



蛋白质相互作用技术——功能基因组学研究利器

冯然^{1,2}, 刘峰¹, 吴苏², 李若菲¹, 松阳洲^{1,2*}

1. 中山大学生命科学学院, 广州 510275;

2. 中山大学孙逸仙纪念医院, 广州 510120

* 联系人, E-mail: songyanz@mail.sysu.edu.cn

收稿日期: 2024-05-11; 接受日期: 2024-07-19; 网络版发表日期: 2024-09-09

摘要 细胞内的生物活动复杂且动态, 依靠蛋白质之间相互作用形成的动态网络完成细胞内与细胞间的信息传递和交流. 因此, 研究蛋白质相互作用对于揭示蛋白质的功能尤为重要. 本团队在蛋白互作研究领域深耕多年, 先后开发并利用多种蛋白系统性互作研究技术对细胞和生命体内的重要蛋白网络进行功能解析和机制探索, 主要研究集中在端粒稳态及调控、基因组损伤修复及细胞内重要信号通路调控机制等方面. 本文首先阐述了目前常用的蛋白质相互作用技术及其特征, 随后以本课题组主要研究成果为例说明这些技术如何推动生物学相关研究领域的发展; 最后总结了人工智能技术对于本领域的促进作用, 以及蛋白互作领域未来的发展方向. 本综述旨在展现蛋白质相互作用技术的开发如何促进功能基因组学研究的深入发展, 以及用实例展现科技人员一如既往坚持科技创新的重要意义.

关键词 蛋白质相互作用技术, 邻近标记, 蛋白互补分析技术

蛋白质是生命活动的基础, 依靠蛋白质之间相互作用(protein-protein interaction, PPI)形成的信号网络完成复杂且动态的生物活动^[1]. 人体内的PPI非常复杂: 目前已知人类基因组中蛋白编码基因为两万多个^[2], UniProt数据库中收录的蛋白则超过20万种^[3], 而PPI的数目则进一步指数递增^[4]. IntAct数据库(<https://www.ebi.ac.uk/intact>)中收录的人源PPI条目超过76万^[5], 而STRING数据库(<https://string-db.org/>)^[6]和BioGRID数据库(<https://thebiogrid.org/>)^[7]中收录的PPI均超过100万. 而植物的蛋白组平均包含约36795个蛋白, 其中已鉴定到的蛋白-蛋白互作对数量在75000~150000之间^[8,9]. PPI的丰富性不仅来源于蛋白种类的多样, 对

于单个蛋白来说, 其可变剪接、翻译后修饰, 以及蛋白突变等因素均可能影响其蛋白互作种类^[10]. 此外, 蛋白互作组具有时空特异性, 在不同组织细胞、不同发育阶段、不同生理病理状态下, 均存在较大差异^[11~13]. 然而, 由于研究方法的局限性, 目前已发现的PPI仅是全部互作的冰山一隅, 瞬时及微弱的蛋白质间相互作用目前仍难以鉴定, 仍有大量未知的蛋白互作有待发掘. 同时, 实验平台和分析方法的差异所产生的大量批次效应和数据噪声, 使得对不同PPI数据集进行均一化仍面临较大困难. 因此, 蛋白互作研究技术的开发将是开展功能基因组学的关键所在, 也直接决定了蛋白组学研究的深度.

引用格式: 冯然, 刘峰, 吴苏, 等. 蛋白质相互作用技术——功能基因组学研究利器. 中国科学: 生命科学, 2024, 54: 1603–1613

Feng R, Liu F, Wu S, et al. Protein-protein interaction techniques—a powerful tool for functional genomics (in Chinese). Sci Sin Vitae, 2024, 54: 1603–1613, doi: 10.1360/SSV-2024-0151

按技术原理可将鉴定PPI的策略大致分为以免疫共沉淀(co-immunoprecipitation, co-IP)、导向性多肽库(oriented peptide library, OPL)为代表鉴定细胞外PPI的技术, 和以酵母双杂交(yeast two-hybrid, Y2H)、蛋白互补分析技术(protein-fragment complementation assay, PCA)、生物素邻近标记技术为代表的鉴定细胞内PPI的技术(表1)。本文以上述PPI互动技术的先后发展历程作为顺序, 回顾蛋白互动研究领域的演变过程。

1 鉴定细胞外蛋白质相互作用

免疫共沉淀是PPI研究中应用最广泛的技术之一, 通过抗原-抗体之间的高亲和力, 将诱饵蛋白bait捕获在固相化基质上, 同时bait的互动蛋白也被结合并沉淀下来(图1)。该方法可偶联质谱使用, 使得一次实验可以高效鉴定到大量目的蛋白的潜在互动蛋白^[14]。2021年Huttlin等人^[13]在HEK293T细胞和HCT116细胞中外源表达HA/FLAG标签的蛋白, 并采用大规模的亲和沉淀偶联质谱(affinity purification-mass spectrometry, AP-MS)实验捕获这些蛋白的互动组, 命名为Bio-Plex 3.0, 其中包含HEK293T细胞中14586个蛋白间的

118162个互动, 以及HCT116细胞中的5522个蛋白的互动组数据。

免疫共沉淀最大的优势在于无需对细胞进行基因层面的操作, 因此对于难以被转染、侵染的原代细胞或者临床组织, 免疫共沉淀是鉴定PPI的首选方法。但是该方法仍存在一些局限, 包括: (i) 裂解液对膜蛋白等部分蛋白溶解性差, 导致其难以被纯化并鉴定; (ii) 该方法检测的蛋白互动发生在裂解液中, 与细胞内并不相同, 一些仅在特定胞内位点或信号通路影响下发生的互动可能被遗漏; (iii) 弱互动、瞬时互动难以被鉴定; (iv) 严格的洗涤条件下部分互动信息会丢失^[4]; (v) 存在非特异性结合, 导致背景蛋白的富集, 影响实验结果的准确性。

此外, 由于体内存在修饰的蛋白质比例较低, 利用免疫共沉淀方法对此类蛋白鉴定存在困难。且由于蛋白质之间的相互作用通常由一些具有特征性的基序介导, 免疫共沉淀方法也无法对此类蛋白进行鉴定。为了弥补免疫共沉淀存在的诸多缺陷, 本课题组早在1993年就利用自主开发的OPL技术, 成功发现了著名的SH2结构域可以与磷酸化酪氨酸基序结合^[15], 并解析了酪氨酸激酶与底物的相互作用特征^[16], 开启了信号

表1 各种蛋白质相互作用方法比较

Table 1 Advantages and disadvantages of research methods based on biochemistry and molecular biology

蛋白质相互作用方法	优势	缺点
免疫共沉淀	检测自然状态下蛋白质的相互作用	不能检测瞬时蛋白互动和低亲和力蛋白互动; 检测到的可能不是蛋白间直接相互作用
酵母双杂交技术	检测活细胞内相互作用; 可以应用于高通量筛选	不能反映人源细胞内相互作用; 无法反映膜蛋白、无序蛋白、分泌蛋白等相互作用; 无法检测蛋白翻译后修饰; 依赖文库进行高通量筛选; 假阳性率高; 不能检测瞬时蛋白互动
荧光共振能量转移	能够实时监测分子相互作用; 灵敏度高; 可以进行定量分析	受到蛋白间距离的限制, 不适用于大蛋白复合物或其他复杂体系; 应用受限于部分细胞结构, 存在非特异荧光信号和背景荧光信号的干扰
双分子荧光互补	直接可视化细胞内蛋白互动; 对细胞扰动小; 操作简单; 灵敏度高	存在假阳性信号; 受细胞环境影响, 可能影响荧光复合物的形成和稳定性; 蛋白质标记片段的插入可能引起蛋白质结构改变; 定型分析蛋白互动存在与否, 无法量化强度
高斯荧光素酶片段互补	分子量小、灵敏度高、半衰期短; 可以检测细胞内动态变化并量化强度	受细胞环境影响, 依赖于蛋白折叠, 存在假阳性信号及背景噪声
邻近连接技术	细胞内可视化蛋白间相互作用	非特异性连接存在, 存在假阳性信号; 标记可能会影响蛋白质的结构和功能
基于过氧化酶的邻近标记技术	分辨率高; 能够反映活细胞原位的蛋白质间相互作用; 灵敏度高; 适用于高通量筛选	具有细胞毒性, 无法应用于个体水平; 较高的背景噪声和非特异性标记; 标记时间过短, 难于捕获生物学过程中的动态互动变化; 融合表达APEX2可能会影响某些蛋白的正确定位
基于生物素连接酶的邻近标记技术	分辨率高; 能够反映活细胞原位的蛋白质间相互作用; 灵敏度高; 适用于高通量筛选; 特异性高; 细胞毒性低	存在一定程度的假阳性(可能是邻近蛋白, 非直接蛋白间互动); 有随机标记产生的背景; 低丰度蛋白的捕捉受到生物素连接酶活性的影响

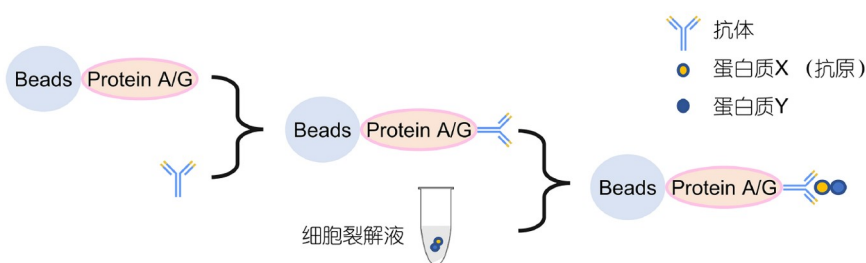


图 1 免疫共沉淀原理

Figure 1 Principle of co-immunoprecipitation

转导通路模块化研究的新时代. 该技术可与蛋白芯片技术结合发展为导向性多肽库芯片(oriented peptide array library, OPAL)^[17], 进一步高效筛选鉴定多种结构域的识别基序和寻找激酶在细胞内的潜在底物. 利用人工合成的库, 能够特异性寻找蛋白质的特征性修饰以及蛋白质相互作用的界面, 并能帮助鉴定蛋白质间瞬时相互作用. 此技术丰富了信号转导研究的蛋白互作基础^[18], 搭建了实验室里可广泛使用的第一代蛋白相互作用平台, 并将其用于探测端粒蛋白间的相互作用^[19].

2 鉴定细胞内蛋白质相互作用

酵母双杂交技术, 最早由Fields和Song在1989年建立^[20]. Y2H的报告系统包括酵母转录因子的DNA结合结构域(binding domain, BD)和转录激活域(activation domain, AD), 以及在完整BD-AD复合体驱动下转录的报告基因; 单独的AD和BD并不能激活转录反应, 当AD和BD分别与互作的蛋白对形成融合蛋白时, 互作蛋白对的结合使AD和BD在空间上靠近, 激活报告基因的表达^[21](图2). 2020年, Luck等人^[22]采用Y2H构建了迄今最大的人类蛋白间二元物理互作组——HuRI, 包含8000种蛋白间的53000个高置信互作. 然而, 酵母中蛋白的翻译后修饰与哺乳动物有很大不同, 同时PPI也可能缺乏必要的辅助因子或上游信号, 使Y2H技术具有一定局限性, HuRI鉴定的高置信互作仅占已知的11%^[23]. 因此, 目前更加倾向于在人类细胞中利用类似酵母双杂交的方法进行蛋白质相互作用组的筛选和鉴定.

随着20世纪90年代对荧光蛋白的结构生物学、生物物理学和生物化学的深入研究, 荧光共振能量转移

(fluorescence resonance energy transfer, FRET)方法被开发出来(图3). 1997年, Miyawaki等人^[24]首先报道了采用不同变体绿色荧光蛋白(green fluorescent protein, GFP)在活体HeLa细胞中实时观察蛋白互作的方法, 绿色发射光GFP、黄色发射光GFP分别与钙调素、钙调素结合多肽M13融合表达, 在钙离子的存在下, 钙调素与M13的相互结合促使两种GFP相互靠近并发生荧光共振能量转移, 导致发射光波长发生改变. 然而, 未发生FRET的发射光与FRET发射光容易形成串扰, 仅在强互作蛋白对中FRET信号才不会被掩盖. 此外, 由于FRET的荧光信号检测需要特定仪器, 所检测分子的空间距离需要<10 nm, 且无法与组学方法联用进行大规模筛选. 因此, 仍需开发可以应用于人类细胞且能兼容大规模筛选的蛋白互作方法.

蛋白互补分析技术是一种基于bait和prey蛋白分别与两个互补的报告蛋白片段融合的技术, 当它们在空间上靠近时重新组装, 并恢复完整蛋白的功能, 此报告蛋白可以是酶或荧光蛋白. 酶活性或荧光信号的恢复可以作为bait和prey蛋白之间相互作用的指标(图4). PCA技术不仅可以在活细胞内检测PPI, 还可以检测包括不同亚细胞区室内或细胞器, 不同刺激条件诱导下形成的PPI. 目前已经开发的PCA技术衍生物有很多, 包括双分子荧光互补(bimolecular fluorescence complementation, BiFC)、高斯荧光素酶蛋白片段互补测定技术(Gaussia luciferase protein-fragment complementation assay, GLuc PCA)等被应用于多种不同研究场景, 用以解决特定问题. 1994年, Johnson等人^[25]通过分裂的泛素酶片段首次建立了PCA. 随后, 二氢叶酸还原酶(dihydrofolate reductase, DHFR)、甘氨酸酰胺核糖核苷酸转化酶、氨基糖苷激酶、TEM β -内酰胺酶^[26-29]等也作为报告蛋白应用于PCA研究中, 特别是

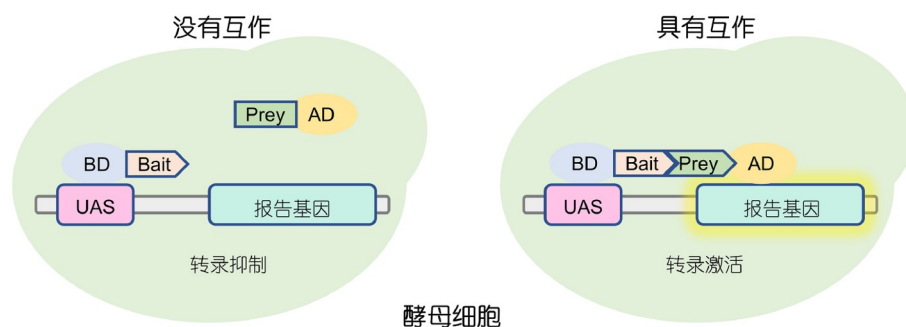


图2 酵母双杂交原理

Figure 2 Principle of yeast two-hybrid system

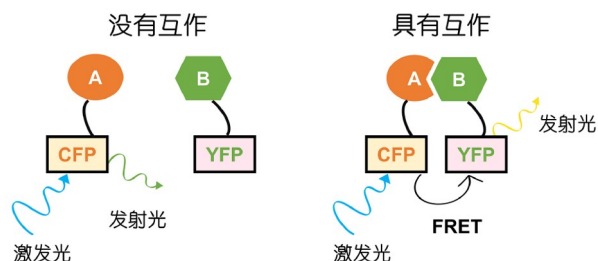


图3 荧光共振能量原理

Figure 3 Principle of fluorescence resonance energy transfer

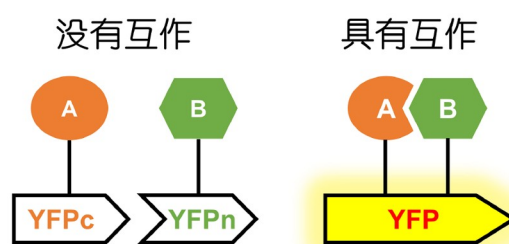


图4 双分子荧光互补原理

Figure 4 Principle of bimolecular fluorescence complementation

用于筛选新的蛋白互作对。此外, 荧光蛋白和荧光素酶是另一组广泛应用的PCA报告蛋白。例如, BiFC是一种较好的在活细胞中基于荧光蛋白互补观察PPI的手段。2000年, Ghosh等人^[30]首先证实将GFP分成两部分后, 能在体外和大肠杆菌中组装成发出绿色荧光的完整蛋白; 随后, Hu等人^[31]将此应用到PPI领域, 他们将YFP分成两部分并与互作蛋白对bZIP、Rel转录因子构建融合蛋白, 观察到完整YFP荧光的形成, 证明了BiFC的可行性。本课题组基于蛋白互补分析技术, 成功改善了BiFC技术, 通过与人类ORFeome文库中的18000个基因cDNA应用相结合搭建了实验室第二代蛋白质互作技术平台。利用此技术本课题组鉴定了一系列端粒结合蛋白、干细胞调控因子及其作用机制^[32-39]。BiFC具有细胞活体蛋白互作可视化的优势, 能直观、快速地分析活细胞中蛋白质之间相互作用及定位, 特别是对瞬时蛋白互作有捕捉效应, 所以结合流式细胞分析能灵敏地检测到微弱和瞬态的相互作用, 在捕捉弱相互作用PPI、观察PPI响应外界条件变化具有无可比拟的优势。本课题组长期致力于重复序列

DNA损伤修复研究, 利用BiFC技术发现TPP1维持端粒稳定性^[40-42], 发现多个蛋白能够结合到重复序列, 如DAXX结合并调节逆转录转座子的表达^[43], 端粒体蛋白可以结合端粒外的重复序列^[44], Bend家族蛋白结合并调控基因组重复序列^[39]等。然而, 由于荧光蛋白片段互补后不可逆转, 容易造成BiFC假阳性信号^[45,46]。此外, 该技术仍不能摆脱对高质量cDNA文库的依赖, 其日常维护难度较高。高斯荧光素酶由于其分子量小、灵敏度高、半衰期短等特点也作为报告基因被广泛使用^[47]。GLuc PCA是一种可逆且灵敏的探测手段, 用于研究动态和瞬时的PPI。将高斯荧光素酶的N端和C端片段分别融合到bait和prey蛋白上, 当bait蛋白与prey蛋白相互作用时, 分裂的荧光素酶片段重构并催化底物发光, 不仅可以检测PPI动力学, 还可以量化PPI的相对强度。相比于BiFC, Gluc PCA互补发光信号的形成是可逆的, 能够实时反映PPI的动态变化, 更具优势。但PCA存在一定程度的背景信号干扰, 以及融合报告系统可能影响蛋白折叠、定位、互作、功能等缺点, 因此研究人员仍在寻找在活细胞内检测瞬时和弱

相互作用且背景较低的PPI方法。

基于邻近位置关系研究蛋白相互作用的策略近年来已经得到充分的拓展。蛋白质会与邻近的分子形成氢键、离子键、范德华力与疏水作用等, 这些构成了PPI的基础。邻近类技术突破了传统技术的诸多局限性, 主要包括基于抗体-DNA复合物的邻近连接分析(proximity ligation assay, PLA)系统、APEX与生物素连接酶邻近标记系统等。

邻近连接分析是一种高灵敏度的免疫分析检测技术, 在体内和体外均可使用。近年来更多地用于检测内源蛋白相互作用(<40 nm)^[48,49]并分析基因状态以及相应蛋白表达之间的关联, 并更透彻地了解癌症及其他复杂疾病。该方法通过一对标记有一段寡聚脱氧核苷酸(单链DNA)的单克隆或者多克隆抗体的探针(PLA probe)识别目的蛋白, 当这两个探针识别互作蛋白对时, 两个探针之间的距离靠近, 产生了所谓的邻近效应(proximity)。此时, 通过加入一段分别与连接在抗体上的DNA互补的连接寡聚脱氧核苷酸(connector oligonucleotides), PLA probe上的DNA就会通过配对互补作用, 与该段DNA互补; 然后在连接酶的作用下, PLA probe上的片段DNA被连接在一起形成一条新的DNA片段。通过荧光PCR可以扩增并对该新的DNA片段进行定量, 从而对应的定量目标蛋白(图5)。2011年, 美国俄亥俄州立大学的研究人员将传统的荧光原位杂交与PLA相结合, 能同时观察细胞中的基因数量和蛋白表达。由于检测到的是放大后的信号, 因此也可以用于检测一些较弱的蛋白质间相互作用。然而PLA技术较大程度上依赖于合适的内源抗体, 同时不能用于检测活细胞内的蛋白质间相互作用。

邻近标记(proximity-dependent biotinylation, PDB)

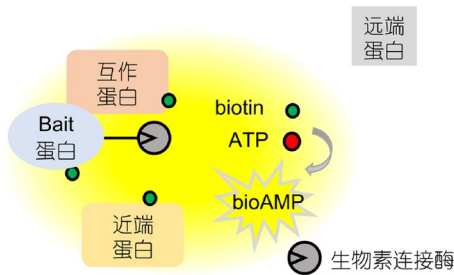


图6 基于邻近依赖性生物素化标记原理

Figure 6 Principle of proximity-dependent biotinylation

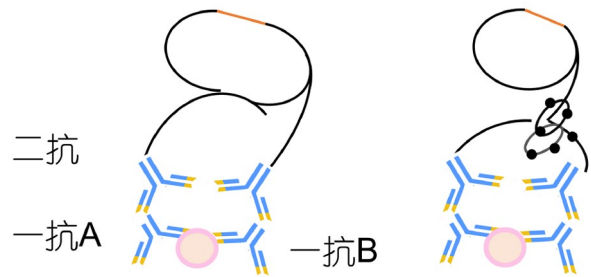
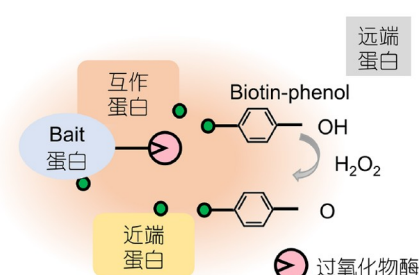


图5 邻近连接测定原理

Figure 5 Principle of proximity ligation assay

技术对bait诱饵蛋白邻近蛋白进行生物素标记, 联合质谱对其互作组进行鉴定, 突破了传统技术的诸多局限性, 可以根据蛋白质与靶标的接近程度来实现对活细胞中蛋白质的原位标记, 并且在寻找一些瞬时或弱相互作用蛋白以及一些不溶性蛋白的研究方面有独特优势。近年来极大地推动了蛋白质互作及细胞器蛋白质组学的进展^[4]。目前应用于邻近标记的酶主要可以分为两类: 以APEX, APEX2, HRP为代表的过氧化物酶和包括BirA^[50], BioID2^[51], BASU^[52], TurboID^[53], miniTurbo^[53], ultraID^[54], AirID^[55], PhastID^[56]等在内的生物素连接酶(图6)。APEX是根据Ting课题组^[57]从豌豆抗坏血酸过氧化物酶(ascorbate peroxidase, APX)开发的。经工程改造后的突变酶APEX(engineered ascorbate peroxidase)可以将许多酚衍生物氧化为苯氧基自由基, 因此可以与富含电子的氨基酸(例如Tyr, Trp, His和Cys)在小半径(<20 nm)内共价反应。当生物素酚作为APEX的底物时, 突变酶能氧化生物素苯酚, 产生高活化的苯氧基-生物素自由基并释放其进攻融合邻近蛋白表面富含电子的氨基酸残基侧链, 使其带上生



物素酚标记, 然后通过与链霉亲和素偶联进行纯化富集^[58]。过氧化物酶的标记半径在20 nm内, 一般的蛋白直径在5~10 nm, 因此除目的蛋白的直接互作蛋白外, PDB能捕获相邻蛋白复合体的其他组分, 甚至其蛋白微环境的信息^[59]。该系统目前也进一步扩展到APEX2^[60]和Split APEX2^[61], 增强了此方法的灵敏度。APEX系统标记最大的优势是标记时间可以短至1分钟, 且可用于标记核酸^[62]。但APEX的过氧化氢细胞毒性强会导致细胞膜的破坏, 因此无法应用于个体水平的邻近组研究^[63]。同时极强的标记效率允许极窄的标记时间窗口, 但也会带来较高的背景噪声和非特异标记; 同样由于标记时间过短, 难以应用于捕获一些生物学过程中的动态互作变化, 因此该系统的应用仍存在局限性。

与APEX系统不同, BioID技术利用蛋白质生物素连接酶可以通过活化生物素(biotin)形成中间产物生物素酰-5'-腺苷酸biotinyl-5'-AMP, 将生物素标记到其底物上^[64]。经过人工突变, 该连接酶失去对底物的特异性, 释放的biotinyl-5'-AMP化学性质极不稳定, 可将邻近的赖氨酸残基侧链的伯胺基团共价标记上生物素酰化修饰。BirA*的标记作用范围大约为10 nm^[50], 该距离恰好是蛋白质-蛋白质相互作用的距离范围, 因此在蛋白质组学中具有广泛的应用前景。

BioID技术首先将目的蛋白与生物素连接酶形成融合蛋白, 经过生物素标记后利用生物素-亲和素系统可以下拉和纯化周围被生物素连接酶标记的蛋白, 再通过高通量的方法进一步鉴定出这些蛋白。目前已报道了来自多种物种的生物素连接酶, 包括BioID(BirA*^[50], 来自*Escherichia coli*)、BioID2^[51](BPL*, 来自*Aquifex aeolicus*)和RapID^[52](来自*Bacillus subtilis*)等。2011年Roux等人^[50]利用大肠杆菌中的生物素连接酶BirA蛋白第118位精氨酸突变成甘氨酸后(R118G, 突变后的连接酶称为BirA*), 与biotinoyl-5' AMP亲和力和明显减弱的特性^[65,66], 建立了初代的BioID技术。BirA*生产并释放高度活化的biotinoyl-5' AMP分子, 后者扩散时与周围蛋白的赖氨酸侧链发生反应, 使该蛋白共价结合生物素, 构成了邻近标记的基础^[50]。原始的BioID使用过程中会发现, 大分子量蛋白与靶标的融合通常会影

响正常蛋白的折叠和定位, 长期标记会降低信噪比。为了降低酶的分子量, 提高靶向效率, 研究人员在嗜嗜热菌*Aquifex aeolicus*中发现了另一种新型的BPL, 这种生物素连接酶与其大肠杆菌同源物相比缺少了DNA结合结构域, 在结构上更加精简, 这种新型突变的生物素连接酶称为BioID2, 其标记强度相比于BirA*有所下降。随着对于高标记活性酶的需求进一步增加, 最近报道的BASU, AirID, TurboID, miniTurbo等新型生物素酶在活性上有明显的提高, 同时标记时间也随之进一步缩短, 从最初BirA*需要24小时, 到BASU, TurboID, UltraID等标记在分钟级别^[53,54], 并且具有不同优势。使用在枯草芽孢杆菌中发现的BASU系统, 在缩短生物素连接酶标记时间的同时, 该系统也已被证明可识别RNA-蛋白质相互作用。通过对大肠杆菌BirA*进行体外分子进化以增强其活性, 产生了TurboID和miniTurbo变体。在目前报道的生物素连接酶中TurboID的标记能力最强, 但它对内源生物素的利用能力较强, 具有一定的细胞毒性。目前BioID类技术已经在许多模式生物和非模式生物的研究中被广泛应用, 如细菌^[53,67]、酵母^[53,68,69]、黏菌^[70,71]、寄生虫^[72~76]、蠕虫^[53]、苍蝇^[53,77,78]和小鼠^[79,80]等, 也广泛应用于细胞信号通路转导关键蛋白的鉴定^[56]。

针对目前BioID技术中各种生物素连接酶的种种缺点, 本课题组^[56]近期开发出具有自主知识产权、基于嗜热古菌*Pyrococcus horikoshii*生物素连接酶的系统PhastID。PhastID所采用的新型生物素酶Ph标记活性极高(最快可以在1分钟内完成标记), 同时分子量小于TurboID和BioID2等目前最常用的系统, 具有更准确的胞内定位及标记特征, 是目前兼具快速与最高信噪比的研究蛋白质间动态相互作用的理想工具。利用PhastID技术, 本课题组对衰老相关信号通路“胰岛素-mTORC1”进行关键蛋白互作网络研究。通过胰岛素刺激15分钟前后的快速标记, 课题组筛选并鉴定了mTORC1关键激活蛋白Rheb的鸟苷酸交换因子(guanine nucleotide exchange factor, GEF)——一种溶酶体上的膜蛋白, 填补了该领域20多年的研究空白。同时, PhastID也能鉴定到与AMPK互作的膜蛋白AMPKIP1通过招募磷酸酶ALPP来负调控AMPK信号通路的功能, 进一步说明其在鉴定信号通路早期事件, 尤其是对膜蛋白的鉴定方面具有独特优势。此外, 本课题组将PhastID应用于热激前后15分钟的核凝集体邻近互作蛋白网络研究, 结合自主开发的RA算法和基于PhastID的人类已知全部核体蛋白互作网络构建的核体全景地图(未发表数据), 清晰描述核凝集体作为无膜细

胞器在蛋白组成与功能上的应激变化趋势,充分说明了PhastID在捕捉蛋白质动态互作中的强大功能. 另外,由于Ph生物素酶的分子量较小,PhastID技术也非常适用于大蛋白的原位蛋白网络捕捉. 本课题组利用同源重组敲入途径将PhastID敲入基因组特定位点成功构建与内源BRCA2的融合蛋白并分析BRCA2的蛋白互作网络,解决了大分子量蛋白研究的难题;并通过PhastID和遗传筛选的方法发现了一个新型的核膜定位蛋白NUMEN,证明其具有核酸内切酶3'到5'核酸外切酶活性,不仅决定了DNA损伤修复途径NHEJ或HR的选择问题,而且通过抑制重复序列的重组保持端粒等重复序列的稳定性,对于异染色质重复序列的修复也起到重要的作用^[81].

3 AI与蛋白质相互作用

计算生物学和机器学习的进步正在帮助预测基于大量生物数据集的PPI,可以帮助解释以前未知的蛋白相互作用. AI通过机器学习算法计算模型,在包含已知PPIs、蛋白质结构、序列和功能注释等大型数据集上训练算法,学习与PPIs相关的模式和特征,使它们能够预测蛋白质之间的潜在相互作用^[82]. 机器学习技术目前被应用于分析复杂的生物网络,如蛋白质相互作用网络,识别并分析参与细胞过程的模块、通路和关键蛋白质. 通过机器学习可以对蛋白质进行功能注释分析,通过整合包括基因组、转录组、蛋白组和通路信息,以注释蛋白质并推断其功能. 此外,机器学习方法可以通过分析蛋白质的序列、结构和进化特征来预测突变对PPIs的影响,帮助理解遗传变异如何影响疾病易感性、药物反应和其他表型特征. 此外,预测PPIs的计算方法在药物发现和开发中发挥着至关重要的作用. 通过识别蛋白质靶点和潜在的药物候选物,加速药物设计和优化的过程,为各种疾病提供了更有效的治疗方案. 总的来说,计算生物学和机器学习技术的整合极大地扩展了人们预测和理解蛋白质相互作用的能力,揭示了复杂的生物过程,并促进了生物医学研究和治疗方法的进步. 与快速增长的人类蛋白互作组的尺度相伴随的是,各类非统一实验平台和分析方法的杂糅所产生的大量批次效应和数据噪声的引入. 因而,一个具有大规模、高质量和均一化特征的PPI数据集以及标准化的数据处理方案,将会极大推动相关研究的发展.

4 未来展望

PPI未来的研究势必是分子生物学、结构生物学、数据科学的相结合. 正如前文中提及的,在过去数十年间,该领域已经积累了以百万计的PPI互作数据,这些数据构成了机器学习的基础. 然而,目前数据科学预测PPI仍以基于序列和基于结构为主,基于现有PPI数据的算法仍鲜有报道^[83]. 一方面,PPI数据集仍存在高背景、均一化差、批次效应强等缺点,这反过来需要实验方法的提升^[84];另外一方面,以AlphaFold2为代表的基于序列和结构的算法模型表现出较大潜力,可以基于序列特征与共进化信息,在结构预测中显示出较好的准确性^[85];有研究者将互作蛋白对的氨基酸序列人为加入linker序列后,使用AlphaFold2成功预测复合物结构,这也催生了AlphaFold-Multimer^[86]多链复合物的结构预测. 基于结构的PPI预测不仅能帮助判断蛋白互作是否存在,还能提供蛋白复合物的结构模型,量化其结合力^[83]. 但目前已有的各种算法暂时无法预测蛋白内在无序区域(intrinsically disordered regions, IDRs)^[87],且对缺乏同源序列作为算法参考的蛋白表现不佳,如抗体可变区-抗原互作^[88]. 同时,既往实验方法对PPI的描述,很难兼顾蛋白质翻译后修饰、蛋白结合辅助因子,以及PPI的细胞状态、生理病理情况等时空背景,这些信息特征的缺失也使得开发快速有效的算法仍存在巨大挑战.

除结构外,我们在对核凝集体的邻近标记数据分析中发现,通过比较靶蛋白邻近蛋白网络与已知核体蛋白组学数据的相似程度,可以帮助判定靶蛋白在核内的具体位置. 同时通过对靶蛋白的邻近蛋白组学数据分析,我们能够获得更多关于它的功能提示和作用机制线索,这非常有助于对这些未知蛋白的潜在功能研究. 通过对比正常蛋白和疾病相关蛋白的邻近蛋白组学数据,我们也能更加精细地解析疾病的致病机理,从而启发未来相关的个性化精准治疗方案. 此外,未来还可以期待在单细胞水平、个体水平、组织器官水平上研究蛋白质相互作用. 基于不同蛋白质相互作用技术间的比较,邻近标记技术可能是目前较为合适的技术,可以进一步拓展邻近标记技术的特异性、标记效率、拓展标记的应用范围,提高其在不同生物系统中的适用性.

参考文献

- 1 Liu X, Salokas K, Tamene F, et al. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat Commun*, 2018, 9: 1188
- 2 Willyard C. New human gene tally reignites debate. *Nature*, 2018, 558: 354–355
- 3 Bateman A, Martin M J, Orchard S, et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*, 2023, 51: D523–D531
- 4 Sharifi Tabar M, Parsania C, Chen H, et al. Illuminating the dark protein-protein interactome. *Cell Rep Methods*, 2022, 2: 100275
- 5 del Toro N, Shrivastava A, Ragueneau E, et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res*, 2022, 50: D648–D653
- 6 Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*, 2023, 51: D638–D646
- 7 Oughtred R, Stark C, Breitkreutz B J, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res*, 2019, 47: D529–D541
- 8 Struk S, Jacobs A, Sánchez Martín-Fontecha E, et al. Exploring the protein-protein interaction landscape in plants. *Plant Cell Environ*, 2019, 42: 387–409
- 9 Ramírez-Sánchez O, Pérez-Rodríguez P, Delaye L, et al. Plant proteins are smaller because they are encoded by fewer exons than animal proteins. *Genomics Proteomics Bioinf*, 2016, 14: 357–370
- 10 Rabbani G, Baig M H, Ahmad K, et al. Protein-protein interactions and their role in various diseases and their prediction techniques. *Curr Protein Pept Sci*, 2018, 19: 948–957
- 11 Gonçalves E, Poulos R C, Cai Z, et al. Pan-cancer proteomic map of 949 human cell lines. *Cancer Cell*, 2022, 40: 835–849.e8
- 12 Kim M, Park J, Bouhaddou M, et al. A protein interaction landscape of breast cancer. *Science*, 2021, 374: eabf3066
- 13 Huttlin E L, Bruckner R J, Navarrete-Perea J, et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, 2021, 184: 3022–3040.e28
- 14 Noone J, Wallace R G, Rochfort K D. Immunoprecipitation: variations, considerations, and applications. In: Loughran S T, Milne J J, eds. *Protein Chromatography. Methods in Molecular Biology*. New York: Humana, 2023. 271–303
- 15 Songyang Z, Shoelson S E, Chaudhuri M, et al. SH2 domains recognize specific phosphopeptide sequences. *Cell*, 1993, 72: 767–778
- 16 Songyang Z, Carraway Iii K L, Eck M J, et al. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature*, 1995, 373: 536–539
- 17 Rodriguez M, Li S S C, Harper J W, et al. An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J Biol Chem*, 2004, 279: 8802–8807
- 18 Songyang Z, Fanning A S, Fu C, et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*, 1997, 275: 73–77
- 19 Liu D, Safari A, O'Connor M S, et al. PTOP interacts with POT1 and regulates its localization to telomeres. *Nat Cell Biol*, 2004, 6: 673–680
- 20 Silva J V, Freitas M J, Felgueiras J, et al. The power of the yeast two-hybrid system in the identification of novel drug targets: building and modulating PPP1 interactomes. *Expert Rev Proteomics*, 2015, 12: 147–158
- 21 Ma J, Ptashne M. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell*, 1988, 55: 443–446
- 22 Luck K, Kim D K, Lambourne L, et al. A reference map of the human binary protein interactome. *Nature*, 2020, 580: 402–408
- 23 Causier B, Davies B. Analysing protein-protein interactions with the yeast two-hybrid system. *Plant Mol Biol*, 2002, 50: 855–870
- 24 Miyawaki A, Llopis J, Heim R, et al. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature*, 1997, 388: 882–887
- 25 Johnsson N, Varshavsky A. Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc Natl Acad Sci USA*, 1994, 91: 10340–10344
- 26 Galarneau A, Primeau M, Trudeau L E, et al. β -Lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein-protein interactions. *Nat Biotechnol*, 2002, 20: 619–622
- 27 Michnick S W, Remy I, Campbell-Valois F X, et al. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol*, 2000, 328: 208–230
- 28 Remy I, Wilson I A, Michnick S W. Erythropoietin receptor activation by a ligand-induced conformation change. *Science*, 1999, 283: 990–993
- 29 Michnick S W, Landry C R, Levy E D, et al. Protein-fragment complementation assays for large-scale analysis, functional dissection, and

- spatiotemporal dynamic studies of protein-protein interactions in living cells. *Cold Spring Harb Protoc*, 2016, 2016: pdb.top083543
- 30 Ghosh I, Hamilton A D, Regan L. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc*, 2000, 122: 5658–5659
 - 31 Hu C D, Chinenov Y, Kerppola T K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell*, 2002, 9: 789–798
 - 32 Chen L Y, Liu D, Songyang Z. Telomere maintenance through spatial control of telomeric proteins. *Mol Cell Biol*, 2007, 27: 5898–5909
 - 33 Liang J, Wan M, Zhang Y, et al. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol*, 2008, 10: 731–739
 - 34 Kim H, Lee O H, Xin H, et al. TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nat Struct Mol Biol*, 2009, 16: 372–379
 - 35 Wu W, Hsu Y M S, Bi L, et al. CARD9 facilitates microbe-elicited production of reactive oxygen species by regulating the LyGDI-Rac1 complex. *Nat Immunol*, 2009, 10: 1208–1214
 - 36 Zhang Y, Wu Y, Mao P, et al. Cold-inducible RNA-binding protein CIRP/hnRNP A18 regulates telomerase activity in a temperature-dependent manner. *Nucleic Acids Res*, 2016, 44: 761–775
 - 37 Li F, Kim H, Ji Z, et al. The BUB3-BUB1 complex promotes telomere DNA replication. *Mol Cell*, 2018, 70: 395–407.e4
 - 38 Liu Y, Liu F, Cao Y, et al. Shwachman-diamond syndrome protein SBDS maintains human telomeres by regulating telomerase recruitment. *Cell Rep*, 2018, 22: 1849–1860
 - 39 Shi G, Bai Y, Zhang X, et al. Bend family proteins mark chromatin boundaries and synergistically promote early germ cell differentiation. *Protein Cell*, 2022, 13: 721–741
 - 40 Xin H, Liu D, Wan M, et al. TPP1 is a homologue of ciliate TEBP- β and interacts with POT1 to recruit telomerase. *Nature*, 2007, 445: 559–562
 - 41 Han X, Liu D, Zhang Y, et al. Akt regulates TPP1 homodimerization and telomere protection. *Aging Cell*, 2013, 12: 1091–1099
 - 42 Zhang Y, Chen L Y, Han X, et al. Phosphorylation of TPP1 regulates cell cycle-dependent telomerase recruitment. *Proc Natl Acad Sci USA*, 2013, 110: 5457–5462
 - 43 He Q, Kim H, Huang R, et al. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. *Cell Stem Cell*, 2015, 17: 273–286
 - 44 Yang D, Xiong Y, Kim H, et al. Human telomeric proteins occupy selective interstitial sites. *Cell Res*, 2011, 21: 1013–1027
 - 45 Miller K E, Kim Y, Huh W K, et al. Bimolecular fluorescence complementation (BiFC) analysis: advances and recent applications for genome-wide interaction studies. *J Mol Biol*, 2015, 427: 2039–2055
 - 46 Kerppola T K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu Rev Biophys*, 2008, 37: 465–487
 - 47 Remy I, Michnick S W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat Methods*, 2006, 3: 977–979
 - 48 Fredriksson S, Gullberg M, Jarvius J, et al. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*, 2002, 20: 473–477
 - 49 Gullberg M, Gústafsdóttir S M, Schallmeiner E, et al. Cytokine detection by antibody-based proximity ligation. *Proc Natl Acad Sci USA*, 2004, 101: 8420–8424
 - 50 Roux K J, Kim D I, Raida M, et al. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*, 2012, 196: 801–810
 - 51 Kim D I, Jensen S C, Noble K A, et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell*, 2016, 27: 1188–1196
 - 52 Ramanathan M, Majzoub K, Rao D S, et al. RNA-protein interaction detection in living cells. *Nat Methods*, 2018, 15: 394
 - 53 Branon T C, Bosch J A, Sanchez A D, et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol*, 2018, 36: 880–887
 - 54 Kubitz L, Bitsch S, Zhao X, et al. Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells. *Commun Biol*, 2022, 5: 657
 - 55 Kido K, Yamanaka S, Nakano S, et al. AirID, a novel proximity biotinylation enzyme, for analysis of protein-protein interactions. *eLife*, 2020, 9: e54983
 - 56 Feng R, Liu F, Li R, et al. The rapid proximity labeling system PhastID identifies ATP6AP1 as an unconventional GEF for Rheb. *Cell Res*, 2024, 34: 355–369

- 57 Martell J, Deerinck T, Sancak Y, et al. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol*, 2012, 30: 1143–1148
- 58 Rhee H W, Zou P, Udeshi N D, et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, 2013, 339: 1328–1331
- 59 Gingras A C, Abe K T, Raught B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr Opin Chem Biol*, 2019, 48: 44–54
- 60 Hung V, Udeshi N D, Lam S S, et al. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc*, 2016, 11: 456–475
- 61 Han Y, Branon T C, Martell J D, et al. Directed evolution of split APEX2 peroxidase. *ACS Chem Biol*, 2019, 14: 619–635
- 62 Wang K, Wang L N, Xu X W, et al. Recent advances in RNA proximity labeling (in Chinese). *Sci Sin Vitae*, 2021, 51: 1657–1667 [王凯, 王丽娜, 许晓亮. RNA邻近标记技术研究进展. 中国科学: 生命科学, 2021, 1657–1667]
- 63 Geoghegan V, Mottram J C, Jones N G. Tag Thy Neighbour: nanometre-scale insights into kinetoplastid parasites with proximity dependent biotinylation. *Front Cell Infect Microbiol*, 2022, 12: 894213
- 64 Prakash O, Eisenberg M A. Biotinyl 5'-adenylate: corepressor role in the regulation of the biotin genes of *Escherichia coli* K-12. *Proc Natl Acad Sci USA*, 1979, 76: 5592–5595
- 65 Kwon K, Beckett D. Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci*, 2000, 9: 1530–1539
- 66 Kwon K, Streaker E D, Ruparella S, et al. Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor.. *J Mol Biol*, 2000, 304: 821–833
- 67 Santin Y G, Doan T, Lebrun R, et al. *In vivo* TssA proximity labelling during type VI secretion biogenesis reveals TagA as a protein that stops and holds the sheath. *Nat Microbiol*, 2018, 3: 1304–1313
- 68 Opitz N, Schmitt K, Hofer-Pretz V, et al. Capturing the Asc1p/receptor for activated C kinase 1 (RACK1) microenvironment at the head region of the 40S ribosome with quantitative BioID in yeast. *Mol Cell Proteomics*, 2017, 16: 2199–2218
- 69 Larochelle M, Bergeron D, Arcand B, et al. Proximity-dependent biotinylation mediated by TurboID to identify protein-protein interaction networks in yeast. *J Cell Sci*, 2019, 132: jcs232249
- 70 Batsios P, Meyer I, Gräf R. Proximity-dependent biotin identification (BioID) in amoebae. *Method Enzymol*, 2016, 569: 23–42
- 71 Pitzen V, Askarzada S, Gräf R, et al. CDK5RAP2 is an essential scaffolding protein of the corona of the dictyostelium centrosome. *Cells*, 2018, 7: 32
- 72 Morriswood B, Havlicek K, Demmel L, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryot Cell*, 2013, 12: 356–367
- 73 Chen A L, Kim E W, Toh J Y, et al. Novel components of the inner membrane complex revealed by BioID. *mBio*, 2015, 6: e02357-14
- 74 McAllaster M R, Ikeda K N, Lozano-Núñez A, et al. Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in *Trypanosoma brucei*. *Mol Biol Cell*, 2015, 26:3013–3029
- 75 Khosh-Naucke M, Becker J, Mesén-Ramírez P, et al. Identification of novel parasitophorous vacuole proteins in *P. falciparum* parasites using BioID. *Int J Med Microbiol*, 2018, 308: 13–24
- 76 Tu V, Tomita T, Sugi T, et al. The *Toxoplasma gondii* cyst wall interactome. *mBio*, 2020, 11: e02699-19
- 77 Shinoda N, Hanawa N, Chihara T, et al. Drone-independent basal executioner caspase activity sustains *Drosophila* imaginal tissue growth. *Proc Natl Acad Sci USA*, 2019, 116: 20539–20544
- 78 Carnesecchi J, Sigismondo G, Domsch K, et al. Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity. *Nat Commun*, 2020, 11: 1388
- 79 Uezu A, Kanak D J, Bradshaw T W A, et al. Identification of an elaborate complex mediating postsynaptic inhibition. *Science*, 2016, 353: 1123–1129
- 80 Dingar D, Kalkat M, Chan P K, et al. BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors. *J Proteomics*, 2015, 118: 95–111
- 81 Chen B, Ge T, Jian M, et al. Transmembrane nuclease NUMEN/ENDOD1 regulates DNA repair pathway choice at the nuclear periphery. *Nat Cell Biol*, 2023, 25: 1384
- 82 Zhou Z, Yan C, Zhang C Y. Artificial intelligence biology—biology V3.0 (in Chinese). *Sci Sin Vitae*, 2022, 52: 291–300 [周祯, 闫超, 张辰宇. 人

人工智能生物学——生物学3.0. 中国科学: 生命科学, 2022, 52: 291–300]

- 83 Grassmann G, Miotto M, Desantis F, et al. Computational approaches to predict protein-protein interactions in crowded cellular environments. *Chem Rev*, 2024, 124: 3932–3977
- 84 Baryshev A, La Fleur A, Groves B, et al. Massively parallel protein-protein interaction measurement by sequencing (MP3-seq) enables rapid screening of protein heterodimers. *bioRxiv*, 2023, doi: 10.1101/2023.02.08.527770
- 85 Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 2021, 596: 583–589
- 86 Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2022, doi: 10.1101/2021.10.04.463034
- 87 Ruff K M, Pappu R V. AlphaFold and implications for intrinsically disordered proteins. *J Mol Biol*, 2021, 433: 167208
- 88 Yin R, Feng B Y, Varshney A, et al. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci*, 2022, 31: e4379

Protein-protein interaction techniques—a powerful tool for functional genomics

FENG Ran^{1,2}, LIU Feng¹, WU Su², LI RuoFei¹ & SONGYANG Zhou^{1,2}

1 School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China;

2 Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

Cellular activities are complex and robust processes that rely on dynamic networks formed by protein-protein interactions (PPIs) for intra- and intercellular signaling and communication. The study of PPIs is, therefore, crucial for understanding protein functions and cell signaling. For the last 30 years, our team has been devoted to developing and employing various PPI techniques to systematically dissect and explore the components and mechanisms of crucial protein networks in human cells. Our primary research focuses on crucial intracellular signaling pathways, including regulatory mechanisms of telomere homeostasis, genome damage repair, and stem cell self-renewal. This review begins by elucidating commonly used PPI tools and their characteristics. Subsequently, it exemplifies how these techniques have propelled advances in biological research using key research findings from our group as an example. Finally, it outlines the potential role of artificial intelligence in advancing this field and future directions in PPI research. This comprehensive review aims to demonstrate how the development of PPI technologies promotes functional genomics research and underscores the significance of technological innovation in biology.

protein-protein interaction, proximity labeling, protein-fragment complementation assay

doi: 10.1360/SSV-2024-0151