



孔 琪, 医学博士, 研究员, 硕士生导师。现任中国医学科学院医学实验动物研究所生物信息研究中心负责人、课题组长, 曾任所长助理。兼任中国实验动物学会实验动物标准化技术委员会秘书长、全国实验动物标准化技术委员会前任秘书长暨现任委员。长期从事实验动物大数据、标准化和学科体系建设, 主持或参与国家863项目、科技重大专项、卫生行业专项、北京市自然科学基金面上项目等20余项课题研究, 组织开发12个实验动物与比较医学大数据平台或数据库, 发表科技文章60余篇, 其中SCI论文18篇。同时, 参与策划编写国家统编研究生教材、医学八年制教材、北京协和医学院特色研究生教材、实验动物科学系列丛书等, 作为副主编或编委参编教材或专著23部; 主编或参编国家标准12项, 团体标准20项。获得教育部、中国实验动物学会等科技奖励8项, 以及中国实验动物学会科学技术奖优秀青年人才奖、中国标准化战略联盟标准化专家等荣誉。

冠状病毒感染动物模型组学数据集成分析

吴 玥¹, 李 璐², 张 阳², 王 珏¹, 冯婷婷¹, 李依桐¹, 王 凯², 孔 琪¹

(1. 中国医学科学院医学实验动物研究所, 国家人类疾病动物模型资源库, 国家动物模型技术创新中心, 国家卫生健康委员会比较医学重点实验室, 新发再发传染病动物模型研究北京市重点实验室, 北京市人类重大疾病实验动物模型工程技术研究中心, 北京 100021; 2. 上海宇道生物技术有限公司, 上海 201210)

[摘要] **目的** 分析各个公共数据库中可感染人的冠状病毒感染动物模型的组学数据资源情况, 包括数据分布、数据集数量、数据类型、物种、品系、研究内容等, 从而深入理解冠状病毒的生物学特征和致病机制, 为研究有效的治疗方案和预防措施奠定基础。**方法** 定义特定病毒名称、时间范围和物种等检索策略与纳入排除标准, 检索GEO、ArrayExpress等大型公共组学数据库。根据不同字段类型进行二次过滤, 获取更精确的数据列表。建立组学数据文本库, 进行文献计量学分析, 构建共现网络图, 分析不同研究主题、技术方法和涉及物种之间的关联强度。同时, 分析研究涉及的细胞类型、器官和参与的生物途径, 以进一步阐明病原体与宿主之间的致病相互作用。**结果** 含有冠状病毒组学数据的公共数据库有20余个, 以新型冠状病毒感染组学数据为主。常用物种为人、小鼠、仓鼠和猴, 常用病毒株为Wuhan-Hu-1和USA-WA1/2020。此外, 人类相关研究主要集中于气道上皮细胞和Calu-3细胞, 动物模型(如小鼠、猕猴与雪貂)则多采用肺组织。表达谱数据显示感染后参与炎症、细胞因子反应、补体途径、细胞损伤、增殖和分化等通路基因显著上调。蛋白组学研究揭示, 在不同感染阶段的患者样本中磷酸化蛋白质组、泛素组和全蛋白质组具有显著变化。特定蛋白质类别, 包括病毒受体和蛋白酶、转录因子、细胞因子、凝血系统相关蛋白、血管生成相关蛋白及纤维化标志物等六类蛋白均在冠状病毒感染后发生改变。此外, 代谢组数据提示磷酸胆碱、磷酸乙醇胺、花生四烯酸和油酸可作为潜在的代谢标志物。表观组学研究结果显示, m6A甲基化在新冠病毒复制、感染和传播过程中发挥作用, 并且对宿主细胞-病毒互作产生影响。N、S、非结构蛋白2和3泛素化最为显著。微生物组学研究趋势表明, 肠道和废水中的微生物群落正在成为新的研究重点。**结论** 冠状病毒组学数据类型丰富, 模型与细胞类型多样。根据不同病毒的特征, 造模物种和技术方法的选择具有差异性。研究冠状病毒感染动物模型的多组学数据可以揭示宿主-病原体之间的关键相互作用, 发现生物标志物和潜在的治疗靶点, 为深入理解冠状病毒的生物学特性和感染机制提供丰富信息。

[关键词] 冠状病毒感染; 动物模型; 组学数据; SARS-CoV-2; SARS-CoV; MERS-CoV; 集成分析

[中图分类号] Q95-33; R-332 **[文献标志码]** A **[文章编号]** 1674-5817(2024)04-0357-17



Integrative Analysis of Omics Data in Animal Models of Coronavirus Infection

[基金项目] 北京市自然科学基金资助项目“基于大数据的新型冠状病毒肺炎动物模型数据库建立”(M21027); 国家重点研发计划重点项目“国家实验动物资源库服务科技创新能力提升关键技术与示范”(2021YFF0702800)

[第一作者] 吴 玥(1993—), 女, 硕士, 助理研究员, 研究方向: 比较医学及相关数据库建设。E-mail: wuyue@cnilas.org. ORCID: 0000-0001-5874-8263

[通信作者] 孔 琪(1978—), 男, 博士, 研究员, 研究方向: 比较医学及生物信息学。E-mail: kongqi@cnilas.org. ORCID: 0000-0003-2867-7382

WU Yue¹, LI Lu², ZHANG Yang², WANG Jue¹, FENG Tingting¹, LI Yitong¹, WANG Kai², KONG Qi¹

(1. Institute of Laboratory Animal Science, CAMS & PUMC, National Human Diseases Animal Model Resource Center, National Center of Technology Innovation for Animal Model, NHC Key Laboratory of Comparative Medicine, Beijing Key Laboratory for Animal Models of Emerging and Reemerging Infectious Diseases, Beijing Engineering Research Center for Experimental Animal Models of Human Critical Diseases, Beijing 100021, China; 2. Nutshell Therapeutics Co., Ltd., Shanghai 201210, China)

Correspondence to: KONG Qi (ORCID: 0000-0003-2867-7382), E-mail: kongqi@cnilas.org

[ABSTRACT] Objective This study analyzes the omics data resources in human-infecting coronavirus animal models collected from various public databases, focusing on data distribution, dataset quantity, data types, species, strains, and research content. It aims to enhance our understanding of biological characteristics and pathogenic mechanisms of coronaviruses, thereby providing a solid foundation for devising effective therapeutic strategies and preventive measures. **Methods** Query strategies, including specific virus names, time ranges, and inclusion and exclusion criteria, were defined to retrieve data from major public omics databases such as GEO and ArrayExpress. Secondary filtering was performed based on different field types to obtain a more accurate data list. An omics data text database was established for bibliometric analysis. Co-occurrence networks were constructed for the analysis of the correlation strengths between different research themes, technical methods, and involved species. The cell types, organs, and biological pathways involved in studies were examined to further elucidate the pathogenic interplay between pathogens and hosts. **Results** About twenty public databases containing coronavirus-related omics data were identified, with a primary focus on novel coronavirus infection. Commonly used species include humans, mice, hamsters, and monkeys, while the commonly used virus strains are Wuhan-Hu-1 and USA-WA1/2020. Lung tissues are primarily used in animal models such as mice, macaques, and ferrets, while airway epithelial cells and Calu-3 cells are predominantly employed in human-related studies. Expression profiling data indicate that gene pathways involved in inflammation, cytokine response, complement pathway, cell damage, proliferation, and differentiation are significantly upregulated after infection. Proteomics studies reveal significant changes in phosphoproteome, ubiquitinome, and total proteome of patient samples at different infection stages. Specific protein categories, including viral receptors and proteases, transcription factors, cytokines, proteins associated with coagulation system, angiogenesis-related proteins, and fibrosis markers, show alterations after coronavirus infection. In addition, metabolomics data suggest that phosphocholine, phosphoethanolamine, arachidonic acid, and oleic acid could serve as potential metabolic markers. Epigenomics research indicates m6A methylation plays a role in SARS-CoV-2 replication, infection, and transmission, affecting host cell-virus interactions. Among these, N, S, and non-structural proteins 2 and 3 exhibit the most significant ubiquitination. Trends in microbiomics research suggest that microbial communities in the gut and wastewater are emerging as new research focuses. **Conclusion** The data types of coronavirus omics are diverse, with a wide variety of models and cell types used. The selection of species and technical methods for modelling varies based on the characteristics of different viruses. Multi-omics data from animal models of coronavirus infection can reveal key interactions between hosts and pathogens, identifying biomarkers and potential therapeutic targets, and provide valuable information for a deeper understanding of biological characteristics and infection mechanisms of coronaviruses.

[Key words] Coronavirus infection; Animal models; Omics data; SARS-CoV-2; SARS-CoV; MERS-CoV; Integrative analysis

冠状病毒是广泛存在于自然界中的一个大型病毒家族，具有高度变异性和潜在致病性，能够引发一系列人类疾病，包括常见的感冒和更为严重的传染病，

如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），以及近年来在全球范围

内暴发的新型冠状病毒感染疾病 (corona virus disease 2019, COVID-19) 等。根据系统发育聚类, 冠状病毒可分为 α 冠状病毒 (alpha coronavirus, α CoV)、 β 冠状病毒 (beta coronavirus, β CoV)、 γ 冠状病毒 (gamma coronavirus, γ CoV) 和 δ 冠状病毒 (delta coronavirus, δ CoV) 共4个属^[1-2]。已知可感染人类的冠状病毒包括 SARS、MERS、HCoV-229E (human coronavirus 229E)、HCoV-OC43 (human coronavirus OC43)、HCoV-NL63 (human coronavirus NL63) 和 HCoV-HKU1 (human coronavirus HKU1) 6种^[3]。

严重急性呼吸综合征冠状病毒2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) 又称新型冠状病毒 (简称新冠病毒), 是已知能够感染人类的第7种冠状病毒^[4]。SARS-CoV-2的传染力强、传播速度快, 给人类健康乃至社会、政治和经济都带来了巨大挑战。冠状病毒研究相关的组学数据可以帮助研究人员更好地了解冠状病毒的进化、传播和变异情况, 从而制定有效的预防措施和治疗方案。组学数据集成可以从更为全面的视角探究新冠病毒的生物学特征和感染机制。此外, 研究人员可以使用这些数据来比较不同地区的病毒流行状况, 以及不同变异病毒的传播机制^[5]。

公开常用的大型组学数据库包括美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的SRA (Sequence Read Archive) 和GEO (Gene Expression Omnibus), 以及欧洲生物信息学研究所 (EMBL's European Bioinformatics Institute, EMBL-EBI) 的ArrayExpress和ENA (European Nucleotide Archive), 此外还有OmicsDI (Omics Discovery Index) 和日本DNA数据银行 (DNA Data Bank of Japan, DDBJ) 的GEA (Genomic Expression Archive) 等。这些数据库收录了不同物种的微阵列芯片、二代测序以及其他形式的测序数据^[6-7]。其中, GEO收录的表达谱数据量最多, 其次是ArrayExpress。ArrayExpress包含从GEO中导入的数据, GEA中的数据与ArrayExpress相关联^[8-9]。这些数据库遵循“FAIR”原则 (可发现 findable、可访问 accessible、可互操作 interoperable 和可重用 reusable), 以确保组学数据的广泛可用性^[10]。本研究总结了公共数据库中SARS-CoV-2、SARS-CoV、MERS-CoV、HCoV-229E、HCoV-OC43、HCoV-NL63 和 HCoV-HKU1 等冠状病毒感染相关组学的数据资源情况, 包含数据集、动物模

型物种分布、品系使用情况、测序类型、研究内容等^[11-13]; 然后, 归纳总结各组学研究所获得的结论, 并利用结构化和非结构化数据, 对动物模型组学数据进行集成分析, 从而更全面地了解组学数据的类型、分布情况以及所用的物种、品系, 涉及的细胞类型、器官和技术方法; 此外, 探究冠状病毒在感染过程中的各种生物标志物和信号通路, 全面揭示其感染机制, 并利用知识图谱挖掘组学研究重点, 通过对相关术语进行提取和整理, 探索研究的主要趋势、技术方法、细胞类型、器官和参与途径, 阐明病原体和宿主之间的致病相互作用, 为研究人员提供参考和借鉴, 从而推动冠状病毒相关科研工作的进展。

1 材料与方法

1.1 检索策略

以病毒或疾病名称 (例如“SARS-CoV-2”或“COVID-19”、“SARS-CoV”或“SARS”、“MERS-CoV”或“MERS”、“HCoV-229E”、“HCoV-OC43”、“HCoV-NL63”、“HCoV-HKU1”) 作为关键词, 在GEO (<https://www.ncbi.nlm.nih.gov/gds/>)、ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>)、DDBJ (<https://www.ddbj.nig.ac.jp/index-e.html>)、OmicsDI (<https://www.omicsdi.org/>) 和GISAID (<https://gisaid.org/>) 等数据库中进行检索, 一个关键词包含两个单词时需用双引号, 否则视为两个独立的关键词。输入相关术语的缩写, 并使用逻辑运算符“or”进行连接, 确保广泛覆盖数据集^[14]。使用各数据库提供的过滤器工具 (additional filter), 根据研究类型、组织或细胞来源、数据类型和技术类型等条件进行二次过滤, 进一步缩小检索范围并优化搜索结果。此筛选机制旨在精确锁定与研究最相关的数据集。采用病毒名称、物种名称和组学类型相结合的复合检索式 (例如“SARS-CoV-2” AND “*Mus musculus*” AND “Genome sequencing”, 其中“AND”与“+”等同) 来检索不同的组学数据集。通过管理过滤器 (manage filter)、限制 (limits) 和高级检索 (advanced search) 功能, 对搜索结果进行筛选。本文数据为截至2024年1月的查询统计结果。

1.2 纳入及排除标准

纳入标准: 样本信息清晰, 含有正确分类的物种名称, 数据完整。排除标准: (1) 含有部分检索词, 但研究内容与检索内容不符; (2) 仅含有数据描述而无具体数据; (3) 在冠状病毒流行背景下进行的其他

病毒或细菌的研究。经过纳入及排除标准筛选后,得到与冠状病毒相关的数据集,在OmicsDI、GISAID、ArrayExpress、GEO和DDBJ数据库中分别有16 417 013、99 090、99、1 293和2 217个。

1.3 数据处理

使用Python 3.9.2进行预处理和文本特征提取。利用NCBI API key,结合E-utilities (Entrez Programming Utilities),根据Entrez.esearch格式(db="database", term="query", retmax=20),执行自动化检索。其中,“db”为查询数据库,“term”为查询关键词,“retmax”为设置返回结果条目的最大数量,默认值为20,上限为100 000。在具体操作中,将“database”设为GEO数据库在eInfoResult中的DbName。在“gds”“term”处填入检索词,并在“retmax”处设定期望返回的记录数,以下载GEO数据库中的标题(title)、实验总体设计(overall design)、生物体信息(organism)、数据类型(type)及平台(platform)等信息。通过“print(record)”命令输出全部信息,并将返回的信息(return value)储存为CSV格式。使用Excel 2019软件进行筛选和统计分析,COUNTIF函数用于特定内容的识别与计数。

采用文本共现网络分析与新冠病毒感染相关的基因组学、转录组学、蛋白质组学、代谢组学和微生物组学相关文献之间的关联性。本研究涉及多组学文献52篇,基因组学文献43篇,转录组学文献97篇,蛋白质组学文献49篇,代谢组学文献47篇,微生物组学文献41篇,表观遗传学文献39篇,共计368篇。将所有文献的PDF格式导入Zotero建立文本库。基于标题和摘要,提取SARS-CoV-2组学研究的相关信息,并生成RIS文件。采用二进制计数法量化单个术语的出现频率。采用VOSviewer (v1.6.14)软件进行文献计量学分析,构建共现网络图,并设置最小出现阈值,确保每个术语至少出现2次^[15-16]。VOSviewer不仅可以基于词语的出现频率来识别关键词之间的关联强度,还可通过计算术语之间的共现次数来揭示它们之间的相互关系。通过这种方法,结合手动合并意义相近的同义词,并剔除与研究主题不相关的术语,可避免信息冗余,提高数据的准确性和分析的清晰度。生成共线性网络图的不同颜色基于“VOS”(visualization of similarities)算法进行聚类,这种算法通过相似性度量(共现次数),将节点分组到不同的簇中。相似性度量通常是基于节点之间的连通性和关联强度。例如,如果两个关键词经常一起出现,它们会被归为同一个簇

中。而VOSviewer会自动为网络图中不同的簇分配不同的颜色。颜色通常随机,但不同簇的颜色会有显著差异,以便于区分和可视化。

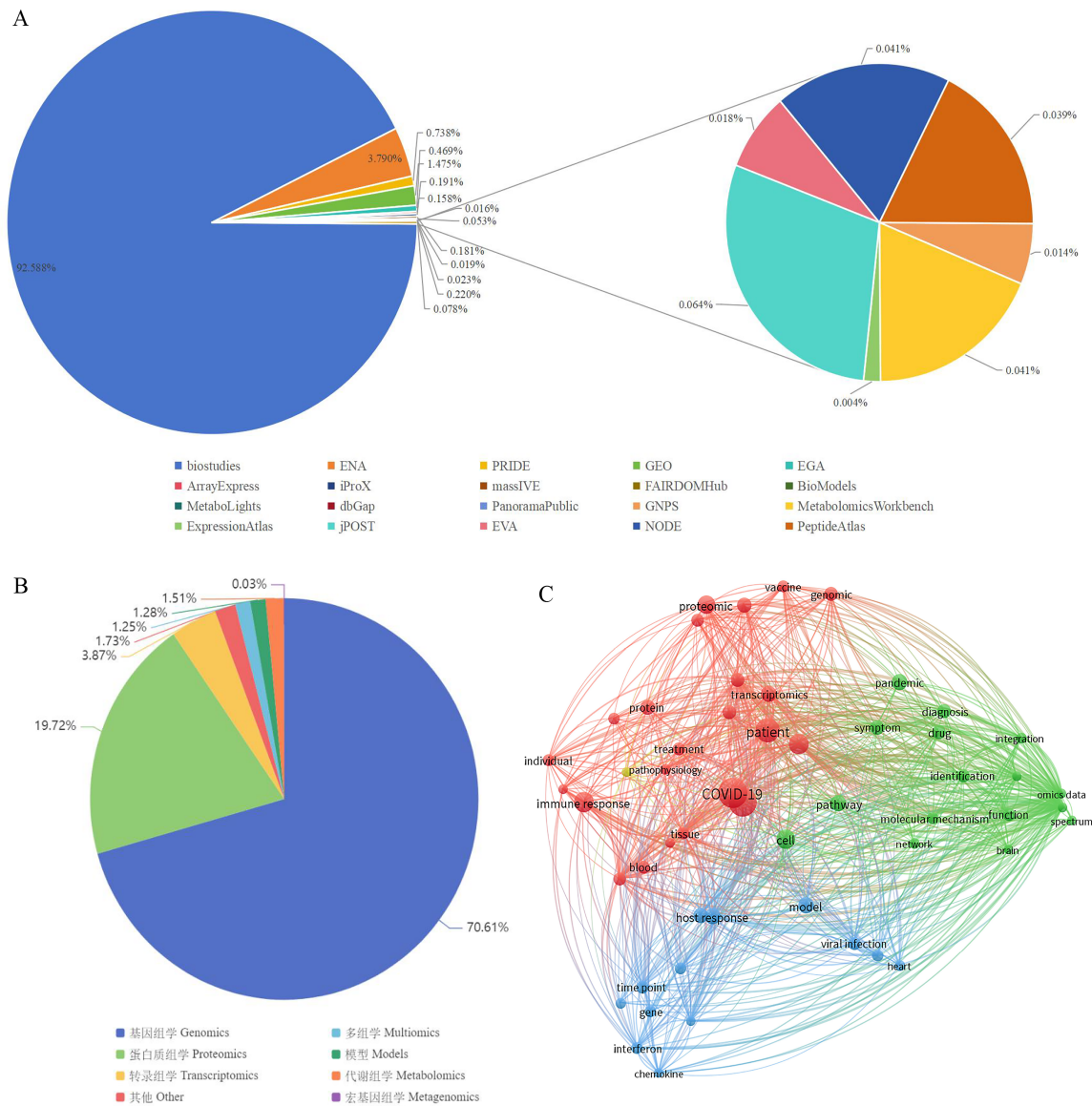
2 结果

2.1 冠状病毒组学数据概况

公共数据库中收录了丰富的冠状病毒组学数据资源,有20余个数据库提供涵盖基因组学、转录组学、蛋白质组学、代谢组学和表观遗传学等多种数据集,其中以新冠病毒数据为主(表1)。在这些数据库中,EBI的Biostudies数据库汇集了大量以新冠病毒感染文献为核心的相关数据集,新冠病毒项目数量突出。其次为ENA数据库、GEO数据库和PRIDE数据库,分别收录了大量新冠病毒序列数据、表达谱数据和蛋白数据(图1A)。OmicsDI对公共数据库中的数据进行了整合^[17],共有99 090个新冠病毒相关项目,主要为基因组学、蛋白质组学和转录组学数据(图1B)。新冠病毒相关项目主要与甲型流感病毒、结核分枝杆菌进行联合研究,建立新冠病毒感染与其他疾病之间的关联,从而有效地指导临床诊断和治疗。

文本共现网络分析结果显示,新冠病毒多组学研究通常涉及多个器官(主要为肺部、脑、心脏和肾脏),并涉及基因组学、转录组学、蛋白质组学和代谢组学的关联性研究。通过联合分析发现免疫调节通路、炎症信号通路和代谢途径在病毒感染和免疫应答中发挥重要作用,并发现特定基因、蛋白质或代谢物与新冠病毒感染严重程度相关。研究发现,病毒感染过程、mRNA剪接、细胞因子和干扰素信号转导以及泛素介导的蛋白质水解是新冠病毒感染的重要途径。基于多组学生物信息学分析的精度高达90%以上,该结果可帮助人们更加全面地了解新冠病毒在人体内的作用机制和影响,从而指导候选药物开发并加速疾病预防工作(图1C)^[18-21]。

ArrayExpress数据库中有60个新冠病毒数据集,这些数据集主要为人类样本,同时也含有小鼠和叙利亚仓鼠(金黄仓鼠)的数据。研究主要在RNA层面,其次为DNA、蛋白和非编码RNA数据(non-coding RNA profiling)(图2A)。该数据库含有29个SARS数据集,主要以人和小鼠为研究对象,同时也包括了部分猕猴、西非绿猴和雪貂的数据,研究重点主要集中在RNA与DNA层面;还有10个MERS数据集,主要涉及物种为人、狨猴、猕猴、绿猴。



注：A，公共数据库中冠状病毒组学数据分布；B，OmicsDI 数据库中新冠病毒数据类型；C，新冠病毒多组学文本共现网络分析。
Note: A, Distribution of coronavirus omics data in public databases; B, Data types of SARS-CoV-2 in OmicsDI; C, Text co-occurrence network analysis of multi-omics related to SARS-CoV-2.

图1 公共数据库中冠状病毒多组学数据统计及文本共现网络分析

Figure 1 Statistics and text co-occurrence network analysis of coronavirus multi-omics data in public databases

NCBI 的 GEO 数据库中，目前共收录 1 293 个冠状病毒相关数据集。其中，新冠病毒相关数据集占绝大多数，数量为 1 212 个（94%）。此外，SARS 数据集为 30 个，MERS 数据集为 33 个，其他类型的冠状病毒数据集为 18 个。新冠病毒数据类型丰富，包括但不限于以下几种：高通量测序表达谱、表达谱芯片、基因组结合/占据谱、非编码 RNA 谱、甲基化谱、蛋白质谱、基因组变异谱和全基因组关联研究。值得注意的是，高通量测序表达谱数据占据了新冠病毒数据集的 72%，

显示出其在新冠病毒研究中的主要地位（图 2B）。

MERS 数据集主要为表达谱芯片、高通量测序表达谱、染色质免疫沉淀测序、甲基化谱、非编码 RNA 谱，其中以表达谱芯片与高通量测序表达谱数据为主（图 2C）。而 SARS 数据集主要为表达谱芯片与非编码 RNA 谱（图 2D）。

在 DDBJ 数据库中，冠状病毒数据以新冠病毒项目为主，共含有 1 981 个项目，其中 MERS 和 SARS 项目分别有 154 和 59 个，其他类型的冠状病毒项目包括

表1 含有冠状病毒组学数据的公共数据库

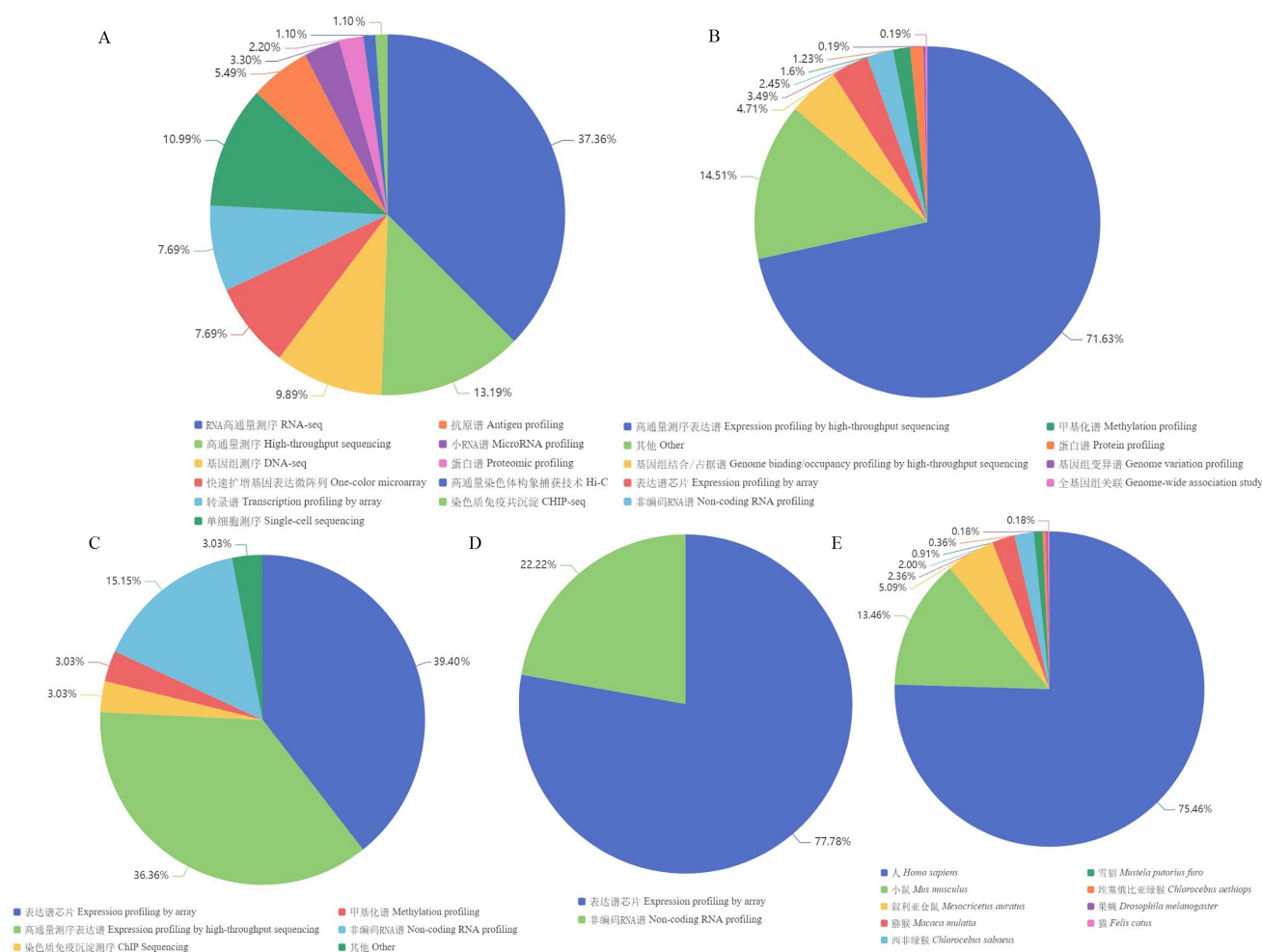
Table 1 Public databases containing coronavirus omics data

数据库名称 Database name	网址 URL	简介 Brief description
Biostudies	https://www.ebi.ac.uk/biostudies/	包含生物学研究数据集, 以及与 EMBL-EBI 数据库或其他公共数据库的数据链接
ENA (European Nucleotide Archive)	https://www.ebi.ac.uk/ena	核酸测序原始数据、序列拼装和功能注释信息
PRIDE (Proteomics Identifications Database)	https://www.ebi.ac.uk/pride	蛋白质组质谱数据存储平台包含蛋白质鉴定、翻译后修饰和光谱数据
GEO (Gene Expression Omnibus)	https://www.ncbi.nlm.nih.gov/geo/	存储各种高通量实验数据的公共数据库, 以表达谱、芯片数据为主
EGA (European Genome-Phenome Archive)	https://ega-archive.org	存储生物医学研究中产生的遗传和表型数据
ArrayExpress	https://www.ebi.ac.uk/arrayexpress/	主要包括微阵列芯片和高通量测序数据, 与 GEO 数据库类似
iProX (Integrated Proteome Resources)	https://www.iprox.org	国家蛋白质科学中心网站, 收录蛋白质组学相关数据集
MassIVE (Mass Spectrometry Interactive Virtual Environment)	https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp	储存蛋白、多肽、质谱数据
FAIRDOMHub	https://fairdomhub.org/	共享多种类型的科学研究数据集、模型或研究过程和成果数据
BioModels	https://www.ebi.ac.uk/biomodels/	储存生物和医学相关的数学模型、生理学和药理学相关机制模型
MetaboLights	https://www.ebi.ac.uk/metabolights/	储存代谢组学和相关衍生信息的数据库, 包含代谢物的结构、浓度、功能数据
dbGap (Database of Genotypes and Phenotypes)	https://dbgap.ncbi.nlm.nih.gov/	储存基因型和表型互作数据
PanoramaPublic	https://panoramaweb.org/	储存蛋白质组学和小分子数据
GNPS (Global Natural Products Social Molecular Networking)	https://gnps.ucsd.edu/	储存原始、处理或注释过的质谱数据
Metabolomics Workbench	https://www.metabolomicsworkbench.org/	代谢组学元数据和实验数据的公共数据库, 包含不同物种的质谱 (MS) 和核磁共振 (NMR) 光谱数据
ExpressionAtlas	https://www.ebi.ac.uk/gxa/home	提供不同条件下基因表达信息, 包含芯片与转录组数据
jPOST (Japan Proteome Standard Repository Database)	https://jpostdb.org/	包含蛋白质组数据及分析结果数据
EVA (European Variation Archive)	https://www.ebi.ac.uk/eva/	包含不同物种遗传变异数据
NODE (National Omics Data Encyclopedia)	https://www.biosino.org/node/	储存多组学数据资源, 包括测序数据、蛋白质组学数据、代谢组学数据及荧光成像数据
PeptideAtlas	https://peptideatlas.org/	收录质谱、蛋白质组学数据。主要为人、小鼠、酵母数据

HCoV-229E (18 个项目)、HCoV-OC43 (2 个项目)、HCoV-NL63 (2 个项目) 和 HCoV-HKU1 (1 个项目) (图 2E)。这些数据涵盖 SRA、GEO 中的部分项目。就物种而言, 数据集主要涉及人、小鼠、叙利亚仓鼠和猕猴。数据类型以基因组学和转录组学数据为主, 同时也包含了部分微生物组学和单细胞测序数据, 而蛋白质组学数据相对较少。

2.2 冠状病毒基因组数据

基因组学测序 (Genomics Sequencing) 是了解冠状病毒感染分子机制、追溯传染源、破译传播途径和检测宿主易感性的基础工具。通过基因组学测序进行流行病学调查, 有助于跟踪调查不同病毒株在不同地区的传播动态, 便于通过更加全面的视角来比较不同国家或地区的疾病流行状况, 有助于发现突变病毒株



注: A, ArrayExpress 数据库中新冠病毒数据类型; B, GEO 数据库中新冠病毒数据类型; C, GEO 数据库中 MERS 数据类型; D, GEO 数据库中 SARS 数据类型; E, DDBJ 数据库中新冠病毒数据的物种分布。

Note: A, Data types of SARS-CoV-2 in ArrayExpress; B, Data types of SARS-CoV-2 in GEO; C, Data types of MERS in GEO; D, Data types of SARS in GEO; E, Species distribution of SARS-CoV-2 data in DDBJ.

图2 ArrayExpress、GEO 及 DDBJ 数据库中冠状病毒组学数据统计及分析

Figure 2 Statistics and analysis of coronavirus omics data in ArrayExpress, GEO, and DDBJ

并探究不同变异病毒株的传播机制, 以及不同物种的感染情况。通过对基因组序列及其特征的深入分析, 如全基因组短串联重复序列和全基因组单核苷酸多态性研究, 可以探究其易感性变异。截至 2024 年 1 月, GISAID 数据库 (<https://gisaid.org/hcov-19-variants-dashboard/>) 中已提交了 16 417 013 个新冠病毒相关序列, 提交数据在 2022 年 2 月前后达到峰值 (图 3A)。ENA 数据库中含有 3 742 908 条新冠病毒序列数据和 5 892 条基因组组装 (Assembly) 数据, 695 条 SARS 序列数据和 4 条 Assembly 数据, 294 条 MERS 序列数据和 2 条 Assembly 数据。而 NCBI 数据库中包括 172 条新冠病毒 Assembly 数据, 6 条 MERS Assembly 数据, 2 条

SARS Assembly 数据, 503 个新冠病毒相关基因, 277 个 SARS 相关基因以及 68 个 MERS 相关基因。DDBJ 数据库中含有 232 个新冠病毒基因组序列和 68 个全基因组项目。样本主要为全血样本、鼻咽拭子和口咽拭子。

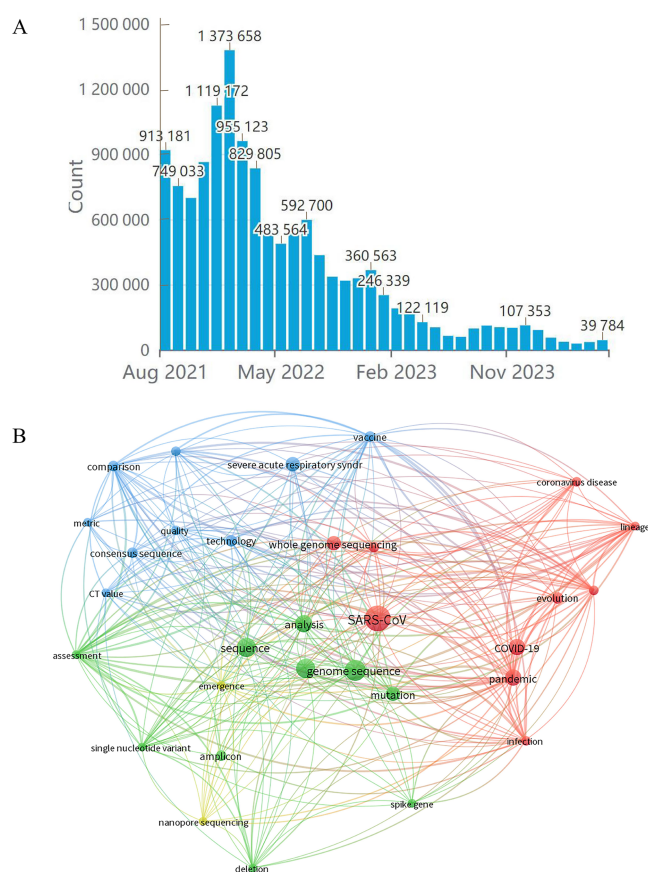
文本共现网络分析显示共识序列 (consensus sequence) 在冠状病毒基因组学研究中发挥突出作用。通过比较不同样本间的基因组序列和共识序列, 可发现新冠病毒的变异情况, 从而推断病毒的演化趋势及传播情况。本研究揭示了冠状病毒强大的跨物种传播能力、广泛的宿主适应性以及周期性暴发的主要驱动因素。这些因素包括病毒基因的重组、高糖基化的同源三聚体 S 蛋白与宿主细胞受体的结合、S1/S2 交界处

的Furin蛋白酶切位点、第一发卡环和开放阅读框8 (open reading frame 8, *ORF8*) 基因等。其中, 重组被认为是最根本的因素。冠状病毒的跳跃式转录和重组具有相同的分子机制, 提示该病毒的暴发具有不可避免性^[22]。此外, RNA 依赖性RNA聚合酶 (RNA-dependent RNA polymerase, *RdRp*) 基因中的14 408 C>T突变可能会影响校对功能或与其他辅因子的结合, 从而影响病毒的突变率。而S蛋白中的23 403 G>A突变可能会增加病毒的传播性^[23]。研究显示了纳米孔测序技术 (nanopore sequencing) 在基因组学研究中的应用, 对于了解病毒的演化趋势、制定有效的防控政策及开发新的治疗手段等具有重要价值 (图3B)。

2.3 冠状病毒表达谱数据

使用Illumina NovaSeq 6000、Illumina NextSeq 500和Illumina HiSeq 2500测序平台进行数据分析, 结果显

示: GEO数据库中含有大量表达谱数据, 其中新冠病毒相关表达谱数据共有1 032个。在这些数据中, 共953个 (92%) 高通量测序表达谱数据; 42个表达谱芯片和37个非编码RNA谱。这些数据涵盖多个物种, 包括人、小鼠、叙利亚仓鼠、猕猴、雪貂、西非绿猴、埃塞俄比亚绿猴、食蟹猴、大鼠、东非狒狒、小毛足鼠、家兔、犬、猫、东北虎、山羊、鹿、蜥蜴、鸽子和穿山甲等 (图4A)。新冠病毒相关表达谱主要为人、小鼠、仓鼠和猴的数据, 其中涉及人类的数据主要为气道上皮细胞、气道基底干细胞、肺泡上皮细胞、肺骨髓细胞、人肺腺癌细胞 (human lung adenocarcinoma cell line, Calu-3)、人胚肾细胞293T (human embryonic kidney 293T cells, 293T)、人结直肠腺癌细胞 (colon adenocarcinoma cells, Caco-2)、血液、肺和胸腔积液样本。绿猴的数据主要来自Vero-E6与COS-7细胞, 以及非洲绿猴的肺组织。小鼠的数据主要包括肺、脑组织、嗅球和嗅觉上皮细胞, 品系主要为年龄在6~8周龄的C57BL/6小鼠、BALB/c小鼠、K18启动子驱动的人类血管紧张素转化酶2 (Keratin 18-human angiotensin-converting enzyme 2, K18-hACE2) 转基因小鼠、表达人源化细胞外基质金属蛋白酶诱导剂基因hCD147的小鼠、MISTRG6人源化小鼠、干扰素诱导跨膜蛋白3 (interferon induced transmembrane protein 3, *IFITM3*) 基因敲除小鼠, 以及载脂蛋白 (apolipoprotein, *APO*) E2、*APOE3*和*APOE4*基因敲入小鼠。感染途径多为鼻道, 性别多为雌性。主要使用的病毒株为Wuhan-Hu-1和USA-WA1/2020。感染检测周期通常在1~7 d (以3 d、6 d居多), 最长可达30 d。主要研究的器官包括肺、肝、肾、心、胃肠道、大脑和呼吸道。在感染新冠病毒后, 人和动物模型的多种基因会上调, 这些基因涉及免疫应答、炎症反应、抗病毒反应以及细胞凋亡等多个方面^[24]。例如, CXC趋化因子配体10 (C-X-C motif chemokine ligand, CXCL10)、C-C趋化因子配体2 (C-C motif chemokine ligand 2, CCL2) 以及参与免疫细胞招募和激活的白细胞介素 (interleukin, IL) -6、IL-8等炎症因子在感染过程中表达上调。1型和3型干扰素 (interferon, IFN) 在受到外界刺激时激活免疫细胞并抑制病毒复制^[25-26]。细胞溶解分子 [例如穿孔素1 (perforin 1, PRF1)] 的表达量增加, 自然杀伤细胞被激活, 促进中性粒细胞等免疫细胞的募集, 从而导致炎症损伤。神经纤毛蛋白1表达升高, 使宿主更易受到新冠病毒感染^[27-28]。

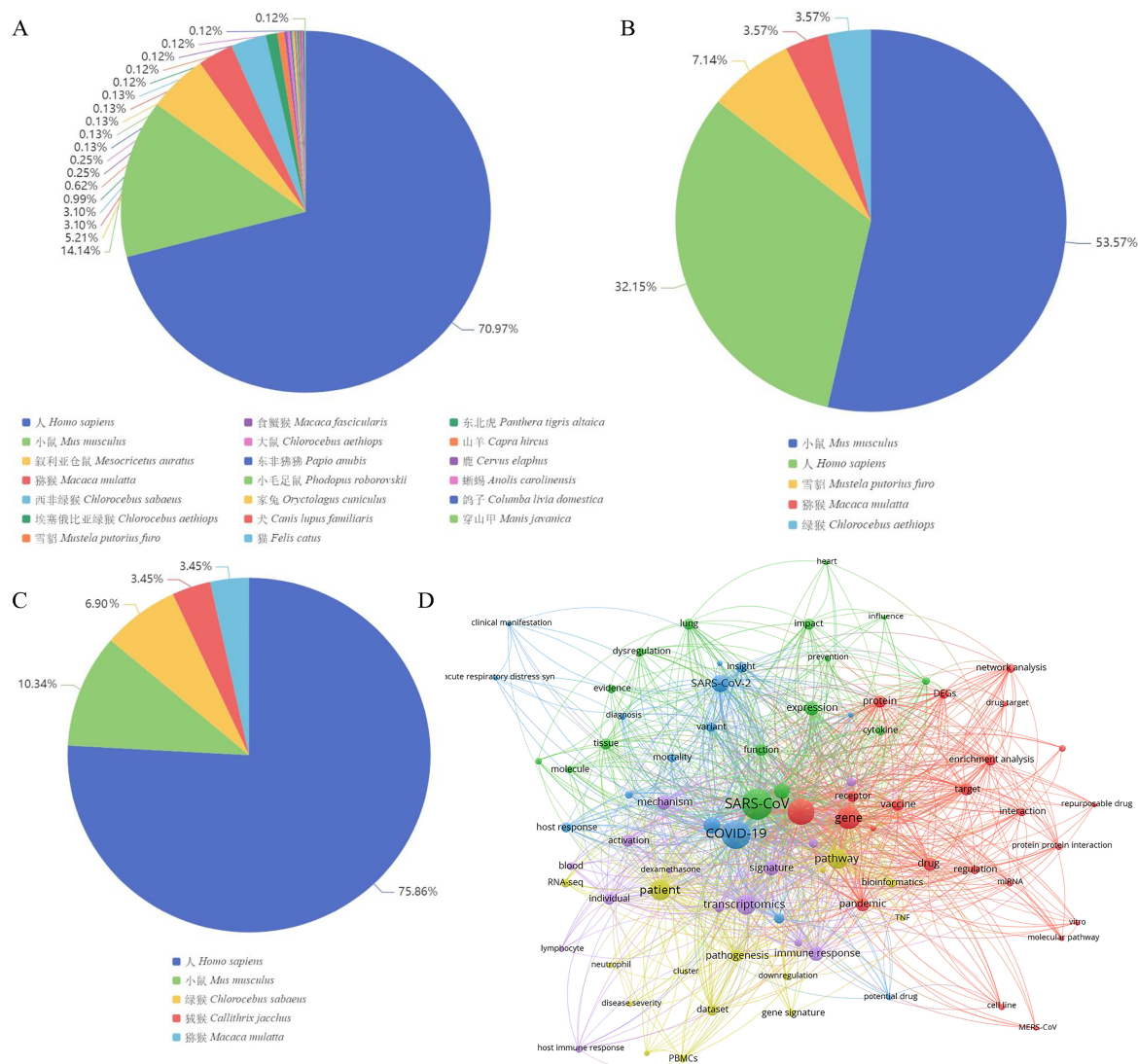


注: A, GISAID 数据库中新冠病毒序列数据变化情况; B, 新冠病毒基因组学文本共现网络分析。

Note: A, Changes of SARS-CoV-2 sequence data in GISAID; B, Text co-occurrence network analysis of SARS-CoV-2 genomics.

图3 新冠病毒基因组学的数据统计及分析

Figure 3 Statistics and analysis of SARS-CoV-2 genomics data



注：A，GEO 数据库中新冠病毒表达谱物种分布；B，GEO 数据库中 SARS 表达谱物种分布；C，GEO 数据库中 MERS 表达谱物种分布；D，新冠病毒转录组学文本共现网络分析。

Note: A, Species distribution of SARS-CoV-2 expression profiles in GEO; B, Species distribution of SARS expression profiles in GEO; C, Species distribution of MERS expression profiles in GEO; D, Text co-occurrence network analysis of SARS-CoV-2 in transcriptomics.

图4 GEO 数据库中冠状病毒表达谱统计及分析

Figure 4 Statistics and analysis of coronavirus expression profiles in GEO

SARS 相关表达谱数据共 30 个，包括 26 个表达谱芯片和 4 个非编码 RNA 谱。这些数据涉及人、小鼠、雪貂、猕猴和绿猴（图 4B）。小鼠主要使用了肺组织，病毒株主要为 MA15，感染途径为鼻道感染，品系包括 BALB/c、C57BL/6 和 129S6/SvEv，敲除基因主要为肿瘤坏死因子受体超家族成员 1B（tumor necrosis factor receptor superfamily member 1B, *TNFRSF1B*）、金属蛋白酶组织抑制剂 1（tissue inhibitor of metalloproteinases 1, *TIMP1*）、Toll 样受体 3（toll-like receptor 3, *TLR3*）、信号转导子和转录激活子 1（signal transducer

and activator of transcription 1, *STAT1*）、干扰素 α 和 β 受体亚基 1（interferon alpha and beta receptor subunit 1, *IFNAR1*）、含三联基序蛋白 55（tripartite motif-containing protein 55, *TRIM55*）和纤溶酶原激活物抑制剂-1（plasminogen activator inhibitor-1, *PAI1*）。小鼠性别以雌性居多，年龄分别为 8 周龄、10 周龄、20 周龄和 1 岁。检测时间通常在感染后 1~28 d，大多数检测集中在感染 2 d 或 4 d。人类、非人灵长类和雪貂使用的病毒株主要为 Urbani（GZ02 或 HC/SZ/6103）和 TOR2。感染时间通常选择 24 h、48 h（2 d）、72 h（3 d）、4 d、

5 d、7 d、11 d、14 d、21 d和28 d, 大多数检测集中在感染24 h、48 h、4 d和7 d。感染途径多为气管和鼻道。猕猴与雪貂的研究中主要使用肺组织, 与人类相关的研究则主要使用气道上皮细胞和Calu-3细胞。绿猴的研究则主要使用Vero-E6细胞。

MERS相关表达谱数据共30个, 包括14个表达谱芯片、11个高通量测序表达谱和5个非编码RNA谱, 涉及的物种有人、小鼠、猕猴、绿猴、猕猴(图4C)。检测时间通常选择感染后3 h、4 h、7 h、12 h、24 h、36 h、48 h、4 d、7 d, 其中选择12 h、24 h和36 h检测的居多。使用的病毒株主要为EMC2012。与人相关的研究主要使用了原代人气道上皮细胞、肺上皮细胞、原代人微血管内皮细胞、原代成纤维细胞、原代人树突状细胞和Calu-3肺腺癌细胞; 小鼠研究中, 主要使用肺细胞; 而对于猕猴与狨猴, 主要使用肺组织; 绿猴的研究则主要为Vero-E6细胞。感染途径多为气管、鼻道、眼部和口腔。

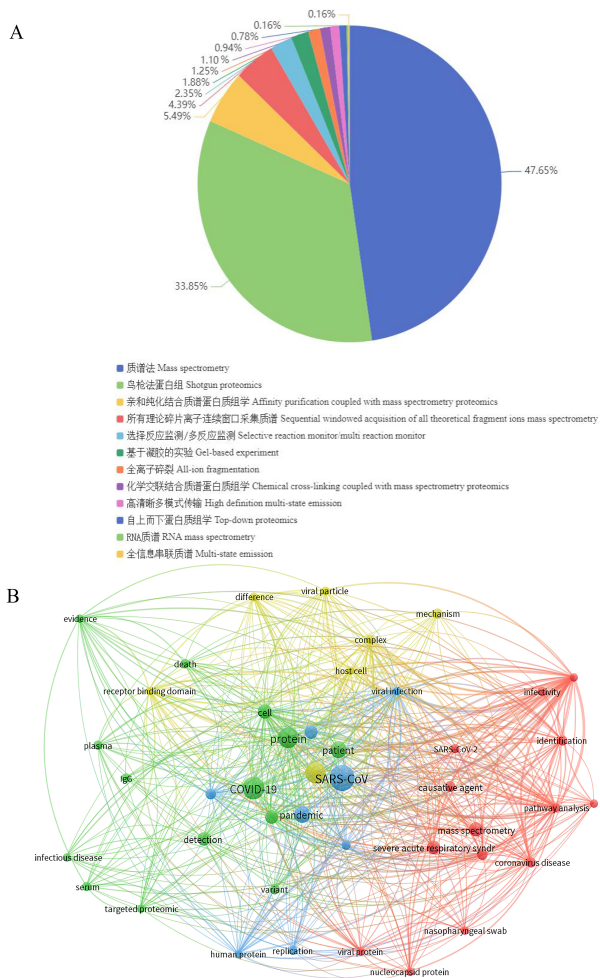
文本共现网络分析显示, 转录组学主要关注新冠病毒感染后宿主细胞的基因表达变化, 通过分析感染后不同时间点的转录组数据, 探究基因的表达模式、宿主应答和免疫应答, 以了解病毒与宿主细胞之间的相互作用。此外, 关于miRNA的研究日益增多, 探究miRNA对病毒基因表达的调控, 深入了解宿主对病毒的应答过程, 有助于了解病毒复制和感染机制, 并为抗病毒药物的设计提供思路(图4D)。通过对重症和轻症新冠病毒感染患者组的差异基因进行基因本体(gene ontology, GO)分析, 发现基因功能主要集中在神经肽、IL-13、趋化因子分泌、T辅助细胞分化和肥大细胞脱颗粒等方面。京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)信号通路分析显示与新冠病毒感染相关的通路, 包括鞘磷脂信号通路、Hippo信号通路和磷脂酶D信号通路, 以及一些与疾病进展相关的通路, 如肥厚型心肌病、催产素信号通路、扩张型心肌病、1型糖尿病和5-羟色胺神经通路等。此外, 还有一些关键临床指标, 如患者入院时的淋巴细胞计数降低、高龄、患基础疾病(如高血压和冠心病)、D-二聚体升高、血红蛋白降低、SOFA评分和APACH评分高等, 都是疾病进展的危险因素^[29]。感染SARS-CoV-2 Delta和Omicron的人、小鼠模型、普通猕猴与食蟹猴模型中, 参与炎症反应、细胞因子反应、补体途径、细胞损伤、增殖和分化等通路基因均上调表达。肺泡二型上皮细胞显著

减少, 而巨噬细胞、成纤维细胞和内皮细胞数量增加^[30]。

2.4 冠状病毒蛋白组数据

PRIDE数据库中含有350个新冠病毒相关蛋白(proteomics)数据集, 包含2个SARS数据集和5个MERS数据集。检测样本多为血清、血浆、肺、口咽和鼻咽拭子, 也有少数尿液和唾液样本。数据以人类为主, 含有少数绿猴、小鼠和猕猴数据。iProX数据库中共有4 219个蛋白数据集, 包含70个新冠病毒数据集和2个MERS数据集, 主要为质谱和鸟枪法蛋白组数据(图5A)。数据集对不同感染时期的患者样本进行磷酸化蛋白质组、泛素组和全蛋白质组分析, 检测与疾病严重程度相关的蛋白质, 观察患者与未感染个体之间显著调节的蛋白和泛素化位点, 以及翻译后修饰情况, 以便找出生物标志物, 特别是参与病毒复制和抑制宿主先天免疫反应的蛋白, 或许可作为预测疾病严重程度早期指标。感染新冠病毒后, 胆固醇代谢和糖酵解相关蛋白受到影响, 建立与病毒相关的蛋白信号网络有助于确定感染后不同阶段相关蛋白质和抗体的变化, 有利于临床诊断和治疗。此外, 数据集中还研究了宿主循环外泌体对新冠病毒感染的反应, 免疫调节机制及其作为疾病严重程度生物标志物的潜在用途。蛋白质组学技术也可应用于病毒检测, 如利用质谱法检测鼻咽黏膜、唾液、痰液和其他体液中的病毒核衣壳N蛋白, 该法在病毒载量低的情况下也可以检测, 有效提高了检测的准确性。

文本共现网络分析显示, 蛋白表达及其功能的研究主要是新冠病毒刺突蛋白和膜蛋白的表达、宿主蛋白的应答反应, 以及免疫应答机制等。其中, 质谱法扮演了重要角色, 帮助了解病毒的复制机制和感染过程, 并为药物设计和疫苗开发提供关键信息(图5B)。新冠病毒感染患者特别是危重患者的支气管肺泡灌洗液中的蛋白质组成发生了显著变化。其中, 病毒受体和蛋白酶、转录因子、细胞因子、凝血系统相关蛋白、血管生成相关蛋白以及纤维化标志物等6类蛋白发生改变。锌指蛋白(zinc finger protein, ZNF)家族基因表达下调, 能够反映感染晚期的严重程度^[31]。干扰素 α 诱导蛋白27(interferon alpha inducible protein 27, IFI27)是血液样本中新冠病毒感染早期高风险的生物标志物^[32]。S100/钙结合蛋白基因如S100A4、S100A6、S100A8和S100A9作为潜在的生物标志物, 能够将白细胞募集到感染区域, 用于区分轻重度新冠



注：A，冠状病毒蛋白质组学数据类型；B，新冠病毒蛋白质组学文本共现网络分析。

Note: A, Types of coronavirus proteomics data; B, Text co-occurrence network analysis of SARS-CoV-2 proteomics.

图5 冠状病毒蛋白质组学数据统计及文本共现网络分析
Figure 5 Statistics and text co-occurrence network analysis of coronavirus proteomics data

病毒感染^[33]。急性炎症蛋白C反应蛋白（C-reactive protein, CRP）和清道夫受体富含半胱氨酸型1蛋白M130（CD163）在感染后上调幅度最大^[34]。此外，上调的差异蛋白与“铁载体-铁进入细胞”、“细胞黏附”和“应激反应”等多个生物学过程密切相关。同时，“宿主-防御蛋白”、“IL信号”和“活性氧解毒作用”等通路也上调。而烟酰胺磷酸核糖转移酶（nicotinamide phosphoribosyl transferase, NAMPT）、核受体亚家族3C组成员1（nuclear receptor subfamily 3, group C, member 1, NR3C1）和干扰素 γ 受体1（interferon gamma receptor 1, IFNGR1）在大多数器官中下调^[35-36]。

2.5 冠状病毒代谢组数据

含有代谢组学（metabolomics）数据的公共数据库中以新冠病毒数据为主，Metabolomics Workbench数据库中含有13个新冠病毒感染相关代谢数据集，OmicsDI中含有47个新冠病毒相关代谢数据集，主要采用液相色谱-质谱法（liquid chromatography-mass spectrometry, LC-MS）、气相色谱-质谱法（gas chromatography-mass spectrometry, GC-MS）和核磁共振（nuclear magnetic resonance, NMR）波谱法。物种主要为人和小鼠。样本类型主要为血浆和血清，细胞主要为Calu-3细胞、气道上皮细胞、微血管内皮细胞和成纤维细胞。

研究主要是检测与疾病进展直接相关的代谢产物变化情况，发现三酰甘油、磷脂酰胆碱、前列腺素E2以及精氨酸水平升高，甜菜碱和腺苷水平降低^[37]；甘油磷脂、鞘脂和脂肪酸下调，同时胆红素和胆汁酸的表达异常，提示可能与肝脏受损相关^[38]；能量代谢途径中非氧化的糖和聚糖代谢在上呼吸道中至关重要^[39]。预测疾病严重程度的生物标志物并分析鉴定出预后标志物，结果显示标志物包括葡萄糖、葡糖醛酸、胆红素降解产物、谷氨酸、精氨酸、鸟氨酸、谷氨酰胺等，参与氮代谢、TCA循环、脂肪酸代谢、氨基酸代谢和胆汁酸代谢，基于代谢特征可以进行预后和治疗评估^[40]。磷酸胆碱、磷酸乙醇胺、花生四烯酸和油酸可作为潜在的代谢标志物。神经酸、亚油酸、 α -亚麻酸、反式维生素酸和棕榈油酸水平可能与新冠病毒感染的不良预后相关。磺胆酸、牛磺脱氧胆酸、甘脱氧胆酸、糖胆酸和熊脱氧胆酸的表达水平随着疾病的严重程度而增加，轻症患者在出院时即可恢复到正常值^[41]。犬尿氨酸、精氨酸、肌氨酸和溶血磷脂酰胆碱是鉴定新冠病毒阳性患者的最佳代谢产物^[42]。

文本共现网络分析显示，代谢组学研究中代谢产物的检测和分析，特别是能量代谢、蛋白质合成和脂肪酸代谢等方面有明显变化。通过与光谱学和脂质组学数据的联动分析，可以更全面地了解生物系统的代谢状态，从而将系统生物学的研究推向更高水平。研究发现，新冠病毒感染轻症、重症和致死队列中的差异代谢物主要集中在嘧啶代谢、葡萄糖/甘露糖代谢和碳代谢。血浆代谢物如苹果酸、天冬氨酸和鸟苷酸的减少会加速疾病的发展进程。此外，乙酸盐、苯丙氨酸在新冠病毒感染过程中的变化受到重点研究（图6A）^[43]。

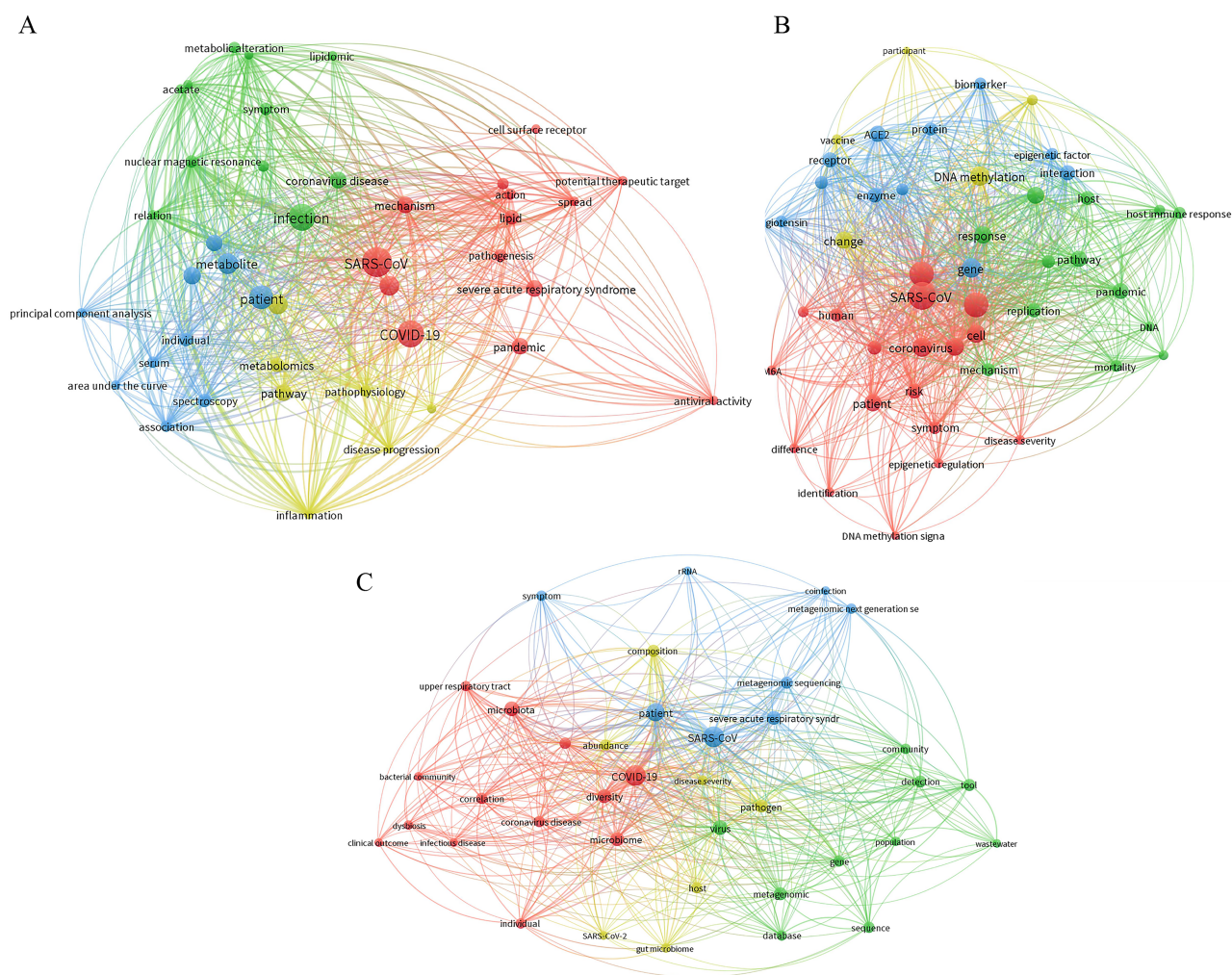


图6 新冠病毒代谢组学(A)、表观组学(B)、微生物组学(C)的文本共现网络分析

Figure 6 Text co-occurrence network analysis of SARS-CoV-2 metabolomics (A), epigenomics (B), and microbiomics (C)

2.6 冠状病毒表观组学

含有表观遗传学测序 (epigenetics sequencing) 的公共数据库中以新冠病毒甲基化数据为主, GEO 数据库中含有 17 个新冠病毒甲基化谱数据。DDBJ 数据库中含有 11 个新冠病毒甲基化谱数据和 1 个 MERS 甲基化数据。物种主要为人、小鼠以及非洲绿猴, 样本主要为外周血和肾脏细胞。主要研究新冠病毒感染后的甲基化组变化、甲基化位点动力学和甲基化模式的免疫学特征。通过 RNA 免疫沉淀测序可检测 SARS-COV-2 RNA 中的 N6-甲基腺苷修饰 (m6A)。免疫细胞类型组成随着年龄的变化而变化, 分析免疫细胞的表观遗传学可以了解疾病的严重程度。表观遗传学参与控制免疫反应和病毒活性, 发现甲基化调节位点, 可为疾病的特异性治疗提供新靶点。

文本共现网络分析显示, 表观遗传学对于 DNA 甲

基化、组蛋白和染色质修饰, 以及非编码 RNA 转录的研究可探究调节、复制和感染过程。感染新冠病毒后宿主细胞内 m6A 的浓度增加, 表明 m6A 甲基化对新冠病毒复制、感染和传播以及对宿主细胞-病毒互作产生影响^[44]。编码免疫调节基因和 ACE2 的多种启动子中差异甲基化 CpG 位点可能是主要的表观遗传特征, CpG 甲基化通常与转录抑制有关, 表明 ACE2 甲基化和翻译后组蛋白修饰可能是导致宿主组织、生物学年龄和性别对病毒感染模式差异的原因 (图 6B)^[45]。活化白细胞黏附分子 (activated leukocyte cell adhesion molecule, ALCAM)、醛脱氢酶 3 家族成员 B1 (aldehyde dehydrogenase 3 family member B1, ALDH3B1)、连环蛋白 $\alpha 1$ (catenin alpha 1, CTNNA1)、内皮分化相关因子 1 (endothelial differentiation-related factor 1, EDF1) 和溶质载体家族 12 成员 2 (solute carrier

family 12 member 2, SLC12A2) 等在感染后泛素化水平下降。基质蛋白 (matrix, M)、核衣壳 (nucleocapsid, N)、刺突 (spike, S)、非结构蛋白 3 和开放读码框 9b (ORF9b) 蛋白发生磷酸化, N、S、非结构蛋白 2 和 3 的泛素化最为显著^[46-47]。

2.7 冠状病毒微生物组学数据

含有微生物组学 (microbiomes) 的公共数据库中以新冠病毒感染数据为主, 其中 SRA 中含有 85 个新冠病毒微生物组学相关项目 (BioProject), 以环境样本为主。EBI 的 BioStudies 数据库中含有 10 个新冠病毒微生物组学相关数据集, DDBJ 中含有 16 个新冠病毒微生物组学相关项目, 主要为宏基因组测序 (metagenome sequencing)。与新冠病毒感染相关的微生物组学测序类型主要为 16S RNA 扩增子测序和宏基因组测序。物种主要为人与小鼠。样本收集部位主要为鼻、口、咽、肺和肠道, 也有环境介质如废水。样本类型主要为粪便、唾液和精液等。主要研究健康个体与新冠病毒感染患者之间的肠道微生物群差异、微生物组成、微生物多样性、微生物相互作用及功能。无偏见的宏基因组学覆盖范围广, 对于快速识别新出现的病原体并准确诊断病原体和混合感染具有无可比拟的优势。微生物组的多样性和组成与疾病的不同病理状况有关, 研究新冠病毒与微生物组的相互作用和微生物组的动态变化, 可以分析微生物群的类别及其相对丰度。其中, 韦荣氏球菌属被确定为新冠肺炎患者组中最突出的生物标志物, 其相对丰度的变化可以为区分新冠病毒感染患者与其他人群提供重要参考^[48]。重症感染组的微生物群特征是肠球菌和乳酸菌增加、粪杆菌和拟杆菌减少, 念珠菌占主导地位^[49]。此外, 克雷伯菌、痤疮杆菌和沙雷菌的比例增加与疾病的严重程度以及全身炎症标志物 (如中性粒细胞-淋巴细胞比率) 的升高相关, 丙酸粪杆菌 (一种抗炎细菌) 的丰度与疾病的严重程度呈负相关^[50-51]。功能注释结果显示, 生化途径主要富集在氨基酸和能量代谢、ABC 转运蛋白、膜转运、毒力、疾病防御、毒力调节、程序性细胞死亡和原发性免疫缺陷相关等方面。研究微生物组和微生物多样性相关特征有助于更好地了解呼吸道感染与肠道微生物组的相关性, 从而开发基于微生物的疾病诊断和治疗方法^[52-53]。

文本共现网络分析显示, 对健康人群和新冠病毒感染患者的人体微生物组学研究主要集中在微生物群落结构和功能方面, 且越来越多侧重于肠道微生物的

研究。废水中的微生物群落结构变化也成为研究重点之一, 研究废水处理过程中微生物群落结构变化有助于了解微生物群落对病毒的抵抗能力及其相互作用机制, 可为优化处理工艺和提高处理效率提供科学依据 (图 6C)。

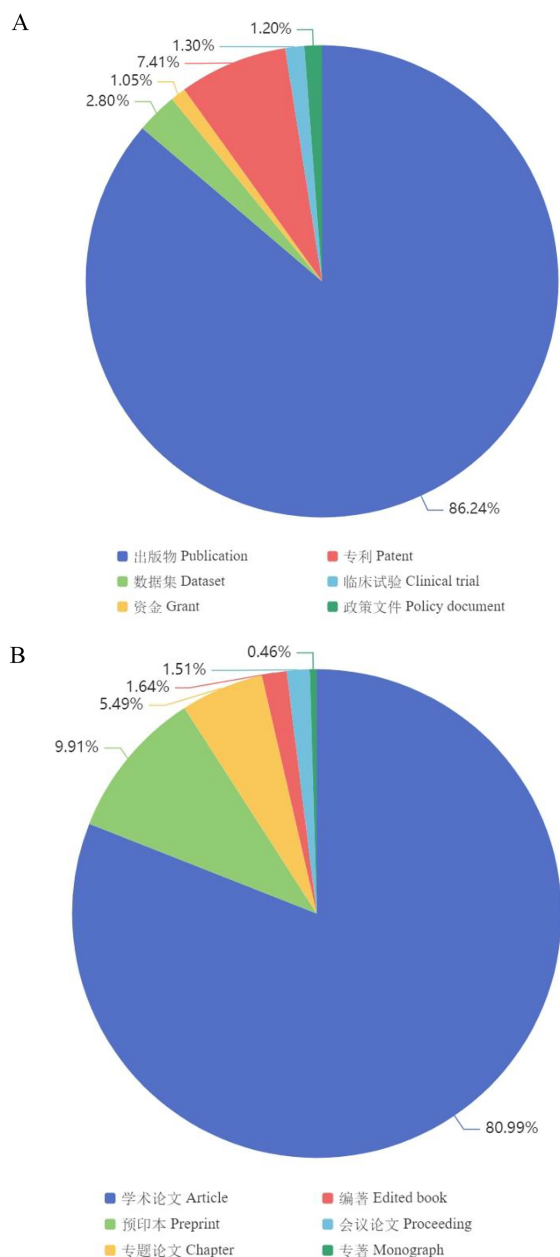
2.8 利用冠状病毒组学数据发表文章分析

PubMed 中关于新冠病毒组学数据发表的文章最多, 其次为 SARS、MERS 和 HCoV-HKU1。Dimension 数据库 (<https://app.dimensions.ai/discover/publication>) 中与冠状病毒相关的数据以文献为主, 包括 675 933 篇出版物、58 055 项专利、21 955 个数据集、10 230 个临床试验、9 380 个相关政策文件和 8 224 个研究基金项目 (图 7A)。其中, 出版物以学术论文为主 (图 7B)。在 X-MOL 数据库 (<https://www.x-mol.com/>) 中, 使用搜索表达式 (“SARS-CoV-2” OR “COVID-19” OR “新冠”) AND (“组学” OR “omics”) 在文章的标题和摘要中进行检索, 得到 4 664 篇新冠病毒组学相关文章。

3 讨论

COVID-19 曾全球肆虐, 尽管近期全国疫情防控形势整体平稳, 但新冠病毒似乎不会从地球上消失, 且突变株奥密克戎 (Omicron) XBB 系列非常容易导致二次感染。基于新冠病毒的研究主要涉及以下几个方面: (1) 病例分析, 即使用病例报告分析新冠病毒的传播特征、潜在危险因素和治疗方法等; (2) 病毒基因组分析, 即研究新冠病毒基因组的结构、变异特征以及病毒的进化趋势; (3) 病毒表面抗原分析; (4) 病毒抗体分析; (5) 病毒毒力分析; (6) 病毒携带者分析, 尤其是新冠病毒携带者的特征以及携带者的传播趋势; (7) 病毒抗药性分析; (8) 病毒抗体检测分析; (9) 病毒抗体及其治疗分析^[54-55]。由于生物安全的限制, 针对冠状病毒感染的研究中存在难以快速收集大量临床样本的问题, 尤其是潜在传染性样本, 因此数据的集成和共享显得尤为重要。

组学研究是一种新兴的研究方法, 其特点是多维度、多层次和多种类型。组学数据的采集和分析可以加速信息的收集、分析和预测, 进而更好地进行可视化^[56-57]。通过利用基因组学、转录组学、表观遗传学、蛋白质组学和代谢组学等多种组学技术, 结合生物信息学、统计学等方法, 可以识别和表征病毒、突变株与宿主之间的相互作用。研究新冠病毒的遗传学、



注：A，Dimension 数据库中冠状病毒相关数据类型；B，Dimension 数据库中冠状病毒相关出版物类型。

Note: A, Types of coronavirus-related data in Dimension; B, Types of coronavirus-related publications in Dimension.

图7 冠状病毒相关数据及文献分布

Figure 7 Distribution of coronavirus-related data and literature

生物学、病理学和免疫学等特性及其对人类健康的影响，可以帮助人们更好地理解该病毒的传播机制、病原体变异、抗药性以及对人体的免疫反应，为病毒的预防、控制和治疗提供重要的科学依据^[58]。目前，基于多组学的研究已广泛应用于新冠病原体、诊疗、发

病机制、预后评估和相关药物靶点挖掘等方面^[59]。其中，基因表达谱分析可以比较不同疾病组织中的基因表达水平，发现与疾病相关的基因表达模式，从而挖掘差异表达基因。基于基因组关联分析，可以比较不同疾病组织样本的基因组变异，从而探究与疾病相关的基因组变异模式。同样，转录组学、蛋白质组学和代谢组学也可以从不同水平比较疾病组织样本情况，从而发现与疾病相关的变异模式^[6]。

在利用组学数据资源方面，应加大多组学数据的整合，加强数据质量控制与标准化，在统筹兼顾的前提下，做到数据的完备化、规范化、标准化和计算化，并结合人工智能和机器学习进行识别与预测，从而获得更为全面的分析结果。充分发挥“互联网+动物模型”的助推作用，可以实现模型数字化，提升可视化水平，开发针对组学数据的专用算法和工具，提高数据分析的效率和准确性，同时注重算法的可解释性可使研究者能够更好地理解和信任分析结果。组学数据的利用、集成和共享有以下作用：（1）更好地支持研究，使研究者能够更快地获得更多的数据，从而全方位多角度地理解和探索科学问题；（2）提高数据分析的准确性；（3）提高研究的可重复性，使研究者能够更容易地重复和验证研究结果；（4）提高研究的可比性，使研究者能够更容易地比较不同研究的结果；（5）将多种类型的组学数据进行整合能够揭示生物体内复杂的网络关系^[60-61]。目前已有的多组学数据和综合分析回顾了不同组学技术在疾病研究中的使用进展，探究病毒传播背后的分子机制，探讨了多组学研究的局限性和未来方向。但对于组学数据库中的数据组成、物种占比和样本类型等详细信息却很少涉及，组学数据的集成分析有助于研究者在研究前选择合适的物种、性别、年龄和感染时间等参数^[54-62]。

中国医学科学院医学实验动物研究所已建立了基于疾病的不同组学数据库（即冠状病毒感染动物模型比较转录组学数据库、心脏病动物模型比较转录组学数据库、流感病毒感染动物模型比较转录组学数据库）以及基于动物模型的实验动物品系数据库、比较医学大数据平台、国家动物模型资源共享信息平台 and 动物模型药筛数据库^[63-69]，通过数据集成及开发比较分析工具，为用户提供方便、快捷、形象化的有效数据，供科研人员进一步了解与研究，推动了由以真实记录动物模型特征为主的研究方法向以多媒体计算机为中心的软件化的特征识别与预测为主的研究方法的转变，

对于国内外科研人员充分利用现有组学数据进行人类疾病研究与防治及科学性研究具有积极意义。

[作者贡献 Author Contribution]

吴玥构建文章主体框架,撰写初稿并修改;
李璐进行文本共现网络分析与echarts作图;
张阳编写Python脚本;
王珏参与数据整理分析;
冯婷婷进行数据调研;
李依桐查找及筛选相关文献;
王凯进行Python脚本调试;
孔琪负责审核、修改文稿。

[利益声明 Declaration of Interest]

所有作者均声明本文不存在利益冲突。

[参考文献 References]

- [1] KHALIFA S A M, MOHAMED B S, ELASHAL M H, et al. Comprehensive overview on multiple strategies fighting COVID-19[J]. *Int J Environ Res Public Health*, 2020, 17(16): 5813. DOI: 10.3390/ijerph17165813.
- [2] SHEHATA M M, GOMAA M R, ALI M A, et al. Middle East respiratory syndrome coronavirus: a comprehensive review [J]. *Front Med*, 2016, 10(2): 120-136. DOI: 10.1007/s11684-016-0430-6.
- [3] RANA R, TRIPATHI A, KUMAR N, et al. A comprehensive overview on COVID-19: future perspectives[J]. *Front Cell Infect Microbiol*, 2021, 11: 744903. DOI: 10.3389/fcimb.2021.744903.
- [4] CHAN J F, KOK K H, ZHU Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan [J]. *Emerg Microbes Infect*, 2020, 9(1): 221-236. DOI: 10.1080/22221751.2020.1719902.
- [5] SHI W F, LI J, ZHOU H, et al. Pathogen genomic surveillance elucidates the origins, transmission and evolution of emerging viral agents in China[J]. *Sci China Life Sci*, 2017, 60(12):1317-1330. DOI: 10.1007/s11427-017-9211-0.
- [6] DAS J K, TRADIGO G, VELTRI P, et al. Data science in unveiling COVID-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing[J]. *Brief Bioinform*, 2021, 22(2):855-872. DOI: 10.1093/bib/bbaa420.
- [7] SAYERS E W, BECK J, BOLTON E E, et al. Database resources of the National Center for Biotechnology Information[J]. *Nucleic Acids Res*, 2024, 52(D1): D33-D43. DOI: 10.1093/nar/gkad1044.
- [8] O'SULLIVAN C, BUSBY B, MIZRACHI I K. Managing sequence data[J]. *Methods Mol Biol*, 2017, 1525:79-106. DOI: 10.1007/978-1-4939-6622-6_4.
- [9] SHUMWAY M, COCHRANE G, SUGAWARA H. Archiving next generation sequencing data[J]. *Nucleic Acids Res*, 2010, 38(Database issue): D870-D871. DOI: 10.1093/nar/gkp1078.
- [10] WILKINSON M D, DUMONTIER M, JAN AALBERSBERG I, et al. Addendum: the FAIR Guiding Principles for scientific data management and stewardship[J]. *Sci Data*, 2019, 6(1):6. DOI: 10.1038/s41597-019-0009-6.
- [11] MONTALDO C, MESSINA F, ABBATE I, et al. Multi-omics approach to COVID-19: a domain-based literature review[J]. *J Transl Med*, 2021, 19(1):501. DOI: 10.1186/s12967-021-03168-8.
- [12] VENKATAKRISHNAN A J, PURANIK A, ANAND A, et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors[J]. *Elife*, 2020, 9: e58040. DOI: 10.7554/eLife.58040.
- [13] KOLENC Ž, PIRIH N, GRETIC P, et al. Top trends in multiomics research: evaluation of 52 published studies and new ways of thinking terminology and visual displays[J]. *OMICS*, 2021, 25(11):681-692. DOI: 10.1089/omi.2021.0160.
- [14] 周大琼, 曹继华, 任力锋, 等. 基因芯片数据库 GEO 与 ArrayExpress 的使用及比较分析[J]. *中国现代医学杂志*, 2014, 24(12):38-42. DOI:10.3969/j.issn.1005-8982.2014.12.009.
- [15] ZHOU D Q, CAO J H, REN L F, et al. Usage and comparative analysis of GEO and ArrayExpress microarray database[J]. *China J Mod Med*, 2014, 24(12):38-42. DOI:10.3969/j.issn.1005-8982.2014.12.009.
- [16] MURILLO J, VILLEGAS L M, ULLOA-MURILLO L M, et al. Recent trends on omics and bioinformatics approaches to study SARS-CoV-2: a bibliometric analysis and mini-review[J]. *Comput Biol Med*, 2021, 128: 104162. DOI: 10.1016/j.combiomed.2020.104162.
- [17] VAN ECK N J, WALTMAN L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. *Scientometrics*, 2010, 84(2):523-538. DOI: 10.1007/s11192-009-0146-3.
- [18] DASS G, VU M T, XU P, et al. The omics discovery REST interface[J]. *Nucleic Acids Res*, 2020, 48(W1): W380-W384. DOI: 10.1093/nar/gkaa326.
- [19] BARH D, TIWARI S, WEENER M E, et al. Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19[J]. *Comput Biol Med*, 2020, 126: 104051. DOI: 10.1016/j.combiomed.2020.104051.
- [20] GUO M Y, XIONG M Y, PENG J Y, et al. Multi-omics for COVID-19: driving development of therapeutics and vaccines[J]. *Natl Sci Rev*, 2023, 10(9): nwad161. DOI: 10.1093/nsr/nwad161.
- [21] DRAKE K A, TALANTOV D, TONG G J, et al. Multi-omic profiling reveals early immunological indicators for identifying COVID-19 Progressors[J]. *Clin Immunol*, 2023, 256: 109808. DOI: 10.1016/j.clim.2023.109808.
- [22] LI C X, GAO J, ZHANG Z C, et al. Multiomics integration-based molecular characterizations of COVID-19[J]. *Brief Bioinform*, 2022, 23(1): bbab485. DOI: 10.1093/bib/bbab485.
- [23] SARAVANAN K A, PANIGRAHI M, KUMAR H, et al. Role of genomics in combating COVID-19 pandemic[J]. *Gene*, 2022, 823:146387. DOI: 10.1016/j.gene.2022.146387.
- [24] ROUCHKA E C, CHARIKER J H, CHUNG D. Variant analysis of 1, 040 SARS-CoV-2 genomes[J]. *PLoS One*, 2020, 15(11): e0241535. DOI: 10.1371/journal.pone.0241535.
- [25] GAO C C, LI M, DENG W, et al. Differential transcriptomic

- landscapes of multiple organs from SARS-CoV-2 early infected rhesus macaques[J]. *Protein Cell*, 2022, 13(12):920-939. DOI: 10.1007/s13238-022-00915-5.
- [25] RUAN Q R, YANG K, WANG W X, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China[J]. *Intensive Care Med*, 2020, 46(5):846-848. DOI: 10.1007/s00134-020-05991-x.
- [26] KIM J A, KIM S H, SEO J S, et al. Temporal transcriptome analysis of SARS-CoV-2-infected lung and spleen in human ACE2-transgenic mice[J]. *Mol Cells*, 2022, 45(12):896-910. DOI: 10.14348/molcells.2022.0089.
- [27] DALY J L, SIMONETTI B, KLEIN K, et al. Neuropilin-1 is a host factor for SARS-CoV-2 infection[J]. *Science*, 2020, 370(6518):861-865. DOI: 10.1126/science.abd3072.
- [28] O'DONNELL K L, PINSKI A N, CLANCY C S, et al. Pathogenic and transcriptomic differences of emerging SARS-CoV-2 variants in the Syrian golden hamster model[J]. *EBioMedicine*, 2021, 73: 103675. DOI: 10.1016/j.ebiom. 2021. 103675.
- [29] DESAI N, NEYAZ A, SZABOLCS A, et al. Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection[J]. *Nat Commun*, 2020, 11(1): 6319. DOI: 10.1038/s41467-020-20139-7.
- [30] OH T, KIM G, BAEK S H, et al. Comparative spatial transcriptomic profiling of severe acute respiratory syndrome coronavirus 2 Delta and Omicron variants infections in the lungs of cynomolgus macaques[J]. *J Med Virol*, 2023, 95(6): e28847. DOI: 10.1002/jmv.28847.
- [31] ESPOSITO S, D'ABROSCA G, ANTOLAK A, et al. Host and viral zinc-finger proteins in COVID-19[J]. *Int J Mol Sci*, 2022, 23(7):3711. DOI: 10.3390/ijms23073711.
- [32] GUPTA R K, ROSENHEIM J, BELL L C, et al. Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-control diagnostic accuracy study[J]. *Lancet Microbe*, 2021, 2(10): e508-e517. DOI: 10.1016/S2666-5247(21) 00146-4.
- [33] ZIEGLER C G K, MIAO V N, OWINGS A H, et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19[J]. *Cell*, 2021, 184(18): 4713-4733. e22. DOI: 10.1016/j. cell.2021.07.023.
- [34] NIE X, QIAN L J, SUN R, et al. Multi-organ proteomic landscape of COVID-19 autopsies[J]. *Cell*, 2021, 184(3):775-791. e14. DOI: 10.1016/j.cell.2021.01.004.
- [35] MESSNER C B, DEMICHEV V, WENDISCH D, et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection[J]. *Cell Syst*, 2020, 11(1): 11-24. e4. DOI: 10.1016/j. cels.2020.05.012.
- [36] BOJKOVA D, KLANN K, KOCH B, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets[J]. *Nature*, 2020, 583(7816):469-472. DOI: 10.1038/s41586-020-2332-7.
- [37] XU J J, ZHOU M, LUO P, et al. Plasma metabolomic profiling of patients recovered from coronavirus disease 2019 (COVID-19) with pulmonary sequelae 3 months after discharge[J]. *Clin Infect Dis*, 2021, 73(12):2228-2239. DOI: 10.1093/cid/ciab147.
- [38] SHEN B, YI X, SUN Y T, et al. Proteomic and metabolomic characterization of COVID-19 patient sera[J]. *Cell*, 2020, 182(1): 59-72.e15. DOI: 10.1016/j.cell.2020.05.032.
- [39] KARPE A V, NGUYEN T V, SHAH R M, et al. A time-series metabolomic analysis of SARS-CoV-2 infection in a ferret model[J]. *Metabolites*, 2022, 12(11): 1151. DOI: 10.3390/metabo12111151.
- [40] JIA H L, LIU C W, LI D T, et al. Metabolomic analyses reveal new stage-specific features of COVID-19[J]. *Eur Respir J*, 2022, 59(2):2100284. DOI: 10.1183/13993003.00284-2021.
- [41] VALDÉS A, MORENO L O, RELLO S R, et al. Metabolomics study of COVID-19 patients in four different clinical stages[J]. *Sci Rep*, 2022, 12(1):1650. DOI: 10.1038/s41598-022-05667-0.
- [42] THOMAS T, STEFANONI D, REISZ J A, et al. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status[J]. *JCI Insight*, 2020, 5(14): e140327. DOI: 10.1172/jci.insight.140327.
- [43] MA J L, DENG Y W, ZHANG M H, et al. The role of multi-omics in the diagnosis of COVID-19 and the prediction of new therapeutic targets[J]. *Virulence*, 2022, 13(1): 1101-1110. DOI: 10.1080/21505594.2022.2092941.
- [44] KGATLE M M, LAWAL I O, MASHABELA G, et al. COVID-19 is a multi-organ aggressor: epigenetic and clinical marks[J]. *Front Immunol*, 2021, 12: 752380. DOI: 10.3389/fimmu. 2021.752380.
- [45] BALNIS J, MADRID A, HOGAN K J, et al. Persistent blood DNA methylation changes one year after SARS-CoV-2 infection[J]. *Clin Epigenetics*, 2022, 14(1): 94. DOI: 10.1186/s13148-022-01313-8.
- [46] STUKALOV A, GIRAULT V, GRASS V, et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV[J]. *Nature*, 2021, 594(7862): 246-252. DOI: 10.1038/s41586-021-03493-4.
- [47] PHILLIPS S, MISHRA T, KHADKA S, et al. Epitranscriptomic N⁶-methyladenosine profile of SARS-CoV-2-infected human lung epithelial cells[J]. *Microbiol Spectr*, 2023, 11(1): e0394322. DOI: 10.1128/spectrum.03943-22.
- [48] MA S L, ZHANG F, ZHOU F X, et al. Metagenomic analysis reveals oropharyngeal microbiota alterations in patients with COVID-19[J]. *Signal Transduct Target Ther*, 2021, 6(1):191. DOI: 10.1038/s41392-021-00614-3.
- [49] MAEDA Y, MOTOOKA D, KAWASAKI T, et al. Longitudinal alterations of the gut mycobiota and microbiota on COVID-19 severity[J]. *BMC Infect Dis*, 2022, 22(1):572. DOI: 10.1186/s12879-022-07358-7.
- [50] ZUO T, ZHANG F, LUI G C Y, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization[J]. *Gastroenterology*, 2020, 159(3):944-955.e8. DOI: 10.1053/j.gastro.2020.05.048.
- [51] LIU Q, MAK J W Y, SU Q, et al. Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome[J]. *Gut*, 2022, 71(3): 544-552. DOI: 10.1136/gutjnl-2021-325989.

- [52] NIES L D, GALATA V, MARTIN-GALLAUSIAUX C, et al. Altered infective competence of the human gut microbiome in COVID-19[J]. *Microbiome*, 2023, 11(1): 46. DOI: 10.1186/s40168-023-01472-7.
- [53] XU X G, ZHANG W, GUO M Q, et al. Integrated analysis of gut microbiome and host immune responses in COVID-19[J]. *Front Med*, 2022, 16(2): 263-275. DOI: 10.1007/s11684-022-0921-6.
- [54] PANG Z Q, ZHOU G Y, CHONG J, et al. Comprehensive meta-analysis of COVID-19 global metabolomics datasets[J]. *Metabolites*, 2021, 11(1):44. DOI: 10.3390/metabo11010044.
- [55] LU T, WANG Y R, GUO T N. Multi-omics in COVID-19: seeing the unseen but overlooked in the clinic[J]. *Cell Rep Med*, 2022, 3(3):100580. DOI: 10.1016/j.xcrm.2022.100580.
- [56] CEN X P, WANG F G, HUANG X H, et al. Towards precision medicine: Omics approach for COVID-19[J]. *Biosaf Health*, 2023, 5(2):78-88. DOI: 10.1016/j.bsheal.2023.01.002.
- [57] MIRALLES A, BRUY T, WOLCOTT K, et al. Repositories for taxonomic data: where we are and what is missing[J]. *Syst Biol*, 2020, 69(6):1231-1253. DOI: 10.1093/sysbio/syaa026.
- [58] HAAS P, MURALIDHARAN M, KROGAN N J, et al. Proteomic approaches to study SARS-CoV-2 biology and COVID-19 pathology[J]. *J Proteome Res*, 2021, 20(2): 1133-1152. DOI: 10.1021/acs.jproteome.0c00764.
- [59] YANG J J, YAN Y Z, ZHONG W. Application of omics technology to combat the COVID-19 pandemic[J]. *MedComm*, 2021, 2(3):381-401. DOI: 10.1002/mco2.90.
- [60] RAY S, SRIVASTAVA S. COVID-19 pandemic: hopes from proteomics and multiomics research[J]. *OMICS*, 2020, 24(8): 457-459. DOI: 10.1089/omi.2020.0073.
- [61] ZHU Z J, ZHANG S N, WANG P, et al. A comprehensive review of the analysis and integration of omics data for SARS-CoV-2 and COVID-19[J]. *Brief Bioinform*, 2022, 23(1): bbab446. DOI: 10.1093/bib/bbab446.
- [62] DONG X J, LIU C Y, DOZMOROV M. Review of multi-omics data resources and integrative analysis for human brain disorders[J]. *Brief Funct Genomics*, 2021, 20(4): 223-234. DOI: 10.1093/bfpg/elab024.
- [63] 吴玥, 向志光, 高苒, 等. 冠状病毒感染动物模型比较转录组学数据库的建立[J]. *中国实验动物学报*, 2022, 30(1):92-99. DOI: 10.3969/j.issn.1005-4847.2022.01.012.
- WU Y, XIANG Z G, GAO R, et al. Establishment of a comparative transcriptomics database of coronavirus infected animal models[J]. *Acta Lab Anim Sci Sin*, 2022, 30(1): 92-99. DOI: 10.3969/j.issn.1005-4847.2022.01.012.
- [64] 吴玥, 王珏, 薛婧, 等. 心脏病动物模型比较转录组学数据库的构建[J]. *中国比较医学杂志*, 2023, 33(3):75-81. DOI: 10.3969/j.issn.1671-7856.2023.03.010.
- WU Y, WANG J, XUE J, et al. Establishment of a comparative transcriptomics database of heart disease animal models[J]. *Chin J Comp Med*, 2023, 33(3):75-81. DOI: 10.3969/j.issn.1671-7856.2023.03.010.
- [65] 孔琪, 夏震宇, 秦川. 实验动物品系数据库的建立[J]. *中国比较医学杂志*, 2015, 25(4):78-83. DOI: 10.3969/j.issn.1671.7856.2015.004.016.
- KONG Q, XIA X Y, QIN C. Establishment of the laboratory animal strain database[J]. *Chin J Comp Med*, 2015, 25(4):78-83. DOI: 10.3969/j.issn.1671.7856.2015.004.016.
- [66] 吴玥, 魏强, 张连峰, 等. 比较医学大数据平台的建立[J]. *中国比较医学杂志*, 2022, 32(12):57-65. DOI: 10.3969/j.issn.1671-7856.2022.12.008.
- WU Y, WEI Q, ZHANG L F, et al. Establishment the comparative medicine big-data platform[J]. *Chin J Comp Med*, 2022, 32(12): 57-65. DOI: 10.3969/j.issn.1671-7856.2022.12.008.
- [67] 吴玥, 薛婧, 魏强, 等. 国家动物模型资源共享信息平台的建立[J]. *中国实验动物学报*, 2022, 30(8):1080-1086. DOI: 10.3969/j.issn.1005-4847.2022.08.009.
- WU Y, XUE J, WEI Q, et al. Establishment of national infrastructure for an animal model resource platform[J]. *Acta Lab Anim Sci Sin*, 2022, 30(8):1080-1086. DOI: 10.3969/j.issn.1005-4847.2022.08.009.
- [68] 吴玥, 王珏, 冯婷婷, 等. 基于动物模型的药物筛选数据库的建立[J]. *中国实验动物学报*, 2023, 31(6):778-786. DOI: 10.3969/j.issn.1005-4847.2023.06.010.
- WU Y, WANG J, FENG T T, et al. Construction of a drug screening database based on animal models[J]. *Acta Lab Anim Sci Sin*, 2023, 31(6): 778-786. DOI: 10.3969/j.issn.1005-4847.2023.06.010.
- [69] WU Y, WANG J, XUE J, et al. Flu-CED: A comparative transcriptomics database of influenza virus-infected human and animal models[J]. *Animal Model Exp Med*, 2024, 2: 2-12. DOI: 10.1002/ame2.12384.

(收稿日期: 2024-01-16 修回日期: 2024-04-10)

(本文编辑: 翟玉凤, 张俊彦, 富群华, 丁宇菁)

[引用本文]

吴玥, 李璐, 张阳, 等. 冠状病毒感染动物模型组学数据集成分析[J]. *实验动物与比较医学*, 2024, 44(4): 357-373. DOI: 10.12300/j.issn.1674-5817.2024.008.

WU Y, LI L, ZHANG Y, et al. Integrative analysis of omics data in animal models of coronavirus infection[J]. *Lab Anim Comp Med*, 2024, 44(4): 357-373. DOI: 10.12300/j.issn.1674-5817.2024.008.