计算机科学
COMPUTER SCIENCE

生物标志物智能识别关键技术:环状RNA与疾病关联预测研究综述

胡学钢, 李扬, 王磊, 李培培, 尤著宏

引用本文

胡学钢, 李扬, 王磊, 李培培, 尤著宏生物标志物智能识别关键技术:环状RNA与疾病关联预测研究综述[J]. 计算机科学, 2023, 50(4): 369-387.

HU Xuegang, LI Yang, WANG Lei, LI Peipei, YOU Zhuhong. Key Technologies of Intelligent Identification of Biomarkers:Review of Research on Association Prediction Between Circular RNA and Disease [J]. Computer Science, 2023, 50(4): 369-387.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于随机洋葱路由的LBS移动隐私保护方案

LBS Mobile Privacy Protection Scheme Based on Random Onion Routing

计算机科学, 2022, 49(9): 347-354. https://doi.org/10.11896/jsjkx.210800077

基于节点相似性和网络嵌入的复杂网络社区发现算法

Complex Network Community Detection Algorithm Based on Node Similarity and Network Embedding

计算机科学, 2022, 49(3): 121-128. https://doi.org/10.11896/jsjkx.210200009

面向申威1621通用矩阵乘算法的实现与优化

Implementation and Optimization of Sunway1621 General Matrix Multiplication Algorithm

计算机科学, 2021, 48(11A): 699-704. https://doi.org/10.11896/jsjkx.201200150

基于核保持嵌入的子空间学习

Kernel-preserving Embedding Based Subspace Learning

计算机科学, 2021, 48(6): 79-85. https://doi.org/10.11896/jsjkx.200900014

基于块对角化表示的多视角字典对学习

Multi-view Dictionary-pair Learning Based on Block-diagonal Representation

计算机科学, 2021, 48(1): 233-240. https://doi.org/10.11896/jsjkx.200800211

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220500114

生物标志物智能识别关键技术:环状 RNA 与疾病关联预测研究综述

胡学钢[1,2,3] **李 扬**[1,2,3] **王 磊**[4,5] **李培培**[1,2,3] **尤著宏**[6]
1 大数据知识工程教育部重点实验室(合肥工业大学) 合肥 230601
2 合肥工业大学计算机与信息学院 合肥 230601
3 合肥工业大学大知识科学研究院 合肥 230601
4 枣庄学院信息科学与工程学院 山东 枣庄 277160
5 广西科学院大数据与智能计算研究中心 南宁 530007
6 西北工业大学计算机学院 西安 710129
(jsjxhuxg@hfut.edu.cn)

摘 要 生物标志物识别是实现"精准医疗"的一大基础,其对复杂疾病诊断、判断疾病分期及评价新药或新疗法在目标人群中的安全性和有效性具有重要作用。作为生物标志物智能识别关键技术,环状 RNA 与疾病关联预测是深入评测和衡量被试个体生物学过程、病理学过程及干预病理学反应的关键,是践行"精准医疗"的有效手段和途径之一。文中对基于生物大数据挖掘的环状 RNA-疾病关联预测计算模型进行了全面梳理和展望。具体地,首先从环状 RNA 的研究背景、理化特性、功能等方面探讨了环状 RNA 与疾病之间的关系;然后调研了环状 RNA 和疾病公共数据库资源;接着从智能计算的角度梳理了环状 RNA-疾病关联预测的 4 类计算方法,并分析了其优势与不足;最后讨论了环状 RNA-疾病关联预测问题目前面临的挑战和未来可能的研究方向。

关键词:精准医疗;生物大数据;环状 RNA;环状 RNA-疾病关联;计算模型
中图法分类号 TP391

Key Technologies of Intelligent Identification of Biomarkers: Review of Research on Association Prediction Between Circular RNA and Disease

HU Xuegang[1,2,3], LI Yang[1,2,3], WANG Lei[4,5], LI Peipei[1,2,3] and YOU Zhuhong[6]
1 Key Laboratory of Knowledge Engineering with Big Data(Hefei University of Technology), Ministry of Education, Hefei 230601, China
2 School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China
3 Research Institute of Big Knowledge, Hefei University of Technology, Hefei 230601, China
4 College of Information Science and Engineering, Zaozhuang University, Zaozhuang, Shandong 277160, China
5 Big Data and Intelligent Computing Research Center, Guangxi Academy of Sciences, Nanning 530007, China
6 School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China

Abstract Biomarker recognition is a major basis for achieving precision medicine, which plays an important role in diagnosing complex diseases, judging disease stages and evaluating the safety and effectiveness of new drugs or therapies in the target population. As the key technology of intelligent identification of biomarkers, the prediction of the association between circular RNA and disease is the key to deeply evaluate and measure the biological process, pathological process and intervention pathological response of subjects, and is one of the effective means and approaches to practice "precision medicine". This paper comprehensively combs and prospects the circular RNA disease association prediction model in biological big data. Specifically, it first discusses the relationship between circRNA and disease in terms of the research background, physicochemical properties and functions of circRNAs. Then, it investigates the public database resources of circRNAs and diseases, and summarizes four computational methods of circRNA-disease association prediction from the perspective of computational models, and analyzes their advantages and shortcomings. Finally, this paper discusses the current challenges and future possible research directions of circRNA-disease associa-

到稿日期:2022-05-13 返修日期:2022-09-19
基金项目:国家重点研发计划(2016YFB1000901);国家自然科学基金(62172355,61806065,61702444);中央高校基本科研业务费专项资金(JZ2020HGQA0186)
This work was supported by the National Key Research and Development Program of China(2016YFB1000901), National Natural Science Foundation of China(62172355,61806065,61702444) and Fundamental Research Funds for the Central Universities(JZ2020HGQA0186).
通信作者:王磊(leiwang@gxas.cn)

tion prediction problem.

Keywords Precision medicine, Biological big data, CircRNA, CircRNA-disease association, Computational model

随着生物信息学的迅速发展，海量生物数据不断积累，如何对这些生物大数据进行有效的挖掘与分析，成为了当前生命科学领域和计算机科学领域急需解决的问题[1]。现有的研究表明，circRNA 可以通过多种途径在生物体内起到不同的作用。大量的证据显示 circRNA 与许多常见的人类疾病密切相关，因此对 circRNA-disease 关联的计算模型进行深入的研究具有重要的意义。circRNA-disease 关联预测的计算模型旨在基于已知的、经过实验验证的 circRNA-disease 的关联数据，开发一个有效且准确的计算框架并预测新的、未知的或潜在的 circRNA-disease 之间的关联。这对于疾病的早期治疗和临床应用具有重要的研究价值。

图 1 为 circRNA 的发展历程图。1976 年，研究人员首次在植物类病毒中发现了环状 RNA，从而提出了 circRNA 的概念。随后，1993 年，人们在小鼠的 Sry 基因中也发现了大量且丰富的环状 RNA。随着测序技术的快速发展，2012 年，通过高通量测序技术，人们在不同物种中相继发现了许多 circRNA，从而使 circRNA 逐渐进入了科研领域。2013 年，Nature 杂志刊发了两篇关于 circRNA 的研究论文，首次证实了 circRNA 作为一种有效的 miRNA 海绵的作用，并由此揭开了 circRNA 的神秘面纱。当前，COVID-19 病毒已经在全球范围内广泛蔓延，给人类的生命健康带来了严重的危害。2020 年，有研究发现，circRNA 大量存在于冠状病毒中，且极有可能是一种在冠状病毒中广泛存在的分子，这为我们更好地预测冠状病毒提供了一种新的途径。近年来，有关 circRNA 的研究迅速增长，使其成为了 RNA 领域的一个新的研究热点，特别是 circRNA-disease 关联预测研究已经成为一个非常有价值的研究课题。此外，已有的研究表明 circRNA 可以作为多种人类疾病的重要的生物标志物，circRNA-disease 关联预测研究对于其他生命科学领域相关问题具有重要意义。

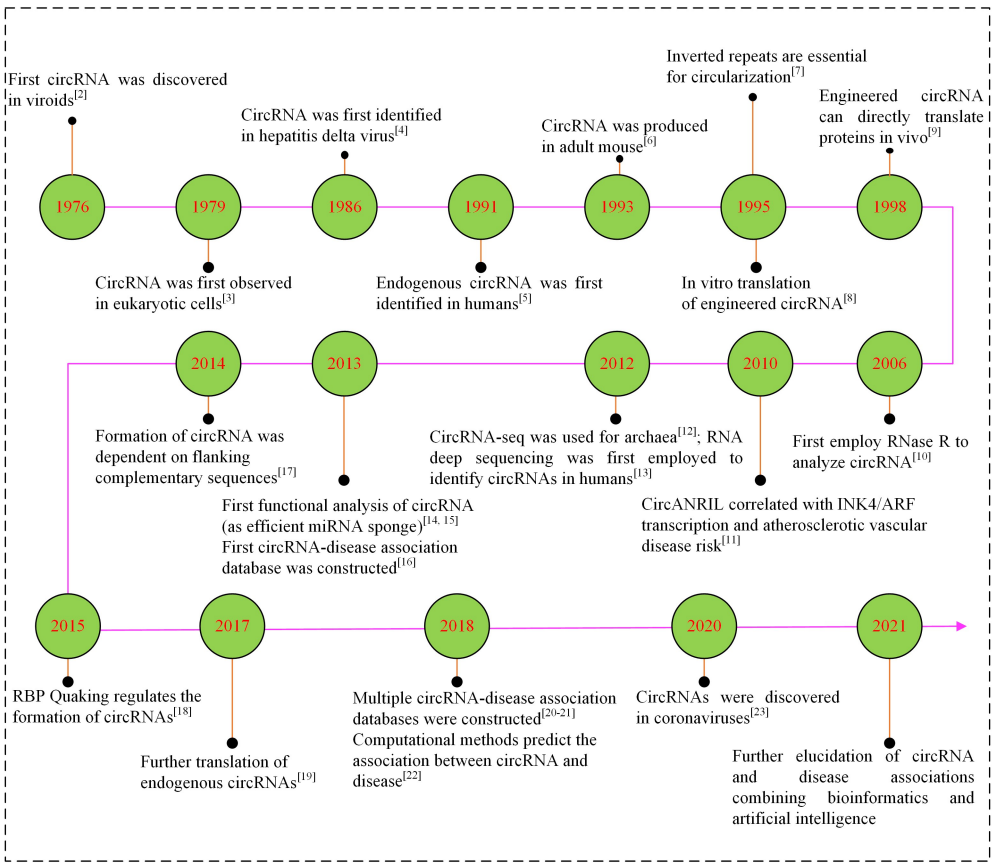

图 1 circRNA 的发展历程图

Fig. 1 Development diagram of circRNA

本文首先对 circRNA 和疾病之间的关系进行了全面的分析，然后详细整理了与 circRNA 和疾病相关的公开数据库资源，接着从计算模型的角度回顾了 circRNA-disease 关联预测现有的计算方法，对现有的 circRNA-disease 关联预测的各种计算方法进行了总结和分类，并对其存在的不足加以探讨。最后对基于计算方法的 circRNA-disease 关联预测研究所面临的挑战进行了讨论。本文旨在通过对当前 circRNA-disease 关联预测的相关计算方法的研究现状加以梳理，希望为后续研究提供可借鉴的思路。本文的具体贡献包含 3 个方面：

(1)对生物大数据中的 circRNA-disease 关联预测进行了全面的梳理，主要从 circRNA 的研究背景、性质、功能、circRNA 与常见疾病的关系、网络数据库以及现存的计算方法等方面详细介绍了 circRNA-disease 关联预测的计算模型。

(2)重点介绍了现有的 circRNA-disease 关联预测的 4 类计算方法，并对其进行了分类整理。

(3)对比了现有的 circRNA-disease 关联预测计算方法的优缺点，并对其未来有待研究的问题进行了分析。

1 相关工作

本节首先介绍了 circRNA 的发现，随后分别对 circRNA 的理化特性、主要功能以及生物学作用进行了概述，最后探讨了 circRNA 和常见疾病的作用机理。

1.1 circRNA 的研究背景

circRNA 是一种具有连续环状结构的内源性非编码 RNA(non-coding RNA，ncRNA)分子，它由下游 5'端剪接位点和上游 3'端剪接位点之间的反向剪接事件产生，目前在多种细胞过程中具有重要作用，是 RNA 家族的新成员[24]。大多数 circRNA 是通过一种可选择的剪接机制，即“反向剪接”(Back-splicing)，或者利用线性剪接过程中产生的内含子套索(Intron lariat)作为中间结构进一步加工而成的[25]。

Sanger 等[2]在 20 世纪 70 年代首次发现了环状 RNA(circluar RNA，circRNA)，并利用电子显微镜观察到植物类病毒中含有 circRNA。尽管 circRNA 被发现的时间较早，但长期以来 circRNA 一直未引起学者们的广泛关注。然而，令人兴奋的是，高通量测序技术的迅速发展为进一步鉴定 circRNA 奠定了基础，大量的 circRNA 可以在多种细胞系和不同物种(如动物、植物)中得到有效识别，因而受到了研究人员的广泛关注[15,26]。因此，新的计算生物学方法与深度测序技术的结合，为 circRNA 物种的全面研究奠定了基础，再一次使 circRNA 成为 RNA 领域研究的前沿，这对开发 circRNA 的新功能和潜在作用有着重要意义[27]。

1.2 circRNA 的理化特性

近年来，circRNA 在越来越多的生物体内被发现，由于 circRNA 具有独特的形成方式和空间结构，使其具有多种不同的生物学特性，了解这些特性对于探究其功能至关重要。目前，CircRNA 普遍具有以下几个特性，下面对其最重要的特性进行详细阐述。

1.2.1 circRNA 具有丰富性

为了探讨 circRNA 的丰富度，文献[28]利用 RNA 深度测序技术在小鼠细胞中进行了大量的实验，研究 N6 甲基腺苷(N6-methyladenosine，m6A)在具有编码潜力的 circRNA 生物发生中的作用机制。实验结果表明，在小鼠精子发生过程中，m6A 依赖的 circRNA 聚集可以弥补精子发育后期线状 mRNA 的降解。这意味着，当晚期粗线精母细胞发育为圆形并延伸成精子细胞时，circRNA 的丰富度会提升，即随着精子发生的进展，circRNA 的表达水平逐渐升高。

1.2.2 circRNA 具有稳定性

与线性 RNA 相比，circRNA 是一种共价闭环结构，这种特定的结构使得 circRNA 在体内比线性 RNA 更稳定，这种稳定性对核酸酶和 RNA 外切酶具有更好的抵抗性[29]。为了延长线性 RNA 的表达时间，文献[30]开发了一种在体外合成外源性 circRNA 的方法，以延长线性 RNA 的蛋白表达时间。实验结果表明，外源性环状 RNA 的体外合成可以在真核细胞中高效、稳定地表达蛋白质，为线性 RNA 的替代提供了可能。

1.2.3 circRNA 具有进化保守性

先前的研究已经表明，circRNA 在哺乳动物(包括人类、小鼠和大鼠)体内都具有高度保守性[31]。文献[32]在哺乳动物的大脑中发现了人类额叶皮层中高度保守的环状 RNA，即 circSLC45A4。经过深入研究发现，circSLC45A4 基因的缺失能够促进人神经母细胞瘤细胞系自发分化。因此，circSLC45A4 对于维持哺乳动物脑神经细胞的祖细胞状态是必需的。

1.2.4 circRNA 具有组织特异性

近年来，随着高通量 RNA 测序技术以及 circRNA 相关生物信息学方法的快速发展，真核生物中已经识别出了数千种 circRNA，其中包括酵母、果蝇、古细菌、小鼠、昆虫、植物及大多数生物，并且发现它们在某些组织的表达中具有特异性[33]。文献[34]采用综合分析和实验鉴定方法，对近 1 000 个人类癌细胞系中 circRNA 的表达进行了研究，发现大多数 circRNA 仅能在一个肿瘤谱系中被识别，并且谱系特异性不依赖亲代 mRNA 基因的表达。这一结果表明，circRNA 在不同的肿瘤谱系中均有特异性的产生，并能很好地改善肿瘤异质性，这进一步提示我们：circRNA 可作为一种有效的癌症诊断标志物，且具有潜在的临床应用价值。

1.3 circRNA 的主要功能

由于 circRNA 具有丰富度高、稳定性好、进化保守性以及表达特异性等优点，再一次证明了 circRNA 是一种广泛存在于生物体剪接过程中的新产物，它并非转录噪声或简单的非功能产物，而是能在生物体内发挥重要作用的非编码 RNA 分子[35]。与此同时，这些特性也为 circRNA 提供了许多潜在的功能，本节将对 circRNA 的主要功能进行详细讨论。

1.3.1 作为 miRNA 的海绵

MicroRNA(miRNA)是一种长度约为 22 个核苷酸的进化保守单链非编码 RNA 分子，在转录后水平调节基因表达，并在疾病的发生、发展、诊断和治疗中发挥重要作用[36]。先前的研究已经证实 circRNA 是一种具有竞争性的内源性 RNA(competing endogenous RNA，ceRNA)，可以通过 miRNA 反应元件(miRNA response elements，MREs)与 miRNA 结合，在转录后水平上吸附 miRNA，从而减弱内源性 miRNA 介导的靶基因的抑制并调节靶基因的表达，这种现象被称为 miRNA 海绵效应[19,37]。近年来，随着研究的深入，在生物体中发现了许多 circRNA 具有 miRNA 海绵效应。例如，文献[38]利用 RNA-seq 技术对胃癌细胞中 circ_0004771(circNRIP1)的表达进行分析，发现敲除胃癌细胞内的 circNRIP1 可以有效地阻断胃癌细胞的增殖，即 circNRIP1 通过海绵效应抑制 miR-149-5p 的表达，从而影响 AKT1 的表达水平，促进胃癌细胞的增殖和发展，这证实了 circNRIP1 具有 miRNA 海绵的作用，同时也预示 circNRIP1 的抑制可能成为胃癌治疗的一个很有前景的靶点。

1.3.2 与 RNA 结合蛋白相互作用

RNA 结合蛋白(RNA binding proteins，RBP)是一种特殊的蛋白质，它可以与单链 RNA 或双链 RNA 结合，从而形成核糖核蛋白(ribonucleoprotein，RNP)复合物。RBP 在转录后

基因表达中扮演着重要角色，包括 mRNA 的转换、输出、修饰、翻译、剪接、定位以及多腺苷酸化等。此外，RBP 还广泛存在于动植物及微生物中，在细胞的增殖、分化、衰老、凋亡等多种生物学过程中也发挥了重要作用[39-40]。已有研究表明，RBP 对 circRNA 的降解、定位、翻译以及 circRNA 在细胞质与细胞核间的运输发挥着直接或间接的调控作用[41]。因此，circRNA 和 RBP 的结合对许多生物学过程具有重要意义。

1.3.3 调控基因的表达

随着对基因表达的分子调控的深入研究，基因调控的每个环节都变得越来越复杂。正如文献[15]所言，circRNA 可以产生大量的转录后调控因子，其中含有内含子的 circRNA 被保留在细胞核中，作为调节基因的转录。因此，circRNA 可以作为一种新的 RNA 或蛋白质的基因调控因子，在转录水平和转录后水平调控基因的表达。文献[42]利用体内 circRNA 的特异性敲除小鼠前额叶皮层中的 circHomer1a，发现 circHomer1a 调节了与突触和精神疾病相关基因的 mRNA 转录本的表达。

1.3.4 作为翻译模板

对于真核生物而言，信使核糖核酸上的帽子和尾部均具有促进翻译的作用，它们可以与翻译起始因子协同作用来刺激翻译。由于 circRNA 是一种没有 5'端和 3'端的环状结构，最初仅被认为是非编码 RNA，因此忽略了其翻译功能。近年来，随着对 circRNA 研究的不断深入，circRNA 可以作为翻译模板的这一功能也得到了越来越多的证实。文献[43]综述了环状 RNA 作为翻译模板的可行性，并对 circRNA 翻译的可能机理和复杂性进行了探讨。尽管已经发现一些具有翻译功能的 circRNA 在人类癌症中发挥着重要作用，但迄今为止关于 circRNA 编码蛋白质在人类癌症中的作用尚不清楚。

1.4 circRNA 的生物学功能

由于 circRNA 具有作为 miRNA 的海绵、与 RNA 结合蛋白相互作用、调控基因的表达、可作为翻译模板等重要功能，因此 circRNA 在机体生理、病理过程中发挥着重要作用。越来越多的研究表明，circRNA 可以作为多种疾病的预测、诊断和治疗的有效生物标志物[44]。近年来，circRNA 大量且广泛存在于人体中，并且通常在疾病的组织和发育阶段特异性地表达。同时，随着对越来越多有前景的环状 RNA 生物标志物的发现和深入研究，circRNA 在调节生物学功能以及作为一种与人类疾病的诊断和治疗有关的生物标记物方面具有巨大的潜力[45]。

1.5 circRNA 和疾病的作用

正如 1.4 节所述，circRNA 可以作为许多人类疾病的生物标志物，这些生物标志物具有稳定性、敏感性和特异性的特点，长期以来被认为是提高病人治疗效果和降低医疗费用的关键。将这些生物标记物应用于临床实践中，将有助于疾病的有效诊断、治疗以及判断患者的异常状况、提供最合适的治疗方法、监测病情的进展。因此，研究 circRNA 与疾病的关系，对于 circRNA 的临床研究及其作为疾病的潜在靶点具有重要意义，可以加快相关疾病的治疗，进一步开发新的候选药物[46-47]。本节将详细讨论 circRNA 与人类常见的六大疾病的关系。

1.5.1 circRNA 与冠状病毒

冠状病毒(Coronavirus)是由数百种病毒组成的大型病毒家族，能引起一系列疾病，包括普通感冒、严重急性呼吸综合征(Severe Acute Respiratory Syndrome，SARS)、中东呼吸综合征(Middle East Respiratory Syndrome，MERS)和新型冠状病毒肺炎(Novel Coronavirus Disease，COVID-19)等[48]。特别是目前正在经历的一场历史性的大流行病 COVID-19，给全球带来了巨大的经济损失。截至 2022 年 3 月 9 日，COVID-19 已在全球范围内感染 448 313 293 人，造成了 6 011 482 人死亡[49]。研究表明，冠状病毒和 circRNA 之间有着密切的联系。文献[50]报道了一种能够编码 SARS-CoV-2 刺突蛋白三聚体 RBD 的环状 RNA 疫苗，该疫苗可通过体外转录快速产生，无需核苷酸修饰，具有很高的稳定性。该疫苗成功地在小鼠体内触发了一种有效且持久的中和抗体，从而阻止了 SARS-CoV-2 的感染，这意味着环状 RNA 疫苗在对抗新冠病毒方面具有很高的价值。

1.5.2 circRNA 与乳腺癌

乳腺癌(Breast Cancer)是全世界女性最常见的恶性肿瘤，在绝大多数国家被诊断为最常见的癌症，也是全球妇女死于癌症的主要原因。根据最近的全球癌症统计数据显示，大约 230 万妇女患有乳腺癌，占所有癌症的 11.7%，其中因乳腺癌而死亡的人数已达 68.5 万[51]。而由于 circRNA 具有疾病诊断的特性，能使疾病在早期得到治愈，目前已有很多研究发现 circRNA 与乳腺癌的发生、发展密切相关，可作为乳腺癌的生物标志物，为进一步的临床治疗提供了新的思路[52]。最近的文献[53]对乳腺癌中恶性程度最高的三阴性乳腺癌进行了全面综述，发现 circRNA 与三阴性乳腺癌密切相关，circRNA 可作为三阴性乳腺癌的潜在治疗靶点，许多 circRNA 参与了三阴性乳腺癌的诊断、预后和治疗，有助于科学实验研究向临床应用的转化。因此，各种迹象表明 circRNA 作为诊断乳腺癌的生物标记物具有很大的应用前景。

1.5.3 circRNA 与阿尔茨海默病

阿尔茨海默病(Alzheimer's Disease，AD)是一种最常见的由脑部老化引起的痴呆症，它包括人的记忆、行为和思维能力的丧失，并伴随着衰老过程，这也是导致全球数百万老年人痴呆的主要原因。遗憾的是，到目前为止，还没有一种可靠的治疗方法能有效阻止阿尔茨海默病的发生，这一复杂难题仍有待于进一步的研究[54]。近年来越来越多的研究发现 circRNA 表达与 AD 严重性的神经病理指标和临床指标密切相关。例如，文献[55]从 AD 患者和非 AD 患者中获得了 RNA-seq 数据，以量化皮质 circRNA 表达。结果表明，circRNA 表达与 AD 诊断、神经病理严重程度以及临床痴呆严重程度有显著相关性，这一发现意味着 circRNA 可作为预测和诊断 AD 的一个有效的生物标志物，还可能是其他神经系统变性疾病的潜在生物标志物。

1.5.4 circRNA 与肺癌

肺癌(Lung Cancer)是一种常见的恶性肿瘤，这也是男性癌症发病率和死亡率增加的主要原因。尽管临床上对肺癌的治疗多采用放疗、化疗、手术切除，但这并不能降低肺癌的发病率，每年的死亡率仍然很高，因此亟需创新的治疗方法。

然而，已有的生物学实验表明，肺癌组织中存在着大量不同的 circRNA，可以利用特定的 circRNA 作为肺癌的生物标志物来进行鉴定，这对于改善肺癌的治疗和预后具有重要意义。文献[56]的细胞实验表明，circSLC25A16 在非小细胞肺癌中有促进糖酵解和增殖的作用。机理实验表明，circSLC25A16 与 miR-488-3p/HIF-1α 相互作用，通过促进乳酸脱氢酶 A 的转录而激活乳酸脱氢酶 A。实验结果揭示了非小细胞肺癌的潜在发病机理，为非小细胞肺癌的精确治疗提供了新思路。

1.5.5 circRNA 与肝细胞癌

肝细胞癌（Hepatocelluar Carcinoma，HCC）是一种由肝组织形成的癌症。根据 2020 年的癌症统计，肝细胞癌被认为是全球第六大最常见的癌症类型，其中约有 906000 个新发肝癌病例和 830000 个肝癌死亡病例，这也是肝硬化患者死亡的主要原因[51]。虽然临床上有多种治疗肝细胞癌的方法，如局部消融、手术切除、肝移植等，但是术后的复发率仍然很高，因此治疗效果并不理想[57]。许多研究表明，circRNA 与肝癌的发生和侵袭有关，尽管其作用尚不明确，但已能反映其作为肝细胞癌诊断和预后生物标志物的潜力。最近，文献[58]为了阐明 circIPO11 在维持肝癌起始细胞中的作用，利用 circRNA 阵列从 3 个 HCC 样本中鉴定出在人和小鼠体内具有高度同源性的 circRNA。研究表明，小鼠 circIPO11 基因的敲除会明显影响肿瘤干细胞自我更新的能力并抑制肿瘤的扩散。总之，该发现为肝癌患者的治疗提供了一个新的潜在靶点，使得环状 RNA 可以作为 HCC 的生物标志物，能够为肝癌提供更精确的诊断和治疗方法。

1.5.6 circRNA 与脑部肿瘤

脑部肿瘤（Brain Tumor）是儿童和成人中常见的实体肿瘤，其以不自然、不受控制的脑细胞生长为主，是儿童和成人癌症死亡的主要原因。尽管已采取一些措施，如神经外科手术、放疗、化疗相结合等，但是脑部肿瘤的治疗方法并未发生太大变化，而且还会产生严重的副作用。近年来，越来越多的研究表明，环状 RNA 在神经细胞和神经胶质细胞中含量丰富，提示其与脑胶质瘤等神经系统疾病密切相关[59]。文献[60]证实了 CircNFIX 在胶质瘤患者体内的表达，发现替莫唑胺耐药患者血清中 CircNFIX 表达显著增高，其高表达可能与替莫唑胺耐药和胶质瘤患者预后不良有关。此外，利用 ROC 曲线评估了 CircNFIX 在脑胶质瘤诊断中的价值，发现 ROC 曲线下面积为 0.855，充分显示了 CircNFIX 作为胶质瘤诊断标志物的潜力，这将为替莫唑胺的耐药治疗提供新的靶点。因此，这进一步说明 circRNA 在脑部肿瘤的诊断、预后和评价治疗方面具有重要价值。

2 公开数据库

本节介绍了公开的数据库，首先解释了 circRNA 和疾病的数据库的来源，然后分别介绍了 circRNA 相关数据库和疾病相关数据库，最后介绍了 circRNA 和疾病关联关系数据库。

2.1 circRNA 和疾病数据库的来源

近年来，由于转录组测序、深度测序、染色质免疫沉淀-测序等高通量测序技术的迅速发展，使得研究人员能够在许多生物体内更容易地检测和鉴定出大量的 circRNA。与此同时，越来越多的生物学实验和临床研究再次证实，许多人类常见疾病与 circRNA 的生物学功能密切相关，circRNA 可以作为多种疾病的预测、诊断和治疗的有效生物标志物，这对今后疾病的进一步治疗和开发新的候选药物至关重要。由于研究的不断深入，传统实验检测技术和临床研究已积累大量 circRNA 以及疾病相关数据，这些数据涉及到 circRNA 的名称、数量、序列信息，疾病的名称、数量，circRNA 与疾病的关系等生物学信息。因此，为了满足基因测序、circRNA 的功能设计、疾病治疗等科学应用的需要，许多国际组织和研究机构建立了大量与 circRNA 和疾病相关的公开、可靠的数据库，用于存储与 circRNA 或疾病研究相关的重要信息，这些信息以不同格式包含在许多大型数据库中，逐渐成为了全球生物医学研究者的宝贵资源，这也为未来 circRNA 或疾病的进一步深入研究提供了充足的数据来源。

2.2 与 circRNA 相关的数据库

为了推进 circRNA 不同方面的研究，目前 circRNA 的研究已经发展和完善了许多数据库。这些数据库在 circRNA 的识别、定位、特征分析以及 circRNA 与靶点的相互作用等方面得到了广泛的应用，这为研究者揭示 circRNA 的潜在机制及其在生理、病理条件下的功能作用提供了良好的机会。为了帮助科研人员更好地研究和开发 circRNA，本节对网络上有关 circRNA 的数据库进行了归纳统计，下面仅对部分主流 circRNA 数据库进行介绍，其他数据库如表 1 所列。

表 1 circRNA 相关数据库列表

Table 1 circRNA-related databases

Database	Website	Short description	Authors，year	Reference
CircBank	http://www.circbank.cn/	Provided a new naming system and included circRNA mutations，miRNA binding sites，circRNA methylation，circRNA conservation，m6A modification of circRNA，etc	Liu 等，2019	[61]
Circ2Traits	http://gyanxet-beta.com/circdb/	Stored information on disease-related miRNA-circRNA-mRNA-lncRNA interaction networks	Ghosal 等，2013	[16]
circRNADb	http://reprod.njmu.edu.cn/circrnadb	Focuses on protein coding annotation information，including genomic information，exon splicing，and predicted IRES	Chen 等，2016	[62]
CircFunBase	http://bis.zju.edu.cn/CircFunBase	Collects high-quality circRNA resources for experimental validation and computational prediction	Meng 等，2019	[63]
CIRCpedia v2	http://www.picb.ac.cn/rnomics/circpedia	Records circRNA expression comparisons，annotation，and conservation between mice and humans	Dong 等，2018	[64]

（续表）

Database	Website	Short description	Authors, year	Reference
circVAR	http://soft.bioinfo-minzhao.org/circvar	Described circRNA genes, as well as SNPs and small deletions and insertions in putative circRNA regions	Zhao 等, 2020	[65]
circBase	http://www.circbase.org/	Records circRNA information from different organisms, including circRNA transcripts, circRNA sequences, etc	Glazar 等, 2014	[66]
exoRBase 2.0	http://www.exoRBase.org	The information of circRNA, mRNA and lncRNA from human blood exosomes	Li 等, 2018	[67]
LncACTdb 2.0	http://www.bio-bigdata.net/LncACTdb/	Provided experimentally supported ceRNA interactions from different species	Wang 等, 2019	[68]
deepBase v3.0	http://rna.sysu.edu.cn/deepbase3/index.html	Collected expression data of circRNAs, sRNAs, miRNAs, and lncRNAs	Xie 等, 2021	[69]
CircRiC	https://hanlab.uth.edu/cRic/	Characterized circRNAs expression profile in 935 cancer cell lines	Ruan 等, 2019	[34]
CSCD	http://gb.whu.edu.cn/CSCD	Collected cancer-specific circRNAs and predicted MREs, RBP sites and ORFs	Xia 等, 2019	[70]
TSCD	http://gb.whu.edu.cn/TSCD	Stored tissue-specific circRNAs in human and mouse tissues	Xia 等, 2017	[71]
EVLncRNAs 2.0	https://www.sdklab-biophysics-dzu.net/EVLncRNAs2/	Provides potential relationships among miRNAs, lncRNAs, genes, proteins, and other functional elements	Zhou 等, 2021	[72]
riboCIRC	http://www.ribocirc.com	Collected experimentally verified translated circRNAs and computationally predicted ribosome-associated circRNAs	Li 等, 2021	[73]
pigcirNet	http://lnc.rnanet.org/circ/	Provided circRNA names, classifications, locations and other genome annotations for pigs	Liang 等, 2017	[74]
CropCircDB	http://deepbiology.cn/crop/	A circRNA database for maize and rice under abiotic stress condition	Wang 等, 2019	[75]
AtCircDB	http://genome.sdau.edu.cn/circRNA	A tissue-specific Arabidopsis circRNA database for plant circRNA research	Ye 等, 2019	[76]
VirusCircBase	http://www.computationalbiology.cn/ViruscircBase/home.html	Stored virus-related circular RNA data	Cai 等, 2021	[77]
SomamiR 2.0	http://compbio.uthsc.edu/SomamiR	Records information on cancer somatic mutations of miRNAs and their target sites	Bhattacharya 等, 2016	[78]
starBase v2.0	http://starbase.sysu.edu.cn/	Provides systematic identification of RNA-RNA and protein-RNA interaction networks	Li 等, 2014	[79]
GreenCircRNA	http://greencirc.cn/	Plant circRNA database acting as miRNA decoys, and containing potential networks of circRNA-miRNA-mRNA	Zhang 等, 2020	[80]
CircInteractome	https://circinteractome.nia.nih.gov	Provided for exploring RBP and miRNA binding sites on human circRNAs and identifying potential IRES	Dudekula 等, 2016	[81]
CircNet	http://circnet.mbc.nctu.edu.tw/	Collected tissue-specific circRNA expression profiles and integrated circRNA-miRNA-gene regulatory networks	Liu 等, 2016	[82]
circMine	http://www.biomedical-web.com/circmine/	A comprehensive database for analyzing and visualizing circRNA transcriptomes related to human diseases	Zhang 等, 2021	[83]
PlantCircNet	http://bis.zju.edu.cn/plantcircnet	Provides plant circRNA-miRNA-gene regulatory networks, and circRNA-related information from eight model plants	Zhang 等, 2017	[84]
PlantcircBase	http://ibi.zju.edu.cn/plantcircbase	Records a plant circRNA database	Chu 等, 2017	[26]
TransCirc	https://www.biosino.org/transcirc/	Collected multi-omics-based evidence supporting the translation potential of circRNAs	Huang 等, 2021	[85]
TRCirc	http://www.licpathway.net/TRCirc	Provided transcriptional regulation of circRNAs and TF-circRNA regulatory relationships	Tang 等, 2019	[86]
MiOncoCirc v2.0	https://mioncocirc.github.io/	Reported the first database consisting of circRNAs directly detected in tumor tissues	Josh 等, 2019	[87]
circlncRNAnet	https://github.com/smw1414/circlncRNAnet	Constructing a co-expression network of lncRNAs and circRNAs from a network perspective	Wu 等, 2018	[88]
NeuroCirc	https://voineagulab.github.io/NeuroCirc/	Reported studies for circRNA expression in human brain	Walsh 等, 2021	[89]
circ2GO	https://github.com/airbox11/circ2GO	Records a database linking circRNAs to gene function	Lyu 等, 2020	[90]
circExp	http://soft.bioinfo-minzhao.org/circexp	Provided a database for circRNA expression profiles in human cancers	Zhao 等, 2021	[91]
ViroidDB	https://viroids.org/	Collected sequences and other related information of viroid-like circRNAs	Lee 等, 2022	[92]

CircBank 数据库提供了一个新的命名系统，它从不同来源收集了超过 140000 个 circRNA，主要包括 circRNA 的染色体位置和宿主基因、circRNA 突变、miRNA 结合位点、circRNA 甲基化、circRNA 的保守性、circRNA 的 m6A 修饰等信息。

Circ2Traits 数据库构建了一个预测长非编码和环状 RNA 基因、miRNA 与蛋白质编码之间相互作用的网络，并存储了疾病相关的 miRNA-circRNA-mRNA-lncRNA 相互作用网络信息。

circRNADb 数据库精选了 32 914 个人类外显子环状 RNA，提供 circRNA 的详细信息，包括基因组序列、基因组信息、外显子剪接、内部核糖体进入位点（Internal Ribosome Entry Site，IRES）以及编码蛋白质的 circRNA。

CircFunBase 数据库收录了 7000 多个经实验证实和计算预测的 circRNA，主要包括人类、小鼠等。此外，还提供了 circRNA 基因组可视化及可视化 circRNA-miRNA 相互作用网络的功能。

CIRCpedia v2 是一个包含 6 个不同物种的 180 多个 RNA-Seq 数据集的 circRNA 数据库，重点介绍了 circRNA 的表达比较、注释以及小鼠和人类之间的保守性。

circVAR 数据库对来自不同数据库的 295 526 个 circRNA 进行了大量注释，公布了所有 circRNA 和相关基因预先计算的遗传数据，形成了 circRNA 基因、单核苷酸多态性的在线遗传资源。

circBase 是一个大规模研究和预测 circRNA 的数据库，它来自以前的出版物，包含了 circRNA 转录本和 circRNA 序列信息，目前已保存大量关于小鼠、秀丽隐杆线虫、人类和矛尾鱼生物的数据。

exoRBase 2.0 是一个包含了 circRNA，mRNA 和 lncRNA 信息的数据库，这些数据来源于人类血液外切体的 RNA 测序分析。此外，已发表的文献中的 circRNA 数据也被纳入其中。

LncACTdb 2.0 是一个提供了不同物种中实验支持的 ceRNA 相互作用的数据库，它还提供了多种在线工具，包括 LncACTFunction，LncACT-Get，LncACTBrowser，LncACT-Network 和 LncACT-Survival，用于进行功能分析、定制化分析、基因组可视化、网络图解和生存分析。

deepBase v3.0 数据库通过深入挖掘组织、癌症和外泌体的深度测序数据，对来自 14 个物种的 67619 个数据集进行了全面的 sRNAs 和 lncRNAs 表达分析，并收录了 circRNAs，miRNAs 和 lncRNAs 表达数据。

2.3 与疾病相关的数据库

当前，各种疾病严重威胁着人类的生命和健康，已成为各国政府和医学研究机构面临的共同挑战[93]。以疾病为中心的相关数据库的建立，可以使研究者获得多种疾病相关的数据信息，对于认识疾病的本质、发病机制以及采取相应的治疗措施具有重要意义[94]。为了推动疾病不同方面的研究，本节总结了目前已存在的疾病相关数据库，以帮助研究者理解和使用这些数据库。下面仅对部分主流疾病数据库进行介绍，其他数据库如表 2 所列。

表 2 疾病相关数据库列表

Table 2 disease-related databases

Database	Website	Short description	Authors，year	Reference
Disease ontology	http://www.disease-ontology.org	A database for disease semantic integration	Kibbe 等，2015	[95]
MeSH	https://www.nlm.nih.gov/mesh/	Provides an authoritative list of medical subject headings	Lipscomb，2000	[96]
OMIM	http://www.ncbi.nlm.nih.gov/omim/	Records an intersecting database of human genes and genetic diseases	Hamosh 等，2005	[97]
DISEASES	http://diseases.jensenlab.org/	Reported a text mining database of disease-gene associations	Sune 等，2015	[98]
DisGeNET	http://www.disgenet.org	A comprehensive analysis platform integrating information on genes and variants of human diseases	Pinero 等，2016	[99]
HPO	www.human-phenotype-ontology.org	Focus on human phenotype ontology，including the standardized vocabulary of disease phenotypes，disease-phenotype annotations，etc	Köhler 等，2021	[100]
DiseaseConnect	http://disease-connect.org/	Developed a web server for comprehensive analysis of shared molecular mechanisms between diseases	Liu 等，2014	[101]
SIDD	https://ngdc.cncb.ac.cn/databasecommons/database/id/4661	Provided a semantically integrated database for human diseases	Cheng 等，2013	[102]
PhenomeCentral	https://phenomecentral.org	Records a database of rare diseases developed for clinicians and researchers	Buske 等，2015	[103]
eDGAR	http://edgar.biocomp.unibo.it/gene_disease_db/	Collected information of disease-gene associations between diseases and genes	Babbi 等，2017	[104]
MalaCards	http://www.malacards.org/	A comprehensive database of human diseases with multiple clinical and genetic annotations	Rappaport 等，2017	[105]
mBodyMap	https://mbodymap.microbiome.cloud	Constructing a database involving human microbes associated with health and disease	Jin 等，2022	[106]

Disease ontology 是一个用来建立疾病数据源的生物医学数据库，并且已经发展成为针对人类常见和罕见疾病的一个标准本体。该数据库旨在通过与生物医学研究者的密切合作，为每一种疾病提供一个明确的定义，使这些疾病术语能够在生物医学界对人类疾病进行统一、可重复和可持续使用的规范命名描述，从而使它们在注释生物医学数据时能够

得到普遍使用和应用。

MeSH 是美国国家医学图书馆编著的生物医学领域非常权威的主题词表，在国际上得到了广泛应用。MeSH 术语的一个最大优势是所有 MeSH 术语都是预定义的，并且包含了同义词，可以在生物医学文献中创建统一、一致的主题索引。此外，MeSH 数据库提供了一个严格的疾病分类系统，它通过有向无环图反映疾病之间的关系。

OMIM 是及时、全面、权威的人类基因和遗传疾病知识数据库，旨在为教育、人类遗传学研究和临床遗传学实践提供数据资源。OMIM 是一个不断更新的疾病与基因的交叉数据库，它能够帮助学生、临床医生和科研人员发现基因与疾病之间的复杂关系。

DISEASES 是一个人类疾病相关数据集成库，用于从生物医学文档中挖掘与疾病有关的基因信息。该数据库包含一个高效搜索栏，用于提取人类疾病和基因的命名实体，从而将文本挖掘的结果与现有数据库中人工收集的疾病基因、癌症突变数据和全基因组数据相结合，实现从医学图谱中准确、高效地提取 disease-gene 关联数据。

2.4 circRNA 与疾病关联的数据库

正如 1.5 节所述，circRNA 与许多常见的人类疾病密切相关，许多证据已经表明，circRNA 可以作为多种疾病诊断和治疗的有效生物标志物，用于对疾病进行预后和治疗。因此，许多研究者也致力于进一步阐明其在病理条件下的作用机理。为了深入研究 circRNA 与疾病之间的关联关系，本节回顾了目前已有的 circRNA-disease 关联关系数据库，以便研究者更好地了解和使用这些数据库。表 3 列出了公开可用的 circRNA-disease 关联关系数据库。

表 3 circRNA-disease 关联关系数据库列表

Table 3 circRNA-disease association databases

Database	Website	Species	circRNAs number	Diseases/ Cancers number	Associations	Authors, year
CircR2Disease v2.0	http://bioinfo.snnu.edu.cn/CircR2Disease_v2.0	Human, Chicken, Mouse, Rat and C. elegans	3077	312	4201	Fan 等, 2021[107]
circAtlas v2.0	http://circatlas.biols.ac.cn/	Human, Chicken, Mouse and Rat	776	117	846	Wu 等, 2020[108]
Circ2Disease	http://bioinformatics.zju.edu.cn/Circ2Disease/index.html	Human	249	60	270	Yao 等, 2018[109]
circRNADisease	http://cgga.org.cn:9091/circRNADisease/	Human	332	48	350	Zhao 等, 2018[20]

CircR2Disease v2.0 是 CircR2Disease 数据库的更新版本，它增加了 circRNA-disease 关联的数量和新的特征，旨在通过建立一个完善的网络平台来揭示 circRNA 与多种疾病之间的关系。与之前的数据库相比，CircR2Disease v2.0 收录了超过其 5 倍的经实验证实的 circRNA-disease 关联关系。该数据库目前包含了 3077 个 circRNA、312 个疾病、1611 个基因以及 4201 个经实验证实的 circRNA-di-sease 关联关系数据。

此外，该数据库还包括了 circRNA-miRNA 相互作用、miRNA 靶点以及 circRNA 结合蛋白等信息。

circAtlas v2.0 整合了一个包含小鼠、大鼠、鸡、猕猴、猪和人类 6 个物种的环状 RNA 数据库，构建了 circRNA 的共表达、优先级排序和功能注释信息。根据我们统计的结果，该数据库目前有 854 个与疾病相关的 circRNA-disease 的关联信息，其中包含了 776 个 circRNA 和 117 个疾病之间的 846 个 circRNA-disease 关联关系数据。

Circ2Disease 是一个人工管理的、实验支持的人类 circRNA 与疾病关联信息数据库，每个关联数据包含 circRNA 的名称、表达模式、疾病的名称、实验方法以及 circRNA 与疾病关系的简要注释信息。根据我们统计的结果，目前该数据库中共有 273 个人工整理、经实验验证的 circRNA-disease 关联信息，其中包含 249 个 circRNA 和 60 个疾病之间的 270 个 circRNA-disease 关联关系数据。此外，该数据库还包含一个完整的 circRNA-miRNA-gene 网络可视化工具，这些信息对于进一步了解 circRNA 在疾病中的转录后调控作用有重要作用。

circRNADisease 数据库中的每个数据都是从 PubMed 数据库中手工获取的，其中包含经实验验证的 circRNA 和相关疾病的关联信息。每个条目都涉及 circRNA-disease 关联的详细信息，如 circRNA ID 或名称、circRNA 表达模式、疾病名称、实验检测技术、宿主基因、物种信息、circRNA 生物功能的简要说明、文献来源等其他注释信息。根据我们统计的结果，目前该数据库中共有 354 个人工整理、经实验验证的 circRNA-disease 关联信息，其中包含 332 个 circRNA 和 48 个疾病之间的 350 个 circRNA-disease 关联关系数据。

3 circRNA 和疾病的计算模型

准确地鉴定和证实 circRNA-disease 关联关系，不仅有助于生物学家理解疾病的致病机理，还有助于人类疾病早期的诊断、预防和治疗。截至目前，尽管已经有一些 circRNA-disease 关联数据得到了实验验证，但是这仅仅是所有关联关系数据中极少的一部分，更多潜在的 circRNA-disease 关联信息尚未被发现。由于 circRNA 在机体生长发育、疾病发生、发展中所起的重要作用尚未得到充分阐明，新的 circRNA 及其疾病的鉴定仍然是一个非常困难的任务，因此众多因素使得 circRNA-disease 关联关系的确定变得日益迫切。为了进一步深入理解 circRNA 与疾病之间的关联关系及其作用机理，研究人员收集了 circRNA 与疾病之间的相关信息，并构建了多个 circRNA-disease 关联关系数据库。这些数据库的建立是正确识别 circRNA-disease 关联关系的首要环节，同时也为 circRNA 和疾病的相关研究提供了宝贵的数据资源。然而，

由于这些高维数据集具有高维、复杂、类型多样等特点，因此我们很难对这些数据进行有效的分析，给研究人员带来了极大的挑战。即便是现在，circRNA-disease 关联的生物实验验证仍然耗时、昂贵，而且仅限于小规模的研究。因此，面对大规模、复杂的 circRNA-disease 关联关系的检测，开发有效、可靠的计算模型显得尤为重要。使用计算方法来预测潜在的 circRNA-disease 关联目前已成为计算生物学的研究热点，它不但可以为生物实验验证提供新的可能的 circRNA-disease 关联候选对象，而且还能极大地降低生物实验所需的时间和成本。

本节介绍了 circRNA 和疾病关联预测的计算模型。首先介绍了 circRNA-disease 关联预测计算模型常用的评价指标，然后介绍了一系列用于预测 circRNA-disease 关联的计算模型，并将其分为 4 类：基于矩阵变换的方法、基于神经网络的方法、基于异构网络的方法和其他计算方法。表 4 详细地列出了一些最新的用于预测 circRNA-disease 关联的计算方法。

表 4　circRNA-disease 关联预测的不同类型的计算方法列表

Table 4　Different types of computational approaches for circRNA-disease association prediction

Model type	Model names	Short description
Matrix transformation-based methods	iCDA-CMG[110], iCircDA-MF[111], NMFCDA[112], DMFMSF[113], HGRTMF[114], DMFCDA[115], MRLDC[116]	Based on the original circRNA-disease association matrix, a new circRNA-disease association matrix was reconstructed to obtain two low-dimensional important and potential feature matrices of circRNA and disease, respectively
Neural network-based methods	GCNCDA[117], CDASOR[118], IMS-CDA[119], Mudiyanselage' s method[120], KGANCDA[121], SGANRDA[122], Yang's method[123], Wang's method[124]	These methods use neural networks as the core to objectively, efficiently, and automatically extract deep and implicit feature representations of the data from the association network of circRNAs and diseases, and make predictions using deep learning algorithms
Heterogeneous network-based methods	KATZCPDA[125], RWLR[126], DWNCPCDA[127], CRPGCN[128], CDWBMS[129], PWCDA[22], BRWSP[130], KATZHCDA[131], NSL2CD[132], IBNPKATZ[133]	These methods are mainly based on the network topology constructed by circRNAs and diseases, and predict the potential association between circRNAs and diseases from some nodes or different types of path information in a given graph network
Other computational methods	ICFCDA[134], MGRCDA[135], iCricDA-LTR[136], iCDA-CGR[137], DWNN-RLS[138]	These methods are not included in the other three methods

3.1　模型的评价指标

有效的评价指标可以进一步衡量预测 circRNA-disease 关联关系的计算模型的好坏，进而对实验结果作出一个合理的评判，以验证所提出的计算模型能否准确地预测 circRNA 和疾病之间的关联关系。因此，评估不同算法的预测性能是十分必要的。

在 circRNA-disease 关联关系预测模型中常见的评价指标为：准确率（accuracy，Acc）、精确率（precision，Pre）、敏感度（sensitivity，Sen）、马修斯相关系数（Matthews Correlation Coefficient，MCC）[139]。Acc，Pre，Sen，MCC 的计算式分别为：

$$Acc=\frac{TN+TP}{TN+TP+FN+FP} \tag{1}$$

$$Pre=\frac{TP}{FP+TP} \tag{2}$$

$$Sen=\frac{TP}{TP+FN} \tag{3}$$

$$MCC=\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)\times(TN+FN)\times(TN+FP)\times(TP+FN)}} \tag{4}$$

其中，TN 是真阴性，表示负样本中模型正确识别的 circRNA-disease 非关联的数量；TP 是真阳性，表示正样本中模型正确识别的 circRNA-disease 关联的数量；FN 是假阴性，表示正样本中模型错误预测为 circRNA-disease 非关联的数量；而 FP 是假阳性，表示负样本中模型错误预测为 circRNA-disease 关联的数量。

此外，K 折交叉验证也被广泛用于评估计算模型的预测性能，其具体步骤如下。首先，将所有的 circRNA-disease 关联对划分为 K 个子集，依次将每个子集单独保存作为测试集，其余 $K-1$ 个子集作为训练集，而不是从中选择一个 circRNA-disease 关联对作为测试样本。值得注意的是，K 通常被设定为 5 或 10，即代表五折交叉验证或十折交叉验证。在对预测模型进行 K 折交叉验证后，为了直观地评估计算模型的预测性能，通常会绘制出 ROC 曲线（Receiver Operating Characteristic Curve）并计算出 ROC 曲线下的面积（Area Under the ROC Curve，AUC），这也是评价二元分类模型的主要方法[140]。

3.2　基于矩阵变换的方法

3.2.1　iCDA-CMG

文献[110]基于具有类似功能的 circRNA 往往与表型相似的疾病有关这一假设，通过整合多源先验信息，包括已知的 circRNA 与疾病关联信息、circRNA 相似性和疾病相似性信息，并构建 circRNA 和疾病的相似性网络，从该网络中分别获得 circRNA 和疾病的特征表示。最后，采用一种基于协同矩阵补全的图学习模型来预测 circRNA 和疾病之间的潜在关联，这一问题可以表述为以下目标函数：

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\hat{\boldsymbol{Y}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{H}\|_F^2+\lambda_c\sum_{i,j=1}^{N_c}\|\boldsymbol{w}_i-\boldsymbol{w}_j\|^2(\boldsymbol{S}_C^*)_{ij}+\lambda_d\sum_{i,j=1}^{N_d}\|\boldsymbol{h}_i-\boldsymbol{h}_j\|^2(\boldsymbol{S}_D^*)_{ij}$$
$$\text{s.t. } \boldsymbol{W}\geqslant 0,\boldsymbol{H}\geqslant 0 \tag{5}$$

其中，$\hat{\boldsymbol{Y}}$ 是 circRNA 与疾病之间的关联矩阵；$\boldsymbol{S}_C$ 和 $\boldsymbol{S}_D$ 分别是 circRNA 的相似性矩阵和疾病的相似性矩阵；$\boldsymbol{W}$ 和 $\boldsymbol{H}$ 是两个非负矩阵，且 $\boldsymbol{w}_i$ 和 $\boldsymbol{h}_j$ 分别表示矩阵 $\boldsymbol{W}$ 和 $\boldsymbol{H}$ 的第 i 列和第 j 列；$\|g\|_F$ 代表 Frobenius 范数；$\boldsymbol{\lambda}_c$ 和 $\boldsymbol{\lambda}_d$ 是两个正则化超参数；而 $\boldsymbol{S}_C^*$ 和 $\boldsymbol{S}_D^*$ 分别代表利用相似网络的结构信息确定的 circRNA 和疾病的稀疏相似性矩阵。然后，通过对矩阵 $\boldsymbol{W}$ 和

$\boldsymbol{H}$ 进行进一步更新，最后可以得到预测的 circRNA-disease 的关联矩阵为 $\boldsymbol{Y}_{new}=\boldsymbol{W}^{T}\boldsymbol{H}$，再利用相应的概率得分来预测 circRNA 与疾病之间的关联关系。

3.2.2 iCircDA-MF

文献[111]提出了一种名为 iCircDA-MF 的计算方法，用于预测 circRNA 与疾病之间的关联关系。该方法首先整合了多种生物学信息，即从疾病语义信息中获得的疾病相似性和 circRNA 相似性矩阵，以及已知的 gene-disease，circRNA-gene 和 circRNA-disease 的关联信息。然后，根据原始关联邻接矩阵中的邻居相互作用谱来融合更多的关联信息，以重新构建一个稀疏的 circRNA-disease 关联邻接矩阵。最后，对更新后的 circRNA-disease 关联矩阵进行非负矩阵分解，计算出关联得分，从而从稀疏矩阵中提取出重要的潜在特征。在基准数据集上的实验结果表明，该方法优于其他方法，能够有效地预测新的 circRNA 与疾病之间的关联。

3.2.3 NMFCDA

文献[112]提出了一种 NMFCDA 的计算模型，它通过将基于随机化的神经网络伪逆学习与非负矩阵分解(Non-Negative Matrix Factorization，NMF)相结合，来预测 circRNA 与疾病的关联，该模型的流程如图 2 所示。

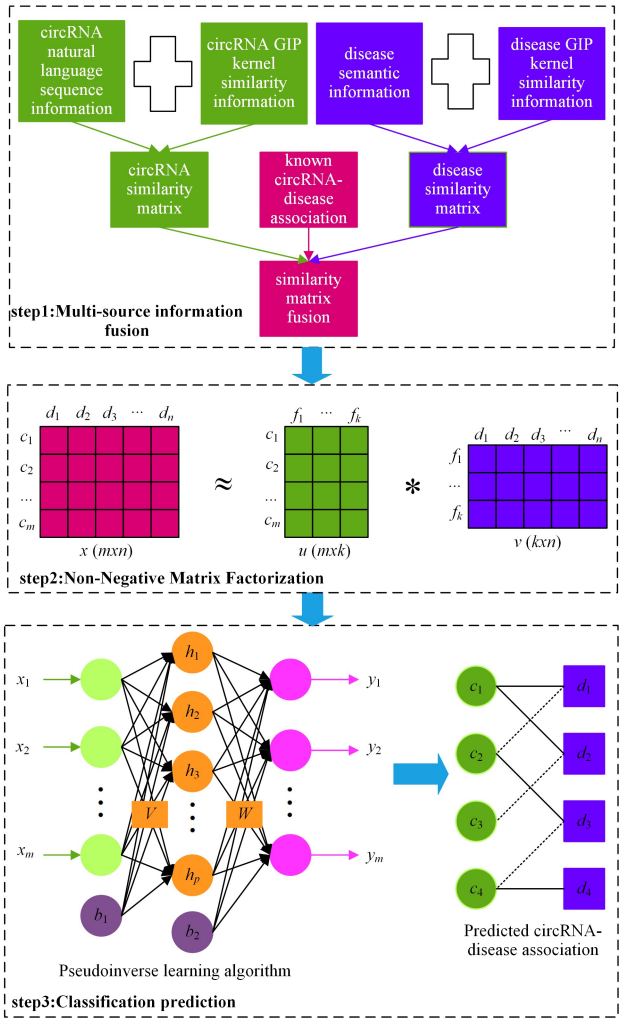

图 2 NMFCDA 模型预测 circRNA-disease 关联的流程图

Fig. 2 Flowchart of NMFCDA model predicting circRNA-disease association

具体而言，该模型首先将疾病语义信息、circRNA 的自然语言序列信息以及 circRNA 与疾病的高斯相互作用谱核相似性信息进行融合，从而使 circRNA 与疾病关联的生物学信息更加丰富。然后，在特征提取过程中，非负矩阵分解算法可以有效地提取 circRNA-disease 关联的基本特征，从而提高模型的预测性能。最后，利用随机化的伪逆学习算法来寻找全局最优解，实现对 circRNA 与疾病关联的准确预测。在 NMFCDA 模型中，NMF 问题可以进一步转化为求解以下目标函数的最小化问题：

$$\min_{U,V}\|\boldsymbol{X}-\boldsymbol{UV}\|_F^2 \quad \text{s.t. } \boldsymbol{U}\geqslant 0, \boldsymbol{V}\geqslant 0 \tag{6}$$

其中，$\boldsymbol{X}$ 是原始矩阵，$\|g\|_F$ 是 Frobenius 范数。NMF 算法的核心思想是可以将 $\boldsymbol{X}$ 分解为两个非负矩阵相乘的形式，从而获得一个重构的矩阵 $\boldsymbol{UV}$。这里，$\boldsymbol{U}$ 是基向量矩阵，$\boldsymbol{V}$ 是系数矩阵，且 $\boldsymbol{X}\approx\boldsymbol{UV}$。NMF 通过乘法迭代更新规则来最小化目标函数，其更新式为：

$$\boldsymbol{U}_{i,k}=\boldsymbol{U}_{i,k}\frac{(\boldsymbol{XV})_{i,k}}{(\boldsymbol{UVV}^{T})_{i,k}} \tag{7}$$

$$\boldsymbol{V}_{k,j}=\boldsymbol{V}_{k,j}\frac{(\boldsymbol{U}^{T}\boldsymbol{X})_{k,j}}{(\boldsymbol{UU}^{T}\boldsymbol{V})_{k,j}} \tag{8}$$

其中，$\boldsymbol{U}_{i,k}$ 和 $\boldsymbol{V}_{k,j}$ 分别表示矩阵 $\boldsymbol{U}$ 的第 i 行第 k 列的元素和矩阵 $\boldsymbol{V}$ 的第 k 行第 j 列的元素。

3.2.4 DMFMSF

文献[113]基于 circRNA 和疾病的相似性信息，提出了一种被称为 DMFMSF 的深度矩阵分解模型，用于预测 circRNA 与疾病的关联关系。首先，该模型整合了多个已知的 circRNA 和疾病的相似度信息。然后，采用了加权 K 近邻算法进行预处理，以消除缺失和未知的 circRNA-disease 关联对所提模型的影响。最后，将奇异值分解与深度矩阵分解相结合，分别提取出线性和非线性两种特征，从而推断 circRNA 与疾病之间的潜在关联。在此，基于 3 个全连接层深度矩阵分解的输出描述如下：

$$\boldsymbol{O}_{x_i}=\sigma(\boldsymbol{W}^{T}\boldsymbol{O}_{x_{i-1}}+\boldsymbol{b}_{x_i}) \tag{9}$$

$$\boldsymbol{O}_{y_i}=\sigma(\boldsymbol{W}_{y_i}\boldsymbol{O}_{y_{i-1}}+\boldsymbol{b}_{y_i}) \tag{10}$$

其中，σ 是非线性激活函数，W_{x_i} 和 W_{y_i} 表示权重矩阵，b_{x_i} 和 b_{y_i} 是偏置项。为了获得非线性特性，使用 ReLU 作为激活函数，其计算式为：

$$ReLU(x)=max(0,x) \tag{11}$$

3.3 基于神经网络的方法

3.3.1 GCNCDA

基于已知的 circRNA-disease 关联信息，结合不同疾病与 circRNA 的高斯相互作用谱核相似度信息，文献[117]提出了一种基于深度学习快速图卷积网络(FastGCN)算法的 GCNCDA 计算模型，以预测潜在的 circRNA 与疾病之间的关联。具体地，该模型通过使用 FastGCN 算法来客观地提取包含在 circRNA 与疾病融合描述符中的高层次的抽象特征，FastGCN 算法可以表示成如下的积分形式：

$$\tilde{h}^{(l+1)}(v)=\int\hat{\boldsymbol{A}}(v,u)h^{(l)}(u)\boldsymbol{W}^{(l)}\mathrm{d}P(u), h^{(l+1)}(v)=\sigma(\tilde{h}^{(l+1)}(v)) \tag{12}$$

其中，u 和 v 是具有相同概率尺度 P 的独立随机变量，函数 $h^{(l)}$ 表示来自第 l 层的嵌入函数。这里，两个连续层的嵌入函数通过卷积相互关联并通过积分变换加以表示，其中核 $\hat{\boldsymbol{A}}(v,u)$ 对应邻接矩阵 $\hat{\boldsymbol{A}}$ 的 (v,u) 元素。最终，森林 PA 分类器被用于执行预测任务，其使用属性的权重从自助抽样中生成决策树，在一个权重范围 WR 内随机产生的属性的权重被定义为：

$$WR^{\lambda}=\begin{cases}[0.0000,\mathrm{e}^{-\frac{1}{\lambda}}], & \text{if } \lambda=1\\ [\mathrm{e}^{-\frac{1}{\lambda-1}}+\rho,\mathrm{e}^{-\frac{1}{\lambda}}], & \text{if } \lambda>1\end{cases} \tag{13}$$

这里，属性的权重是根据属性在最新树中的测试级别 λ 而生成的。如果一个属性出现在根节点中，那么 $\lambda=1$；同样，如果在根节点的子节点测试到一个属性，那么 $\lambda=2$。

3.3.2 CDASOR

文献[118]基于 circRNA 的序列和疾病的本体表示信息，并结合卷积神经网络和循环神经网络设计了一个名为 CDASOR 的深度学习模型，用于预测 circRNA 和疾病的关联。特别地，在 circRNA 中，采用 k-mers 算法将其序列信息编码成低维向量，然后通过卷积神经网络进行局部特征提取，并利用双向长短期记忆网络来学习序列之间的长期依赖信息。在疾病中，将疾病本体序列化为含有本体论的语句，并通过 GloVe 算法得到每个疾病术语的本体嵌入，再利用长短期记忆网络学习这些术语间的依赖关系。最后，将各个语句表示的信息整合到关联预测的全连接层中进行分类。

3.3.3 IMS-CDA

根据功能相似的 circRNA 倾向于与表型相似的疾病相关的假设，文献[119]利用多源生物信息提出了一种用于预测潜在的 circRNA-disease 关联的 IMS-CDA 计算模型。该模型通过使用疾病的语义相似性、已知的 circRNA 和疾病关联、circRNA 和疾病的 Jaccard 相似性、circRNA 和疾病的 GIP 核相似性等信息整合了 circRNA 和疾病的相似性矩阵。然后，采用深度学习中的堆叠自动编码器算法，从融合的 circRNA 和疾病的相似性矩阵中客观、高效地提取其隐含特征。最后，旋转森林分类器用于快速准确地预测潜在的 circRNA-disease 关联，它可以描述为：

$$\boldsymbol{G}_i=\begin{bmatrix}\mathcal{N}_{i,1}^{(1)},\cdots,\mathcal{N}_{i,1}^{(R_1)} & 0 & \cdots & 0\\ 0 & \mathcal{N}_{i,2}^{(1)},\cdots,\mathcal{N}_{i,2}^{(R_2)} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \mathcal{N}_{i,k}^{(1)},\cdots,\mathcal{N}_{i,k}^{(R_k)}\end{bmatrix} \tag{14}$$

其中，$\mathcal{N}_{i,k}^{(1)},\cdots,\mathcal{N}_{i,k}^{(R_k)}$ 是矩阵的系数，$\boldsymbol{G}_i$ 表示一个稀疏的旋转矩阵。在分类过程中，给定一个未知的测试样本 x，通过决策树所产生的 $d_{i,j}(\boldsymbol{X}\boldsymbol{G}_i^{j})$ 来决定 x 所属的类别。最后，采用平均组合法来计算给定测试样本中每个类的置信度，计算式如下：

$$\theta_j(x)=\frac{1}{P}\sum_{i=1}^{P}d_{i,j}(\boldsymbol{X}\boldsymbol{G}_i^{j}) \tag{15}$$

其中，P 是决策树的个数，$\theta_j(x)$ 是所有概率的平均值。

3.3.4 Mudiyanselage's method

文献[120]提出了一种消息传递图卷积网络结构来识别 circRNA 与疾病之间的关联，该模型整合了 circRNA 序列相似性、circRNA 和疾病的 GIP 核相似性信息，建立了用于图中学习的多源相似性特征。其次，针对构建好的图结构网络，采用消息传递图卷积来聚合邻域信息，通过学习图结构来抽取出 circRNA-disease 网络中各节点具有重要意义的深层特征向量，从而利用学习到的潜在特征表示来预测 circRNA 与疾病之间的新的关联。具体地，消息传递图卷积运算的计算式如下：

$$\boldsymbol{h}_v^k=f(\boldsymbol{W}_k\sum_{u\in N(v),\boldsymbol{A}}\boldsymbol{h}_u^{k-1}) \tag{16}$$

其中，$\boldsymbol{h}_v^k$ 是第 k 个卷积层的输出，输入是 $\boldsymbol{h}_u^{k-1}$，u 表示节点 v 的邻居，$\boldsymbol{W}_k$ 是第 k 层的可学习参数，而 f 是激活函数。

3.3.5 KGANCDA

文献[121]首次将知识图注意力网络用于识别 circRNA 与疾病之间的关联。具体地，该模型首先从多个不同的数据源收集了 circRNA，miRNA，lncRNA 与疾病之间的生物关联信息，并构建了 circRNA-disease 的知识图谱。然后利用图注意力网络来获得不同实体的高质量的嵌入向量表示，在聚合所有邻居之后，$\boldsymbol{V}_{N_h}$ 表示实体 h 的嵌入，聚合公式为：

$$\boldsymbol{V}_{N_h}=\sum_{(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\in N_h}\boldsymbol{a}(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})\boldsymbol{V}_t \tag{17}$$

其中，$\boldsymbol{a}(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ 表示每个邻居的关注度，它决定了邻居传播的信息量；$\boldsymbol{V}_t$ 表示邻居尾部嵌入。最后，将 circRNA 和疾病的嵌入向量表示相结合，并将其输入到多层感知器中，以预测 circRNA 与疾病之间的潜在关联，采用多层感知器嵌入可以表示为：

$$\hat{\boldsymbol{y}}_{ij}=\boldsymbol{W}^l\sigma^l(\boldsymbol{W}^{l-1}\sigma^{l-1}(\cdots(\boldsymbol{W}^1(\boldsymbol{V}_{c_i}\parallel\boldsymbol{V}_{d_j})+\boldsymbol{b}^1)\cdots)+\boldsymbol{b}^{l-1})+\boldsymbol{b}^l \tag{18}$$

其中，σ^l 是第 l 层的激活函数，$\boldsymbol{W}^l$ 和 $\boldsymbol{b}^l$ 分别表示多层感知器中的可学习权重矩阵和偏置，$\boldsymbol{V}_{c_i}$ 代表 circRNA 各层的嵌入连接，而 $\boldsymbol{V}_{d_j}$ 则表示疾病各层的嵌入连接，$\parallel$ 是连接操作符。

3.3.6 SGANRDA

文献[122]提出了一种用于预测 circRNA 与疾病关联的半监督生成对抗网络（Generative Adversarial Network，GAN）模型，该模型的流程图如图 3 所示。在这个模型中，首先将 circRNA 序列作为一种自然语言，利用自然语言处理技术来获取 circRNA 的抽象特征。然后将 circRNA 序列的自然语言特征与两种疾病的语义相似性、circRNA 和疾病的 GIP 核相似性特征相结合，在已知的 circRNA-disease 关联信息的基础上来实现不同特征之间的互补融合。接着将所有 circRNA 和疾病的融合特征输入到半监督深度学习 GAN 中进行预训练，从而使模型能够抽取出更具表现力的特征信息。最后利用极限学习机（Extreme Learning Machine，ELM）分类器对 circRNA 与疾病之间的关联进行准确鉴定。GAN 的优化目标函数可以定义为：

$$\min_G\max_D\boldsymbol{V}(D,G)=\boldsymbol{E}_{x\sim P_{\text{data}}(x)}[\log D(x)]+\boldsymbol{E}_{\boldsymbol{z}\sim P_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-D(G(\boldsymbol{z})))] \tag{19}$$

其中，G 是一个生成模型，它接收一个随机噪声 $\boldsymbol{z}$，并通过这个噪声产生样本，记为 $G(\boldsymbol{z})$；P_{data} 是样本 x 的真实分布；D 是一个判别模型，用于判断样本是否真实，而输出 $D(x)$ 则表示 x 是真实样本的概率。最后，极限学习机被作为分类器来执行预测任务，它可以表示为：

$$\sum_{i=1}^{P}\boldsymbol{\gamma}_i f(\boldsymbol{W}_i\times\boldsymbol{X}_j+\boldsymbol{b}_i)=\boldsymbol{c}_j,j=1,\cdots,N \tag{20}$$

其中，$\boldsymbol{\gamma}_i$ 是输出权重矩阵，$\boldsymbol{W}_i=[\boldsymbol{w}_{i1},\boldsymbol{w}_{i2},\cdots,\boldsymbol{w}_{in}]^{\mathrm{T}}$ 是输入权重矩阵，$\boldsymbol{X}_j$ 表示样本属性，$\boldsymbol{W}_i\cdot\boldsymbol{X}_j$ 是 $\boldsymbol{W}_i$ 和 $\boldsymbol{X}_j$ 的内积，$\boldsymbol{b}_i$ 表示第 i 个隐藏层的偏置，$f(x)$ 代表激活函数，N 是样本的数量，$\boldsymbol{c}_j$ 表示输出。

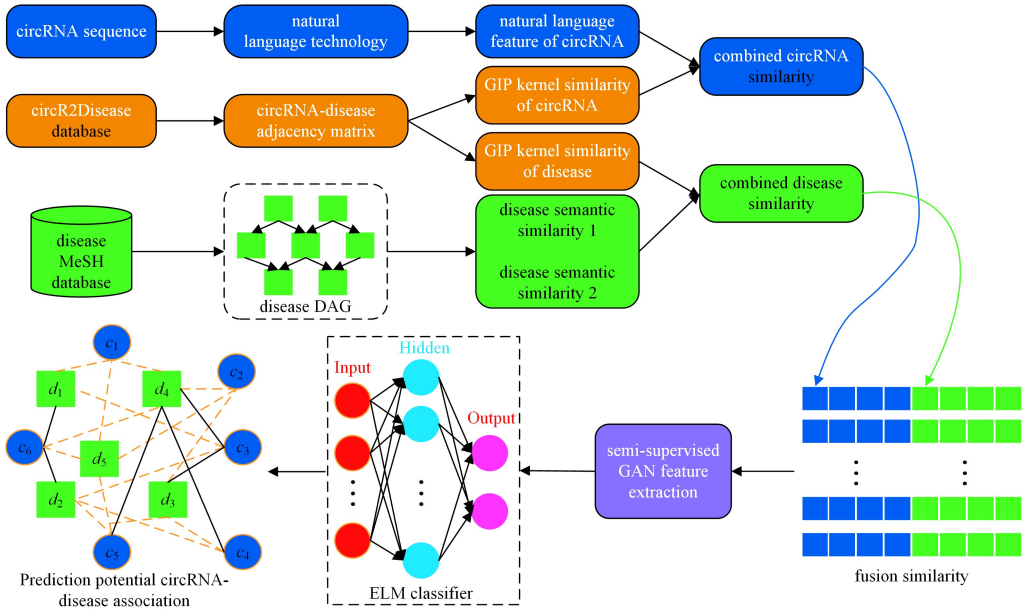

图 3 SGANRDA 模型预测 circRNA-disease 关联的流程图

Fig. 3 Flowchart of SGANRDA model predicting circRNA-disease association

3.4 基于异构网络的方法

3.4.1 KATZCPDA

基于与同一蛋白质相关的 circRNA 倾向于与蛋白质相关的疾病相关联这一假设，文献[125]收集了 3 种不同的关联网络，即 circRNA-protein 关联网络、protein-disease 关联网络以及来自 circR2Disease 数据库中的已知的 circRNA-disease 关联网络的信息，提出了一种名为 KATZCPDA 的计算模型来预测 circRNA 与疾病之间的关联，该模型的示意图如图 4 所示。然后，通过合并 circRNA 相似性网络、蛋白质相互作用网络、疾病相似性网络以及经实验验证的 circRNA-disease 关联网络，形成了多个异构信息网络。最后，通过 KATZ 算法将该关联预测问题转换成异构网络中节点间的相似度计算问题来推断 circRNA 与疾病的关联，circRNA 节点 $c(i)$ 与疾病节点 $d(j)$ 之间最终的潜在关联的定义如下：

$$\boldsymbol{S}(c(i),d(j))=\sum_{l=1}^{n}\boldsymbol{\beta}_l\boldsymbol{A}^p(i,j) \tag{21}$$

其中，n 是从 circRNA 节点 $c(i)$ 到疾病节点 $d(j)$ 的路径的最大长度；非负系数序列 $\boldsymbol{\beta}_l$ 是贡献参数；$\boldsymbol{A}^p(i,j)$ 代表从 $c(i)$ 节点到 $d(j)$ 节点的路径 p 中的行走次数。

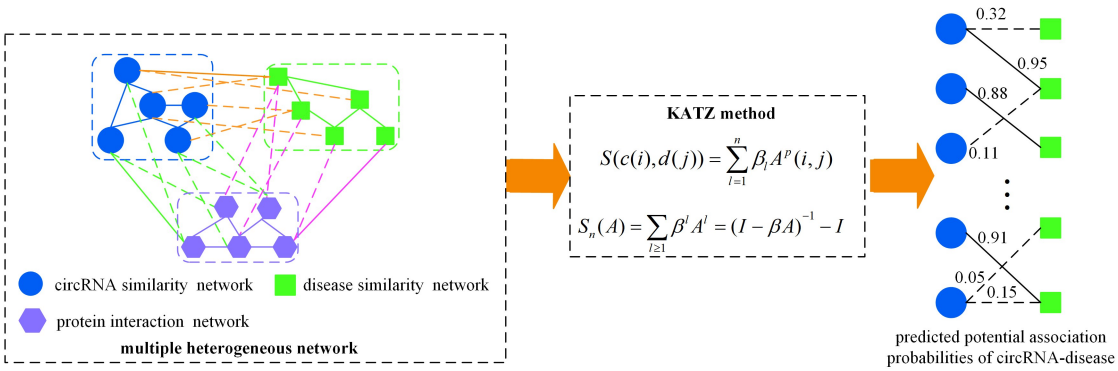

图 4 KATZCPDA 模型预测 circRNA-disease 关联的框架示意图

Fig. 4 Framework of KATZCPDA model predicting circRNA-disease association

3.4.2 RWLR

鉴于随机游走算法能够从网络节点中获得全局的特征信息，文献[126]将随机游走和逻辑回归模型相结合，提出了一种 RWLR 的计算模型，用于预测 circRNA 与疾病之间的关联。该模型首先使用 Jaccard 相似度算法计算 circRNA 之间的功能相似性，并建立了 circRNA-circRNA 相似性网络。然后将带重启的随机游走应用于 circRNA 相似网络中，获得每个 circRNA 的全局信息。随后根据随机游走达到稳定状态后的结果和 circRNA-disease 关联矩阵，提取到每个 circRNA-disease 对的特征向量，带重启的随机游走算法的定义如下：

$$\boldsymbol{p}(t+1)=(1-r)*\boldsymbol{M}*\boldsymbol{p}(t)+\boldsymbol{r}*\boldsymbol{p}(0) \tag{22}$$

其中，r 是重新启动概率；$\boldsymbol{M}$ 是转换矩阵，这里指代列归一化的 circRNA-circRNA 功能相似性矩阵；$\boldsymbol{p}(t)$ 是一个概率向量，$\boldsymbol{p}(0)$ 是初始概率向量。最终，该模型采用二元逻辑回归分类器来预测 circRNA-disease 的潜在关联，具体可以表示为：

$$\boldsymbol{P}(y=1|x)=\frac{\exp(\boldsymbol{w}\cdot\boldsymbol{x})}{1+\exp(\boldsymbol{w}\cdot\boldsymbol{x})} \tag{23}$$

$$\boldsymbol{P}(y=0|\boldsymbol{x})=\frac{1}{1+\exp(\boldsymbol{w}\cdot\boldsymbol{x})} \tag{24}$$

其中，$\boldsymbol{x}$ 是样本的特征向量，y 是样本特征所对应的标签，而 $\boldsymbol{w}$ 代表需要训练的权重向量。

3.4.3 DWNCPCDA

文献[127]将深度游走和网络一致性投影算法相结合，提出了一种名为 DWNCPCDA 的计算模型，用来预测 circRNA 与疾病之间的关联。该模型利用深度学习中的深度游走网络嵌入算法，从已知的 circRNA- disease 关联网络中学习各节点的嵌入特征，从而分别获得 circRNA 和疾病的拓扑相似性信息。然后，从网络的角度出发，采用网络一致性投影方法从 circRNA 和疾病的拓扑相似性网络中预测尚未被发现的 circRNA-disease 关联。

3.4.4 CRPGCN

文献[128]将 circRNA-disease 邻接矩阵、circRNA GIP 核相似性矩阵、基于基因的 circRNA 相似性矩阵、基于序列的 circRNA 相似性矩阵以及疾病语义相似性矩阵相融合，构建了异构邻接矩阵和异构特征矩阵，提出了一种 CRPGCN 的计算方法，用于预测 circRNA-disease 关联。该方法采用 RWR 算法使每个节点都可以从邻近的节点中获得相关性高的信息，然后利用 PCA 方法对矩阵拼接带来的不可避免的噪声进行降维和特征提取，使得具有较高相似性的 circRNA 与疾病之间的连接更加紧密。最后，图卷积网络被用来预测每个 circRNA-disease 对的相似性得分。

3.4.5 CDWBMS

文献[129]通过在异构网络上采用改进的加权带偏置的元结构来预测 circRNA 与疾病之间的关联。为了增加 circRNA 的数量，首先从 exoRBase 数据库获得了 1 511 个 circRNA 的表达谱数据；同时，通过整合其他 4 个 circRNA-disease 关联数据库，得到了 1 511 个已知的 circRNA 与疾病的关联信息。然后将 circRNA 整合相似性网络、疾病整合相似性网络和 circRNA-disease 关联网络相结合，构建了一个异构网络。最后通过在所构建的异构网络中引入一个改进的加权带偏置的元结构算法，得到了每个 circRNA-disease 关联对的预测概率。在元结构算法中，可利用如下公式来计算某个 circRNA-disease 关联对($\boldsymbol{c}_i$，$\boldsymbol{d}_j$)的预测概率：

$$M(c_i,d_j)=\sum_{n=1}^{N}\sum_{k=1}^{K}\boldsymbol{\theta}_n^k M_n^k(c_i,d_j) \tag{25}$$

其中，N 是元结构类型的数量，K 表示某个元结构类型下的元结构数量，θ 是不同元结构类型的偏置，$M_n^k(c_i,d_j)$代表在 c_i 和 d_j 之间属于第 n 个元结构类型的第 k 个元结构的贡献值。

3.5 其他计算的方法

3.5.1 ICFCDA

基于将 circRNA 视为项目、将疾病视为用户，这样就可以将不同的项目推荐给不同用户，即将不同的 circRNA 推荐给不同疾病的这一理念，文献[134]提出了一种名为 ICFCDA 的计算模型，首次将协同过滤推荐算法应用于预测 circRNA 与疾病之间的关联。该模型计算了不同 circRNA 相似性网络和多个疾病相似性网络，并分别对 circRNA 和疾病的相似性进行了整合。最后利用协同过滤的方法计算出每个 circRNA-disease 对的推荐得分。

3.5.2 MGRCDA

文献[135]将关联预测问题视为一个系统推荐问题，从推荐系统的角度提出了一种名为 MGRCDA 的计算模型，用于预测 circRNA-disease 之间的关联。具体而言，该模型首先利用 circRNA 和疾病相似性所构建的异构生物网络设计了一系列元图，将元路径作为元图的特殊情况并从中进行学习。然后利用元图推荐算法在异构生物网络中进行迭代搜索，计算出不同类型元图对同一 circRNA-disease 对的贡献。最后将各元图所做的贡献累加，并将其作为 circRNA-disease 对的最终得分，以此来判断它们之间是否存在关联。这里，通过迭代搜索并累加所有元图的贡献，得到 circRNA 和疾病的关联预测得分 PS，可以表示为：

$$PS(c_i,d_j)=\sum_{u=1}^{n}\sum_{r=1}^{m}J_u^r PS_u^r(c_i,d_j) \tag{26}$$

其中，n 是元图类型的数量；m 是 circRNA-disease 对所包含的元图的数量；J 是元图的权重参数，取值范围为[0，1]；$S_u^r(c_i,d_j)$表示第 u 个元图在第 r 次迭代中的贡献值。

3.5.3 iCricDA-LTR

基于 circRNA-disease 关联识别任务与主题-文档对搜索任务之间的相似性，文献[136]提出了一种名为 iCricDA-LTR 的计算模型，用于预测 circRNA 与疾病之间的关联。在该模型中，为了增强 circRNA-disease 对的特征描述能力，作者构建了 6 种不同的基于机器学习的模型来计算 circRNA-disease 对的评分；此外，还计算了疾病的属性特征。最后，针对不同的预测器和特征表示，采用一种学习排序算法对 circRNA-disease 关联对进行排序，以捕获这些关联对之间的排名信息，从而发现新的 circRNA 与相关疾病之间的潜在关联。

4 分析与展望

4.1 分析

基于已有的关于 circRNA-disease 关联的研究工作及模型，本节将分别探讨基于矩阵变换的方法、基于神经网络的方法、基于异构网络的方法和其他计算的方法的优点和不足。同时，给出了 circRNA-disease 关联预测问题中目前所面临的主要挑战。

(1)基于矩阵变换的方法。基于矩阵变换的方法来预测 circRNA-disease 之间的关联，首先需要给出 circRNA 相似矩阵、疾病相似矩阵以及 circRNA-disease 关联矩阵，然后将这些相似矩阵进行整合，使 circRNA 与疾病相关的生物学信息更加丰富。最后通过对融合后的矩阵进行不同的数学变换，来将 circRNA-disease 关联预测问题进一步转化为求解多种目标函数的最优化问题。目前，根据 circRNA-disease 关联数据库的记录以及现有的 circRNA 相关的实验研究，学者们仅证实了实验支持的 circRNA-disease 关联数据，即阳性样本，而 circRNA-disease 的非关联数据并未在相关文献中有过讨论，也没有被纳入到相应的数据库中。因此，阴性样本，即 circRNA-disease 非关联数据通常是很难获得的。此时，基于

矩阵变换的方法的优点在于它不需要阴性样本，仅依赖于现有的阳性样本即可预测 circRNA-disease 关联。然而，该方法的不足之处在于，对于高维、复杂的矩阵，目标函数的参数很难确定且算法的求解效率往往不高。此外，如何选取合适的求解策略，以使不同的目标函数都能获得最优解，这也是一个极为复杂的问题。

(2)基于神经网络的方法。基于神经网络的方法来预测 circRNA-disease 之间的关联，主要是采用以神经网络为核心的深度学习算法，它能够客观、高效地从 circRNA 与疾病的关联网络中抽取出数据中更深层的隐含特征表示，并通过构建不同的分类器来预测 circRNA 和疾病的关联。该方法的优点是采用了多层次结构，能够从数据中自动学习并更新大量的神经元参数，从而提高了学习和特征表示的能力。同时，在模型训练的过程中，该类方法充分考虑了 circRNA 和疾病之间关联的先验知识，因此也能够对新的未知的 circRNA-disease 的潜在关联进行预测。此外，目前大部分的深度学习方法都是以端对端的方式进行，因此可以很容易地对整个模型进行训练和测试。然而，该方法的不足之处在于，在对整个模型的学习过程中，不仅需要提供阳性样本，还需要阴性样本。由于实验证实的 circRNA-disease 非关联数据很难获得，因此现有文献中的解决方案都是随机选取与阳性样本相同数量的非关联数据作为阴性样本。尽管在某些情况下，随机选取的阴性样本可能会对模型的预测性能有一定的影响，但是它仍是目前最常用的解决方案。此外，在目前已知的关联数据库中，经过人工验证的 circRNA-disease 关联数据相对较少，在现有的大多数有监督学习模式下，该方法很容易导致网络模型的过拟合。

(3)基于异构网络的方法。基于异构网络的方法来预测 circRNA-disease 之间的关联，首先基于 circRNA 与疾病的网络拓扑结构信息或者整合多种数据来源的生物数据来构建 circRNA-disease 的异构网络。例如，使用 circRNA-protein 关联网络、protein-disease 关联网络和已知的 circRNA-disease关联网络来构建多个异构网络。然后利用不同的网络算法，从该异构网络的各个节点中预测 circRNA 与疾病之间的潜在关联。该方法的优势在于，一方面，通过充分利用不同类型生物数据构建异构网络，融合多源异构网络信息，有效地减少了二元网络在预测 circRNA-disease 时的误报率，提高了模型的预测精度。另一方面，对于某个不存在与之相关的疾病的 circRNA，只要网络中至少有一条可行的路径，就可以从已知的异构网络中获得其潜在的相关疾病。然而，如何有效地整合来自不同异构网络的多种类型的特征是一个需要深入研究的问题。

(4)其他的计算方法。利用其他的计算方法来预测 circRNA-disease 之间的关联，主要是将 circRNA-disease 关联预测问题转化为其他问题，如转化为协同过滤、推荐系统、主体文档搜索、混沌游戏表示等来进行建模，从而预测 circRNA-disease 之间的关联。该方法的优点在于，它从不同的角度出发来解决 circRNA-disease 关联预测问题，为研究人员提供了新的思路和想法，同时也有利于计算机科学与生物信息学的交叉融合，为其他相关问题的研究提供了借鉴。然而，该方法的不足之处在于，虽然它可以引入一些流行的计算框架或者将新兴的技术用于该问题的研究中，但是所构建模型的预测效果不一定很好，这还需要研究人员不断地进行测试。

4.2 未来展望

尽管 circRNA 与疾病关联预测问题的研究已经取得许多优异的成果，但 circRNA-disease 关联预测计算模型在预测性能、模型的鲁棒性、方法的可扩展性等方面仍无法满足大规模 circRNA-disease 关联预测的需求。circRNA-disease 的关联预测问题仍然面临着很多挑战，尤其是在以下 6 个方面。

(1)circRNA 和疾病的命名尚未统一。虽然目前已经开发许多关于 circRNA 的数据库，但是每个 circRNA 数据库都有其独特的命名规则，而且对于从不同的数据库中收集的 circRNA 数据，其命名规则、表达方式、描述重点也各不相同。例如，circGFRA1 在 circRNADisease 数据库中的名称为 circGFRA1，而在 circAtlas v2.0 数据库中的名称为 hsa_circ_005239。此外，同一疾病在疾病数据库中也有多个不同的名称，它们之间也会出现同义词、大小写不一致的问题，甚至还可能发生表述矛盾的错误。如何合理地整合并充分利用这些数据信息，是当前所面临的一个难题。目前，除了传统的手工处理方式，还可以尝试采用人工智能技术来实现自动化的提取。因此，研究者在使用这些数据库时要更加谨慎，并尽可能地为今后的研究提供一个统一的命名体系。

(2)circRNA-disease 样本的类别不平衡问题。在有监督的机器学习模型中，构建计算模型的训练样本集既需要阳性样本(关联的 circRNA-disease 对)，也需要阴性样本(非关联的 circRNA-disease 对)。然而，实验证实的非关联 circRNA-disease 对通常很难获得，因此 circRNA-disease 关联预测中存在着阳性样本和阴性样本数量不平衡的问题。由于半监督机器学习方法可以很好地处理类不平衡问题，因此研究人员可以尝试在未来利用半监督学习算法，来建立新的计算模型，以预测 circRNA-disease 之间的关联。

(3)已知的 circRNA-disease 关联数量很少。目前，circRNA-disease 关联数据库中所收录的 circRNA-disease 的关联数据相对较少，而未知或潜在关联的数据占绝大多数，这不但造成了数据稀疏的问题，而且还极大地限制了现有计算模型的预测精度。在 circRNA-disease 关联中，如何合理地选择高质量的输入数据，并充分利用 circRNA-disease 的相关信息，以构建其丰富的特征描述符，对于 circRNA-disease 的关联预测至关重要。因此，未来的工作重点是收集更多的经生物学实验证实的 circRNA-disease 关联数据，以及其他相关类型的生物学数据信息。我们认为，新的 circRNA-disease 关联数据应当从专用或人工管理的数据库中获取，特别是应该舍弃掉那些不明确或表述不一致的数据，而不应该为了达到数据的广度而加以保留，这样才能获得高质量且高可信的数据信息，从而有望在未来提出更有效、更合理的计算模型，以提高 circRNA-disease 关联的预测性能。

(4)circRNA 和疾病特征表示的挑战。在大多数机器学习算法中，优秀的特征表示是模型获得良好预测结果的关键，也是所有基于计算的 circRNA-disease 关联预测模型所面临的共同挑战。在特征提取的过程中，最大的难点在于如何

才能从数据中既挖掘出浅层的特征表示，又获得更深层次的特征隐喻，从而全方位地反映出生物的隐含特征。由于不同生物个体的生物学特性各不相同，因此在进行生物医学分析时，往往需要综合不同生物和生物医学实体的特征。在未来，可以根据各种计算方法的特点，通过整合更有效的生物数据，选取合适的相似性度量指标和相应的特征提取策略，并提出不同的特征融合方法，从而获得包含丰富信息的 circRNA 与疾病之间的特征表示，这将在某种程度上提高模型的预测效果。

(5)构建更易使用的 Web 服务器。目前，研究者们已经提出多种不同的计算方法，用于预测 circRNA-disease 之间的关联。然而，易于使用的 circRNA-disease 的 Web 服务器较少。尽管在 github 等公开网站中提供了一些计算模型的源程序，但是这些源程序在大规模的 circRNA-disease 检测中，很难得到广泛的应用。因此，未来应该构建一些更易于使用的 Web 服务器，以便生物学家进一步证实 circRNA 和疾病之间的潜在关联。

(6)预测结果的可解释性。circRNA-disease 关联预测一直是目前医学和生命科学领域的研究热点。深度学习已被证实是一种有效的学习方法，可以用来预测 circRNA-disease 关联。但是在某些情况下，它却不能对所抽取的特征信息进行合理的解释，导致生物与医学研究者对其生物学意义的认识不足。因此，实验结果所获得的高预测精度不是预测模型较好的充分条件，其预测结果也应易于理解。可见，对新的预测结果进行合理的分析与解释，也是 circRNA-disease 关联预测中的一项很有挑战性的工作。尽管目前在获得潜在的候选关联时，研究人员通常会寻找实证文献进行案例分析，但是目前对任何计算中发现的问题加以解释都是一项复杂而艰巨的任务。因此，我们希望在未来能够有更多的生物学依据或生物医学方面的解释，以便生物学家们选择某种 circRNA 或疾病来进行生物实验验证。

综上所述，circRNA-disease 关联预测的计算模型还有待进一步的完善和优化，以实现更高的预测精度、更好的泛化能力和更广泛的适用性。通过解决这些未知问题，circRNA 有望成为治疗各种病理疾病的一种新的诊断方式，这将会是一项令人振奋且富有挑战性的研究课题。

结束语 生物信息学和计算生物学的迅猛发展使生命科学领域进入了大数据时代。随着高通量测序技术的不断发展与更新，circRNA 的发现与识别日益引起人们的广泛关注。大量的研究结果表明，circRNA 与多种复杂疾病的发生与发展息息相关，是疾病的一种新的生物标志物。因此，研究 circRNA 与复杂疾病之间的关联对于疾病的早期诊断、预后和治疗具有重要的意义。本文从 circRNA 的发展历程、理化特性、主要功能以及 circRNA 与常见疾病的关系方面阐述了 circRNA 与疾病的重要性，并指出了计算模型预测 circRNA-disease 关联的必要性。然后，对 circRNA 与疾病之间的公开数据库进行了归纳和整理，并对当前 4 种 circRNA 与疾病的计算方法进行了详细的梳理。最后分析了 4 种计算方法在 circRNA-disease 关联预测中的优缺点，并对其未来可能的研究方向进行了探讨和展望。

参 考 文 献

[1] ZHANG G Q, LI Y X, WANG Z F, et al. New Challenges and Trends in Bio-Med Big Data[J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 853-860.

[2] SANGER H L, KLOTZ G, RIESNER D, et al. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures[C]// Proceedings of the National Academy of Sciences. 1976: 3852-3856.

[3] HSU M T, COCA-PRADOS M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells [J]. Nature, 1979, 280(5720): 339-340.

[4] KOS A, DIJKEMA R, ARNBERG A, et al. The hepatitis delta (δ) virus possesses a circular RNA[J]. Nature, 1986, 323(6088): 558-560.

[5] NIGRO J M, CHO K R, FEARON E R, et al. Scrambled exons [J]. Cell, 1991, 64(3): 607-613.

[6] CAPEL B, SWAIN A, NICOLIS S, et al. Circular transcripts of the testis-determining gene Sry in adult mouse testis[J]. Cell, 1993, 73(5): 1019-1030.

[7] DUBIN R A, KAZMI M A, OSTRER H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript [J]. Gene, 1995, 167(1/2): 245-248.

[8] CHEN C Y, SARNOW P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs[J]. Science (New York, NY), 1995, 268(5209): 415-417.

[9] PERRIMAN R, ARES M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo[J]. Rna, 1998, 4(9): 1047-1054.

[10] SUZUKI H, ZUO Y, WANG J, et al. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing[J]. Nucleic Acids Research, 2006, 34(8): e63.

[11] BURD C E, JECK W R, LIU Y, et al. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk[J]. PLoS Genetics, 2010, 6(12): e1001233.

[12] DANAN M, SCHWARTZ S, EDELHEIT S, et al. Transcriptome-wide discovery of circular RNAs in Archaea[J]. Nucleic aAids Research, 2012, 40(7): 3131-3142.

[13] SALZMAN J, GAWAD C, WANG P L, et al. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types[J]. PloS One, 2012, 7(2): e30733.

[14] HANSEN T B, JENSEN T I, CLAUSEN B H, et al. Natural RNA circles function as efficient microRNA sponges[J]. Nature, 2013, 495(7441): 384-388.

[15] MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495(7441): 333-338.

[16] GHOSAL S, DAS S, SEN R, et al. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits[J]. Frontiers in Genetics, 2013, 4: 283.

[17] ZHANG X O, WANG H B, ZHANG Y, et al. Complementary sequence-mediated exon circularization[J]. Cell, 2014, 159(1):

134-147.

[18] CONN S J, PILLMAN K A, TOUBIA J, et al. The RNA binding protein quaking regulates formation of circRNAs[J]. Cell, 2015, 160(6): 1125-1134.

[19] PAMUDURTI N R, BARTOK O, JENS M, et al. Translation of circRNAs[J]. Molecular Cell, 2017, 66(1): 9-21.

[20] ZHAO Z, WANG K, WU F, et al. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations[J]. Cell Death & Disease, 2018, 9(5): 1-2.

[21] FAN C, LEI X, FANG Z, et al. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases[J/OL]. https://doi.org/10.1093/database/bay044.

[22] LEI X, FANG Z, CHEN L, et al. PWCDA: path weighted method for predicting circRNA-disease associations[J]. International Journal of Molecular Sciences, 2018, 19(11): 3410.

[23] CAI Z, LU C, HE J, et al. Identification and characterization of circRNAs encoded by MERS-CoV, SARS-CoV-1 and SARS-CoV-2[J]. Briefings in Bioinformatics, 2021, 22(2): 1297-1308.

[24] ZHANG Y, ZHANG X O, CHEN T, et al. Circular intronic long noncoding RNAs[J]. Molecular Cell, 2013, 51(6): 792-806.

[25] CHEN L L. The expanding regulatory mechanisms and cellular functions of circular RNAs[J]. Nature Reviews Molecular Cell Biology, 2020, 21(8): 475-490.

[26] CHU Q, ZHANG X, ZHU X, et al. PlantcircBase: a database for plant circular RNAs[J]. Molecular Plant, 2017, 10(8): 1126-1128.

[27] LEI X J, ZHANG W X, LIU L. Prediction of circRNA-disease associations based on multiple biological data[J]. Scientia Sinica Informationis, 2021, 51(6): 927-939.

[28] TANG C, XIE Y, YU T, et al. m6A-dependent biogenesis of circular RNAs in male germ cells[J]. Cell Research, 2020, 30(3): 211-228.

[29] ENUKA Y, LAURIOLA M, FELDMAN M E, et al. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor[J]. Nucleic Acids Research, 2016, 44(3): 1370-1383.

[30] WESSELHOEFT R A, KOWALSKI P S, ANDERSON D G. Engineering circular RNA for potent and stable translation in eukaryotic cells[J]. Nature Communications, 2018, 9(1): 1-10.

[31] RYBAK-WOLF A, STOTTMEISTER C, GLAŽAR P, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed[J]. Molecular Cell, 2015, 58(5): 870-885.

[32] SUENKEL C, CAVALLI D, MASSALINI S, et al. A highly conserved circular RNA is required to keep neural cells in a progenitor state in the mammalian brain[J]. Cell Reports, 2020, 30(7): 2170-2179.

[33] WESTHOLM J O, MIURA P, OLSON S, et al. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation[J]. Cell Reports, 2014, 9(5): 1966-1980.

[34] RUAN H, XIANG Y, KO J, et al. Comprehensive characterization of circular RNAs in ~1000 human cancer cell lines[J]. Genome Medicine, 2019, 11(1): 1-14.

[35] CUI Q H. Function of non-coding RNAs and relationship to disease[J]. Progress in Physiological Sciences, 2016, 47(3): 161-161.

[36] BARTEL D P. MicroRNAs: genomics, biogenesis, mechanism, and function[J]. Cell, 2004, 116(2): 281-297.

[37] THOMSON D W, DINGER M E. Endogenous microRNA sponges: evidence and controversy[J]. Nature Reviews Genetics, 2016, 17(5): 272-283.

[38] ZHANG X, WANG S, WANG H, et al. Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway[J]. Molecular Cancer, 2019, 18(1): 1-24.

[39] KEENE J D. RNA regulons: coordination of post-transcriptional events[J]. Nature Reviews Genetics, 2007, 8(7): 533-543.

[40] LI X, LIU C X, XUE W, et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection[J]. Molecular Cell, 2017, 67(2): 214-227.

[41] HUANG A, ZHENG H, WU Z, et al. Circular RNA-protein interactions: functions, mechanisms, and identification[J]. Theranostics, 2020, 10(8): 3503-3517.

[42] ZIMMERMAN A J, HAFEZ A K, AMOAH S K, et al. A psychiatric disease-related circular RNA controls synaptic gene expression and cognition[J]. Molecular Psychiatry, 2020, 25(11): 2712-2727.

[43] LEI M, ZHENG G, NING Q, et al. Translation and functional roles of circular RNAs in human cancer[J]. Molecular Cancer, 2020, 19(1): 1-9.

[44] LI Y, ZHENG Q, BAO C, et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis[J]. Cell Research, 2015, 25(8): 981-984.

[45] VAN ZONNEVELD A J, KOLLING M, BIJKERK R, et al. Circular RNAs in kidney disease and cancer[J]. Nature Reviews Nephrology, 2021, 17(12): 814-826.

[46] POSTE G. Bring on the biomarkers [J]. Nature, 2011, 469(7329): 156-157.

[47] HE A T, LIU J, LI F, et al. Targeting circular RNAs as a therapeutic approach: current strategies and challenges[J]. Signal Transduction and Targeted Therapy, 2021, 6(1): 1-14.

[48] SINGHAL T. A review of coronavirus disease-2019(COVID-19)[J]. The Indian Journal of Pediatrics, 2020, 87(4): 281-286.

[49] WHO. WHO Coronavirus Disease(COVID-19) Dashboard[EB/OL]. https://covid19.who.int/.

[50] QU L, YI Z, SHEN Y, et al. Circular RNA vaccines against SARS-CoV-2 and emerging variants[J]. Cell, 2022, 185(10): 1728-1744. e16.

[51] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2021, 71(3): 209-249.

[52] JAHANI S, NAZERI E, MAJIDZADEH-A K, et al. Circular RNA; a new biomarker for breast cancer: A systematic review[J]. Journal of Cellular Physiology, 2020, 235(7/8): 5501-5510.

[53] TIAN T, ZHAO Y, ZHENG J, et al. Circular RNA: A potential

diagnostic，prognostic，and therapeutic biomarker for human triple-negative breast cancer [J]. Molecular Therapy-Nucleic Acids，2021，26：63-80.

[54] BONDI M W，EDMONDS E C，SALMON D P. Alzheimer's disease：past，present，and future[J]. Journal of the International Neuropsychological Society，2017，23(9/10)：818-831.

[55] DUBE U，DEL-AGUILA J L，LI Z，et al. An atlas of cortical circular RNA expression in Alzheimer disease brains demonstrates clinical and pathological associations[J]. Nature Neuroscience，2019，22(11)：1903-1912.

[56] SHANGGUAN H，FENG H，LV D，et al. Circular RNA circSLC25A16 contributes to the glycolysis of non-small-cell lung cancer through epigenetic modification[J]. Cell Death & Disease，2020，11(6)：1-10.

[57] YANG J D，HAINAUT P，GORES G J，et al. A global view of hepatocellular carcinoma：trends，risk，prevention and management[J]. Nature Reviews Gastroenterology & Hepatology，2019，16(10)：589-604.

[58] GU Y，WANG Y，HE L，et al. Circular RNA circIPO11 drives self-renewal of liver cancer initiating cells via Hedgehog signaling[J]. Molecular Cancer，2021，20(1)：1-20.

[59] SUN J，LI B，SHU C，et al. Functions and clinical significance of circular RNAs in glioma[J]. Molecular Cancer，2020，19(1)：1-18.

[60] DING C，YI X，WU X，et al. Exosome-mediated transfer of circRNA CircNFIX enhances temozolomide resistance in glioma [J]. Cancer Letters，2020，479：1-12.

[61] LIU M，WANG Q，SHEN J，et al. Circbank：a comprehensive database for circRNA with standard nomenclature[J]. RNA Biology，2019，16(7)：899-905.

[62] CHEN X，HAN P，ZHOU T，et al. circRNADb：a comprehensive database for human circular RNAs with protein-coding annotations[J]. Scientific Reports，2016，6(1)：1-6.

[63] MENG X，HU D，ZHANG P，et al. CircFunBase：a database for functional circular RNAs[J/OL]. https://doi.org/10.1093/database/baz003.

[64] DONG R，MA X K，LI G W，et al. CIRCpedia v2：an updated database for comprehensive circular RNA annotation and expression comparison[J]. Genomics，Proteomics & Bioinformatics，2018，16(4)：226-233.

[65] ZHAO M，QU H. circVAR database：genome-wide archive of genetic variants for human circular RNAs[J]. BMC Genomics，2020，21(1)：1-8.

[66] GLAŽAR P，PAPAVASILEIOU P，RAJEWSKY N. circBase：a database for circular RNAs[J]. Rna，2014，20(11)：1666-1670.

[67] LI S，LI Y，CHEN B，et al. exoRBase：a database of circRNA，lncRNA and mRNA in human blood exosomes[J]. Nucleic Acids Research，2018，46(D1)：D106-D112.

[68] WANG P，LI X，GAO Y，et al. LncACTdb 2.0：an updated database of experimentally supported ceRNA interactions curated from low-and high-throughput experiments[J]. Nucleic Acids Research，2019，47(D1)：D121-D127.

[69] XIE F，LIU S，WANG J，et al. deepBase v3.0：expression atlas and interactive analysis of ncRNAs from thousands of deep-sequencing data[J]. Nucleic Acids Research，2021，49(D1)：D877-D883.

[70] XIA S，FENG J，CHEN K，et al. CSCD：a database for cancer-specific circular RNAs [J]. Nucleic Acids Research，2018，46(D1)：D925-D929.

[71] XIA S，FENG J，LEI L，et al. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes [J]. Briefings in Bioinformatics，2017，18(6)：984-992.

[72] ZHOU B，JI B，LIU K，et al. EVLncRNAs 2.0：an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments[J]. Nucleic Acids Research，2021，49(D1)：D86-D91.

[73] LI H，XIE M，WANG Y，et al. riboCIRC：a comprehensive database of translatable circRNAs [J]. Genome Biology，2021，22(1)：1-11.

[74] LIANG G，YANG Y，NIU G，et al. Genome-wide profiling of Sus scrofa circular RNAs across nine organs and three developmental stages[J]. DNA Research，2017，24(5)：523-535.

[75] WANG K，WANG C，GUO B，et al. CropCircDB：a comprehensive circular RNA resource for crops in response to abiotic stress [J/OL]. https://doi.org/10.1093/database/baz053.

[76] YE J，WANG L，LI S，et al. AtCircDB：a tissue-specific database for Arabidopsis circular RNAs[J]. Briefings in Bioinformatics，2019，20(1)：58-65.

[77] CAI Z，FAN Y，ZHANG Z，et al. VirusCircBase：a database of virus circular RNAs [J]. Briefings in Bioinformatics，2021，22(2)：2182-2190.

[78] BHATTACHARYA A，CUI Y. SomamiR 2.0：a database of cancer somatic mutations altering microRNA-ceRNA interactions[J]. Nucleic Acids Research，2016，44(D1)：D1005-D1010.

[79] LI J H，LIU S，ZHOU H，et al. starBase v2.0：decoding miRNA-ceRNA，miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data [J]. Nucleic Acids Research，2014，42(D1)：D92-D97.

[80] ZHANG J，HAO Z，YIN S，et al. GreenCircRNA：a database for plant circRNAs that act as miRNA decoys[J/OL]. https://doi.org/10.1093/database/baaa039.

[81] DUDEKULA D B，PANDA A C，GRAMMATIKAKIS I，et al. CircInteractome：a web tool for exploring circular RNAs and their interacting proteins and microRNAs[J]. RNA Biology，2016，13(1)：34-42.

[82] LIU Y C，LI J R，SUN C H，et al. CircNet：a database of circular RNAs derived from transcriptome sequencing data[J]. Nucleic Acids Research，2016，44(D1)：D209-D215.

[83] ZHANG W，LIU Y，MIN Z，et al. circMine：a comprehensive database to integrate，analyze and visualize human disease-related circRNA transcriptome [J]. Nucleic Acids Research，2022，50(D1)：D83-D92.

[84] ZHANG P，MENG X，CHEN H，et al. PlantCircNet：a database for plant circRNA-miRNA-mRNA regulatory networks[J/OL]. https://doi.org/10.1093/database/bax089.

[85] HUANG W，LING Y，ZHANG S，et al. TransCirc：an interac-

tive database for translatable circular RNAs based on multi-omics evidence[J]. Nucleic Acids Research, 2021, 49(D1): D236-D242.

[86] TANG Z, LI X, ZHAO J, et al. TRCirc: a resource for transcriptional regulation information of circRNAs[J]. Briefings in Bioinformatics, 2019, 20(6): 2327-2333.

[87] VO J N, CIESLIK M, ZHANG Y, et al. The landscape of circular RNA in cancer[J]. Cell, 2019, 176(4): 869-881.

[88] WU S M, LIU H, HUANG P J, et al. circlncRNAnet: an integrated web-based resource for mapping functional networks of long or circular forms of noncoding RNAs[J/OL]. https://doi.org/10.1093/gigascience/gix118.

[89] WALSH K, GOKOOL A, ALINEJAD-ROKNY H, et al. NeuroCirc: an integrative resource of circular RNA expression in the human brain[J]. Bioinformatics, 2021, 37(20): 3664-3666.

[90] LYU Y, CAUDRON-HERGER M, DIEDERICHS S. circ2GO: a database linking circular RNAs to gene function[J]. Cancers, 2020, 12(10): 2975.

[91] ZHAO M, LIU Y, QU H. circExp database: an online transcriptome platform for human circRNA expressions in cancers[J/OL]. https://doi.org/10.1093/database/baab045.

[92] LEE B D, NERI U, OH C J, et al. ViroidDB: a database of viroids and viroid-like circular RNAs[J]. Nucleic Acids Research, 2022, 50(D1): D432-D438.

[93] CHEN R S. Precision medicine is just on the road [J]. China Awards for Science and Technology, 2017, 3: 6-6.

[94] CHEN K X. Precision Medicine and Innovative Development of Traditional Chinese Medicine[J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2017, 19(1): 7-18.

[95] KIBBE W A, ARZE C, FELIX V, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data[J]. Nucleic Acids Research, 2015, 43(D1): D1071-D1078.

[96] LIPSCOMB C E. Medical subject headings(MeSH)[J]. Bulletin of the Medical Library Association, 2000, 88(3): 265-266.

[97] HAMOSH A, SCOTT A F, AMBERGER J S, et al. Online Mendelian Inheritance in Man(OMIM), a knowledgebase of human genes and genetic disorders[J]. Nucleic Acids Research, 2005, 33: D514-D517.

[98] PLETSCHER-FRANKILD S, PALLEJ A, TSAFOU K, et al. DISEASES: Text mining and data integration of disease-gene associations[J]. Methods, 2015, 74: 83-89.

[99] PI ERO J, BRAVO À, QUERALT-ROSINACH N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants[J]. Nucleic Acids Research, 2016, 45: D833-D839.

[100] KHLER S, GARGANO M, MATENTZOGLU N, et al. The human phenotype ontology in 2021[J]. Nucleic Acids Research, 2021, 49(D1): D1207-D1217.

[101] LIU C C, TSENG Y T, LI W, et al. DiseaseConnect: a comprehensive web server for mechanism-based disease-disease connections[J]. Nucleic Acids Research, 2014, 42: W137-W146.

[102] CHENG L, WANG G, LI J, et al. SIDD: a semantically integrated database towards a global view of human disease[J]. PloS One, 2013, 8(10): e75504.

[103] BUSKE O J, GIRDEA M, DUMITRIU S, et al. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases[J]. Human Mutation, 2015, 36(10): 931-940.

[104] BABBI G, MARTELLI P L, PROFITI G, et al. eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes[J]. BMC Genomics, 2017, 18(5): 25-34.

[105] RAPPAPORT N, TWIK M, PLASCHKES I, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search[J]. Nucleic Acids Research, 2017, 45: D877-D887.

[106] JIN H, HU G, SUN C, et al. mBodyMap: a curated database for microbes across human body and their associations with health and diseases[J]. Nucleic Acids Research, 2022, 50: D808-D816.

[107] FAN C, LEI X, TIE J, et al. CircR2Disease v2. 0: An Updated Web Server for Experimentally Validated circRNA-disease Associations and Its Application [J/OL]. https://doi.org/10.1016/j.gpb.2021.10.002.

[108] WU W, JI P, ZHAO F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes[J]. Genome Biology, 2020, 21(1): 1-14.

[109] YAO D, ZHANG L, ZHENG M, et al. Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease[J]. Scientific Reports, 2018, 8(1): 1-6.

[110] XIAO Q, ZHONG J, TANG X, et al. iCDA-CMG: identifying circRNA-disease associations by federating multi-similarity fusion and collective matrix completion[J]. Molecular Genetics and Genomics, 2021, 296(1): 223-233.

[111] WEI H, LIU B. iCircDA-MF: identification of circRNA-disease associations based on matrix factorization[J]. Briefings in Bioinformatics, 2020, 21(4): 1356-1367.

[112] WANG L, YOU Z H, ZHOU X, et al. NMFCDA: Combining randomization-based neural network with non-negative matrix factorization for predicting CircRNA-disease association[J]. Applied Soft Computing, 2021, 110: 107629.

[113] XIE G, CHEN H, SUN Y, et al. Predicting circRNA-Disease Associations Based on Deep Matrix Factorization with Multi-source Fusion[J]. Interdisciplinary Sciences: Computational Life Sciences, 2021, 13(4): 582-594.

[114] WANG H, TANG J, DING Y, et al. Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment [J/OL]. https://doi.org/10.1093/bib/bbaa409.

[115] LU C, ZENG M, ZHANG F, et al. Deep matrix factorization improves prediction of human circRNA-disease associations[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(3): 891-899.

[116] XIAO Q, LUO J, DAI J. Computational prediction of human disease-associated circRNAs based on manifold regularization learning framework[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 23(6): 2661-2669.

[117]WANG L,YOU Z H,LI Y M,et al. GCNCDA:A new method for predicting circRNA-disease associations based on Graph Convolutional Network Algorithm[J]. PLOS Computational Biology,2020,16(5):e1007568.

[118]LU C,ZENG M,WU F X,et al. Improving circRNA-disease association prediction by sequence and ontology representations with convolutional and recurrent neural networks[J]. Bioinformatics,2020,36(24):5656-5664.

[119]WANG L,YOU Z H,LI J Q,et al. IMS-CDA:prediction of CircRNA-disease associations from the integration of multisource similarity information with deep stacked autoencoder model[J]. IEEE Transactions on Cybernetics,2020,51:5522-5531.

[120]MUDIYANSELAGE T B,LEI X,SENANAYAKE N,et al. Graph Convolution Networks Using Message Passing and Multi-Source Similarity Features for Predicting circRNA-Disease Association[C]// Proceedings of IEEE International Conference on Bioinformatics and Biomedicine. IEEE,2020:343-348.

[121]LAN W,DONG Y,CHEN Q,et al. KGANCDA:predicting circRNA-disease associations based on knowledge graph attention network[J/OL]. https://doi.org/10.1093/bib/bbab494.

[122]WANG L,YAN X,YOU Z H,et al. SGANRDA:semi-supervised generative adversarial networks for predicting circRNA-disease associations[J/OL]. https://doi.org/10.1093/bib/bbab028.

[123]YANG J,LEI X. Predicting circRNA-disease associations based on autoencoder and graph embedding[J]. Information Sciences,2021,571:323-336.

[124]WANG L,YOU Z H,HUANG Y A,et al. An efficient approach based on multi-sources information to predict circRNA-disease associations using deep convolutional neural network[J]. Bioinformatics,2019,36(13):4038-4046.

[125]DENG L,ZHANG W,SHI Y,et al. Fusion of multiple heterogeneous networks for predicting circRNA-disease associations[J]. Scientific reports,2019,9(1):1-10.

[126]DING Y,CHEN B,LEI X,et al. Predicting novel CircRNA-disease associations based on random walk and logistic regression model[J]. Computational Biology and Chemistry,2020,87:107287.

[127]LI G,LUO J,WANG D,et al. Potential circRNA-disease association prediction using DeepWalk and network consistency projection[J]. Journal of Biomedical Informatics,2020,112:103624.

[128]MA Z,KUANG Z,DENG L. CRPGCN:predicting circRNA-disease associations using graph convolutional network based on heterogeneous network[J]. BMC Bioinformatics,2021,22(1):1-23.

[129]LEI X J,BIAN C,PAN Y. Predicting CircRNA-Disease Associations Based on Improved Weighted Biased Meta-Structure[J]. Journal of Computer Science and Technology,2021,36(2):288-298.

[130]LEI X,ZHANG W. BRWSP:predicting circRNA-disease associations based on biased random walk to search paths on a multiple heterogeneous network[J/OL]. https://doi.org/10.1155/2019/5938035.

[131]FAN C,LEI X,WU F X. Prediction of CircRNA-disease associations using KATZ model based on heterogeneous networks[J]. International Journal of Biological Sciences,2018,14(14):1950-1959.

[132]XIAO Q,FU Y,YANG Y,et al. NSL2CD:identifying potential circRNA-disease associations based on network embedding and subspace learning[J/OL]. https://doi.org/10.1093/bib/bbab177.

[133]ZHAO Q,YANG Y,REN G,et al. Integrating bipartite network projection and KATZ measure to identify novel CircRNA-disease associations[J]. IEEE Transactions on Nanobioscience,2019,18(4):578-584.

[134]LEI X,FANG Z,GUO L. Predicting circRNA-disease associations based on improved collaboration filtering recommendation system with multiple data[J/OL]. https://doi.org/10.3389/fgene.2019.00897.

[135]WANG L,YOU Z H,HUANG D S,et al. MGRCDA:Metagraph Recommendation Method for Predicting CircRNA-Disease Association[J/OL]. https://doi.org/10.1109/TCYB.2021.3090756.

[136]WEI H,XU Y,LIU B. iCircDA-LTR:identification of circRNA-disease associations based on Learning to Rank[J]. Bioinformatics,2021,37:3302-3310.

[137]ZHENG K,YOU Z H,LI J Q,et al. iCDA-CGR:Identification of circRNA-disease associations based on Chaos Game Representation[J]. PLoS Computational Biology,2020,16(5):e1007872.

[138]YAN C,WANG J,WU F X. DWNN-RLS:regularized least squares method for predicting circRNA-disease associations[J]. BMC Bioinformatics,2018,19(19):73-81.

[139]WANG L,YOU Z H,LI L P,et al. Predicting circRNA-disease associations using deep generative adversarial network based on multi-source fusion information[C]// Proceedings of IEEE International Conference on Bioinformatics and Biomedicine. IEEE,2019:145-152.

[140]WANG L,XU T,SONG C D,et al. Prediction Algorithm of Association Between miRNAs and Diseases Based on Deep Learning[J]. Acta Electronica Sinica,2020,48(5):870-877.

HU Xuegang, born in 1961, Ph.D, professor, Ph.D supervisor. His main research interests include data mining and knowledge engineering.

WANG Lei, born in 1982, Ph.D, professor, Ph.D supervisor. His main research interests include big data analysis, data mining and its applications in bioinformatics.

(责任编辑:何杨)