本科毕业论文

论文题目：　　物种保守环形 RNA 的泛癌分析

姓　　名：　刘铠铭　　学　　号：　20307110190

院　　系：　生命科学学院

专　　业：　生态学

指导教师：　杨力　　职　　称：　研究员

单　　位：　复旦大学生物医学研究院

完成日期：　2024 年 5 月 21 日

物种保守环形 RNA 的泛癌分析

完成人

刘铠铭

指导小组成员

杨力　研究员

目　录

摘　要

外显子反向剪接来源的环形 RNA（circular RNA，circRNA）是一类在多个物种中广泛存在的非编码 RNA。近年来，随着对 circRNA 功能机制研究的深入，其在癌症发生发展中的功能也逐渐得到关注。保守性是预测 RNA 功能的重要指标，保守性 circRNA 往往可能存在重要功能。本课题关注了物种保守 circRNA 在多类型癌症中的表达特征及其潜在功能，针对人类和小鼠、大鼠、斑马鱼和线虫四种模式生物鉴定了物种保守 circRNA。进一步地，本课题基于 RNA-seq 数据使用 CIRCexplorer3/CLEAR 工具对 circRNA 在宫颈癌、结肠癌、前列腺癌常用细胞系中的表达进行定量分析，发现在上述细胞系中物种保守 circRNA 的表达水平显著高于非保守 circRNA，揭示了物种保守 circRNA 在癌症发展中的潜在功能。此外，本课题通过 MiOncoCirc 数据库中的患者癌症组织和正常组织样本数据，进行差异表达分析和功能注释，识别了与癌症密切相关的差异表达 circRNA 及其保守性；结合 GO 和 KEGG 通路富集分析，对差异表达 circRNA 的潜在功能进行了探索。本课题在癌症细胞系及组织水平上系统构建了物种保守 circRNA 的泛癌表达谱，初步解析了物种保守 circRNA 与癌症发生发展的关联，为 circRNA 在癌症诊断和治疗中的应用提供了理论基础，也为未来的 circRNA 功能研究提供了新的方向和思路。

关键词：环形 RNA，物种保守性，泛癌，表达谱

Abstract

Circular RNAs derived from exon back-splicing (circRNAs) are a class of non-coding RNAs, widely spread in multiple species. In recent years, as research on the functional mechanisms of circRNAs has deepened, their role in cancer development has gradually garnered researchers' attention. RNA conservation is often associated with important functions, suggesting that conserved circRNAs might possess significant roles. Our study focuses on the expression characteristics and potential functions of conserved circRNAs in various cancers, identifying conserved circRNAs across human and four model organisms: mouse, rat, zebrafish, and nematodes. Furthermore, using RNA-seq data and the CIRCexplorer3/CLEAR tools, our study quantitatively analyzes the expression of circRNAs in commonly used cell lines of cervical cancer, colon cancer, and prostate cancer. We found that the expression levels of conserved circRNAs in these cell lines were significantly higher than those of non-conserved circRNAs, indicating their potential roles in cancer development. Additionally, through differential expression analysis of data from the MiOncoCirc database, which includes cancer tissues and normal tissues from patients, circRNAs related to cancer were identified. Through GO and KEGG pathway enrichment analyses, we explored the potential functions of these differentially expressed circRNAs. Our study systematically constructs a pan-cancer expression profile of conserved circRNAs at the levels of cancer cell lines and tissues, initially elucidating the association between conserved circRNAs and cancer development. It provides a theoretical foundation for the application of circRNAs in cancer diagnosis and treatment, and offers new directions and insights for future research on circRNA functions.

Key words： circular RNA, species conservation, pan-cancer, expression profile

一、前　言

1.1 研究背景

越来越多的研究表明，非编码 RNA（non-coding RNA，ncRNA）在基因表达调控中发挥了广泛而重要的作用。根据其结构特征、功能机制和空间分布等，非编码 RNA 可以被划分为多种类型，包括核糖体 RNA（ribosomal RNA, rRNA）、转运 RNA（transfer RNA, tRNA）、长非编码 RNA（long non-coding RNA，lncRNA）、小核 RNA（small nuclear RNA，snRNA）、微 RNA（micro RNA，miRNA）、小干扰 RNA（small interfering RNA，siRNA）和环形 RNA（circular RNA）等[1]。近年来，研究人员广泛揭示了不同类型的非编码 RNA 在多种癌症（如结肠癌、肺癌和乳腺癌等）中的作用及机制[2]。例如，长非编码 RNA *SEMA3B-AS1* 通过 miR-3940 海绵抑制 KLLN 降解从而抑制三阴性乳腺癌（triple-negative breast cancer，TNBC）的发展[3]；miRNA miR-193a 靶向 *Caprin1* 蛋白抑制结肠癌细胞的增殖[4]；环形 RNA *circNDUFB2* 通过充当支架增强 TRIM25 和 癌症发展正向调节剂 IGF2BPs 之间的相互作用从而抑制非小细胞肺癌（non-small cell lung cancer，NSCLC）细胞的增殖[5]。这些发现提示着非编码 RNA 与癌症的发生和发展密切相关,凸显了非编码 RNA 在癌症治疗等领域的研究潜力。

环形 RNA（circular RNA）是一类具有特殊共价闭合环状结构的非编码 RNA[6]。在真核生物中，根据其生成方式可分为外显子反向剪接来源的 circRNA 及内含子套索逃脱脱分支作用形成的 ciRNA（circular intronic RNA）两大类[7]。虽然在上个世纪 70 年代，Sanger 等人在类病毒中发现了共价闭合的单链环状 RNA 分子[8]，但是直到上个世纪 90 年代 Cocquerelle 等[9]和 Capel 等[10]才发现了真核生物内源的环形 RNA 分子[7, 11]。进入 21 世纪后，尤其是随着 2010 年以来 RNA-seq 测序技术的进步以及计算分析流程的发展，大量的 circRNA 被发现和报道[12]，目前已发现的 circRNA 多达数十万种[13]并广泛存在于自然界多个物种及组织中，其功能及机制研究逐渐成为领域内的研究热点。

近年来的研究表明，circRNA 在分子水平上可以通过多种作用机制在生理和病理条件下发挥重要作用[14]。首先，一些 circRNA 可以作为 miRNA 海绵，影响

miRNA 与 mRNA 之间的相互作用，从而参与转录后基因调控，例如：*circHIPK3* 能够分别与肿瘤抑制性的 miR-124、miR - 193、miR-379 和 miR-654 等多种 miRNA 结合，从而影响一系列抑制癌症细胞增殖的基因，如 IL6R 和 DLX2 等[15]。同时，某些 circRNA 还可以与一些 RNA 结合蛋白相互作用形成 circRNP 复合体发挥调控作用，例如 *circFOXO3* 可以与抗衰老蛋白 ID-1 和抗应激蛋白 FAK、HIF1α 相互作用调节上述蛋白质的易位[16]。部分 circRNA 还可以与蛋白质和 mRNA 形成三元复合体，调节 mRNA 的稳定性及翻译。例如：*circNSUN2*-IGF2BP2-HMGA2 复合体促进了 IGF2BP2 与 HMGA2 的相互作用，从而增强 HMGA2 mRNA 的稳定性[17]；*circYAP* 与 YAP mRNA 以及 PABP 和 eIF4G 两个翻译起始蛋白形成的复合体可以阻断 YAP 蛋白的翻译起始[18]等。此外，某些 circRNA 可以翻译形成蛋白质发挥调控功能，例如：*circFBXW7* 编码蛋白质 FBXW-185aa，后者通过与 USP28 相互作用促进 c-Myc 蛋白质的降解，进而抑制了胶质瘤细胞的增殖[19]。最新的研究表明，一些 circRNA 可以编码形成肿瘤特异性抗原肽进肿瘤免疫的进行，如 *circFAM53B* 在乳腺癌中编码与 HLA-1 结合的隐性抗原肽，可以有效地引发新生 CD4+和 CD8+ T 细胞并诱导抗肿瘤免疫[20]。随着针对 circRNA 在癌症中功能及调控研究的深入，其在癌症预测、诊断和靶向药物治疗等方面也显示出了突出的应用前景，因此，针对癌症细胞或组织进行 circRNA 的功能分析显得尤为重要。

目前，针对全转录组 circRNA 的鉴定、识别和定量分析往往要依赖于实验生物学和计算生物学的方法相结合。在实验方面，poly(A)– RNA-seq 和 ribo– RNA-seq 是目前较常使用的 circRNA 鉴定方法[21]；此外，RNase R 处理后的 RNA-seq 技术可以特异性地富集 circRNA[22]，因此也被广泛应用于 circRNA 相关研究；随着测序技术的发展，Oxford Nanopore 等长读长测序平台也逐步应用于 circRNA 的研究[23]。在计算方面，基于上述 RNA-seq 方法获取的高通量测序数据，可以通过检测 circRNA 反向剪接位点（back-splicing junction, BSJ）的读序数量来完成 circRNA 的鉴定和定量分析，常用的计算方法包括 CIRCexplorer3/CLEAR[24]、DCC[25]和 MapSplice[26]等。基于上述实验及计算方法，目前已经建立了许多 circRNA 数据库，如 CIRCpedia[27]、CircAtlas[28]和 MiOncoCirc[29]等，其中记载了

多个物种、细胞系的 circRNA 的基因组位置及表达信息，这为本课题的后续分析提供了一定的数据基础。

尽管已有部分研究在胃癌[30]、前列腺癌[31]等癌症中构建了全转录组水平的 circRNA 表达图谱，但面对海量的多类型、多来源 circRNA 数据，如何有效地整合并进行泛癌分析，以深入探索 circRNA 的功能与作用机制是目前领域内亟待解决的问题。当前，已有针对部分 circRNA 表达与功能的泛癌分析研究证实了其在多种癌症中的重要功能，如 *circCDR1as* 在 CD8+ T 细胞、活化 NK 细胞、M2 巨噬细胞、癌相关成纤维细胞和内皮细胞等多个肿瘤细胞的微环境改变中发挥媒介作用[32]等，提示了寻找普适性调控癌症发生发展的 circRNA 的潜在价值。因此，利用泛癌分析方法解析更多种类的 circRNA 功能也成为揭示 circRNA 与癌症关联的重要研究方向。

通常来说，保守性是识别功能性 RNA 的重要经验指标，许多具有较强的物种保守性的 RNA 往往在物种间可能存在着重要且普适性的生物学功能[33, 34]，例如参与翻译过程（rRNA[35]和 tRNA[36]）、RNA 降解（RNase P[37]等）和蛋白质定位（SRP[36]）等。类似地，在真核生物中也存在许多具有物种保守性的 circRNA，这提示其可能具有重要的生物学功能，因此具有重要的研究价值。

针对物种保守 circRNA 的研究已经有一定的基础。Peter L. Wang 等人整合并刻画了 circRNA 在部分真核生物中的表达情况，发现 circRNA 在包括后生动物、原生动物、植物和真菌中均有表达[38]。针对脊椎动物，Zhao 等人构建了包含六个物种的 circRNA 数据库——CircAtlas，并使用 MCS 方法进行了 circRNA 的保守性分析，初步筛选了脊椎动物中具有保守性的 circRNA[28]。然而，这些研究主要集中在物种保守 circRNA 的筛选和保守性评估上，对 circRNA 在不同生理及病理条件下，尤其是不同肿瘤类型中的表达及功能研究仍然较为匮乏。此外，模式生物是现代生物科学研究中的重要研究对象，已有的关于遗传、发育、生理学以及潜在的细胞和分子层次生物过程的大部分知识都来自对模式生物的研究[39]，针对涵盖多种模式生物的物种保守性 circRNA 研究能够为 circRNA 的功能探索提供支持，具有重要的研究意义。

1.2 主要研究内容

本课题主要针对外显子反向剪接来源的 circRNA 开展研究，以模式生物中具有物种保守性的 circRNA 为研究对象进行泛癌分析，以期发现与探索新的与癌症密切相关的 circRNA 分子并初步了解其作用机制。

首先，本课题基于 CIRCpedia v2 数据库进行了模式生物中物种保守 circRNA 的筛选，获取了在人类与小鼠、大鼠、斑马鱼和线虫等多种模式生物中保守的 circRNA 数据，并在这一基础上分析归纳了物种保守 circRNA 的来源及序列特征，为后续的功能分析提供了数据基础。此外，为了构建物种保守 circRNA 在多种癌症细胞系中的表达谱，本课题选取了来自宫颈癌细胞系 HeLa，结肠癌细胞系 HT29，前列腺癌细胞系 42D、PC3 和 V16A 的 RNA-seq 测序数据，使用 CIRCexplorer3/CLEAR 工具对其中物种保守 circRNA 的表达进行了定量分析，并筛选出高表达的物种保守 circRNA 进行功能分析，以探索物种保守 circRNA 在多类型癌症中的作用机制。进一步地，为了探索物种保守 circRNA 与在真实癌症组织中的功能调控，本课题系统分析了 MiOncoCirc 数据库提供的患者癌症组织和正常组织样本的 circRNA 表达数据，鉴定出了一系列差异表达的 circRNA，并分析了其表达特征和保守性特征。最后，本课题通过基因富集分析对上述差异表达 circRNA 的潜在功能进行了探索，并着重对其中具有物种保守性的 circRNA 进行了功能分析。

综上，本课题在细胞系及组织水平上系统构建了物种保守 circRNA 的泛癌表达谱，初步解析了物种保守 circRNA 与癌症发生发展的关联，有助于更好地理解 circRNA 在癌症发生发展的功能调控。本课题鉴定到的与癌症相关的物种保守 circRNA 分子，不仅为后续构建 circRNA 与癌症关联研究的模式生物提供了一定的参考，也为未来 circRNA 用于癌症诊断和治疗的应用提供了新的思路。

二、材料与方法

2.1 材料

2.1.1 物种保守 circRNA 筛选所使用的 circRNA 数据

表 1 circRNA 物种保守性分析所使用的数据概览

物种	样本数量（ribo–）	circRNA 数量
Human	54	112394
Mouse	51	53448
Rat	6	10197
Zebrafish	1	87
Worm	6	113

注：circRNA 数量统计已进行去重，对来自不同组织的同一 circRNA 进行了合并。

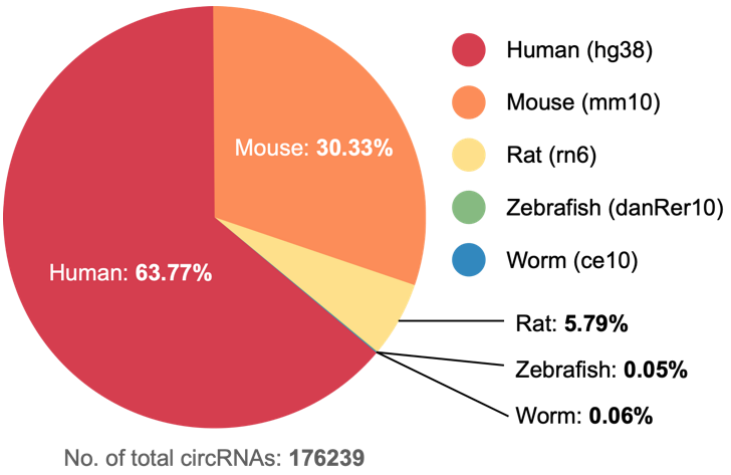

图 1 物种保守性分析中 circRNA 的物种来源分布

图中所示百分比代表该物种来源的 circRNA 在物种保守性分析中总 circRNA 的占比，图中不同颜色对应不同物种：红色，人类；橙色，小鼠；黄色，大鼠；绿色，斑马鱼；蓝色，线虫。

本课题选取了人类及包括小鼠、大鼠、斑马鱼和线虫在内的四种常见遗传学模式生物，从 CIRCpedia v2 数据库中获取了上述物种的 circRNA 表达及基因组位置信息。CIRCpedia v2 数据库收集了来自 Gene Expression Omnibus（GEO）、Encyclopedia of DNA Elements（ENCODE）项目和 EMBL-European Bioinformatics Institute（EMBL-EBI）数据库中的 RNA-seq 数据集，共 185 个，其中包含有 ribo−、poly(A)−、ribo−/RNase R 和 poly(A)−/RNase R 等 RNA-seq 方法获得的数据集[27]。

为了保证环形 RNA 数据的准确性以及后续对其亲本线性 RNA 表达的分析，本课题仅采用了来自 118 个 ribo– RNA-seq 样本中的 circRNA 数据，共 176,239 个 circRNA（表 1）。研究数据中绝大多数 circRNA（99.89%）来自人类（63.77%）、小鼠（30.33%）和大鼠（5.79%），而极少数来自斑马鱼和线虫（图 1）。

2.1.2 多类型癌症细胞系 RNA-seq 数据

本课题从 NCBI GEO 数据库中获取了包括宫颈癌、结肠癌和前列腺癌三种癌症的常用细胞系的 RNA-seq 数据，分别来源于 HeLa（宫颈癌）、HT29（结肠癌）细胞系和 42D、PC3 和 V16A（前列腺癌）细胞系，共计来自七个样本的 RNA-seq 数据（表 2）。其中，本课题选取的 HeLa 细胞系和 HT29 细胞系的 RNA-seq 数据来源于 GEO: GSE149641 数据集，该数据集中提取了 HeLa 细胞和 HT29 细胞的 total RNA，并使用 RiboMinus 试剂盒去除了核糖体 RNA 用以制备 ribo– RNA-seq 库，使用 Illumina Hiseq X ten 平台进行深度测序[40]；本课题选取的 42D、PC3 和 V16A 细胞系来源于 GEO: GSE113120 数据集，该数据集提取了 42D、PC3 和 V16A 等前列腺癌细胞的 total RNA，同样使用 RiboMinus 试剂盒进行核糖体 RNA 的去除并制备了 ribo– RNA-seq 文库，使用 Illumina Hiseq 2000 平台进行了深度测序[41]。本课题选取的上述数据均为 ribo– RNA-seq 数据，适用于 circRNA 及亲本线性 RNA 的表达分析，同时各数据的总读序数量均在 4×10^7 到 8×10^7 左右，数据量充足，可保证后续分析的顺利进行。

表 2 从 NCBI GEO Datasets 数据库中获取的多类型癌症细胞系 RNA-seq 数据

癌症类型	细胞系	数据集	样本	样本编号	总读序数量
宫颈癌	HeLa	GSE149691	GSM4509043	HeLa_Rep1	58,091,594
宫颈癌	HeLa	GSE149691	GSM4509044	HeLa_Rep2	62,437,222
结肠癌	HT29	GSE149691	GSM4509039	HT29_Rep1	54,427,103
结肠癌	HT29	GSE149691	GSM4509040	HT29_Rep2	47,053,475
前列腺癌	42D	GSE113120	GSM3097219	42D_NoRep	81,834,221
前列腺癌	PC3	GSE113120	GSM3097221	PC3_NoRep	55,613,861
前列腺癌	V16A	GSE113120	GSM3097224	V16A_NoRep	76,719,348

注：样本编号是本课题基于样本来源进行的重新编号，仅在本课题中使用。

2.1.3 物种保守 circRNA 与癌症关联分析所使用的患者癌症组织和正常组织 circRNA 表达数据

表 3 MiOncoCirc 数据库中癌症组织和正常组织样本的 circRNA 表达数量

组别	癌症类型	样本量	circRNA 数量
PRAD	Prostate Adenocarcinoma 前列腺癌	217	75,272
BRCA	Breast Cancer 乳腺癌	118	70,247
SARC	Sarcoma (Osteo, Fibro) 肉瘤（骨肉瘤/纤维肉瘤）	78	59,896
MISC	Rare Cancer 罕见癌症	56	57,389
HNSC	Head and Neck Cancer 头颈癌	31	34,443
SECR	Glandular Cancer 腺癌	27	41,664
CHOL	Cholangiocarcinoma 胆管癌	26	29,592
LUNG	Lung Adenocarcinoma 肺腺癌	26	27,113
PAAD	Pancreatic Cancer 胰腺癌	26	27,695
Normal	Normal Tissue 正常组织	25	22,488
ALL	Acute Lymphoblastic Leukemia 急性淋巴细胞白血病	21	50,900
BLCA	Bladder Cancer 膀胱癌	16	23,555
OV	Ovarian Cancer 卵巢癌	14	22,488
SKCM	Skin Cancer 皮肤癌	14	15,026
COLO	Colon Cancer 结肠癌	13	19,129
NRBL	Neuroblastoma 成神经细胞瘤	13	27,659
ESCA	Esophageal Cancer 食管癌	12	26,899
KDNY	Kidney Cancer 肾癌	12	28,032
STAD	Stomach Cancer 胃癌	12	16,642
ACC	Adrenal Carcinoma 肾上腺癌	11	23,852
AML	Acute Myeloid Leukemia 急性髓系白血病	10	31,068
HCC	Liver Cancer 肝癌	10	35,180
RHABDO	Rhabdomyosarcoma 横纹肌肉瘤	10	20,901
MBL	Medulloblastoma 成神经管细胞瘤	7	24,161
JMML	Leukemia 白血病	6	19,696
THCA	Thyroid Cancer 甲状腺癌	6	6,678
GBM	Glioblastoma 成胶质细胞瘤	5	14,460

为进一步探究 circRNA 在癌症组织中的表达特征，本课题对 MiOncoCirc 数据库中来自患者癌症组织和正常组织的 circRNA 进行了差异表达分析。MiOncoCirc 数据库共包含来自 868 个样本的 RNA-seq 数据，按照癌症类型进行分组，共有 39 个组别（其中 38 组对应不同癌症类型，来自正常组织的样本归类为 1 组）[29]。该数据库将 RNA-seq 数据使用 STAR 方法比对到人类基因组后，使用 CIRCexplorer[42]和 CODAC 计算流程统计了各样本中 circRNA 的表达数据，输出为各样本中比对到每个 circRNA 的读序数量[29]。MiOncoCirc 数据库在其 v0 版本中提供了详细的样本来源和分组信息，在 v1 版本补充了部分样本，因此本课题选取了来自 v0 版本的 38 种癌症类型进行分析，在数据库中获取了来自两个版本的共计 845 个样本（癌症组织共包含来自 38 种癌症类型的 820 个样本，正常组织共包含 25 个样本）。

为了减少低表达 circRNA 对后续差异表达分析的影响，本课题选取了同时在至少 5 个样本中可以被检测的 circRNA 用于后续分析。此外，本课题选取了样本数量最多的前列腺癌（PRAD）和乳腺癌（BRCA）两种癌症，对其中差异表达 circRNA 的亲本基因进行 GO 和 KEGG 通路富集分析，以保证富集分析的准确性。

2.1.4 参考基因组信息和相关基因注释文件

本课题在物种保守性 circRNA 的筛选过程中使用了来自 USCS 基因组数据库的参考基因组，具体包括人类参考基因组（版本：GRCh38/hg38）、小鼠参考基因组（版本：GRCm38/mm10）、大鼠参考基因组（版本：RGSC6/rn6）、斑马鱼参考基因组（版本：GRCz10/danRer10）和线虫参考基因组（版本：WS220/ce10）。在多类型癌症细胞系 RNA-seq 数据的分析过程中，本课题在 rRNA 序列去除过程中来利用来自 NCBI 数据库中的人类 rRNA 参考序列文件 GCF_000001405.40_GRCh38.p14_rna_from_genomic.fna（包含人类 5S、5.8S、18S、28S、45S rRNA 和线粒体的 12S、16S rRNA 序列）通过 Linux 脚本和 seqkit 工具根据上述 rRNA 参考序列文件中的序列信息提取了人类 rRNA 的参考序列；在 HISAT2 序列联配中使用了来自 GENCODE 数据库的人类基因组注释文件 gencode.v41.annotation_spsites.txt。

2.2 方法

2.2.1 物种保守 circRNA 的筛选及初步分析

本课题从 CIRCpedia v2 数据库中获取了来自人类及包括小鼠、大鼠、斑马鱼和线虫等四种模式生物样本的 circRNA 表达和基因组位置信息。首先，本课题根据 circRNA 的基因组位置信息信息使用 Linux 脚本分别对来自各物种样本中不同组织及细胞系的 circRNA 进行去重，进一步选取其中来自 ribo– RNA-seq 样本的 circRNA 数据，并将其基因组位置信息转换为 BED 格式的 circRNA 注释文件。之后，使用 liftOver 工具将各 circRNA 注释文件进行基因组坐标转换，转换为人类基因组（版本：hg38）的基因组坐标。（参数：-bedPlus=3 -tab -minMatch=0.1 -minBlocks=1）需要注意的是，由于某些其它物种来源的 circRNA 在转换到人类基因组时可能同时存在对应多个不同的基因组位置而无法确定其实际对应关系，为保证后续分析的准确性，本课题对该部分 circRNA 进行了过滤和排除。本课题基于以下标准筛选了物种保守 circRNA：对于某物种中的某一环形 RNA，若其在人类基因组坐标的上下游 5 bp 范围内存在人源的环形 RNA，则认为该环形 RNA 在人类和该物种间保守[27]。最终，本课题基于 CIRCpedia v2 数据库中的 circRNA 注释文件，筛选出了在人类与小鼠、大鼠、斑马鱼和线虫四个物种间物种保守的 circRNA，输出为包含物种保守 circRNA 基因组位置信息的 BED 格式文件。本课题基于上述物种保守 circRNA 进行了初步分析，使用 Linux 和 R 语言脚本统计了物种保守 circRNA 的物种来源比例。同时，本课题还利用 bedtools 工具提取了物种保守 circRNA 的原始序列，使用 seqkit 针对物种保守 circRNA 的序列特征进行提取和分析，包括序列碱基比例等。

2.2.2 多类型癌症细胞系中 circRNA 的鉴定与表达图谱构建

本课题从 NCBI GEO 数据库中的 GSE149641 和 GSE113120 数据集中分别获取了来自 HeLa 和 HT29，以及 42D、PC3 和 V16A 共 5 个细胞系的 7 个样本的 RNA-seq 数据（表 2），原始测序数据为 SRA 文件格式。本课题首先使用 fasterq-dump 工具将 SRA 文件格式转换为 FASTQ 格式，使用 FastQC 工具进行测序质量控制，获取测序质量控制报告。之后，使用 Trimmomatic（v0.39）工具去除了读序中来自 Illumina 平台的接头片段去除接头的读序数据比对到从 NCBI 数据库中提取的人类 rRNA 序列上，分析其读序数据中 rRNA 来源的比例以减少 rRNA 序列在后续 circRNA 分析中的干扰。进一步地，本课题使用 HISAT2（v2.1.0）工

具将去接头后的读序数据比对到人类参考基因组（版本：GRCh38/hg38）上，使用 samtools 工具将输出的 SAM 格式比对结果文件转化为 BAM 格式文件，再使用 featureCounts（version 2.0.1）工具统计测序数据中线性 RNA 的表达数据，以 FPKM（fragments per kilobase per million）为指标计算其表达量。此外，本课题使用了 CLEAR/CIRCexplorer3 计算流程[24]进行 circRNA 的表达分析，统计了每个细胞系中 circRNA 及其对应线性 RNA 的表达水平，以 FPB_{circ}（fragments per billion mapped bases of circRNA）和 FPB_{linear}（fragments per billion mapped bases of linear RNA）指标进行评估，以 CIRCscore（FPB_{circ} 和 FPB_{linear} 的比值）衡量两者的表达差异。本课题统计了在各癌症细胞系中的物种保守和非保守 circRNA 的平均表达水平，筛选出其中表达水平位于前 10 位的物种保守 circRNA 借助文献资料等分析了其功能特征。

2.2.3 circRNA 的物种保守性及癌症关联分析

MiOncoCirc 数据库提供了来自多种癌症患者的癌症组织和正常组织样本的 circRNA 表达数据，其中 circRNA 的表达水平以样本中比对到该 circRNA 的读序数量表示。由于每个分组中各样本的测序数据量均有所差异，本课题使用各分组中样本测序数据中总读序数量的中位数（约 50,000,000）对 circRNA 表达水平进行标准化，每个标准化后的读序相当于每 50,000,000 个比对到线性基因上的读序中发现的一个反向剪接读序[29]，计算方法如下所示。

$$Normalized\ Reads\ Count = \frac{Actual\ Reads\ Count}{Sequencing\ depth\ of\ the\ sample} \times Median\ of\ sequencing\ depth\ of\ all\ samples$$

为去除低表达 circRNA 对后续分析的影响，本课题仅选用了同时在至少 5 个样本中出现的 circRNA 用于后续计算分析。基于每个样本中筛选出的 circRNA 表达数据，本课题使用 Perl 语言脚本 combine_diff.pl 针对不同癌症分组构建了包含组内所有样本中 circRNA 表达矩阵，并使用 edgeR 等 R 包进行了差异表达分析，筛选出在各癌症组织和正常组织间差异表达的 circRNA。本课题借助 ClusterProfiler、org.Hs.eg.db、topGO、DOSE 等 R 包对差异表达的 circRNA 的亲本基因进行了功能富集分析，根据其基因功能和相关文献资料来分析推测癌症中 circRNA 的潜在功能。本课题提取了 GO term 和 KEGG pathway 中的前 10 位，基于其中的生物学过程和相关通路进一步探讨 circRNA 在相关过程中的功能与机制。

三、研究结果

3.1 物种保守 circRNA 的筛选

3.1.1 物种保守 circRNA 的数量及分布

本课题共筛选出 17,934 个物种保守 circRNA，仅占 CIRCpedia v2 数据库中的人类来源的总 circRNA 约 15.96%，提示在进化过程中 circRNA 的产生具有一定的复杂性和多样性。在以上物种保守 circRNA 中，大多数仅在人和另外一个物种间保守（85.10%），少数在人和另外两个物种间保守（14.90%）。本课题针对在人和另外一个物种间保守的 circRNA 分析了其物种来源，发现绝大多数为在人和小鼠间保守的环形 RNA（91.00%），少数为在人和大鼠间保守的环形 RNA（8.99%），极少数为在人和斑马鱼间保守的环形 RNA（< 0.01%），未发现仅在人和线虫中保守的环形 RNA。参考图 1 所示 CIRCpedia v2 数据库中各物种的 ribo– RNA-seq 样本数量，推测上述结果可能是除小鼠和大鼠外其它物种的样本量较少导致。因此，本课题在后续中针对人鼠之间保守的环形 RNA 进行重点分析。

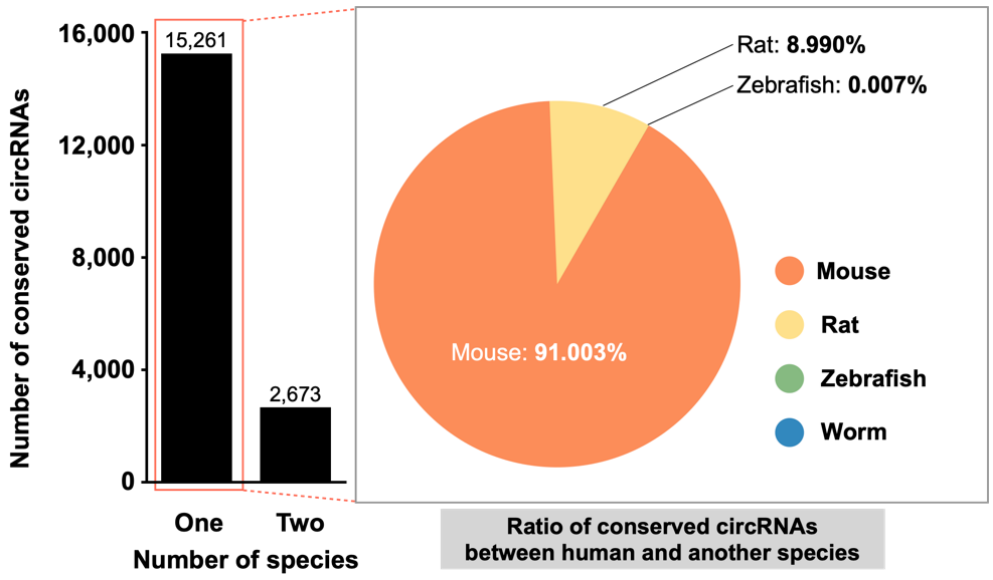

图 2 人类和其它物种之间保守性环形 RNA 的数量

左图，在人类和其它多个物种间保守的 circRNA 数量，横轴表示物种数目，纵轴表示在某一数目的物种中保守的 circRNA 数量；右图，在人类和其它 1 个物种间保守的 circRNA 的物种来源分布，图中百分比代表该物种来源的 circRNA 在人类和其它 1 个物种间保守的 circRNA 中的占比，图中不同颜色对应不同物种：橙色，小鼠；黄色，大鼠；绿色，斑马鱼；蓝色，线虫

3.1.2 物种保守 circRNA 的序列特征

本课题从碱基组成角度分析了物种保守 circRNA 的序列特征。如图 3 所示，结果表明物种保守 circRNA 的 GC 含量的平均值约为 39.61%，非物种保守 circRNA 的 GC 含量平均值约为 41.00%，t-test 检验结果表明两者间存在显著差异（$p < 2.2 \times 10^{-16}$）且物种保守 circRNA 的 GC 含量相对较低。基于 GC 含量对 DNA 稳定性的正相关，我们推测较低 GC 含量的保守性 circRNA 对应基因的转录能力更强从而引发了保守 circRNA 的高表达，这一假设有待进一步实验验证。

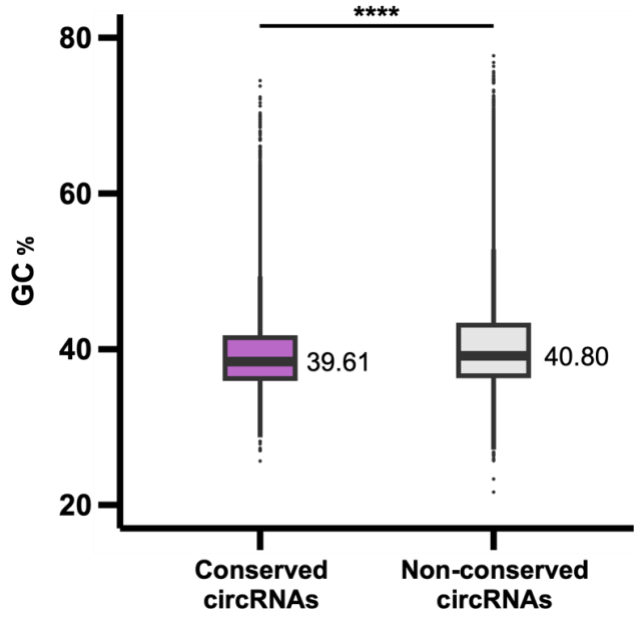

图 3 物种保守和非保守 circRNA 间的 GC 含量差异

物种保守 circRNA 和非保守 circRNA 序列中 GC 含量的差异，横轴为 circRNA 保守性，纵轴为序列 GC 含量（百分比），不同颜色代表 circRNA 保守性类型：紫色，物种保守 circRNA；灰色，非保守 circRNA

3.2 多类型癌症中物种保守 circRNA 表达分析

3.2.1 多类型癌症细胞系 RNA-seq 数据中线性 RNA 和 circRNA 表达分析

本课题对来自宫颈癌、结肠癌和前列腺癌的五种癌症细胞系的七个样本 RNA-seq 数据进行了分析。针对原始测序数据的处理和初步分析表明（步骤详见 2.2.1），上述去除接头序列步骤对测序质量成功进行了优化，且在数据中比对到 rRNA 的读序比例均较低，符合 circRNA 表达分析的需求。本课题在七个样本的 RNA-seq 数据中发现的 circRNA 个数在 4,000 ~ 8,500 不等，其中物种保守 circRNA 的数量在 1000 个左右，对应的表达基因数量在 10000 个左右，数据充足而可信度较高（表 2）。

表 2 多类型癌症细胞系中的 circRNA 数量统计

样本编号	总读序数	去接头后的总读序数	比对到 rRNA 的总读序比例	比对率	表达基因数量 (FPKM>1)	circRNA 数量	物种保守 circRNA 数量
HeLa_Rep1	58,091,594	57,628,966	1.32%	85.14%	13,830	8,138	1,463
HeLa_Rep2	62,437,222	62,114,528	1.03%	90.78%	9,749	4,963	950
HT29_Rep1	54,427,103	53,686,937	2.19%	88.70%	11,956	5,774	1,087
HT29_Rep2	47,053,475	46,317,149	1.67%	88.51%	12,327	5,356	1,057
42D_NoRep	81,834,221	77,512,349	9.27%	93.96%	11,577	4,814	985
PC3_NoRep	55,613,861	52,667,624	11.94%	81.04%	9,563	5,673	989
V16A_NoRep	76,719,348	75,387,825	15.63%	92.73%	11,968	6,306	1,108

注：样本编号是本课题基于样本来源进行的重新编号，仅在本课题中使用。

3.2.2 来自癌症细胞系 HeLa 和 HT29 样品重复组的相关性分析

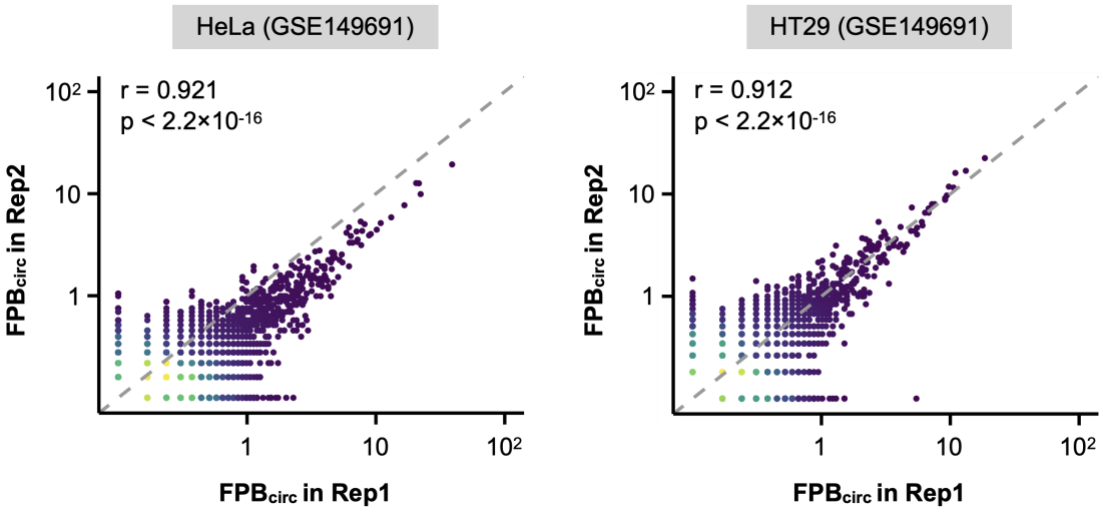

图 4 癌症细胞系 HeLa 与 HT29 中重复组的相关性分析

左图，HeLa 细胞系的两个重复组（样本编号 HeLa_Rep1、HeLa_Rep2）之间的表达相关性，r 为 Pearson 相关系数，p 为 Pearson 相关性检验的 p-value，散点颜色代表邻近点的密集程度，由黄色向紫色逐渐降低；右图，HT29 细胞系的两个重复组（样本编号 HT29_Rep1、HT29_Rep2）之间的表达相关性，r 为 Pearson 相关系数，p 为 Pearson 相关性检验的 p-value，散点颜色代表邻近点的密集程度，由黄色向紫色逐渐降低

本课题中采用的 HeLa 和 HT29 癌症细胞系测序数据分别来源于对应处理的 2 个重复组。本课题在每个细胞系中两个重复组间的相关性检验结果（图 4）显示，样品间相关系数分别为 r = 0.921（HeLa）；r = 0.912（HT29），且 $p < 2.2 \times 10^{-16}$，表明每个细胞系中两个重复组间有显著相关性。因此，为综合评估该细胞系整体的 circRNA 表达情况，可以使用两者表达量的平均值代表该细胞系 circRNA 的表达水平。

3.2.3 多类型癌症细胞系中物种保守 circRNA 的表达水平

本课题统计了 circRNA 在上述癌症细胞系中的表达水平，以 FPB_{circ} 指标表示。t-test 检验结果显示，在上述癌症细胞系中物种保守 circRNA 的表达水平均显著高于非保守 circRNA（图 5）。该结果进一步说明保守性 circRNA 很可能在相应癌症细胞系中有着重要功能。

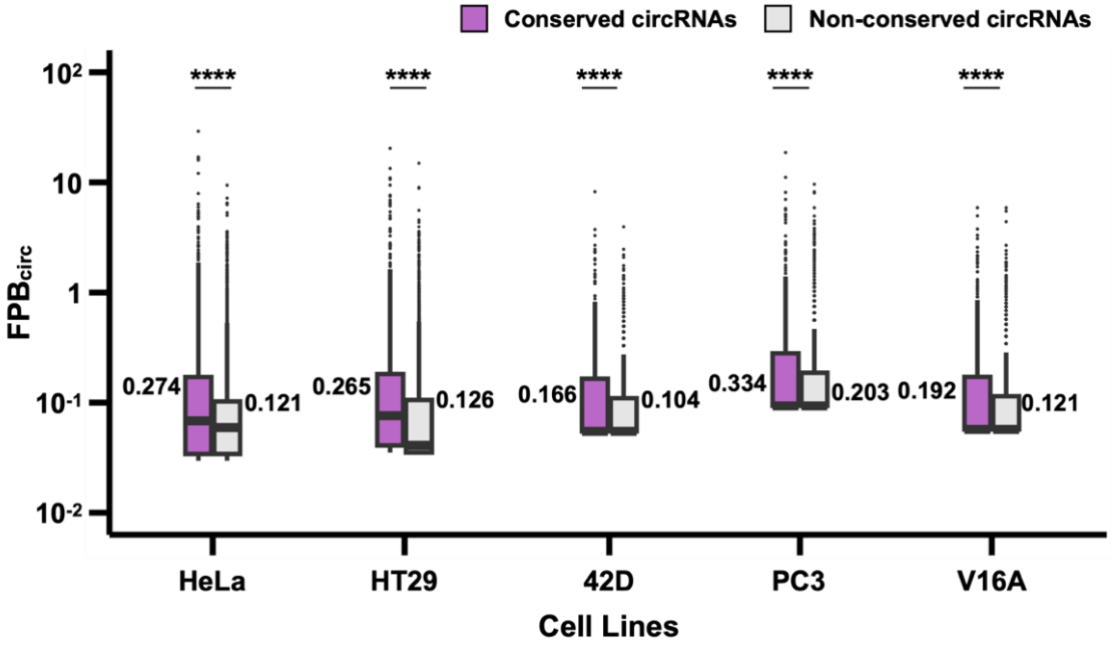

图 5 多类型癌症细胞系中的 circRNA 表达水平

物种保守 circRNA 和非保守 circRNA 在 HeLa、HT29、42D、PC3 和 V16A 细胞系中的表达水平，横轴为细胞系，纵轴为表达水平 FPB_{circ}，不同颜色代表 circRNA 的保守性：紫色，物种保守 circRNA；灰色，非保守 circRNA

3.2.4 多类型癌症细胞系中高表达物种保守 circRNA 功能分析

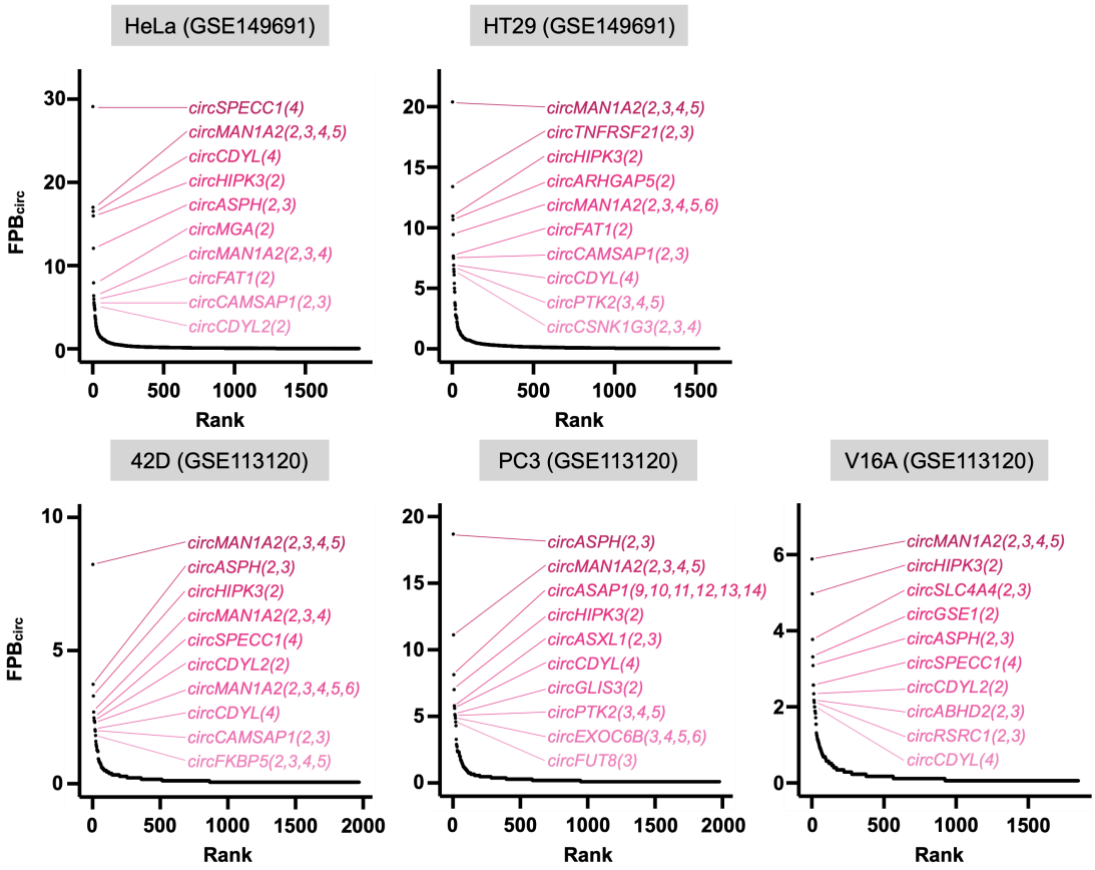

图 6 多类型癌症细胞系中表达水平位于前 10 位的物种保守 circRNA

在 HeLa、HT29、42D、PC3 和 V16A 五种癌症细胞系中表达水平分别位于细胞系内所有 circRNA 前 10 位的 circRNA，每一细胞系对应的散点图中横轴为表达水平排名，纵轴为表达水平 FPB_{circ}

本课题在每个癌症细胞系中筛选了高表达的物种保守 circRNA，以表达量前10位为筛选标准，共27个circRNA（图6）。其中，*circMAN1A2(2,3,4,5)*、*circHIPK3(2)* 和 *circCDYL(4)* 在所有五个癌症细胞系中表达量均位于前 10 位，表明上述 circRNA 很可能在癌症细胞中存在普适性的调控功能。因此，结合文献资料针对对其中表达量更高的 *circMAN1A2(2,3,4,5)* 和 *circHIPK3(2)* 进行了进一步的功能分析。

*circMAN1A2(2,3,4,5)*在人类、小鼠和大鼠中保守，在小鼠和大鼠中均存在对应的 circRNA 即 *circMan1a2(2,3,4,5)*。*circMAN1A2(2,3,4,5)* 的表达水平在五种癌症细胞系中均位于前两位，在 HT29 细胞系和 42D 细胞系中表达量显著高于其他 circRNA。同时，*circMAN1A2(2,3,4,5)* 相较于对应的线性 RNA 表达水平更高（在 5 中癌症细胞系中的 CIRCscore 均大于或接近 1）。上述结果表明，

*circMAN1A2(2,3,4,5)*在癌症细胞系中可能存在重要功能。已有研究证明，*circMAN1A2(2,3,4,5)*在胃癌[43]、卵巢癌[44]和鼻咽癌[45]等多种癌症类型中可以通过其 miRNA 海绵的功能促进肿瘤的发展。

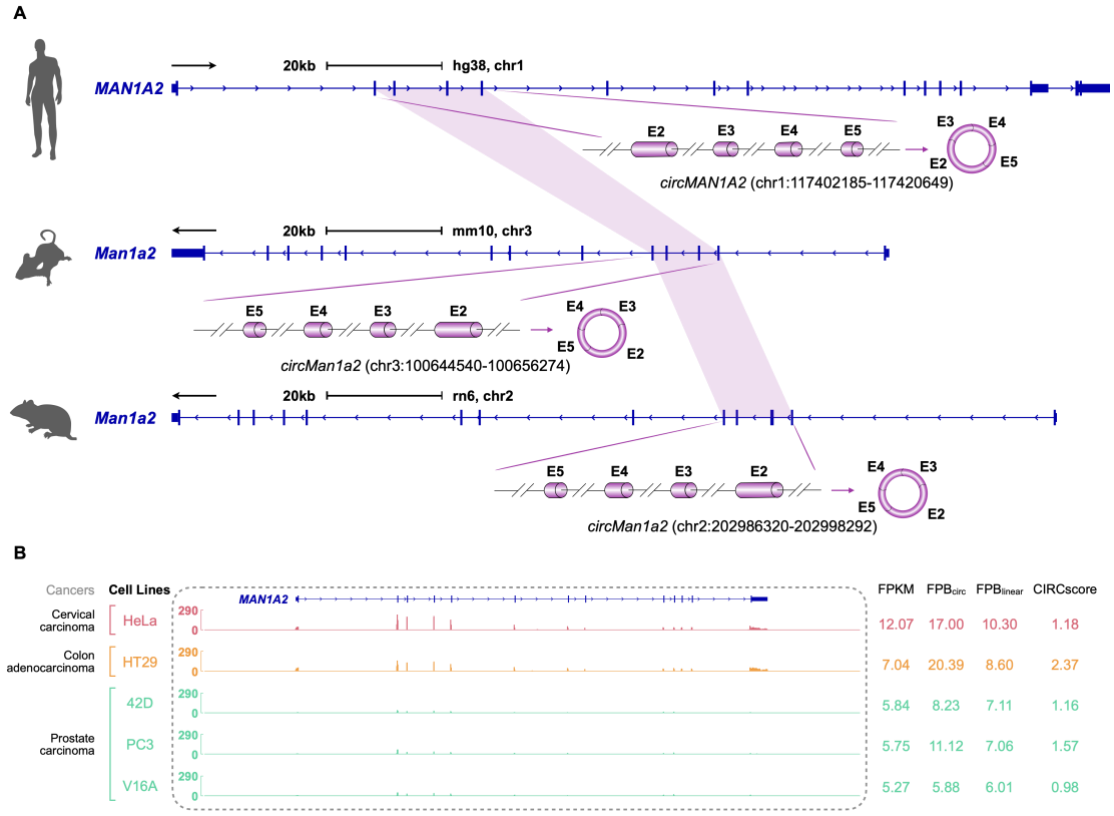

图 7 高表达 circRNA *circMAN1A2(2,3,4,5)*的保守性和表达分析

图 A，*circMAN1A2(2,3,4,5)*及其在小鼠和大鼠中对应的 circRNA *circMan1a2(2,3,4,5)* 间的保守性示意图；图 B，*circMAN1A2(2,3,4,5)*及其对应的线性 RNA 在不同细胞系中的表达水平

*circHIPK3(2)*在人类、小鼠和大鼠均保守，在小鼠和大鼠中均存在对应的 circRNA 即 *circHipk3(2)*。*circHIPK3(2)*的表达水平在五种癌症细胞系中相对较高，均位于前四位，提示其在肿瘤细胞中可能发挥着重要作用。已有研究报道，*circHIPK3* 可以作为 miRNA 海绵调节目的基因，调控肿瘤细胞的增殖、入侵和迁移等[46]。例如，在乳腺癌、结肠癌、前列腺癌等 14 余种人类癌症中，*circHIPK3* 可以作为多种 miRNA 的海绵促进癌症细胞的生长[15, 46]。

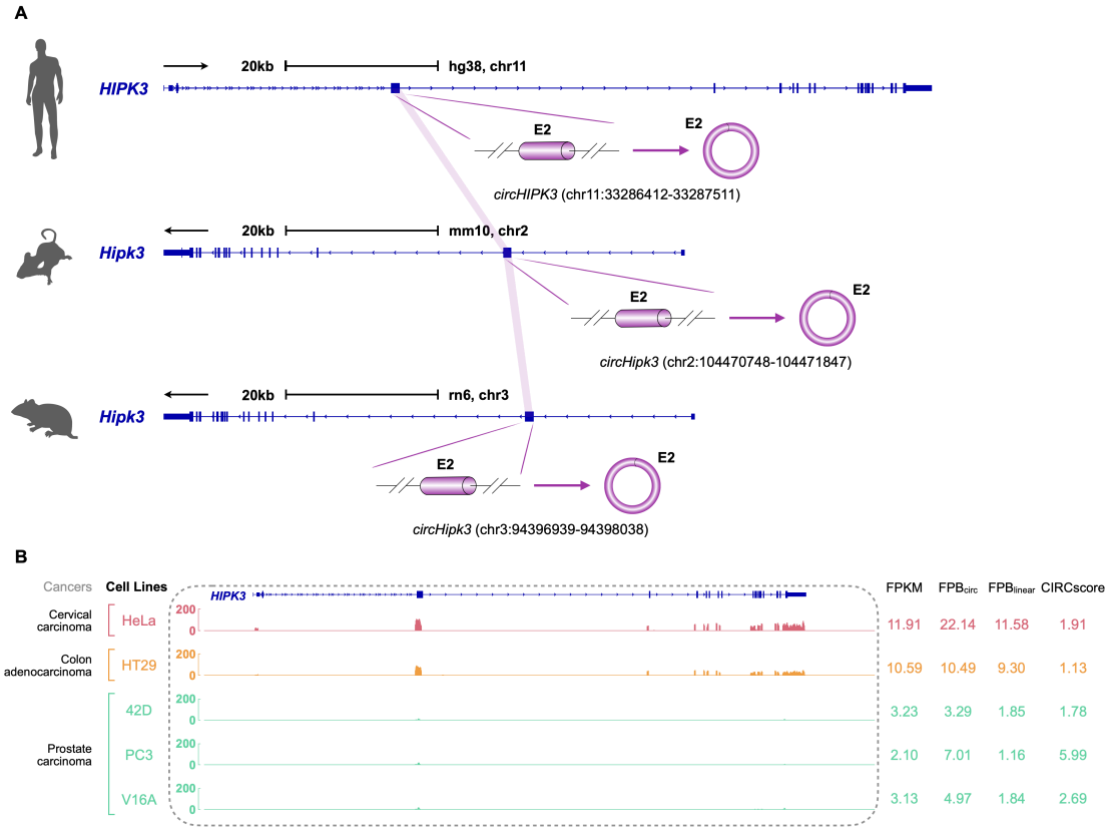

图 8 高表达 circRNA *circHIPK3(2)*的保守性和表达分析

图 A，*circHIPK3(2)*及其在小鼠和大鼠中对应的 circRNA *circHipk3(2)*间的保守性示意图；图 B，*circHIPK3(2)*及其对应的线性 RNA 在不同细胞系中的表达水平

3.3 物种保守 circRNA 与癌症关联分析

3.3.1 癌症组织与正常组织间 circRNA 差异表达分析

本课题针对 MiOncoCirc 数据库中的患者癌症组织和正常组织的 circRNA 表达数据进行了 circRNA 差异表达分析。本课题选取了样本量 ≥ 5 的癌症类型进行差异表达分析，共有 26 种（表 4）。结果显示，相较于正常组织，在所有的 26 种癌症类型中，绝大多数（24 种）癌症组织中拥有更多上调的 circRNA，仅有头颈癌（Head and Neck Cancer，HNSC）和白血病（Leukemia，JMML）两种类型中呈现出较多的下调 circRNA。本课题进一步对差异表达的 circRNA 的物种保守性进行分许，发现除前列腺癌（Prostate Adenocarcinoma，PRAD）外，在其它癌症类型中所有差异表达的 circRNA 均为物种保守 circRNA。上述结果表明，物种保守 circRNA 可能在癌症组织中发挥着重要功能而呈现出表达水平差异。值得注意的是，在前列腺癌组织中差异表达的 circRNA 中物种保守 circRNA 仅占 21.23%（表达上调的 circRNA 中占比为 19.96%，表达下调的 circRNA 中占比为 25.35%）。通过对前列腺癌中差异表达 circRNA 的表达水平进行分析，发现约 82%的

circRNA 表达量 ≤ 5 normalized reads，因而推测在前列腺癌中出现差异表达的 circRNA 中物种保守 circRNA 占比较低，是由于该癌症类型样本数量大（217 个样本）而导致大部分低表达 circRNA 被同样应用到差异表达分析中，后续应当设定表达量阈值对该推测进行进一步验证和分析。

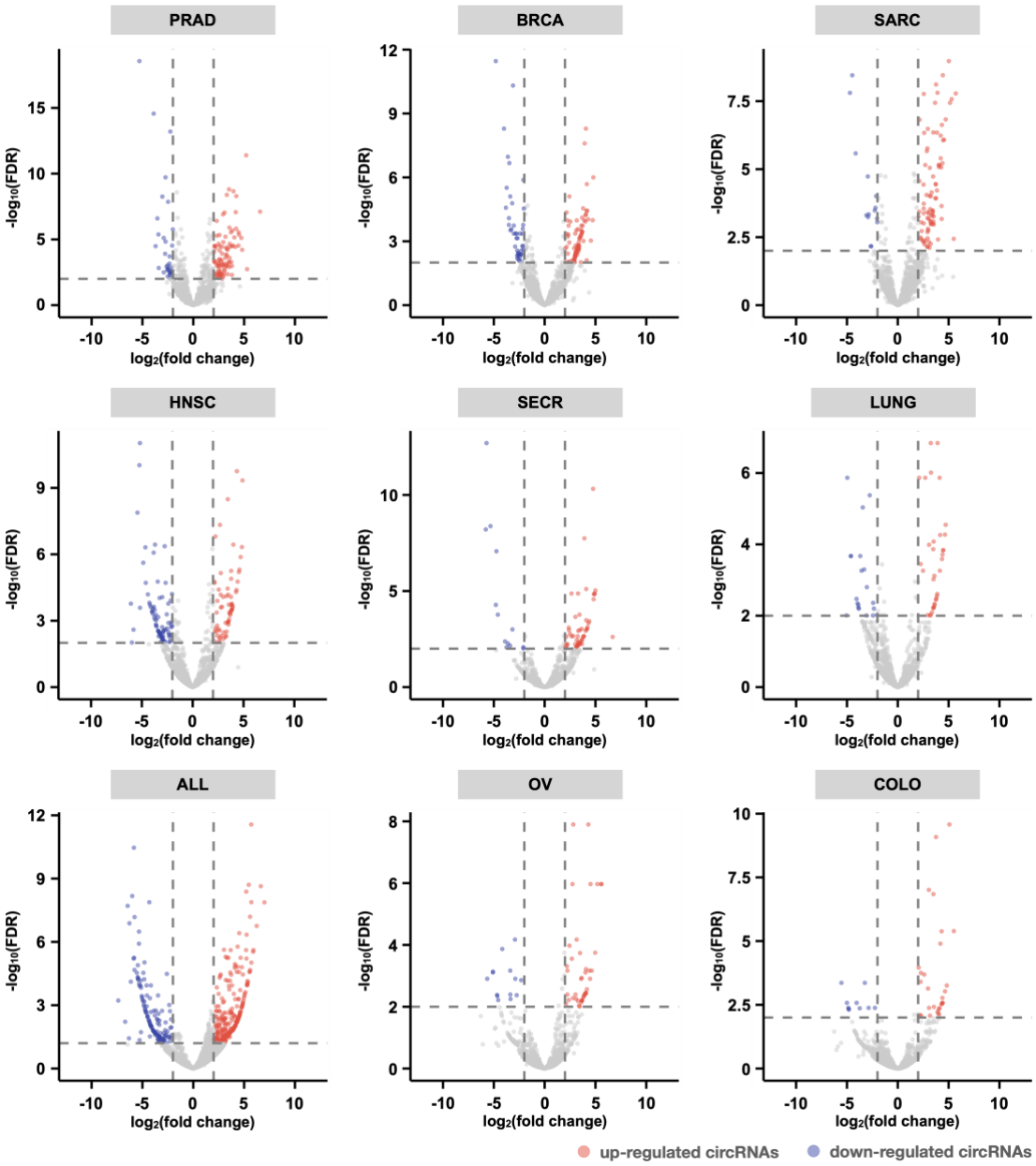

图 9 部分类型癌症组织与正常组织的 circRNA 差异表达分析

前列腺癌（PRAD）、乳腺癌（BRCA）、肉瘤（骨肉瘤/纤维肉瘤，SARC）、头颈癌（HNSC）、腺癌（SECR）、肺癌（LUNG）、急性淋巴细胞白血病（ALL）、卵巢癌（OV）和结肠癌（COLO）组织与正常组织中差异表达的 circRNA：红色，表达上调的 circRNA；蓝色，表达下调的 circRNA

表 4 部分类型癌症组织与正常组织的 circRNA 差异表达分析数据统计

癌症类型	样本数量	circRNA 数量	过滤低表达后的 circRNA 数量	上调 circRNA 数量	下调 circRNA 数量	上调保守 circRNA 比例	下调保守 circRNA 比例	差异表达 circRNA 比例
正常组织数据（用于差异表达分析的对照）								
Normal	25	22,488	5,347	0	0	/	/	/
拥有较多上调的 circRNA 的癌症组织类型								
PRAD	217	75,272	4,415	461	142	19.96%	25.35%	21.23%
BRCA	118	70,247	4,765	82	44	100.00%	100.00%	100.00%
SARC	78	59,896	5,227	86	14	100.00%	100.00%	100.00%
MISC	56	57,389	5,805	76	11	100.00%	100.00%	100.00%
SECR	27	41,664	5,900	63	13	100.00%	100.00%	100.00%
CHOL	26	29,592	3,860	53	23	100.00%	100.00%	100.00%
LUNG	26	27,113	3,507	34	20	100.00%	100.00%	100.00%
PAAD	26	27,695	3,914	34	10	100.00%	100.00%	100.00%
ALL	21	50,900	9,445	370	146	100.00%	100.00%	100.00%
BLCA	16	23,555	3,514	26	19	100.00%	100.00%	100.00%
OV	14	22,488	3,773	43	14	100.00%	100.00%	100.00%
SKCM	14	15,026	2,394	8	12	100.00%	100.00%	100.00%
COLO	13	19,129	3,029	29	10	100.00%	100.00%	100.00%
NRBL	13	27,659	4,591	82	11	100.00%	100.00%	100.00%
ESCA	12	26,899	4,570	58	9	100.00%	100.00%	100.00%
KDNY	12	28,032	3,594	42	13	100.00%	100.00%	100.00%
STAD	12	16,642	2,591	18	3	100.00%	100.00%	100.00%
ACC	11	23,852	3,786	53	14	100.00%	100.00%	100.00%
AML	10	31,068	5,037	81	64	100.00%	100.00%	100.00%

HCC	10	35,180	1,898	15	4	100.00%	100.00%	100.00%
RHABDO	10	20,901	2,898	30	6	100.00%	100.00%	100.00%
MBL	7	24,161	1,580	18	3	100.00%	100.00%	100.00%
THCA	6	6,678	10	0	18	/	100.00%	100.00%
GBM	5	14,460	53	1	0	100.00%	/	100.00%
拥有更多下调的 circRNA 的癌症组织类型								
HNSC	31	34,443	3,989	69	84	100.00%	100.00%	100.00%
JMML	6	19,696	2,167	38	43	100.00%	100.00%	100.00%

3.3.2 前列腺癌和乳腺癌中差异表达 circRNA 对应基因的功能富集分析

本课题对前列腺癌和乳腺癌中差异表达 circRNA 的对应基因进行了 GO 和 KEGG 通路富集分析，进一步探究 circRNA 在肿瘤发生发展过程中的潜在功能。在前列腺癌癌症组织中，GO 差异分析表明差异表达 circRNA 对应的基因主要与蛋白丝氨酸/苏氨酸激酶活性、核苷三磷酸酶调节因子活性、高尔基体囊泡运输、组蛋白结合与组蛋白修饰，以及脂质转运蛋白活性、细胞核维甲酸受体结合、ABC 型外源转运蛋白活性等过程相关。某些基因对应的 circRNA 在相应功能或通路中存在调控作用，具体来讲：1）蛋白质的丝氨酸和苏氨酸激酶参与细胞中多个关键的信号通路，而某些来源于对应基因的 circRNA 可以对该过程进行调控，例如 *circAKT3* 编码的 AKT3-174aa 在功能上与其对应的蛋白丝氨酸/苏氨酸激酶 AKT3 拮抗从而调控 ATK3 的作用效果[47]。2）核苷三磷酸酶参与癌症细胞中包括 DNA 复制和修复、转录、核糖体生物生成和翻译后蛋白糖基化等多种核苷酸代谢过程，其对应 circRNA 对癌症细胞也存在调控作用，例如 *circARHGAP26* 与其对应的 GTP 酶调节因子编码基因 ARHGAP26 对胃癌肿瘤细胞的调控作用相互拮抗[48]等，但许多该类 circRNA 具体作用机制还未得到充分阐明。KEGG 差异分析揭示了与差异表达 circRNA 对应基因相关联的通路，包括内吞作用、胰岛素抵抗、单磷酸腺苷活化蛋白激酶（AMP-activated protein kinase，AMPK）信号通路、胰高血糖素信号通路、神经营养因子信号通路、松弛素信号通路、长寿调节通路、多梳蛋白抑制复合物、ABC 转运蛋白和可卡因成瘾等。

在乳腺癌癌症组织中，GO 差异分析表明差异表达 circRNA 对应基因与有丝分裂细胞周期 G1/S 转变、DNA 代谢过程的调控、外源性凋亡信号通路的调控、

基因表达的表观遗传调控、生物肢体和附器的形态发生以及 RSC 型复合物等生物功能相关。某些基因对应的 circRNA 能够对上述过程起到调控作用：例如 *circSMARCC1* 可以通过诱导前列腺癌细胞的 G1/S 转变来促进细胞生长[49]，*circBIRC6* 通过 miRNA 海绵的功能促进肝细胞癌的细胞凋亡[50]等。KEGG 差异分析表明差异表达 circRNA 对应基因与多梳蛋白抑制复合物，肌动蛋白细胞骨架的调节，赖氨酸降解，肝细胞癌，突触引导，ErbB 信号通路，胆碱能突触，生长激素的合成、分泌和作用，VEGF 信号通路，硫辛酸代谢等通路相关。相应的，某些基因产生的 circRNA 也对来源基因参与的通路存在调控功能：*circEZH2* 在胶质瘤中可以作为 miRNA 海绵促进多梳蛋白抑制复合物（Polycomb repressive complex，PRC） CBX3 的表达从而促进癌症细胞发展[51]；*circSUZ12* 通过招募 FUS 蛋白来调多梳蛋白抑制复合物 SUZ12 的表达[52]；*circPTK2* 可以对与肌动蛋白细胞骨架（actin cytoskeleton）相关的上皮-间充质转化（epithelial-to-mesenchymal transition，EMT）过程产生抑制作用[53]等。

综合上述结果，在前列腺癌和乳腺癌中差异表达的 circRNA 中，一部分差异表达 circRNA 在功能上与其基因功能相关的生物过程或通路有着直接或间接的调控关系, 并有相关研究进行了报道。然而，仍有较多的差异表达 circRNA 未被系统性地研究，其对肿瘤发生发展的具体调控机制仍然需要进一步探索。

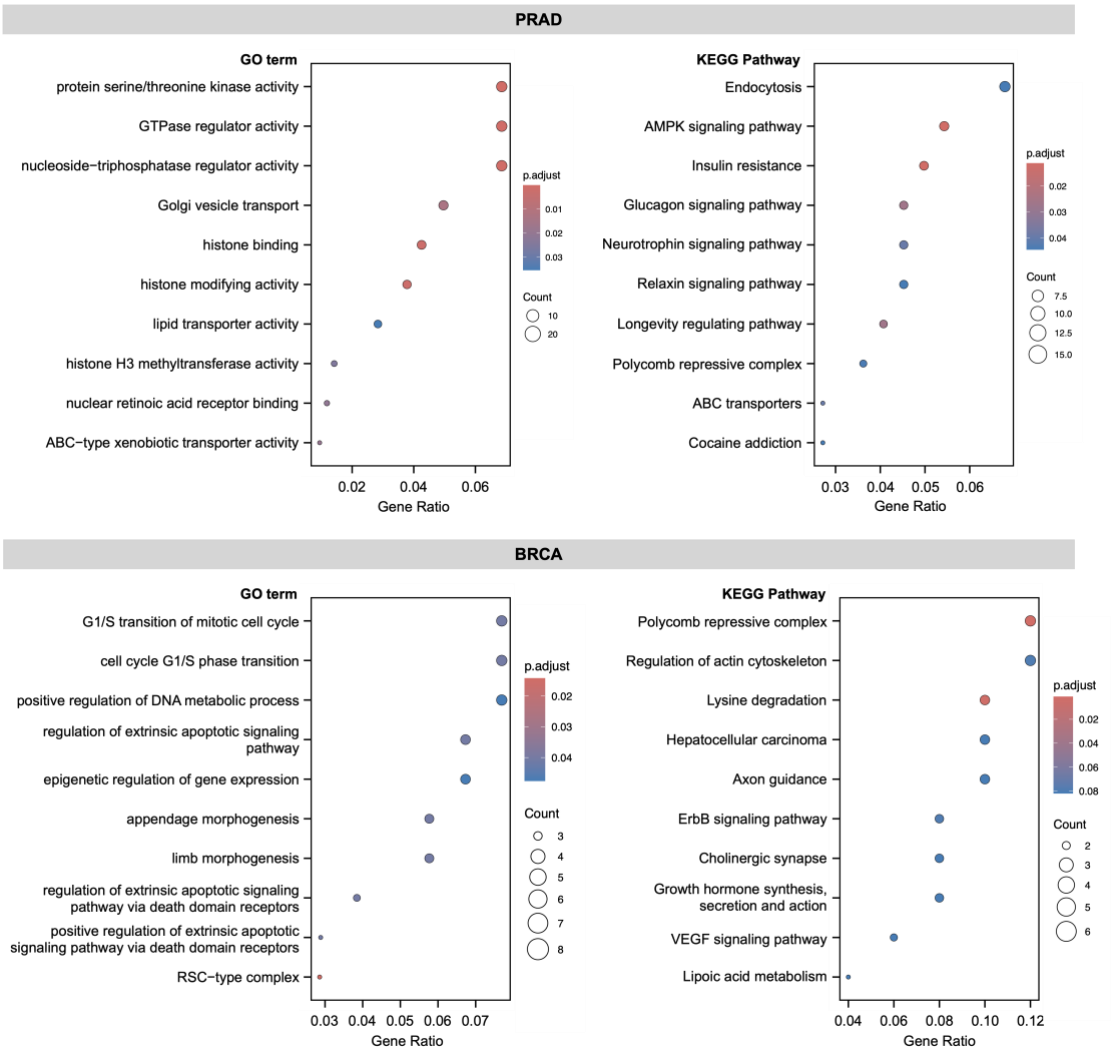

图 10 差异表达 circRNA 对应基因的 GO 和 KEGG 通路富集分析

上图，前列腺癌（PRAD）癌症组织与正常组织间差异表达 circRNA 对应基因的 GO 和 KEGG 通路富集分析结果；下图，乳腺癌（BRCA）癌症组织与正常组织间差异表达 circRNA 对应基因的 GO 和 KEGG 通路富集分析结果；p.adjust 代表校正后的富集 p-value，Count 为对应某一 GO term 或 KEGG pathway 的基因数目

3.4 总结

本课题首先基于 CIRCpedia v2 数据库鉴定了人类与小鼠、大鼠、斑马鱼和线虫等四种模式生物中共 17,934 个保守 circRNA，并初步探索了物种保守 circRNA 的序列特征。进一步地，为了解析物种保守 circRNA 与癌症的关联，本课题利用 HeLa、HT29、42D、PC3 和 V16A 等五种常用癌症细胞系的 RNA-seq 数据，构建细胞系水平的物种保守 circRNA 表达图谱，评估了物种保守 circRNA 与非保守 circRNA 的表达差异，物种保守 circRNA 在癌症细胞系中的高表达进一步提示了其在癌症的发生发展中扮演着重要的角色。此外，本课题还利用

MiOncoCirc 数据库在癌症组织和正常组织间进行了 circRNA 的差异表达分析，并对鉴定出的差异表达 circRNA 进行了保守性评估。在绝大多数的癌症中（除低表达 circRNA 较多的前列腺癌外），差异表达 circRNA 均为物种保守 circRNA，进一步提示了物种保守 circRNA 可能参与调控癌症的发生发展。最后，本课题针对差异表达 circRNA 对应的基因进行了 GO 和 KEGG 通路富集分析，并结合相关文献资料解析了部分 circRNA 的相关功能。已有报道表明，一部分差异表达 circRNA 在功能上与其基因功能相关的生物过程或通路有着直接或间接的调控关系。然而，仍有较多的差异表达 circRNA 未被系统性地研究，其对肿瘤发生发展的具体调控过程及机制需要进一步探索。

综上，本课题在癌症细胞系及组织水平上系统构建了物种保守 circRNA 的泛癌表达谱，初步解析了物种保守 circRNA 与癌症发生发展的关联，有助于更好地理解 circRNA 在癌症发生发展过程中参与的功能调控。本课题鉴定到的与癌症相关的物种保守 circRNA 分子，不仅为后续构建 circRNA 与癌症关联研究的模式生物提供了一定的参考，也为未来 circRNA 用于癌症诊断和治疗的应用提供了新的思路。

四、讨　论

外显子反向剪接来源的环形 RNA 是一类广泛分布的非编码 RNA，可以通过 miRNA 海绵、与 RBPs 结合等机制发挥重要的调控功能。近年来，circRNA 在癌症发生发展中的功能机制也逐渐得到了解析，在癌症诊断和治疗领域显示出发展前景。非编码 RNA 的物种保守性常对应其在各物种间重要且普适性的功能，提示着物种保守 circRNA 可能存在着潜在的重要功能。本课题鉴定了来自五种生物的 17,934 个物种保守 circRNA，有利于 circRNA 功能机制探索的所需模式生物的构建。此外，本课题关注了物种保守 circRNA 在常用癌症细胞系中的表达特征，并初步分析了高表达物种保守 circRNA 的功能机制，揭示了物种保守 circRNA 在癌症细胞中的重要功能。进一步地，本课题通过患者癌症组织和正常组织间的差异表达分析和保守性分析验证了物种保守 circRNA 与癌症的关联，获取了与癌症相关的差异表达 circRNA，有利于癌症关联 circRNA 功能机制研究的开展。最后，本课题根据基因富集分析结果对上述差异表达 circRNA 的功能进行了简要分析，探索了 circRNA 功能与其基因功能的关联，基于基因功能为 circRNA 功能探索的开展提供了建议。总体上，本课题鉴定了与癌症相关的物种保守 circRNA，为未来进行基于模式生物的 circRNA 功能探索提供了支持。

本课题在物种保守 circRNA 的筛选及其与癌症的关联分析过程中也发现了一些问题。首先，在模式生物的选取过程中，常见的模式生物果蝇由于缺乏对应的 chain file 供 liftOver 工具参考以进行基因组坐标转换，导致无法参与后续物种保守 circRNA 的筛选。因此，目前的研究结果在以果蝇为模式生物的诸多相关研究缺乏参考价值。在未来的研究中，对保守 circRNA 来源物种库的扩充是重要的研究方向，在扩充果蝇等更多模式生物的同时，可以参考 CircAtlas 等 circRNA 数据库中的更多物种，结合物种进化关系对物种保守 circRNA 进行深入探究。其次，在保守性 circRNA 的筛选过程中，本课题仅通过 2.2.1 中所示的筛选标准鉴定了物种保守 circRNA，在后续研究中可以参考多重保守评分（multiple conservation score, MCS）[28]等多种方法进一步验证所得 circRNA 的保守性，为未来的研究提供更加准确的数据。此外，在物种保守 circRNA 序列特征分析的过

程中，本课题着重关注了物种保守 circRNA 与非保守 circRNA 的 GC 含量差异，而翻译起始位点、RBP 结合位点等序列和结构特征仍然有待后续探索。最后，在差异表达分析的过程中，本课题发现在有且仅有前列腺癌的差异表达 circRNA 中存在非保守性 circRNA，推测这一结果可能是由于其样本量大而低表达量 circRNA 占比较高导致的，在后续研究中应当根据表达量阈值过滤后再次进行差异表达分析以验证这一推测。

关于物种保守 circRNA 在癌症中的功能调控，还有许多问题值得进一步研究和讨论，如针对高表达 circRNA 在癌症组织中功能通路，物种保守 circRNA 在癌症细胞中的定位与功能的联系等。针对物种保守 circRNA 和癌症的关联研究有利于进一步完善对 circRNA 功能机制的认识，为基于 circRNA 的癌症治疗方案做出贡献。

参考文献

1. Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning[M]. // SLABY O, CALIN G A. Non-coding RNAs in Colorectal Cancer. City: Springer International Publishing, 2016: 3-17.
2. Goodall G J, Wickramasinghe V O. 2021. RNA in cancer[J]. Nature Reviews Cancer, 21(1): 22-36.
3. Hu J, Huang H, Xi Z, et al. 2022. LncRNA SEMA3B-AS1 inhibits breast cancer progression by targeting miR-3940/KLLN axis[J]. Cell Death & Disease, 13(9): 800.
4. Teng Y, Ren Y, Hu X, et al. 2017. MVP-mediated exosomal sorting of miR-193a promotes colon cancer progression[J]. Nature Communications, 8(1): 14448.
5. Li B, Zhu L, Lu C, et al. 2021. circNDUFB2 inhibits non-small cell lung cancer progression via destabilizing IGF2BPs and activating anti-tumor immunity[J]. Nature Communications, 12(1): 295.
6. Chen L-L, Yang L. 2015. Regulation of circRNA biogenesis[J]. RNA Biology, 12(4): 381-8.
7. Li X, Yang L, Chen L-L. 2018. The Biogenesis, Functions, and Challenges of Circular RNAs[J]. Molecular Cell, 71(3): 428-42.
8. Sanger H L, Klotz G, Riesner D, et al. 1976. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures[J]. Proceedings of the National Academy of Sciences, 73(11): 3852-6.
9. Cocquerelle C, Mascrez B, Hétuin D, et al. 1993. Mis-splicing yields circular RNA molecules[J]. The FASEB Journal, 7(1): 155-60.
10. Capel B, Swain A, Nicolis S, et al. 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis[J]. Cell, 73(5): 1019-30.
11. Chen L-L. 2020. The expanding regulatory mechanisms and cellular functions of circular RNAs[J]. Nature Reviews Molecular Cell Biology, 21(8): 475-90.
12. Patop I L, Wüst S, Kadener S. 2019. Past, present, and future of circRNAs[J]. The EMBO Journal, 38(16): e100836.
13. Chen L-L, Bindereif A, Bozzoni I, et al. 2023. A guide to naming eukaryotic circular RNAs[J]. Nature Cell Biology, 25(1): 1-5.
14. Misir S, Wu N, Yang B B. 2022. Specific expression and functions of circular RNAs[J]. Cell Death & Differentiation, 29(3): 481-91.
15. Zheng Q, Bao C, Guo W, et al. 2016. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs[J]. Nature Communications, 7(1): 11215.
16. Du W W, Yang W, Chen Y, et al. 2017. Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses[J]. European Heart Journal, 38(18): 1402-12.

17. Chen R-X, Chen X, Xia L-P, et al. 2019. N6-methyladenosine modification of circNSUN2 facilitates cytoplasmic export and stabilizes HMGA2 to promote colorectal liver metastasis[J]. Nature Communications, 10(1): 4695.
18. Wu N, Yuan Z, Du K Y, et al. 2019. Translation of yes-associated protein (YAP) was antagonized by its circular RNA via suppressing the assembly of the translation initiation machinery[J]. Cell Death & Differentiation, 26(12): 2758-73.
19. Yang Y, Gao X, Zhang M, et al. 2018. Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis[J]. JNCI: Journal of the National Cancer Institute, 110(3): 304-15.
20. Huang D, Zhu X, Ye S, et al. 2024. Tumour circular RNAs elicit anti-tumour immunity by encoding cryptic peptides[J]. Nature, 625(7995): 593-602.
21. Ma X-K, Zhai S-N, Yang L. 2023. Approaches and challenges in genome-wide circular RNA identification and quantification[J]. Trends in Genetics, 39(12): 897-907.
22. Xiao M-S, Wilusz J E. 2019. An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3′ ends[J]. Nucleic Acids Research, 47(16): 8755-69.
23. Rahimi K, Færch Nielsen A, Venø M T, et al. 2021. Nanopore long-read sequencing of circRNAs[J]. Methods, 196: 23-9.
24. Ma X-K, Wang M-R, Liu C-X, et al. 2019. CIRCexplorer3: A CLEAR Pipeline for Direct Comparison of Circular and Linear RNA Expression[J]. Genomics, Proteomics & Bioinformatics, 17(5): 511-21.
25. Cheng J, Metge F, Dieterich C. 2016. Specific identification and quantification of circular RNAs from sequencing data[J]. Bioinformatics, 32(7): 1094-6.
26. Wang K, Singh D, Zeng Z, et al. 2010. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery[J]. Nucleic Acids Research, 38(18): e178-e.
27. Dong R, Ma X-K, Li G-W, et al. 2018. CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison[J]. Genomics, Proteomics & Bioinformatics, 16(4): 226-33.
28. Wu W, Ji P, Zhao F. 2020. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes[J]. Genome Biology, 21(1): 101.
29. Vo J N, Cieslik M, Zhang Y, et al. 2019. The Landscape of Circular RNA in Cancer[J]. Cell, 176(4): 869-81.e13.
30. Shao Y, Li J, Lu R, et al. 2017. Global circular RNA expression profile of human gastric cancer and its clinical significance[J]. Cancer Medicine, 6(6): 1173-80.
31. Xia Q, Ding T, Zhang G, et al. 2018. Circular RNA Expression Profiling Identifies Prostate Cancer- Specific circRNAs in Prostate Cancer[J]. Cellular Physiology and Biochemistry, 50(5): 1903-15.
32. Zou Y, Zheng S, Deng X, et al. The Role of Circular RNA CDR1as/ciRS-7 in Regulating Tumor Microenvironment: A Pan-Cancer Analysis[J/OL]. Biomolecules, 2019,9(9): [

33. Seemann S E, Mirza A H, Hansen C, et al. 2017. The identification and functional annotation of RNA structures conserved in vertebrates[J]. Genome Res, 27(8): 1371-83.
34. Rivas E. 2021. Evolutionary conservation of RNA sequence and structure[J]. Wiley Interdiscip Rev RNA, 12(5): e1649.
35. Woese C R, Fox G E, Zablen L, et al. 1975. Conservation of primary structure in 16S ribosomal RNA[J]. Nature, 254(5495): 83-6.
36. Xue H, Shen W, Giegé R, et al. 1993. Identity elements of tRNA(Trp). Identification and evolutionary conservation[J]. Journal of Biological Chemistry, 268(13): 9316-22.
37. Kazantsev A V, Pace N R. 2006. Bacterial RNase P: a new view of an ancient enzyme[J]. Nature Reviews Microbiology, 4(10): 729-40.
38. Wang P L, Bao Y, Yee M-C, et al. 2014. Circular RNA Is Expressed across the Eukaryotic Tree of Life[J]. PLOS ONE, 9(3): e90859.
39. Müller B, Grossniklaus U. 2010. Model organisms — A historical perspective[J]. Journal of Proteomics, 73(11): 2054-63.
40. Gao X, Ma X-K, Li X, et al. 2022. Knockout of circRNAs by base editing back-splice sites of circularized exons[J]. Genome Biology, 23(1): 16.
41. Chen S, Huang V, Xu X, et al. 2019. Widespread and Functional RNA Circularization in Localized Prostate Cancer[J]. Cell, 176(4): 831-43.e22.
42. Zhang X-O, Wang H-B, Zhang Y, et al. 2014. Complementary Sequence-Mediated Exon Circularization[J]. Cell, 159(1): 134-47.
43. Guo R, Cui X, Li X, et al. 2022. CircMAN1A2 is upregulated by Helicobacter pylori and promotes development of gastric cancer[J]. Cell Death & Disease, 13(4): 409.
44. Li B, Hu C, Zhao D, et al. 2024. Circular RNA circMAN1A2 promotes ovarian cancer progression through the microRNA-135a-3p/IL1RAP/TAK1 pathway[J]. PeerJ, 12: e16967.
45. Dang Q-Q, Li P-H, Wang J, et al. 2023. CircMAN1A2 contributes to nasopharyngeal carcinoma progression via enhancing the ubiquitination of ATMIN through miR-135a-3p/UBR5 axis[J]. Human Cell, 36(2): 657-75.
46. Zhang Y, Liu Q, Liao Q. 2020. CircHIPK3: a promising cancer-related circular RNA[J]. Am J Transl Res, 12(10): 6694-704.
47. Xia X, Li X, Li F, et al. 2019. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-dependent Kinase-1[J]. Mol Cancer, 18(1): 131.
48. Zhang L, Zhou A, Zhu S, et al. 2022. The role of GTPase-activating protein ARHGAP26 in human cancers[J]. Molecular and Cellular Biochemistry, 477(1): 319-26.
49. Xie T, Fu D-J, Li Z-M, et al. 2022. CircSMARCC1 facilitates tumor progression by disrupting the crosstalk between prostate cancer cells and tumor-associated macrophages via miR-1322/CCL20/CCR6 signaling[J]. Molecular Cancer, 21(1): 173.

50. Yang G, Wang X, Liu B, et al. 2019. circ-BIRC6, a circular RNA, promotes hepatocellular carcinoma progression by targeting the miR-3918/Bcl2 axis[J]. Cell Cycle, 18(9): 976-89.
51. Gao F, Du Y, Zhang Y, et al. 2020. Circ-EZH2 knockdown reverses DDAH1 and CBX3-mediated cell growth and invasion in glioma through miR-1265 sponge activity[J]. Gene, 726: 144196.
52. Li L, Li C, Cao S, et al. 2023. Circ-SUZ12 Protects Cardiomyocytes from Hypoxia-Induced Dysfunction Through Upregulating SUZ12 Expression to Activate Wnt/β-catenin Signaling Pathway[J]. International Heart Journal, 64(6): 1113-24.
53. Wang L, Tong X, Zhou Z, et al. 2018. Circular RNA hsa_circ_0008305 (circPTK2) inhibits TGF-β-induced epithelial-mesenchymal transition and metastasis by controlling TIF1γ in non-small cell lung cancer[J]. Molecular Cancer, 17(1): 140.

致　谢

行文至此，皆为终章。本科的四年时光如白驹过隙，终于要画上最后的句号。回首过去的四年，不禁感慨万千，有太多的人和事值得我用最真挚的语言去铭记。

感谢杨力老师对我毕业论文的悉心指导。您在毕业论文的整个过程中为我指引了方向，带领我逐步走进了科研的大门，真正领略到科研的魅力。您的学术热情和工作态度也鼓励着我在科研路上更进一步。

感谢袁国华师兄和 YangLab 课题组的大家，你们在我毕业课题的过程中给予了我无尽的包容与耐心，为我答疑解惑，你们的帮助让我感受到了课题组的温暖，让我在初入门槛时少了些迷茫。

感谢我的父母在本科阶段的一贯支持。你们一直是我最坚实的后盾，最温暖的港湾，你们的宽慰、鼓励支撑着我克服困难，不断前行。

感谢我在本科阶段的所有老师和同学们，本科的学习生活有你们的陪伴才让我在的大学生涯中不再孤单。感谢贺强老师在 DreamLab 科创项目中给予我的帮助，这次比赛经历帮助我积累了宝贵的科研经验。

感谢我的小兔在一年来的陪伴，在我最焦虑的时光有你来治愈。

感谢复旦大学自行车协会和复旦大学公路车队的大家，你们的活力与能量让我的大学时光更加多彩。

不知不觉，我走到了本科阶段的 ending，虽然有不舍、有遗憾，但都已经成为了过去式。正如杨老师说的那样，现在只是 the end of the beginning，我的人生才刚刚开始，还有很长的路等待我去探索，还有很大的留白等待我去填充。

今天是 5 月 20 日，祝我自己爱我所爱，毕业快乐。