy based on interface-focused high-throughput mutational scanning[J]. Biochemical and Biophysical Research Communications, 2012, 428(3): 395-400.
[22] LEDSGAARD L, LJUNGARS A, RIMBAULT C, SØRENSEN CV, TULIKA T, WADE J, WOUTERS Y, MCCAFFERTY J, LAUSTSEN AH. Advances in antibody phage display technology[J]. Drug Discovery Today, 2022, 27(8): 2151-2169.
[23] AHMED S, BHASIN M, MANJUNATH K, VARADARAJAN R. Prediction of residue-specific contributions to binding and thermal stability using yeast surface display[J]. Frontiers in Molecular Biosciences, 2022, 8: 800819.
[24] KOCH P, SCHMITT S, HEYNISCH A, GUMPINGER A, WÜTHRICH I, GYSIN M, SHCHERBAKOV D, HOBBIE SN, PANKE S, HELD M. Optimization of the antimicrobial peptide Bac7 by deep mutational scanning[J]. BMC Biology, 2022, 20(1): 1-21.

[25] KIM I, MILLER CR, YOUNG DL, FIELDS S. High-throughput analysis of *in vivo* protein stability[J]. Molecular & Cellular Proteomics, 2013, 12(11): 3370-3378.
[26] ARAYA CL, FOWLER DM. Deep mutational scanning: assessing protein function on a massive scale[J]. Trends in Biotechnology, 2011, 29(9): 435-442.
[27] van DIJK EL, JASZCZYSZYN Y, NAQUIN D, THERMES C. The third revolution in sequencing technology[J]. Trends in Genetics, 2018, 34(9): 666-681.
[28] WANG YH, ZHAO Y, BOLLAS A, WANG YR, AU KF. Nanopore sequencing technology, bioinformatics and applications[J]. Nature Biotechnology, 2021, 39(11): 1348-1365.
[29] FOWLER DM, ARAYA CL, GERARD W, FIELDS S. Enrich: software for analysis of protein function by enrichment and depletion of variants[J]. Bioinformatics, 2011, 27(24): 3430-3431.
[30] BLOOM JD. Software for the analysis and visualization of deep mutational scanning data[J]. BMC Bioinformatics, 2015, 16(1): 1-13.
[31] FAURE AJ, SCHMIEDEL JM, BAEZA-CENTURION P, LEHNER B. DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies[J]. Genome Biology, 2020, 21(1): 1-23.
[32] RUBIN AF, GELMAN H, LUCAS N, BAJJALIEH SM, PAPENFUSS AT, SPEED TP, FOWLER DM. A statistical framework for analyzing deep mutational scanning data[J]. Genome Biology, 2017, 18(1): 1-15.
[33] HOSKINS I, SUN S, COTE A, ROTH FP, CENIK C. Satmut_utils: a simulation and variant calling package for multiplexed assays of variant effect[J]. Genome Biology, 2023, 24(1): 1-27.
[34] BRABERG H, ECHEVERRIA I, KAAKE RM, SALI A, KROGAN NJ. From systems to structure—using genetic data to model protein structures[J]. Nature Reviews Genetics, 2022, 23(6): 342-354.
[35] MCLAUGHLIN RN, POELWIJK FJ, RAMAN A, GOSAL WS, RANGANATHAN R. The spatial architecture of protein function and adaptation[J]. Nature, 2012, 491(7422): 138-142.
[36] ADKAR BV, TRIPATHI A, SAHOO A, BAJAJ K, GOSWAMI D, CHAKRABARTI P, SWARNKAR MK, GOKHALE RS, VARADARAJAN R. Protein model discrimination using mutational sensitivity derived from deep sequencing[J]. Structure, 2012, 20(2): 371-381.
[37] ROLLINS NJ, BROCK KP, POELWIJK FJ, STIFFLER MA, GAUTHIER NP, SANDER C, MARKS DS. Inferring protein 3D structure from deep mutation scans[J]. Nature Genetics, 2019, 51(7): 1170-1176.
[38] SCHMIEDEL JM, LEHNER B. Determining protein structures using deep mutagenesis[J]. Nature Genetics, 2019, 51(7): 1177-1186.
[39] BOLOGNESI B, FAURE AJ, SEUMA M, SCHMIEDEL JM, TARTAGLIA GG, LEHNER B. The mutational landscape of a prion-like domain[J]. Nature Communications, 2019, 10: 4162.
[40] GRAY VE, SITKO K, KAMENI FZN, WILLIAMSON M, STEPHANY JJ, HASLE N, FOWLER DM. Elucidating the molecular determinants of aβ aggregation with deep mutational scanning[J]. G3: Genes, Genomes, Genetics, 2019, 9(11): 3683-3689.
[41] EBO JS, GUTHERTZ N, RADFORD SE, BROCKWELL DJ. Using protein engineering to understand and modulate aggregation[J]. Current Opinion in Structural Biology, 2020, 60: 157-166.
[42] NEWBERRY RW, LEONG JT, CHOW ED, KAMPMANN M, DEGRADO WF. Deep mutational scanning reveals the structural basis for α-synuclein activity[J]. Nature Chemical Biology, 2020, 16(6): 653-659.
[43] ARNOLD FH. Innovation by evolution: bringing new chemistry to life (Nobel lecture)[J]. Angewandte Chemie International Edition, 2019, 58(41): 14420-14426.
[44] BLANCO C, JANZEN E, PRESSMAN A, SAHA R, CHEN IA. Molecular fitness landscapes from high-coverage sequence profiling[J]. Annual Review of Biophysics, 2019, 48: 1-18.
[45] CANALE AS, COTE-HAMMARLOF PA, FLYNN JM, BOLON DN. Evolutionary mechanisms studied through protein fitness landscapes[J]. Current Opinion in Structural Biology, 2018, 48: 141-148.
[46] FIRNBERG E, LABONTE JW, GRAY JJ, OSTERMEIER M. A comprehensive, high-resolution map of a gene's fitness landscape[J]. Molecular Biology and Evolution, 2016, 33(5): 1378.
[47] STIFFLER MA, HEKSTRA DR, RANGANATHAN R. Evolvability as a function of purifying selection in TEM-1 β-lactamase[J]. Cell, 2015, 160(5): 882-892.
[48] NARAYANAN KK, PROCKO E. Deep mutational scanning of viral glycoproteins and their host receptors[J]. Frontiers in Molecular Biosciences, 2021,

8: 636660.
[49] BURTON TD, EYRE NS. Applications of deep mutational scanning in virology[J]. Viruses, 2021, 13(6): 1020.
[50] FRANCINO-URDANIZ IM, WHITEHEAD TA. An overview of methods for the structural and functional mapping of epitopes recognized by anti-SARS-CoV-2 antibodies[J]. RSC Chemical Biology, 2021, 2(6): 1580-1589.
[51] STARR TN, GREANEY AJ, HILTON SK, ELLIS D, CRAWFORD KHD, DINGENS AS, NAVARRO MJ, BOWEN JE, ALEJANDRA TORTORICI M, WALLS AC, KING NP, VEESLER D, BLOOM JD. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding[J]. Cell, 2020, 182(5): 1295-1310.e20.
[52] WRENBECK EE, KLESMITH JR, STAPLETON JA, ADENIRAN A, TYO KEJ, WHITEHEAD TA. Plasmid-based one-pot saturation mutagenesis[J]. Nature Methods, 2016, 13(11): 928-930.
[53] FORSYTH CM, JUAN V, AKAMATSU Y, DUBRIDGE RB, DOAN M, IVANOV AV, MA ZY, POLAKOFF D, RAZO J, WILSON K, POWERS DB. Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing[J]. mAbs, 2013, 5(4): 523-532.
[54] FINDLAY GM, BOYLE EA, HAUSE RJ, KLEIN JC, SHENDURE J. Saturation editing of genomic regions by multiplex homology-directed repair[J]. Nature, 2014, 513(7516): 120-123.
[55] HUNDLEY FV, TOCZYSKI DP. Chemical-genetic CRISPR-Cas9 screens in human cells using a pathway-specific library[J]. STAR Protocols, 2021, 2(3): 100685.
[56] MA LY, BOUCHER JI, PAULSEN J, MATUSZEWSKI S, EIDE CA, OU JH, EICKELBERG G, PRESS RD, ZHU LJ, DRUKER BJ, BRANFORD S, WOLFE SA, JENSEN JD, SCHIFFER CA, GREEN MR, BOLON DN. CRISPR-Cas9-mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(44): 11751-11756.
[57] KURATA M, YAMAMOTO K, MORIARITY BS, KITAGAWA M, LARGAESPADA DA. CRISPR/Cas9 library screening for drug target discovery[J]. Journal of Human Genetics, 2018, 63(2): 179-186.
[58] JANG DE, LEE JY, LEE JH, KOO OJ, BAE HS, JUNG MH, BAE JH, HWANG WS, CHANG YJ, LEE YH, LEE HW, YEOM SC. Multiple sgRNAs with overlapping sequences enhance CRISPR/Cas9-mediated knock-in efficiency[J]. Experimental & Molecular Medicine, 2018, 50(4): 1-9.
[59] MASON DM, WEBER CR, PAROLA C, MENG SM, GREIFF V, KELTON WJ, REDDY ST. High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis[J]. Nucleic Acids Research, 2018, 46(14): 7436-7449.
[60] LI S, SU WJ, ZHANG CZ. Linear double-stranded DNAs as innovative biological parts to implement genetic circuits in mammalian cells[J]. The FEBS Journal, 2019, 286(12): 2341-2354.
[61] ZHAO YJ, WANG Y, SU WJ, LI S. Construction of synthetic nanobody library in mammalian cells by dsDNA-based strategies[J]. ChemBioChem, 2021, 22(20): 2957-2965.
[62] MASON DM, FRIEDENSOHN S, WEBER CR, JORDI C, WAGNER B, MENG SM, EHLING RA, BONATI L, DAHINDEN J, GAINZA P, CORREIA BE, REDDY ST. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence *via* deep learning[J]. Nature Biomedical Engineering, 2021, 5(6): 600-612.
[63] STENSON PD, MORT M, BALL EV, HOWELLS K, PHILLIPS AD, ST THOMAS N, COOPER DN. The human gene mutation database: 2008 update[J]. Genome Medicine, 2009, 1(1): 1-6.
[64] RICHARDS S, AZIZ N, BALE S, BICK D, DAS S, GASTIER-FOSTER J, GRODY WW, HEGDE M, LYON E, SPECTOR E, VOELKERDING K, REHM HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology[J]. Genetics in Medicine, 2015, 17(5): 405-424.
[65] MANOLIO TA, FOWLER DM, STARITA LM, HAENDEL MA, MACARTHUR DG, BIESECKER LG, WORTHEY E, CHISHOLM RL, GREEN ED, JACOB HJ, MCLEOD HL, RODEN D, RODRIGUEZ LL, WILLIAMS MS, COOPER GM, COX NJ, HERMAN GE, KINGSMORE S, LO C, LUTZ C, et al. Bedside back to bench: building bridges between basic and clinical genomic research[J]. Cell, 2017, 169(1): 6-12.

[66] HICKS MA, HOU CYC, IRANMEHR A, MAROSI K, KIRKNESS E. Target discovery using biobanks and human genetics[J]. Drug Discovery Today, 2020, 25(2): 438-445.
[67] LIVESEY BJ, MARSH JA. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations[J]. Molecular Systems Biology, 2020, 16(7): 1-12.
[68] STARITA LM, ISLAM MM, BANERJEE T, ADAMOVICH AI, GULLINGSRUD J, FIELDS S, SHENDURE J, PARVIN JD. A multiplex homology-directed DNA repair assay reveals the impact of more than 1 000 BRCA1 missense substitution variants on protein function[J]. The American Journal of Human Genetics, 2018, 103(4): 498-508.
[69] KRAIS JJ, JOHNSON N. BRCA1 mutations in cancer: coordinating deficiencies in homologous recombination with tumorigenesis[J]. Cancer Research, 2020, 80(21): 4601-4609.
[70] FINDLAY GM, DAZA RM, MARTIN B, ZHANG MD, LEITH AP, GASPERINI M, JANIZEK JD, HUANG XF, STARITA LM, SHENDURE J. Accurate classification of BRCA1 variants with saturation genome editing[J]. Nature, 2018, 562(7726): 217-222.
[71] MAJITHIA AR, DIABETES CONSORTIUM UM, TSUDA B, AGOSTINI M, GNANAPRADEEPAN K, RICE R, PELOSO G, PATEL KA, ZHANG XL, BROEKEMA MF, PATTERSON N, DUBY M, SHARPE T, KALKHOVEN E, ROSEN ED, BARROSO I, ELLARD S, KATHIRESAN S, O'RAHILLY S, CHATTERJEE K, et al. Prospective functional classification of all possible missense variants in PPARG[J]. Nature Genetics, 2016, 48(12): 1570-1575.
[72] HOLOHAN C, van SCHAEYBROECK S, LONGLEY DB, JOHNSTON PG. Cancer drug resistance: an evolving paradigm[J]. Nature Reviews Cancer, 2013, 13(10): 714-726.
[73] PINES G, FANKHAUSER RG, ECKERT CA. Predicting drug resistance using deep mutational scanning[J]. Molecules, 2020, 25(9): 2265.
[74] WAGENAAR TR, MA LY, ROSCOE B, PARK SM, BOLON DN, GREEN MR. Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain[J]. Pigment Cell & Melanoma Research, 2014, 27(1): 124-133.
[75] SARFATI H, NAFTALY S, PAPO N, KEASAR C. Predicting mutant outcome by combining deep mutational scanning and machine learning[J]. Proteins: Structure, Function, and Bioinformatics, 2022, 90(1): 45-57.
[76] GELMAN S, FAHLBERG SA, HEINZELMAN P, ROMERO PA, GITTER A. Neural networks to learn protein sequence-function relationships from deep mutational scanning data[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(48): 1-12.
[77] GREIFF V, WEBER CR, PALME J, BODENHOFER U, MIHO E, MENZEL U, REDDY ST. Learning the high-dimensional immunogenomic features that predict public and private antibody repertoires[J]. The Journal of Immunology, 2017, 199(8): 2985-2997.
[78] VANELLA R, KOVACEVIC G, DOFFINI V, FERNÁNDEZ de SANTAELLA J, NASH MA. High-throughput screening, next generation sequencing and machine learning: advanced methods in enzyme engineering[J]. Chemical Communications, 2022, 58(15): 2455-2467.
[79] FERNANDEZ-de-COSSIO-DIAZ J, UGUZZONI G, PAGNANI A. Unsupervised inference of protein fitness landscape from deep mutational scan[J]. Molecular Biology and Evolution, 2021, 38(1): 318-328.
[80] TAFT JM, WEBER CR, GAO B, EHLING RA, HAN J, FREI L, METCALFE SW, OVERATH MD, YERMANOS A, KELTON W, REDDY ST. Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain[J]. Cell, 2022, 185: 4008-4022.e14.

(本文责编 郝丽芳)