

中国博士后科学基金面上资助 申 请 书（流动站）

第 69 批

姓 名 孙磊

进站单位 清华大学生命学院

申报项目 利用人工智能方法整合 RNA
结构信息研究原始精原细胞发
育中 RNA 结合蛋白的作用

项目所属
一级学科 生物学

中国博士后科学基金会

须 知

1. 本表由申请者在“中国博士后科学基金管理信息系统”中填报并自动生成，网下生成无效。
2. “一、个人信息”中“1. 申请人基本情况”由系统自动填报。
3. “二、项目信息”中不得填写个人信息，包括申请人姓名、设站单位名称、合作导师姓名等，否则评审专家可视为申请人故意泄露个人信息，计0分。
4. 中文或英文均可。

| | | | | | |
|-------------------------|------|-------|------|--------|-------|
| 国家或部级 项目/课题 情况 | 下达时间 | 项目/课题 | 下达部门 | 经费(万元) | 负责情况 |
| | | | | | |
| 出版专著情 况 | 出版时间 | 书 名 | 出版社 | | 作者排名 |
| | | | | | |
| 已取得的专 利 | 取得时间 | 名 称 | 类 型 | 授权编号 | 批 准 国 |
| | | | | | |
| 获得国际、 国家及 部 委奖励情况 | 获奖时间 | 名 称 | 授予单位 | | 排 名 |
| | | | | | |

二、项目信息

注：请勿泄露个人信息，包括申请人姓名、设站单位名称、合作导师姓名等。可参考使用“本人”“本单位”“本实验室”等表述。

| | | | |
|----------|---|----------|-------|
| 项目名称 | 利用人工智能方法整合 RNA 结构信息研究原始精原细胞发育中 RNA 结合蛋白的作用 | | |
| 项目来源 | 自选 <input type="checkbox"/> 合作导师项目 <input checked="" type="checkbox"/> 自主获得的国家或省部级课题 <input type="checkbox"/> 其他 <input type="checkbox"/> | | |
| 项目所属一级学科 | 生物学 | 项目所属二级学科 | 发育生物学 |
| 交叉一级学科 | | 交叉二级学科 | |

1. 选题依据（国内外研究现状及选题价值，限 1000 字）

1.1 研究 RNA 及 RNA 结构的必要性与重要性

人类基因组中转录为 RNA 的区域，只有不到 2%是编码信使 RNA 的，其余部分编码长非编码 RNA，核糖体 RNA，转运 RNA，环状 RNA 等[1, 2]。这些种类丰富的 RNA 参与到细胞转录以及转录后调控的各个方面。RNA 调控与功能的实现，与 RNA 结构尤其是二级结构有密切关系[3, 4]。例如，Xist 的 repA 区域包含多个 stem-loop 结构单元，可以招募 PRC2 复合体发挥沉默 X 染色体的作用[5, 6]。然而，现在的很多研究依靠预测的 RNA 二级结构，活细胞内 RNA 结构的研究还存在巨大空白。申请人利用探测细胞内 RNA 结构的方法 icSHAPE，研究了细胞内不同组分 RNA 结构的动态变化[7]，RNA 结构改变对 RBP 结合影响[8]以及 SARS-CoV-2 RNA 结构对感染的影响[9]等重要的科学问题。

1.2 RNA 结合蛋白在生殖细胞分化中的重要作用

RNA 从转录产生，处理，转运，定位，出核，翻译到降解，都受到精密的调控[10, 11]。而这些调控的进行，与不同功能的 RNA 结合蛋白（RBP）在 RNA 上的识别结合密切相关[12, 13]。研究 RNA 与 RBP 的相互作用，对于深入研究 RNA 调控机制以及下游生物功能至关重要。

生殖细胞是个体发育的起源，生殖细胞的正常发育对于物种的延续至关重要。生殖细胞中含有多种 RBP[14]。RBP 对于原始生殖细胞发生以及配子发生至关重要。例如，RNA 结合蛋白 DAZL 对于人原始生殖细胞形成是必须的[15-17]。然而，受限于生殖细胞数目较少，更多的 RBP 功能还没有进行深入的研究。

1.3 深度学习方法在生物学中的应用

深度学习方法是受人类大脑连接和整合信息的多级神经元启发，利用多层神经网络构建模型，学习复杂任务的方法。2015 年，DeepBind 首次利用卷积神经网络（CNN）研究蛋白质与 RNA 或者 DNA 的相互作用[18]。在这之后，不同的深度学习方法被引入到生物学的研究中。例如，Ben-Bassat 等通过卷积神经网络以及递归神经网络（RNN）研究 RBP 与 RNA 相互作用[19]。上海交通大学潘小勇等通过整合卷积神经网络与双向长短时记忆网络（BLSTM），利用 RNA 序列以及预测的 RNA 结构，研究 RBP 与 RNA 的相互作用[20]。申请人作为主要参与者开发的基于 RNA 序列信息以及细胞内 RNA 结构信息的 RBP 结合位点预测模型 PrismNet，不仅可以反映特定的细胞内环境，而且对于预测结果的生物学意义有更好的解释[8]。

综上所述，我们利用体外构建的生殖细胞诱导系统，获得充足的胚胎干细胞以及原始精原细胞，并

解析其细胞内全转录组 RNA 二级结构。在现有的 RBP 结合数据基础上，训练不同 RBP 对于 RNA 结合序列以及结构的偏好性。利用训练好的模型，在胚胎干细胞及原始精原细胞中预测不同 RBP 的结合情况。通过分析 RBP 结合的差异，探索在生殖细胞分化中发挥重要功能的 RBP，并通过实验验证 RBP 的功能。

2. 研究内容（研究对象，拟解决的关键科学问题，研究目标，限 2000 字）

研究目标：

构建由胚胎干细胞体外培养诱导原始精原细胞（pro spermatogonia）的系统，通过 icSHAPE 方法测定不同细胞的全转录组 RNA 结构及 RNA 结构的动态变化。利用已知其他细胞系的 168 个人类 RBP 的结合位点信息（通过已发表 CLIP 数据整理所得），训练 RBP 与 RNA 结合的包含序列与结构信息的深度学习模型。在生殖细胞系统中，基于不同时期 RNA 结构的差异，预测 RBP 结合的差异。探索 RBP 在生殖细胞分过程中的功能以及可能的机制并通过后续实验进行验证。

研究方法：

1) 人的胚胎干细胞分化为原始精原细胞

哺乳动物原始生殖细胞由原肠胚时期内胚层（近端上胚层）细胞分化而来，在发育过程中迁移到生殖嵴处，结合不同的生殖嵴中胚层细胞（性腺体细胞），决定分化为精原细胞或卵原细胞从而完成性别分化。雄性哺乳动物的个体性腺中原始生殖细胞发育为原始精原细胞（pro spermatogonia）。出生后的原始精原细胞进入有丝分裂期产生两种不同的细胞命运，一种是精原细胞（spermatogonia）经过数次有丝分裂分化为精母细胞，进入精子发生过程。精母细胞通过减数分裂过程产生单倍体精细胞，通过精子分化变形变为成熟的精子，完成一个完整的生精周期。另一种是原始精原细胞发育为未分化的精原细胞——精原干细胞，之后精原干细胞通过自我更新维持自身数量的稳定并分化产生各级生精细胞。

本研究通过干性培养基体外培养人胚胎干细胞 H1，然后利用分化培养基定向分化原始精原细胞。通过检测关键调控因子的基因表达与蛋白表达，确定分化的效率。

2) 解析胚胎干细胞与原始精原细胞内 RNA 结构信息

细胞内 RNA 结构信息可以反映细胞内环境，同时，不同的细胞内环境也会影响 RNA 的二级结构。利用 icSHAPE 技术，可以通过穿膜的化学小分子 NAI-N₃，特异性地修饰细胞内单链 RNA 的 2' 羟基，标识活细胞内的 RNA 结构信息。在 RNA 逆转录过程中，化学修饰可以阻断逆转录的进行，产生一个逆转录停止位点（RT stop），该停止位点的前一位即细胞内的单链 RNA 区域。通过高通量测序的方法，可以同时获取全转录组上万条 RNA 的二级结构信息。

本研究利用上述 icSHAPE 方法，针对人胚胎干细胞 H1 以及诱导分化的原始精原细胞，探测不同细胞内 RNA 结构信息。因为人胚胎干细胞以及原始精原细胞比普通细胞更难获得，我们会优化 icSHAPE 实验步骤，使用更低起始量 RNA 进行实验。

3) 构建更准确的基于结构的深度学习模型

在前期研究中，通过 CLIPdb，ENCODE 等数据库，我们已经收集了 168 个人类 RNA 结合蛋白的 CLIP 数据，获得了其特异的结合位点。同时，我们也解析了包括 HEK293 细胞系，K562 细胞系，HepG2 细胞系，HeLa 细胞系等不同的细胞系的 RNA 结构，获得了与 CLIP 数据对应的细胞系的 RNA 二级结构信息。通过整合 RBP 结合的序列信息与结构信息，利用深度学习模型，我们构建了不同 RBP 识别 RNA 的序列与结构整合基序（sequence and structure integrative motif）。

之前研究的模型并没有考虑蛋白的表达水平，以及不同蛋白质之间的相互作用。在之前研究的基础上，我们将进一步整合蛋白表达水平以及不同 RBP 的结合信息，以更准确的预测 RBP 的结合位点。

4) 预测原始生殖细胞内重要的 RNA 结合蛋白

基于上述方法获得的全转录组的 RNA 二级结构信息以及更准确的基于 RNA 结构的 RBP 结合位点预测模型，我们可以预测 RNA 结合蛋白在不同细胞环境下结合的差异。通过整合 RBP 结合差异信息，蛋白质表达信息以及 RBP 结合转录本的位点信息，我们可以预测在原始生殖细胞分化与维持中起重要作用的 RNA 结合蛋白。

5) 验证原始生殖细胞内重要 RNA 结合蛋白的功能

现在已经发现大量在生殖细胞发育过程中发挥重要作用的 RNA 结合蛋白，比如 DAZL, LIN28A, DND1, NANOS3 以及 VASA 等。在这些 RNA 结合蛋白中，有的是原始生殖细胞形成必须的，如 DAZL[13-15]；有的可以在转录调控以及转录后调控同时起作用，如 LIN28A[16, 17]；有的 RBP 可能间接的影响生殖细胞的发育，如 NANOS3 高表达与 DAZL 的 mRNA 表达下降有关[21]，暗示可能对 DAZL 有潜在抑制作用。

针对上述预测的有潜在功能的 RBP，我们可以通过 RBP 敲低与敲除实验，验证 RBP 是否会影响胚胎干细胞到生殖细胞的分化。同时，还可以通过设计特定位点的反向互补短核苷酸序列(antisense oligo, ASO)，阻断 RBP 的结合，从而验证 RBP 起作用的具体机制。另外，可以通过体外构建包含特定 RBP 结合位点的表达体系，验证 RBP 起作用的具体机制，比如影响 RNA 的稳定性或者 RNA 的翻译速率。

拟解决的关键科学问题：

- 1. 胚胎干细胞与原始精原细胞内 RNA 二级结构以及动态变化。
- 2. 通过深度学习模型整合细胞内不同 RBP 的表达与结合信息，以及 RNA 序列与结构信息，预测 RBP 在胚胎干细胞与原始精原细胞内的结合变化。
- 3. 通过整合预测的 RBP 结合差异信息，蛋白质表达信息以及在转录本上的结合位点信息，预测并验证在原始生殖细胞分化与维持中发挥重要作用的 RNA 结合蛋白。

3. 研究方案（限 2000 字）

整体技术路线：

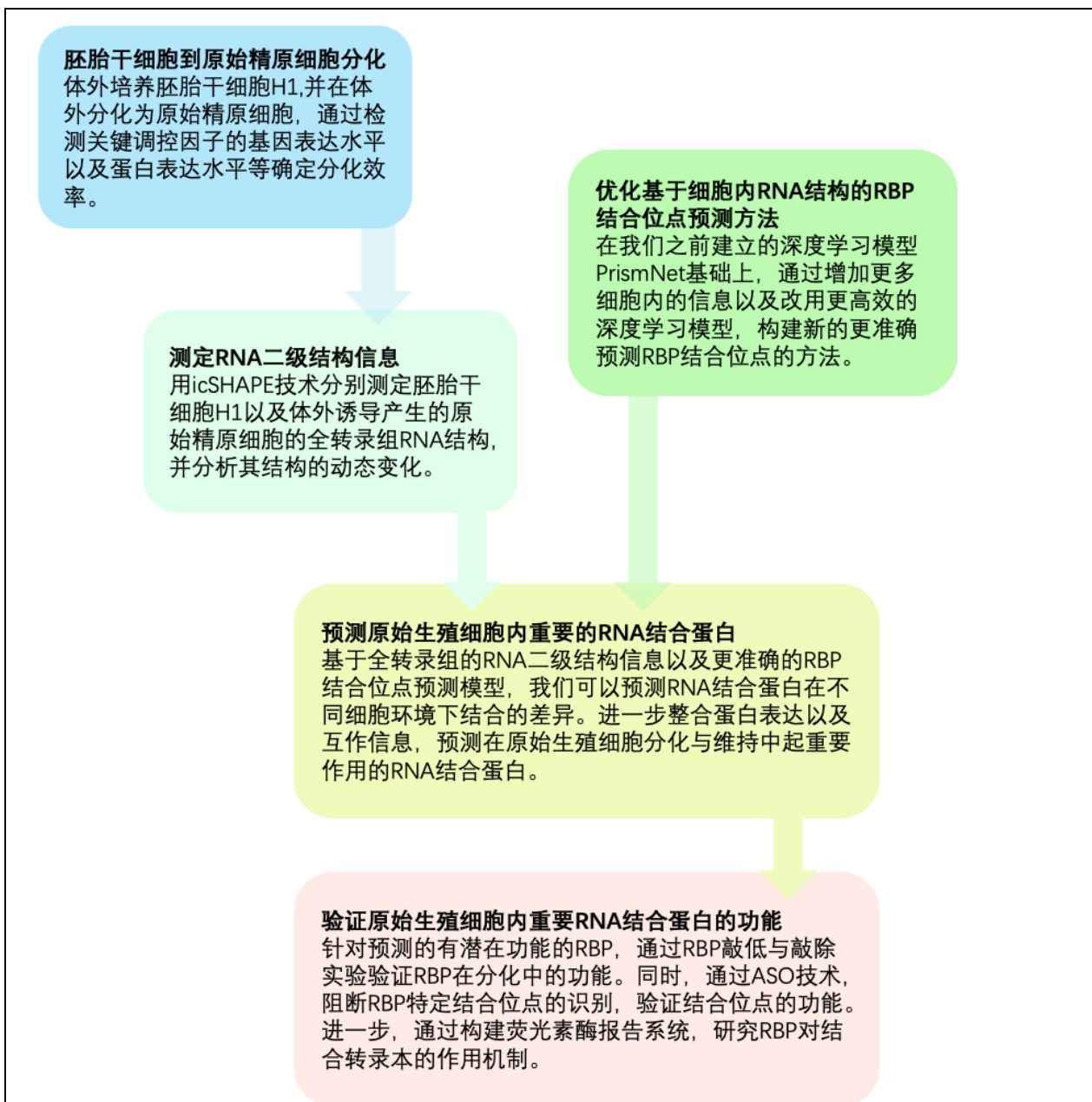


图1 研究技术路线

实验手段与关键技术:

1. 从胚胎干细胞向原始精原细胞定向分化

- 利用干细胞培养基进行胚胎干细胞 H1 培养。将干细胞分为 A, B, C, D 4 组, 其中 A, B 组只进行胚胎干细胞培养。C, D 组继续进行定向分化。
- 添加 glial-derived neurotrophic factor(GDNF)等生长因子在分化培养基中, 定向分化原始精原细胞, 并在不同的时间点检测原始精原细胞分化比例。
- 当原始精原细胞比例到达 80%以上, 进行接下来的 RNA 二级结构测定。

2. 解析 RNA 二级结构

- 对培养完成的胚胎干细胞以及分化的原始精原细胞, 实验组 A, C 组利用可以特异性标记单链 RNA 的 NAI-N₃ 修饰, 对照组 B, D 组利用 DMSO 修饰。所有修饰都是在活细胞状态下进

行。

- b) 利用 TRIzol™ 试剂提取 RNA，对于提取的 RNA 利用 polyA 进行富集，去除核糖体 RNA。利用 2100 检测 RNA 质量以及核糖体 RNA 去除效率。
- c) 将 RNA 打断，进行随机引物逆转录，cDNA 停止位点记录了单链修饰位点的信号。然后对 cDNA 进行末端加接头，合成双链，末端修复，双链接头连接等步骤，构建 icSHAPE 文库。
- d) 对大于 175nt 的文库进行切胶回收，并通过 2100 与 qPCR 质控文库，然后上机测序。
- e) 对原始测序数据进行质控，对于质量合格的数据利用 icSHAPE-pipe 进行分析，获得全转录组 RNA 二级结构。
- f) 通过比较胚胎干细胞与原始精原细胞的转录组数据，识别 RNA 结构动态变化区域，这些区域有更高的 RBP 结合差异可能性。

3. 建立更完善的深度学习网络

- a) 针对模型的数据量，我们已经开发的 PrismNet 方法包含了 168 个人类 RBP 的训练模型，在此基础上在收集更多新的 RBP 结合位点训练模型。
- b) 针对模型的训练维度，将更多细胞内信息作为新模型的输入进行训练，主要包括蛋白质的表达量以及所有已知 RNA 结合蛋白的结合位点信息。
- c) 针对模型的结构，采用最新的 Transformer 模型框架。在之前的 PrismNet 模型中我们已经引入“关注机制”(attention-mechanism) 寻找 RBP 结合高响应的区域。通过 Transformer 框架，可以更精准的定位高响应的序列以及结构位置。
- d) 将 RBP 结合位点的序列信息，RNA 结构信息以及该位点其他 RBP 的结合信息作为输入，整合新的 Transformer 深度学习框架，训练新的深度学习模型。

4. 预测并验证原始精原细胞分化过程中关键 RNA 结合蛋白

- a) 通过整合 RBP 结合差异比例，RNA 结合蛋白表达差异以及具体的 RNA 结合位点的转录本，对原始精原细胞的重要 RBP 进行预测排序。
- b) 利用 siRNA 对排名靠前的 RBP 进行敲低实验，然后重复从胚胎干细胞向原始精原细胞定向分化的实验，验证是否会影响原始精原细胞分化。
- c) 利用 ASO 特异性的阻断关键 RBP 的结合位点，或者敲低特异性的 RNA，验证 RBP 是否通过特异的 RNA 结合行使作用。
- d) 通过构建荧光素酶报告系统，插入诱导分化关键的转录本，研究特定 RBP 对结合转录本的作用机制。是否影响了 RNA 的翻译或者降解过程。

4. 特色与创新之处 (限 1000 字)

本项目生物信息学分析与生物实验验证相结合，通过整合高通量测序方法、深度学习算法、体外干细胞定向分化以及传统的验证实验，解决关于生殖以及基因表达调控的分子机制，其特色与创新之处有：

- a) 通过体外培养胚胎干细胞以及体外定向分化原始精原细胞，获取充足的实验材料。利用 icSHAPE 测定细胞内全转录组 RNA 结构需要较多的细胞。利用 CLIP 实验检测特定 RBP 的结合位点也需要大量的细胞。受限于体内的生殖细胞有限，很多相关的分子调控机制没有深入研究。通过体外实验系统，可以有效弥补实验材料不足。
- b) 通过深度学习模型，对所有已知的 RNA 结合蛋白在原始精原细胞进行系统性的结合预测，可以同时研究所有 RBP 可能在原始精原细胞内的作用，而通过实验方法很难完成如此大规模的筛选。

- c) 通过整合高通量测序，深度学习模型以及实验验证，提供了一种系统研究 RNA 结合蛋白作用的可行思路。在很多其他的类似的系统上，都可以利用此方法，将之前无法研究或者不能系统性研究的问题进行系统性的研究。比如，在病毒侵染细胞过程中，宿主蛋白的结合发生了怎么样的变化？这些 RBP 的结合差异会对病毒的复制扩增有怎样的影响？RBP 的结合差异对宿主蛋白发挥作用有怎么样的影响？这些问题，都可以通过我们提供的思路进行研究。

5. 研究计划及预期成果（限 500 字）

研究计划：

2021 年

- 1) 完成胚胎干细胞到原始精原细胞的定向分化。
- 2) 完成相关细胞的 RNA 二级结构测定。

2022 年

- 1) 完成对 PrismNet 的模型优化，预测胚胎干细胞和原始精原细胞中所有的 RBP 结合位点的结合预测。
- 2) 预测并验证在原始精原细胞中发挥重要作用的 RNA 结合蛋白。
- 3) 撰写文章。

预期成果：

- 1) 解析胚胎干细胞以及原始精原细胞内全转录组 RNA 结构图谱；
- 2) 构建整合细胞内序列信息，结构信息，蛋白表达信息的预测 RBP 在细胞内结合的深度学习模型；
- 3) 发现并验证在原始精原细胞中发挥作用的新的 RNA 结合蛋白；
- 4) 在国际权威期刊发表具有影响力的文章；
- 5) 参加国际会议，特别是 RNA 领域的会议，提高本工作的影响力。

6. 研究基础（与本项目相关的研究工作积累和已取得的研究工作成绩；已具备的科研条件，尚缺少的科研条件和拟解决的途径；正在承担的与本项目相关的科研项目情况，限 1000 字。）

与本项目相关的研究工作积累和已取得的研究工作成绩：

- 1) 申请人解析了 6 个人类细胞系内 RNA 全转录组 RNA 二级结构，从 POSTAR, ENCODE 等数据库收集了所有 RBP 结合位点信息。通过整合 RBP 结合位点的序列信息以及结构信息，构建了可以准确预测 RBP 结合位点的深度学习模型 PrismNet。该方法不仅可以准确预测 RBP 结合位点，还可以预测同一物种，不同细胞状态的 RNA 结合位点差异，比如本申请书讨论的胚胎干细胞与原始精原细胞内 RBP 结合的差异。该文章已经发表在 *Cell Research* 杂志。申请人是共同一作排名第一。
- 2) 利用开发的 PrismNet 方法，通过 icSHAPE 探测了 SARS-CoV-2 RNA 的二级结构，通过整合 RNA 序列信息以及 RNA 结构信息，我们在 SARS-CoV-2 RNA 上预测 RBP 的结合位点。我们发现了 42 个可以结合在 SARS-CoV-2 非编码区的 RBP。通过 RNA pull-down、质谱分析、RNA 敲低、ASO 阻断等实验方法，验证了我们预测结果的准确性。另外，通过利用预测的 RBP，寻找重定向的 FDA 批准的药物，找到并验证了多种可以显著抑制病毒扩增的药物。该文章已经发表在 *Cell* 杂志。申请人是共同一作排名第一。

已取得的初步试验结果:

1) 建立了完善的胚胎干细胞到原始生殖细胞分化的体系。利用上述方法，将胚胎干细胞培养为原始精原细胞，并且通过细胞分选法检测细胞分化的效率（图 2）。

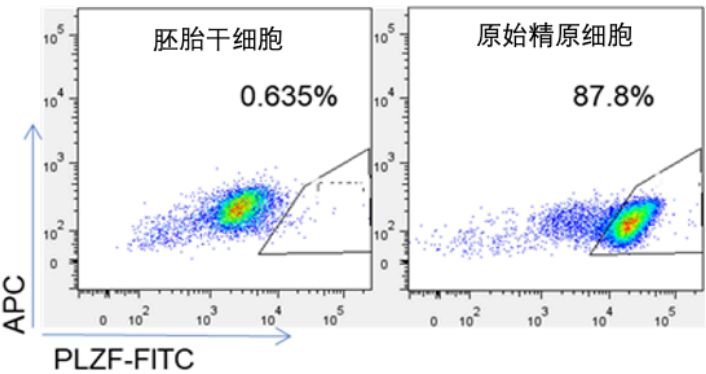


图2 通过细胞分选检测胚胎干细胞分化为原始精原细胞的比例。

2) 建立了完善的测定细胞内 RNA 二级结构的方法。

申请人熟悉利用 icSHAPE 技术测定细胞内 RNA 结构的实验流程。在获得需要的细胞系的基础上，我们已经初步测定了一部分胚胎干细胞 H1 以及原始生殖细胞的 RNA 结构。测序的质量控制如表 1，核糖体 RNA18S 的结构图谱如图 3。

| 细胞类型 | H1 | | | |
|----------|----------|----------|-----------|-----------|
| 处理 | DMSO1 | DMSO2 | NAI1 | NAI2 |
| 原始测序数据 | 38221909 | 58767675 | 253389030 | 201431609 |
| 去重复后数据 | 34687120 | 54164401 | 217453689 | 174105174 |
| 去接头后数据 | 33542217 | 51760304 | 140562390 | 115101719 |
| 回帖到转录组数据 | 23043485 | 35339485 | 98155241 | 80447745 |
| 回帖到转录组比例 | 68.70% | 68.28% | 69.83% | 69.89% |

表 1 将胚胎干细胞 H1 建库测序所得 4 个不同处理的测序数据。分析质控数据发现，我们所做样本去重复，去接头，去核糖体 RNA 之后，仍然有充足的数据进行后续分析。

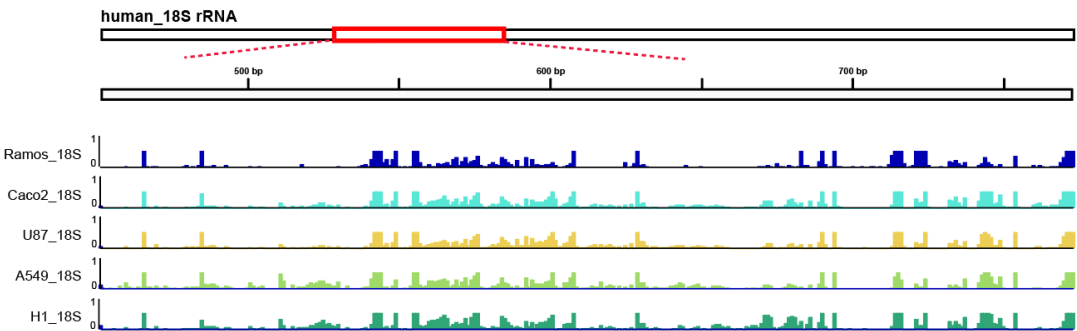


图3 18S 核糖体 RNA 在 Ramos, Caco2, U87, A549 以及 H1 细胞系的 icSHAPE 值。如上图所示，H1 细胞与其他的细胞的核糖体 RNA 的 icSHAPE 值比较一致，体现了方法的稳定性。

3) 初步尝试增加更多的细胞内信息预测 RBP 结合位点。

在 PrismNet 方法的基础上，我们首先尝试了增加更多细胞内的信息来提高预测的准确性。在每个结合位点的训练输入数据中，通过增加其他蛋白的结合位点的信息，可以显著增加预测的准确性（图 4）。

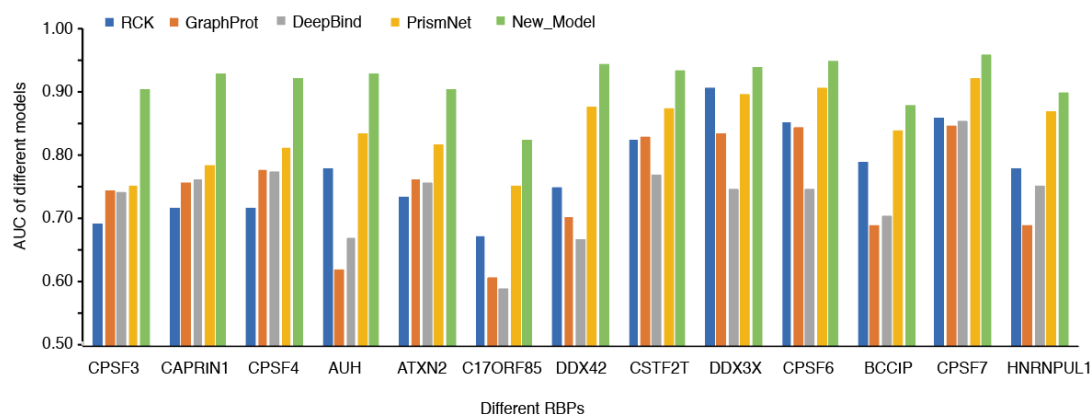


图 4 不同模型预测 RBP 结合位点的准确性 (AUC)。上图展示了 RCK, GraphProt, DeepBind, PrismNet 以及我们新优化的方法 (New_Model) 在 12 个 RBP 上预测结合位点的准确性。其中 RCK 与 GraphProt 是利用序列以及预测的 RNA 结构训练模型; DeepBind 是只利用 RNA 序列训练模型; PrismNet 是利用 RNA 序列以及细胞内的 RNA 结构作为输入训练模型; 我们新的模型不仅仅利用了 RNA 序列与细胞内 RNA 结构, 而且利用了其他 RBP 在 RNA 上的结合信息。

参考文献

1. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. Science, 2005. **309**(5740): p. 1559-63.
2. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
3. Pan, T. and T. Sosnick, *RNA folding during transcription*. Annu Rev Biophys Biomol Struct, 2006. **35**: p. 161-75.
4. Warf, M.B. and J.A. Berglund, *Role of RNA structure in regulating pre-mRNA splicing*. Trends Biochem Sci, 2010. **35**(3): p. 169-78.
5. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome*. Science, 2008. **322**(5902): p. 750-6.
6. Lu, Z., et al., *RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure*. Cell, 2016. **165**(5): p. 1267-79.
7. Sun, L., et al., *RNA structure maps across mammalian cellular compartments*. Nat Struct Mol Biol, 2019. **26**(4): p. 322-330.
8. Sun, L., et al., *Predicting dynamic cellular protein-RNA interactions using deep learning and in vivo RNA structure*. Cell Research, 2021.
9. Sun, L., et al., *In vivo structural characterization of the whole SARS-CoV-2 RNA genome identifies host cell target proteins vulnerable to re-purposed drugs*. Cell, 2021.
10. Sharp, P.A., *The centrality of RNA*. Cell, 2009. **136**(4): p. 577-80.
11. Martin, K.C. and A. Ephrussi, *mRNA localization: gene expression in the spatial dimension*. Cell, 2009. **136**(4): p. 719-30.
12. Lunde, B.M., C. Moore, and G. Varani, *RNA-binding proteins: modular design for efficient function*. Nat Rev Mol Cell Biol, 2007. **8**(6): p. 479-90.
13. Glisovic, T., et al., *RNA-binding proteins and post-transcriptional gene regulation*. FEBS Lett, 2008. **582**(14): p. 1977-86.

14. Lesch, B.J. and D.C. Page, *Genetics of germ cell development*. Nat Rev Genet, 2012. **13**(11): p. 781-94.
15. Chen, H.H., et al., *DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells*. Stem Cell Reports, 2014. **3**(5): p. 892-904.
16. Gill, M.E., et al., *Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells*. Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7443-8.
17. Lin, Y., et al., *Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos*. Science, 2008. **322**(5908): p. 1685-7.
18. Alipanahi, B., et al., *Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning*. Nat Biotechnol, 2015. **33**(8): p. 831-8.
19. Ben-Bassat, I., B. Chor, and Y. Orenstein, *A deep neural network approach for learning intrinsic protein-RNA binding preferences*. Bioinformatics, 2018. **34**(17): p. i638-i646.
20. Pan, X. and H.B. Shen, *RNA-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach*. BMC Bioinformatics, 2017. **18**(1): p. 136.
21. Shinoda, G., et al., *Lin28a regulates germ cell pool size and fertility*. Stem Cells, 2013. **31**(5): p. 1001-9.
22. Zeng, Y., et al., *Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression*. Mol Cell, 2016. **61**(1): p. 153-60.
23. Panula, S., et al., *Over Expression of NANOS3 and DAZL in Human Embryonic Stem Cells*. PLoS One, 2016. **11**(10): p. e0165268.