

# RNA-蛋白质复合物规模化富集 与鉴定新方法

董沛滢<sup>1</sup>, 刘彤<sup>2</sup>, 秦伟捷<sup>1,2</sup>

(1. 安徽医科大学基础医学院, 合肥 230032;

2. 医学蛋白质组全国重点实验室, 北京蛋白质组研究中心, 国家蛋白质科学中心(北京),  
军事科学院军事医学研究院, 北京 102206)

**摘要** RNA-蛋白质复合物调控着生物体内各种生理过程, 在RNA从合成到降解的生理活动历程中发挥关键作用. 为了更全面解析RNA-蛋白质复合物(RNA-protein complexes, RPC)相互作用网络, 近年来研究者除了开发RNA直接结合蛋白(RNA-binding proteins, RBP)规模化富集方法, 也将目光转向了与RBP存在相互作用的RNA间接结合蛋白(RNA-associated proteins, RAP), 二者与RNA共同组成了结构复杂、功能各异的RPC. 但以往方法均将RPC中包含的所有蛋白视为整体, 统一鉴定, 缺乏能够规模化区分RBP和RAP的有效途径. 基于此, 本文联合使用254 nm紫外交联法和二硫双琥珀酰亚胺丙酸酯(DSP)蛋白交联剂, 通过不同反应原理将RBP与RAP分别富集于链霉素磁珠, 进而通过分步洗脱方法分别获取纯化的RBP和RAP, 并结合质谱分析实现二者的各自独立鉴定. 采用该方法在HeLa细胞中共鉴定到2007个高置信RBP, 927个高置信RAP, 其中有243个RAP为首次鉴定. 此方法的建立有助于细化RPC中各组元间的相互作用网络, 为后续的生物应用提供方法和数据基础.

**关键词** RNA结合蛋白; 蛋白质组学; 富集; 分离; 质谱

中图分类号 O657 文献标志码 A doi: 10.7503/cjcu20240091

## A New Method for Large-scale Enrichment and Stepwise Identification of RNA-protein Complexes

DONG Peiying<sup>1</sup>, LIU Tong<sup>2\*</sup>, QIN Weijie<sup>1,2\*</sup>

(1. School of Basic Medical Science, Anhui Medical University, Hefei 230032, China;

2. State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center,  
National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China)

**Abstract** RNA-protein complexes (RPC) regulate various physiological processes in organisms and play a crucial role in the synthesis and degradation of RNA. For deep RPC profiling in addition to develop RNA-binding proteins (RBP) enrichment methods, some researchers also turned their attention to RNA-associated proteins (RAP), which interact with RBP *via* protein-protein interaction. However, previous methods often enrich all proteins in the RPC as a whole and cannot differentiate RBP from RAP. Therefore, we reported a new method using combination of 254 nm UV crosslinking of RBP and RNA and DSP crosslinking of RBP and RAP. Next, RBP and RAP can be eluted in a stepwise way and separately identified by mass spectrometry. A total of 2007 high confidence RBP and 927 high confidence RAP were identified in HeLa cells, of which 243 RAP were identified for the first time. This method aims

收稿日期: 2024-02-23. 网络首发日期: 2024-03-16.

联系人简介: 秦伟捷, 男, 博士, 研究员, 主要从事蛋白质分离分析新技术新方法研究. E-mail: aunp\_dna@126.com

刘彤, 女, 博士, 主要从事蛋白质分离分析新技术研究. E-mail: liutong19920918@163.com

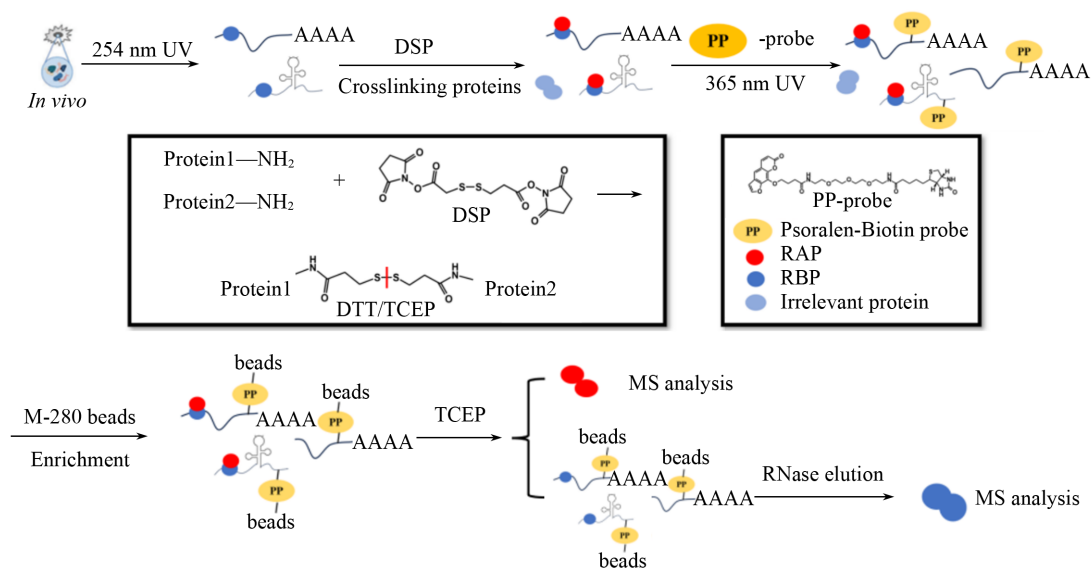
基金项目: 国家重点研发计划项目(批准号: 2021YFA1302604)资助.

Supported by the National Key Research and Development Program of China(No.2021YFA1302604).

to refine the RNA-protein interaction network and provide a favorable foundation for subsequent biological investigations.

**Keywords** RNA binding protein; Proteomics; Enrichment; Separation; Mass spectrometry

RNA(脱氧核糖核酸)在真核细胞中以RNA-蛋白质复合物(RNA-protein complexes, RPC)的形式存在,且主要以该形式发挥功能<sup>[1-5]</sup>.在组成RPC的蛋白质中,直接与RNA作用的蛋白被称为RNA直接结合蛋白(RNA-binding proteins, RBP),而通过与RBP结合以蛋白-蛋白相互作用参与进RPC中的蛋白被称为RNA相关蛋白(RNA-associated proteins, RAP).其中,RPC深入参与到RNA生命周期的各个阶段<sup>[6,7]</sup>,目前已有大量研究表明RPC组成和含量的变化不仅会影响细胞稳态<sup>[8]</sup>,还会参与如癌症<sup>[9,10]</sup>、抗病毒反应<sup>[11]</sup>、神经和肌肉退行性疾病<sup>[12,13]</sup>等多种疾病的发生进程.因此,在不同生理和病理条件下规模化地解析RPC的蛋白质组成近年来受到广泛关注.随着蛋白质组学的发展,人们对RPC的研究内容是逐级探索、逐渐扩大的.首先,与高通量测序技术一同发展面世的是基于poly(A)序列和互补链的mRNA及其结合蛋白的富集方法<sup>[14,15]</sup>,这种方法将RBP的鉴定规模从单一蛋白扩展到规模化鉴定,但因其富集把手依赖RNA的poly(A)尾结构,因此不能完成对全种类RNA的RBP富集.为进一步扩充RNA结合蛋白种类,研究人员随后开发了基于炔基尿苷类似物代谢标记串联点击化学反应的富集方法<sup>[16-18]</sup>、基于RNA、蛋白质和DNA的不同密度进行分离的相分离富集方法<sup>[19-21]</sup>以及基于硅亲和的固相萃取法<sup>[22,23]</sup>.以上3种方法不受RNA种类的限制,可以无偏性富集与RNA直接结合的RBP.但是代谢标记法因其代谢核苷酸类似物标记周期长、实验成本高;相分离法因其分离蛋白质和RNA时边界不清晰等问题依然有待改进.此外,随着蛋白相互作用研究的不断深入,研究人员发现RNA等遗传物质并非蛋白质生理活动唯一的调控者,蛋白-蛋白相互作用在细胞生理活动及疾病发展中也起着重要作用<sup>[24]</sup>.同时有研究<sup>[25]</sup>表明,RPC的形成是通过多价RNA-RNA、RNA-蛋白质以及蛋白-蛋白相互作用实现的.因此,除了与RNA直接作用的RBP外,与RBP形成蛋白-蛋白作用的RAP同样有鉴定和研究意义.为此,研究人员进一步拓展了RPC的鉴定边界,建立了基于密度梯度离心的R-Deep法<sup>[26]</sup>和基于免疫纯化法富集RPC的方法<sup>[27]</sup>,这两种方法均将在RPC中发生相互作用的蛋白笼统地看作一个整体,无区分地富集与RNA存在直接相互作用或者是通过蛋白相互作用参与的RPC,缺乏对RBP和RAP鉴定的准确性和针对性.针对上述研究需要和技术困境,本文建立了一个两步交联,分步释放的RPC规模化富集鉴定新方法.如Scheme 1所示,首先采用254 nm紫外交联,固定RNA与直接结合蛋白质间的相互作用;进而引入蛋白-蛋白交联剂二硫双琥珀酰亚胺丙酸酯(DSP),利用其N-羟基丁二酰亚胺(NHS)端与蛋白N端和侧链氨基的反应形成的稳定共价键,固定与RBP发生相互作用的RAP.然后,利用补



**Scheme 1** Schematic overview of the RPC enrichment for large-scale RBP and RAP identification by MS

骨脂素-生物素探针(PP探针<sup>[28]</sup>)的补骨脂素端与嘧啶碱基在365 nm紫外光照激活下发生的共价偶联,以及生物素端与链霉亲和素磁珠的特异性吸附,从细胞中富集出完整的RPC. 在洗脱阶段先加入TCEP还原剂破坏DSP的二硫键,从而分离出RAP;而后加入RNase酶断裂RNA,释放出RBP,达到分步分离RBP和RAP的目的. 最后,分别对二者进行基于质谱的蛋白质组分析,获得RBP和RAP的定性定量信息.

## 1 实验部分

### 1.1 试剂与仪器

尿素(UA, 纯度99%)、三(2-羧乙基)膦溶液(TCEP, 0.5 mol/L)、2-氯乙酰胺(CAA, 纯度98%)和三乙基碳酸氢铵(TEAB, 1 mol/L, pH=8.5), Sigma-Aldrich公司; 乙腈(ACN, 纯度99%), 质谱纯, Merck公司; 二硫双琥珀酰亚胺丙酸酯(DSP, 纯度99%)、甲醛(CH<sub>2</sub>O, 10%~15%)、氘代甲醛(CD<sub>2</sub>O, 20%)、氰基硼氢化钠(NaBH<sub>3</sub>CN, 纯度99%)、氨水(NH<sub>3</sub>·H<sub>2</sub>O, 99%)、三氟乙酸(TFA, 1 mol/L)、TritonX-100、RNase A (10 mg/mL)、RNase I (10 mg/mL)、链霉亲和素磁珠(M-280 beads)和补骨脂素-生物素探针(EZ-LINK Psoralen-PEG3-Biotin), 美国Thermo公司; RNase抑制剂(Ribonucleoside Vanadyl Complex), 美国NEB公司; 无EDTA蛋白酶抑制剂, Roche公司; 银染试剂盒, CWBIO公司; DEPC处理水, 中国生工; DMEM培养基、磷酸盐缓冲溶液(PBS)和胎牛血清(FBS), Gibico公司; 双抗(青霉素和链霉亲和素), HyClone公司; 胰蛋白酶(Trypsin), 测序级, 美国Promega公司. HeLa细胞购于ATCC公司; TXL-10枪头(Pipet tips), 美国Axygen公司; C<sub>18</sub>膜(C<sub>18</sub>-empore), 德国3M公司; 10 kDa超滤管, 美国Millipore公司; 细胞培养皿和细胞刮, Corning公司.

Sartorius BP211d型分析天平, 德国Sartorius公司; HERACELL VIOS 160i型细胞培养箱、Easy nLC1000纳升级液相色谱系统、Orbitrap Fusion™Tribid™型质谱仪、金属震荡仪和低温离心机, 美国Thermo Fisher Scientific公司; DTY-1310型超净工作台, 北京德天佑科技; 垂直电泳仪和电泳槽和Gel Doc EZ系统, Bio-Rad公司; CL-1000型254 nm紫外交联仪, 美国UVP公司; 365 nm紫外交联仪, 深圳市兰普里克科技有限公司; Vortex QL-901型涡旋混合器, 海门市其林贝尔仪器制造有限公司.

### 1.2 实验方法

1.2.1 HeLa细胞的培养 参照文献[29]方法, 在添加有10%FBS及1%双抗(100 U/mL青霉素和100 mg/mL链霉素)的DMEM培养基中培养HeLa细胞, 培养环境为37 °C, 含有5%CO<sub>2</sub>的培养箱.

1.2.2 RPC的分步偶联 当15 cm培养皿中长满培养的HeLa细胞后, 使用5 mL冷PBS缓冲液清洗3次, 将皿内液体全部吸干, 放入254 nm紫外交联仪中(光照能量为0.25 J/cm<sup>2</sup>)于冰上交联. 然后将新配制的100×DSP(40.44 mg/mL)用PBS缓冲液稀释为1×DSP, 每盘细胞使用3mL 1×DSP在室温下孵育30 min. 孵育完成后, 每盘细胞加入2 mL冷PBS缓冲液并用细胞刮收集, 将收集到的细胞放入2 mL离心管中, 以1000g转速离心5 min取沉淀. 此后均使用DEPC处理过的水配制溶液.

1.2.3 PP探针标记富集RPC和分步洗脱 获取细胞沉淀后, 先加入250 μL裂解液1(1×PBS, pH=7.4+终浓度0.5%的SDS+RNA酶抑制剂+无EDTA的蛋白酶抑制剂). 用1 mL注射器抽析使其形成匀浆, 在4 °C下温和旋转孵育20 min. 继续加入1 mL裂解液2(1×PBS, pH=7.4+终浓度为0.2%的Triton+RNA酶抑制剂+无EDTA的蛋白酶抑制剂). 用1 mL注射器抽析使其形成匀浆, 在4 °C下温和旋转孵育20 min后, 以14000g转速离心10 min, 离心后取上层清液加入6孔板. 加入终浓度为5 μmol/L的PP探针在4 °C摇床上避光温和孵育30 min, 随后使用365 nm紫外交联仪于冰上交联3 min进行PP探针对RNA的标记. 标记完成后, 以14000g转速离心, 将裂解液分批转移到10 kDa的超滤管中. 并在转移后用1×PBS缓冲液清洗3遍以除去过量的PP探针. 使用2 mol/L UA(1×PBS缓冲液配制)将超滤管中液体调整至500 μL, 并转移到1.5 mL RNase-free离心管中, 加入预处理的50 μL链霉亲和素磁珠在4 °C下温和旋转孵育1 h. 富集完成后, 将对照组磁珠与实验组磁珠分别进行处理. 首先将对照组磁珠与清液使用磁力架磁性分离, 弃去上层清液后, 使用200 μL 1×PBS缓冲液清洗2次后加入200 μL提前配制的

RNase(0.01  $\mu\text{g}/\mu\text{L}$ , 用纯水配制), 于37 °C金属浴振荡1 h后弃去上层清液. 将已清洗1次的对照组磁珠依次用0.2%SDS, 2 mol/L UA, 6 mol/L UA和50 mmol/L TEAB各清洗2次. 实验组磁珠依次用1×PBS, 0.2%SDS, 2 mol/L UA, 6 mol/L UA和50 mmol/L TEAB各清洗2次, 除去磁珠非特异性吸附(上述试剂皆采用无RNase的PBS溶液配制). 清洗完成后, 在实验组和对照组中同步加入25  $\mu\text{L}$  100 mmol/L TCEP(pH=8.5), 于37 °C金属浴振荡1 h, 使RPC中的RAP率先分离, 并收集其所在的上层清液; 随后向磁珠中加入提前配制的25  $\mu\text{L}$  RNase(0.01  $\mu\text{g}/\mu\text{L}$ , 用50 mmol/L TEAB配制)继续于37 °C金属浴振荡1 h洗脱RBP(洗脱步骤对照组和实验组操作完全一致).

1.2.4 RPC的蛋白质组学分析 按照1.2.3节步骤得到实验组RAP和实验组RBP. 对于对照组, 前期细胞交联、裂解、标记、富集步骤与实验组完全一致, 不同的是对照组富集后先使用0.1  $\mu\text{g}/\mu\text{L}$ 的RNase A溶液于37 °C孵育1 h, 目的是断裂RNA链以洗脱全部RNA相关蛋白, 方便后期与实验组进行差异筛选. 在与实验组同步完成SDS和UA清洗后, 先加入20  $\mu\text{L}$  100 mmol/L TCEP溶液取上层清液后作为RAP的对照; 再加入20  $\mu\text{L}$  0.01  $\mu\text{g}/\mu\text{L}$  RNase A溶液重悬磁珠, 作为RBP的对照. 向实验组与对照组中分别加入80  $\mu\text{L}$  50 mmol/L TEAB溶液至100  $\mu\text{L}$ 体系后, 进行蛋白质组样品的制备, 具体步骤如下: 加入终浓度为40 mmol/L的CAA和终浓度为20 mmol/L的TCEP, 于37 °C金属浴振荡中进行还原烷基化反应1.5 h后, 按照1:100的质量比加入适量的胰蛋白酶溶液, 于37 °C酶切12 h后, 再加入相同质量的胰蛋白酶, 于37 °C继续酶切4 h, 随后将磁珠与清液磁性分离后收集上层清液, 并用50  $\mu\text{L}$  0.2%TFA溶液清洗磁珠2次, 合并清洗液与上层清液(共200  $\mu\text{L}$ )加入到1.5 mL蛋白低吸离心管中, 用真空旋转干燥仪于45 °C热干即得最终肽段. 随后, 对肽段进行稳定同位素二甲基化标记: 首先将肽段分别重溶于200  $\mu\text{L}$  100 mmol/L TEAB溶液中, 实验组RBP和RAP中加入8  $\mu\text{L}$  4%CH<sub>2</sub>O溶液, 对照组RAP和RBP则加入8  $\mu\text{L}$  4%CD<sub>2</sub>O溶液, 同时每个样品中均加入8  $\mu\text{L}$  0.6 mol/L NaBH<sub>3</sub>CN溶液, 涡旋混匀后于室温反应1 h. 反应完成后, 每个样品加入4  $\mu\text{L}$  30%氨水溶液终止反应. 最后, 每个样品中加入4  $\mu\text{L}$  10%TFA溶液酸化样本后, 将对应的轻、重同位素标记的实验组和对照组混合[RBP对照组(重标)+实验组(轻标)/RAP对照组(重标)+实验组(轻标)]. 再使用Stage Tip C<sub>18</sub>脱盐并使用真空旋转干燥仪浓缩热干, 于-80 °C保存, 准备质谱分析.

1.2.5 RPC的LC-MS/MS鉴定分析 将得到的最终肽段重溶于0.1%TFA溶液中, 以14000g转速离心10 min后取上层清液, 加载于2 cm自装预柱(内径100  $\mu\text{m}$ , 并装有内径为3  $\mu\text{m}$ 的C<sub>18</sub>填料)中, 同时使用内径150  $\mu\text{m}$ , 长度15 cm的自制反相分析柱(填有粒径1.9  $\mu\text{m}$ 的Ultimate XB-C<sub>18</sub>填料)进行分离. 使用Easy-nLC1000纳升级液相色谱系统, 液相梯度为78 min(流动相A为0.1%FA水溶液, 流动相B为含有0.1%FA的ACN溶液, 流速600 nL/min), 洗脱梯度: 0~8 min, 5%~8%B; 8~58 min, 8%~22%B; 58~70 min, 22~32%B; 70~71 min, 32%~90%B; 71~78 min, 90%B. 洗脱后的肽段通过电离子喷雾离子源依次进入Orbitrap Fusion™ Tribid™质谱仪中进行分析. 喷雾电压为2 kV, 数据采集模式为数据依赖模式(Data-Dependent Acquisition, DDA), 扫描方式为正离子扫描模式, 离子传输管温度设定为320 °C, 一级全扫描检测范围 $m/z$  300~1400, 扫描分辨率120000, 最大注入时间100 ms, 自动增益控制(AGC)设置为 $5 \times 10^5$ , 采用高能碰撞诱导解离模式(HCD)将母离子碎裂成二级碎片, 碰撞能量为32%, 使用线性离子阱快速模式进行二级数据采集, AGC设置为5000, 最大注入时间设为35 ms, 动态排除时间设为18 s.

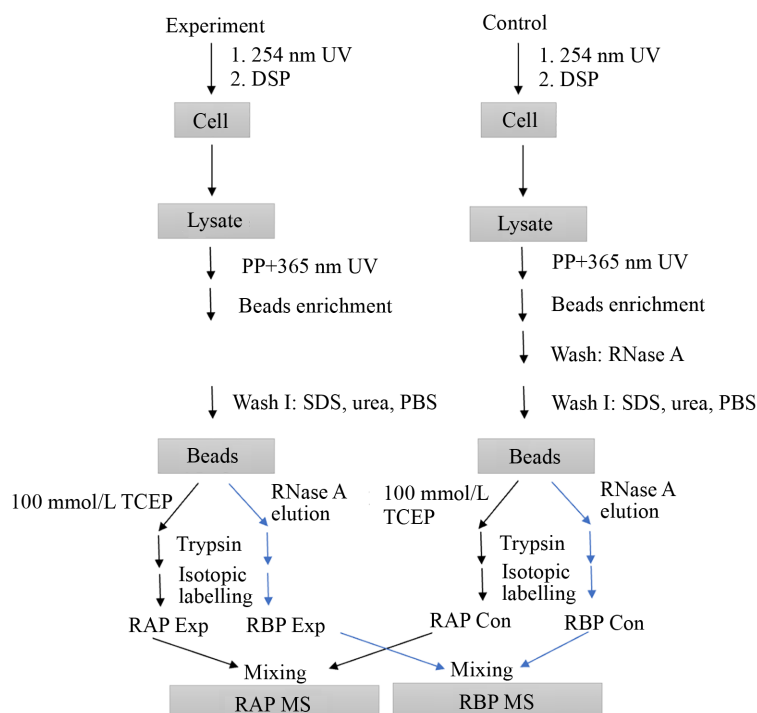
1.2.6 质谱数据检索与分析 对通过质谱得到的原始数据使用MaxQuant软件(Version1.16.7.0)进一步分析, 使用Uniprot Human(Release on 2022, 20198entry)数据库进行检索. 选择酶解模式为胰蛋白酶, 最大允许漏切位点设置为2, 每个肽段最少含有的氨基酸个数为6, 半胱氨酸脲甲基化(Cysteine carbamidomethylation)为固定修饰, 甲硫氨酸氧化(Methionine oxidation)及N-末端乙酰化(N-terminal acylation)设为可变修饰. 在二甲基化标记方面, 标记个数勾选为2, 轻标勾选DimethLys0和DimethNter0; 重标勾选DimethLys4和DimethNter4, 母离子及二级碎片离子的最大质量容差为 $\delta$  20和0.5 Da, 蛋白水平与谱图的假阳性率(False Discovery Rate, FDR) $\leq$ 1%. 按照正态分布原则对缺失值进

行填充, 再通过Persus软件计算每个蛋白的富集倍数(Fold change, 实验组与对照组信号强度的比值)及用 $P$ 值表示T-test检验计算富集显著的程度, $P$ 值越小说明实验组与对照组差异越显著. 获得的数据通过如下原则筛选被认为是高置信的RPC:(1) 3次实验中至少有2次鉴定到的蛋白;(2) 蛋白最少包含2条特异肽段(Unique peptide);(3) 实验组/对照组富集倍数不低于2倍(Fold change)且 $P < 0.01$ .

1.2.7 鉴定RNA结合蛋白的功能分析 通过在线软件DAVID(<https://david.ncifcrf.gov/tools.gvp>)使用UNIPROT\_ACCESSION数据库对鉴定到的RPC进行GO分析, 通过Reactome进行通路(Pathway)分析. 使用Origin软件对数据进行火山图绘制. 通过在线网站BioLadder(<https://www.bioladder.cn/>)绘制相关性分析图.

## 2 结果与讨论

在254 nm紫外偶联RNA和RBP的基础上, 引入蛋白-蛋白交联剂DSP偶联RBP和RAP, 富集RPC后使用TCEP和RNase酶先后断裂DSP的二硫键和RNA链以达到分步分离RAP和RBP的目的. 分离完成后, 使用基于质谱的蛋白质组学技术对富集出的RAP和RBP进行定性、定量分析和生物信息分析, 为后续的应用提供基础. 为除去实验过程中不可避免的非特异性吸附蛋白造成RBP和RAP鉴定的假阳性影响, 通过对实验组和对照组分别进行二甲基化同位素标记后, 利用相对定量法筛选RPC蛋白. 实验完整流程如Scheme 2所示, 对照组和实验组同步完成富集流程后, 先使用RNase A清洗对照组, 洗去RNA及与其结合的RPC. 对于随后的清洗(Wash I)、分步洗脱、酶切和二甲基化标记流程, 实验组和对照组都保持一致. 在二甲基化标记时,  $CD_2O$ 重标对照组包含的是无法通过清洗除去非特异吸附蛋白;  $CH_3O$ 轻标实验组包含的是无法通过清洗除去非特异吸附蛋白和RPC. 获取质谱数据后, 在数据处理阶段, 可以通过相对定量法差异筛选排除非特异吸附蛋白干扰, 获得高置信的RPC蛋白.

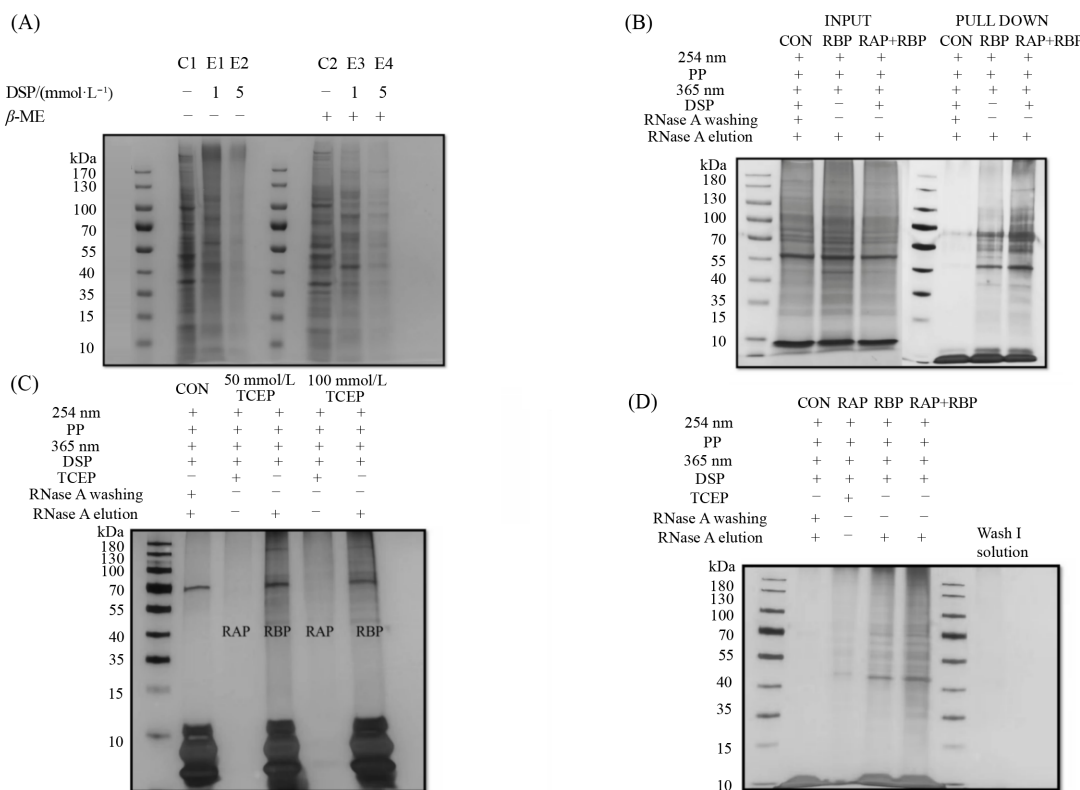


Scheme 2 Experimental design of quantitative proteomic comparison between the experiment group and control group for RPC identification

### 2.1 富集RBP与RAP的SDS-PAGE银染表征

DSP作为常用的蛋白交联剂, 在以往研究中被广泛应用于细胞<sup>[30]</sup>和组织内<sup>[31]</sup>的蛋白交联, 但应用于RAP富集尚属首次. 因此本文在建立方法时, 在此前使用的DSP方法基础上进一步来验证了DSP在

活细胞内交联蛋白的能力及最适浓度. 如图1(A)所示, 在活细胞内加入DSP后, E1中蛋白条带相对于未加入DSP的C1有明显上移现象, 说明蛋白发生交联形成了分子量更大的复合物. 加大DSP浓度后E2中的蛋白条带分布虽然相较E1无明显变化, 但强度明显降低, 这是由于DSP造成大规模蛋白交联沉淀, 导致样品损失. 在加入 $\beta$ -巯基乙醇( $\beta$ -ME)还原剂断裂DSP的二硫键后, E3和E4中的交联蛋白重新分散, 导致上移的蛋白条带恢复为交联前的初始状态. 因此可以推断得出结论, DSP具有交联活细胞内蛋白的能力, 而且其二硫键可以被还原剂断裂重新释放交联的蛋白. 在活细胞中使用DSP交联的最适浓度为1 mmol/L. 随后在DSP交联后的细胞内对RPC进行了富集实验, 结果如图1(B)所示, 交联后的RPC条带明显深于对照组未加DSP交联的RBP条带, 验证了DSP能交联RBP和潜在的RAP的可能性. 为高效释放所富集的RAP, 进一步考察了合适的还原剂浓度. 结果如图1(C)所示, 利用经254 nm紫外光照与DSP交联以及PP探针和链霉素磁珠富集出的RPC, 可以通过先后加入TCEP断裂DSP的二硫键和RNase酶水解RNA达到分步分离的目的, 从而分别富集出RBP和RAP用于独立鉴定分析. 并且通过对比图1(C)中2个RAP条带发现, 100 mmol/L TCEP洗脱下的RAP蛋白条带更深, 因此适合TCEP发挥还原作用的条件为100 mmol/L, pH=8.5. 在条件优化完成后, 根据Scheme 2所示实验流程对结果的特异性进一步验证. 富集蛋白条带如图1(D)所示, 对照组条带基本不可见, 说明清洗有效, 残留的非特异性吸附蛋白较少, 可以通过定量差异筛选进一步去除, 从而确保所得RPC蛋白的可靠性; RAP和RBP条带都清晰可见, 且RAP+RBP条带明显深于RBP. 上述实验结果验证了本文提出的两步交联, 分步分离RAP和RBP方法的可行性.



**Fig. 1** Investigation on the optimal concentration of DSP for crosslinking intracellular proteins(A), enrichment of RBP and RPC using 254 nm UV, and 254 nm UV+DSP(B), investigation on the optimal concentration of TCEP for separating RAP and RBP(C) and stepwise separation of RAP and RBP(D)

## 2.2 RPC的蛋白质组鉴定及功能分析

为了降低实验过程中的随机误差, 每次实验设置3次生物学重复. 为了考察3次重复实验是否具有较大差异, 将每次实验3组生物学重复中所得到的RBP或RAP的强度值(Intensity)分别进行两两比

较并绘制相应的散点图[图2(A)和(B)]. 可见, 3组实验中获得定量结果均具有较好的相关性, 计算得到的Pearson系数均在0.85以上, 说明同次实验的3组生物学重复所得结果具有相似的富集强度值, 证明该富集方法具有较好的重现性, 得到的RBP和RAP可信度较高. 此外, 以蛋白质的 $\log_2$  fold change为横轴、 $-\lg P$ 为纵轴绘制了火山图[图2(C)和(D)]以满足1.2.6节所述的三重卡值标准进行差异筛选: 对照组与实验组无明显差异的部分即为不与RNA结合的磁珠非特异性吸附蛋白, 记为黑色圆点; 对照组与实验组有明显差异, 且满足1.2.6节所述具有2个以上Unique peptide的即为高置信的RAP和RBP, 记为红色圆点. 最终, 共鉴定到高置信的RBP 2007个, 高置信RAP 927个.

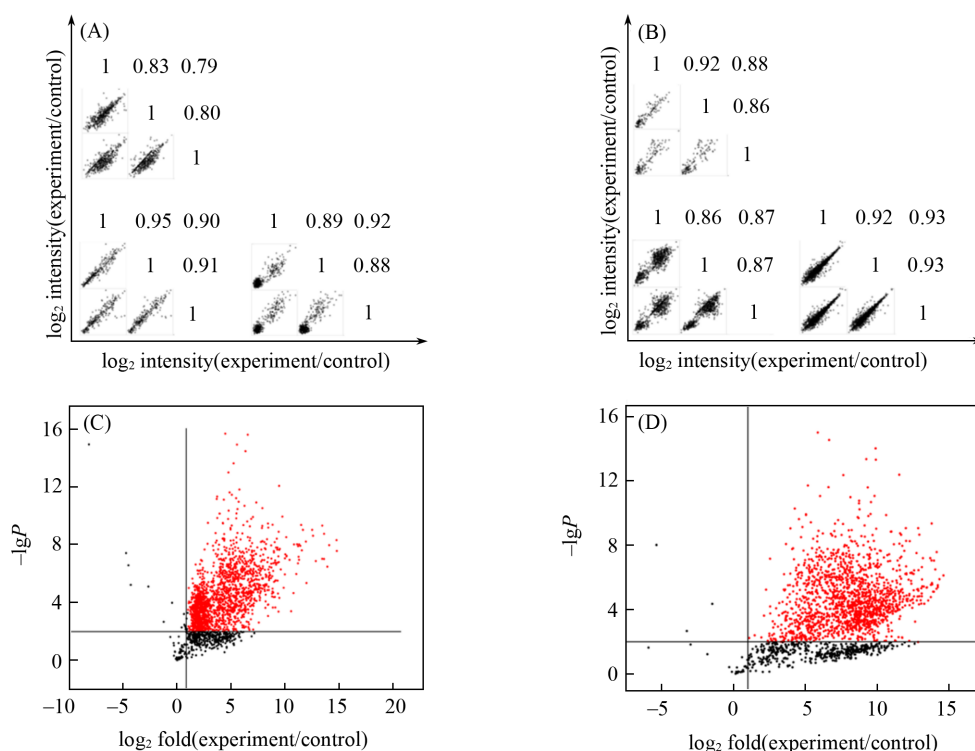
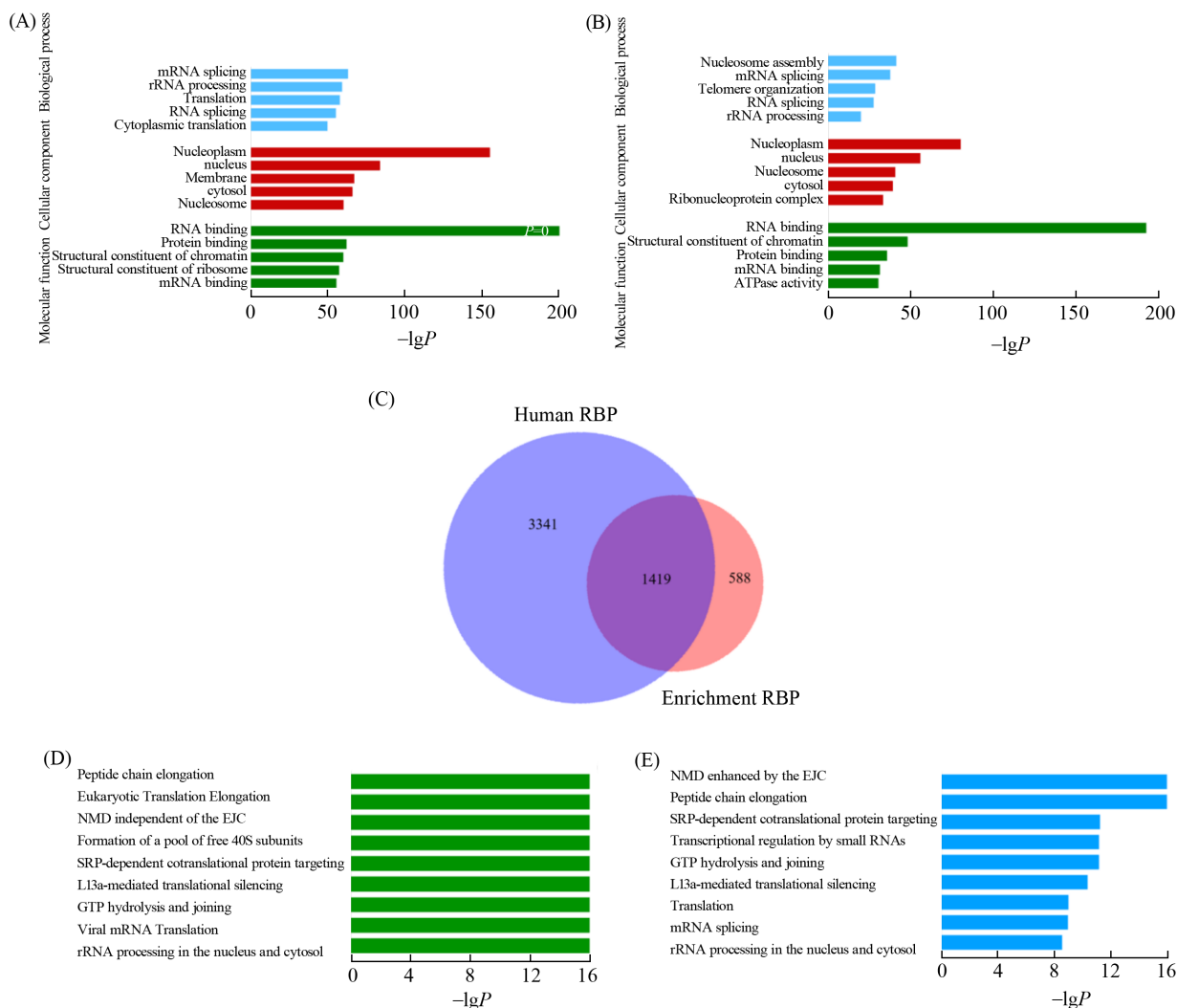


Fig. 2 Correlation analysis chart(A, B) and volcano plot(C, D) of RBP(A, C) and RAP(B, D)

对富集到的高置信RBP和RAP分别进行了GO(Gene ontology)分析, 其结果如[图3(A)和(B)]所示. 相应的分析条目按照生物学过程(Biology process)、细胞组成(Cellular component)和分子功能(Molecular function)依次列出. 本文分别列出了RBP和RAP排在前五位的富集条目. 对于RBP, 这些条目几乎都与RNA的加工进程有关, 例如BP条目下mRNA(信使RNA)的剪接, rRNA(核糖体RNA)的加工和翻译等都与RNA的生理过程息息相关; 在CC条目下排名靠前的核质(Nucleoplasm)、核内(Nucleus)、细胞膜(Membrane)和胞质(Cytosol)词条也说明了富集出的蛋白主要位置和RNA的生理功能和集中分布区域重合; 在MF条目中, 排名第一的RNA binding(RNA结合)词条也能直接说明富集出的蛋白与RNA的相关性, 进一步证实了富集出的RBP具有较高的可信度和后续研究价值. 对富集出的RAP的GO条目进行解析发现, 除了与RNA高度相关的条目之外, 在MF条目下的ATPase activity(ATP酶活性)类别中, 发现了多种与以往研究<sup>[32]</sup>报道一致的与核苷酸和ATP结合的蛋白, 这一发现可能有助于未来研究中深入探索RAP对RNA的功能调控作用. 除了GO条目的分析, 为了进一步证实富集出的RBP与此前报道的已知RBP的一致性, 绘制了韦恩图[图3(C)]. 将近5年来经典文献中鉴定到的人类RBP汇总, 与本文在HeLa细胞中富集到的2007个RBP进行比对发现, 本文鉴定的RBP有71%的蛋白与人类已知RBP重叠, 证明了本文富集出的RBP的可信性的同时, 也提供了一些未被报道的潜在新RBP. 除了上述分析, 为了进一步验证本文富集鉴定出的RBP和RAP的可靠性并发掘其潜在生物学价值, 分别对富集出的高置信RBP和RAP做了通路(pathway)分析[图3(D)和(E)], 图中列出了排在前九位的条目. 通过分析图中所示条目可以发现, 其中排名靠前的核糖体亚基的形成与水解、mRNA翻译、

沉默介导的衰变和 rRNA 加工这些都与 RNA 加工进程有关, 进一步证实了该方法富集 RPC 的可靠性. 此外, 在图 3(E) 中出现的肽链延长 (Peptide chain elongation) 条目, 也能说明 RAP 与蛋白-蛋白相互作用的相关性.



**Fig. 3** GO enrichment analysis of RBP(A), GO enrichment analysis of RAP(B), Venn diagram showing overlap of the RBP identified in this article, reported known human RBP(C), RBP identified in this article pathway analysis(D) and RAP identified in this article pathway analysis(E)

### 3 结 论

设计了一种基于两步交联、分步分离的细胞内 RPC 规模化富集方法, 实现了 RAP 和 RBP 的准确区分. 应用此方法从 HeLa 细胞中鉴定到 2007 个高度可信的 RBP, 覆盖了之前已知人类 RBP 的 71%. 此外, 还鉴定到与 RNA 高度相关的 927 个高置信 RAP, 为后续针对 RPC 的生物学研究提供了新的手段. 本文方法与此前建立的依赖 poly(A) 的 mRNA 结合蛋白富集方法相比, 具有无偏性富集所有种类 RNA 及其 RPC 的优势. 与依赖 RNA 物理性质富集 RBP 的分离方法相比, 本文方法通过共价反应偶联 RNA 与 RBP 和 RAP, 不但为 RPC 的富集鉴定提供了新的手段与角度, 与之前的方法互为补充, 在复杂生物样本体系中也具有更高的选择性. 虽然本文仅将此方法应用于规模化富集 HeLa 细胞的 RPC, 但理论上基于 254 nm 和 DSP 交联的 RPC 固定手段可以应用于所有细胞样本, 同时也因为 254 nm 紫外光不足以穿透组织, 所以本方法在生物组织层面富集 RPC 还存在一些局限性.

参 考 文 献

- [ 1 ] Müller-McNicoll M., Neugebauer K. M., *Nat. Rev. Genet.*, **2013**, *14*, 275—287
- [ 2 ] Chen Y. Y., Gui Z., Hu D., Chen M. Y., He J., Yu S. Y., Feng Y. Q., Wang J., Yuan B. F., *Chinese Chem. Lett.*, **2023**, *35*, 108522
- [ 3 ] Guo C., Zhang X. X., Hong X. J., Hu Y. Q., Mao L. N., Jiang K. Z., *Chinese Chem. Lett.*, **2023**, *35*, 108867
- [ 4 ] Wang Q., Jiang Q., Li D., Yang Z., Gao L., Liu F., Li C., Feng Y., He Z., Luo C., Sun J., *Chinese Chem. Lett.*, **2023**, *35*, 108683
- [ 5 ] Hu B., Kong S., Weng Y., Zhao D., Hussain A., Jiao Q., Zhan S., Qiu L., Lin J., Xie M., Li B., Huang Y., *Chinese Chem. Lett.*, **2023**, *34*, 108210
- [ 6 ] Hocq R., Paternina J., Alasseur Q., Genovesio A., Le Hir H., *Nucleic Acids Res.*, **2018**, *46*, 11553—11565
- [ 7 ] Hudson W. H., Ortlund E. A., *Nat. Rev. Mol. Cell Biol.*, **2014**, *15*, 749—760
- [ 8 ] Al-Husini N., Tomares D. T., Pfaffenberger Z. J., Muthunayake N. S., Sa-mad M. A., Zuo T., Bitar O., Aretakis J. R., Bharmal M. -H. M., Gega A., *Mol. Cell*, **2020**, *78*, 670—682
- [ 9 ] Valentin-Vega Y. A., Wang Y. D., Parker M., Patmore D. M., Kanagaraj A., Moore J., Rusch M., Finkelstein D., Ellison D. W., Gilbertson R. J., *Sci. Rep.*, **2016**, *6*, 25996
- [ 10 ] Somasekharan S. P., El-Naggar A., Leprivier G., Cheng H., Hajee S., Grunewald T. G. P., Zhang F., Ng T., Delattre O., Evdokimova V., *Cell Biol.*, **2015**, *208*, 913—929
- [ 11 ] Brownsword M. J., Locker N., *Rev. RNA*, **2022**, *14*, e1741
- [ 12 ] Verdile V., de Paola E., Paronetto M. P., *Front. Genet.*, **2019**, *10*, 173
- [ 13 ] Ramaswami M., Taylor J. P., Parker R., *Cell*, **2013**, *154*, 727—736
- [ 14 ] Castello A., Fischer B., Eichelbaum K., *Cell*, **2012**, 1393—1406
- [ 15 ] He C., Sidoli S., Warneford Thomson, *Mol. Cell*, **2016**, 416—430
- [ 16 ] Jao C. Y., Salic A., *Proc. Natl. Acad. Sci. USA*, **2008**, 15779—15784
- [ 17 ] Urdaneta E. C., Vieira-Vieira C. H., Hick T., Wessels H., Figini D., Moschall R., Medenbach J., Ohler U., Granneman S., Selbach M., Beckmann B. M., *Nat. Commun.*, **2019**, *10*(1), 990
- [ 18 ] Chomczynski P., Sacchi N., *Anal. Biochem.*, **1987**, 156—159
- [ 19 ] Chomczynski P., Sacchi N., *Nat. Protoc.*, **2006**, *1*, 581—585
- [ 20 ] Smith T., Villanueva E., Queiroz M. L., Dawson C. S., Elzek M., Urdaneta E. C., *Current Opinion in Chem. Bio.*, **2020**, *54*, 70—75
- [ 21 ] Trendl J., Schwarzl T., Horos R., *Cell*, **2018**, *37*(2), 169—178
- [ 22 ] Castello A., Horos R., Strein C., *Nat. Protoc.*, **2013**, *8*(3), 491—500
- [ 23 ] Bae J. W., Kwon S. C., Na Y., *Nat. Struct. Mol. Biol.*, **2020**, *27*(7), 678—682
- [ 24 ] Ryan D. P., Matthews J. M., *Current Opinion in Structural Bio.*, **2005**, *15*, 441—446
- [ 25 ] Ripin N., Parker R., *Cell*, **2023**, *186*, 4737—4756
- [ 26 ] Caudron-Herger M., Rusin S. F., Adamo M. E., Barreau E., Kettenbach A. N., Diederichs S., *Mol. Cell*, **2019**, *75*, 184—199
- [ 27 ] Street L., Rothamel K., Brannan K., Jin W., Bokor B., Dong K., Rhine K., *bioRxiv*, **2023**, 544225
- [ 28 ] Zhang Z., Liu T., Dong H., Li J., Sun H., *Nucleic Acids Research*, **2021**, *49*(11), e65
- [ 29 ] Li Y., Xia C., Zhao H., Xie Y., Zhang Y., Zhang W., Yu Y., Wang J., Qin W., *Chinese Chem. Lett.*, **2022**, *34*, 107377
- [ 30 ] Parimalam S. S., Oguchi Y., Abdelmoez M. N., Tsuchida A., *Anal. Chem.*, **2018**, *90*, *21*, 12512—12518
- [ 31 ] Xiang C. C., Mezey E., Chen M., Key S., Ma L., Brownstein M. J., *Nucleic Acids Research*, **2004**, *32*(22), e185
- [ 32 ] Mugler C. F., Hondele M., Heinrich S., Sachdev R., Vallotton P., Koek A. Y., Chan L. Y., Weis K., *eLife*, **2016**, *5*, e18746

(Ed.: N, K)