

HE Minmin, ZHU Yufeng, LÜ Xuan, TANG Guangyan, QUAN Yuan. Mechanism of the Effect of Ionizing Radiation on Human B Cells Based on Network-guided Random Forest (in Chinese). *Chinese Journal of Space Science*, 2025, 45(2): 506–516. CSTR:32142.14.cjss.2024-0055. DOI:10.11728/cjss2025.02.2024-0055

基于网络引导随机森林的电离辐射 对人类 B 细胞影响机制*

何敏敏 朱玉锋 吕 萱
唐广燕 全 源

(华中农业大学信息学院 武汉 430070)

摘 要 在空间环境中, 电离辐射可破坏细胞 DNA 和其他分子结构, 导致细胞突变或死亡, 增加癌症及其他疾病风险. 解析电离辐射对人类细胞的影响机制已成为航天医学领域亟待解决的问题. 近年来, 空间组学数据的大量累计以及生物信息学技术的发展为该问题的解决提供了可行途径. 本文使用网络引导随机森林 (Network Guided Random Forest, NGF) 算法研究人类 B 细胞对电离辐射的响应机制. 基于基因功能富集分析发现人类 B 细胞在经过大剂量电离辐射后难以对受损伤的 DNA 进行正常修复, 大量细胞发生凋亡或癌变. 基于 cMap (Connectivity Map) 数据库进行潜在抗电离辐射药物筛选, 结果显示紫杉醇和醉椒素等天然产物或可辅助人体抵抗电离辐射伤害. 研究结果将为空间环境对人体影响机制的解析奠定基础, 辅助开展航天员抗空间逆境策略研究.

关键词 电离辐射, 人类 B 细胞, 网络引导随机森林算法, 基因功能分析, 药物筛选

中图分类号 V7

Mechanism of the Effect of Ionizing Radiation on Human B Cells Based on Network-guided Random Forest

HE Minmin ZHU Yufeng LÜ Xuan
TANG Guangyan QUAN Yuan

(College of Information, Huazhong Agricultural University, Wuhan 430070)

Abstract With the intensification of the problems of the Earth's population, energy and ecological environment, space immigration may become a feasible solution. In the space environment, ionizing radiation will damage the DNA and other molecular structures of human cells, causing cell mutation or death, increasing the risk of cancer and other diseases. The mechanism of the effect of ionizing radiation on hu-

* 中国科协青年人才托举工程项目 (2023QNRC001) 和国家自然科学基金青年项目 (32300545) 共同资助

2024-04-08 收到原稿, 2024-09-14 收到修定稿

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>)

man cells has become an urgent problem in the field of space medicine. In recent years, a large amount of spatial omics data has accumulated, and the development of bioinformatics technology provides a feasible way to solve the above problems. In this study, the Network Guided Random Forest (NGF) algorithm was used to investigate the response mechanism of human B cells to ionizing radiation. Based on gene function enrichment analysis, it was found that human B cells could not repair the damaged DNA normally after high dose of ionizing radiation, and a large number of cells suffered apoptosis or cancer. In addition, screening of potential anti-ionizing radiation drugs based on the cMap (Connectivity Map) database showed that natural products such as paclitaxel and zonicin may assist the human body in resisting ionizing radiation damage. This paper will lay a foundation for the analysis of the influence mechanism of space environment on human beings, and help astronauts to study the strategies of resisting space adversity.

Key words Ionizing radiation, human B cells, Network Guided Random Forest (NGF), Gene function analysis, Drug screening

0 引言

未来随着更多的空间探索任务远离地球磁场的保护,航天员将会遭受比之前空间飞行更高剂量的电离辐射。电离辐射对人体伤害极大,可以损伤生物体中多种活性物质,尤其是生物大分子^[1]。辐射危害一般以两种形式产生:一是直接作用于生物体活性物质(例如 DNA),细胞内 DNA 对辐射很敏感,可直接因为吸收辐射能量造成损伤,导致 DNA 单链或者双链断裂;二是辐射对生物大分子中其他分子起作用,产生自由基,使得生物大分子受损,这是辐射对人体伤害的间接损伤^[2]。长时间高剂量的电离辐射会导致人体产生许多健康问题。例如,高剂量电离辐射会诱导淋巴细胞凋亡,导致免疫力下降^[3]。在现实生活中,1 Gy 是引起急性辐射综合症的最低剂量,可使得患致命癌症的风险从约 22% (美国平均癌症风险)增加到约 27%,而暴露于 6 Gy 或 8 Gy 以上剂量的受害者在接受治疗的前提下也极有可能死亡^[4]。

B 细胞是人体免疫系统的重要组成部分,负责产生抗体并参与免疫应答。在空间环境中,航天员免疫系统面临极大的挑战,电离辐射会显著影响 B 细胞的功能。研究表明,电离辐射会导致 B 细胞凋亡,抑制抗体产生,进而削弱免疫系统的整体防御能力^[5]。这不仅增加了航天员感染疾病的风险,还可能导致长期健康问题,如癌症和免疫系统紊乱^[6]。因此,深入研究电离辐射对 B 细胞的影响,对于保障航天员的生理健康至关重要^[7]。

综上所述,人类 B 细胞对电离辐射的响应机制研

究可以帮助空间工作人员保持身体健康。现有研究已经识别出一些电离辐射对细胞的损伤机制,例如 DNA 损伤、氧化蛋白损伤、细胞凋亡和 p53 信号通路^[5]。

传统的基因表达分析方法和差异表达基因筛选技术往往难以全面揭示复杂的辐射响应机制。这些方法通常只能捕捉到表达变化显著的基因,而对于那些在基因调控网络中起关键作用但是表达变化不显著的基因则容易被忽视。此外,现有研究大多集中于单一基因或信号通路,缺乏系统性和整体性,难以全面解析电离辐射对细胞的多层次影响^[8]。

为了解决这些问题,引入了网络引导随机森林 (Network Guided Random Forest, NGF) 算法。NGF 作为一种新兴的生物信息学技术,提供了一种更为全面和准确的研究手段。NGF 算法将基因表达谱与蛋白质相互作用网络相结合,通过机器学习方法识别出与电离辐射响应相关的重要基因^[9]。与传统方法相比,NGF 算法具有更高的生物学解释性,能够捕捉到那些在基因调控网络中起重要作用但表达变化不显著的基因,从而提供更为准确的基因功能注释和通路分析。NGF 算法已经被用于组织发育、癌症以及植物免疫调控等相关研究中关键基因的鉴定,例如将 NGF 算法应用于蛋白质网络的逻辑功能研究,揭示了其在细胞发育和癌症中的关键作用。Dong 等^[10]则利用 NGF 算法研究了植物免疫响应中的基因网络组织,发现了一些关键调控基因。

研究应用 NGF 算法将人类 B 细胞在 10 Gy 电离辐射后 2 h 和 6 h 的基因表达谱数据和蛋白质相

互作用关系作为随机森林输入, 得到电离辐射相关基因重要性排序列表, 从而解析电离辐射响应机制, 筛选潜在抗辐射药物. 分析结果表明, 经电离辐射后初期的 B 细胞有抗辐射性, 但是细胞难以修复大剂量电离辐射导致的 DNA 损伤, 最终细胞存在向癌细胞转化的趋势. 同时用 NGF 算法研究人类 B 细胞对电离辐射的响应机制, 通过该算法可以鉴定出人体细胞中与空间电离辐射显著相关的多个基因, 并进行差异表达分析以及功能分析, 进一步阐明造成电离辐射损伤的机制. 进一步基于 cMap 数据库^{*} (Connectivity Map) 进行抗辐射损伤药物筛选, 得到多个潜在抗辐射损伤小分子药物, 推进人类抗电离辐射损伤研究.

1 材料与方法

1.1 数据来源与预处理

1.1.1 基因芯片数据

研究数据来源于基因表达综合数据库 (Gene Expression Omnibus, GEO) 中的基因芯片数据集^{**}, 注册号为 GSE36910. 95 个相互之间没有联系的 CEpH-Utah B 细胞中基因表达水平被测量. 用 10 Gy 电离辐射处理每个个体, 并在处理后 0, 2, 6 h 测量基因表达水平. 基因表达水平使用 Human U133 A 2.0 (Affymetrix) 平台测量, 使用 MAS 5.0 算法对基因表达信号进行均一化, 表达强度被缩放至 500, 并进行 log₂ 转换^[11].

使用探针平台注释文件对基因探针进行注释, 若一个探针名对应多个基因名则删除该数据, 而对于一个基因名对应多个探针名的情况, 则对表达值取平均值作为该基因最终表达值^[12].

1.1.2 蛋白相互作用数据

人类蛋白质互作数据从 String 数据库中获取^{***}, 蛋白互作信息来源包括实验数据、文本挖掘数据、数据库数据、基因邻接、基因融合和基因共表达^[13]. 该数据库对不同来源得到的基因相互作用信息赋予权重, 最终给出一个综合打分. 在本研究中, 选取打分大于 400 的蛋白相互作用关系进行后续分析.

1.1.3 cMap 基因芯片数据

药物基因组学数据库 cMap 是一个在线药物基因组数据库, 用于对来自用各种化学物质 (包括各种植物化学物质) 分别处理过的培养细胞基因表达数据进行分类. cMap 数据库包含用多种小分子药物处理 5 种人类肿瘤细胞系前后的基因芯片数据. 处理条件多种多样, 包括不同药物、不同浓度、不同处理时长等等, 大多数处理条件都有 3 个处理样本以及对应的 3 个及以上空白对照组. 基因表达谱建立了基因、疾病和药物关联, 可有效地帮助在药物开发领域快速利用基因表达谱的数据比对出与疾病具有高关联性的药物, cMap 是发现新化学预防药物的有效工具.

1.1.4 GO 和 KEGG 功能富集分析

DAVID 数据库 (Database for Annotation, Visualization and Integrated Discover) 集成了多种生物学数据资源和分析注释工具^{****}, 能够系统挖掘大规模基因/蛋白列表中的包含信息. DAVID 是最常用的功能富集分析网页版工具^[14]. 研究采用 DAVID 网站对网络引导随机森林算法鉴定出的与电离辐射相关的重要基因进行 GO 和 KEGG 功能富集, 分析每个基因涉及的生物学功能和代谢通路.

1.2 实验方法

1.2.1 网络引导随机森林算法

NGF 算法使用两种经过不同处理的基因表达谱数据和蛋白相互作用网络作为输入, 训练一个随机森林分类模型, 用来准确预测表达谱相关表型, 并根据每个基因或相互作用对模型分类准确性的提高程度对其赋予重要性系数 (Importance Score, IS). NGF 将随机森林 (Random Forest, RF) 关键思想与蛋白质相互作用网络生物学约束相结合, 这是蛋白质网络在集成学习中的首次应用.

研究使用 10 Gy 电离辐射处理后 0, 2, 6 h 的基因表达谱以及蛋白相互作用数据来训练电离辐射相关的两个不同时间点的 NGF 模型. 在构建模型过程中, 每棵树都使用分类回归树算法构建. 为生成决策树, 首先对所有表达谱子集进行随机抽样并替换, 以产生与原始表达谱大小相同的新表达谱数据集. 从所

* <https://www.broadinstitute.org/connectivity-map-cmap>

** <https://www.ncbi.nlm.nih.gov/geo/>

*** <https://string-db.org/>

**** <https://david.ncifcrf.gov>

有 N 个基因中随机选择 $\sqrt{N}$ 基因作为根节点的候选分裂节点, 并要求根节点在蛋白互作网络中度数至少为 5. 在树生长过程中, 使用 Gini 系数作为纯度度量, 判定基因和相应表达阈值是否可以很好地区分样品的表型类别, 并选择最能区分表型类别的基因作为最终选择节点, 迭代重复, 在每个分支上选择带来最大 Gini 系数降低的节点, 直至 Gini 系数降低小于 0.02. 重复上述步骤 10000 次, 共建立 10000 个不同的决策树, 形成与电离辐射相关的 NGF 模型. 决策树中节点分裂导致输出样本基尼系数减少, 因此获得关键调控基因及其重要性得分, 其定义如下:

$$\text{var}(S) = \text{Gini}(S) = 1 - \left(\frac{n_P}{N_S}\right)^2 - \left(\frac{n_N}{N_S}\right)^2, (1)$$

$$I_S(g) = \text{var}(T) - \frac{n_U}{N_T} \text{var}(U) - \frac{n_D}{N_T} \text{var}(D). (2)$$

其中, $\text{var}(T)$ 为由分割节点 g 测试所有样本的方差; $\text{var}(U)$ 为样本子集 U 的方差, 其中分割节点 g 的表达水平超过阈值 (上调); $\text{var}(D)$ 为样本子集 D 的方差, 其中节点 g 的表达水平低于或等于阈值 (下调). I_S 是根据每个基因或相互作用对模型分类准确性的提高程度对其赋予的重要性系数. N_T , n_U 和 n_D 是相应样本集大小. 基尼系数用于计算样本方差, N_S 代表样本集 S 中的总样本数, 其中 n_P 和 n_N 分别表示样本集 S 中正样本和负样本的数量.

1.2.2 五折交叉验证

为了验证 NGF 算法模型的性能, 并评估其是否存在过拟合问题, 本研究采用了 5 折交叉验证 (5-fold cross-validation) 的方法. 五折交叉验证是一种常用的模型评估技术, 通过将数据集分成多个子集, 分别用于模型训练和验证, 以获得更稳健的模型性能评估结果. 具体而言, 本研究将整个数据集随机分成 5 个大小相近的互不重叠的子集 (Folds). 在每一轮迭代中, 选取其中的 4 个子集作为训练集, 剩余 1 个子集作为验证集. 通过这种方式, 每个子集在整个过程中恰好作为一次验证集, 其余 4 次作为训练集, 充分利用了所有数据进行模型评估. 在每次迭代中, 记录模型在验证集上的多种评价指标, 包括准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 得分 (F1 Score). 这些评价指标分别反映模型的整体分类性能、对正类样本预测的准确性、对正类样本的召回能力以及精确率和召回率的综合表现. 最后, 计算所

有迭代的平均值, 作为模型的整体性能评估. 这些平均值可以反映模型在不同子集上的表现是否一致, 从而评估模型的泛化能力及鲁棒性^[15]. 如果训练集与验证集的误差差异显著, 则存在过拟合问题; 如果模型在不同子集上的表现差异很大, 则表明模型可能对数据的变化较为敏感, 鲁棒性不足^[16].

1.2.3 差异表达分析

差异表达分析是比较两个条件下的基因表达差异, 通过一定的统计学方法和手段, 识别出与条件相关的特异性基因, 并进一步分析特异性基因的生物学意义.

使用 R 包 limma 分别筛选 2 h 和 6 h 电离辐射与对照组之间的差异表达基因, 主要使用此包中 lmFit 方法对基因拟合线性模型, 并使用 eBayes 经验贝叶斯方法根据前者拟合结果进行统计推断, 计算矫正 t 统计量、矫正 F 统计量和差异表达标度^[14].

数据筛选标准是 $P < 0.05$ 且 $|\text{FC}| > 0.05$, 筛选出的基因被视为差异表达基因.

1.3 结果分析方法

1.3.1 基因功能富集分析

基因功能富集分析是指在对一组基因进行富集分析后, 可以发现其中有着显著富集的特定生物学通路, 从而揭示其中的生物学分子机制. 本研究中对与人体 B 细胞中 NGF 鉴定出的电离辐射相关重要基因进行功能富集分析, 分析基因表达信息, 发现其中的分子功能和生物学过程, 以更深入地了解细胞应对电离辐射时的响应机制^[17].

为深入分析 2 h 和 6 h 电离辐射后人体 B 细胞相关重要调控基因功能和参与代谢通路, 使用 R 包 clusterProfiler 和基因功能注释工具 DAVID, 根据 KEGG 与 GO 数据库对 NGF 重要性打分前 5% 的基因进行生物学富集分析.

在 RStudio 软件中, 以 $P < 0.05$ 作为富集筛选标准对富集结果进行筛选, 同时在 R 中使用 ggplot2 包对富集分析结果进行可视化处理.

1.3.2 潜在抗辐射小分子药物鉴定

以 Query Signature 文件格式输入 cMap 数据库中 2 h 和 6 h 电离辐射处理后的样本与对照组进行差异表达分析, 得到上调和下调基因, 并与数据库中药物基因表达值进行比较, 进而得到与电离辐射表达谱相似和相反的小分子及药物. 将结果按 P 大小进行排

序, 选取 $P < 0.05$ 作为筛选阈值。

富集打分可以评估基因与小分子药物的相似性, 该富集打分也称为关联度 (Connectivity score)。富集得分大于 0, Connectivity score 为正值代表正相关, 其假设药物扰动表达谱与疾病扰动表达谱正相关, 可以简单地认为药物能够引起或者加剧该疾病状态, 表明该药物小分子对电离辐射具有潜在协同作用, 为正相关药物分子; 而富集得分小于 0, Connectivity score 为负值代表负相关, 其假设药物扰动表达谱与疾病扰动表达谱负相关, 可以简单地认为药物能够减轻甚至逆转该疾病状态, 表明该药物小分子具有潜在拮抗作用, 为负相关药物分子, 该药物小分子或可用作抗辐射损伤药物^[18,19]。

选取 $P < 0.05$ 且富集得分小于 0 的药物小分子作为潜在抗辐射损伤药物。图 1 给出了具体分析流程。

2 实验结果

2.1 NGF 算法鉴定电离辐射相关重要基因

NGF 根据基因节点对分类准确性的贡献进行打分, 在 10 Gy 电离辐射后 2 h, 重要性打分前 100 的基因大多数不属于差异表达基因; 而在 10 Gy 电离辐射后 6 h, 重要性打分前 100 的基因大多数属于差异表达基因, 有 6 个不属于差异表达基因。存在 NGF 打分很高、差异表达不显著的基因, 却表现出与电离辐射响应相关的功能, 例如电离辐射后 6 h, RAS 癌基因家族成员 RAB1 A 没有显著差异表达, 然而其具有较高的 NGF 打分。这种现象与大剂量电

离辐射后细胞癌变相关。通过比较 NGF 打分前 100 的基因和差异表达分析中变化倍数最大的 100 个基因的基因功能富集分析结果, 发现 NGF 方法能够得到更多的电离辐射响应相关基因。例如电离辐射后, p53 信号通路被激活, NGF 方法能够找到 13 个该通路的基因而差异表达分析只能找到 10 个, 并且 NGF 方法找到的基因能够富集到 DNA 修复条目, 而差异表达分析找到的基因则没有富集到该条目, 然而电离辐射会对 DNA 产生损伤, 细胞通过 DNA 修复来抵抗该损伤。通过以上分析可以看出, NGF 方法能够准确对实验组和对照组的基因表达谱进行分类, 捕获到差异表达分析漏掉的部分基因信息。利用 ggplot2 数据包对网络引导随机森林算法中重要性打分前 100 的基因进行可视化处理 (见图 2)。

2.2 电离辐射相关重要基因功能富集分析

在 10 Gy 电离辐射处理后 2 h 和 6 h 的基因表达谱数据的 GO 和 KEGG 功能富集分析中, 经过电离辐射处理后, B 细胞主要通过响应 DNA 损伤来抵抗辐射逆境。具体而言, 细胞周期阻滞和 DNA 修复相关通路被显著激活, 尤其是 p53 信号通路的激活, 这与细胞凋亡和 DNA 修复密切相关。

在 10 Gy 电离辐射后 2 h 的 GO 分析中, 富集的生物学过程包括 DNA 修复、细胞周期调控和凋亡, 这表明细胞在初期尝试通过修复受损的 DNA 来维持生存。在 KEGG 分析中, 富集的通路主要是 p53 信号通路和细胞周期通路, 这些通路在细胞应对 DNA 损伤和凋亡过程中起着重要作用。在 10 Gy 电离辐射后 6 h 的 GO 分析中, 除了继续富集在 DNA 修复和细胞周期调控外, 还出现了与细胞分化和发育相关的通路。这表明随着时间的推移, 电离辐射可能引发更广泛的细胞反应。在 KEGG 分析中, 除了之前提到的 p53 信号通路和细胞周期通路, 还出现了与癌症相关的信号通路, 暗示着细胞可能向癌变方向发展。

为了清晰反映这些结果, 利用 ggplot2 数据包分别对 2 h 和 6 h 电离辐射后 GO 和 KEGG 富集结果进行可视化处理 (见图 3 和图 4)。在图 3 和图 4 中富集条目越多, 说明该时间点下与电离辐射响应相关的基因越多。这些富集条目通过颜色和大小来表示显著性和富集度, 越显著的条目颜色越深, 富集度越高的条目尺寸越大。具体来说, 图 3 给出了 2 h 后富集的 GO 条目和 KEGG 通路, 而图 4 则给出了 6 h 后的结果。

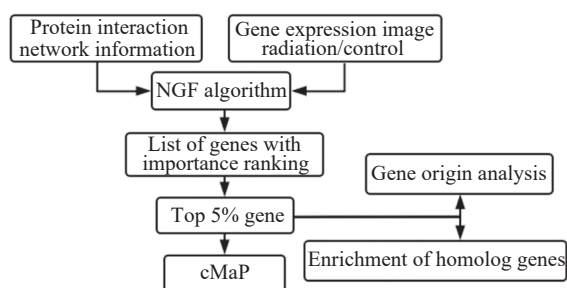


图 1 基于差异表达分析和 NGF 算法研究人 B 细胞对电离辐射响应机制的方法流程

Fig. 1 Flow chart of the method to study the response mechanism of human B cells to ionizing radiation based on differential expression analysis and network-guided random forest algorithm

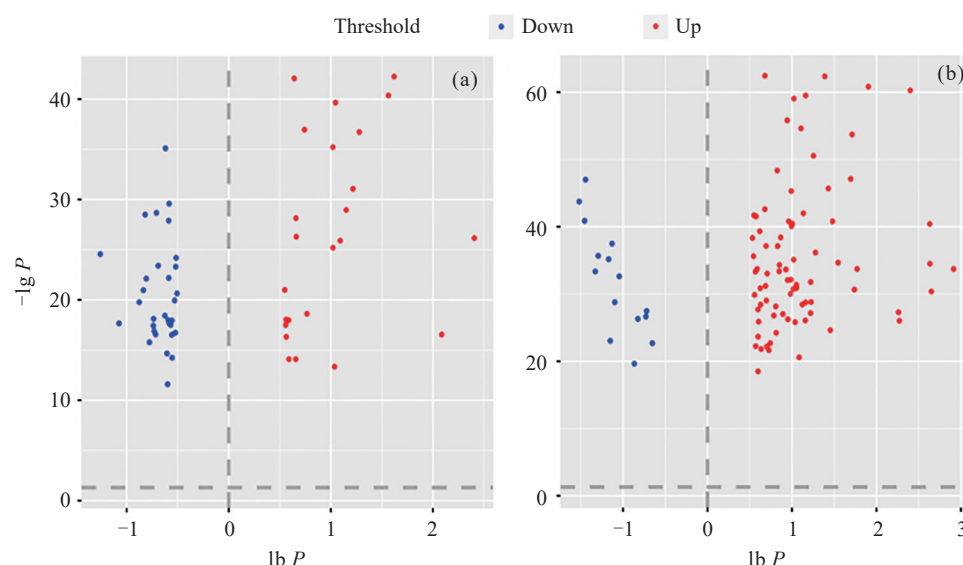


图 2 10 Gy 电离辐射后 2 h (a) 和 6 h (b) NGF 重要性打分前 100 的基因差异可视化处理
Fig. 2 Volcano plot of gene differential expression in the top 100 NGF importance scores
2 h (a) and 6 h (b) after 10 Gy ionizing radiation

根据分析结果, 从 10 Gy 电离辐射后 2 h 到 6 h 的响应基因功能变化来看, 人类 B 细胞在大剂量电离辐射后难以正常修复 DNA 损伤, 大量细胞发生凋亡和癌变. Ao 等^[20]提出的内源性网络理论可以反映复杂疾病背后控制各种表型的核心调控网络. 将本研究得到的人类 B 细胞电离辐射响应基因与 Ao 等构建的白血病内源性网络中核心调控基因进行比较, 发现 10 Gy 电离辐射后 6 h 细胞状态和白血病细胞相差较大, 表明 6 h 细胞处于向癌细胞转化过程中, 仅有部分癌症相关基因被激活或抑制, 因此通过干预表达尚未显著改变的基因可以抑制细胞进一步向癌细胞转化, 例如激活细胞周期负调节因子 RB1 基因表达.

2.3 电离辐射相关重要基因进化分析

靶标发现是药物研发的关键步骤之一, 将进化生物学应用到医学领域, 分析成药靶标进化特征, 评价在研靶标成靶潜力, 将为药物研发提供一个新准则, 且有助于解决药物耐药性问题. 研究表明, 早起源基因成靶潜力更大, 且靶标治疗与其进化历史相关. 使用 Liebeskind 等研究中得到的基因起源数据, 将 NGF 得到的重要性打分前 5%(10 Gy 电离辐射后 6 h) 的基因视为电离辐射相关重要基因, 并对以上基因进行基因起源分析. 结果表明, 在细胞生物共同祖先、真核生物和古菌共同祖先以及真核细胞阶段中,

显著富集与电离辐射相关的早期起源基因, 证明与电离辐射相关早期起源基因具有更高成靶潜力^[20].

全基因组倍增 (Whole Genome Duplication, WGD) 是脊椎动物进化史上的重大事件, 人类祖先基因组经历过两次 WGD, 人类基因组中约有 30% 的蛋白编码基因与 WGD 有关, 这些与 WGD 有关的基因被称为 Ohnologs. 研究发现 Ohnologs 对生物体发育和转录调控起着重要作用, 并且 Ohnologs 具有剂量敏感性, 与疾病密切相关, 该特性使得 Ohnologs 作为药物靶标的可能性更高. 有研究分析成功药物靶标在 Ohnolog 和 Non-Ohnolog 基因中的百分比, 发现药物靶标显著富集在 Ohnolog 基因中^[21].

对电离辐射响应重要基因进行基因起源和 Ohnolog 基因富集分析, 使用 *R* 中超几何分布 *phyper* 函数计算 *P*, 发现以上基因起源较早, 主要富集在 Cellular_organisms, Euk+Bac 和 Eukaryota 三个阶段. 对 NGF 鉴定的电离辐射相关重要基因进行 Ohnolog 基因富集分析, 结果显示基因显著富集在 Ohnolog 基因, 进一步证明了这些基因的重要性和成靶潜力 (见表 1 和图 5).

2.4 抗辐射药物筛选

为减少辐射对人体不可逆的伤害, 有效的抗辐射损伤药物研发非常重要. 目前有研究发现一些药物分子可以帮助减轻辐射对人体的伤害, 包括细胞因子类

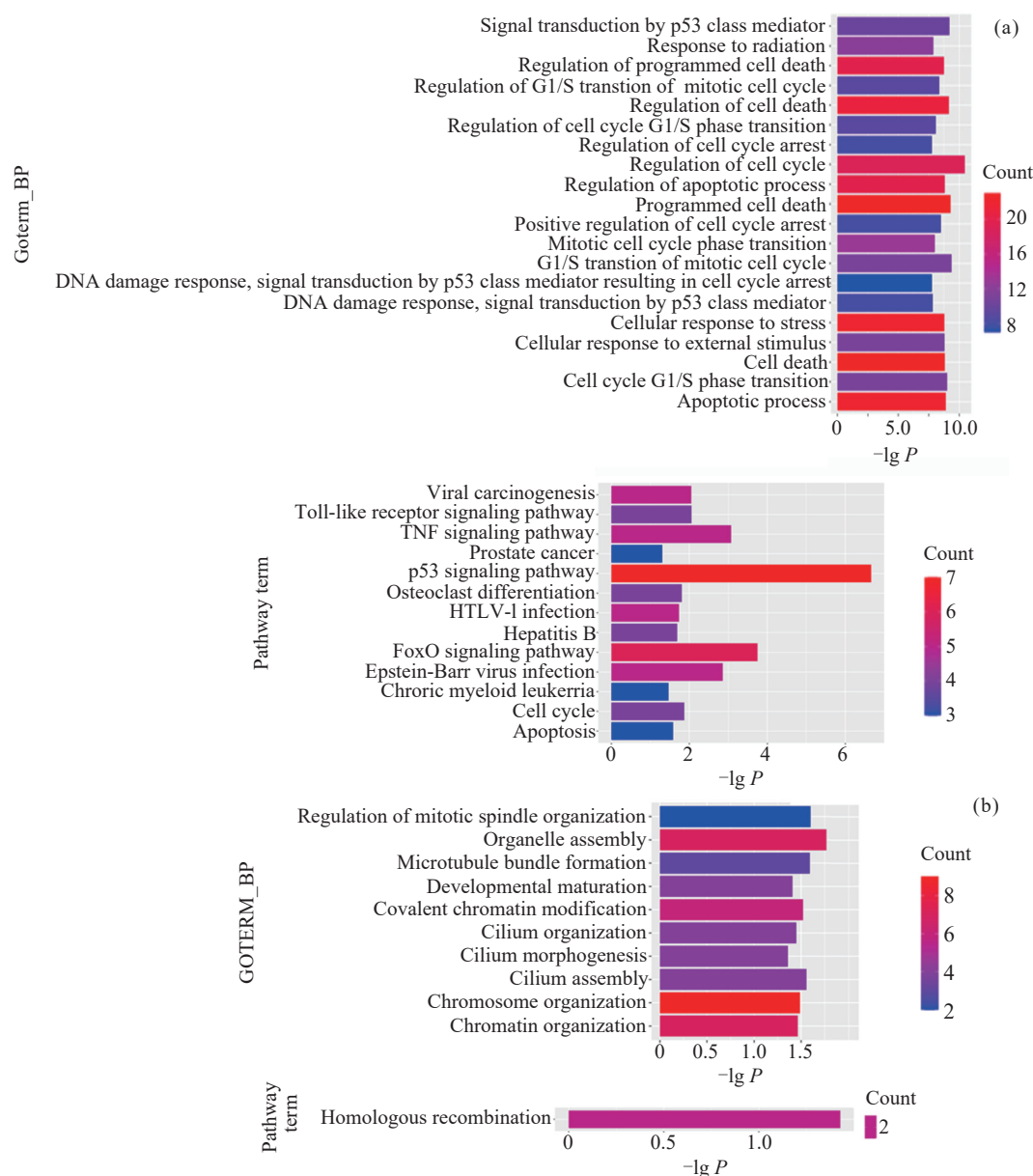


图3 10 Gy 电离辐射 2 h 后的 GO 分析 (a) 和 KEGG 分析 (b)

Fig. 3 GO (a) and KEGG (b) analysis after 2 hours of 10 Gy ionizing radiation

药物、激素类药物和天然药物等。已知的辐射防护剂虽多,然而很多具有较强的副作用,这些副作用将影响药物大范围使用,因此发现更有效的副作用更小的天然抗辐射损伤药物十分重要。

为鉴定抗辐射损伤药物,使用 cMap 数据库上传电离辐射相关重要基因中的下调和上调基因列表,与 cMap 数据库中 1309 个药物小分子相关基因表达谱进行比较,得到相似或相反作用的药物小分子,最后选择 P -value 较小的 6 种药物小分子作为潜在抗辐射损伤药物,分别是紫杉醇、阿苯达唑、利美索龙、氯倍

他索、普萘洛尔、醉椒素,其中紫杉醇和醉椒素等为天然产物(见表 2)。

紫杉醇是众所周知的具有独特作用机理的抗癌药物,被认为是最成功的天然抗癌药物之一。其是从紫杉树中提取的,可以有效抑制细胞周期进程,防止有丝分裂并抑制癌细胞的生长。紫杉醇可用于治疗卵巢癌、乳腺癌、肺癌及卡波西氏肉瘤等,可以提供给空间工作人员日常摄入,帮助其抵抗电离辐射损伤^[22]。

醉椒素是从卡瓦胡椒根部提取的有效成分之一,具有抵抗炎症和抗氧化作用。醉椒素在制备治疗骨质



图 4 10 Gy 电离辐射 6 h 后的 GO 分析 (a) 和 KEGG 分析 (b)

Fig. 4 GO (a) and KEGG (b) analysis after 6 hours of 10 Gy ionizing radiation

疏松药物方面有应用, 通过抑制钙介导的 NFATc1 和 MAPK 信号通路抑制破骨和骨吸收, 钙信号是另一种调节破骨细胞发生的关键途径^[23].

此外, 通过筛选富集系数为正的 药物, 可以鉴别多个潜在辐射增敏剂, 这些药物大部分已被证实属于抗癌药物, 这些抗癌药物可以改变肿瘤细胞对放射的反应性, 显著提高癌症的控制率和治愈率. 例如白藜芦醇, 这是目前已被申请用于人体临床试验的辐射增敏剂, 可以显著提升癌症放射性治疗疗效, 白藜芦醇对实验鼠肝细胞癌和白血病等多种肿瘤细胞均有显著抑制作用 (见表 3).

3 分析与讨论

综上所述, 在电离辐射对人类细胞影响的生物信息学分析方面取得了多项重要结果.

将网络引导随机森林 (NGF) 算法应用于电离辐射响应机制研究. NGF 算法结合了基因表达谱与蛋

白相互作用网络, 能够识别出在传统差异表达分析中被忽略的重要基因. 尽管这些基因在表达水平上变化不显著, 但是在基因调控网络中发挥了关键作用, 从而揭示了更为复杂和全面的生物学过程. 通过基因功能富集分析, 详细揭示了电离辐射对 B 细胞的多层次影响, 包括 DNA 损伤修复障碍、细胞凋亡和癌变机制. 这些发现为理解电离辐射的生物学效应提供了新的视角, 有助于开发新的防护和治疗策略.

基于 cMap 数据库的药物筛选, 鉴定了几种可能具有抗辐射作用的药物, 例如紫杉醇和醉椒素. 通过基因功能分析, 这些药物显示出具有帮助细胞抵抗电离辐射损伤的潜力, 为未来开发抗辐射药物提供了科学依据.

尽管取得一些重要进展, 但是仍存在不足之处. 本研究使用的数据集主要来自基因表达综合数据库 (GEO) 和 cMap 数据库. 由于样本数量有限且实验条件相对单一, 可能无法全面反映实际环境空间中电离

表 1 电离辐射响应重要基因起源分析
Table 1 Analysis of the origin of important genes in response to ionizing radiation

基因起源阶段	基因数	P-value
Cellular_organisms	35	0.0424
Euk_Archaea	8	0.308
Euk+Bac	60	0.0101
Eukaryota	229	0.000000819
Eumetazoa	129	0.0949
Mammalia	30	1.00
Opisthokonta	30	0.0716
Vertebrata	48	1.00

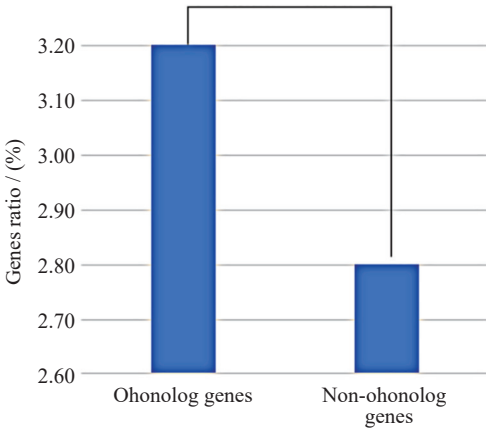


图 5 Ohnolog 基因富集分析
Fig. 5 Ohnolog gene enrichment analysis

表 2 基于 cMap 数据库筛选的抗辐射药物
Table 2 Anti-radiation drugs screened based on cMap database

药物名称	P-value	药物疗效
紫杉醇	0.00177	通过抑制微管网络的形成, 抑制肿瘤细胞增殖分裂
阿苯达唑	0.00004	通过与蠕虫β-微管蛋白结合, 抑制其聚合或组装成微管
利美索龙	0.00275	糖皮质激素受体激动剂治疗眼部炎症
氯倍他索	0.00559	治疗硬皮病
普萘洛尔	0.00641	肾上腺素受体拮抗剂, 治疗心律失常
醉椒素	0.00703	抗惊厥, 治疗肌肉松弛

表 3 基于 cMap 数据库筛选的辐射增敏剂
Table 3 Radiation sensitizers screened based on cMap database

药物名称	P-value	药物疗效
莫诺苯宗	0	治疗黑色素沉积
盐酸苯氧苄胺	0	治疗高血压和嗜铬细胞瘤
依米丁	0	抑制蛋白质合成
海恩酮	0	DNA合成抑制剂, 抗血吸虫
依托泊苷	0	细胞周期特异性抗肿瘤药物, 作用于DNA拓扑异构酶 II
三氟胸苷	0	可在DNA复制过程中使用, 阻止碱基对形成
白藜芦醇	0	抗氧化、抗炎、抗癌及心血管保护等
伊利替康	0.00008	由喜树碱制成, 可导致DNA损伤
甲氨蝶呤	0.00012	抗叶酸类抗肿瘤药
毒胡萝卜素	0.00172	诱导内质网应激反应性细胞凋亡
洛莫司汀	0.00422	烷化剂类抗肿瘤药, 使DNA链断裂, 可通过血脑屏障

辐射对人体细胞的影响. 未来的研究应收集更多样本和更广泛的实验数据, 以提高研究的代表性和结果的可靠性.

此外, NGF 算法在基因节点打分过程中涉及复杂的计算和模型训练, 虽然提高了分类准确性和生物学解释性, 但是也增加了计算成本和时间. 模型参数的选择和优化对最终结果有较大影响, 需要进一步研究和验证. 另外研究主要关注 B 细胞对电离辐射的响应机制, 而忽略了其他类型细胞的响应. 不同类型细胞可能对电离辐射有不同的响应机制. 未来研究应扩展到其他类型细胞, 以更全面地了解电离辐射的生物学效应.

未来的研究应收集和分析更多的基因表达数据, 尤其是来自不同个体和不同实验条件的数据. 这将有助于提高模型的泛化能力, 减少数据偏差带来的影响, 从而提供更可靠的研究结论. 此外, 应拓展研究范围, 涵盖其他类型的免疫细胞 (例如 T 细胞、NK 细胞) 以及非免疫细胞, 全面揭示电离辐射对人体各类细胞的影响机制. 这将有助于制定更为有效的防护措施和治疗方案, 保障航天员的健康.

除了 NGF 算法外, 还应考虑结合其他生物学方法 (例如机器学习、深度学习等), 提高模型的准确性和预测能力. 通过多种方法的结合, 可以更全面地解析电离辐射的生物学效应, 提供更加多维度的研究视角. 对于筛选出的潜在抗辐射药物, 应进行进一步的实验验证. 通过体内和体外实验, 确认这些药物

的有效性和安全性, 为实际应用提供可靠的依据, 这将有助于将生物信息学研究成果转化为实际的医学应用. 通过这些拓展研究, 进一步深入理解电离辐射对人类细胞的影响机制, 为未来空间探索中的人体健康保障提供理论和实践支持.

参考文献

[1] REYNOLDS R J, BUKHTIYAROV I V, TIKHONOVA G I, *et al.* Contrapositive logic suggests space radiation not having a strong impact on mortality of US astronauts and Soviet and Russian cosmonauts[J]. *Scientific Reports*, 2019, **9**(1): 8583

[2] NILSSON R, LIU N A. Nuclear DNA damages generated by reactive oxygen molecules (ROS) under oxidative stress and their relevance to human cancers, including ionizing radiation-induced neoplasia part I: physical, chemical and molecular biology aspects[J]. *Radiation Medicine and Protection*, 2020, **1**(3): 140-152

[3] ZHU M Q, YANG M D, ZHANG J J, *et al.* Immunogenic cell death induction by ionizing radiation[J]. *Frontiers in Immunology*, 2021, **12**: 705361

[4] HOLMBERG O, PINAK M. How often does it happen? A review of unintended, unnecessary and unavoidable high-dose radiation exposures[J]. *Journal of Radiological Protection*, 2021, **41**(4): R189-R201

[5] LUMNICZKY K, IMPENS N, ARMENGOL G, *et al.* Low dose ionizing radiation effects on the immune system[J]. *Environment International*, 2021, **149**: 106212

[6] SPIELMANN G, AGHA N, KUNZ H, *et al.* B cell homeostasis is maintained during long-duration spaceflight[J]. *Journal of Applied Physiology*, 2019, **126**(2): 469-476

[7] PAUL A M, OVERBEY E G, DA SILVEIRA W A, *et al.*

- Immunological and hematological outcomes following protracted low dose/low dose rate ionizing radiation and simulated microgravity[J]. *Scientific Reports*, 2021, **11**(1): 11452
- [8] GHOSH S, GHOSH A. Activation of DNA damage response signaling in mammalian cells by ionizing radiation[J]. *Free Radical Research*, 2021, **55**(8): 814-827
- [9] ALHARBI F, VAKANSKI A. Machine learning methods for cancer classification using gene expression data: A review[J]. *Bioengineering*, 2023, **10**(2): 173
- [10] LIU D F, ZHANG X H, ZHENG T, *et al.* Optimisation and evaluation of the random forest model in the efficacy prediction of chemoradiotherapy for advanced cervical cancer based on radiomics signature from high-resolution T2 weighted images[J]. *Archives of Gynecology and Obstetrics*, 2021, **303**(3): 811-820
- [11] SIGALOVA O M, SHAEIRI A, FORNERIS M, *et al.* Predictive features of gene expression variation reveal mechanistic link with differential expression[J]. *Molecular Systems Biology*, 2020, **16**(8): e9539
- [12] BRUGALETTA R, SANTORO A, ALISI A, *et al.* 1283 Gene expression analysis of blood cells in radiation health care workers occupationally exposed to ionising radiation[J]. *Occupational & Environmental Medicine*, 2018, **75**(S2): A428
- [13] SZKLARCZYK D, GABLE A L, NASTOU K C, *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. *Nucleic Acids Research*, 2021, **49**(D1): D605-D612
- [14] LUO Y B, SUN F L, PENG X W, *et al.* Integrated bioinformatics analysis to identify abnormal methylated differentially expressed genes for predicting prognosis of human colon cancer[J]. *International Journal of General Medicine*, 2021, **14**: 4745-4756
- [15] DEIST T M, DANKERS F J W M, VALDES G, *et al.* Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers[J]. *Medical Physics*, 2018, **45**(7): 3449-3459
- [16] HONG S, LEE S, LEE J, *et al.* Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: model development and retrospective clinical validation study[J]. *JMIR Medical Informatics*, 2020, **8**(8): e15932
- [17] GUPTA M, LIU X F, TERAOKA S N, *et al.* Genes affecting ionizing radiation survival identified through combined exome sequencing and functional screening[J]. *Human Mutation*, 2021, **42**(9): 1124-1138
- [18] DERLIN T, BOGDANOVA N, OHLENDORF F, *et al.* Assessment of γ -H2AX and 53BP1 foci in peripheral blood lymphocytes to predict subclinical hematotoxicity and response in somatostatin receptor-targeted radionuclide therapy for advanced gastroenteropancreatic neuroendocrine tumors[J]. *Cancers*, 2021, **13**(7): 1516
- [19] JENSEN M K, PERS T H, DWORZYNSKI P, *et al.* Protein interaction-based genome-wide analysis of incident coronary heart disease[J]. *Circulation: Cardiovascular Genetics*, 2011, **4**(5): 549-556
- [20] SU H, WANG G, ZHU X, XU M, AO P. Endogenous molecular-cellular network theory: A system-biomedical perspective towards complex diseases[J]. *Chinese Journal of Nature*, 2015, **37**(6): 448-454.
- [21] WU Y D, BYRNE E M, ZHENG Z L, *et al.* Genome-wide association study of medication-use and associated disease in the UK Biobank[J]. *Nature Communications*, 2019, **10**(1): 1891
- [22] SHARIFI-RAD J, QUISPE C, PATRA J K, *et al.* Paclitaxel: application in modern oncology and nanomedicine-based cancer therapy[J]. *Oxidative Medicine and Cellular Longevity*, 2021, **2021**: 3687700
- [23] LIU X N, LI Z, LIU H, *et al.* Low concentration flufenamic acid enhances osteogenic differentiation of mesenchymal stem cells and suppresses bone loss by inhibition of the NF- κ B signaling pathway[J]. *Stem Cell Research & Therapy*, 2019, **10**(1): 213

作者简介



何敏敏(共同第一作者) 女, 1998年12月出生于山西省运城市, 毕业于华中农业大学信息学院生物信息系。主要研究方向为空间疾病致病机制解析、药物发现等。

E-mail: 1173505363@qq.com



朱玉锋(共同第一作者) 男, 2001年11月出生于江苏省泰州市, 毕业于华中农业大学信息学院生物信息系。主要研究方向为空间生物医学、药物筛选等。

E-mail: neverbo1101@outlook.com



全源(通信作者) 女, 1990年7月出生于广西壮族自治区柳州市, 现为华中农业大学信息学院副研究员, 博士生导师, 主要研究方向为空间生物信息学、空间医学、药物发现等。

E-mail: quanyuan@mail.hzau.edu.cn