



Multomics integration-based molecular characterizations of COVID-19

Chuan-Xing Li [†], Jing Gao [†], Zicheng Zhang, Lu Chen, Xun Li, Meng Zhou and Åsa M. Wheelock

Corresponding authors: Xun Li, the First Hospital of Lanzhou University, Lanzhou, China. Tel/Fax: +86 931 8619797; E-mail: lxdx21@126.com; Meng Zhou, School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, China. Tel/Fax: +86 577 88068272; E-mail: zhoumeng@wmu.edu.cn; Åsa M. Wheelock, Respiratory Medicine Unit, Department of Medicine & Centre for Molecular Medicine, Karolinska Institutet, and Department of Respiratory Medicine and Allergy, Karolinska University Hospital, Stockholm, Sweden. Tel: +46-(0)70-2200308; E-mail: asa.wheelock@ki.se.

[†]These authors contributed equally to this work.

Abstract

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), rapidly became a global health challenge, leading to unprecedented social and economic consequences. The mechanisms behind the pathogenesis of SARS-CoV-2 are both unique and complex. Omics-scale studies are emerging rapidly and offer a tremendous potential to unravel the puzzle of SARS-CoV-2 pathobiology, as well as moving forward with diagnostics, potential drug targets, risk stratification, therapeutic responses, vaccine development and therapeutic innovation. This review summarizes various aspects of understanding multiomics integration-based molecular characterizations of COVID-19, which to date include the integration of transcriptomics, proteomics, genomics, lipidomics, immunomics and metabolomics to explore virus targets and developing suitable therapeutic solutions through systems biology tools. Furthermore, this review also covers an abridgment of omics investigations related to disease pathogenesis and virulence, the role of host genetic variation and a broad array of immune and inflammatory phenotypes contributing to understanding COVID-19 traits. Insights into this review, which combines existing strategies and multiomics integration profiling, may help further advance our knowledge of COVID-19.

Keywords: COVID-19, multiomics integration, molecular characteristics, severity, outcome, single-cell omics

Introduction

Coronavirus disease 2019 (COVID-19) was first detected in China in December 2019, and was subsequently declared a global pandemic by the World Health Organization in March 2020 [1–3]. Severe acute respiratory syndrome (SARS) coronavirus 2 (SARS-CoV-2) is the etiological driver of COVID-19 [4], representing a major global health challenge that follows outbreaks of the Middle East respiratory syndrome CoV (MERS-CoV) and severe acute respiratory syndrome CoV (SARS-CoV) [2, 3, 5]. SARS-CoV-2 is a positive-sense single-stranded RNA virus with enveloped virions, which primarily target the epithelial cells, typically causing respiratory tract illnesses potentially leading to severe pneumonia and

acute respiratory distress syndrome (ARDS) [2, 6]. As of September 2021, more than 232 million COVID-19 cases and over 4 million associated fatalities have been reported affecting 220 countries and territories [7]. The entire world continues to battle the wave of COVID-19, although extraordinary efforts to prevent and treat the disease have been carried out globally to date, including lockdowns of cities and countries, as well as the widespread rollout of vaccines, supportive therapy, and treatment to prevent respiratory failure [8, 9]. SARS-CoV-2-induced COVID-19 in many ways resembles SARS caused by SARS-CoV given the acute clinical presentation, although with the addition of previously unknown peculiar pathogenetic, epidemiological and

Chuan-Xing Li, PhD, is a researcher at Karolinska Institutet (Sweden) and the First Hospital of Lanzhou University (China). Her research interest lies in computational precision medicine and multiomics integration.

Jing Gao, PhD, MD, is a researcher at the Karolinska Institutet (Sweden), the University of Helsinki and Helsinki University Hospital (Finland), and the First School of Clinical Medicine at Lanzhou University (China). Her research interests include respiratory medicine and bioinformatics.

Zicheng Zhang is a graduate student in the School of Biomedical Engineering, Wenzhou Medical University (China). His research interests include bioinformatics and immunoinformatics.

Lu Chen is a graduate student in the School of Biomedical Engineering at Wenzhou Medical University (China). Her research interests include bioinformatics and disease systems biology.

Xun Li, PhD, MD, is a professor at the Department of General Surgery and Key Laboratory of Biotherapy and Regenerative Medicine of Gansu Province at the First Hospital of Lanzhou University (China). His research interests extend across multiple disease fields, including COVID-19.

Meng Zhou is an associate professor in the School of Biomedical Engineering at Wenzhou Medical University (China). His research interests include bioinformatics, computational precision medicine and immuno-oncology.

Åsa M. Wheelock, PhD, is an associate professor at the Respiratory Medicine Unit, Department of Medicine and the Centre for Molecular Medicine at the Karolinska Institutet, Stockholm, Sweden. Her research interests involve molecular sub-phenotyping of heterogeneous diagnoses of obstructive lung disease, such as COPD, asthma and as of recent also post-acute COVID syndrome (PACS) using multiomics integration and systems medicine approaches.

Received: August 2, 2021. **Revised:** October 4, 2021. **Accepted:** October 23, 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

clinical features, thereby adding to the complexity of COVID-19 pathogenesis [3]. The estimated COVID-19 mortality is approximately 2.3%, lower than that for SARS (9.5%) and MERS (34.4%) [3], with the risk of COVID-19 mortality increasing with age and comorbidities, such that more than 20% of all reported disease-associated deaths have occurred in nursing homes in some countries [10]. COVID-19 has been shown to spread more readily in the community than MERS and SARS, primarily due to three major factors: (i) a large proportion of SARS-CoV-2 infected people display mild-to-no COVID-19 clinical symptoms [11]; (ii) SARS-CoV-2 appears to be contagious prior to symptom onset [12]; (iii) SARS-CoV-2 appears to primarily be transmitted through aerosols from exhaled air, which can linger in poorly ventilated areas for several hours, thereby contributing to so-called super-spreader events [13]. Further considering inconsistencies in the reporting of data from varying sources, tracing and estimating the spread of the pandemic remains challenging [10]. The manifestations of COVID-19 exhibit a range of differences from asymptomatic to exceptionally severe, with the possibility of presenting multiple systemic symptoms, including common symptoms such as a fever, shortness of breath, a persistent dry cough, chills, muscle pain, headache, a loss of taste or smell, renal dysfunction and gastrointestinal symptoms [3, 14, 15]. Age- and sex-related differences in both presentation and severity of COVID-19 disease are well documented [10, 16]. Multi-organ damage in COVID-19 patients has also been reported [1]. Furthermore, elderly COVID-19 patients with other comorbidities may present with severe respiratory manifestations [17]. Hence, an increased and comprehensive understanding of COVID-19 characteristics is crucial to guiding the pandemic response [10], and integrating multiple data modalities from the same patients in combination with appropriate patient stratification may prove advantageous in understanding COVID-19 disease.

Despite the rapid global scientific response to COVID-19, disease management and scientific research remain emergency-level challenges. Evidence indicates that both viral and host factors play critical roles in COVID-19, with host factors closely impacting disease severity [18, 19]. Multiomics datasets of high-throughput data, protein interactions and functional annotation profiling and approaches may provide a comprehensive understanding of the heterogeneity and diversity of data types and complex internal relationships in COVID-19. Identifying these characteristics using multiomics integration profiling covering specific stages and severities of COVID-19 disease may lead to a better understanding of specific virus variant characteristics and individual responses, disease manifestations and therapeutic opportunities, all of which facilitate the accurate prediction of patient outcomes as well as prevention and intervention in a personalized medicine manner. Moreover, novel multiomics approaches could contribute to increased

statistical power in small cohorts [20] and reduce the risk of overfitting statistical analyses. By summarizing the multiomics integration profiles of COVID-19, we may identify preventive and therapeutic targets along with early diagnostic tools, and more effectively predict critical or severe outcomes, better select drugs, and further our understanding of comorbid conditions and the long-term complications associated with COVID-19. In this review, we summarize the multiomics integration data published to date, including the integration of proteome, metabolome, lipidome, transcriptome, interactome, genome, secretome and cytokine, immune-related signatures and drug omics data to characterize a range of acute to long-term manifestations of COVID-19. We first summarize the applications of multiomics-based molecular characterizations in COVID-19 based on clinical stages and questions, and then, we group the major bioinformatics approaches used in multiomics integration analyses of COVID-19. Finally, we discuss various trends and challenges, such as the rapid growth of single-cell omics and network-based approaches. We argue that the use of multimodality integration situates COVID-19 within the purview of precision medicine, potentially improving disease management in the future.

Application of multiomics-based molecular characterization to COVID-19

A striking feature of COVID-19 is its heterogeneity [21], involving diverse underlying pathophysiological processes [22]. Clinical manifestations differ with age, although populations of all ages are susceptible to SARS-CoV-2 infection [17]. For instance risk factors for developing severe respiratory disease include higher age, male sex, and a number of pre-existing chronic disease, particularly multi-morbidity [17]; female gender is a risk factor for the long-term symptoms of fatigue combined with vagal and/or respiratory dysfunction associated with post-acute COVID syndrome (PACS), whereas the majority of previously healthy young adults, youth and children present with mild or even asymptomatic disease [23, 24]. Multiple symptoms have been reported in COVID-19, including but not limited to fever, cough, myalgia, hemoptysis, diarrhea and olfactory and taste disorders [25–29]. Multiomics data integration analyses could contribute to a better understanding of the molecular intricacy and variations of disease, providing new opportunities for studying COVID-19 in a more comprehensive manner given the unprecedented dimensionality and diversity of data available [30]. In this review, we highlight investigations of COVID-19 combining samples from multiple anatomical compartments, such as blood, bronchoalveolar lavage fluid (BALF), cell lines, throat swabs and tissues, combined with analyses from multiple molecular levels, such as DNA, mRNA, miRNA, proteins and metabolites. In addition, we provide an overview of how multiomics integration techniques contributed to the discovery of the molecular

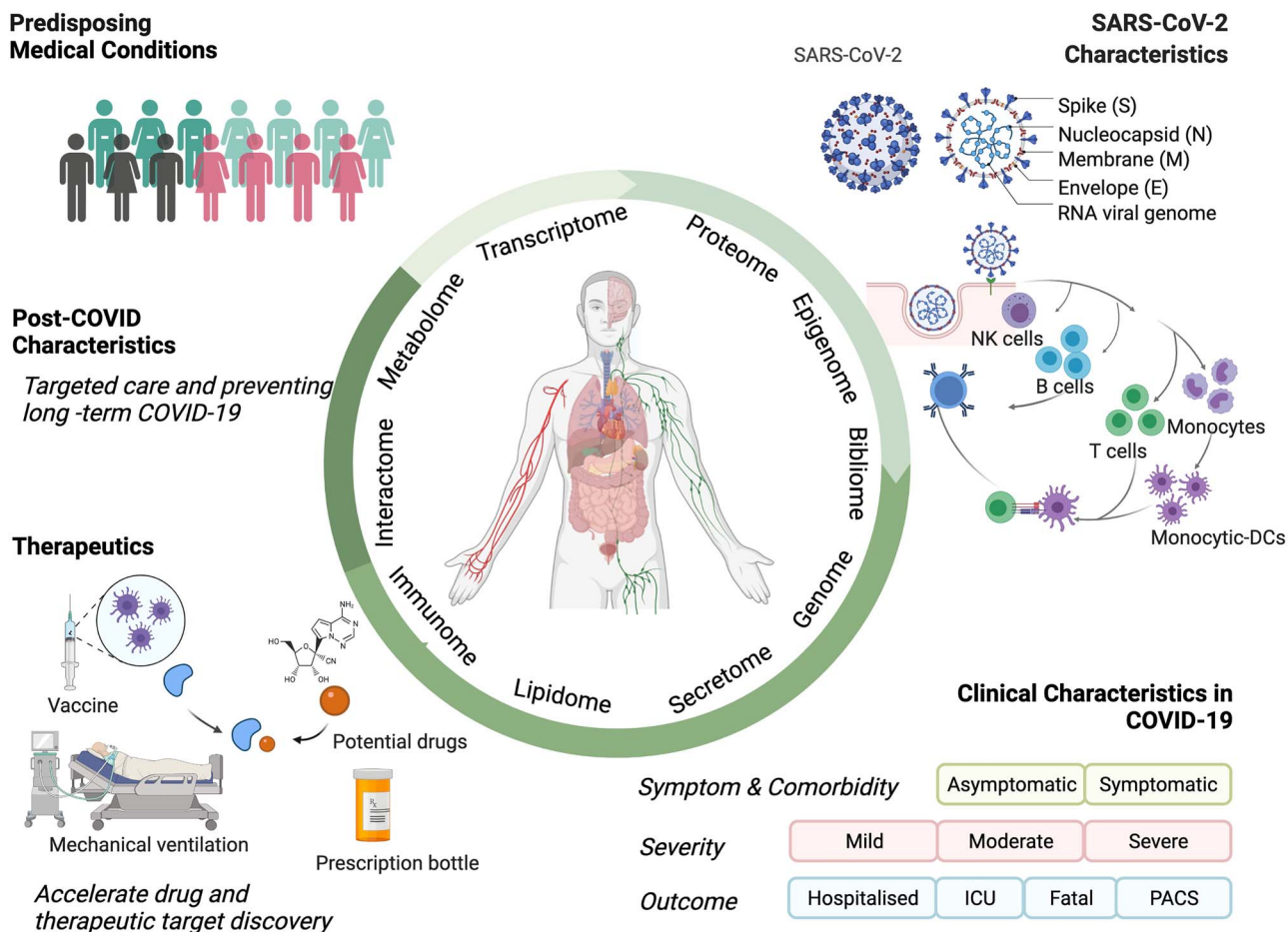


Figure 1. Applications of the multiomics integration-based molecular characterization of COVID-19. COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; MERS-CoV, Middle East respiratory syndrome CoV; ICU, intensive care unit; PACS, post-acute COVID syndrome. Created using [BioRender.com](https://www.biorender.com).

characteristics of COVID-19, including the virus and its functions, in addition to disease manifestations, severity and progression. Specifically, we summarize how multiple omics data can be applied to address issues such as predicting patient outcomes, identifying novel molecular targets for therapeutic interventions, and providing insights into data on COVID-19 [30] (Figure 1 and Table 1).

Multiomics investigation of SARS-CoV-2 characteristics and host responses

An integrated approach to investigations of SARS-CoV-2 could facilitate the study of complex biological processes holistically, while the integration of multiomics data highlights the interrelationships amongst the involved biomolecules and their functions [30]. As an unbiased data-driven technique [31], high-throughput omics can explore the complex characteristics of SARS-CoV-2 from multiple levels. Hence, multiomics and subsequent integrated analyses provide an opportunity to better understand SARS-CoV-2, facilitating insights into the pathophysiological heterogeneity of COVID-19 patients. Here, we summarize several examples to further explain how a multiomics approach helped

advance knowledge of SARS-CoV-2. In this review, we summarize methods and results from various studies through an integrative omics approach, clarifying their ability to address applications in SARS-CoV-2, such as examining mutations, host responses, infection biology, transmission characteristics and candidate drugs.

To explore SARS-CoV-2 mutation and host responses

Identifying the characteristics of COVID-19 strains helps better understand the virus' successful invasion and transmission, potentially aiding an understanding of the population-specific variations leading to a high rate of SARS-CoV-2 infections [32]. For example, Gupta et al. [32] presented a global view of the mutational pattern of SARS-CoV-2 through a multiomics levels (genome, proteome and interactome) using comparative genomics and integrated network approach, unveiling the phylogenetic patterns, co-mutational hot spots, functional cross-talk and regulatory interactions in SARS-CoV-2. Coronaviruses generally mutate to a limited degree compared to other types of viruses due to the presence of a high fidelity proofreading function during RNA replication. However, the high case load worldwide has resulted in an accumulation of single nucleotide

Table 1. Summary of the main biospecimen type and multiple omics data blocks from COVID-19 multiomics publications

Authors	Main biospecimen types	Proteome	Metabolome	Lipidome	Transcriptome	Interactome	Genome	Secretome/Cytokine	Immunome/Signature	Other omics
Su et al. [37]	Plasma	bulk	bulk					sc	TCR and BCR	
Overmyer et al. [18]	PBMCs Plasma	sc bulk	bulk	bulk						
Chen et al. [1]	Leukocyte Plasma	bulk	bulk		bulk exRNA bulk					
Thomas et al. [14]	Blood sample	bulk	bulk		bulk and sc					
Zhou et al. [54]	Red blood cells Bronchial epithelial cells Cell line (Caco-2 cells)	bulk	bulk	bulk						
Tomazou et al. [57]	Unspecified Serum BALF PBMCs Cell lines (A549, Calu-3 and NHBE)	bulk	bulk			PPI				
Barh et al. [21]	Lung sample				bulk					
Barh et al. [38]	Unspecified					PPI	GWAs			
Barh et al. [46]	Cell line (Caco-2 cells)	bulk								
Zhao et al. [73]	BALF				bulk					
Yang et al. [59]*	PBMCs				bulk					
Gupta et al. [32]	Cell lines ^a Unspecified Colostrum Unspecified SARS-CoV-2 viruses	bulk	bulk bulk	bulk	bulk	PPI	bulk			Bibliome
Shen et al. [70]	Unspecified					PPI				
Song et al. [69]	Serum	bulk	bulk							
Young et al. [33]	Plasma	bulk	bulk	bulk					immune mediators	
Bruzzone et al. [39]	Plasma	bulk	bulk	bulk						
Barberis et al. [78]	Respiratory samples									
Islam et al. [35]	Serum Plasma Nasopharyngeal samples Lung sample Cell lines (Calu-3 cells and NHBE)	bulk bulk	bulk bulk	bulk bulk	bulk bulk bulk				bulk (Δ 382) bulk (Δ 382)	
Chen et al. [58]	Unspecified Plasma Leukocyte Cell line (Caco-2 cells)	bulk	bulk	bulk	bulk bulk	PPI				

(Continued)

Table 1. Continued

Authors	Main biospecimen types	Proteome	Metabolome	Lipidome	Transcriptome	Interactome	Genome	Secretome/Cy-tokine	Immunome/Signature	Other omics
Sciacchitano et al. [55]	PBMCs				bulk		bulk	cytokine	immunological profiles immuno-transcriptome	
Muthuramalingam et al. [60]	Blood sample Unspecified				bulk	PPI				
Bernardes et al. [49]	Blood sample PBMCs Serum				bulk sc				bulk BCR-seq scBCR-seq	Epigenome
Sun et al. [36]	Serum	bulk	bulk		bulk			cytokine	antiviral antibodies	
Terracciano et al. [71]	Unspecified				bulk	PPI				
Stukalov et al. [40]	PBMCs Cell line (A549 cells)	bulk bulk			bulk					Ubiquitinome; Phospho- proteome
Chen et al. [79]**	Unspecified					PPI				
Mc Reynolds et al. [50]	Lung sample	bulk			bulk				immune cells	
Chen et al. [80]	Plasma	Bulk		Bulk				Cytokine		
Ahmed et al. [41]	Serum Cell line (NHBE)	Bulk			Bulk					
Galbraith et al. [22]	Unspecified Plasma PBMCs	Bulk Bulk	Bulk		Bulk	PPI; miRomics		Bulk	Seroconversion immune mapping	Bibliome
Stephenson et al. [47]	Blood sample				Bulk					
Liu et al. [56]	Red blood cells		Bulk		sc				TCR and BCR	
	PBMCs	sc			sc				TCR and BCR	
	PBMCs	sc			sc				TCR and BCR	
	Blood sample	Bulk						Bulk		

Note: Unspecified biospecimen types represent data resources from multiple or unspecified samples such as PPI or literature (bibliome). PBMCs, peripheral blood mononuclear cells; sc, single-cell; TCR, T-cell receptor; BCR, B-cell receptor; BALF, bronchoalveolar lavage fluid; PPI, protein-protein interaction; NHBE, normal human bronchial epithelial cells; GWAs, association results from the genome-wide association study; Δ382, 382-nucleotide deletion; exRNA, extracellular RNA; Cell lines^a, human lung epithelium-derived cell lines. ^aPerformed metabolomic profiling of Qingfei Paidu Decoction and transcriptomic profiling of a pneumonia rat model [81]. ^{**}Multiple proteome data blocks from several longitudinal timepoints. The full list and descriptions for all these publications including the purposes and summary of results, number of samples or cell lines are provided in Supplementary Table 1. The online searchable format shiny app can be found at 'Collection of multiomics datasets in COVID-19' (<https://zhougroupp.shinyapps.io/moCOVID/>).

polymorphisms (SNPs) resulting in the occurrence of a number of new variants [32]. The comparative genomics and integrated network study presented five central clades (a, b, c, d and e (e1 and e2)), which were identified and distinguished through co-mutational combinations in COVID-19. This includes clades d and e2, found only in the US strains [32]. The 382-nucleotide deletion (Δ 382) variant of SARS-CoV-2 was presented to the world, and an observational study using multiomics data clarified the association between the Δ 382 variant of SARS-CoV-2 and mild infection [33]. Moreover, van Dorp et al. [34] reported mutations to SARS-CoV-2 at both the genomic and protein levels. Yet, a multiomics integration study showed that around 67% of SNP mutations occurred at the amino acid level in COVID-19, although the distribution of SNP mutations and amino acid variations (AAVs) in the virus were not uniform [32]. In addition, the proteomic and genomic study identified critical proteins in COVID-19, whilst high-frequency AAV mutations were present in these proteins [32]. A mutation in SARS-CoV-2 proteins inhibits viral penetration to the host, and multiomics integration studies describe the viral behavior, the host response, and the virus–host interaction following SARS-CoV-2 infection [32, 35, 36]. Transcriptomes and a comparative study revealed disparate host responses against SARS-CoV-2, wherein innate immunity, interferon and cytokine stimulation served as key factors in the host-induced response to COVID-19 [35]. Multiomics results demonstrated that MYO-5 (A, B and C) proteins served as key host partners in COVID-19, strongly interacting with viral proteins (N, S and M) [32]. The viral proteins play critical roles in conferring SARS-CoV-2 pathogenicity, such as the presence of CpG dinucleotides in N and Nsp1 proteins, which may be involved in regulating pathogenesis [32]. Furthermore, virus–host protein interactions play a critical role in regulating the viral life cycle in COVID-19, where the generation of protein–protein interactions revealed that SARS-CoV-2 hijacked and/or altered cellular processes [36]. Comparative genomics and an integrated network established a host–pathogen interaction (HPI) model, which could serve as the fundamental basis for the structure-guided pathogenesis process inside the host cell in COVID-19 [32]. An multiomics integration approach described the observed SARS-CoV-2 virus characteristics and transmission globally, contributing to the development of tailor-made strategies to COVID-19 [32, 33, 35–37].

To understand SARS-CoV-2 biology and similarities

Compared to single omics research in COVID-19, integrated proteome, transcriptome, interactome and bibliome data from COVID-19 research details the biological factors associated with SARS-CoV-2 infection, including more enriched processes such as neutrophil degranulation in the virus [38]. For instance, Barh et al. [38] integrated multiomics (interactome, proteome, transcriptome and bibliome profiles) with host genetic

information, ultimately identifying the SARS-CoV-2 infection biology, potential drugs and prophylaxis agents against this virus. Furthermore, that study demonstrated that the primary interactions of the virus are related to the innate immune pathways, the host translation machinery and the Cullin ubiquitin ligase complex, whereas the critical pathway in SARS-CoV-2 infection is associated with the virus process, mRNA splicing, cytokine and interferon signaling, and ubiquitin-mediated proteolysis [38]. Dysregulated metabolomic and lipidomic profiles in SARS-CoV-2 infection revealed that the pathogenic redistribution of the lipoprotein is related to a high risk of atherosclerosis, whilst increased levels of ketone bodies associate with liver damage [39]. Furthermore, a multiomics integration (transcriptome, proteome, ubiquitinome and phosphoproteome) study of SARS-CoV-2 and SARS-CoV identified both distinct and shared molecular mechanisms in these related coronaviruses [40]. Notably, the two viruses exhibited different abilities to modulate both mitochondrial function and homeostasis through nonstructural protein 2 (NSP2), whilst dysregulating the TGF- β pathway and autophagy in SARS-CoV-2 respectively refer to the open reading frame 8 (ORF8) and ORF3, indicative of the important characteristics of SARS-CoV-2 biology [40]. A lung-specific protein interactomes study identified the hub proteins based on the expression data of SARS-CoV-2 infections, and the hub proteins are merely associated with MERS and human coronaviruses [41]. In addition, the multiomics-based identification of SARS-CoV-2 infection demonstrated that the virus shared a pathway with other diseases, including other virus-related diseases (i.e. influenza A, hepatitis virus, human T-lymphotropic virus 1, the Epstein–Barr virus and measles) as well as non-virus-related diseases (i.e. protozoan, bacterial and parasitic) [38]. By clarifying the biology and similarities of SARS-CoV-2, such findings could not only promote drugs development to block multiple infection pathways, but also guide virus- and host-dependent therapies in COVID-19 [40].

Identification of disease manifestations

To distinguish symptoms of COVID-19

The clinical presentation of COVID-19 consists of asymptomatic and symptomatic patients, but asymptomatic patients present with positive detections of the nucleic acid of SARS-CoV-2 and can transmit the virus to others [11]. A previous study summarized the most common short-term clinical symptoms of COVID-19, which included fever, cough, fatigue, dyspnea, sputum, diarrhea, nasal congestion and emotional disturbances [42–44]. Multiple pathophysiological processes have been implicated in the etiology of severe COVID-19 symptoms, for which integrated omics data could elucidate the interplay between these mutual processes [36, 45]. Barh et al. [46] identified mild and severe symptoms,

associated comorbid conditions, and short- and long-term complications of COVID-19, using a multiomics (proteome, transcriptome, interactome and bibliome) based bioinformatics approach, an approach with a precision exceeding 90%. In the multiomics-associated prediction model, 36 viral, 53 short-term, 62 short- to mid-long-term, 194 mid-long-term and 57 congenital conditions were identified amongst specific conditions [46]. Using this model, an average of 93% enriched conditions were identified as associated with COVID-19 [46]. Notably, a dry cough and loss of taste were excluded, although a cough has been reported as a common clinical symptom of COVID-19, highlighting the importance of the comprehensive knowledge needed to capture the clinical picture of COVID-19 [46]. In addition, a single-cell multiomics study presented the symptom-related mechanics of COVID-19, specifically identifying the relative loss of IgA2 in symptomatic disease [47]. These multiomics-based results describe symptom-related risk factors in individuals, thereby helping to fill current gaps in personalized approaches to COVID-19.

To determine severity in COVID-19

The clinical course of COVID-19 infection is highly variable in severity, initially falling into four types: mild, moderate, severe and critical [11]. Multiple symptoms accompany COVID-19, ranging from mild to critical [11, 48]. Patients with pneumonia show abnormal chest CT findings, with severe patients frequently requiring artificial ventilation [27, 48]. Multiomics research could provide better insight into the molecular features with cross-ome correlations further explaining any observed differences [18]. Quantified transcriptomic, proteomic, metabolic and lipidemic profiles uncovered correlations between various biomolecule classes, and allowed for the further mapping of critical molecular features of COVID-19, particularly in terms of status and severity [18]. The features were closely correlated to complementary activation, neutrophil activation and dysregulated lipid transport in COVID-19 [18]. Anticoagulation strategies remain contraindicated in COVID-19 with varying severities, whereas an integrated omics study provided novel insights into the field and identified a COVID-19 phenotype characterized by hypercoagulation [18]. Su et al. [37] reported a sharp disease-state shift between mild and moderate COVID-19 using integrated bulk and single-cell multiomics data, whereby therapeutic interventions are likely to be most effective for moderate cases, including an increase in inflammatory signaling and a loss of some sets of metabolites and metabolic processes. Furthermore, multiple unusual immune cell phenotypes emerged in moderate COVID-19 and amplified with increasing disease severity [37]. Interestingly, these novel subsets did not appear with mild COVID-19, whilst sharp differences emerged when comparing mild and moderate disease, involving inflammation signals, immune cell function and plasma metabolite composition [37]. Another multiomics dataset showed that tissue-specific proteins and extracellular RNA

(exRNA) expression significantly differed between mild and severe COVID-19 patients [1]. Proteomic, metabolic, transcriptomic and cytokine data identified milder disease symptoms, which accompanied significant T-cell responses [1], whilst single-cell multiomics data revealed that circulating follicular helper T-cells accompanied mild COVID-19 [47]. Moreover, a single-cell multiomics study also demonstrated an increased level of T-cells in severe COVID-19, including an increased ratio of CD8+ effector T-cells to effector memory T-cells and expanded CD8+ T-cells [47]. Nevertheless, several biomarkers were identified in severe COVID-19 based on multiomics studies, such as megakaryocytes, erythroid cells, plasmablasts and plasma linoleate diols [49, 50]. Thus, identifying features and differences amongst patients with varying severities of COVID-19 based on all omics data may prove more effective in understanding the potential therapeutic targets to modulate and assist in developing of predictive models of disease severity, resulting in patient better outcomes [18].

To investigate comorbidity in COVID-19

Hypertension, cardiovascular diseases (CVDs), diabetes mellitus chronic kidney diseases (CKDs), cancers and chronic obstructive pulmonary disease (COPD) are common comorbidities or pre-existing conditions in COVID-19 associated with worse outcomes [21, 51–53]. The risk of mortality in COVID-19 increases with age, gender and comorbidities, whilst age, male gender, shortness of breath, cerebrovascular disease and COPD were mortality-associated factors [16]. A multi-country study revealed that survival time from symptom onset was significantly shorter in elderly versus young patients, males versus females and patients with versus patients without comorbidities [16]. Multiomics studies contributed to identifying the pathway crosstalk between COVID-19 and comorbidities, helping to clarify these shared pathways and to identify gene-based targets for COVID-19 [21]. An integrated omics approach identified common comorbidities associated with COVID-19 including multiple systematic diseases or disorders, such as immune, pulmonary, cardiovascular, metabolic, liver, kidney and other systems [46]. Similarly, a high-throughput experimental data and protein interactome study identified the hub proteins in COVID-19, indicating the likelihood of symptoms similar to other lung diseases, such as asthma, COPD and pneumonia [41]. Meanwhile, a network medicine approach indicated that COVID-19 shared an intermediate inflammatory molecular profile with asthma [54]. Multiomics integration studies further identified the pathobiology, the viral modulated functional hubs of COVID-19 and pathway crosstalk between COVID-19 and other diseases, specifically considering pre-existing comorbid conditions [21, 41, 54]. Such findings demonstrated that COVID-19 shared pathobiology with inflammatory bowel disease [54], whilst the protein hubs in viral replication were correlated with hypertension, CVD and diabetes [41], and pathway crosstalk was identified between

COVID-19 and diabetes, hypertension, CVDs and chronic kidney disease [21]. In addition, a pathway crosstalk was identified between COVID-19 and cancer [21], whilst the immune and genomic results demonstrated that a low triiodothyronine (T3) syndrome could coexist [55]. These multiomics-based results describe various risk factors in individuals, improving current gaps in personalized approaches to COVID-19.

To predict disease progression and outcomes in COVID-19

Multiomics integration data could be used to determine a biosignature to identify COVID-19 progression, prognostics and predicting outcomes amongst heterogeneous patients and guide therapeutic interventions. For example, multiomics profiles revealed that features of the stages of seroconversion could be used to stratify COVID-19 patients and correlate with different pathophysiological states, involving multiple underlying pathophysiological processes such as hyperinflammation and thrombotic microangiopathy [22]. Low antibody titers correlated with hyperactive T-cells, and the depletion of neutrophils, lymphocytes as well as platelets. Upon seroconversion, these biosignatures decrease or fully reverse, leading to an increase in B-cell subsets [22]. Meanwhile, D-dimer and hypoalbuminemia increased upon seroconversion microangiopathy [22]. Moreover, a multiomics study by Chen et al. [1] using transcriptomic, proteomic and metabolomic profiles provided the molecular features of the pathophysiology in disease progression and clinical outcomes, allowing for the construction of a COVID-19 prognostic classification model. Integrating these plasma omics profiles with multiple datasets as well as the features of expressed genes, proteins, exRNAs and biochemical parameters as potential biomarkers can help predict varying prognoses in COVID-19 patients and stratify those patients into groups [1]. Furthermore, novel blood biomarkers identified through a multiomics integration model revealed a clear distinction between a good and poor prognosis in COVID-19 [1]. For instance, integrating the clinical and proteomics datasets, increased blood clotting factor levels and a decline in the expression of the coagulation factor XIII A chain (F13A1) marker correlated with a poor prognosis in COVID-19 patients [1]. Meanwhile, megakaryocyte and erythroid cell derived co-expression modules derived from multiomics data were identified as predictors of fatal outcomes in COVID-19 [49]. Furthermore, a multiomics transition study demonstrated fluctuations in red blood cell (RBC) and hemoglobin (Hb) levels between survivors and non-survivors [36]. A network analysis revealed a late juncture in fatal COVID-19 based on the simultaneous assessment of proteins, transcriptomes and T-cell receptor sequences, whereby distinct circulating protein trajectories appeared useful in predicting recovery versus fatal outcomes [56]. These findings refine our comprehensive understanding of the pathophysiology and clinical progression of COVID-19, and ultimately could be used to improve COVID-19 interventions and therapies.

Accelerated development of drugs, prevention tools and therapies for COVID-19

Despite multiple vaccines being widely and globally deployed, the pandemic continues to progress, and effective treatment remains largely unavailable. One of the primary purposes of understanding the characteristics of COVID-19 through multiomics integration data is to accelerate the development of drugs, vaccines, prevention tools and therapies for COVID-19. Multiomics data suggest that a high-frequency amino acid variation (AAV) in mutations could be considered for the novel design of a vaccine [32], whilst the clinical impact of deletions in ORF8 could be used to develop treatments and vaccines for COVID-19 [33]. Multiomics data with multiple source integration from patients could drive multiplex drug repurposing and offer rapid mapping and drug prioritization for COVID-19 [57]. Specifically, multiomics integrative analyses with a multiplex drug repurposing (DR) approach yielded a highly informed shortlist of drug candidates against COVID-19 and its causative virus, including drugs aiming to reverse COVID-19-induced perturbations (e.g. immunomodulatory and anti-inflammatory drugs) and compounds with a direct antiviral activity [57]. For instance, Src tyrosine kinase inhibitors hold a promising potential against COVID-19 [57]. Integrated interatomic, proteomic, transcriptomic and bibliomic profiles showed several prophylaxis agents (e.g. curcumin, vitamin D and melatonin) and candidate drugs (e.g. betamethasone and cyclosporin A) against COVID-19 [38]. A multimodal data harmonization approach identified a list of 84 drug target candidates and 7 high-confidence targets as potential starting points for drug therapy and development in COVID-19 [58], and also presented the potential candidate in phytochemicals and Chinese medicine [59, 60]. Meanwhile, multiomics and systems pharmacological findings revealed 154 compounds that targeted 13 immune genes involved in diverse signaling pathways, suggesting novel potential targets for COVID-19 treatment and prevention [60]. Thus, multivariate approaches to harmonize multiomics data demonstrated advantages in developing COVID-19 prevention and treatment methods, an approach that effectively worked in other diseases.

Post-acute COVID-19 syndrome (PACS) characteristics

Mounting evidence suggests that symptoms may last for months after recovery from the initial COVID-19 infection [9, 61]. Whilst definitions vary with symptoms persisting for more than 4 to 12 weeks, a condition termed post-acute COVID syndrome (PACS) [23] or in some contexts simply long COVID has been defined [9, 62]. Whereas this group also includes patients with lingering symptoms after an initially acute severe disease which required hospitalization, the majority of this group consists of patients that initially experienced mild disease with little to no symptoms [25]. The latter group

typically consists of patients who are young, previously healthy, and mentally and physically strong [63, 64]. In addition, the female gender appears to be twice as likely to develop long COVID as males [24]. The most common long-term sequelae symptoms include fatigue, dyspnea, chest pain, recurring fever, neurological complaints, olfactory dysfunctions, tachycardia, intestinal disorders and skin manifestations, and possibly unrelated to the severity of the initial acute infection [9, 24, 62–64]. In the Coverscan study [65], damage to at least one organ was observed in 70% of patients, with multiple organ damage appearing in 25% of patients with long COVID. Furthermore, long COVID does not appear to consist of prolonged recovery from the initial infection, but rather a cyclic condition with fatigue and a recurring fever exacerbated after activity. The upregulated omics profiles in COVID-19 might remain switched on, due to viral RNA remaining active in survivors, suggesting that a corresponding non-communicable disease may develop and lead to long-term consequences [46], whilst the long-term symptoms may change in nature over time [62]. A comprehensive understanding of the characteristics in post-COVID remains unmet, whereas very limited multiomics studies in this field have been reported. For instance, Barh et al. [46] used a multiomics approach to predict possible long-term complications of COVID-19, achieving high accuracy of over 90%. A multiomics-based bioinformatics approach identified the most common conditions accompanying COVID-19, noting that these play a potential role in the long-term consequences for COVID-19 survivors [46]. Hence, multiomics datasets may truly predict the growing effect of the disease beyond hospitalization and mortality, and accurately target care for survivors through the development of preventive and effective treatments for COVID-19 [9].

Multiomics integration approaches

The development of multiomics integration bioinformatics approaches has provided an opportunity to comprehensively understand a complex disease involving multiple molecular-level changes and their interactions, representing an important direction in systems and precision medicine. In previous reviews, these approaches could be grouped based on their purpose: disease subtyping and patient stratification; diagnostic and prognostic predictive model construction; and driving insight into disease biology through the identification of biomarkers, pathways and disease networks [30]. Such approaches could also be distinguished based on algorithms, such as multivariate, machine learning, network, correlation or similarity and a fusion of approaches using either multistaged (linear or sequential integration) or meta-dimensional (or simultaneous integration) integrative strategies. (Further details and a comparison of these latter approaches can be found in reviews presented elsewhere [30, 66, 67]). Given the exceptional heterogeneity and diversity

of multiomics datasets as well as the widespread availability of external databases, multiomics integration processes are quite project-oriented and specific. Here, we summarize several commonly used approaches in the multiomics molecular characterization of COVID-19 (see Figure 2).

Multivariate analyses

Multivariate analyses such as partial least squares (PLS) and its extension multi-block PLS are supervised multivariate analyses seeking predictable features and attempting to build predictive models for both the severity and outcomes of COVID-19. Liu et al. [56] built a logistic PLS model based on changing protein expressions in a longitudinal cohort to predict fatal outcomes in COVID-19. Similarly, Chen et al. performed a supervised PLS on each single omics dataset and multiblock PLS (DIABLO) to integrate the multiomics dataset, including proteome, transcriptomes, translome, lipidome and metabolome data. They revealed a harmonized multimodal signature that can further assist in the discovery of drug targets [58]. In addition, mixOmics is a commonly used tool for multiomics integration, which also includes additional multivariate methods for future applications [68]. One advantage to multivariate analysis is its simultaneous ability to select features and generate predictable models. However, it carries a limited predictive power when the data are complex, such as with nonlinear relationships.

Machine learning in predictive models

Machine learning approaches have a higher potential to construct efficient predictive models for complex and ambiguous data. Specifically, such approaches could be used in unsupervised COVID-19 subtyping and patient stratification, and in constructing supervised diagnostic and prognostic predictive models. For instance, Overmyer et al. [18] constructed an ExtraTrees classifier based on combining four omics datasets of metabolites, lipids, proteins and transcripts to predict disease severity. In a study by Song et al. [69], they first selected significantly differentiated lipids and polar metabolites comparing healthy and COVID-19 patients, and then constructed a predictive logistic regression model for COVID-19 with high accuracy (AUC=0.975). Similarly, Chen et al. [1] selected variables and trained prognostic models using multiple machine learning algorithms, including the nearest mean classification (NMC), k-nearest neighbors (KNNs), support vector machine (SVM) and random forest (RF), through the integration of clinical measurements, exRNA-seq, mRNA-seq and proteomics datasets. Relatedly, Shen et al. [70] built an RF model based on proteomic and metabolomic data from 18 non-severe and 13 severe patients, further prioritizing 22 proteins and 7 metabolites using the mean decrease in accuracy. Generally, machine-learning approaches have the advantage of high accuracy, flexible requirements and hypotheses related to data distribution, and

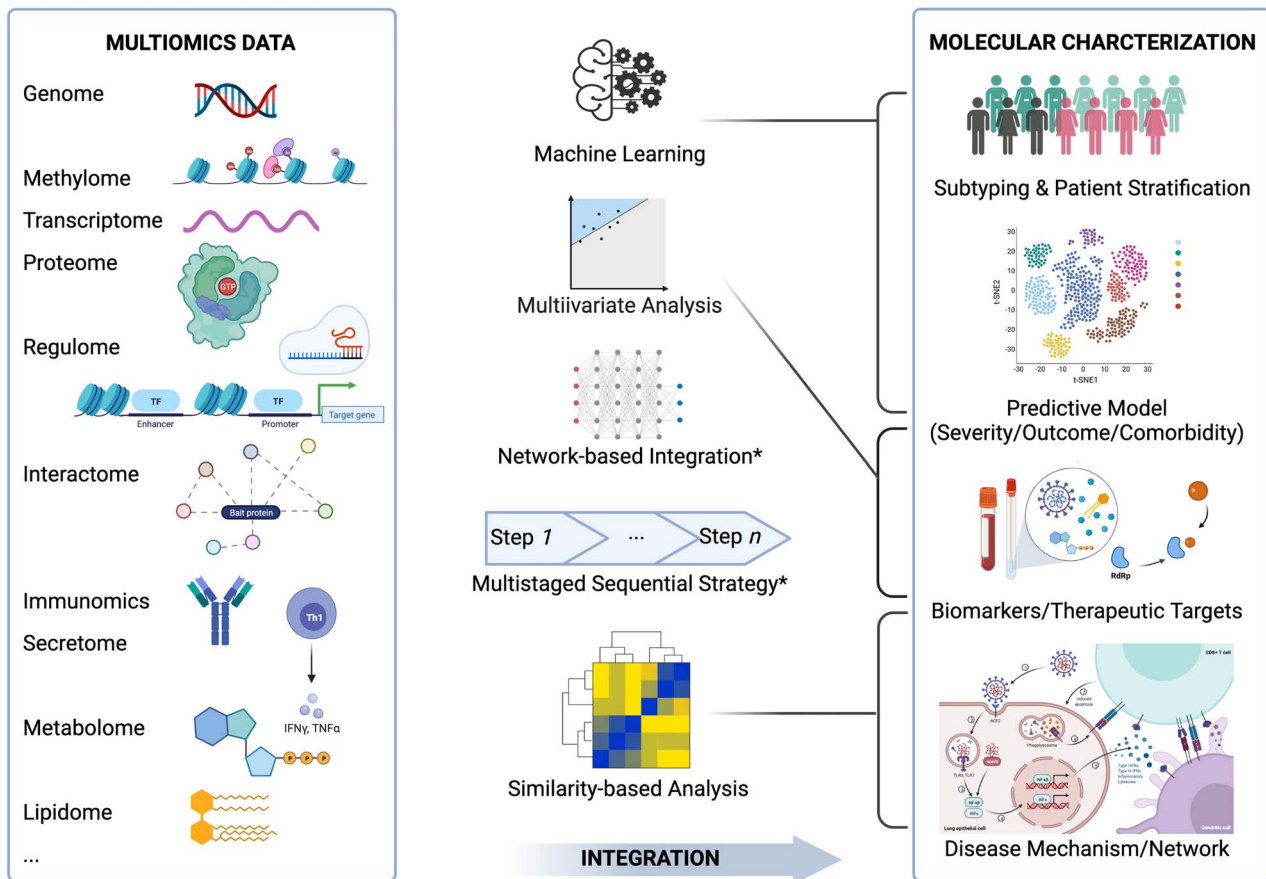


Figure 2. Summary of five categories of multiomics integration strategies and their application in the molecular characterization of COVID-19. Commonly used multiomics data (left box) are integrated through five categories of integration approaches (middle) to investigate four major applications in the molecular characterization of COVID-19 (right box). The grey lines from the middle to the right represent the major applications of approaches for specific purposes. Both network-based and multistaged strategies have been performed for all four applications. Created using BioRender.com.

high compatibility with data types. Nevertheless, such approaches require a greater effort to avoid overfitting, a larger sample size, and further downstream explanations to link the predictable features or models with biological insights.

Network-based multiomics integration to understand COVID-19 infection and the immune response

Network-based integration approaches have been increasingly applied in multiomics COVID-19 research given that such approaches match the nature of multilevel, complex and crosstalk interaction networks involved in SARS-CoV-2 infection and COVID-19 disease progression. Network approaches could illustrate and explain interactions amongst multilevel molecules, including DNA, RNA, protein, metabolites and their interactomes, as well as cellular interactions and immune responses across all processes from SARS-CoV-2 exposure to COVID-19 disease and post-COVID syndrome. Terracciano et al. [71] utilized a multilevel approach to mapping the SARS-CoV-2–host protein–protein interactome using affinity purification mass spectrometry with a graphical network

representation combined with the simultaneous analysis of the host transcriptome, proteome, ubiquitinome and phospho-proteome following viral infections to discover host-directed anti-SARS-CoV-2 therapeutics. Similarly, Zhou et al. [26] incorporated SARS-CoV-2 virus–host protein–protein interactions, transcriptomics and proteomics into the human interactome to reveal the underlying pathogenesis. They further investigated COVID-19-associated disease manifestations by evaluating the network-based relationships of 64 diseases across enriched function categories and drug repurposing for COVID-19 [54]. In addition, Stukalov et al. [40] integrated the interactomes of both SARS-CoV-2 and SARS-CoV, examining their influence on the transcriptome, proteome, ubiquitinome and phosphoproteome of a lung-derived human cell line to identify host permutations in multilevel proteomics through a network diffusion approach. Understanding the immune response and system changes represents an important step to gaining insight into COVID-19, and identifying predictive and therapeutic signatures and targets. Additionally, Liu et al. [56] performed conditional independence network analysis to create a severity network through the

identification of direct and indirect associations between enriched differentially expressed gene sets and a disease severity metric using a Lasso regression model. The development of single-cell omics technology allows for investigating the microenvironment and changes to the immune cell and their communication. In doing so, the B-cell receptor (BCR) and T-cell receptor (TCR) represent the most interesting and useful signatures for B-cell and T-cell responses and clonotypes. For example, Stephenson et al. [47] constructed a single-cell BCR clonotype network using adjacency matrices computed from the pairwise Levenshtein distance of the full amino acid sequence alignment for BCRs contained in every pair of cells from the peripheral blood mononuclear cells (PBMCs) within each COVID-19 severity cohort. Based on the clonotype network, Stephenson et al. [47] discovered some evidence of class switching within symptomatic COVID-19 groups, but not in asymptomatic or healthy individuals, whilst also observing some difference between men and women. Similarly, Tomazou et al. [57] devised a network-based integration of multiomics data, combining proteomics and metabolomics, transcriptomics, genomics and the pathogen–host network from COVID-19 patients and a drug repurposing cell line, attempting to prioritize the most important genes related to COVID-19 (gene–disease association ranking), subsequently reranking the identified candidate drugs. Whilst multilevel molecular and cellular networks have been investigated, there is enormous potential for network-based multiomics molecular research in COVID-19. Alongside the molecular network, a network-based patient similarity network has proven efficient in identifying disease subtypes and further predicting and stratifying patients in multiple cancers and chronic obstructive pulmonary disease. This line of inquiry represents a candidate for future research in COVID-19 patient diagnosis and outcome stratification.

Multistaged sequential integration to identify disease-related biomarkers, pathways, or networks

Multistaged sequential approaches are typically used to integrate parallel analyses from each single-omics or layer of data. One method involves identifying differentially expressed signatures or genes from each single-omics platform, then merging these lists based on their direct overlap or further enriched downstream functions or pathways. For example, Chen et al. [1] identified severity-related multiomics signatures and their enriched functions and pathways, and then, further associated potential multi-organ damage based on tissue-enhanced proteins and functions. Interactome (both virus–host and human interactome), regulome, bibliome, gene ontology and pathways are commonly used external databases, combined in multiomics research. The more heterogeneous and diverse the data type, the greater the number of requirements

to the bioinformatics approaches. Barh et al. [38] developed a strategy including gene set enrichment, lung-specific protein–drug network and candidate gene analysis to identify SARS-CoV-2 infection biology and candidate drugs to counter COVID-19 using interactome, bibliome, longitudinal proteome and transcriptome data. A multistaged method could flexibly integrate several analytical modules based on a specific research question and require cautious efforts to reproduce results.

Similarity and correlation-based therapeutic targets and drug selection

The similarity between molecular changes and drug responses from a disease view relies on the basic premise of similarity or correlation-based analysis in COVID-19. For instance, Overmyer et al. [18] performed a cross-ome correlation analysis, correlating plasma proteins, metabolites and lipids omics data, identifying clusters further associated with COVID-19 patient status. Similarly, Song et al. [69] performed a differential correlation analysis of plasma lipids in mild COVID-19 cases relative to healthy controls, constructing a correlation network in mild COVID-19 cases. Su et al. [37] constructed a COVID-19 severity-dependent cross-omic interaction network, which included clinical measurements, plasma metabolomic and proteomic data. They further resolved an orchestrated response gene module that correlated with clinical features by integrating both high-dimensional bulk and single-cell multiomics profiles across cell types relying on a surprisal analysis [37]. This module could further characterize coordinated changes to cell types across COVID-19 patients. Similarity and correlation-based approaches could fast track links between COVID-19 and existing findings related to disease mechanisms, therapeutic targets and drugs from other similar diseases, thereby accelerating the development of therapeutics as well as prevention and diagnostic tools for both pre-COVID-19 medical conditions and post-COVID syndrome.

Trends and challenges in multiomics research on COVID-19

COVID-19 represents a challenging emergency both in terms of disease management and scientific research. Multiomics integration molecular characterization of COVID-19 patients will greatly assist preventive and personalized medicine efforts. In our review of 32 multiomics studies on COVID-19 published before July 2021 featuring at least two integrated omics datasets, we identified some trends and challenges in multiomics research on COVID-19.

Systematic similarity-based association assumptions to improve understandings of COVID-19 mechanisms

Similarity-based assumptions, ranging from virus and disease phenotype to drug response correlation at the

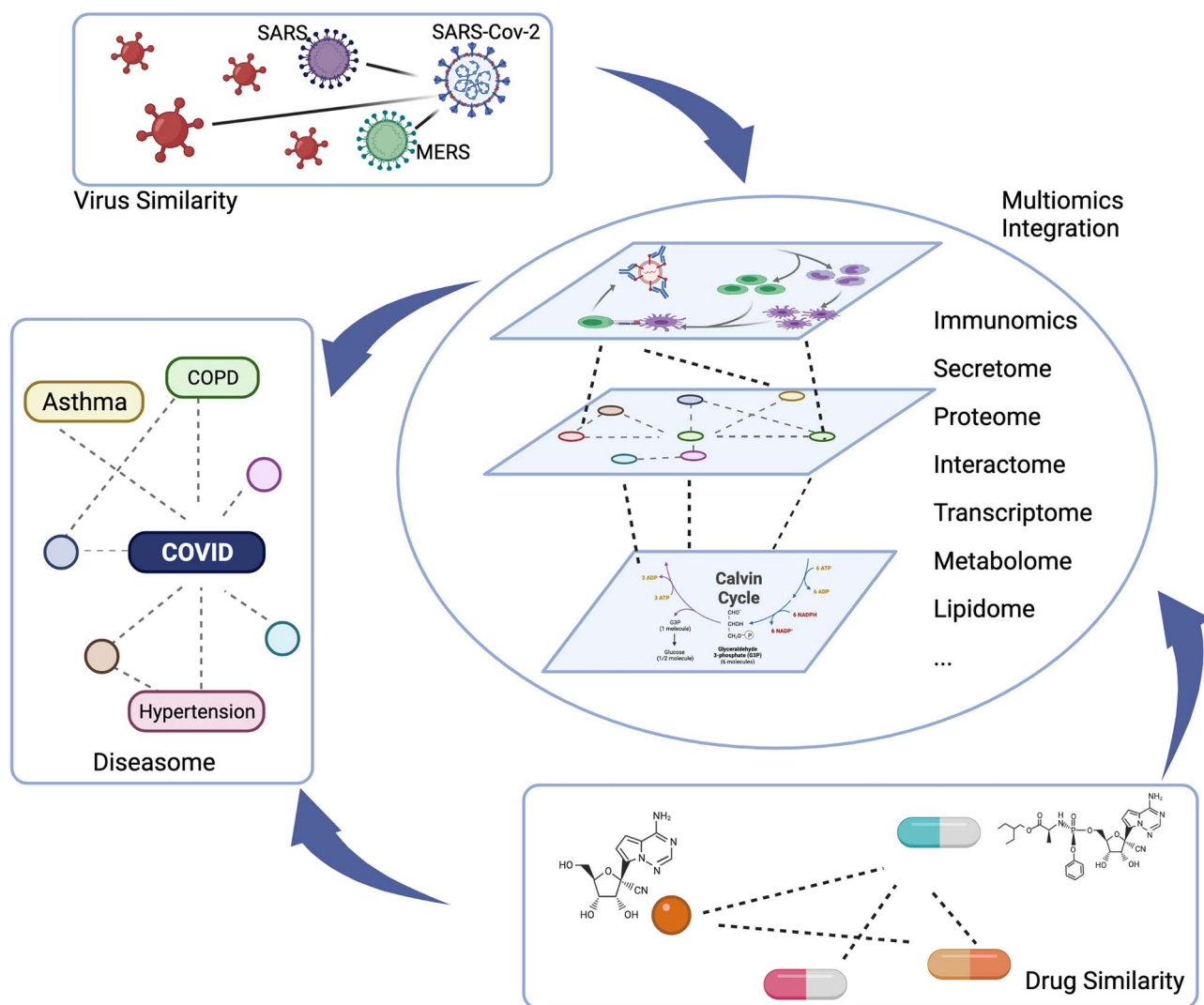
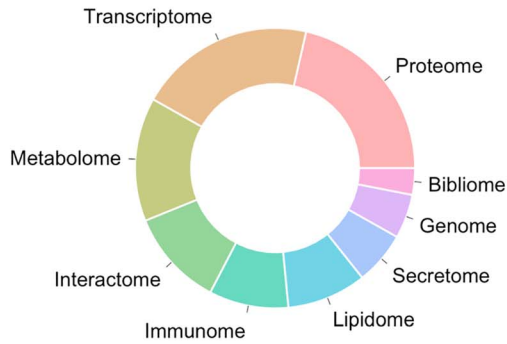


Figure 3. Illustration of the similarity-based assumptions for the transfer of previous knowledge to multiomics integration in COVID-19. Multiomics integration from immunomics, secretome, proteome, interactome, transcriptome, metabolome and lipidome amongst others could provide a systematic understanding of viral infection and COVID-19 disease progression and processes. SARS-CoV-2 is similar to SARS and MERS, as well as other viruses. Based on their similarities, the virus–host response and potential diagnostic and therapeutic targets derived from multiomics analyses could be transferred to and prioritized within COVID-19 research. Based on diseaseome, given the similarity with known diseases, the candidate genes from diseases similar to COVID-19 could be analyzed and examined, particularly in relation to a predisposition to disease, comorbidities and in predicting long-COVID and characterizing patients. Drug similarities in terms of the chemical effects as well in the multiomics-level response could be used to prioritize candidate drugs and therapeutic targets. Created using [BioRender.com](https://www.biorender.com).

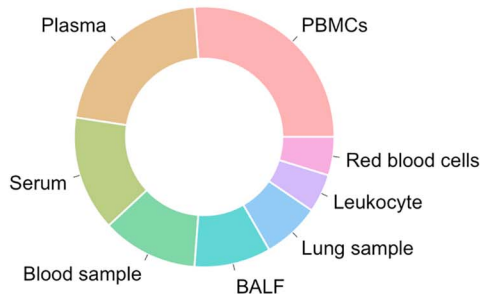
multiomics level, represent one fundamental hypothesis, which applies previous knowledge to COVID-19 research (see [Figure 3](#)). Briefly, multiomics-level reflections from immunomics, secretome, proteome, interactome, transcriptome, metabolome, lipidome and epigenome amongst others could provide a systematic understanding of viral infection and COVID-19 disease progression and processes. In addition, three levels of similarities could accelerate the transfer of previous knowledge to multiomics integration in COVID-19. SARS-CoV-2 is similar to SARS and MERS, as well as other viruses. Based on their similarities, the virus–host response and potential diagnostic and therapeutic targets from multiomics analyses could be transferred to

and prioritized within COVID-19 research. Based on diseaseome as the similarity of known diseases, the candidate genes from diseases similar to COVID-19 could be analyzed and examined, particularly in relation to a predisposition to disease, comorbidities, and in predicting long-COVID and characterizing patients. Drug similarities in terms of chemical effects as well in the multiomics-level response could be used to prioritize candidate drugs and therapeutic targets. Overall, the knowledge linking and transfer under the systematic similarity-based assumptions, encompassing virus and drug similarities, diseaseome and multiomics molecular networks, may accelerate our understanding of COVID-19 mechanisms.

A. Frequency of omics data



B. Frequency of biospecimen type



C. Combinations of omics data

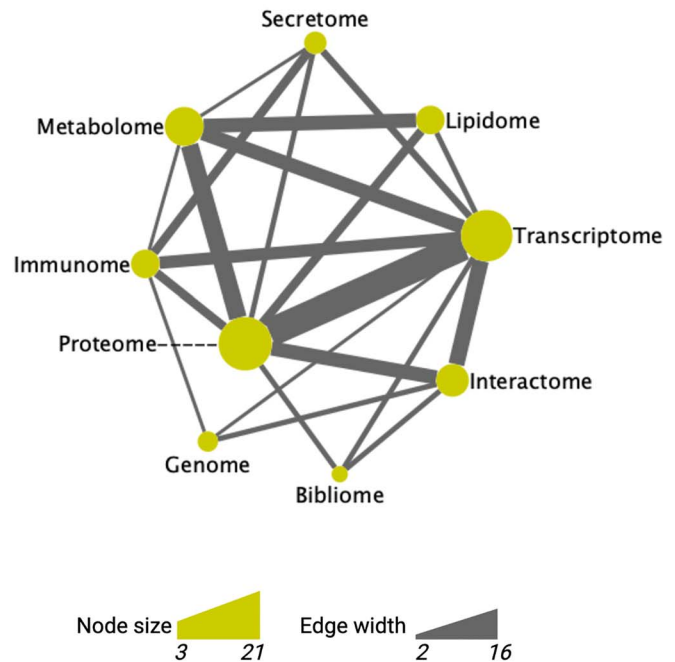


Figure 4. Summary of platform co-appearance in 32 multiomics studies of COVID-19. Pie charts of the prevalence of various omics platforms (A) and biospecimen types (B) that appeared in at least two publications, respectively. C) A network plot of the co-appearance of omics platforms, in which omics (node) appeared in at least two publications and omics pairs (edge) co-appeared in at least two publications. The size of the nodes corresponds to their appearance number in these publications (3 to 21). The width of the edge is related to the number of the co-appearances in these publications (2 to 16). Detailed information is available in [Supplementary Tables 2–4](#). PBMCs, peripheral blood mononuclear cells; BALF, bronchoalveolar lavage fluid. Pie charts and network are created using R version 3.6.0. and Cytoscape version 3.8.2, respectively.

Diversity and complexity of multiomics data and bioinformatics approaches pose a challenge to reproducibility

More than ten different kinds of omics or molecular signatures from over 15 biospecimen types have been used in the 32 multiomics studies discussed in this review (omics approach and biospecimen type used in each publication are detailed and summarized in [Supplementary Tables 2 and 3](#)). Nine of the 12 omics data types were used in at least 3 of 32 studies, including proteome, transcriptome, metabolome, interactome, immunome/signatures, lipidome, secretome/cytokine, genome and bibliome ([Figure 4A](#)). To highlight this data crossover, we constructed an omics co-appearance network, consisting of omics (node) that appeared in at least two publications and omics pairs (edge) that co-appeared in at least two publications. The size of nodes and the widths (weights) of edges correspond to the number of publications in which they were used (node) and co-appeared (edges), respectively. There are dense and complex combinations between these omics datasets. The seven most frequently used biospecimen types, used in at least two publications are visualized in [Figure 4B](#), including PBMCs (corresponding to single-cell omics), plasma, serum, blood sample, BALF, lung sample, leukocyte and red blood cells (excluded unspecified/multiple recourse such as database and cell lines). Notably, epigenetics and epigenomics play

important roles in SARS-CoV-2 infection and COVID-19 pathogenesis, as discussed in Milad Shirvaliloo's article, 'Epigenomics of COVID-19 and the link between DNA methylation, histone modifications and SARS-CoV-2 infection' [72]. In multiomics research, Zhao et al. [73] identified broad cellular effects from SARS-CoV-2 infection beyond adaptive immune cells through DNA methylome and transcriptome profiles. In addition, Shen et al. [70] revealed the crosstalk between perturbations taking place upon infection with SARS-CoV-2 and SARS-CoV at different levels, whereby multiomics were involved, including ubiquitinome and phosphoproteome. Multiple investigative opportunities on COVID-19 multiomics research exist in different multiomics combinations with diverse bioinformatics approaches, given the variety of clinical backgrounds.

One challenge when summarizing the multiomics data stems from the diversity and the ambiguous definition of omics data. For example, 32 cytokines expression levels are referred to as secretome in Su et al.'s study [37], but McReynolds et al. [50] referred to these as simply cytokines even though they analyzed 50 cytokines. Here, we grouped data into each omics type using an inclusive definition, whereby both Su et al. and McReynolds et al.'s studies are mapped into secretome/cytokine and both bulk and single-cell omics are attributed equal importance. It should be noted that the ambiguous definitions discussed above may have

given rise to a certain level of inconclusiveness in our literature mining.

Another layer of challenges inherent in attempts to summarize such studies lies in the variety and complexity of bioinformatics approaches used. Multiomics integration is still an emerging and dynamic field, and standardized or generalized methods regarding the minimal number of combinations of omics, biospecimen types, clinical backgrounds, and other data modalities required to use the terminology, or approaches for mapping to PPI or pathways exist as of yet. The diversity and complexity in multiomics integration, on the one hand, provide greater possibilities to improve the statistical power and molecular depth of COVID-19 investigations, while on the other hand, the propensity for overfitting and difficulty in finding matching validation cohort may increase the difficulty in reproducing multiomics research, as well as accurately estimating systematic errors and bias overlap between data collected from different molecular levels and anatomical locations.

The availability of high-throughput multiomics datasets both in bulk and at the single-cell level, in combination with the availability of relatively comprehensive databases for protein interaction, functional annotation and molecular pathways offer vast opportunities for mechanistic interpretations of multiomics integration networks. Conversely, heterogeneous and diverse data types with complex internal relationships give rise to challenges to the computational approaches. Machine learning approaches along with more recently developed multiomics integration approaches such as Similarity Network Fusion [74] and iCluster [75] have proven highly efficient, but with a high risk of overfitting if the study design and cohort size is inappropriate for the chosen method. Another challenge lays in the gaps between the availability of powerful statistical methods based on thousands of molecular features, and a lack of matching clinical and biological explanations.

Single-cell omics analyses in COVID-19

Single-cell omics provides an unprecedented potential for exploring biological systems also when sample sizes are scarce, and represents an emerging trend in COVID-19 research as well as in many other areas [76]. To name a few, Su et al. [37] performed single-cell multiomics analyses of PBMCs using whole transcriptome, surface proteins, secreted proteins and TCR and BCR gene sequence data, and thereby identified a sharp disease-state shift between mild and moderate COVID-19, suggesting that moderate COVID-19 may provide the most effective setting for therapeutic intervention [37]. Stephenson et al. performed single-cell transcriptome investigations of the immune response in COVID-19 of 130 patients with varying severities of COVID-19, highlighting the comprehensive landscape of a coordinated immune response contributing to COVID-19 pathogenesis [47]. In a longitudinal multiomics COVID-19 study, the authors

employed parallel scRNA-seq, single-cell BCR profiling, bulk mRNA sequencing (RNA-seq), BCR amplicon sequencing and multicolour flow cytometry from baseline to follow-up [49]. Through integrations of single-cell and bulk sample data, their analysis identified a few hallmarks of severe COVID-19, such as megakaryocytes, erythroid cells and plasmablasts [49]. A more in-depth analysis on the subject on single-cell analyses in COVID-19 is provided in a recent review by Huo et al.'s review [77]. The integration of bulk- and single-cell multiomics also pose a new set of challenge for bioinformatics algorithms.

The importance of cohort design in multiomics investigations

In this review, we discuss the current status of and potential for multiomics integration in COVID-19 research. The investigations performed to date primarily focused on predicting SARS-CoV-2 infection and COVID-19 severity. However, in order to harness the full potential of multiomics investigations, great care must be placed on the study- and cohort design. We have previously shown that multiomics integration can increase the statistical power to classify small sub-groups of patients greatly, reducing the required group size to reach 95% power from $n=30$ for single omics to $n=6$ for 5–7-tuple omics integration. However, this level of statistical power was achieved in a rather homogeneous cohort (the Karolinska COSMIC cohort) resulting from very strict exclusion criteria in terms of co-morbidities, treatments, etc. [20]. In contrast, many of the COVID-19 studies discussed in this review were limited to 10–50 subjects, with 3 or fewer omics platforms included. In order to reach the full statistical potential in COVID-19 related multiomics integration studies, cohorts must either be designed strictly with specific sub-groups of patients and symptoms in mind, ideally collecting samples not only systemically but also at the site of injury, or with very large numbers. With these factors met, future breakthroughs may include identifying preventive and therapeutic targets, early diagnostic tools specifically for asymptomatic subjects and high-risk populations with medical conditions, effective predictive models of critical and severe outcomes, and the selection of drugs as well as further understanding comorbid conditions and the long-term complications and consequences of COVID-19.

Development of user-friendly tools and database

Many multiomics datasets are available for further analysis, including several web-based dataset summaries (<https://opendata.ncats.nih.gov/covid19/omics>), as well as analytical and visualization tools, which are friendly for non-bioinformaticians. These tools help to quickly apply results and make available original data to the clinical and research communities. Given the diversity and complexity of multiomics, we found many studies associated with several databases or resources, along with varying depths of procession. Well-processed data

in an integrated, standardized format will benefit the re-analysis and re-integration of data. Thus, we developed an online Shiny App termed 'Collection of multiomics datasets in COVID-19' (available via <https://zhougroup.shinyapps.io/moCOVID/>) to aid researchers and clinicians in searching COVID-related multiomics investigations (Table 1 and Supplementary Table 1), as well as a corresponding downloadable database and online tools.

Conclusions

Like many other complex diseases, COVID-19 is an umbrella diagnosis with patient outcomes ranging from asymptomatic, to acute respiratory distress syndromes and cytokine storms requiring hospitalization and ICU treatment, to the debilitating long term symptoms of fatigue, vagal dysfunction, brain fog, postural orthostatic tachycardia syndrome (POTS), etc., experienced by patients with PACS. Identification of the molecular characteristics of this broad range of COVID-19 patients, powered by multiomics integration studies could offer a tremendous potential to unravel SARS-CoV-2 pathobiology, disease progression and resolution, predicting outcomes, prevention and therapy. Many lessons learned from multiomics integration in like various cancer types and respiratory disease can potentially transfer to COVID-19 research [20, 30, 66, 67]. New multiomics integration approaches, including multilevel virus–host interactions, handling of longitudinal data and dynamic models is needed. User-friendly, open access software programs and visualization tools are also desirable to increase accessibility of integration strategies the to a broader researcher base, beyond highly trained bioinformaticians, to facilitate efficient implementation of these results to clinical and biological research.

Key Points

- An increased and comprehensive understanding of COVID-19 characteristics is crucial to guiding the pandemic response.
- Identification of the molecular characteristics of COVID-19 patients from multiomics integration studies offers tremendous potential to unravel the puzzles related to SARS-CoV-2 pathobiology and COVID-19 progression.
- Multiomics integration-based molecular characterizations of COVID-19 could answer multiple important questions, specifically those related to disease severity, outcome prediction, prevention and therapy.
- To date, more than ten types of omics data and five categories of multiomics integration approaches have been applied to COVID-19 research, with further research potential and challenges in the biological and bioinformatics areas.

Supplementary data

Supplementary data are available online at <https://academic.oup.com/bib>.

Acknowledgments

We extend our thanks to Siqi Bao, Ke Li and Congcong Yan from the School of Biomedical Engineering, Wenzhou Medical University, Wenzhou, China, for their careful review of all omics datasets. We also extend our gratitude to Vanessa L Fuller, from Language Services at the University of Helsinki (Finland), for assistance with the initial language revision of this manuscript.

Funding

The Swedish Heart–Lung Foundation (to C.X.L., J.G. and Å.M.W.), the Swedish Research Council (to Å.M.W.), the Swedish Asthma and Allergy Foundation (to J.G.), the Sigrid Jusélius Foundation, and the Väinö and Laina Kivi Foundation; the Major Science and Technology Projects of Gansu Province (China; no. 20ZD2FA001 and 20YF2FA008 to X.L.) and Lanzhou Municipal Chenguan District Science and Technology Program (China; no. 2020JSCX0019).

References

1. Chen YM, Zheng Y, Yu Y, et al. Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *EMBO J* 2020;**39**:e105896.
2. NIH. <https://www.niaid.nih.gov/diseases-conditions/coronaviruses>.
3. Petrosillo N, Viceconte G, Ergonul O, et al. COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect* 2020;**26**: 729–34.
4. Thomas T, Stefanoni D, Dzieciatkowska M, et al. Evidence for structural protein damage and membrane lipid remodeling in red blood cells from COVID-19 patients. *J Proteome Res* 2020;**19**: 4455–69.
5. Mohamadian M, Chiti H, Shoghli A, et al. COVID-19: virology, biology and novel laboratory diagnosis. *J Gene Med* 2021;**23**:e3303.
6. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;**17**:181–92.
7. Coronavirus Resource Center JHUM. <https://coronavirus.jhu.edu/map.html>.
8. Habas K, Nganwuchu C, Shahzad F, et al. Resolution of coronavirus disease 2019 (COVID-19). *Expert Rev Anti Infect Ther* 2020;**18**:1201–11.
9. Salamanna F, Veronesi F, Martini L, et al. Post-COVID-19 syndrome: the persistent symptoms at the post-viral stage of the disease. A systematic review of the current data. *Front Med (Lausanne)* 2021;**8**:653516.
10. O'Driscoll M, Ribeiro Dos Santos G, Wang L, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 2021;**590**:140–5.
11. Gao Z, Xu Y, Sun C, et al. A systematic review of asymptomatic infections with COVID-19. *J Microbiol Immunol Infect* 2021;**54**:12–6.
12. Gao Y, Shi C, Chen Y, et al. A cluster of the corona virus disease 2019 caused by incubation period transmission in Wuxi, China. *J Infect* 2020;**80**:666–70.

13. Meselson M. Droplets and aerosols in the transmission of SARS-CoV-2. *N Engl J Med* 2020;**382**:2063.
14. Thomas T, Stefanoni D, Dzieciatkowska M, et al. Evidence of structural protein damage and membrane lipid Remodeling in red blood cells from COVID-19 patients. *J Proteome Res* 2020;**19**:4455–69.
15. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;**382**:1708–20.
16. Li G, Liu Y, Jing X, et al. Mortality risk of COVID-19 in elderly males with comorbidities: a multi-country study. *Aging (Albany NY)* 2020;**13**:27–60.
17. Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 2021;**19**:141–54.
18. Overmyer KA, Shishkova E, Miller JJ, et al. Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst* 2021;**12**:23–40.e7.
19. Zhang X, Tan Y, Ling Y, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* 2020;**583**:437–40.
20. Li CX, Wheelock CE, Skold CM, et al. Integration of multi-omics datasets enables molecular classification of COPD. *Eur Respir J* 2018;**51**:1701930.
21. Barh D, Aljabali AA, Tambuwala MM, et al. Predicting COVID-19-comorbidity pathway crosstalk-based targets and drugs: towards personalized COVID-19 management. *Biomedicine* 2021;**9**:556.
22. Galbraith MD, Kinning KT, Sullivan KD, et al. Seroconversion stages COVID19 into distinct pathophysiological states. *Elife* 2021;**10**:e65508.
23. Nalbandian A, Sehgal K, Gupta A, et al. Post-acute COVID-19 syndrome. *Nat Med* 2021;**27**:601–15.
24. Ortona E, Malorni W. Long COVID: to investigate immunological mechanisms and sex/gender related aspects as fundamental steps for a tailored therapy. *Eur Respir J* 2021;**58**:2102245. <https://doi.org/10.1183/13993003.02245-2021>.
25. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020;**323**:1239–42.
26. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;**395**:507–13.
27. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;**395**:497–506.
28. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020;**323**:1061–9.
29. Adorni F, Prinelli F, Bianchi F, et al. Self-reported symptoms of SARS-CoV-2 infection in a nonhospitalized population in Italy: cross-sectional study of the EPICoVID19 web-based survey. *JMIR Public Health Surveill* 2020;**6**:e21866.
30. Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 2020;**14**:117793221989905.
31. Badhwar A, McFall GP, Sapkota S, et al. A multiomics approach to heterogeneity in Alzheimer's disease: focused review and roadmap. *Brain* 2020;**143**:1315–31.
32. Gupta V, Haider S, Verma M, et al. Comparative genomics and integrated network approach unveiled undirected phylogeny patterns, co-mutational hot spots, functional cross talk, and regulatory interactions in SARS-CoV-2. *mSystems* 2021;**6**(1):e00030–21.
33. Young BE, Fong SW, Chan YH, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* 2020;**396**:603–11.
34. van Dorp L, Richard D, Tan CCS, et al. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* 2020;**11**:5986.
35. Islam A, Khan MA, Ahmed R, et al. Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2. *J Transl Med* 2021;**19**:32.
36. Sun C, Sun Y, Wu P, et al. Longitudinal multi-omics transition associated with fatality in critically ill COVID-19 patients. *Intensive Care Med Exp* 2021;**9**:13.
37. Su Y, Chen D, Yuan D, et al. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* 2020;**183**:1479–1495 e1420.
38. Barh D, Tiwari S, Weener ME, et al. Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19. *Comput Biol Med* 2020;**126**:104051.
39. Bruzzone C, Bizkarguenaga M, Gil-Redondo R, et al. SARS-CoV-2 infection dysregulates the Metabolomic and Lipidomic profiles of serum. *iScience* 2020;**23**:101645.
40. Stukalov A, Girault V, Grass V, et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* 2021;**594**:246–52.
41. Ahmed S, Paramasivam P, Raj K, et al. Regulatory cross talk between SARS-CoV-2 receptor binding and replication machinery in the human host. *Front Physiol* 2020;**11**:802.
42. Alimohamadi Y, Sepandi M, Taghdir M, et al. Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. *J Prev Med Hyg* 2020;**61**:E304–12.
43. Vakili K, Fathi M, Pezeshgi A, et al. Critical complications of COVID-19: a descriptive meta-analysis study. *Rev Cardiovasc Med* 2020;**21**:433–42.
44. Mizrahi B, Shilo S, Rossman H, et al. Longitudinal symptom dynamics of COVID-19 infection. *Nat Commun* 2020;**11**:6208.
45. Tay MZ, Poh CM, Renia L, et al. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* 2020;**20**:363–74.
46. Barh D, Tiwari S, Andrade BS, et al. A novel multi-omics-based highly accurate prediction of symptoms, comorbid conditions, and possible long-term complications of COVID-19. *Mol Omics* 2021;**17**:317–37.
47. Stephenson E, Reynolds G, Botting RA, et al. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med* 2021;**27**:904–16.
48. Wolff D, Nee S, Hickey NS, et al. Risk factors for Covid-19 severity and fatality: a structured literature review. *Infection* 2021;**49**:15–28.
49. Bernardes JP, Mishra N, Tran F, et al. Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and Plasmablasts as hallmarks of severe COVID-19. *Immunity* 2020;**53**:1296–1314.e9.
50. McReynolds CB, Cortes-Puch I, Ravindran R, et al. Plasma linoleate diols are potential biomarkers for severe COVID-19 infections. *Front Physiol* 2021;**12**:663869.
51. Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its impact on patients with COVID-19. *SN Compr Clin Med* 2020;**2**:1069–76.
52. Ng WH, Tipih T, Makoah NA, et al. Comorbidities in SARS-CoV-2 patients: a systematic review and meta-analysis. *MBio* 2021;**12**(1):e03647–20. <https://doi.org/10.1128/mBio.03647-20>.

53. Radwan NM, Mahmoud NE, Alfaifi AH, et al. Comorbidities and severity of coronavirus disease 2019 patients. *Saudi Med J* 2020;**41**:1165–74.
54. Zhou Y, Hou Y, Shen J, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLoS Biol* 2020;**18**:e3000970.
55. Sciacchitano S, De Vitis C, D'Ascanio M, et al. Gene signature and immune cell profiling by high-dimensional, single-cell analysis in COVID-19 patients, presenting low T3 syndrome and coexistent hematological malignancies. *J Transl Med* 2021;**19**:139.
56. Liu C, Martins AJ, Lau WW, et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* 2021;**184**:1836–1857.e22.
57. Tomazou M, Bourdakou MM, Minadakis G, et al. Multi-omics data integration and network-based analysis drives a multiplex drug repurposing approach to a shortlist of candidate drugs against COVID-19. *Brief Bioinform* 2021;**22**. <https://doi.org/10.1093/bib/bbab114>.
58. Chen T, Philip M, Le Cao KA, et al. A multi-modal data harmonisation approach for discovery of COVID-19 drug targets. *Brief Bioinform* 2021;**22**. <https://doi.org/10.1093/bib/bbab185>.
59. Yang R, Liu H, Bai C, et al. Chemical composition and pharmacological mechanism of Qingfei Paidu decoction and ma Xing Shi Gan decoction against coronavirus disease 2019 (COVID-19): in silico and experimental study. *Pharmacol Res* 2020;**157**:104820.
60. Muthuramalingam P, Jayasri R, Valliammai A, et al. Global multi-omics and systems pharmacological strategy unravel the multi-targeted therapeutic potential of natural bioactive molecules against COVID-19: an in silico approach. *Genomics* 2020;**112**:4486–504.
61. Kamal M, Abo Omirah M, Hussein A, et al. Assessment and characterisation of post-COVID-19 manifestations. *Int J Clin Pract* 2021;**75**:e13746.
62. Shah W, Hillman T, Playford ED, et al. Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. *BMJ* 2021;**372**:n136.
63. Yelin D, Margalit I, Yahav D, et al. Long COVID-19-it's not over until? *Clin Microbiol Infect* 2021;**27**:506–8.
64. Yelin D, Wirthheim E, Vetter P, et al. Long-term consequences of COVID-19: research needs. *Lancet Infect Dis* 2020;**20**:1115–7.
65. Lacobucci G. Long covid: damage to multiple organs presents in young lrp. *BMJ* 2020;**371**:m4470.
66. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet* 2017;**8**:84.
67. Sathyanarayanan A, Gupta R, Thompson EW, et al. A comparative study of multi-omics integration tools for cancer driver gene identification and tumour subtyping. *Brief Bioinform* 2020;**21**:1920–36.
68. Le Cao KA, Gonzalez I, Dejean S. integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 2009;**25**:2855–6.
69. Song JW, Lam SM, Fan X, et al. Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. *Cell Metab* 2020;**32**:188–202.e5.
70. Shen B, Yi X, Sun Y, et al. Proteomic and Metabolomic characterization of COVID-19 patient sera. *Cell* 2020;**182**:59–72.e15.
71. Terracciano R, Preiano M, Fregola A, et al. Mapping the SARS-CoV-2-host protein-protein interactome by affinity purification mass spectrometry and proximity-dependent biotin labeling: a rational and straightforward route to discover host-directed anti-SARS-CoV-2 therapeutics. *Int J Mol Sci* 2021;**22**:532.
72. Shirvaliloo M. Epigenomics in COVID-19; the link between DNA methylation, histone modifications and SARS-CoV-2 infection. *Epigenomics* 2021;**13**:745–50.
73. Zhao Y, Shang Y, Ren Y, et al. Omics study reveals abnormal alterations of breastmilk proteins and metabolites in puerperant women with COVID-19. *Signal Transduct Target Ther* 2020;**5**:247.
74. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;**11**:333–7.
75. Shen R, Mo Q, Schultz N, et al. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One* 2012;**7**:e35236.
76. Grun D, van Oudenaarden A. Design and analysis of single-cell sequencing experiments. *Cell* 2015;**163**:799–810.
77. Huo L, Jiao Li J, Chen L, et al. Single-cell multi-omics sequencing: application trends, COVID-19, data analysis issues and prospects. *Brief Bioinform* 2021;**22**. <https://doi.org/10.1093/bib/bbab229>.
78. Barberis E, Timo S, Amede E, et al. Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2. *Int J Mol Sci* 2020;**21**:8623.
79. Chen XJ, Li K, Xu L, et al. Novel insight from the first lung transplant of a COVID-19 patient. *Eur J Clin Invest* 2021;**51**:e13443.
80. Chen Y, Yao H, Zhang N, et al. Proteomic analysis identifies prolonged disturbances in pathways related to cholesterol metabolism and myocardium function in the COVID-19 recovery stage. *J Proteome Res* 2021;**20**:3463–74.
81. OpenData Portal of National Center for Advancing Translational Sciences, <https://opendata.ncats.nih.gov/covid19/omics>.