

中国科学技术大学

University of Science and Technology of China

硕士学位论文



论文题目 基于代谢分析抑制病毒复制的通用靶

点预测方法研究

作者姓名 齐瑜鹏

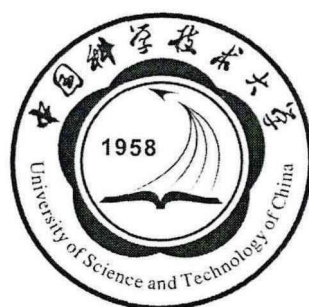
学科专业 计算机应用技术

导师姓名 郑浩然 副教授

完成时间 二〇二四年五月

中国科学技术大学

硕士学位论文



基于代谢分析抑制病毒复制的通用靶点 预测方法研究

作者姓名： 齐瑜鹏

学科专业： 计算机应用技术

导师姓名： 郑浩然 副教授

完成时间： 二〇二四年五月二十一日

University of Science and Technology of China
A dissertation for master's degree



Research on Generic Target Prediction Methods for Inhibiting Viral Replication Based on Metabolic Analyzing

Author: Yupeng Qi

Speciality: Computer Application Technology

Supervisor: Associate Prof. Haoran Zheng

Finished time: May 21, 2024

摘要

病毒种类多样,感染人体多种细胞组织,且具备极高的变异性和传染性,给人类生命健康安全带来了巨大的困扰。由于病毒不具备自身的代谢系统,一旦侵入宿主细胞,只能利用宿主细胞的代谢系统合成所需物质,以完成自身的复制和生长。因此,从代谢网络出发预测抗病毒靶点是十分有意义的。当前针对病毒性疾病代谢网络模型的研究普遍面临几个局限。首先,这些研究的分析方法只适用于特定的病毒类型,不存在通用的病毒性疾病靶点预测方法。其次,这些方法使用的代谢网络模型受到组织特异性以及原发组织特征的影响。此外,现有的靶点发现过程往往忽略了对正常细胞组织是否具有危害性的考虑。

本文提出了一种基于代谢分析的通用病毒性疾病靶点预测方法。从病毒基因表达数据以及病毒蛋白与人类蛋白之间的相互作用数据出发,分析病毒入侵宿主细胞后的异常代谢。通过消除这部分异常代谢活动,达到抑制病毒复制的目的。在对各种病毒分别进行靶点预测的同时,希望寻找到病毒性疾病的广谱性靶点,为未来的治疗策略提供创新的思路。主要工作包括:

(1) 设计并实现了一种基于基因表达数据的病毒性疾病通用靶点预测方法。通过分析病毒感染细胞与正常细胞的基因表达水平差异,重构出病毒性疾病的差异代谢网络模型,克服组织特异性与原发组织特征的影响。并将基于代谢网络模型展开的代谢分析过程建模成一个线性规划计算模型,通过调整模型中的约束条件,模拟基因敲除,毒性测试等生物学实验。寻找抑制病毒生物物质生长,但对正常细胞无毒的靶点基因。发现的抗病毒靶点中的一部分与现有文献知识吻合,其中具有广谱性的靶点 CTH 和 DCTD 也被证实与病毒共有代谢活动相关。

(2) 设计并实现了一种基于病毒蛋白与人类蛋白相互作用数据的病毒性疾病通用靶点预测方法。尝试将蛋白组学的数据整合到代谢网络模型中,使得基因表达数据缺失时,使用相互作用数据也能完成靶点预测。通过从强相互作用数据中提取代谢特征,重构出病毒异常代谢部分的代谢网络模型。将基于代谢网络模型展开的生物学分析建模成线性规划计算模型,通过调整线性规划约束,模拟基因敲除和毒性测试实验,预测无毒的抗病毒靶点。预测的靶点结果与使用基因表达数据的预测结果基本一致,且已被文献证实与病毒的致病机制相关,进一步印证了方法的可靠性。

(3) 设计并完成了一种不依赖病毒生物物质函数的通用靶点预测方法。在方法(1)的基础上,建立了基因表达水平与代谢反应通量之间一致性最高的约束计算模型。通过计算感染状态与正常状态之间的代谢通量距离,代替病毒生物物质反应通量进行病毒抑制效果的评估,从而消除靶点预测过程中对病毒生物物质函数的

依赖,克服了病毒生物物质函数构造过程中数据缺失的问题。实验结果表明不同病毒性疾病的靶点结果相似性较小,但 ALPL 在多种冠状病毒中共同表现出抗病毒性质,两者之间的相关性已被现有文献证实。

关键词: 代谢; 代谢网络; 病毒性疾病; 基因靶点; 广谱靶点

ABSTRACT

Viruses are diverse and can infect multiple human cell tissues, possessing high variability and infectiousness, presenting significant challenges to human health and safety. Since viruses do not have their own metabolic systems, they rely on the host cell's metabolic system to synthesize the necessary substances for their replication and growth once they invade the host cells. Therefore, predicting antiviral targets from the metabolic network is very meaningful. Current studies on metabolic network models for viral diseases face several limitations. Firstly, the analytical methods of these studies are only applicable to specific types of viruses and lack a universal method for predicting targets for viral diseases. Secondly, the metabolic network models used are influenced by tissue specificity and characteristics of the original tissues. Moreover, the existing target discovery processes often neglect whether they pose a risk to normal cellular tissues.

This thesis proposes a universal method for predicting targets for viral diseases based on metabolic analysis. Starting from viral gene expression data and the interactions between viral proteins and human proteins, it analyzes the abnormal metabolism after viruses invade host cells. By eliminating these abnormal metabolic activities, it aims to inhibit viral replication. While predicting targets for various viruses, it also seeks to identify broad-spectrum targets for viral diseases, providing innovative ideas for future treatment strategies. The main work includes:

(1) A universal target prediction method for viral diseases based on gene expression data has been designed and implemented. By analyzing the differences in gene expression levels between virus-infected cells and normal cells, a differential metabolic network model for viral diseases is reconstructed, overcoming the influences of tissue specificity and original tissue characteristics. The metabolic analysis process based on the metabolic network model is modeled into a linear programming calculation model, by adjusting the constraints in the model to simulate biological experiments such as gene knockouts and toxicity tests. This seeks to find target genes that inhibit viral biomass growth but are non-toxic to normal cells. Some of the antiviral targets discovered are consistent with existing literature knowledge, including broad-spectrum targets CTH and DCTD, which have also been confirmed to be related to shared metabolic activities with viruses.

(2) A universal target prediction method based on the interactions between viral

proteins and human proteins has been designed and implemented. It attempts to integrate proteomics data into the metabolic network model so that target prediction can be completed even when gene expression data is missing. By extracting metabolic features from strong interaction data, a metabolic network model of the viral abnormal metabolism is reconstructed. The biological analysis based on the metabolic network model is modeled into a linear programming calculation model, by adjusting the linear programming constraints to simulate gene knockout and toxicity testing experiments, predicting non-toxic antiviral targets. The predicted targets are generally consistent with those predicted using gene expression data and have been documented in literature as related to the virus's pathogenic mechanisms, further verifying the reliability of the method.

(3) A universal target prediction method that does not rely on viral biomass functions has been designed and completed. Building on method (1), a constrained calculation model has been established that aligns gene expression levels with metabolic reaction fluxes most consistently. By calculating the metabolic flux distance between infected and normal states instead of using viral biomass reaction flux for evaluating viral inhibition effects, the dependency on viral biomass functions in the target prediction process is eliminated, overcoming the issue of data absence in constructing viral biomass functions. Experimental results show that the similarity of target results for different viral diseases is small, but ALPL has shown antiviral properties in multiple coronaviruses, and the correlation between them has been confirmed by existing literature.

Key Words: Metabolism; Metabolic networks; Viral diseases; Gene targets; Broad-spectrum targets

目 录

第 1 章 绪论.....	1
1.1 研究背景.....	1
1.2 研究现状.....	5
1.3 本文主要工作.....	7
1.4 内容安排.....	9
第 2 章 基于基因表达数据的病毒性疾病通用靶点预测方法	11
2.1 背景及意义.....	11
2.2 数据与方法.....	12
2.2.1 数据与预处理.....	13
2.2.2 差异表达分析.....	14
2.2.3 “DELTA”网络模型重构.....	15
2.2.4 基因必要性测试.....	16
2.2.5 细胞毒性测试.....	18
2.3 实验结果.....	19
2.3.1 差异表达基因.....	19
2.3.2 “DELTA”网络模型结构与分析.....	21
2.3.3 抗病毒靶点预测结果.....	24
2.3.4 对病毒变异体的鲁棒性.....	25
2.4 本章小结.....	27
第 3 章 基于相互作用数据的病毒性疾病通用靶点预测方法	29
3.1 背景及意义.....	29
3.2 数据与方法.....	30
3.2.1 数据与预处理.....	31
3.2.2 代谢网络模型重构与抗病毒靶点预测	32
3.3 实验结果.....	34
3.3.1 相互作用相关基因.....	34
3.3.2 重构代谢网络模型结构与分析.....	35
3.3.3 抗病毒靶点预测结果.....	37
3.4 本章小结.....	40
第 4 章 不依赖 VBOF 的病毒性疾病通用靶点预测方法	41
4.1 背景及意义.....	41
4.2 数据与方法.....	42
4.2.1 代谢网络模型重构.....	42
4.2.2 基因表达水平与代谢反应通量一致性最高的约束计算模型	

构建	43
4.2.3 抗病毒靶点预测.....	45
4.3 实验结果.....	46
4.3.1 重构代谢网络模型结构与分析.....	46
4.3.2 基于约束计算模型的代谢通量分布.....	48
4.3.3 抗病毒靶点预测结果.....	49
4.4 本章小结.....	51
第 5 章 总结与展望	52
5.1 工作总结.....	52
5.2 工作展望.....	53
参考文献.....	55

插图清单

图 1.1	病毒对宿主细胞代谢重编程原理图	2
图 1.2	研究工作关系图	8
图 2.1	基于基因表达数据的通用抗病毒靶点预测方法流程图	12
图 2.2	“DELTA”网络模型重构示意图	15
图 2.3	基因敲除过程示意图	17
图 2.4	15 种病毒性疾病的差异表达基因规模	21
图 2.5	15 种病毒性疾病的差异表达基因出现频次	21
图 2.6	对 15 种病毒性疾病的差异上调基因的 GO 分析结果	22
图 2.7	对 15 种病毒性疾病的差异上调基因的 KEGG Pathway 分析结果	22
图 2.8	15 种病毒性疾病的“DELTA”网络模型规模	23
图 2.9	15 个“DELTA”网络模型的管家反应消除率结果对比图	24
图 2.10	15 种病毒性疾病的必要性基因和靶点基因规模	25
图 2.11	5 种 SARS-CoV-2 变异株 VBOF 的系数差异	26
图 2.12	对 5 种 SARS-CoV-2 变异株的序列分析结果	27
图 3.1	PPI 结构示意图	29
图 3.2	从病毒序列数据出发进行抗病毒靶点预测流程示意图	30
图 3.3	基于病毒 PPI 数据获取强相互作用代谢反应的流程图	33
图 3.4	PPI 相关的人类基因频次图	35
图 3.5	对 11 种病毒性疾病相互作用的人类基因的 GO 分析结果	36
图 3.6	对 11 种病毒性疾病相互作用的人类基因的 KEGG Pathway 分析结果	36
图 3.7	基于 PPI 数据重构的 11 个代谢网络模型规模	37
图 3.8	17 个共有代谢反应的通量分布热图	39
图 3.9	11 个代谢网络模型的必要性基因和靶点基因规模	40
图 4.1	不依赖 VBOF 的通用抗病毒靶点预测方法示意图	42
图 4.2	15 种病毒性疾病的代谢网络模型规模	47
图 4.3	管家反应消除率结果对比图	48
图 4.4	16 个共有代谢反应的通量分布热图	50

表 格 清 单

表 2.1	15 种病毒性疾病基因表达数据详情	14
表 2.2	15 个“DELTA”网络模型对管家反应的消除率详情	23
表 3.1	11 种病毒性疾病相互作用数据详情	32
表 3.2	11 个代谢网络模型共有的 17 个代谢反应	38
表 4.1	15 个代谢网络模型对管家反应的消除率详情	47
表 4.2	15 个代谢网络模型共有的 16 个代谢反应	49
表 4.3	四种冠状病毒的靶点基因预测结果	51

算 法 清 单

2.1	基于基因表达数据的通用抗病毒靶点预测算法	13
2.2	“DELTA”网络模型重构算法	16
2.3	基因必要性测试算法	18
2.4	细胞毒性测试算法	20
3.1	基于 PPI 数据的通用抗病毒靶点预测算法	31
3.2	病毒蛋白与人类蛋白相互作用数据预处理算法	32
4.1	不依赖 VBOF 的通用抗病毒靶点预测算法	43
4.2	基于约束模型的抗病毒靶点预测算法	46

第1章 绪 论

1.1 研究背景

病毒性疾病对人类生活造成了广泛而深远的影响^[1]。病毒性疾病的病原体种类繁多^[2-4]，不同类型的病毒对宿主组织的感染方式和效果各不相同。有些病毒会攻击呼吸道、消化道或神经系统等特定的组织，而另一些则可能在全身引发系统性感染^[5]。病毒的传染性和变异性是另一个令人担忧的问题。由于其高度变异的基因组，病毒能够快速适应不同的环境，并且在宿主间迅速传播^[6]。这种变异性是病毒性疾病预防和治疗中的一大挑战，因为它可能导致现有的疫苗或药物失效^[7]，甚至产生全新的致病性病毒株。这些特性使得病毒性疾病对人类的生命健康和经济生活构成严重威胁。

当前针对病毒性疾病的治疗方案通常包括预防^[8]、药物治疗^[9]和疫苗接种^[10]。研究病毒性疾病治疗药物和疫苗的原理始于对病毒生命周期和宿主细胞互动机制的深入探索。在过去的几十年里，随着基因组学^[11]和蛋白质组学^[12]等技术的发展，科学家们逐渐揭示了病毒与宿主细胞之间复杂的相互作用网络^[13]，为研发针对病毒性疾病的药物和疫苗奠定了基础。当前的病毒性疾病治疗方式主要聚焦在基因和蛋白质组学层面^[14]。研究人员通过深入了解病毒基因组结构和宿主细胞的蛋白质组学特征，试图找到干预病毒感染的关键靶点。其中，一种主要的策略是通过切断病毒和宿主细胞表面受体蛋白之间的联系来阻止病毒入侵宿主细胞，从而阻断病毒感染的初期阶段^[15-16]。此外，研究人员也致力于发现和开发能够抑制病毒复制过程的药物，以减轻病毒感染对宿主细胞的影响。例如当前针对 HIV 等逆转录病毒的药物，作用机制主要是抑制病毒的逆转录酶，阻断病毒复制过程^[17]。

然而，目前的治疗方案主要是支持性的，即针对病毒感染的症状进行缓解和控制，而非针对病毒本身的根本治疗。这是因为病毒性疾病的种类繁多，且每种病毒都具有独特的生命周期和致病机制，因此很难找到通用的治疗方案。要实现有效的治疗，需要深入了解每种病毒的详细结构和生物学特征，以开发针对特定病毒的精准治疗方案。

代谢作为生物体最基本的活性特征，在生命过程中扮演着至关重要的角色^[18]。代谢是指生物体内部通过一系列化学反应来实现生命活动所需要的能量和物质交换过程，从而维持人体内环境的稳态^[19-20]。它包括物质的合成、降解和转运等过程，是细胞内部各种生物学功能的基础。

病毒作为一种没有完整结构的亚微观粒子，主要由蛋白质外壳和遗传物质组成^[21-22]。由于不具备自身的代谢机制与酶系统，在病毒感染过程中，病毒利

用宿主细胞的代谢系统来满足其自身的复制和生长需求。病毒入侵宿主细胞后具有强大的代谢重编程能力^[23-25]，抑制了宿主细胞的正常代谢活动，使其转向更适合病毒复制生长的方向，从而引发疾病的发生（图 1.1）。例如，HIV 感染会导致宿主细胞的葡萄糖摄取增加，从而提供充足的能量和原料来支持病毒的复制和生长^[26]。一些病毒还会干扰宿主细胞的核酸和脂质代谢，以及细胞信号传导通路，从而有效地逃避宿主的免疫监视^[27-29]。

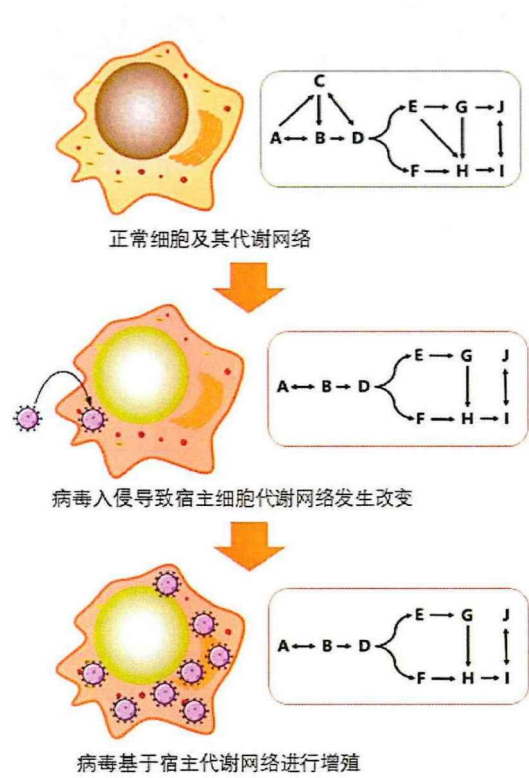


图 1.1 病毒对宿主细胞代谢重编程原理图

基于病毒入侵后对宿主细胞代谢重编程的特性，越来越多的研究展开了对病毒性疾病的代谢分析。通过分析病毒感染后宿主细胞的代谢组学信息，可以揭示病毒与宿主细胞之间的相互作用机制，发现潜在的治疗靶点。例如，针对病毒感染后宿主细胞的代谢通路进行干预，可能有助于阻断病毒的复制和生长，从而减轻病毒性疾病的严重程度。

当前，关于病毒性疾病的代谢组学研究主要集中在特定病毒的研究上，缺乏一种通用的寻找抗病毒靶点的方法^[30]。这些研究通常针对单个病毒，试图分析病毒感染后宿主细胞的代谢活动，发现治疗该特定病毒的新靶点。然而，由于不同病毒的生命周期和致病机制各不相同，这种单一病毒的研究方法并不能提供

适用于其他病毒性疾病。鉴于病毒感染宿主细胞过程中的共性特点,即病毒在感染过程中需要借助宿主细胞的代谢系统来完成其复制和生长,从代谢网络出发设计通用的病毒性疾病靶点预测方法是一个可行的方向。

当前关于病毒性疾病的数据类型众多,但存在病毒种类少,分布零散,涉及的生物学实验过程复杂等问题,使得获取关于多种病毒的全面数据十分困难。基因表达数据和病毒蛋白与人类蛋白相互作用数据是两种涉及多种病毒类型,数据形式规范且比较容易获取的数据类型,也是当前在病毒性疾病分析中使用较多的两种数据。

基因表达数据在生物学研究中扮演着重要角色,它反映了基因转录产物 mRNA 在细胞中的丰度,可以揭示基因在特定条件下的活性水平^[31]。换句话说,基因表达数据可以量化成千上万个基因的表达水平。在早期,通常使用 cDNA 微阵列^[32]和寡核苷酸芯片^[33]来获取基因表达数据。随着高通量杂交测序和基于测序技术的分子生物学实验技术的快速发展与普及^[34],第二代测序的 RNA-seq 技术^[35]开始更广泛地应用于获取基因表达数据。这种方法具有更高的通量、更低的成本和更少的噪声,并且在识别差异表达基因方面更加准确。当前已有一些基于基因表达数据来预测抗病毒靶点的研究。例如,Andrea 等人^[36]通过分析 COVID-19 患者与健康对照的外周血单个核细胞和死后肺组织的基因表达数据,发现多种趋化因子的表达在 2019 冠状病毒疾病患者的肺组织和气道中均显著增加,包括 CCL2, CCL3L1, CCL7, CCL8 和 CXCL10。这些基因也是推动 COVID-19 发病的重要原因。相关技术的成熟,为从基因表达数据出发,分析病毒性疾病的异常代谢状态提供了数据支持。

除了病毒感染后的基因表达情况,病毒蛋白作为生物学中基因的下游产物,在病毒的感染和致病机制中扮演着至关重要的角色。^[37-39]这些蛋白质不仅是病毒复制和组装的关键组成部分,而且还能通过与宿主细胞内的各种分子进行相互作用,从而精确操控宿主细胞的内部机制,以促进病毒的生存和繁殖。例如,Tsuyoshi 等人^[40]的研究发现一种新的病毒蛋白表达的核糖体移码 PA-X 在流感病毒诱导的宿主关闭中起主要作用。这种宿主关闭被认为是允许病毒逃脱宿主的抗病毒反应,使得病毒能够在宿主体内长期存活而不被清除。Yupeng 等人^[41]发现丙型肝炎病毒的非结构 5A (NS5A) 蛋白与生长因子受体结合蛋白 2(Grb2) 衔接蛋白相互作用以抑制有丝分裂信号传导,导致细胞功能失调,甚至细胞死亡,从而引起疾病的发生。因此,理解病毒蛋白与人类蛋白之间的相互作用对于揭示病毒如何感染宿主、如何影响宿主细胞的功能,以及如何导致疾病具有至关重要的意义。这种相互作用关系可以帮助科学家理解病毒的复制机制,通过分析特定的病毒蛋白与人类蛋白的结合模式,研究人员可以预测出可能的抗病毒药物靶标,这对于开发新的抗病毒策略是非常重要的。

病毒蛋白与人类蛋白之间的相互作用数据还在一定程度上揭示了病毒如何操控宿主细胞的代谢路径，促进自己的复制而不利于宿主细胞的正常功能^[42]。这种对宿主代谢的异常调控是病毒感染的一个重要方面，也是研究病毒病理机制和发展治疗策略的关键点。当前，已经有很多研究通过构建病毒蛋白与人类蛋白质之间的相互作用网络的方式，分析病毒感染后异常代谢的活动通路。例如，Gordon 等人^[43]使用亲和纯化质谱法鉴定了与每种 SARS-CoV-2 蛋白有物理关联的人源蛋白，鉴定出 332 种 SARS-CoV-2 与人源蛋白之间高置信度的蛋白-蛋白相互作用。并基于构建的相互作用网络，发现了 mRNA 翻译抑制剂、sigma-1 和 sigma-2 受体调节剂是两组具有抗病毒活性的药剂。

充分利用这些数据来分析病毒性疾病的代谢活动是一项极具挑战性的任务，也是了解疾病发病机制和发展治疗策略的关键一步。传统的生物学分析方法无法整合多组学数据，且很难发现这些数据之间的关联。系统生物学作为一门跨学科的研究领域^[44]，正是基于这些数据和技术手段而发展起来的。它将类型复杂且数量巨大的多组学实验数据整合在一起，通过建立数学模型和计算模拟，揭示生物系统的结构和功能，从而帮助研究者更全面地理解生物系统的复杂性^[45-46]。因此，通过构建计算模型来预测和验证抗病毒策略的应用越来越广泛。

基因组规模代谢模型（Genome-Scale Metabolic Models, GEMs）是一种系统生物学的重要工具，用于描述和模拟生物体内代谢网络的结构和功能^[47]。GEMs 的建立基于对生物体代谢网络的深入理解和大量生物数据的积累。通过整合基因组学、蛋白质组学和代谢组学等多组学数据，研究者可以构建出生物体的基因组规模代谢模型。GEMs 是对生物学意义上代谢网络的模拟，其本质上是与生物体代谢相关的信息集合，包括了生物体内所有已知的代谢基因、反应、代谢物的信息。其拓扑结构如图 1.1 方框中的网络所示，其中节点表示网络中的代谢物，例如葡萄糖，氨基酸等。边表示代谢物之间的代谢反应，由特定的酶进行催化。一系列的代谢反应组合起来组成代谢通路，通路中的反应共同完成一个特定的生物学功能，如能量的糖酵解路径。GEMs 能够定量描述生物体内代谢通路的结构和功能，包括各个代谢反应的反应类型，涉及的反应物与生成物以及相关的系数关系，反应速率的热力学限制等。此外，GEMs 还可以模拟生物体在不同生理条件下的代谢状态，并预测新的代谢通路和代谢产物。在生物医学领域，GEMs 被广泛用于疾病的机制研究、药物设计和个体化治疗方案的开发^[48]。通过分析疾病相关的基因组规模代谢模型，研究者可以揭示疾病的发病机制，发现新的治疗靶点，并预测药物的疗效和副作用。

从基因表达数据和蛋白-蛋白相互作用数据出发，以一种适用于所有病毒性疾病的通用方法，构建出病毒感染后宿主细胞代谢系统异常代谢部分的代谢网络模型。并通过建立计算模型展开对代谢网络模型的代谢分析，将生物学问题转

化为计算问题进行求解，从而预测潜在的抗病毒靶点，是本文的重点研究内容。同时，基于此种通用靶点预测流程，本研究期望发掘具有广谱性的病毒性疾病的靶点，为寻找或设计可复用的病毒性疾病的药物提供理论指导。

1.2 研究现状

病毒的种类多样，不同病毒入侵不同的人体组织，感染方式也各不相同。此外，病毒极强的传染性与变异性可能会使得现有药物失效，这对病毒性疾病的治疗的时效性造成较大困扰。目前关于病毒性疾病的的治疗药物和疫苗主要集中在基因组学与蛋白质组学层面。例如，SARS-CoV-2 相关疫苗预防感染的原理在于切断 SARS-CoV-2 的刺突蛋白（S 蛋白）与人体血管紧张素转化酶 2（Angiotensin-Converting Enzyme 2, ACE2）之间的联系^[49]，阻止病毒进入宿主细胞内。治疗流感病毒的特效药物奥司他韦是一种抗病毒神经氨酸酶抑制剂，用于用流感病毒 A（包括大流行的 H1N1）治疗和预防感染^[50]，通过抑制其抗病毒活性来抑制病毒神经酶的活性。其原理是在病毒的表面上，阻止了宿主细胞的出芽，从而抑制了病毒的复制和感染进程。

近年来，病毒性疾病的代谢相关研究日益增多，引起了人们的广泛关注。人们逐渐意识到，在病毒感染过程中，宿主细胞的代谢状态发生了显著变化，这种异常代谢状态与病毒的复制和传播密切相关^[23]。病毒通过干扰宿主细胞的代谢通路，改变其能量代谢、合成代谢和信号传导等关键生物学过程，为自身复制提供必要的营养和环境。Willig 等人^[26]的研究发现 HIV 感染后，宿主细胞的能量代谢途径发生改变，通常表现为对糖类代谢的增强和对脂类代谢的抑制。此外，一些病毒感染后还会影响宿主细胞的合成代谢途径和信号传导通路，以满足病毒的生长和复制需求。

值得注意的是，异常的代谢状态不仅是病毒性疾病的病理特征，还可能成为疾病治疗和预防的靶点^[51-52]。因此，越来越多的研究关注于利用代谢相关信息开发新的治疗方法和药物。通过深入理解病毒感染引起的代谢变化机制，可以针对性地设计药物，干扰病毒与宿主细胞之间的代谢相互作用，从而阻断病毒复制和传播的途径。

由于传统药物开发过程的时间周期长，成本高昂，越来越多的研究开始着手于使用计算机工具构建计算模型，用计算的方式模拟传统的生物学过程，从而在更短的时间内展开疾病致病机制的分析与治疗方案的設計工作。同时，构造计算模型的方式使研究者可以整合多组学、多类型的生物实验数据，从更全面的角度分析疾病的发病流程及基因、蛋白、代谢等上下游之间的联系。

随着系统生物学的快速发展，基因组规模代谢模型（Genome-Scale Metabolic

Models, GEMs) 作为代谢组学中的一种重要工具, 其在疾病靶点预测和致病机制方面的应用越来越受到重视。在 2007 年, 第一个人类全基因组规模的代谢网络 Recon1^[53] 被构建了出来。它是基于文献的全面基因组级别的代谢网络, 通过对基因组文献数据的人工逐项筛选, 系统级地描述了人类代谢层面的生物化学、遗传结构等信息。该模型包含了 1496 个基因、2004 个蛋白质、2233 个代谢反应和 1078 个转运反应等。此后, 相关领域的研究人员通过整合多组学数据, 陆续构建了 HMR2.0^[54], Recon2^[55], Recon3D^[56] 等更高质量的人类全基因组代谢网络, 帮助研究人员避免了单一角度研究代谢的局限性。此外, 基因组规模代谢模型的应用领域也在不断拓展。除了传统的微生物代谢工程和药物设计领域, 这些模型还被应用于疾病的机制研究、个体化治疗方案的制定和环境污染物的代谢分解等方面。通过模拟和分析生物体内代谢网络, 研究者可以揭示疾病的发病机制, 发现新的治疗靶点。

然而, 全基因组规模的代谢网络模型包含了人体所有组织的代谢信息。对于特定疾病来说, 其中存在着太多无关组织的冗余信息, 直接使用全基因组规模的代谢网络模型来进行计算机仿真实验, 会产生较大的偏差。因此需要在 GEMs 的基础之上, 重构出特定疾病的特异性代谢网络模型。目前比较常用的重构算法有 FASTCORE^[57]、iMAT^[58]、iINIT^[59] 等。这些算法根据不同的策略, 获得一个关于 GEM 的子集, 去除了其中不活跃的反应。

利用基因组规模代谢网络模型 (GEMs)^[47] 来分析病毒感染后宿主细胞的异常代谢活动是当前系统生物学中预测抗病毒靶点的一种主流方法。此种方法的重要性在于其提供了一种从分子和代谢层面理解病毒与宿主相互作用的工具, 为识别新的抗病毒药物靶点提供了科学依据。通过预测哪些代谢途径在病毒感染过程中被激活或抑制, 研究者可以确定哪些代谢途径最有可能成为抗病毒治疗的有效目标。这不仅加速了抗病毒药物的发现过程, 而且还有助于理解病毒如何操纵宿主细胞的代谢来促进其复制和传播, 从而为开发新的治疗策略提供了新的视角。例如, Renz 等人^[60-61] 在人肺泡代谢网络的基础上, 插入了 SARS-CoV-2 的生物物质函数, 并通过通量平衡分析 (Flux Balance Analysis, FBA)^[62] 和通量变异性分析 (Flux Variability Analysis, FVA)^[63] 的方式分析代谢网络模型中代谢反应与病毒生物物质生长通量之间的关系, 预测出鸟苷酸激酶是抗 SARS-CoV-2 的潜在靶点。Kishk 等人^[64] 使用 FASTCC 算法^[57] 重构了关于 SARS-CoV-2 感染后肺部组织的特异性代谢网络模型, 发现了肺组织中与 COVID-19 严重程度相关的代谢通路和氧化应激、铁变态反应以及嘧啶代谢有关。

在当前关于病毒性疾病代谢网络模型的相关研究中, 病毒的生物物质合成函数 (Virus Biomass Objective Function, VBOF) 是分析代谢网络模型中异常代谢的重要目标, 表征了病毒复制生长过程中对代谢物原料、能量等物质与环境需求。

Aller 等人^[65]的研究提供了构建 VBOF 的一种方法, 已经被很多基于 GEM 展开的病毒性疾病相关研究采取。他们通过分析病毒基因组、编码的蛋白质及其拷贝数, 以及构建这些蛋白质所需的肽键和磷酸二酯键的能量需求, 推导出病毒代谢的化学计量学关系。Aller 等人构造了 Chikungunya、登革热和寨卡病毒这三种病毒的 VBOF, 之后将这些 VBOF 依次单独整合到人类巨噬细胞的 GEM 中, 通过对三个整合模型的分析, 成功预测了一系列涉及氨基酸和核苷酸代谢途径的反应, 这些反应具有潜在的广泛或特异性抗病毒能力。

然而, 已有的关于病毒性疾病代谢网络模型相关的研究, 在以下几个方面存在局限性: (1) 这些研究的对象都是特定的病毒类型, 所得到的实验结果都是针对特定病毒感染设计得到的, 缺乏一种通用的, 适用于所有病毒类型的靶点预测方法; (2) 现有研究重构的病毒性疾病代谢网络模型受组织特异性的影响, 且模型中会包含较多的原发组织特征; (3) 当前的研究并没有考虑到抗病毒靶点对于正常细胞的危害, 这样得到的抗病毒策略可能对人体产生较大的副作用, 不能应用于病毒性疾病的治疗, 增加了临床试验与药物测试的成本。

因此, 设计一种通用的病毒性疾病靶点预测方法是十分重要且有意义的。本文旨在从病毒感染后对宿主细胞代谢重编程的共性出发, 重构能够克服组织特异性和原发组织特征影响的代谢网络模型。将基于代谢网络模型的代谢分析建模成线性规划计算模型, 通过将生物学实验转化为计算问题进行求解, 预测具有抗病毒性质且对正常细胞无毒的靶点。这样的通用靶点预测流程能够发掘可能存在的广谱抗病毒靶点, 使得某种药物可以复用治疗多种病毒性疾病, 为病毒性疾病相关研究提供理论指导。

1.3 本文主要工作

本文提出了一种通用的病毒性疾病靶点预测方法流程。从目前比较容易获取的基因表达数据和相互作用数据出发, 经过一系列处理与映射, 得到与病毒感染异常代谢相关的核心代谢反应, 进行代谢网络模型的重构。这样的代谢网络重构方法关注病毒感染后代谢系统发生异常代谢的部分, 一定程度上克服了组织特异性, 且尽可能消除了重构代谢网络模型中的原发组织特征。将基于重构代谢网络模型展开的代谢分析过程建模成线性规划计算模型, 通过调整模型中的约束条件, 模拟单基因敲除、毒性测试生物学实验, 进行抗病毒靶点的预测。过程中考虑了靶点对正常细胞组织是否具有危害。整个流程通过消除病毒异常代谢活动, 达到抑制病毒复制的目的。本文研究工作的关系如图 1.2 所示。

本文主要研究内容包括:

(1) 越来越多的研究证明, 代谢重编程是病毒性疾病的重要标志。病毒感染

宿主细胞后，会调控宿主细胞的代谢系统，使其朝着更适合病毒复制生长的方向发展。当前关于病毒性疾病代谢方面的研究均针对特定病毒设计靶点发现方法，无法适用于其他病毒性疾病。因此本文设计实现了一种基于基因表达数据的病毒性疾病通用靶点预测方法，使用美国国家生物技术信息中心（National Center of Biotechnology Information, NCBI）的基因综合表达数据库（Gene Expression Omnibus, GEO）^[66]，从感染前后基因表达水平的差异出发，重构了 15 种病毒的特异性代谢网络模型。这些模型反映了病毒感染前后宿主代谢系统发生变化的反应和通路，能很好克服组织特异性和原发组织特征的影响。基于代谢网络模型的系数矩阵和信息集合，构建线性规划计算模型，将代谢分析过程转化为优化求解问题。并通过调整线性规划模型中的约束条件，模拟单基因敲除与毒性测试生物学实验。将敲除后病毒代谢网络模型中 VBOF 通量趋近于 0，且正常细胞模型对于预定代谢任务具有可行解的基因预测为靶点基因。整个预测流程对病毒性疾病具有通用性，可能寻找到针对病毒性疾病的广谱靶点。预测的抗病毒靶点被已有的文献证明与病毒性疾病密切相关，其中具有一定广谱性的靶点也被证明与病毒共同的基础代谢活动有关，证明了结果的正确性。

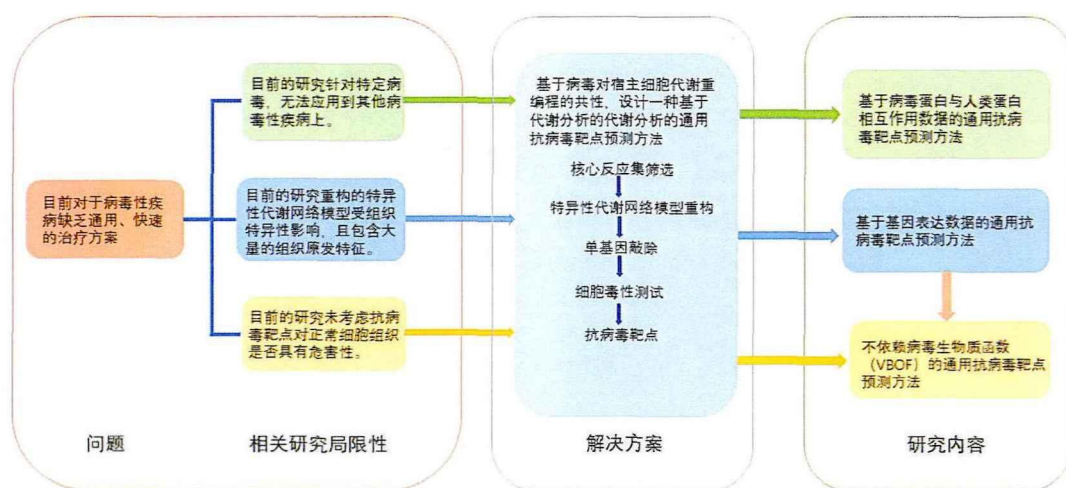


图 1.2 研究工作关系图

(2) 当前，病毒蛋白与人类蛋白相互作用数据是除了基因表达数据之外另一种容易获取的数据类型，反映了病毒蛋白质与人类蛋白之间的紧密联系。因此，本文设计实现了一种基于相互作用数据的病毒性疾病通用靶点预测方法，使得基因表达数据缺失的情况下，使用相互作用数据也能实现抗病毒靶点的预测。使用 IntAct 分子相互作用数据库 (IntAct Molecular Interaction Database)^[67] 中病毒蛋白与人类蛋白之间的相互作用数据，提取与病毒异常代谢相关的代谢特征，重构了 11 种病毒的代谢网络模型。基于代谢网络模型中包含的数据信息，将生物

学分析建模成线性规划计算模型。通过调整计算模型中的约束条件,模拟单基因敲除和细胞毒性测试等生物实验,预测抑制病毒生物物质生长且对正常细胞无毒的抗病毒靶点。该方法将蛋白相互作用数据整合到代谢网络模型中,从多组学的角度分析了病毒性疾病的异常代谢。预测的基因靶点得到了相关文献的证明,其中广谱性靶点与使用基因表达数据的结果相似,进一步印证了实验结果的准确性与可靠性。

(3) 前两种预测流程都需要借助病毒生物物质合成函数。受限于当前关于病毒蛋白拷贝数以及脂质信息的获取难度与准确性,本文在方法(1)的基础之上,设计实现了一种不依赖 VBOF 的靶点预测流程。使用 GEO 数据库中病毒感染前后的基因表达数据,重构了 15 种病毒性疾病的代谢网络模型。通过以基因表达水平与代谢通量之间一致性最大为优化目标,建立线性规划约束计算模型,将基于代谢网络模型的生物学分析转化为优化求解问题。通过调整计算模型中的约束条件,模拟单基因敲除和毒性测试实验。通过计算基因敲除前后感染细胞和正常细胞代谢通量距离的变化情况,将敲除后感染代谢通量显著接近正常代谢通量,且对正常细胞无毒的基因预测为抗病毒靶点。该方法通过计算通量距离,代替计算 VBOF 通量进行病毒抑制效果的评估,消除了对 VBOF 的依赖,进一步提高了靶点预测流程的泛用性,为病毒性疾病的治疗提供了全新的研究思路。

1.4 内容安排

第一章: 绪论

本章主要介绍了病毒性疾病对人类生命安全造成的重大影响以及当前基于代谢网络模型展开的病毒性疾病研究的背景与现状,并总结了本课题的主要研究内容。

第二章: 基于基因表达数据的病毒性疾病通用靶点预测方法

本章设计并实现了一种基于基因表达数据的病毒性疾病通用靶点预测方法。基于多种病毒的基因表达数据,重构出病毒感染前后代谢系统存在差异部分的代谢网络模型,克服组织特异性与原发组织特征的影响。基于代谢网络模型中的系数矩阵与信息集合,将生物学分析过程建模成线性规划计算模型。通过调整计算模型的约束条件,模拟单基因敲除、细胞毒性测试生物实验,预测无害的抗病毒靶点。实验预测的靶点结果已经被一些文献证明与对应病毒相关,说明了方法的准确性与可行性。

第三章: 基于病毒蛋白与人类蛋白相互作用数据的病毒性疾病通用靶点预测方法

本章设计并实现了一种基于相互作用数据的病毒性疾病通用靶点预测方法。

使得基因表达数据缺失的情况下,使用相互作用数据也能完成靶点预测。通过提取病毒蛋白与人类蛋白的相互作用数据中的代谢特征,进行代谢网络模型的重构。将基于代谢网络模型的生物学分析建模成线性规划计算问题,通过调整线性规划约束,模拟单基因敲除和毒性测试过程,预测无毒的抗病毒靶点。预测的靶点结果与使用基因表达数据的结果基本一致,且已被文献证明与病毒相关,进一步印证了靶点的可靠性。

第四章:不依赖 VBOF 的病毒性疾病通用靶点预测方法

本章设计并实现了一种不依赖于 VBOF 的病毒性疾病通用靶点预测方法。通过建立基因表达水平与代谢反应通量一致性最大的约束计算模型,模拟代谢网络模型中的生物学分析过程。以计算感染细胞与正常细胞之间的代谢通量距离,替代前两章中计算 VBOF 通量的方式来评估病毒抑制效果,消除靶点预测过程对 VBOF 的依赖,克服了 VBOF 构造过程中数据缺失的问题。该方法进一步提高了病毒性疾病靶点预测的泛用性,为未来研究提供了新的思路。

第五章:总结与展望

本章对本课题整体的研究进行了总结,分析本课题的优缺点并对未来的工作进行了展望。

第2章 基于基因表达数据的病毒性疾病通用靶点预测方法

2.1 背景及意义

基于病毒感染后会对宿主细胞代谢系统进行代谢重编程的特点，越来越多的研究开始关注病毒感染之后宿主细胞的代谢状态。当前关于病毒性疾病的代谢分析研究大多基于代谢网络模型展开，通过重构出病毒感染后宿主细胞的代谢网络模型，分析有价值的生物学知识，从而发掘病毒性疾病发病过程中的关键代谢通路、反应以及潜在的抗病毒靶点。而当前的研究均以特定病毒为研究对象，所设计的分析过程具有极强的针对性，无法应用到其他病毒性疾病上去，缺少一种通用的抗病毒靶点发现方法。

对于病毒性疾病代谢网络模型重构，现有研究主要采用两种方法。第一种方法通过将病毒生物物质函数（VBOF）插入到已有的人类特定组织的代谢网络模型（GEM）中，以模拟病毒感染后的宿主细胞代谢网络^[60-61]。然而，这种方法实际上并不涉及到真正的网络模型重构，主要分析了 VBOF 对人类组织 GEM 的影响。这种方法使用的代谢网络模型包含了人类原发组织的所有代谢信息，无法有效聚焦于病毒感染引起的宿主细胞代谢异常。此外，病毒通常感染人类多种组织细胞，随着选取不同组织的人类 GEM，代谢分析的结果也可能不同，即代谢分析结果受到组织特异性的影响。第二种方法则基于一组病毒感染后显著活跃的代谢反应集合，以人类全基因组规模的代谢网络模型为背景进行重构^[64]。相比第一种方法，这种方式能更有效地减少信息的冗余，更精确地表征病毒感染后的活跃代谢系统。尽管如此，部分代谢通路在非感染细胞中也可能表现为活跃状态，从而干扰模型的精确重构。由于不同组织中活跃代谢反应不尽相同，此种代谢网络模型重构方法一定程度上也会受到组织特异性的干扰。

此外，现有研究在代谢网络模型上对病毒性疾病的分析方式较为单一，常采用通量平衡分析（FBA）^[62]或通量变异分析（FVA）^[63]来评估模型中的代谢通量分布。这些分析将与病毒生物物质生长反应（VBOF）相关的代谢通路和反应视为影响病毒复制的关键因素，并作为潜在的抗病毒靶点进行生物实验验证其有效性。这一过程大量依赖于生物实验和生物学知识以解释靶点的抗病毒特性。此外，现有研究往往忽视了抗病毒靶点对正常细胞组织的潜在危害性的考虑，预测的靶点可能存在严重的副作用，而无法应用于病毒性疾病的治疗。

针对已有研究中存在的问题，本章旨在从基因表达数据出发，以代谢网络模型为工具，设计一种适用于所有病毒性疾病的通用靶点预测方法。使用 GEO 数

数据库提供的基因表达数据,分析感染前后宿主细胞中显著差异表达的基因。并将这些基因映射到它们调控的代谢反应上,重构出多种病毒性疾病的差异代谢网络模型,称为“DELTA”网络模型。“DELTA”网络模型表征了感染细胞相对正常细胞代谢异常活跃的代谢活动,消除了绝大部分的原发组织特征。从差异的角度进行代谢网络模型重构,很大程度上减少了组织特异性活跃的一些代谢特征,能更准确地聚焦于由病毒感染引起的异常代谢活动。

此外,本研究利用重构的“DELTA”网络模型中的代谢反应系数矩阵、热力学信息等数据,将代谢分析过程建模成一个线性规划计算模型。通过调整计算模型中的约束条件,模拟生物学意义上的基因敲除及代谢任务测试过程,进行单基因敲除和毒性测试仿真实验。通过将生物学实验转化为线性规划优化求解问题,预测对正常细胞无害的抗病毒靶点。这种靶点发现方法不仅考虑了对正常细胞组织的影响,而且整个流程在计算机工具上实现,提高了研究的生物学解释性和操作的可行性。本章从基因表达数据出发,采用通用的流程进行各种病毒性疾病的靶点预测,为病毒性疾病的治疗方案提供新的思路和线索。

2.2 数据与方法

为消除组织特异性以及原发特征的影响,本章从病毒入侵宿主细胞前后的代谢差异出发,构建特异性的差异代谢网络模型。将基于差异代谢网络模型展开的代谢分析过程建模成线性规划计算模型,通过调整约束条件模拟单基因敲除和毒性测试生物实验,预测对正常细胞无害的抗病毒靶点。使用的分析方法大致分为四步:(1)数据预处理,对初始数据集进行预处理,分别获取病毒感染细胞与正常对照的基因表达数据;(2)差异表达分析,对预处理阶段获得的两组数据进行差异分析,得到差异表达基因集;(3)“DELTA”网络模型重构,将差异上调基因映射到一组核心的代谢反应集合,以此为基础重构一致性的“DELTA”网络模型;(4)基于“DELTA”网络的代谢分析及靶点预测,通过分析“DELTA”网络中每个基因对病毒生物物质生长的必要性以及其对正常组织的毒性测试,预测无害的抗病毒靶点基因。整个方法流程如图2.1和算法2.1所示。

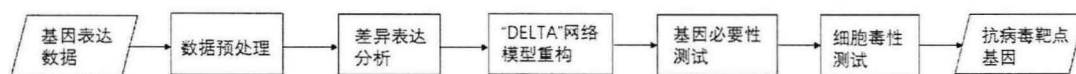


图 2.1 基于基因表达数据的通用抗病毒靶点预测方法流程图

算法 2.1 基于基因表达数据的通用抗病毒靶点预测算法

Input: 病毒性疾病的基因表达数据; 人类全基因组代谢网络 Recon3D; 病毒的生物物质合成函数 VBOF

Output: 病毒性疾病的靶点基因集合 Target

- 1 数据预处理, 获得正常细胞和感染细胞的基因表达数据 NormalData, TumorData;
- 2 差异表达分析, 根据差异分析获取感染细胞相对正常细胞显著上调的基因集合 Gup;
- 3 核心反应集获取, 将差异上调基因映射到一组核心代谢反应上 Rup;
- 4 代谢网络重构, 基于 Rup, VBOF 以及 Recon3D, 使用 FASTCORE 算法重构出“DELTA”网络模型;
- 5 基因必要性测试, 在“DELTA”网络模型上进行单基因敲除实验, 获得必要基因集合 essentialGenes;
- 6 细胞毒性测试, 剔除 essentialGenes 中对正常细胞组织有害的基因, 获得靶点基因集合 Target;
- 7 return Target;

2.2.1 数据与预处理

国家生物技术信息中心 (NCBI) 的基因表达综合数据库 (Gene Expression Omnibus, GEO)^[66] 是一个公共功能基因组数据库, 其中收录了多种疾病 (如癌症、病毒感染) 相关信息, 包含动物细胞的 RNA-seq、DNA 甲基化、生物标记物以及染色质可及性等数据。GEO 数据库的主要应用之一是在疾病研究和药物开发中, 通过比较疾病状态与健康状态下的基因表达差异, 识别出潜在的生物标志物和治疗靶点。例如, 在癌症研究中, 分析 GEO 中的数据可以揭示特定类型癌症的关键基因表达模式, 从而有助于理解癌症的发病机制和寻找新的治疗方法。GEO 数据库中包含了大量各种病毒性疾病的基因表达数据集。通过整合和分析 GEO 数据库中的基因表达数据, 研究者能够在全基因组水平上进行研究, 为病毒性疾病靶点预测的相关研究提供了重要的数据支撑。

从 GEO 数据库中下载了关于 15 种病毒性疾病的基因表达数据, 这些病毒为人们所熟知或对人们的生命健康造成了重大影响。为了分析病毒感染前后的差异表达, 除了需要感染细胞的基因表达数据之外, 还需要足够数量的正常细胞的表达数据作为对照组。

使用的 15 种病毒性疾病的 RNA-seq 基因表达数据是 GEO 数据库提供的 level 3 层面的 read count 数据, 其通过计算测序过程中生成的短序列片段与参考基因组或转录本的匹配程度来获得。相比于原始的测序 reads 数据, level 3 的 read count 数据经过了一系列生物信息学处理, 使其更适合用于后续的统计分析和功能解析, 减轻了研究人员在数据处理上的工作量。因其可靠性和稳定性, read count 数据当前被认为是差异分析结果最准确的数据。对于 15 种病毒性疾病的数据集, 其基因表达数据获取时的实验方案设计不同, 有些实验涉及处理时间、人体组织、以及药物处理等不同变量, 因此需要进行不同的预处理操作。以 SARS-CoV-2 为

例, GEO 数据库中序列号为 GSE163959 的数据集, 其中包含了 SARS-CoV-2 感染和 A 型流感病毒感染的人鼻甲和肺组织的基因表达数据。在预处理过程中, 需要将数据集中 A 型流感病毒相关的数据剔除, 并将 SARS-CoV-2 病毒感染细胞和正常细胞的基因表达数据分开, 便于后续的差异表达分析。预处理后的 15 种病毒性疾病的基因表达数据详情如表 2.1 所示。

表 2.1 15 种病毒性疾病基因表达数据详情

病毒名称	数据集	感染细胞样本数	正常细胞样本数
基孔肯雅热病毒 (Chikungunya virus)	GSE197744	2	2
乙肝病毒 (Hepatitis B virus)	GSE223073	24	22
丙肝病毒 (Hepatitis C virus)	GSE211161	2	2
人冠状病毒 (Human coronavirus)	GSE155986	12	3
人疱疹病毒 1 (Human herpesvirus 1)	GSE184463	6	6
人体免疫缺陷病毒 1 (Human immunodeficiency virus 1)	GSE152788	4	2
人类 SARS 冠状病毒 (Human SARS coronavirus)	GSE148729	8	6
甲型流感病毒 (Influenza A virus)	GSE206606	10	12
日本脑炎病毒 (Japanese encephalitis virus)	GSE221680	3	3
麻疹病毒 (Measles morbillivirus)	GSE131840	6	6
中东呼吸综合征冠状病毒 (MERS coronavirus)	GSE56192	36	9
新型冠状病毒 (SARS-CoV-2)	GSE163959	5	5
西尼罗河病毒 (West Nile virus)	GSE136342	10	10
扎伊尔埃博拉病毒 (Zaire ebolavirus)	GSE100839	3	3
寨卡病毒 (Zika virus)	GSE160201	4	4

2.2.2 差异表达分析

基因表达数据的差异分析是生物学研究中常见的任务之一, 它可以揭示不同样本之间的基因表达水平差异, 有助于识别与特定条件或疾病相关的潜在生物标志物。Limma (Linear Models for Microarray Data)^[68-69] 工具包是一种常用的用于差异表达分析的 R 语言软件包, 其基于线性模型和经验贝叶斯方法, 能够在小样本量下有效地分析基因表达数据。使用 Limma 工具包提供的 voom 函数对 read count 数据进行归一化操作, 并调整数据分布。对每一种病毒性疾病的基因表达数据, 本章以 $|\log FC| > 1.0$ 和 $p - value < 0.05$ 作为基因表达显著差异的阈值, 筛选出感染细胞与正常细胞表达水平存在显著差异的基因集合。其中,

logFC 表示了基因在不同条件下的相对表达水平的倍数变化, p-value 是统计假设检验的结果之一, 用于衡量差异的显著性水平。较大的 logFC 值表示基因表达差异更明显, 而较小的 p-value 表示这种差异更显著。

2.2.3 “DELTA”网络模型重构

现有研究重构的病毒性代谢网络模型会受到组织特异性与原发组织特征的干扰, 从而影响后续靶点发现的结果。本章通过分析感染细胞与正常细胞之间的差异, 提出了一种重构病毒性疾病代谢网络模型的新思路。基于差异上调基因集合, 重构出病毒感染细胞相对于正常细胞代谢异常活跃部分的代谢网络模型, 能消除绝大部分的原发组织特征, 称为“DELTA”网络模型。从感染前后的代谢差异出发重构代谢网络模型, 可以减少组织特异性活跃的特征, 即一些在特定组织中表现活跃的反应, 有效克服组织特异性的影响。

选用人类全局代谢网络模型 Recon3D^[56]作为“DELTA”网络模型重构的基础, 结合病毒生物物质函数 VBOF, 分析病毒感染后发生异常代谢的系统部分。Recon3D 是一款高质量的、覆盖人类全基因组规模的代谢网络模型, 代表了目前最为全面和详尽的代谢通路模型之一。关于 VBOF 的构造, 采用了 Aller 等人^[65]的方法, 基于病毒的基因序列和蛋白质序列计算了核苷酸、氨基酸组成及所需能量供应的化学计量系数, 从而形成病毒生物物质生长的具体反应式。

对于“DELTA”网络模型的重构, 主要分为以下几个步骤, 流程如图 2.2 所示。首先提取上调基因中出现在 Recon3D 模型基因集合的部分, 该部分基因被认为与人类代谢相关。

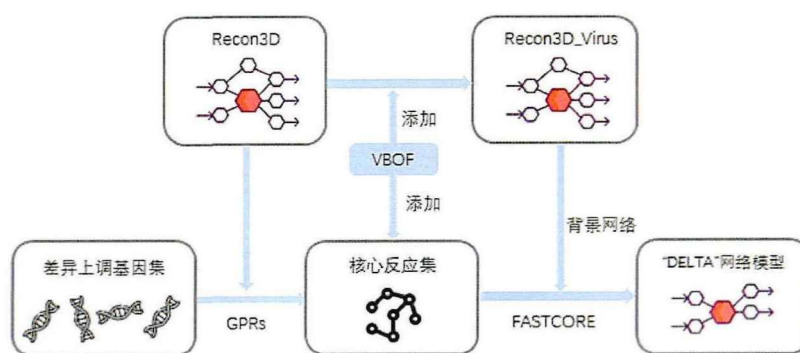


图 2.2 “DELTA”网络模型重构示意图

之后将这部分上调基因映射到一组代谢反应上。借助 Recon3D 模型中的基因-反应 (Gene-Reaction, GR) 映射规则实现, 其通过逻辑运算符 (如 AND、OR、NOT) 详细描述了基因和代谢反应之间的相互关系。使用二进制编码的方式表

示各个基因的状态, 用 1 表示感染细胞相对正常细胞差异上调的基因, 0 表示其他基因, 得到能够表征所有基因状态的向量。在基因状态向量上运用 GR 映射规则, 计算出关于代谢反应状态的 0-1 向量, 将其中状态为 1 的代谢反应认为是感染细胞相对正常细胞异常活跃的代谢反应。

为了分析病毒感染对人类代谢网络的影响, 向 Recon3D 模型中手动引入了病毒生物物质生长函数 VBOF, 命名为 Recon3D_Virus, 作为病毒感染异常代谢网络模型重构的背景网络模型。此外, 为确保重构的代谢网络模型能够支持病毒的生物物质生长, 需要将病毒的生物物质生长函数 VBOF 集成到核心反应集中。

接着, 利用 Cobra Toolbox 3.0^[70]中的 FASTCORE 算法^[57], 以添加了 VBOF 的核心反应集为基础, 在 Recon3D_Virus 上搜索识别出一个最小的、满足通量一致性的子集模型, 即为重构得到的“DELTA”网络模型。整个代谢网络的重构和靶点预测过程均在 MATLAB 2016a 环境下完成。相关流程如算法 2.2 所示。

算法 2.2 “DELTA”网络模型重构算法

Input: 病毒性疾病的差异上调基因集合 G; 人类全基因组代谢网络 Recon3D; 病毒的生物物质合成函数 VBOF

Output: “DELTA”网络模型 DELTA_model

```
1 Gene_Up=intersect(Gi,Recon3D.genes);
2 Rxn_Up=Map_gene2rxn(Gene_Up,Recon3D.grRules,Recon3D.rxns);
3 Rxn_Up.add(VBOF);
4 Recon3D_Virus=addReaction(Recon3D,VBOF);
5 DELTA_model=FASTCORE(Rxn_Up,Recon3D_Virus);
6 return DELTA_model;
```

2.2.4 基因必要性测试

现有研究在重构代谢网络模型上展开的病毒性疾病的靶点预测, 大多使用 FBA 和 FVA 分析方法, 计算代谢网络模型中的反应通量分布。通过筛选出其中与 VBOF 通量相关性高的代谢反应和代谢通路, 配合生物实验验证其对病毒的抑制性, 很大程度上仍然需要借助生物学实验和知识。

本章围绕“DELTA”网络模型展开病毒性疾病的代谢分析。具体而言, 通过计算“DELTA”网络模型中 VBOF 的通量值, 也就是生物意义上的反应速率, 来分析病毒生物物质生长的情况。利用“DELTA”网络模型中的反应系数矩阵和热力学信息, 构建线性规划计算模型进行代谢反应通量分析, 将生物学问题转化为计算问题进行求解。详情如公式 2.1 至 2.3 所示。

$$\max \quad VBOF \quad (2.1)$$

$$s.t. \quad S \cdot v = 0 \quad (2.2)$$

$$v_{\min} \leq v \leq v_{\max} \quad (2.3)$$

其中 S 是“DELTA”网络模型中大小为 $m \times n$ 的化学计量系数矩阵，包含了所有代谢反应涉及的代谢物（反应物与生成物）及系数关系， $S_{i,j}$ 表示反应 i 中的第 j 个代谢物的系数。 v_{max} 和 v_{min} 分别表示代谢反应的反应速率上/下界。 v 表示代谢网络中所有代谢反应的通量向量。“DELTA”网络模型的目标是病毒生物物质生长最大化，因此优化模型的目标为 VBOF 通量最大化（公式 2.1）。基于以上约束条件和优化目标，通过线性规划求解器可以求得“DELTA”网络中 VBOF 通量值。

通过调整线性规划计算模型中的约束条件，限制待敲除基因调控的代谢反应通量，达到基因敲除的效果。对于一个待敲除基因，利用 GR 映射规则计算得到该基因调控的所有代谢反应。在线性规划模型中，将这些代谢反应的通量值约束为 0（公式 2.7），模拟“DELTA”网络模型中该反应速率为 0 的效果，实现单基因敲除（图 2.3）。线性规划模型如公式 2.4 至 2.7 所示。其中 R_{knock} 表示待敲除基因映射而来的所有代谢反应。

$$\max \quad VBOF \quad (2.4)$$

$$s.t. \quad S \cdot v = 0 \quad (2.5)$$

$$v_{min} \leq v \leq v_{max} \quad (2.6)$$

$$v_i = 0, i \in R_{knock} \quad (2.7)$$

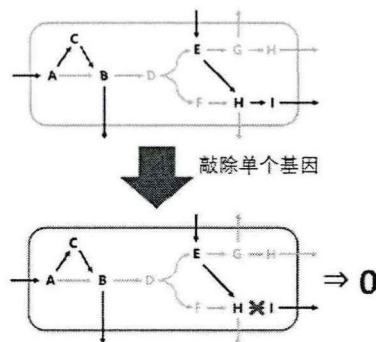


图 2.3 基因敲除过程示意图

调整了线性规划计算模型的解空间之后，计算“DELTA”网络模型中 VBOF 的通量。如果一个基因敲除后 VBOF 通量值为 0(或接近 0)，认为该基因当前“DELTA”网络的必要基因，其缺失会导致“DELTA”网络模型无法正常进行病毒生物物质的合成，使得病毒无法复制增殖。反之，如果一个基因敲除后对 VBOF 通量的影响较小，说明该基因对“DELTA”网络而言不是必要的。对“DELTA”网络模型中的所有基因逐一进行敲除，筛选出所有的必要基因。基因必要性测试

的整体流程如算法 2.3所示。

算法 2.3 基因必要性测试算法

Input: 特异性代谢网络模型 *Specific_Model*
Data: lb 和 ub 代表代谢反应速率下界和上界;
 特异性模型中病毒的生物质合成反应 VBOF;
 判断 VBOF 通量受到影响的阈值 ϵ , 设置为 $1e-4$
Output: *Specific_Model* 的必要基因 *essentialGenes*

```

1 essentialGenes=[];
2 genes=Specific_Model.genes;
3 rxns=Specific_Model.rxns;
4 foreach g in genes do
5     tempModel=Specific_Model;
6     rxn_konck=map_gene2rxn(g,Specific_Model.grRules);
7     r_index=rxns.indexof(rxn_konck);
8     tempModel.lb[r_index]=tempModel.ub[r_index]=0;
9     Object_val=FBA(tempModel,VBOF);
10    if object_val< $\epsilon$  then
11        | essentialGenes.add(g);
12    end
13 end
14 return essentialGenes;
```

2.2.5 细胞毒性测试

当前基于代谢网络模型展开的病毒性疾病相关研究,在靶点发现的过程中均未考虑对正常细胞组织是否具有危害性,从而导致靶点不可用的情况。通过模拟细胞毒性测试生物学实验,旨在寻找无害的抗病毒靶点。

在预测病毒性疾病的潜在靶点时,不能简单地将“DELTA”网络的所有必要基因视为候选靶点。特别是那些在正常组织代谢中发挥重要作用的基因,其缺失可能对正常细胞产生毒性,因此不宜作为靶点。从 Metabolic Atlas^[71]中下载了 83 个正常细胞代谢网络模型,在这些正常细胞代谢网络模型上对“DELTA”必要基因依次展开细胞毒性测试,剔除对人类正常细胞有害的基因。

对每个正常细胞代谢网络模型,依次敲除“DELTA”网络的必要基因,对敲除后的模型进行了 56 项预先定义的代谢任务的检测^[59],分析基因敲除对正常细胞组织的影响。预定义的代谢任务涉及能量和氧化还原、内部转化、底物利用和产物生物合成等一些人类细胞基本的代谢需求,是维持细胞正常状态必不可少的。生物学上关于代谢任务的测试通常使用标记实验和生物化学测定的方式,从代谢物的识别角度测试一些代谢任务能否正常完成,实验操作复杂且花费较长时间。

对于正常细胞的代谢分析与病毒不同,构建的线性规划模型也有所差异,主要体现在优化目标上。正常代谢网络模型以人类生物质生长为目标,线性规划的求解目标为人类生物质反应通量最大化(公式 2.8)。线性规划计算模型的形式如

公式 2.8至 2.10所示。其中 biomass 为人类的生物质反应通量, b 是代谢物的摄入与输出量信息。

$$\max \quad biomass \quad (2.8)$$

$$s.t. \quad S \cdot v = b \quad (2.9)$$

$$v_{min} \leq v \leq v_{max} \quad (2.10)$$

通过将代谢任务中的生物要求, 转化成一组线性规划模型中的约束条件来定义每个代谢任务, 包括摄入量约束、输出约束、反应速率界限约束等。将复杂的代谢任务测试过程转化为线性规划计算问题, 通过调整计算模型中的约束条件, 达到模拟代谢任务的效果。可以通过调整稳态约束 (公式 2.9), 从而限制特定代谢物的吸收和输出量, 模拟细胞对代谢物的摄入和生成情况。此外, 可以通过对交换反应的代谢通量约束 (公式 2.10) 的调整, 限制特定代谢反应的反应速率。基于代谢任务定义的约束调整, 若线性规划问题仍旧存在可行解, 则认为敲除该基因后正常细胞可以完成当前代谢任务。

在 83 个正常细胞组织的代谢网络模型中, 依次敲除“DELTA”必要基因, 并测试敲除这些基因后预定义代谢任务的完成情况, 分析敲除该基因是否对正常细胞组织具有毒性。对敲除基因是否有毒的判断规则如下:

(i) 如果模型中没有包含该基因, 则该基因对模型的毒性是未知的, 认为其对模型来说是不安全的。

(ii) 基因敲除后, 如果模型仍能完成所有 56 项代谢任务, 则该基因对模型是安全的, 否则该基因是有毒的。

(iii) 一个必要基因可以作为候选靶基因的条件是它在超过一半的正常细胞模型 (≥ 42) 中是安全的。候选基因靶点的计算逻辑如算法 2.4 所示。

2.3 实验结果

2.3.1 差异表达基因

对采集到的 15 组不同病毒感染细胞与正常细胞的基因表达数据进行了差异分析, 旨在识别特定病毒感染下差异表达的基因集合。为了深入理解病毒感染宿主细胞后代谢系统的异常活跃的代谢活动, 本章专注于分析差异上调基因。各组基因表达数据的差异上/下调基因规模及频次如图 2.4 和图 2.5 所示。

统计结果显示, CMPK2 和 ASNS 两种基因在超过三分之二的病毒中呈现出差异上调的表达模式。其中 CMPK2 基因编码一种参与核苷酸合成挽救途径的酶, 可能在单核细胞的末端分化中发挥作用。Pawlak 等人^[72]的研究指出, CMPK2 基

因对多种黄热病毒,例如 ZIKA 病毒的翻译过程产生影响。病毒感染后,宿主细胞会释放 I 型干扰素,与细胞表面的受体结合,导致包括 CMPK2 在内的 IFN 刺激基因 (ISGs) 蛋白的表达增加,从而使细胞进入抗病毒状态,抑制病毒的复制和传播。

算法 2.4 细胞毒性测试算法

Input: 特异性代谢网络模型的必要基因集合 *essentialGenes*; 正常组织细胞代谢网络模型 *Normal_Models*; 预定义的代谢任务 *tasks*
Output: 对人类正常组织无毒的靶点基因集 *targetGenes*

```

1 targetGenes=[];
2 foreach g in essentialGenes do
3     safe_score=0;
4     foreach model in Normal_Models do
5         temp_model=removeGenes(model,g);
6         check_result=True;
7         foreach t in tasks do
8             check=CheckTask(temp_model,t);
9             if !check then
10                 check_result=False;
11                 break;
12             end
13         end
14         if check_result then
15             safe_score+=1;
16         end
17     end
18     if safe_score>=length(Normal_Models)/2 then
19         targetGenes.add(g);
20     end
21 end
22 return targetGenes;

```

另一个在多数病毒性疾病中表现出显著上调的基因 *ASNS* 编码的蛋白参与了天冬酰胺的合成,这是人类巨细胞病毒 (HCMV) 复制所需的关键宿主因子。Lee 等人^[73]的研究表明,在整个单循环感染过程中,敲除 *ASNS* 基因会抑制 HCMV 的初级复制。两个基因分别在宿主细胞对病毒感染的免疫反应和病毒本身复制阶段的阶段表现出异常的高表达状态。尽管当前研究仅发现了两者对特定类型病毒的依赖性,但它们与病毒性疾病之间的关联为我们寻找各种病毒共有特征提供了良好的起点。

此外,为了更加全面深入地了解差异表达基因的功能和相关代谢通路,基于 DAVID 数据库 (<https://david.ncifcrf.gov/>)^[74]对 15 组病毒数据得到的所有差异上调基因进行了基因本体 (Gene Ontology,GO) 分析和 KEGG Pathway 分析,均以 $p\text{-value} < 0.05$ 作为分析阈值。根据 GO 分析的结果 (图 2.6),从细胞组分角度看,上调基因主要集中在膜的组成成分与细胞质上,这说明病毒感染后,会修改宿主细胞的细胞质环境与膜的构成,促进病毒复制和感染其他细胞的进程。而在

细胞功能角度，上调基因主要集中在蛋白质、ATP 等物质的结合上，这与病毒感染后快速复制的能量和物质需求相符。KEGG 分析的结果显示(图 2.7)，15 组病毒上调基因在嘌呤代谢和辅因子的生物合成等通路上有一定程度的富集，但未表现得如 GO 分析的结果般显著，这可能与病毒本身感染机制与特异性相关。

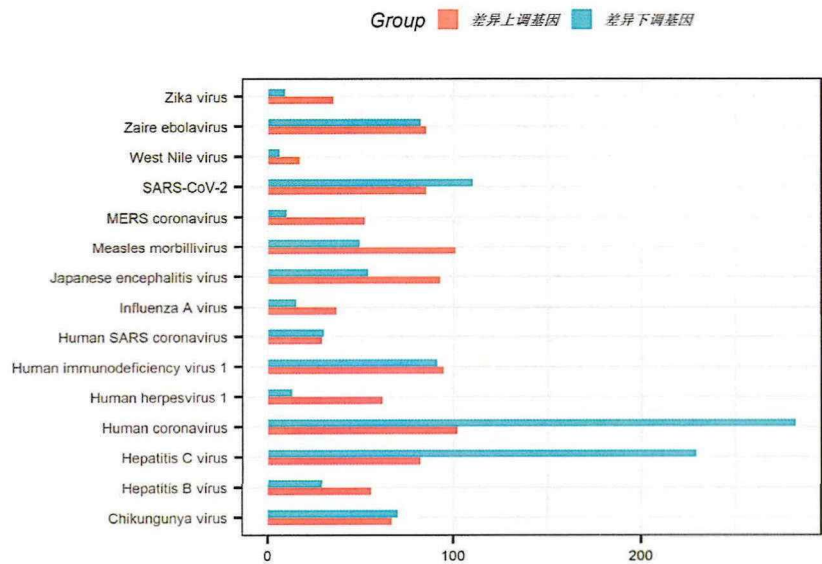


图 2.4 15 种病毒性疾病差异表达基因规模

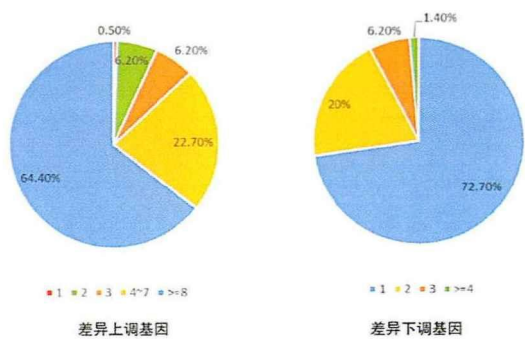


图 2.5 15 种病毒性疾病差异表达基因出现频次

2.3.2 “DELTA” 网络模型结构与分析

15 种病毒性疾病重构的“DELTA”网络模型的规模大小如图 2.8所示，相较于包含了 2248 个代谢基因，10600 个代谢反应和 5835 个代谢物的 Recon3D，15

个“DELTA”网络平均包含了 888 个代谢基因, 1512 个代谢反应和 1261 个代谢物。其中 HIV 对应的“DELTA”网络规模最大, 表明其在感染人体细胞后对人体代谢系统调控程度是最高的, 人体代谢系统的变化最大。相反地, SARS 病毒重构得到的“DELTA”网络规模最小, 表明其对人体代谢系统的改变最小。这与前面所获得的各种病毒差异基因的规模是基本一致的。

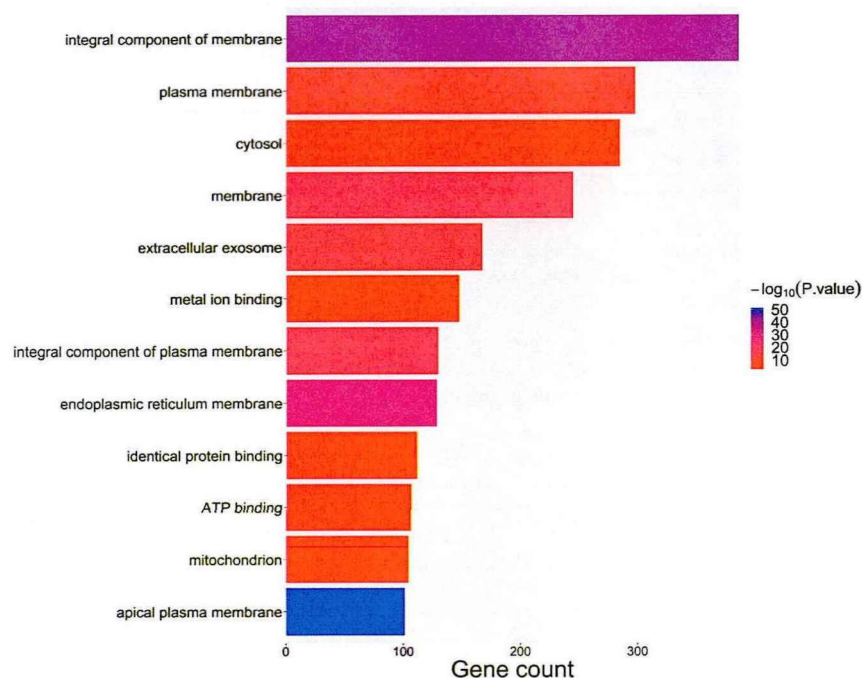


图 2.6 对 15 种病毒性疾病的差异上调基因的 GO 分析结果

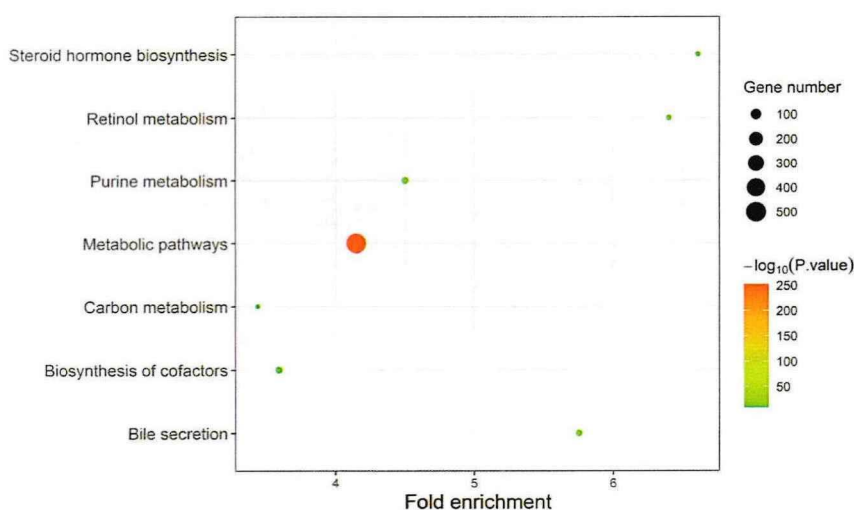


图 2.7 对 15 种病毒性疾病的差异上调基因的 KEGG Pathway 分析结果

不同的病毒感染人体不同的组织细胞, 且现有的基于代谢网络模型展开的

病毒性疾病研究大多基于对应人体组织 GEM，构建感染细胞的特异性代谢网络模型展开分析。因此需要格外注意组织特异性以及代谢系统中包含的原发组织特征的影响。考虑到这些因素，通过构建“DELTA”网络模型的方式，从基因表达数据出发，对多种病毒以一种通用流程完成特异性网络模型的重构，达到消除组织特异性的目的。此外，以感染细胞相对于正常细胞差异上调的基因为核心，构建“DELTA”网络模型，可以很好减少重构模型中的原发组织特征。

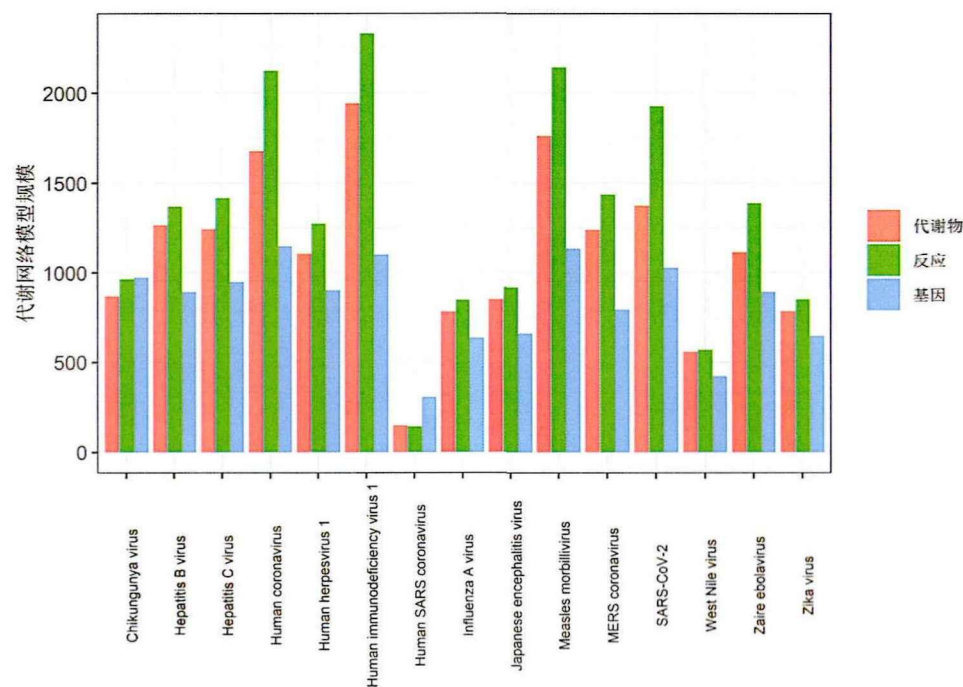


图 2.8 15 种病毒性疾病重构的“DELTA”网络模型规模

为了验证“DELTA”网络模型对原发组织特征的消除情况，从人类蛋白质图谱（The Human Protein Atlas,HPA）^[75]下载了管家蛋白组相关的基因列表，这 8899 个管家基因在人体的每种细胞类型中表达。通过将管家基因通过 GR 规则映射到代谢反应中，作为人体系统中代谢系统的管家反应，并计算各“DELTA”网络模型对管家反应的消除情况，相关指标如表 2.2所示。

表 2.2 15 个“DELTA”网络模型对管家反应的消除率详情

消除率指标	平均值	方差	最大值	最小值
值	82.84%	8.5×10^{-3}	98.70%	67.66%
消除率指标	上四分位数	中位数	下四分位数	四分位距
值	89.59%	83.02%	76.77%	12.82%

15种病毒“DELTA”网络模型对于人类管家反应的平均消除率为82.84%，其中SARS对应“DELTA”网络消除率最高，为98.70%，HIV对应“DELTA”网络消除率最低，为67.66%，分别对应了规模最小与最大的“DELTA”网络。将这些“DELTA”网络模型与Kishk等人^[64]重构的SARS-CoV-2特异性模型和Anoop等人^[76]重建的HIV-1特异性模型进行对比，两者的平均消除率分别为37.61%和62.65%，结果如图2.9所示。可以看出，“DELTA”网络模型关于管家反应的消除效果明显优于两者。15个“DELTA”网络模型相较于Recon3D，网络规模显著减小，且消除了绝大部分原发组织的代谢特征，其目标集中在感染细胞相较于正常细胞发生代谢变化的通路和子系统上，符合预期设想。

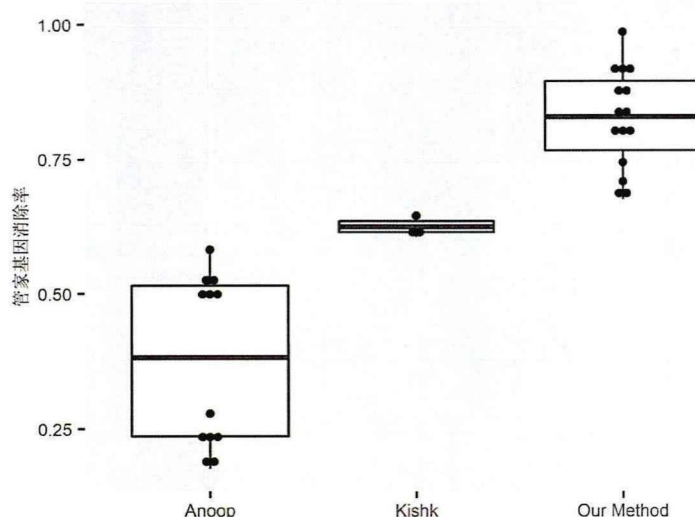


图 2.9 15 个“DELTA”网络模型的管家反应消除率结果对比图

2.3.3 抗病毒靶点预测结果

对15种病毒的重构“DELTA”网络模型进行了基因必要性测试和细胞毒性测试。图2.10展示了15个“DELTA”网络中实验得到的必要性基因和潜在靶点基因的规模分布。可见，这些病毒的必要基因和靶点基因在数量上存在显著差异，这与它们的差异基因集以及对应的VBOF有关。观察靶点基因集时发现，其中一些基因与相关病毒感染机制密切相关。例如，PLPBP基因在SARS-CoV-2的“DELTA”网络中被预测为靶点基因。Kimura等人^[77]的研究表明，SARS-CoV-2病毒入侵人体后的暴发性感染期间需要大量的D-丙氨酸来维持其代谢活动和复制进程。PLPBP基因调控的反应将L-丙氨酸转化为D-丙氨酸，敲除PLPBP基因抑制了D-丙氨酸的合成，从而抑制了SARS-CoV-2病毒的复制进程。被预测为HIV-1靶点基因的SLC29A2基因同样与HIV-1感染过程密切相关。Jordi等

人^[78]认为, HIV-1 感染与该基因的 mRNA 水平上调有关。肿瘤坏死因子- α (TNF- α) 的作用导致 SLC29A2 基因 mRNA 水平增加, 这正是 HIV-1 感染后脂肪组织表达细胞因子增强的结果。

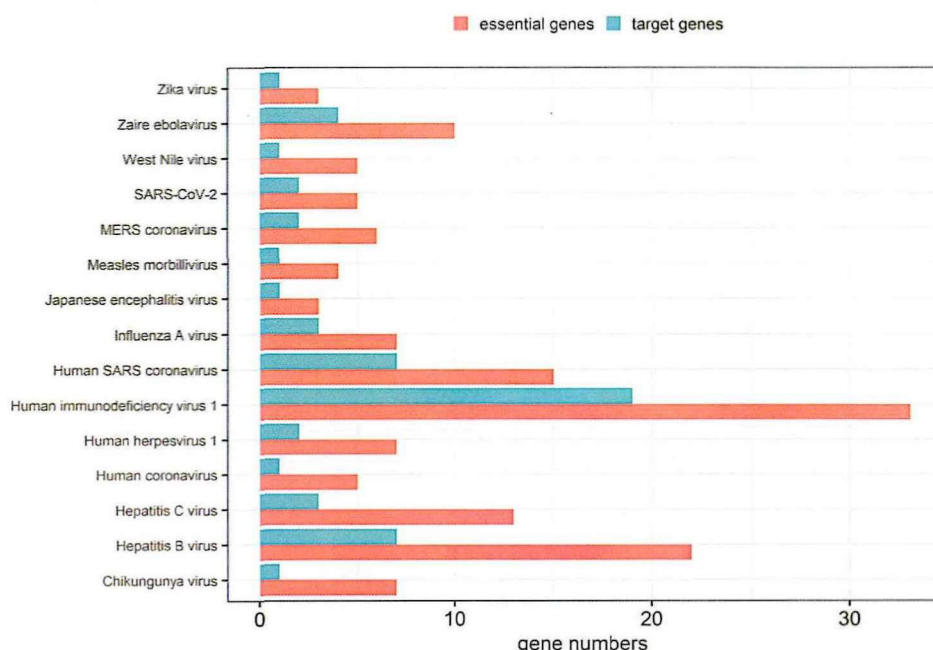


图 2.10 15 种病毒性疾病的必要性基因和靶点基因规模

然而, 15 组病毒数据预测的基因靶点在某种程度上仍存在共性。其中, CTH 和 DCTD 是出现频次最高的两个基因, 分别在 10 个和 8 个“DELTA”网络中被预测为潜在靶点。更重要的是, 在当前全部 15 个“DELTA”网络中, CTH 和 DCTD 单独或组合出现在靶点基因集合中, 这表明这两种基因在病毒异常代谢中起着重要作用。通过 KEGG 数据库^[79]查找 CTH 和 DCTD 两个基因的代谢通路, 可以发现 CTH 基因的代谢活动主要集中在甘氨酸、丝氨酸等氨基酸代谢和生物合成通路上, 而 DCTD 基因的代谢活动主要涉及嘧啶代谢和核苷酸代谢等通路。这两个基因分别与氨基酸和核苷酸的代谢活动密切相关, 而氨基酸和核苷酸正是 VBOF 的两个基本组成部分。敲除 CTH 或 DCTD 基因会影响“DELTA”网络中某些氨基酸/核苷酸代谢通路, 抑制病毒的生物质生长, 而对人体正常代谢无影响。因此, 可以认为 CTH 基因和 DCTD 基因是具有广谱性的抗病毒靶点基因, 值得进一步探究其在病毒性疾病致病机制上的调控作用与原理。

2.3.4 对病毒变异体的鲁棒性

病毒的变异性可能会使得当前治疗方案失效, 因此理想的抗病毒靶点需要对病毒的变异体表现出良好的稳定性。本章以 SARS-CoV-2 为例, 考察了“DELTA”

网络抗病毒靶点流程对变异体的稳健性。

为了探究病毒变异对 VBOF 的结构造成差异的情况，从 GenBank 数据库^[80]获取了 SARS-CoV-2 的五种常见变体（alpha, beta, delta, gamma 和 omicron）^[81-82]的基因序列及对应蛋白质序列，每个变体随机选取了 10 个完整且高质量的序列。然后使用 Aller 等人的方法计算各变异株序列的 VBOF，并将其化学计量系数的平均值与野生型（Wild Type, WT）VBOF 进行比较，五种变异体 VBOF 与 WT 之间的系数差异热图如图 2.11 所示。五种变异体与 WT 之间的系数差异主要集中在 ATP 和 ADP 上，但总体而言，差异很小。

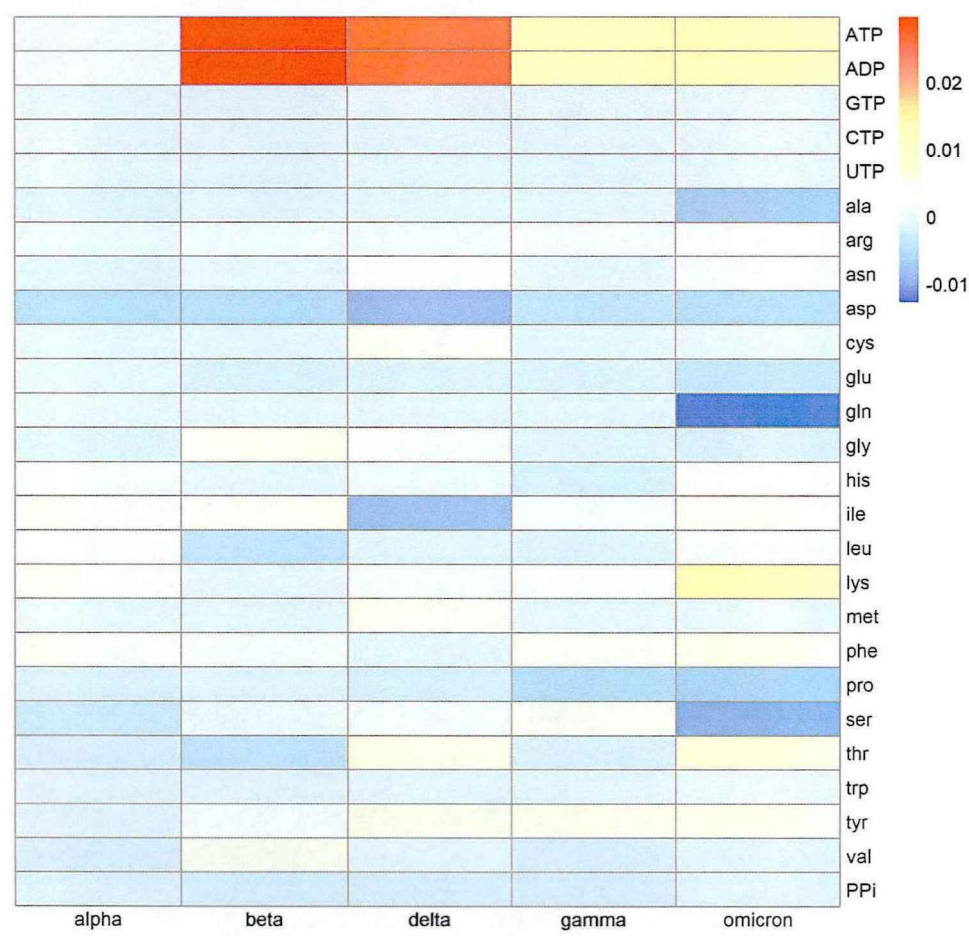


图 2.11 5 种 SARS-CoV-2 变异株 VBOF 的系数差异

此外，对每种变异体的多个样本进行了序列比对，筛选出突变位点较多的蛋白质。结果显示，S 蛋白是突变位点最多的蛋白质，表明 S 蛋白与 SARS-CoV-2 的突变密切相关。序列比对的结果如图 2.12 所示。

为验证 DELTA 网络的预测结果不受 SARS-CoV-2 突变的影响，检查了 S 蛋白关联的基因对“DELTA”网络目标的影响。从 BioGrid 数据库^[83]获取了 SARS-

CoV-2 蛋白质与人类基因相互作用数据,并提取了与 S 蛋白相关的数据。之后根据实验系统进行分组,筛选出与 S 蛋白强相互作用的 23 个代谢基因。在 SARS-CoV-2 的“DELTA”网络基础上,对这 23 个基因进行了单基因敲除实验。结果显示,23 次单基因敲除后,“DELTA”网络中 VBOF 的通量值与未敲除时几乎相同。这表明, S 蛋白关联的基因对 SARS-CoV-2 的生物合成几乎没有影响,也证明了 SARS-CoV-2 的突变不会影响“DELTA”网络的靶点预测结果。

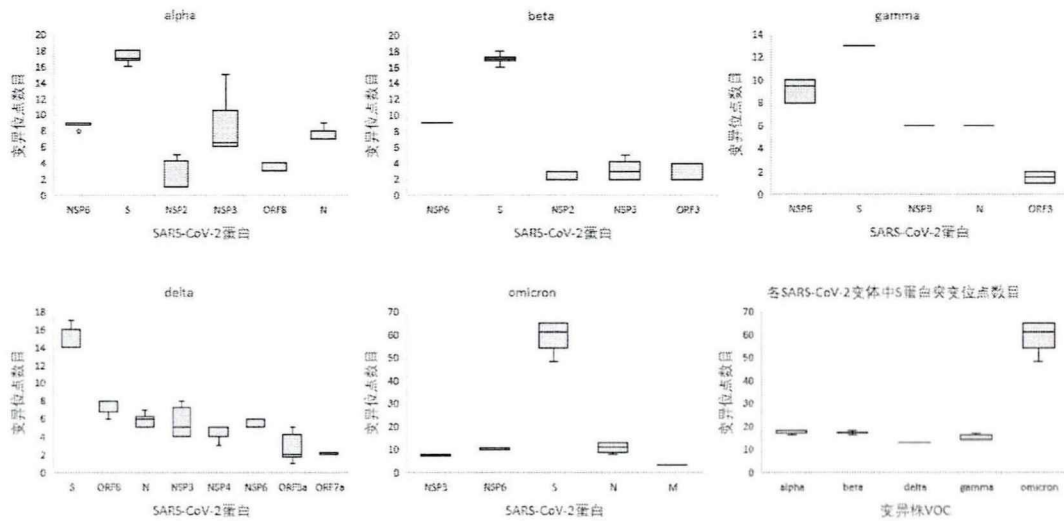


图 2.12 对 5 种 SARS-CoV-2 变异株的序列分析结果

2.4 本章小结

基于病毒感染后会对宿主细胞代谢系统进行代谢重编程的特点,越来越多的研究使用代谢网络模型对病毒性疾病展开代谢分析。然而,现有的研究通常专注于特定病毒,其分析方法具有较强的针对性和特异性,这限制了其在其他病毒性疾病中的应用。此外,这些研究重构的代谢网络模型受到组织特异性与原始组织特征的影响,且在靶点发现过程中忽略了对正常细胞组织可能造成的危害。

为了克服现有研究中的局限性,本章设计并实现了一种适用于所有病毒性疾病的通用靶点预测方法。使用 GEO 数据库中的病毒基因表达数据,通过比较感染细胞与正常细胞之间的差异表达,成功重构了“DELTA”网络模型。该模型显著反映了宿主细胞在病毒感染后的异常活跃代谢活动,能较好地克服组织特异性与消除原发组织特征。利用“DELTA”网络模型中的代谢反应系数矩阵以及热力学信息等数据,将代谢分析过程建模成一个以 VBOF 通量最大为优化目标的线性规划计算模型。通过调整计算模型上的约束条件,模拟单基因敲除和毒性测试的生物学实验,将生物学问题转化为优化求解问题,预测出无毒的抗病毒靶

点。其中，CTH 和 DCTD 基因在多个“DELTA”模型中被识别为关键靶点，显示出一定的广谱抗病毒性质。已有文献证实两者调控的代谢通路主要涉及氨基酸和核苷酸，与病毒共有代谢活动相关。整个靶点预测方法适用于所有的病毒性疾病，发现的广谱靶点具有重要的研究价值，为设计有效的病毒性疾病治疗方案提供了坚实的理论基础。

第 3 章 基于相互作用数据的病毒性疾病通用靶点预测方法

3.1 背景及意义

虽然某些研究领域可以访问到海量数据，但在病毒性疾病研究领域，需要面临的挑战是数据类型繁多而相对零散，样本量有限，且这些数据通常与生物实验的方案设计紧密相关。当前，除了病毒的序列信息，基因表达数据^[31]是相对容易获取的，这些数据包含多种病毒信息，格式规范，并已被广泛应用于抗病毒策略的发现，也是当前关于病毒性疾病代谢分析使用的主流数据类型。尽管如此，当前收录的基因表达数据仍未覆盖所有的病毒类型，仅从基因表达数据出发，使得一些病毒性疾病的靶点预测无法开展。病毒蛋白与人类蛋白的相互作用数据^[42]是另一种涉及多种病毒性疾病的数据类型，当前已有很多研究在蛋白质组学层面分析病毒感染后的致病机制^[42-43]。因此，基于病毒与人类的相互作用数据提取病毒感染相关的代谢特征，并设计通用的病毒性疾病靶点预测方法，具有重要的科学意义。这使得在某种病毒性疾病基因表达数据缺失的情况下，可以从相互作用数据出发，实现抗病毒靶点的预测。此外，蛋白质作为基因的直接产物，在生物学意义上，病毒蛋白与宿主人类蛋白的相互作用在很大程度上依赖于基因表达情况。从基因表达数据和蛋白-蛋白相互作用数据出发，不仅可以进行靶点发现，还可以通过这些数据的相互印证，增强发现的靶点的说服力和科学有效性。

近年来一些研究已经开始尝试使用神经网络等先进技术来构建蛋白-蛋白相互作用（PPI）网络（图 3.1）^[84-85]。这种方法的发展预示着不久的将来可能直接从病毒序列数据出发，构建病毒蛋白与人类蛋白的 PPI 网络，从而便于进行抗病毒靶点的预测，流程示意如图 3.2 所示。因此，从病毒 PPI 数据出发设计病毒性疾病的通用靶点预测方法，不仅可行，而且具有重要的科学和应用价值。

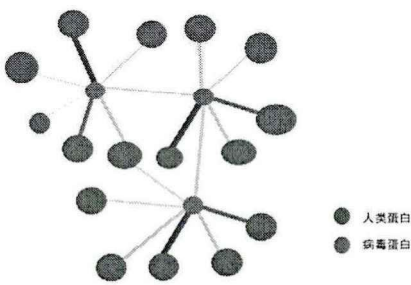


图 3.1 PPI 结构示意图

注：其中连接线的颜色和粗细表示两者相互作用强度

当前关于病毒性疾病 PPI 网络相关的研究,大多通过识别网络中的关键节点(如中心蛋白或桥接蛋白),来预测潜在的药物靶标^[86]。也有一些研究通过网络动态模拟或机器学习方法,预测对病毒感染关键的模块^[87-88]。这些研究的重心仍放在蛋白质组学分析上,而未有研究将病毒 PPI 网络与代谢网络模型结合,展开抗病毒靶点的预测。因此,从病毒与人类的 PPI 网络中提取出与病毒感染致病机制相关的代谢特征,是本章的关注重点,也是需要解决的问题。

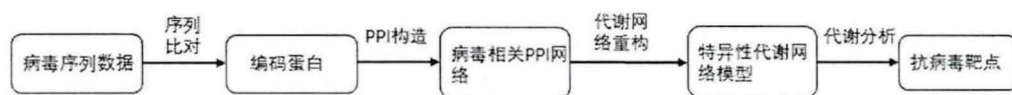


图 3.2 从病毒序列数据出发进行抗病毒靶点预测流程示意图

本章设计并实现了一种基于相互作用数据的病毒性疾病通用靶点预测方法,使得基因表达数据缺失的情况下,使用相互作用数据也能进行抗病毒靶点的预测。从病毒蛋白与人类蛋白之间强相互作用的数据中提取关键的代谢特征,完成代谢网络模型的重构。病毒的 PPI 数据总结了人体多个组织的相互作用关系,不会受到组织特异性的影响。同时,这一重构工作侧重于表征病毒为了便于其自身复制增殖而对宿主代谢系统进行的调控,有效地消除了原发组织的代谢特征的影响。在靶点发现环节,本章采用了与第二章相同的方法,通过代谢网络模型中的信息集合,构建以 VBOF 通量最大化为优化目标的线性规划计算模型进行代谢分析。并通过调整线性规划约束,模拟基因敲除和毒性测试的生物学实验过程,从而预测出既具有抗病毒性质且对正常细胞组织无毒的潜在靶点。未来从序列信息出发构造 PPI 网络的可能性赋予了该方法在抗病毒靶点预测应用上更高的潜力。此外,使用 PPI 数据预测的抗病毒靶点与基于基因表达数据预测的结果的互相验证,增强了靶点的可靠性与准确性。

3.2 数据与方法

本章利用 IntAct 数据库中提供的病毒蛋白与人类蛋白的相互作用数据,挑选出表现出强烈相互作用的人类蛋白质。将这些筛选出的人类蛋白质,映射到与病毒感染异常代谢活动相关的代谢反应上去,进行病毒性疾病代谢网络模型的重构。将病毒性疾病和正常细胞的代谢网络模型中的代谢分析过程建模成线性规划计算模型,通过调整模型中的约束条件,模拟基因必要性测试和细胞毒性测试生物学实验,预测无毒的抗病毒靶点。总体流程如算法 3.1 所示。

算法 3.1 基于 PPI 数据的通用抗病毒靶点预测算法

Input: 病毒蛋白与人类蛋白相互作用数据, 包括相关人类蛋白 P , 及这些蛋白对应的相互作用得分 S_p ; 人类全基因组规模代谢网络模型 Recon3D; 病毒的生物合成函数 VBOF

Output: 病毒性疾病的靶点基因集合 Target

- 1 核心反应集获取, 将 PPI 数据中的人类蛋白 P 及相互作用得分 S_p 映射到一组核心代谢反应 $core_rxns$, 计算其得分 $rxnScore$;
- 2 代谢网络重构, 基于 Recon3D, VBOF 和 $core_rxns$, 使用 FASTCORE 算法重构出病毒性疾病的代谢网络模型 $Specific_Model$;
- 3 基因必要性测试, 在 $Specific_Model$ 上进行单基因敲除实验, 获得必要基因集合 $essentialGenes$;
- 4 细胞毒性测试, 剔除 $essentialGenes$ 中对正常细胞组织有害的基因, 获得靶点基因集合 Target;
- 5 return Target;

3.2.1 数据与预处理

病毒蛋白与宿主体蛋白之间的相互作用在很大程度上可以揭示病毒感染后对宿主细胞代谢异常的调控机制。IntAct 数据库^[67]是一个开源的、公共可用的数据库, 旨在存储、共享和查询蛋白质-蛋白质相互作用 (Protein-Protein Interaction, PPI) 数据。该数据库由欧洲生物信息学研究所 (EBI)^[89]维护, 收集了来自文献报道和实验研究的 PPI 数据, 覆盖了各种生物体, 包括病毒和人类。IntAct 数据库中丰富的病毒相关 PPI 数据使其成为研究病毒异常代谢过程和病毒疾病机制的一个重要工具, 使得从病毒与人类相互作用关系出发研究宿主细胞的代谢异常提供了数据支持^[90-91]。

本章使用 IntAct 数据库中关于 11 种病毒性疾病的病毒蛋白与人类蛋白的相互作用数据, 旨在预测抗病毒疾病的潜在靶点。这些数据源自多种实验技术, 如酵母双杂交、共免疫沉淀和质谱分析, 还有一些数据源于对现有文献的总结。IntAct 数据库将这些来源不同的数据集成在一个数据库中, 建立病毒蛋白与人类蛋白之间的 PPI 网络, 供相关研究人员分析。这些相互作用数据包含了相互作用双方的名称、类型、所属物种、检测方法、相互作用类型以及相互作用得分 (衡量相互作用强弱的属性) 等信息。通过分析这些相互作用, 试图了解病毒感染后宿主细胞的异常代谢状态。

IntAct 数据库收集了包括病毒与病毒, 病毒与人类之间的相互作用数据, 涵盖了蛋白质、DNA/RNA、小分子、肽段等多种实体。通过将特定病毒作为关键词在 IntAct 数据库门户网站进行搜索, 获取该病毒相关的所有相互作用数据。鉴于 IntAct 数据库的下载限制, 采用网络爬虫技术, 利用 Selenium 模块模拟用户与浏览器的交互过程, 包括输入病毒名称、应用网站过滤器进行初步筛选等操作, 以此来访问所需数据页面。之后从网页源码中提取每条相互作用数据的详细信息, 如相互作用双方的名称、物质类型、所属物种及相互作用得分等。接下来,

对收集到的数据进行了进一步的处理。通过筛选,专注于那些相互作用双方分别为病毒和人类的数据条目,并进一步细化到仅包含蛋白质相互作用的部分。这些筛选后的相互作用数据,特别是人类蛋白名称和相互作用得分,为后续的研究和分析研究提供了基础。对11种病毒性疾病的PPI数据所进行的总体预处理流程如算法3.2所示。预处理后的数据详情如表3.1所示。

算法 3.2 病毒蛋白与人类蛋白相互作用数据预处理算法

Input: 11种病毒性疾病的PPI数据 `ppiData`, 包含作用双方的名称 `idA,idB`, 类别 `typeA,typeB`, 所属物种 `specieA,specieB` 以及相互作用得分 `score`; 人类全基因组代谢网络 `Recon3D`

Output: 预处理后的数据, 包含人类基因 `G` 及其对应的相互作用得分 `S`

```
1 ppiData=ppiData[specieA=Human && specieB=Virus];
2 ppiData=ppiData[typeA=protein && typeB=protein];
3 ppiData[hGene]=map_protein2gene(ppiData[idA]);
4 # 使用 DAVID 数据库将人类蛋白映射到对应编码该蛋白的基因;
5 ppiData=ppiData[ppiData[hGene] in Recon3D.genes];
6 G=ppiData[hGene];
7 S=ppiData[score];
8 return G,S;
```

表 3.1 11种病毒性疾病的相互作用数据详情

病毒名称	相互作用数目
丙肝病毒 (Hepatitis C virus)	688
D 型肝炎病毒 (Hepatitis delta virus)	279
人冠状病毒 (Human coronavirus)	396
人体免疫缺陷病毒 1 (Human immunodeficiency virus 1)	1825
人类乳头瘤病毒 16 型 (Human papillomavirus type 16)	735
人类 SARS 冠状病毒 (Human SARS coronavirus)	2297
甲型流感病毒 (Influenza A virus)	1096
中东呼吸综合征冠状病毒 (MERS coronavirus)	612
新型冠状病毒 (SARS-CoV-2)	5895
扎伊尔埃博拉病毒 (Zaire ebolavirus)	262
寨卡病毒 (Zika virus)	448

3.2.2 代谢网络模型重构与抗病毒靶点预测

本章希望从病毒与人类宿主细胞之间的PPI数据出发,进行代谢网络模型的重构。重构的代谢网络模型能够精确表征病毒感染人体后对宿主细胞代谢系统的调控,使宿主细胞的代谢朝着更适合病毒自身的生物质合成需求的方向发展。

此外，这些模型的设计也考虑到了尽可能减少原发组织特征带来的干扰。

目前，并未发现有研究将病毒 PPI 数据集成到代谢网络模型中进行分析。从病毒蛋白与人类蛋白的相互作用数据中提取病毒感染相关的代谢特征是构建代谢网络模型的核心步骤，也是整个工作的重点。具体来说，从病毒的 PPI 数据中识别出一组关键的病毒感染过程中的核心代谢反应集合，这不仅是代谢网络模型重构的关键，也构成了整个靶点预测流程中的一大挑战。

通过以下几个关键步骤将 PPI 数据映射到一组核心代谢反应上去，流程如图 3.3 所示。首先基于预处理的 PPI 数据，将其中与病毒蛋白存在相互作用的人类蛋白映射到相应的编码基因上，并记录人类基因的相互作用得分。从蛋白质到编码基因的映射，需要从蛋白质的氨基酸序列出发，进行 blast 比对^[92]找到相应的基因序列，从而得到对应的编码基因。DAVID 数据库^[74]中整合了包括基因、蛋白质序列数据库以及功能注释资料等多种信息，利用其基因 ID 转换工具可以直接实现蛋白质名称到对应编码基因名的映射。这部分人类基因的得分与对应蛋白质的相互作用得分相同，是一个 0 到 1 之间的实数。

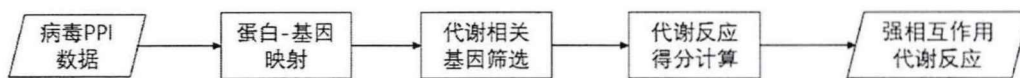


图 3.3 基于病毒 PPI 数据获取强相互作用代谢反应的流程图

之后筛选出与代谢过程紧密相关的人类基因。对于第一步获得的基因集合，筛选出包含在 Recon3D 模型基因集合的部分。该部分基因被认为与人类代谢相关。最后需要从病毒相互作用的人类代谢基因出发，得到一组与病毒强相互作用的人类代谢反应。由于一个基因通常受到多个基因的共同调控，而各个基因的相互作用得分各不相同，会对代谢反应的相互作用得分产生影响。因此，不能使用第二章的基因-反应映射方式，使用简单的布尔计算将基因集合映射到代谢反应集合上，这会造成相互作用得分数据的丢失。对基因-反应（GR）映射规则中的 or/and 逻辑关系进行了调整，使得代谢反应的相互作用得分计算依赖于相关基因的得分数据。具体来说，对于 or 逻辑关系的基因，取两个基因之间较高的相互作用得分作为 or 运算的结果；对于 and 关系，则采用两个基因之间较低的相互作用得分。相互作用得分的具体计算公式见公式 3.1 和 3.2。

$$gene1 \text{ or } gene2 = \max(gene1, gene2) \quad (3.1)$$

$$gene1 \text{ and } gene2 = \min(gene1, gene2) \quad (3.2)$$

将与病毒蛋白相互作用无关的人类代谢基因得分设置为 0，对于每一个代谢反应，利用调整后的 GR 规则，计算出该代谢反应的相互作用得分。本章设定 0.75 作为阈值，仅将得分超过此阈值的代谢反应视为与病毒蛋白与人类蛋白之间发生强相互作用的代谢调控反应。

通过三个关键步骤成功地将病毒 PPI 数据映射到一组代谢反应集合，并计算了对应的相互作用得分。筛选出其中相互作用得分高的代谢反应，将其作为代谢网络模型重构过程中的核心反应集。为确保重构的代谢网络模型能够准确反映病毒本身的基本物质合成需求，将病毒的生物质生长函数 (VBOF) 手动添加到核心反应集中。基于核心反应集，以 Recon3D_Virus 为背景模型，使用 FASTCORE 算法重构出了一个满足通量一致性的最小代谢网络模型。该模型综合反映了病毒感染人体细胞后与人体细胞发生强烈相互作用的异常代谢部分，为进一步的病毒行为分析和抗病毒策略开发提供了重要基础。

本研究旨在通过降低模型中的 VBOF 反应通量，从而减缓病毒生物质合成的速度，达到抑制病毒复制的目的。与第二章相同，基于重构代谢网络模型建立线性规划计算模型，通过调整约束条件模拟基因敲除和毒性测试实验，以预测具有抗病毒特性且对正常细胞无害的潜在靶点。具体的算法流程分别在算法 2.3 和算法 2.4 中进行了详细说明。

3.3 实验结果

3.3.1 相互作用相关基因

基于病毒蛋白与人类蛋白之间的相互作用数据，探讨了病毒侵入人体细胞后所引发的代谢异常状态。分析病毒感染宿主细胞后对宿主代谢系统的调节作用时，病毒蛋白相互作用的人类蛋白质具有重要意义。通过对编码这些蛋白的人类基因进行分析，旨在揭示病毒性疾病致病机制中的关键共有基因。对与病毒相互作用相关的人类基因进行了统计分析，各基因出现的频次分布如图 3.4 所示。结果显示，HACD3 基因在多数病毒性疾病的蛋白-蛋白相互作用 (PPI) 网络中频繁出现，表明其在多种病毒感染机制中扮演着重要角色。

进一步的研究发现，HACD3 基因与病毒入侵宿主细胞后的蛋白质水平和基因组复制的正向调节关系密切。特别是，Taguwa 等人^[93]的研究揭示了在丙型肝炎病毒复制相关的 NS5A 宿主蛋白结合中，HACD3 基因编码的人丁酸盐诱导转

录物 1 (hB-ind1) 与 FKBP8 结合, 并通过其 N 末端 p23 同源性领域的 FXXW 基序与 HSP90 相互作用, 对 HCV 的 RNA 复制及 JFH1 病毒的传播起到至关重要的作用。此外, Milev 等人^[94]的研究表明, HACD3 基因与 HIV-1 复制之间也存在联系, 其涉及的脂质代谢途径可能间接影响 HIV 的生命周期, 尤其是在病毒的组装、成熟及释放阶段。HIV 病毒粒子的组装过程需在宿主细胞的膜上进行, 而这一过程依赖于细胞内脂质的调动。

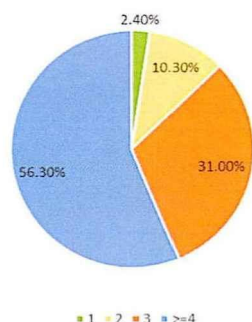


图 3.4 PPI 相关的人类基因频次图

为了深入掌握与病毒存在相互作用的人类基因在功能和代谢通路上的特性, 借助 DAVID 数据库的功能注释工具进行了分析。通过设定 $p\text{-value} \leq 0.05$ 为具有统计学显著性的门槛, 对 11 种病毒性疾病相关的人类基因进行了 GO 分析和 KEGG Pathway 分析, 旨在探究这些基因在基因功能和代谢路径上的富集情况。GO 分析的结果如图 3.5 所示, 与病毒互作的人类基因主要功能富集在蛋白质结合、ATP 结合、金属离子结合过程以及多种膜结构的构成, 这表明病毒感染人体后, 其关键活动之一是改变宿主细胞膜的组成, 并利用宿主的物质和能源进行病毒的复制和增殖。

KEGG 分析的结果, 如图 3.6 所示, 揭示了这些基因在碳水化合物代谢、产热、氧化磷酸化等代谢路径上的富集, 以及在某些疾病路径 (例如朊病毒病、亨廷顿症) 上的富集。这表明病毒侵入宿主细胞后, 通过代谢重编程干扰宿主的代谢系统, 并可能影响与其他疾病相关的代谢途径, 从而引起并发症。

3.3.2 重构代谢网络模型结构与分析

基于蛋白-蛋白相互作用 (PPI) 数据, 重构了 11 种病毒性疾病的代谢网络模型, 其规模在图 3.7 中展示。这些代谢网络模型平均包含 591 个基因、943 个代谢反应及 849 个代谢物, 与 Recon3D 的规模相比, 显著减小, 有效去除了大量冗余信息。

同时, 这些模型确保了代谢通量的一致性以及病毒生物物质合成与增长的能力, 准确反映了病毒与人类细胞相互作用下的代谢系统特征。特别地, MERS 冠

状病毒（MERS-CoV）和 SARS-CoV-2 的特异性模型规模最大，而扎伊尔埃博拉病毒（Zaire ebolavirus）与寨卡病毒（Zika Virus）的模型规模最小，这与它们与人类蛋白之间的相互作用数量相符。

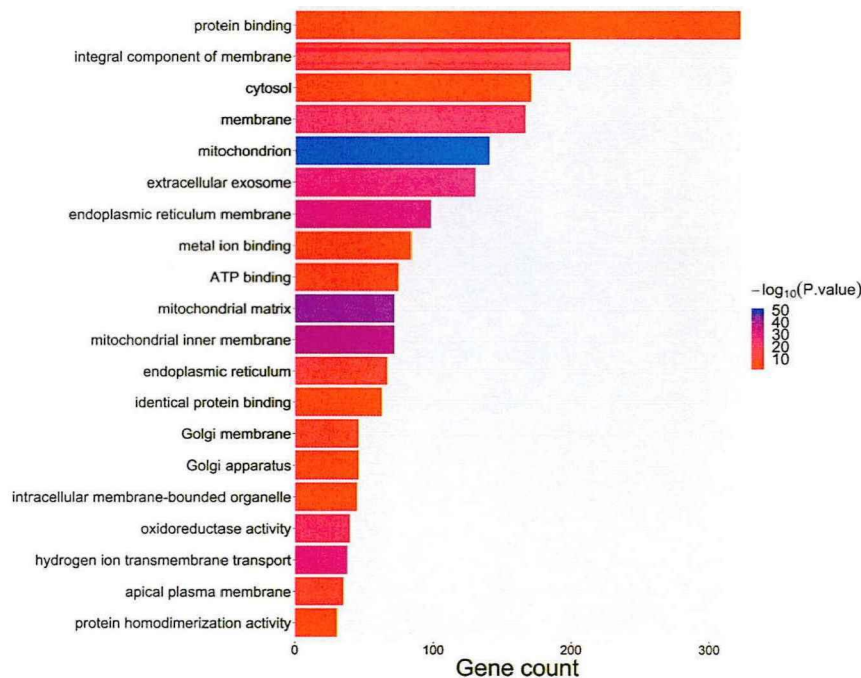


图 3.5 对 11 种病毒性疾病相互作用的人类基因的 GO 分析结果

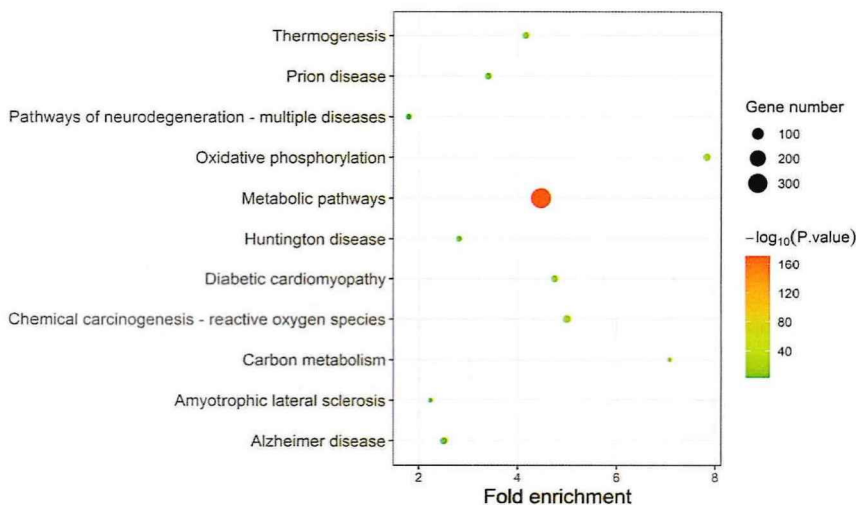


图 3.6 对 11 种病毒性疾病相互作用的人类基因的 KEGG Pathway 分析结果

为深入理解 11 种病毒性疾病在宿主代谢系统中的异常调控共性，对这 11 个重构代谢网络模型中的共有代谢反应进行了统计，共识别出 17 种代谢反应，其详细信息列于表 3.2。这些共有反应包括 3 种交换需求反应、2 种糖酵解和糖异

生反应、3种氧化磷酸化反应、4种运输反应（涉及氧气、ATP/ADP以及线粒体与细胞质之间的运输），以及一些关键的氨基酸和核苷酸转换反应。这些代谢反应直接与病毒所需的基本物质和能量需求相匹配，揭示了病毒在宿主细胞内维持生命状态所进行的基本代谢活动。

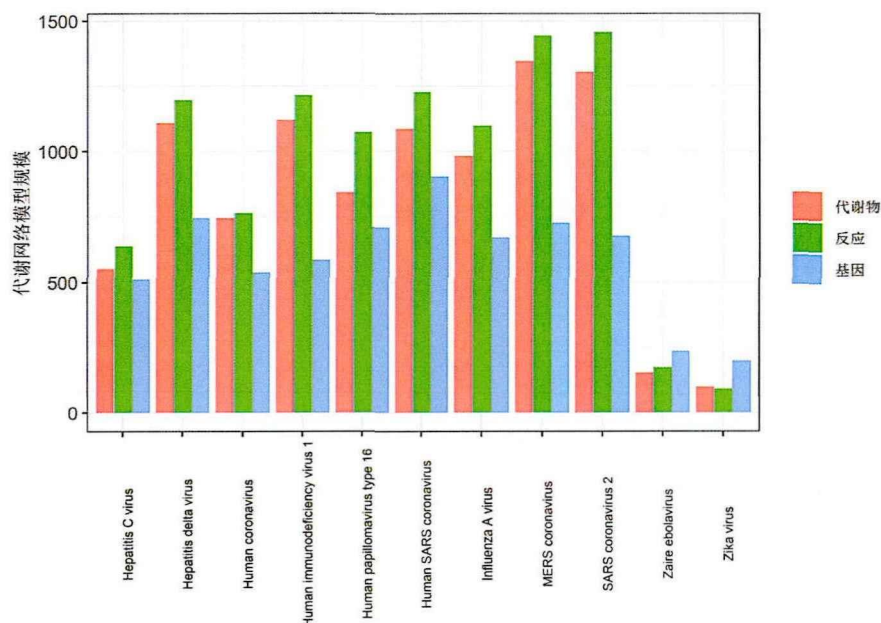


图 3.7 基于 PPI 数据重构的 11 个代谢网络模型规模

此外，以 VBOF 作为代谢网络模型的代谢目标，使用通量平衡分析方法计算了这些共有代谢反应的通量分布情况，结果如图 3.8 所示。值得注意的是，这 17 个共有代谢反应在不同代谢网络模型上的通量方向相同，即对于每一个共有代谢反应，其在 17 个模型上以相同的方向进行物质转换，这说明病毒感染后在宿主细胞中的异常代谢调控大致类似，值得进一步探究。其中 EX_o2[e]，O2t 和 ATPS4mi 三个代谢反应在绝大多数代谢网络模型上表现出了很高的通量，反映了病毒复制增殖过程中对氧气以及能量的大量需求。

3.3.3 抗病毒靶点预测结果

基于蛋白-蛋白相互作用（PPI）数据，成功重构了 11 种病毒性疾病的代谢网络模型，并利用模拟进行了基因必要性测试和细胞毒性测试，旨在预测针对这些病毒的潜在抗病毒靶点基因。各病毒性疾病模型所对应的必要性基因和靶点基因数量展示在图 3.9 中。结果显示，不同病毒模型中预测出的必要性基因与靶点基因数量有明显差异，这一差异反映了各自网络模型的规模和构成的不同。

对 11 个代谢网络模型预测的靶点基因进行综合分析时，发现了病毒感染后

对宿主代谢系统异常调控的关键共性基因。特别是, CTH 和 DCTD 这两个基因在所有的 11 种病毒性疾病的靶点基因集合中单独或组合出现。CTH 或 DCTD 的敲除实验显示, 两者对某些氨基酸/核苷酸代谢通路产生影响, 从而抑制了病毒生物质的生长, 而对宿主的正常代谢过程未产生不良影响。这一靶点预测结果与使用基因表达数据进行的预测结果高度一致, 证明了 CTH 和 DCTD 基因在病毒感染导致的代谢异常调控中的关键作用, 并从另一数据角度验证了靶点预测结果的正确性和可靠性。

表 3.2 11 个代谢网络模型共有的 17 个代谢反应

反应名称	反应式	反应类型
AMPDA	$\text{h2o}[\text{c}]+\text{h}[\text{c}]+\text{amp}[\text{c}]\rightarrow\text{nh4}[\text{c}]+\text{imp}[\text{c}]$	核苷酸转换
ATPtm	$\text{adp}[\text{c}]+\text{atp}[\text{m}]\rightarrow\text{atp}[\text{c}]+\text{adp}[\text{m}]$	ADP/ATP 转运
CATm	$\text{h2o2}[\text{m}]\rightarrow\text{h2o}[\text{m}]+\text{o2}[\text{m}]$	氧气供应
O2tm	$\text{o2}[\text{c}]\rightarrow\text{o2}[\text{m}]$	氧气运输
PPAm	$\text{h2o}[\text{m}]+\text{ppi}[\text{m}]\rightarrow\text{h}[\text{m}]+\text{pi}[\text{m}]$	氧化磷酸化
EX_ahdt[e]	$\text{ahdt}[\text{e}]\rightarrow$	交换需求反应
EX_HC01361[e]	$\text{HC01361}[\text{e}]\rightarrow$	交换需求反应
r0276	$\text{nh4}[\text{c}]+\text{nadp}[\text{c}]+\text{imp}[\text{c}]\rightarrow\text{h}[\text{c}]+\text{nadph}[\text{c}]+\text{gmp}[\text{c}]$	嘌呤分解代谢
r0707	$\text{h2o}[\text{e}]+\text{ahdt}[\text{e}]\rightarrow\text{h}[\text{e}]+\text{pi}[\text{e}]+\text{HC01361}[\text{e}]$	其他
r1156	$\text{uri}[\text{c}]+\text{dutp}[\text{c}]\rightarrow\text{h}[\text{c}]+\text{dudp}[\text{c}]+\text{ump}[\text{c}]$	核苷酸转换
PPItm	$\text{ppi}[\text{c}]\rightarrow\text{ppi}[\text{m}]$	线粒体运输
Ex_o2[e]	$\text{o2}[\text{e}]\rightarrow$	交换需求反应
GK1	$\text{atp}[\text{c}]+\text{gmp}[\text{c}]\rightarrow\text{adp}[\text{c}]+\text{gdp}[\text{c}]$	糖酵解和糖异生
NDPK6	$\text{atp}[\text{c}]+\text{dudp}[\text{c}]\rightarrow\text{adp}[\text{c}]+\text{dutp}[\text{c}]$	糖酵解和糖异生
O2t	$\text{o2}[\text{e}]\rightarrow\text{o2}[\text{c}]$	氧气运输
ATPS4mi	$\text{adp}[\text{m}]+\text{pi}[\text{m}]+\text{h}[\text{i}]\rightarrow\text{h2o}[\text{m}]+\text{h}[\text{m}]+\text{atp}[\text{m}]$	氧化磷酸化
CYOR_u10mi	$\text{h}[\text{m}]+\text{ficytC}[\text{m}]+\text{q10h2}[\text{m}]\rightarrow\text{q10}[\text{m}]+\text{focytC}[\text{m}]+\text{h}[\text{i}]$	氧化磷酸化

在探索现有药物对病毒性疾病的治疗潜力方面, 利用 DrugBank 数据库^[95]分别针对 CTH 和 DCTD 基因的相关药物进行了研究, 旨在评估这些药物对抗病毒性疾病的可复用性。尽管目前尚未发现针对 CTH 基因的药物与病毒性疾病的直接联系, 但对于 DCTD 基因而言, 情况则不同。当前 DrugBank 中存在一种针对 DCTD 基因的靶向药物 Cytarabine, 它是一类嘧啶核苷类似物, 已展现出抗病毒和免疫抑制的特性^[96]。它通过与人体正常的核苷酸竞争, 干扰病毒 DNA/RNA 链的复制过程, 可能通过提前终止或延长 DNA/RNA 链的复制, 或诱导致命突变

来抑制病毒的复制。

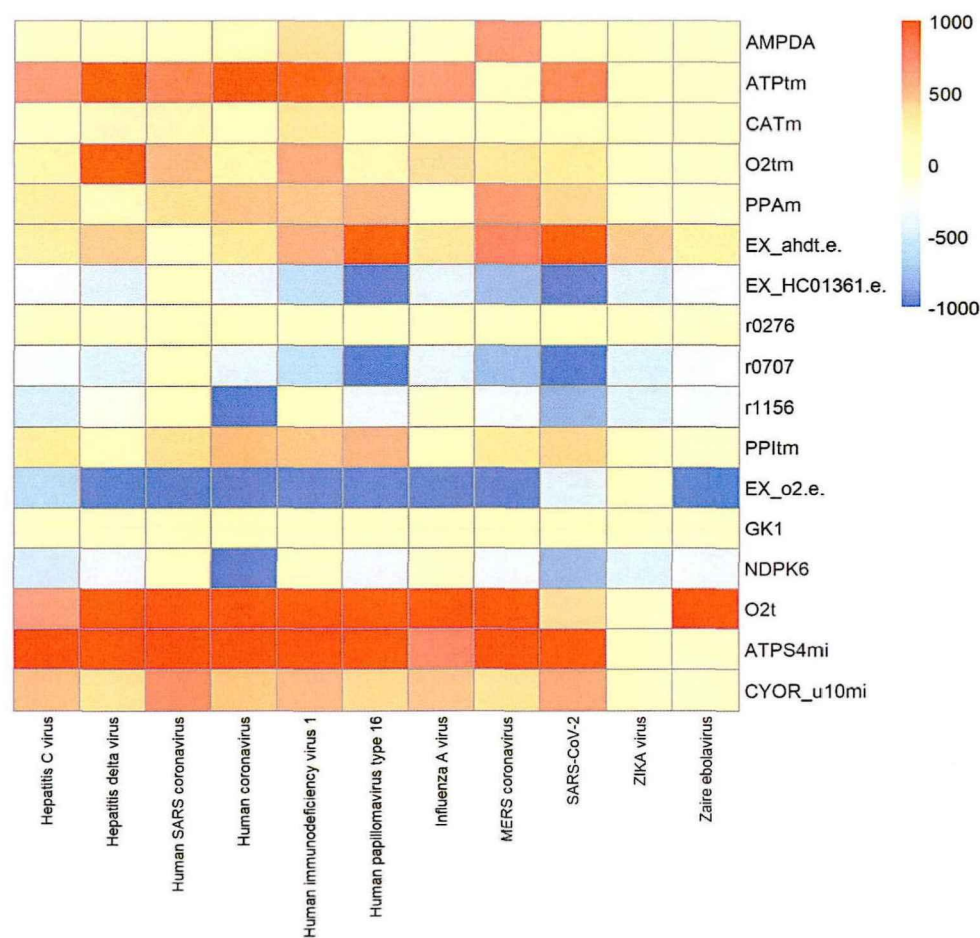


图 3.8 17 个共有代谢反应的通量分布 heatmap

核苷类似物已成为抗病毒药物领域最有希望的类别之一^[97]，用于治疗包括疱疹病毒、人类免疫缺陷病毒、乙型肝炎病毒等在内的多种病毒感染。以近些年全球范围内极具影响力的 SARS-CoV-2 病毒为例，核苷类似物 Remdesivir 已成为其主要的治疗药物之一^[98]，通过与 ATP 竞争抑制病毒 RNA 聚合酶的活性，从而达到抑制病毒复制的目的。

通过观察仿真实验结果发现 CTH 和 DCTD 在 DELTA 网络中对多种病毒显示出抑制作用，这暗示它们可能作为广谱病毒性疾病的潜在基因靶点。因此，深入研究 CTH 和 DCTD 的作用机制，对于开发新型抗病毒药物和治疗策略具有至关重要的意义。尤其是 DCTD 基因相关药物 Cytarabine 已在多种疱疹病毒、腺病毒、痘病毒感染中显示出明显的抑制效果，在其他病毒性疾病中的药物重复使用潜力值得进一步探索。

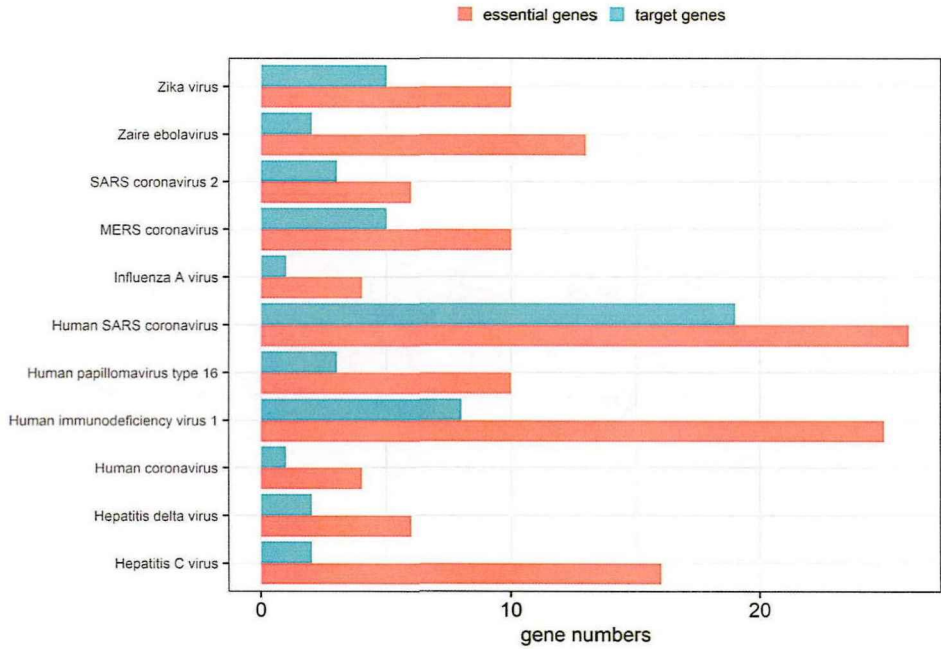


图 3.9 11 个代谢网络模型的必要性基因和靶点基因规模

3.4 本章小结

病毒性疾病相关的数据通常分布零散且依赖于复杂的生物实验设计，增大了病毒性疾病通用研究的难度。当前，基因表达数据与病毒蛋白与人类蛋白相互作用（PPI）数据较为容易获取，并在生物学上存在明显的上下游关系，这使得基于这些数据预测的靶点可以相互验证，增强预测结果的准确性和有效性。本章设计了一种基于 PPI 数据的病毒性疾病通用靶点预测方法，使得病毒基因表达数据缺失时，使用 PPI 数据也能够实现抗病毒靶点的预测。从病毒蛋白与人类蛋白的相互作用数据出发，重构了针对 11 种病毒性疾病的代谢网络模型。利用代谢网络模型中包含的信息，将代谢分析过程建模成一个线性规划计算模型。通过调整线性规划约束条件，模拟基因必要性和毒性测试生物学实验，预测对正常细胞无害的抗病毒靶点。实验结果表明，CTH 和 DCTD 两个基因在多种病毒性疾病中被预测为靶点基因，对抑制病毒复制进程具有一定的广谱性。这与基于基因表达数据进行的抗病毒靶点预测结果高度一致，进一步验证了实验方法的有效性和准确性，也为病毒性疾病靶点预测的研究开辟了新的路径和视角。

第4章 不依赖 VBOF 的病毒性疾病通用靶点预测方法

4.1 背景及意义

前两章的研究实现了基于基因表达数据和蛋白-蛋白相互作用数据来预测病毒性潜在靶点的方法。这些方法的共同之处在于，它们都将病毒生物量增长的反应视为重构的代谢网络模型中的主要代谢目标，通过计算基因敲除后 VBOF 的通量值，将抑制病毒生物量生长且对正常细胞无毒的基因预测为靶点。这样的策略使得 VBOF 成为抗病毒靶点预测方法中不可或缺的一部分。

尽管 Aller 等研究者的方法能够在尽可能的范围内模拟病毒生物合成过程中的物质与能量需求，并已在众多研究中得到应用^[60,99-100]，其方法论本身仍然存在一定的局限性。这些局限主要集中在病毒包膜组成和脂质代谢的考虑不足，以及病毒蛋白质拷贝数数据的准确性和一致性问题上。首先，由于缺乏针对病毒相关脂质的详尽化学计量学信息，以及模型本身不包含动力学特性，Aller 等人在计算病毒生物合成代谢流（VBOF）时未能考虑病毒包膜的具体组成，包括脂质成分。因此，病毒的进入和释放过程，特别是包膜获取过程，未在模型中得到模拟。其次，在关于病毒蛋白质拷贝数的信息收集方面，Aller 等人根据文献中关于病毒总体结构的描述，推断了结构性与非结构性多肽之间的比例，未能充分考虑不同结构性/非结构性蛋白之间可能存在的差异性。此外，实验手段获取病毒蛋白拷贝数的方法可能因实验操作、设备差异及样本选择等多种因素而有所不同。例如，在针对 SARS-CoV-2 的基因编辑模型（GEM）相关研究中，Renz 等人^[61]与 Delattre 等人^[101]所报道的 SARS-CoV-2 蛋白质拷贝数数据之间的差异，可能会对抗病毒靶点的预测结果造成影响。

为了解决由于 VBOF 构造过程中数据缺失而无法进行抗病毒靶点预测的问题，本章在第二章方法的基础上，构建了一种新的代谢反应通量的计算模型，消除现有方法进行抗病毒靶点预测时对 VBOF 的依赖。具体地，基于基因表达数据识别出感染前后表达差异显著的基因集，以 Recon3D 作为背景网络模型，重构了针对 15 种病毒疾病的代谢网络模型。基于代谢网络模型所包含的代谢反应系数矩阵以及热力学信息等数据，构建了一个以基因表达水平与代谢反应通量一致性最高为优化目标的线性规划计算模型。结合病毒感染前后宿主细胞中的基因表达差异，使用线性规划模型计算感染细胞与正常细胞的代谢通量分布。并通过调整线性规划模型中的约束条件，模拟基因敲除的生物效果，重新计算感染细胞与正常细胞的代谢通量分布，识别出能显著降低感染细胞与正常细胞之间代谢通量差异的基因。最后，通过细胞毒性测试的仿真实验，鉴定出对正常细胞组织无毒的靶点基因。此前的方法主要依赖于降低病毒生物量优化函数（VBOF）

的通量以达到抑制病毒复制的目的，本章提出的方法改进了前两章的研究策略。新的方法将基因敲除后缩短感染状态与正常状态下代谢通量距离作为评估抗病毒特性的新标准，消除了靶点预测过程中对 VBOF 的依赖。

4.2 数据与方法

为了克服在对病毒性疾病进行代谢分析时对病毒生物物质合成函数（VBOF）的依赖，本章设计了一个新颖的基于约束的计算模型，旨在计算感染前后整个代谢系统的通量变化，来替代通量平衡分析这种以物质增长等特定反应为代谢目标的计算方法。通过这种方法，旨在深入理解病毒感染对宿主细胞代谢系统的全局影响，从而揭示病毒感染后代谢变化的细节。方法流程如图 4.1 和算法 4.1 所示。本章使用的数据及预处理方式均与第二章相同，使用了从 NCBI 的 GEO 数据库中下载的关于 15 种病毒性疾病的基因表达数据，并根据实验方案设计不同，将数据分成感染细胞与正常对照两组，数据详情如表 2.1 所示。

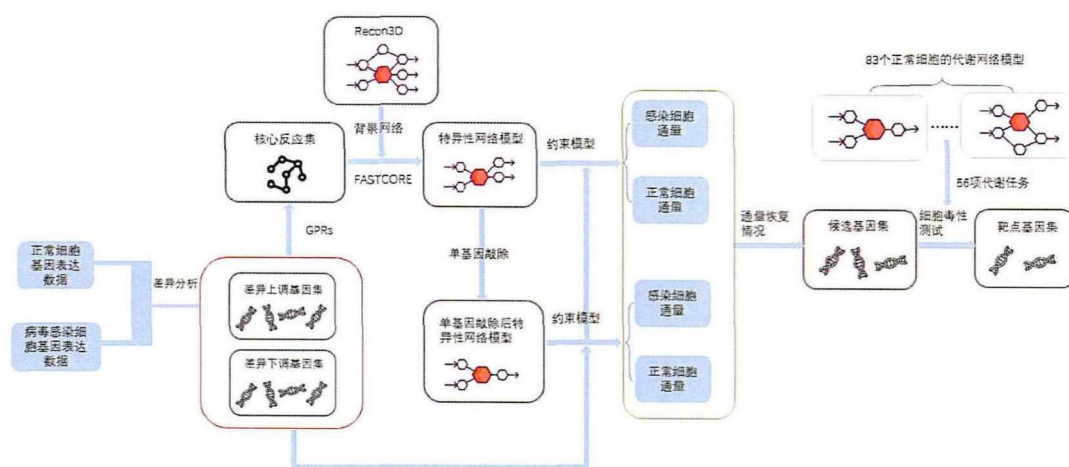


图 4.1 不依赖 VBOF 的通用抗病毒靶点预测方法示意图

4.2.1 代谢网络模型重构

本章后续的分析工作均以重构的代谢网络模型为基础展开，因此代谢网络模型重构是整个靶点预测流程的基础。整个代谢网络模型的重构流程如下：首先采用 Limma 工具包^[68-69]对病毒性疾病相关的基因表达数据执行差异表达分析，目的是鉴定在正常细胞与病毒感染细胞之间存在显著表达差异的基因集，涵盖了表达水平上调及下调的基因。相比之下，本章与第二章中“DELTA”网络模型的重构流程存在细微差异，“DELTA”网络模型以病毒生物量合成为代谢目标，主要关注感染细胞中促进病毒复制与生长的代谢途径的活跃性。然而，为了摆脱

对病毒生物量合成函数 (VBOF) 的依赖, 采用约束模型来分析代谢系统的整体通量变化, 并通过还原这种通量变化来抑制病毒的复制过程。在此背景下, 不能忽略在感染细胞中相对于正常细胞显著下调的基因。

算法 4.1 不依赖 VBOF 的通用抗病毒靶点预测算法

Input: 病毒性疾病的基因表达数据; 人类全基因组代谢网络 Recon3D; 病毒的物质合成函数 VBOF

Output: 病毒性疾病的靶点基因集合 Target

- 1 数据预处理, 获得正常细胞和感染细胞的基因表达数据 NormalData, TumorData;
 - 2 差异表达分析, 根据差异分析获取感染细胞相对正常细胞显著上/下调的基因集合 Gup, Gdw;
 - 3 核心反应集获取, 根据 Recon3D 的 GR 映射规则获取 Gup 和 Gdw 对应的显著上调反应集 Rup, Rdw 作为核心反应集;
 - 4 代谢网络重构, 使用 FASTCORE 算法和核心反应集 Rup, Rdw 重构出特异性代谢网络模型 Specific_Model;
 - 5 使用约束计算模型计算 Specific_Model 中感染状态和正常状态的代谢通量分布;
 - 6 单基因敲除实验, 再次使用约束计算模型, 筛选出敲除后感染状态和正常细胞代谢通量差异显著减小的基因 candidates;
 - 7 细胞毒性测试, 剔除 candidates 中对正常细胞组织有害的基因, 获得靶点基因集合 Target;
 - 8 return Target;
-

通过筛选出差异表达的上调/下调基因中出现在 Recon3D 模型中的部分, 即与人类代谢相关的差异表达基因, 将这部分基因经 GR 规则映射到一组核心代谢反应集合上。基于获取的核心反应集, 以 Recon3D 作为背景网络模型, 使用 FASTCORE 代谢网络重构算法, 为 15 种病毒性疾病重构了代谢网络模型。这些模型旨在揭示病毒感染后宿主细胞中的代谢活动异常情况。

4.2.2 基因表达水平与代谢反应通量一致性最高的约束计算模型构建

在代谢系统中, 基因对相关的代谢反应其调控作用, 其中基因的表达水平直接影响代谢系统中对应的代谢反应速率, 即代谢通量。为深入分析病毒侵染宿主细胞前后代谢系统通量的变化, 建立了一种约束计算模型。该计算模型基于差异表达基因与代谢网络模型, 旨在寻找一个满足计量学及热力学约束条件的稳态通量分布。使得在此分布条件下, 基因表达水平的变化与相应代谢反应通量的变动之间的一致性最高。换言之, 本章期望通过该模型能够确保基因表达水平的调整 (上调或下调) 能在最大程度上反映为系统下游代谢反应通量的增加或减少。约束模型的具体数学描述如公式 4.1 至 4.6 所展示。

$$\min \left(\sum_{i \in R_{up} \cup R_{dw}} IS_i \right) \quad (4.1)$$

$$s.t. \quad S \cdot v^T = 0, S \cdot v^N = 0 \quad (4.2)$$

$$v_{min}^T \leq v^T \leq v_{max}^T, v_{min}^N \leq v^N \leq v_{max}^N \quad (4.3)$$

$$IS_i \geq \epsilon - (|v_i^T| - |v_i^N|), i \in R_{up} \quad (4.4)$$

$$IS_i \geq \epsilon - (|v_i^N| - |v_i^T|), i \in R_{dw} \quad (4.5)$$

$$IS_i \geq 0, i \in R_{up} \cup R_{dw} \quad (4.6)$$

其中, S 是代谢网络模型中的化学计量矩阵, 包含了各代谢反应中参与的反应物与生成物及其系数之间的关系。代谢反应通量向量 v^T 与 v^N 分别代表感染细胞和正常细胞的代谢反应速率。不一致性得分 IS 旨在量化基因表达水平与代谢通量之间的差异, 模型致力于最小化该得分, 如公式 4.1 所指示。

公式 4.2 和 4.3 定义了代谢系统的质量平衡约束和热力学约束, 确保系统在物质的生成与消耗方面能够达到稳态, 并且保证代谢反应速率遵循一定的上下限制, 这些上下限制信息 v_{min} 与 v_{max} 在相应的网络模型中有所体现。基因表达水平与代谢通量变化之间不一致性的量化通过公式 4.4 至 4.6 进行, 其中阈值 ϵ 被设定为 1, 用于判定反应通量的显著变化。

对于差异基因的处理, 本研究对差异上调和下调基因集合分别进行了分析。以差异上调基因为例, 通过 GR 规则将差异表达基因映射到相应的代谢反应集合 R_{up} 。若感染细胞的代谢反应通量 $|v_i^T|$ 显著高于正常细胞的 $|v_i^N|$ (即差值大于 ϵ), 结合公式 4.6 的约束条件, 相应反应的不一致性得分 IS_i 将为 0; 若不满足此条件, 则表明感染细胞的代谢通量未显著高于正常细胞, 导致基因表达水平与代谢通量出现不一致, 此时公式 4.4 中的 IS_i 将为一个由通量差异驱动的正值。差异下调基因及其对应的代谢反应集合的处理逻辑与此相似, 约束条件见公式 4.5。

代谢网络中的反应可分为可逆反应和不可逆反应两大类。对于可逆反应, 由于反应方向的差异, 其代谢通量可表现为正或负值。然而, 标准线性规划求解器处理绝对值运算存在困难。为解决此问题, 本研究对原有约束模型进行了适当修改, 以便于问题的求解。具体策略包括将可逆反应的代谢通量细分为正向通量 $v_{i,forward}$ 和逆向通量 $v_{i,backward}$, 两者都需遵循热力学约束。对于任一可逆反应, 确保在同一时刻, $v_{i,forward}$ 与 $v_{i,backward}$ 中最多仅有一个通量值大于零, 如公式 4.7 至 4.9 所定义。此种方法的引入, 有效地解决了可逆反应通量及其绝对值的表示问题, 具体表达式见公式 4.10 与 4.11。

$$0 \leq v_{i,forward} \leq y_i^+ \cdot v_{max} \quad (4.7)$$

$$0 \leq v_{i,backward} \leq y_i^- \cdot v_{max} \quad (4.8)$$

$$y_i^+ + y_i^- \leq 1, y_i^+, y_i^- \in [0, 1] \quad (4.9)$$

通过对上述线性规划约束模型的求解, 本研究能够依托于差异表达基因集, 计算得到基因表达水平与代谢反应通量一致性方面表现最优的感染前后代谢反应通量分布。即不一致性得分 IS 达到最小值时, 感染细胞的代谢反应通量向量 v^T 以及正常细胞的通量向量 v^N 。此结果为后续的靶点预测工作奠定了基础, 为理解病毒感染对宿主代谢网络影响提供了重要的量化依据。

$$v_i = v_{i,forward} - v_{i,backward} \quad (4.10)$$

$$|v_i| = v_{i,forward} + v_{i,backward} \quad (4.11)$$

4.2.3 抗病毒靶点预测

本章研究通过构建一个以基因表达水平与代谢反应通量一致性最高的约束计算模型, 旨在深入分析病毒侵染人体宿主细胞前后的代谢通量变化。主要目标是通过模拟基因敲除实验, 缩短感染细胞与正常细胞在代谢网络模型中的代谢反应通量的距离, 从而恢复宿主细胞的正常代谢状态。整个靶点预测流程详细阐述于算法 4.2。

抗病毒靶点预测的步骤如下: 首先, 基于重构的代谢网络模型的反应系数矩阵及热力学数据等信息, 构建的线性规划约束计算模型, 计算得到感染细胞与正常细胞初始状态下的代谢反应通量向量 v^T 与 v^N , 为评估两者通量距离的变化提供基准。随后, 通过调整计算模型的线性规划约束条件, 依次模拟对每个基因进行单基因敲除实验, 计算单基因敲除后, 感染细胞与正常细胞在基因表达水平与代谢反应通量一致性方面表现最优的通量向量 v_{knock}^T 与 v_{knock}^N 。

通过对比单基因敲除实验前后的代谢通量分布, 可以评估单个基因在病毒侵染人体细胞的异常代谢调控中的作用重要性。采用曼哈顿距离作为量化感染细胞与正常细胞之间代谢通量距离的指标, 来识别对宿主细胞正常代谢状态恢复具有显著影响的基因。具体计算方法如公式 4.12 所述。若某基因敲除后, 感染细胞与正常细胞的代谢通量距离相比敲除前缩短了 70%, 则表明敲除该基因在很大程度上可以恢复宿主细胞的正常代谢状态, 并可能对病毒的复制进程产生抑制作用。随后, 对这些基因进行细胞毒性测试, 确保其对人体正常组织无负面作用。只有通过细胞毒性测试的基因, 才被认为是潜在的抗病毒靶点基因。细胞毒性测试的具体流程展示于算法 2.4。

$$Dist(v^T, v^N) = \sum_i |v_i^T - v_i^N| \quad (4.12)$$

算法 4.2 基于约束模型的抗病毒靶点预测算法

Input: 病毒性疾病重构的代谢网络模型 Specific_Model; 差异上/下调基因集合 Gene_up, Gene_dw

Data: lb 和 ub 代表代谢反应速率下界和上界; 83 个人类正常组织细胞代谢网络模型 Normal_Models; 56 项预定义的代谢任务 tasks

Output: 抗病毒靶点基因集合 Target

```

1 Target=[];
2 genes=Specific_Model.genes;
3 rxns=Specific_Model.rxns;
4 vTumor,vNormal=Contrained_Model(Specific_Model,Gene_up,Gene_dw);
5 Target=[];
6 foreach g in genes do
7     rxn_Knock=map_gene2rxn(g,Specific_Model.grRules);
8     tempModel=Specific_Model;
9     foreach r in rxn_Knock do
10         rIndex=rxns.indexof(r);
11         tempModel.lb[rIndex]=tempModel.ub[rIndex]=0;
12     end
13     vTumor',vNormal'=Contrained_Model(tempModel,Gene_up,Gene_dw);
14     rate=Dist(vTumor',vNormal')/(vTumor,vNormal);
15     if rate<=0.3 then
16         | essentialGenes.add(g);
17     end
18 end
19 Target=toxicityTest(essentialGenes,Normal_Models,tasks);
20 return Target;
```

4.3 实验结果

4.3.1 重构代谢网络模型结构与分析

在现有的研究范畴内, 多数研究倾向于将病毒生物量合成功能定位为代谢网络分析的主要目标, 进而将抗病毒策略的焦点集中于抑制病毒增殖, 导致这些方法中 VBOF 是不可缺少的一部分。然而 VBOF 的构造过程可能出现数据缺失或信息不准确的问题, 从而对抗病毒靶点预测流程造成影响。通过建立约束计算模型, 计算病毒侵染前后宿主细胞代谢通量的分布, 从而摆脱了对病毒生物量合成函数依赖的限制。与基于“DELTA”网络模型的分析不同, 本章研究特别强调利用差异表达的上调/下调基因作为关键基因, 以重构代谢网络模型。重构的 15 种病毒性疾病的代谢网络模型规模如图 4.2 所示。这些重构模型平均包含 1114 个基因、2124 个代谢反应及 1726 个代谢物, 规模相较于“DELTA”网络模型有所扩大, 这一变化主要是考虑到了差异下调基因集的影响。尽管如此, 与 Recon3D

模型相比，规模显著缩减，这一优化能够更加精确地关注病毒性疾病引起的异常代谢路径与系统整体变化。

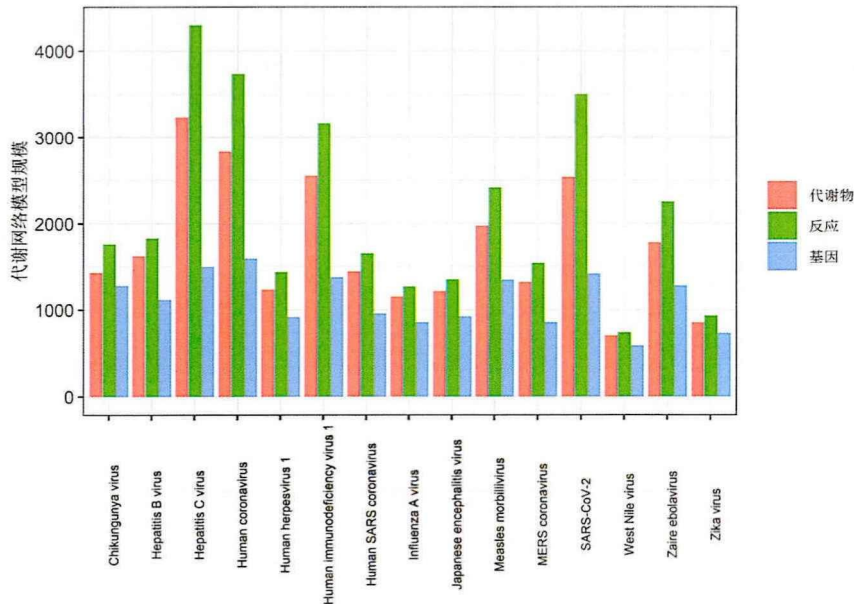


图 4.2 15 种病毒性疾病重构的代谢网络模型规模

通过计算 15 个代谢网络模型进行了管家反应消除率，观察其对组织原发特征的消除情况，相关指标如表 4.1 所示。同样地，将 15 个重构的代谢网络模型与 Kishk^[64]和 Anoop 等人^[76]的病毒性疾病代谢网络模型进行管家反应消除率的比较，结果如图 4.3 所示。可以看出，重构的代谢网络模型关于管家反应的平均消除率为 72.97%，虽然相较“DELTA”网络模型包含了较多的原发特征，但仍然优于已有的病毒性疾病代谢网络模型。

表 4.1 15 个代谢网络模型对管家反应的消除率详情

消除率指标	平均值	方差	最大值	最小值
值	72.97%	2.2×10^{-2}	91.74%	46.21%
消除率指标	上四分位数	中位数	下四分位数	四分位距
值	82.79%	79.77%	63.65%	19.14%

在本研究中，为了探索 15 种病毒性疾病入染宿主细胞后在代谢系统中发生的共性变化，共识别出 16 个在所有代谢网络模型的反应集中共同存在的代谢反应，详细情况列示于表 4.2。这 16 个代谢反应涵盖了 3 个需求交换反应、3 个氧化磷酸化反应、2 个糖酵解及糖异生反应以及 3 个涉及氧气、线粒体等相关的运输反应，还包括了一些与核苷酸转换相关的代谢反应。这些反应凸显了病毒复制

进程中对能量需求及物质转运、转换的关键依赖,体现了病毒入侵宿主细胞后对其代谢系统执行的异常调控。值得注意的是,其中 AMPDA, ATPtm 等 11 个代谢反应也在基于相互作用数据重构的特异性代谢网络模型中被认定为共有反应,这一发现进一步验证了这些代谢反应在病毒感染宿主细胞并进行代谢调控中的关键作用。因此,这些代谢反应在研究病毒性疾病的致病机制上,具有深入探究的科学价值,可能成为理解病毒性疾病共性致病机制和开发新型抗病毒治疗策略的重要基础。

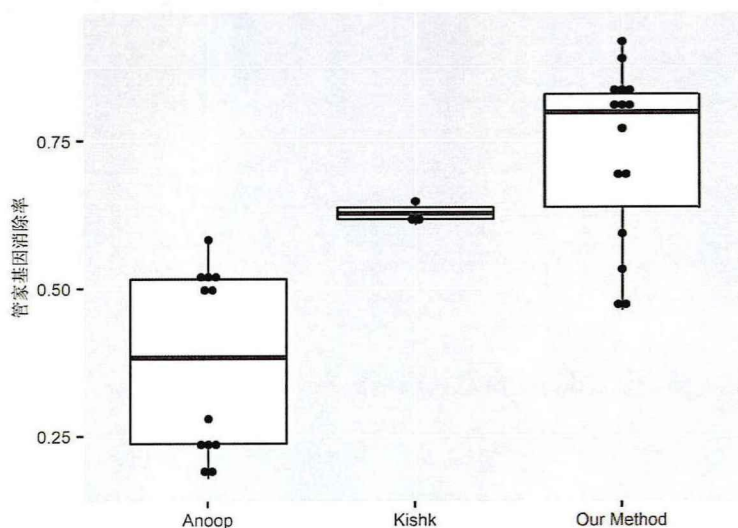


图 4.3 管家反应消除率结果对比图

4.3.2 基于约束计算模型的代谢通量分布

通过构建基因表达水平与代谢通量一致性最高的约束计算模型,成功计算出了病毒侵染前后宿主细胞代谢系统的反应通量及其变化,进行抗病毒靶点的预测。计算所得的代谢反应通量向量,提供了一个直观的窗口,以量化的方法分析病毒入侵宿主细胞后代谢水平的改变。

以所有代谢网络模型中共有的代谢反应为目标,对这些反应在病毒感染状态下的反应通量进行了计算与统计分析。通过这些反应在不同模型中的代谢通量分布(图 4.4),发现了一个有趣的现象:除了 O2tm 反应在若干代谢网络模型中显示出不同的反应通量方向外,大多数共有代谢反应在所有代谢网络模型中均呈现相同的方向性。这一现象指明,病毒感染对宿主细胞代谢系统的异常调控,在很大程度上是沿着相同的方向发展的。特别值得注意的是, ATPtm 和 ATPS4mi 这两个反应在多数代谢网络模型中展现出了显著的通量,这两个反应关乎能量的生成与转运,强调了病毒在宿主细胞内快速复制与增殖对大量能量的需求。

表 4.2 15 个代谢网络模型共有的 16 个代谢反应

反应名称	反应式	反应类型
AMPDA	$\text{h2o}[\text{c}]+\text{h}[\text{c}]+\text{amp}[\text{c}]\rightarrow\text{nh4}[\text{c}]+\text{imp}[\text{c}]$	核苷酸转换
ATPtm	$\text{adp}[\text{c}]+\text{atp}[\text{m}]\rightarrow\text{atp}[\text{c}]+\text{adp}[\text{m}]$	ADP/ATP 转运
EX_utp[e]	$\text{utp}[\text{e}]\rightarrow$	交换需求反应
O2tm	$\text{o2}[\text{c}]\rightarrow\text{o2}[\text{m}]$	氧气运输
PPAm	$\text{h2o}[\text{m}]+\text{ppi}[\text{m}]\rightarrow\text{h}[\text{m}]+\text{pi}[\text{m}]$	氧化磷酸化
EX_HC00250[e]	$\text{HC00250}[\text{e}]\rightarrow$	交换需求反应
r0276	$\text{nh4}[\text{c}]+\text{nadp}[\text{c}]+\text{imp}[\text{c}]\rightarrow\text{h}[\text{c}]+\text{nadph}[\text{c}]+\text{gmp}[\text{c}]$	嘌呤分解代谢
r0892	$\text{utp}[\text{c}]\rightarrow\text{utp}[\text{e}]$	核苷酸转换
r0940	$\text{HC00250}[\text{c}]\rightarrow\text{HC00250}[\text{e}]$	促进扩散
r1156	$\text{uri}[\text{c}]+\text{dutp}[\text{c}]\rightarrow\text{h}[\text{c}]+\text{dudp}[\text{c}]+\text{ump}[\text{c}]$	核苷酸转换
PPItm	$\text{ppi}[\text{c}]\rightarrow\text{ppi}[\text{m}]$	线粒体运输
Ex_uri[e]	$\text{uri}[\text{e}]\rightarrow$	交换需求反应
GK1	$\text{atp}[\text{c}]+\text{gmp}[\text{c}]\rightarrow\text{adp}[\text{c}]+\text{gdp}[\text{c}]$	糖酵解和糖异生
NDPK6	$\text{atp}[\text{c}]+\text{dudp}[\text{c}]\rightarrow\text{adp}[\text{c}]+\text{dutp}[\text{c}]$	糖酵解和糖异生
ATPS4mi	$\text{adp}[\text{m}]+\text{pi}[\text{m}]+\text{h}[\text{i}]\rightarrow\text{h2o}[\text{m}]+\text{h}[\text{m}]+\text{atp}[\text{m}]$	氧化磷酸化
CYOR_u10mi	$\text{h}[\text{m}]+\text{ficytC}[\text{m}]+\text{q10h2}[\text{m}]\rightarrow\text{q10}[\text{m}]+\text{focytC}[\text{m}]+\text{h}[\text{i}]$	氧化磷酸化

4.3.3 抗病毒靶点预测结果

使用基因表达数据,成功重构了 15 种病毒性疾病的代谢网络模型。通过这些模型上执行单基因敲除实验和细胞毒性测试,旨在预测潜在的抗病毒靶点基因。区别于前两章的研究方法,本章并未采用计算 VBOF 通量的方式,预测基因对病毒生物物质生长的潜在影响。基于建立的约束计算模型,计算了感染前后代谢系统的通量变化,旨在通过消除这些通量变化来恢复代谢系统的正常功能。这一方法绕过 VBOF 构建过程中可能存在的数据缺失问题,直接利用基因表达数据快速进行病毒性疾病的靶点预测。此外,采用的新策略对预测结果产生了显著影响,特别是在预测不同病毒性疾病的抗病毒目标时,展现出更高的特异性。与前两章相比,未能识别出类似 CTH 和 DCTD 这类具有广泛抗病毒作用的目标基因。这一现象与 VBOF 的结构组成及之前章节的预测流程紧密相关,后者主要考虑了 VBOF 通量。由于 VBOF 的构建基于基因和蛋白质序列,其主要成分包括核苷酸和氨基酸,多种病毒的 VBOF 结构因此显示出相似性。在代谢分析中,调控核苷酸/氨基酸路径对 VBOF 通量的影响较大,导致了较高的共性。然而,本章仅基于基因表达数据,并以基因表达水平与代谢通量的一致性为目标,分析了

整个代谢系统的通量变化，使得预测结果更加依赖于特定病毒性疾病的基因表达特征，从而表现出更高的特异性。

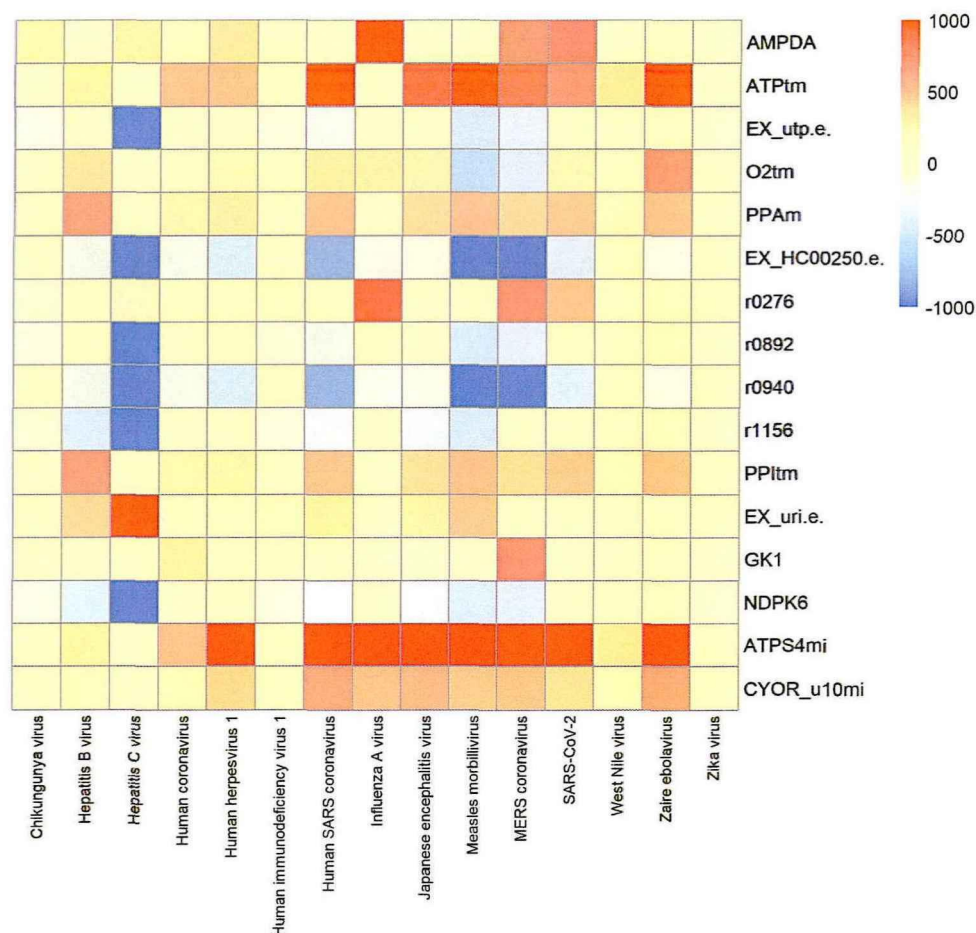


图 4.4 16 个共有代谢反应的通量分布 heatmap

值得强调的是，SLCO1A2 和 ALPL 这两个基因在多个代谢网络模型中被识别为关键靶点。特别是，SLCO1A2 基因在针对 SARS-CoV-2、HIV-1 及 Chikungunya 病毒的特异性模型中被预测为潜在靶点。作为 SLC21A 家族溶质载体成员之一，SLCO1A2 经历选择性剪接，产生多种转录变体。研究人员 Telbisz^[102] 与 Annaert^[103] 等人分别指出，由 SLCO1A2 基因编码的 OATP 蛋白与抑制 SARS-CoV-2 和 HIV-1 的药物间存在强烈相互作用，这可能影响药物的效力。然而，对于这些病毒性疾病的致病机理，目前的研究尚未提供明确的解释。

另一方面，ALPL 基因被预测为多种冠状病毒的潜在抗病毒靶点，四种冠状病毒的靶点预测结果如表 4.3 所示。对 ALPL 与冠状病毒关系的研究揭示，该基因的高表达可能导致肺部组织纤维化，这是冠状病毒感染人体后的主要组织损伤和致病表现。Rajaneesh 等人^[104] 采用了一种创新的基于“基因标签”的药物重

定位策略,通过运用广泛认可的基因排序算法,鉴定出一系列针对 COVID-19 临床试验靶点的抑制剂,包括 ALPL。其中,硫酸锌作为一种相关药物,已被用于 COVID-19 的临床试验中。

表 4.3 四种冠状病毒的靶点基因预测结果

病毒名称	靶点基因
人冠状病毒 (Human coronavirus)	ATP2B1,ALPL
人类 SARS 冠状病毒 (Human SARS coronavirus)	GPI,ALPL
中东呼吸综合征冠状病毒 (MERS coronavirus)	ACOT8,ACOT4,ACOT1,ACOT6, HSD17B10,HADH
新型冠状病毒 (SARS-CoV-2)	SLCO1A2,SLC29A2,ALPL

这些预测结果不仅为病毒性疾病的治疗提供了新的视角,也表明了进一步研究和优化的巨大潜力。通过深入探索这些靶点的作用机制和相应的治疗策略,为抗病毒疗法的发展开辟新的途径。

4.4 本章小结

目前大多数基于代谢网络展开的病毒性疾病研究以及本文的前两章工作,都将抑制病毒生物物质生长定位为通过代谢网络模型预测抗病毒靶点的目标,使得分析过程中 VBOF 是必不可少的。然而,构造 VBOF 时,可能存在病毒蛋白拷贝数和脂质组成数据缺失和不准确的问题,加之缺少动力学特征,对整体抗病毒靶点预测流程造成干扰。

本章利用基因表达数据,结合差异表达分析及代谢网络重构算法,成功重构了 15 种病毒性疾病的代谢网络模型。针对 VBOF 数据缺失的问题,构建了一种以基因表达水平与代谢通量之间的一致性最高为优化目标的线性规划约束计算模型。以计算感染前后代谢网络中的反应通量,代替现有方法计算 VBOF 通量的方式,消除对 VBOF 的依赖。通过调整计算模型中的约束条件模拟基因敲除实验,寻找可以显著缩短感染状态与正常状态下代谢通量距离的基因,结合毒性测试预测出无害的抗病毒靶点基因。与前两章的研究相比,实验预测的抗病毒靶点在不同的病毒中显示出更高的特异性。尽管如此,ALPL 基因在多种冠状病毒中表现出一定的广谱性,已有研究证明了其在冠状病毒感染的致病机制上的作用。该方法通过绕过 VBOF 的构造,直接基于基因表达数据快速预测抗病毒靶点,为病毒性疾病治疗提供了新的策略。

第5章 总结与展望

5.1 工作总结

病毒性疾病历来对人类健康和安全构成了广泛的威胁。鉴于病毒种类繁多,感染人体的组织和方式各异,当前对于病毒性疾病的治疗策略主要集中于支持性治疗,旨在针对病毒的具体结构和感染机理进行干预,而普遍缺乏通用的治疗方案和病毒靶点的预测方法。随着系统生物学的迅速进展,利用计算机模型预测并验证潜在的抗病毒策略已经成为加速抗病毒药物发展和再利用的一种高效手段。特别是,应用基因组规模的代谢网络模型(GEMs)进行病毒感染后代谢系统的分析,已成为代谢组学中预测抗病毒靶点的一种普遍方法。然而,当前关于病毒性疾病代谢网络模型相关的研究都存在一定的局限性。主要表现在以下几个方面。首先,现有研究大多针对特定病毒设计靶点预测方法,缺少一种通用的,适用于所有病毒类型的靶点预测方法。其次,现有研究使用的代谢网络模型会受到组织特异性的干扰,且包含较多的原发组织特征,对靶点预测结果产生干扰。此外,现有研究的靶点发现过程未考虑对正常细胞组织是否具有危害性,使得预测的靶点存在无法应用的风险。

当前关于病毒性疾病的数据存在类型杂,分布零散,涉及的生物学实验复杂等特点,给相关数据获取增大了难度。本文旨在使用两类当前容易获取且广泛应用的数据类型:基因表达数据以及病毒与宿主蛋白相互作用的数据,从病毒感染后对宿主细胞进行代谢重编程的共性出发,设计一种适用于所有病毒性疾病通用靶点预测方法。通过重构病毒感染后发生异常代谢部分的代谢网络模型,克服组织特异性与原发组织特征的影响。将基于代谢网络模型的代谢分析过程建模成一个线性规划计算模型,通过调整模型中的约束条件,模拟单基因敲除和毒性测试生物实验,预测出具有抑制病毒异常代谢且对正常细胞组织无害的靶点。本文完成的主要工作包括:

(1) 设计并实现了一种基于基因表达数据的病毒性疾病通用靶点预测方法。本文使用 GEO 数据库中病毒性疾病的基因表达数据,基于病毒感染细胞与正常细胞的基因表达水平差异,重构了 15 种病毒性疾病的“DELTA”网络模型。这些“DELTA”网络模型表征了病毒感染后与正常细胞相比异常活跃的系统部分,不会受到组织特异性的干扰,且能较好地消除原发组织特征。将在“DELTA”网络模型和正常细胞代谢网络模型中的代谢分析过程建模成线性规划计算模型,通过调整线性规划约束条件模拟单基因敲除和代谢任务测试的生物过程,以计算的方式完成基因必要性测试和细胞毒性测试。通过计算基因敲除后“DELTA”网络模型中 VBOF 的通量,判断基因敲除对病毒生物物质生长的抑制效果。通过计算

基因敲除后线性规划问题是否存在可行解,判断基因敲除对正常细胞是否存在危害。从而预测出具有抗病毒性质且对正常细胞无害的靶点基因。实验结果表明,CTH和DCTD基因在多种病毒性疾病中显示出广泛的抑制作用。针对DCTD的现有药物Cytarabine是一种具有抗病毒和免疫抑制特性的核苷类似物,在治疗其他病毒性疾病中展现出了潜在的再利用价值。已有文献证明这两个基因调控的代谢途径主要集中在氨基酸和核苷酸的合成上,对病毒生物物质合成至关重要,在病毒性疾病致病机制中发挥的作用值得进一步研究。

(2) 设计并实现了一种基于蛋白-蛋白相互作用数据的病毒性疾病通用靶点预测方法。使得在基因表达数据缺失时,可以使用另一种数据形式完成抗病毒靶点预测。利用IntAct数据库提供的病毒蛋白与人类蛋白的相互作用数据,通过提取与病毒强相互作用的代谢反应,重构了11种病毒性疾病的代谢网络模型。与使用基因表达数据进行靶点预测的流程类似,将基于代谢网络模型进行的代谢分析建模成线性规划计算模型,通过调整线性规划约束条件模拟基因敲除和毒性测试的生物学实验,进而预测潜在的抗病毒靶点。值得注意的是,CTH与DCTD被识别为多种病毒性疾病的广谱靶点,与使用基因表达数据的预测结果相一致,进一步印证了预测结果的可靠性。

(3) 设计并实现了一种不依赖VBOF的病毒性疾病的通用靶点预测方法。当前基于GEM展开的病毒相关研究,大多将病毒生物物质生长定位为代谢系统的主要目标,利用通量平衡分析等手段探究基因或代谢途径对病毒生物物质生长的影响。本文前两部分的靶点预测也依赖于计算VBOF通量判断基因敲除对于病毒抑制的效果。然而,当前关于病毒蛋白及脂质相关的定量数据获取困难,以及缺乏相关的动力学特征,使VBOF的构造存在一定的局限性。因此,通过假定基因表达水平和代谢通量之间的一致性最高,建立了一种约束计算模型,计算出病毒感染前后代谢系统中的通量分布。通过调整约束条件模拟基因敲除和毒性测试过程,使得敲除后感染细胞与正常细胞代谢通量距离显著缩小,从而恢复宿主细胞的正常的代谢状态。以缩短感染细胞与正常细胞之间的通量距离,代替降低VBOF通量作为抗病毒性质的评价标准,消除靶点预测过程中对VBOF的依赖。虽然实验结果未发现对所有病毒性疾病均具有抗病毒性质的靶点基因,但ALPL被预测为多种冠状病毒的靶点基因,且已有文献证实ALPL与冠状病毒的致病机制息息相关。

5.2 工作展望

本研究后续工作展望如下:

(1) 搜集更多可用的病毒性疾病相关数据。当前感染人体细胞的病毒种类已

有上百种，本研究中受时间以及数据搜集难度的限制，仅选取了部分热门、影响力较大的病毒种类进行研究。持续搜集更多种类的病毒性疾病数据，进行抗病毒靶点的预测，并将整个靶点预测流程与预测结果集成为一个公共信息平台，是后续研究的重点与目标。

(2) 整合更多关于病毒结构信息相关的数据以提高 VBOF 构造的准确性。当前由于病毒蛋白拷贝数以及脂质相关信息的缺乏，VBOF 构造未考虑病毒包膜的具体构成；当前模型缺乏动力学特性使得病毒进入和释放过程未能得到很好的模拟。后续将持续跟踪相关定量技术的发展及信息获取，从而完善 VBOF 的构造，使靶点预测的结果更加准确。

(3) 采用更多样的生物信息学仿真实验进行抗病毒靶点的预测。由于考虑到实验复杂程度及时间限制，当前对于潜在靶点基因的预测主要采用单基敲除实验。后续研究将尝试设计合理的组合基因敲除策略，探究多基因敲除对代谢网络分析的影响。同时也会在当前的约束模型中结合更多的组学信息和动力学信息，提升结果的准确性与可靠性。

参考文献

- [1] EVANS A S. Viral infections of humans: epidemiology and control[M]. Springer Science & Business Media, 2013.
- [2] TAUBENBERGER J K, MORENS D M. The pathology of influenza virus infections[J]. Annu. Rev. Pathol. Mech. Dis., 2008, 3: 499-522.
- [3] HUI-JUN W, SI-HAO D, XIA Y, et al. Review and prospect of pathological features of corona virus disease.[J]. Journal of Forensic Medicine/Fayixue Zazhi, 2020, 36(1).
- [4] MALAVIGE G, FERNANDO S, FERNANDO D, et al. Dengue viral infections[J]. Postgraduate medical journal, 2004, 80(948): 588-601.
- [5] RAMOS-CASALS M, CUADRADO M J, ALBA P, et al. Acute viral infections in patients with systemic lupus erythematosus: description of 23 cases and review of the literature[J]. Medicine, 2008, 87(6): 311-318.
- [6] ROUSSELET M C, MICHALAK S, DUPRÉ F, et al. Sources of variability in histological scoring of chronic viral hepatitis[J]. Hepatology, 2005, 41(2): 257-264.
- [7] PAREDES R, LALAMA C M, RIBAUDO H J, et al. Pre-existing minority drug-resistant hiv-1 variants, adherence, and risk of antiretroviral treatment failure[J]. The Journal of infectious diseases, 2010, 201(5): 662-671.
- [8] SAJED A N, AMGAIN K. Corona virus disease (covid-19) outbreak and the strategy for prevention[J]. Europasian Journal of Medical Sciences, 2020, 2(1): 1-3.
- [9] DE CLERCQ E. Strategies in the design of antiviral drugs[J]. Nature Reviews drug discovery, 2002, 1(1): 13-25.
- [10] LEMON S M, THOMAS D L. Vaccines to prevent viral hepatitis[J]. New England journal of medicine, 1997, 336(3): 196-204.
- [11] LESK A M. Introduction to genomics[M]. Oxford University Press, 2017.
- [12] PATTERSON S D, AEBERSOLD R H. Proteomics: the first decade and beyond[J]. Nature genetics, 2003, 33(3): 311-323.
- [13] FRANZOSA E A, XIA Y. Structural principles within the human-virus protein-protein interaction network[J]. Proceedings of the National Academy of Sciences, 2011, 108(26): 10538-10543.
- [14] SALAZAR G, ZHANG N, FU T M, et al. Antibody therapies for the prevention and treatment of viral infections[J]. npj Vaccines, 2017, 2(1): 19.
- [15] JIN Y H, CAI L, CHENG Z S, et al. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-ncov) infected pneumonia (standard version)[J]. Military

- medical research, 2020, 7(1): 1-23.
- [16] SHE J, JIANG J, YE L, et al. 2019 novel coronavirus of pneumonia in wuhan, china: emerging attack and management strategies[J]. Clinical and translational medicine, 2020, 9: 1-7.
- [17] SLUIS-CREMER N, TACHEDJIAN G. Mechanisms of inhibition of hiv replication by non-nucleoside reverse transcriptase inhibitors[J]. Virus research, 2008, 134(1-2): 147-156.
- [18] GROFF J L, GROPPER S S. Advanced nutrition and human metabolism.: Ed. 3[M]. 1999.
- [19] WALSH C T, TU B P, TANG Y. Eight kinetically stable but thermodynamically activated molecules that power cell metabolism[J]. Chemical reviews, 2018, 118(4): 1460-1494.
- [20] LUNT S Y, VANDER HEIDEN M G. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation[J]. Annual review of cell and developmental biology, 2011, 27: 441-464.
- [21] KLOSE T, RETENO D G, BENAMAR S, et al. Structure of faustovirus, a large dsdna virus [J]. Proceedings of the National Academy of Sciences, 2016, 113(22): 6206-6211.
- [22] TAYLOR M W, et al. Viruses and man: A history of interactions[R]. Springer, 2014.
- [23] MAYNARD N D, GUTSCHOW M V, BIRCH E W, et al. The virus as metabolic engineer [J]. Biotechnology journal, 2010, 5(7): 686-694.
- [24] ALLEN C N, ARJONA S P, SANTERRE M, et al. Hallmarks of metabolic reprogramming and their role in viral pathogenesis[J]. Viruses, 2022, 14(3): 602.
- [25] PANT A, DSOUZA L, YANG Z. Alteration in cellular signaling and metabolic reprogramming during viral infection[J]. MBio, 2021, 12(5): 10-1128.
- [26] WILLIG A L, OVERTON E T. Metabolic complications and glucose metabolism in hiv infection: a review of the evidence[J]. Current Hiv/aids Reports, 2016, 13: 289-296.
- [27] MORENO-ALTAMIRANO M M B, KOLSTOE S E, SÁNCHEZ-GARCÍA F J. Virus control of cell metabolism for replication and evasion of host immune responses[J]. Frontiers in cellular and infection microbiology, 2019, 9: 95.
- [28] NELEMANS T, KIKKERT M. Viral innate immune evasion and the pathogenesis of emerging rna virus infections[J]. Viruses, 2019, 11(10): 961.
- [29] LI F, DING S W. Virus counterdefense: diverse strategies for evading the rna-silencing immunity[J]. Annu. Rev. Microbiol., 2006, 60: 503-531.
- [30] SUMBRIA D, BERBER E, MATHAYAN M, et al. Virus infections and host metabolism—can we manage the interactions?[J]. Frontiers in immunology, 2021, 11: 594963.
- [31] BRAZMA A, VILO J. Gene expression data analysis[J]. FEBS letters, 2000, 480(1): 17-24.
- [32] DERISI J, PENLAND L, BITTNER M, et al. Use of a cDNA microarray to analyse gene expression[J]. Nat. genet, 1996, 14: 457-460.
- [33] HACIA J G. Resequencing and mutational analysis using oligonucleotide microarrays[J].

- Nature genetics, 1999, 21(1): 42-47.
- [34] ROHLAND N, REICH D. Cost-effective, high-throughput dna sequencing libraries for multiplexed target capture[J]. Genome research, 2012, 22(5): 939-946.
- [35] WANG Z, GERSTEIN M, SNYDER M. Rna-seq: a revolutionary tool for transcriptomics [J]. Nature reviews genetics, 2009, 10(1): 57-63.
- [36] DAAMEN A R, BACHALI P, OWEN K A, et al. Comprehensive transcriptomic analysis of covid-19 blood, lung, and airway[J]. Scientific reports, 2021, 11(1): 7052.
- [37] DE CHASSEY B, NAVRATIL V, TAFFOREAU L, et al. Hepatitis c virus infection protein network[J]. Molecular systems biology, 2008, 4(1): 230.
- [38] KLIMKAIT T, STREBEL K, HOGGAN M D, et al. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release[J]. Journal of virology, 1990, 64(2): 621-629.
- [39] MORADPOUR D, PENIN F. Hepatitis c virus proteins: from structure to function[J]. Hepatitis C virus: from molecular virology to antiviral therapy, 2013: 113-142.
- [40] HAYASHI T, MACDONALD L A, TAKIMOTO T. Influenza a virus protein pa-x contributes to viral growth and suppression of the host antiviral and immune responses[J]. Journal of virology, 2015, 89(12): 6442-6452.
- [41] HE Y, NAKAO H, TAN S L, et al. Subversion of cell signaling pathways by hepatitis c virus nonstructural 5a protein via interaction with grb2 and p85 phosphatidylinositol 3-kinase[J]. Journal of virology, 2002, 76(18): 9207-9217.
- [42] ZHANG X, WANG L, YAN Y. Identification of potential key genes and pathways in hepatitis b virus-associated hepatocellular carcinoma by bioinformatics analyses[J]. Oncology Letters, 2020, 19(5): 3477-3486.
- [43] GORDON D E, JANG G M, BOUHADDOU M, et al. A sars-cov-2 protein interaction map reveals targets for drug repurposing[J]. Nature, 2020, 583(7816): 459-468.
- [44] ALON U. An introduction to systems biology: design principles of biological circuits[M]. Chapman and Hall/CRC, 2019.
- [45] NIELSEN J. Systems biology of lipid metabolism: from yeast to human[J]. FEBS letters, 2009, 583(24): 3905-3913.
- [46] BRUGGEMAN F J, WESTERHOFF H V. The nature of systems biology[J]. TRENDS in Microbiology, 2007, 15(1): 45-50.
- [47] THIELE I, PALSSON B Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction[J]. Nature protocols, 2010, 5(1): 93-121.
- [48] ZHANG C, HUA Q. Applications of genome-scale metabolic models in biotechnology and systems medicine[J]. Frontiers in physiology, 2016, 6: 174546.

- [49] ZHOU Y W, XIE Y, TANG L S, et al. Therapeutic targets and interventional strategies in covid-19: mechanisms and clinical studies[J]. Signal transduction and targeted therapy, 2021, 6(1): 317.
- [50] MCCLELLAN K, PERRY C M. Oseltamivir: a review of its use in influenza[J]. Drugs, 2001, 61: 263-283.
- [51] PAVLOVA N N, THOMPSON C B. The emerging hallmarks of cancer metabolism[J]. Cell metabolism, 2016, 23(1): 27-47.
- [52] HANAHAN D, WEINBERG R A. Hallmarks of cancer: the next generation[J]. cell, 2011, 144(5): 646-674.
- [53] DUARTE N C, BECKER S A, JAMSHIDI N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data[J]. Proceedings of the National Academy of Sciences, 2007, 104(6): 1777-1782.
- [54] MARDINOGLU A, AGREN R, KAMPF C, et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease[J]. Nature communications, 2014, 5(1): 3083.
- [55] THIELE I, SWAINSTON N, FLEMING R M, et al. A community-driven global reconstruction of human metabolism[J]. Nature biotechnology, 2013, 31(5): 419-425.
- [56] BRUNK E, SAHOO S, ZIELINSKI D C, et al. Recon3d enables a three-dimensional view of gene variation in human metabolism[J]. Nature biotechnology, 2018, 36(3): 272-281.
- [57] VLASSIS N, PACHECO M P, SAUTER T. Fast reconstruction of compact context-specific metabolic network models[J]. PLoS computational biology, 2014, 10(1): e1003424.
- [58] ZUR H, RUPPIN E, SHLOMI T. imat: an integrative metabolic analysis tool[J]. Bioinformatics, 2010, 26(24): 3140-3142.
- [59] AGREN R, MARDINOGLU A, ASPLUND A, et al. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling[J]. Molecular systems biology, 2014, 10(3): 721.
- [60] RENZ A, WIDERSPICK L, DRÄGER A. Fba reveals guanylate kinase as a potential target for antiviral therapies against sars-cov-2[J]. Bioinformatics, 2020, 36(Supplement_2): i813-i821.
- [61] RENZ A, WIDERSPICK L, DRÄGER A. Genome-scale metabolic model of infection with sars-cov-2 mutants confirms guanylate kinase as robust potential antiviral target[J]. Genes, 2021, 12(6): 796.
- [62] ORTH J D, THIELE I, PALSSON B Ø. What is flux balance analysis?[J]. Nature biotechnology, 2010, 28(3): 245-248.
- [63] GUDMUNDSSON S, THIELE I. Computationally efficient flux variability analysis[J]. BMC

- bioinformatics, 2010, 11: 1-3.
- [64] KISHK A, PACHECO M P, SAUTER T. Dccov: Repositioning of drugs and drug combinations for sars-cov-2 infected lung through constraint-based modeling[J]. *Iscience*, 2021, 24(11).
- [65] ALLER S, SCOTT A, SARKAR-TYSON M, et al. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against chikungunya, dengue and zika viruses[J]. *Journal of The Royal Society Interface*, 2018, 15(146): 20180125.
- [66] BARRETT T, TROUP D B, WILHITE S E, et al. Ncbi geo: mining tens of millions of expression profiles—database and tools update[J]. *Nucleic acids research*, 2007, 35(suppl_1): D760-D765.
- [67] HERMIAKOB H, MONTECCHI-PALAZZI L, LEWINGTON C, et al. Intact: an open source molecular interaction database[J]. *Nucleic acids research*, 2004, 32(suppl_1): D452-D455.
- [68] RITCHIE M E, PHIPSON B, WU D, et al. limma powers differential expression analyses for rna-sequencing and microarray studies[J]. *Nucleic acids research*, 2015, 43(7): e47-e47.
- [69] SMYTH G. Limma: linear models for microarray data. bioinformatics and computational biology solutions using r and bioconductor. edited by: Gentleman r, carey v, dudoit s, irizarry r, huber w[M]. New York: Springer, 2005.
- [70] HEIRENDT L, ARRECKX S, PFAU T, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0[J]. *Nature protocols*, 2019, 14(3): 639-702.
- [71] ROBINSON J L, KOCABAŞ P, WANG H, et al. An atlas of human metabolism[J]. *Science signaling*, 2020, 13(624): eaaz1482.
- [72] PAWLAK J B, HSU J C C, XIA H, et al. Cmpk2 restricts zika virus replication by inhibiting viral translation[J]. *PLoS pathogens*, 2023, 19(4): e1011286.
- [73] LEE C H, GRIFFITHS S, DIGARD P, et al. Asparagine deprivation causes a reversible inhibition of human cytomegalovirus acute virus replication[J]. *MBio*, 2019, 10(5): 10-1128.
- [74] DENNIS G, SHERMAN B T, HOSACK D A, et al. David: database for annotation, visualization, and integrated discovery[J]. *Genome biology*, 2003, 4: 1-11.
- [75] UHLÉN M, FAGERBERG L, HALLSTRÖM B M, et al. Tissue-based map of the human proteome[J]. *Science*, 2015, 347(6220): 1260419.
- [76] AMBIKAN A T, SVENSSON-AKUSJÄRVI S, KRISHNAN S, et al. Genome-scale metabolic models for natural and long-term drug-induced viral control in hiv infection[J]. *Life Science Alliance*, 2022, 5(9).
- [77] KIMURA-OHBA S, ASAKA M N, UTSUMI D, et al. d-alanine as a biomarker and a therapeutic option for severe influenza virus infection and covid-19[J]. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 2023, 1869(1): 166584.

- [78] GUALLAR J P, CANO-SOLDADO P, AYMERICH I, et al. Altered expression of nucleoside transporter genes (slc28 and slc29) in adipose tissue from hiv-1-infected patients[J]. *Antiviral therapy*, 2007, 12(6): 853-864.
- [79] KANEHISA M. The kegg database[C]// 'In silico' simulation of biological processes: Novartis Foundation Symposium 247: Vol. 247. Wiley Online Library, 2002: 91-103.
- [80] SAYERS E W, CAVANAUGH M, CLARK K, et al. Genbank[J]. *Nucleic acids research*, 2020, 48(D1): D84.
- [81] CAMPBELL F, ARCHER B, LAURENSEN-SCHAFER H, et al. Increased transmissibility and global spread of sars-cov-2 variants of concern as at june 2021[J]. *Eurosurveillance*, 2021, 26(24): 2100509.
- [82] The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2[J]. *Nature microbiology*, 2020, 5(4): 536-544.
- [83] STARK C, BREITKREUTZ B J, REGULY T, et al. Biogrid: a general repository for interaction datasets[J]. *Nucleic acids research*, 2006, 34(suppl_1): D535-D539.
- [84] MOON J H, LIM S, JO K, et al. Pintnet: construction of condition-specific pathway interaction network by computing shortest paths on weighted ppi[J]. *BMC systems biology*, 2017, 11: 1-13.
- [85] MOHAMMADI JENGHARA M, EBRAHIMPOUR-KOMLEH H, PARVIN H. Dynamic protein-protein interaction networks construction using firefly algorithm[J]. *Pattern Analysis and Applications*, 2018, 21: 1067-1081.
- [86] YANG S, FU C, LIAN X, et al. Understanding human-virus protein-protein interactions using a human protein complex-based analysis framework[J]. *MSystems*, 2019, 4(2): 10-1128.
- [87] DE CHASSEY B, MEYNIEL-SCHICKLIN L, VONDERSCHER J, et al. Virus-host interactomics: new insights and opportunities for antiviral drug discovery[J]. *Genome medicine*, 2014, 6: 1-14.
- [88] YANG X, YANG S, LIAN X, et al. Transfer learning via multi-scale convolutional neural layers for human-virus protein-protein interaction prediction[J]. *Bioinformatics*, 2021, 37(24): 4771-4778.
- [89] MCWILLIAM H, LI W, ULUDAG M, et al. Analysis tool web services from the embl-ebi [J]. *Nucleic acids research*, 2013, 41(W1): W597-W600.
- [90] MULLOWNEY M W, DUNCAN K R, ELSAYED S S, et al. Artificial intelligence for natural product drug discovery[J]. *Nature Reviews Drug Discovery*, 2023, 22(11): 895-916.
- [91] RASOOLY D, PELOSO G M, PEREIRA A C, et al. Genome-wide association analysis and mendelian randomization proteomics identify drug targets for heart failure[J]. *Nature Communications*, 2023, 14(1): 3826.

- [92] YE J, MCGINNIS S, MADDEN T L. Blast: improvements for better sequence analysis[J]. *Nucleic acids research*, 2006, 34(suppl_2): W6-W9.
- [93] TAGUWA S, OKAMOTO T, ABE T, et al. Human butyrate-induced transcript 1 interacts with hepatitis c virus ns5a and regulates viral replication[J]. *Journal of virology*, 2008, 82(6): 2631-2641.
- [94] MILEV M P, RAVICHANDRAN M, KHAN M F, et al. Characterization of staufen1 ribonucleoproteins by mass spectrometry and biochemical analyses reveal the presence of diverse host proteins associated with human immunodeficiency virus type 1[J]. *Frontiers in microbiology*, 2012, 3: 33581.
- [95] WISHART D S, FEUNANG Y D, GUO A C, et al. Drugbank 5.0: a major update to the drugbank database for 2018[J]. *Nucleic acids research*, 2018, 46(D1): D1074-D1082.
- [96] KRAUSS A C, GAO X, LI L, et al. Fda approval summary:(daunorubicin and cytarabine) liposome for injection for the treatment of adults with high-risk acute myeloid leukemia[J]. *Clinical Cancer Research*, 2019, 25(9): 2685-2690.
- [97] JORDHEIM L P, DURANTEL D, ZOULIM F, et al. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases[J]. *Nature reviews Drug discovery*, 2013, 12(6): 447-464.
- [98] MALIN J J, SUÁREZ I, PRIESNER V, et al. Remdesivir against covid-19 and other viral diseases[J]. *Clinical microbiology reviews*, 2020, 34(1): 10-1128.
- [99] GU C, KIM G B, KIM W J, et al. Current status and applications of genome-scale metabolic models[J]. *Genome biology*, 2019, 20: 1-18.
- [100] YANESKE E, ZAMPIERI G, BERTOLDI L, et al. Genome-scale metabolic modelling of sars-cov-2 in cancer cells reveals an increased shift to glycolytic energy production[J]. *FEBS letters*, 2021, 595(18): 2350-2365.
- [101] DELATTRE H, SASIDHARAN K, SOYER O S. Inhibiting the reproduction of sars-cov-2 through perturbations in human lung cell metabolic network[J]. *Life science alliance*, 2021, 4(1).
- [102] TELBISZ Á, AMBRUS C, MÓZNER O, et al. Interactions of potential anti-covid-19 compounds with multispecific abc and oatp drug transporters[J]. *Pharmaceutics*, 2021, 13(1): 81.
- [103] ANNAERT P, YE Z, STIEGER B, et al. Interaction of hiv protease inhibitors with oatp1b1, 1b3, and 2b1[J]. *Xenobiotica*, 2010, 40(3): 163-176.
- [104] GUPTA R K, NWACHUKU E L, ZUSMAN B E, et al. Drug repurposing for covid-19 based on an integrative meta-analysis of sars-cov-2 induced gene signature in human airway epithelium[J]. *Plos one*, 2021, 16(9): e0257784.