



OPEN ACCESS

EDITED BY
Anis Rageh Al-Maleki,
University of Malaya, Malaysia

REVIEWED BY
Juan F. González,
Nationwide Children's Hospital,
United States

*CORRESPONDENCE
Niyaz Ahmed
niyaz.ahmed@uohyd.ac.in

†These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Clinical Microbiology,
a section of the journal
Frontiers in Cellular and
Infection Microbiology

RECEIVED 04 October 2022
ACCEPTED 26 October 2022
PUBLISHED 21 November 2022

CITATION
Singh A, Ambaru B, Bandsode V and
Ahmed N (2022) Panomics to decode
virulence and fitness in
Gram-negative bacteria.
Front. Cell. Infect. Microbiol.
12:1061596.
doi: 10.3389/fcimb.2022.1061596

COPYRIGHT
© 2022 Singh, Ambaru, Bandsode and
Ahmed. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Panomics to decode virulence and fitness in Gram-negative bacteria

Anuradha Singh[†], Bindu Ambaru[†], Viraj Bandsode
and Niyaz Ahmed*

Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, University of
Hyderabad, Hyderabad, India

KEYWORDS

Gram-negative bacteria, antimicrobial resistance, multi-omics, integrative data
analytics, host-pathogen interactions

Introduction

Over the millennia, Gram-negative bacteria (GNB) have evolved to become one of the leading causes of fatalities across the globe. These bacterial species range from colonizers of the mammalian gut to pathogenic clones, often implicated in foodborne outbreaks and hospital-associated infections (HAIs) (Janda and Abbott, 2021; Ruegsegger et al., 2022). Eradication of these pathogens is further challenged by the emergence of multidrug-resistant (MDR) phenotypes and lack of novel drugs in the discovery pipeline (Laxminarayan et al., 2016). Recent estimates show that bacterial antimicrobial resistance (AMR) was responsible for 4.95 million recorded death cases in 2019, ranking third among all other global disease burdens (GBD) (Murray et al., 2022). It is noteworthy that six of the twelve pathogens mentioned by the WHO (<https://www.who.int/initiatives/glass/glass-routine-data-surveillance>; Veeraraghavan and Walia, 2019) were GNB, highlighting the need for a deeper understanding of the likely molecular mechanisms behind the propensity and fitness of these pathogens including their interactions with hosts.

With leaping whole genome sequencing (WGS) data, understanding of the molecular and genetic mechanisms underlying the evolution of bacterial pathogens from

Abbreviations: GNB, Gram-negative bacteria; HAIs, Hospital-Associated infections; MDR, Multidrug-Resistant; AMR, Antimicrobial Resistance; GBD, Global Disease Burdens; WHO, World Health Organization; WGS, Whole Genome Sequencing; HGT, Horizontal Gene Transfer; MGEs, Mobile Genetic Elements; GIs, Genomic Islands; ST, Sequence Type; *pks*, Polyketide Synthase; EHEC, Enterohemorrhagic *E. coli*; LGI, Lectin-Glycan Interaction network; GEMs, Genome-Scale Metabolic models; XDR, Extremely Drug-Resistant; LPS, Lipopolysaccharide; CCA, Canonical Clustering Analysis; MOMA, Multi-Omics Model and Analytics.

commensals to pathogens has considerably improved over the past two decades. Acquisition of genetic variation through horizontal gene transfer (HGT) and genome reduction are two major events responsible for bacterial evolution and colonization in diverse host and environmental contexts (Ahmed et al., 2008). Amongst GNBs such as *Escherichia coli*, *Salmonella* spp., *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Acinetobacter baumannii*, HGT-mediated acquisition of mobile genetic elements (MGEs) such as plasmids, phages and, genomic islands (GIs) remains the predominant mechanism of genome evolution (Hawken and Snitkin, 2019). These MGEs harbor genes encoding virulence factors and AMR to avert host-defense mechanisms and environmental vulnerabilities thus providing a survival advantage. The evolution of *E. coli* strains into diverse sequence types (STs) such as ST73, ST131, and ST95 provides substantial evidence regarding the role of HGT in genomic fine-tuning and pathogenicity (Forde et al., 2019; Shaik et al., 2022). High-throughput computational studies from our previous work (Suresh et al., 2021) demonstrated HGT-mediated dissemination of polyketide synthase (*pks*) island across different STs and serotypes of *E. coli*, which is often implicated in colorectal malignancies. Genome reduction is another major evolutionary force observed among Mycobacteria. Compared to other Gram-negative pathogens and host generalist species of *Salmonella*, pseudogenization-mediated metabolic fine-tuning in the immediate host niche appears to be a predominant mechanism of the genome evolution in *Salmonella* Typhi strains (Baddam et al., 2014).

Given that complex adaptive processes exhibited by the bacteria operate in a network of interactions spanning several molecular layers, the response of an entire cellular system to a given perturbation cannot be adequately captured from a single layer. Deep and accurate knowledge to develop a holistic molecular perspective of a biological system requires not just one but several omics analyses. Heterogeneous datasets derived from different omics platforms such as genomics, transcriptomics, proteomics, metabolomics, metagenomics, meta-transcriptomics, meta-proteomics, and meta-metabolomics may complement each other and offer an attractive approach to understand the organisms as well as their interactions with corresponding hosts. Hence, the aim of this review is to discuss and build a narrative on multi-omics/panomics research on Gram-negative priority pathogens and to emphasize upon the need to harness integrated omics analyses to comprehend and control life-threatening infections.

Multi-omics analyses to decode intricate biological processes of Gram-negative bacteria

The current state of art elucidates several omics methodologies and interdisciplinary approaches which have

surpassed the traditional ones. This section highlights some of the studies that have used two or more omics layers to shed light on the GNB's dynamic biological processes. We discuss them taking examples from some of the known priority pathogens.

Escherichia coli

Humans are susceptible to a wide range of intestinal and extraintestinal diseases and infections caused by *E. coli*. The enteric *E. coli* are divided into different pathotypes such as enteropathogenic, enterotoxigenic, enteroinvasive, enterohemorrhagic, enteroaggregative, and diffusely adherent based on their virulence traits (Kaper et al., 2004; Ahmed et al., 2008). Antibiotic-resistant *E. coli* is commonly employed as a model organism in structural and functional investigations to comprehend the physiology and gene expression of MDR bacteria. By using an integrated multi-omics approach that includes the genomic, transcriptomic, and proteomic data of enterohemorrhagic *E. coli* (EHEC) EDL933, Cho and colleagues investigated the interactions between host mucin and pathogen proteins, providing a valuable resource for the creation of Lectin-Glycan Interaction Network (LGI) of *E. coli* (Cho et al., 2020). Extracting critical phenotypic alterations responsible for drug resistances was made possible by the integration of transcriptomics and genomics data (Suzuki et al., 2014). Comparative genomics, transcriptomics, and functional characterization (Hazen et al., 2017) demonstrated that hybrid-pathogenic strains of *E. coli* are capable of expressing the virulence genes from various pathovars. Some of the transcriptomics and fluxomics studies (Fong et al., 2006) enabled new insights into the evolutionary dynamics of *E. coli* by demonstrating the flexibility of the metabolic network to counteract genetic perturbations and also emphasized the advantage of combining multiple omics datasets to differentiate between causal and noncausal mechanistic changes. Another important work (Piazza et al., 2018) predicted a network of interactions and binding sites in *E. coli* using a metabolomics and proteomics approach, thus allowing the discovery of novel enzyme-substrate interactions. Genome-scale metabolic models (GEMs) of *E. coli* B and *E. coli* K12 constructed by integrating the comparative analyses of genomes, transcriptomes, proteomes, and phenomes provided the basis for differentiating the two strains. Similar studies providing insights into cellular physiology and metabolism could be relevant for engineering microorganisms for bioprocess applications as well as towards understanding the virulence mechanisms of various pathogens (Yoon et al., 2012).

Salmonella spp.

Salmonella's rising antibacterial resistance and the lack of novel antimicrobials on the horizon are being addressed via multi-omics studies. Proteomics, metabolomics, glycomics, and metagenomics were used in a multi-omics 'systems' approach

(Deatherage Kaiser et al., 2013) to investigate the molecular interactions between *Salmonella enterica* serovar Typhimurium (S. Typhimurium), the murine host, and the microbiome during intestinal infection with S. Typhimurium. Proteogenomics was employed recently (Karash et al., 2017) to identify the potential genes and proteins that play a role in S. Typhimurium's resistance to H₂O₂, thus deepening the current understanding of S. Typhimurium's survival mechanisms in macrophages. Another research (Crouse et al., 2020) focused on integrating WGS techniques into food safety practices could establish links between virulence and genetic diversity in *Salmonella*. They have also presented a novel approach for risk assessment of particular strains as well as for improved monitoring and source tracking during outbreaks. By utilizing metabolomics and transcriptomics, it has been possible to understand that both glycolysis and lipid metabolism were regulated by SlyA in *Salmonella* (Tian et al., 2021). Another study based on high throughput analyses (Hossain et al., 2017) harnessed the advantage of genomics, gene expression analysis, proteomics, metabolic pathways, and subcellular localization to discover 52 distinct essential proteins in the target proteome of the S. *enterica* that could be used as novel targets to develop newer drugs. Utilizing metabolomics and transcriptomics, it was possible to assess adaptation of S. Typhimurium to essential oils (thyme and cinnamon) and to study the induced resistance as well as the underlying adaptive mechanisms (Chen et al., 2022). Recently, a promising therapeutic target that activates immune response against the extremely drug-resistant (XDR) strain called S. Typhi H58 has been successfully identified using a comprehensive strategy of computational reverse vaccination along with subtractive genomics (Khan et al., 2022).

Klebsiella pneumoniae

It is challenging to treat infections caused due to MDR and highly virulent *K. pneumoniae* strains, highlighting the urgent need to discover novel and effective therapeutics against this pathogen. This was addressed (Ramos et al., 2018) by integrating various multi-omics data like genomics, transcriptomics, metabolomics, and protein structure information to delineate 29 proteins with preferential properties for therapeutic development against *Klebsiella*. This work also provided insights into *K. pneumoniae* metabolism under various host-imitating circumstances. Recently, a gene and metabolite-centric network-based method (Cesur et al., 2019) identified potential therapeutic targets for *K. pneumoniae*, MGH 78578. A thorough assessment of the identification of pharmacological targets and their implications in the therapeutic management of *Klebsiella* infections was presented (Ali et al., 2022) using a multi-omics perspective.

Pseudomonas aeruginosa

AMR nosocomial pathogen *P. aeruginosa* is currently posing unwavering and increasing threats to humans. Grady and colleagues (Grady et al., 2017) integrated the results from studies including RNA-Seq, proteomics, ribosome footprinting, and small molecule LC-MS, to compare the gene expression of *P. aeruginosa*. Collectively, their findings unleash the mechanisms underlying the bacteria's ability to grow and survive on n-alkanes. Integrated analysis of transcriptomics and metabolomics revealed that polymyxin therapy significantly altered lipid, lipopolysaccharide, and peptidoglycan biosynthesis as well as central carbon metabolism and oxidative stress (Han et al., 2019). This study also demonstrated the systems-level dynamics of polymyxin-induced cellular responses, highlighting the need for combination therapy to reduce resistance to the last-resort therapeutic option, polymyxins. Further, it was possible (Filho et al., 2021) to integrate transcriptome data with genome-scale metabolic networks of *P. aeruginosa* to identify potential therapeutic targets. Rashid and colleagues (Rashid et al., 2017) used a comprehensive subtractive genome and proteome computational framework in their investigation to predict potential *P. aeruginosa* vaccine candidates. Recently, a multi-omics based investigation (Molina-Mora and García, 2021) incorporating genomics, phenomics, comparative genomics, transcriptomics, and proteomics provided new insights about molecular determinants of antibiotic resistance in a MDR strain of *P. aeruginosa* (PaeAG1).

Acinetobacter baumannii

A. baumannii's exceptional propensity to quickly acquire resistance determinants to a wide range of antibiotics has made it a significant global cause of HAIs. Understanding the pathophysiology and evolution of AMR can help us fight illnesses caused by *A. baumannii*. Clinical isolates of *A. baumannii* have been reported to be resistant to triclosan (Chen et al., 2009). A multi-omics investigation employing WGS, transcriptomics, and proteomics was carried out to better understand the global alterations in protein expression in the triclosan-resistant mutant strain, AB042 to understand the mechanisms of resistance (Fernando et al., 2017). According to their findings, *A. baumannii* reacts to triclosan by changing the expression of genes related to amino acid and fatty acid metabolism, and AMR. The colistin resistance mechanism in MDR-ZJ06, an MDR clinical strain of *A. baumannii*, was elucidated (Hua et al., 2017) by combining genomics, transcriptomics, and proteomics. The loss of bacterial lipopolysaccharide (LPS) caused by ISAbal insertion in lpxC

was identified in their investigation as the resistance mechanism of the colistin-resistant strain. Through the integration of various data sources, including the co-expression, operon organization, and associated protein structural data of genes in *A. baumannii* (Xie et al., 2020), a co-functional network was built with potential AMR and virulence related features.

Public data sharing is an essential component of research to fight against pathogens. The growing accessibility of microbial omics data combined with heterogeneous metadata is revolutionizing the study of infectious diseases and numerous resources are being created to organize such enormous amounts of data. The prominent microbiological databases that incorporate multi-omics and multi-(meta) omics datasets as well as the specialized databases that focus on a particular GNB are listed in Table 1. In addition, it also includes technologies/resources available for data integration. A more

comprehensive list can be found in the database-focused annual edition of Nucleic Acids Research (Rigden and Fernández, 2022).

Holo-omics approach for deciphering the host-pathogen interactions

Extremely intricate interactions exist between microorganisms and host cells, and these interactions are not always uniform or linear in nature. Pathogens alter the primary metabolic processes in themselves as well as in the host cell based on the nutrient sources prevailing in the infected host niche. This necessitates a comprehensive strategy that can take into account the various

TABLE 1 Prominent microbiological databases and resources available for data integration.

Database	Omics data types	Functionality	Organisms	URL	Reference
Bacterial and Viral Bioinformatics Resource Centre (BV-BRC)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics	Provides access to a variety of data for the National institute of Allergy and Infectious Diseases (NIAID's) priority pathogens.	NIAID category A to C/ emerging/ reemerging pathogens	https://www.bv-brc.org/	(Davis et al., 2020)
Omics Discovery Index (Omicsdi)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics	An open-source platform that provides access, discovery and dissemination of omics data sets. Currently, 11 different repositories, hosted on 4 continents are included in this database.	Various microorganisms	http://www.omicsdi.org	(Perez-Riverol et al., 2017)
National Centre for Biotechnology Information (NCBI)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics Metatranscriptomics Metaproteomics	NCBI establishes standards for data deposition and exchange for the scientific and medical sectors, as well as access to a variety of databases and tools. NCBI provides access to a variety of databases and software, promotes standards for data deposition and exchange for the scientific and medical communities	Various microorganisms	https://www.ncbi.nlm.nih.gov/	(Schuler et al., 1996)
European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics Metatranscriptomics Metaproteomics	Provides bioinformatics resources for promoting research, and disseminates cutting-edge technologies to the academic community and industry.	Various microorganisms	https://www.ebi.ac.uk/ https://elixir-europe.org/	(Kanz et al., 2005)
European Life-Science infrastructure (ELIXIR)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics Metatranscriptomics Metaproteomics	Provides services for depositing and retrieving sequencing data, software tools for analyzing biological data	Various microorganisms	https://www.ddbj.nig.ac.jp/index-e.html	(Tateno et al., 2002)
DNA Databank of Japan (DDBJ)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics Metatranscriptomics Metaproteomics	Provides services for depositing and retrieving sequencing data, software tools for analyzing biological data	Various microorganisms	https://www.ddbj.nig.ac.jp/index-e.html	(Tateno et al., 2002)
China National GenBank (CNGB)	Genomics Transcriptomics Proteomics	A unified platform created for the research community's application services and sharing of biological big data.	Various microorganisms	https://db.cngb.org/	(Guo et al., 2020)

(Continued)

TABLE 1 Continued

Database	Omics data types	Functionality	Organisms	URL	Reference
	Metabolomics Metagenomics Metatranscriptomics Metaproteomics				
Joint Genomic Institute Integrated Microbial genomes (JGI-IMG)	Genomics Transcriptomics Proteomics Metabolomics Metagenomics Metatranscriptomics Metaproteomics	Supports the annotation, analysis and distribution of microbial genome and microbiome datasets.	Various microorganisms	https://img.jgi.doe.gov/	(Markowitz et al., 2014)
Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST)	Genomics Transcriptomics Metagenomics Metatranscriptomics	A Metagenomic analysis server for microbial communities	Various microorganisms	https://www.mg-rast.org/	(Keegan et al., 2016)
BioCYC	Genomics Transcriptomics Proteomics Metabolomics Regulatory networks	The tools for omics data analysis, comparative genomes, and comparative pathway analysis are all provided by BioCyc. It has distinct home pages for the different organisms like EcoCyc.org for <i>E. coli</i> ; Salmonella.biocyc.org for <i>Salmonella</i> spp	Various microorganisms	https://biocyc.org/	(Karp et al., 2019; Keseler et al., 2021)
Ecomics	Transcriptomics Proteomics Metabolomics Fluxomics Phenomics	Multi-omics compendium for <i>E. coli</i>	<i>E. coli</i>	http://prokaryomics.com/	(Kim et al., 2016)
SalmoNet	Protein-protein interactions, transcriptional regulatory interactions and enzyme-enzyme interactions	A molecular interaction database providing network resource containing regulatory, metabolic and protein-protein interactions	<i>Salmonella</i>	http://salmonet.org/	(Métris et al., 2017)
SYSTEMONAS	Genomics Transcriptomics Proteomics Metabolome Gene regulatory network	A source for <i>Pseudomonas</i> systems biology analysis	<i>Pseudomonas</i>	http://www.systemonas.de .	(Choi et al., 2007)
Klebnet	Genomics	A platform for genomic surveillance with analytics, specifically designed for the complex of <i>K. pneumoniae</i> species	<i>Klebsiella</i>	https://klebnet.org/	(Lee et al., 2019)
ABviresDB	Resistance or virulence features as well as their co-functional interactions.	ABviresDB will be useful in revealing the mechanisms of bacterial resistance and virulence and for the network study of bacterial infection	<i>Acinetobacter</i>	https://acba.shinyapps.io/ABviresDB/	(Xie et al., 2020)
Resources for omics data integration					
Omics integrator	Genomics Transcriptomics Proteomics Metabolomics Pathway analysis Visualization	Holistic analysis of different types of omics datasets	Various Microorganisms	http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/	(Tuncbag et al., 2016)
Paintomics 3.0 (web based)	Genomics Proteomics Metabolomics Pathway analysis	A web application for visual representation of integrated view of several omic datasets	Various Microorganisms	https://www.paintomics.org/	(García-Alcalde et al., 2011)
integrOmics (R package)	Genomics Proteomics Metabolomics Microbiome	Two 'omics' variables that are measured on the same samples are effectively integrated by integrOmics.	Various Microorganisms	http://CRAN.R-project.org/	(Lê Cao et al., 2009)

(Continued)

TABLE 1 Continued

Database	Omics data types	Functionality	Organisms	URL	Reference
mixOmics (R package)	Genomics Transcriptomics Proteomics Metabolomics Visualization	Focused on data exploration, dimension reduction, and visualisation with a particular emphasis on multivariate analysis of biological data sets.	Various Microorganisms	http://www.bioconductor.org/packages/release/bioc/html/mixOmics.html	(Rohart et al., 2017)
ProteoClade (Python)	Genomics Proteomics Metaproteomics	Associate taxonomic studies of several species with proteomic data	Various Microorganisms	http://github.com/HeldLab/ProteoClade	(Mooradian et al., 2020)
Metaboanalyst 5.0	Proteomics Metabolomics Pathway analysis Visualaization	Analysis, interpretation, and integration of metabolomics data with other omics data	Various Microorganisms	https://www.metaboanalyst.ca/	(Pang et al., 2022)
Qiime2	Metagenomics Metatranscriptomics Metaproteomics Metabolomics	Open-source microbiome analysis tool that transforms unprocessed sequence data into understandable visualisations and statistical findings	Various Microorganisms	https://qiime2.org/	(Caporaso et al., 2010)

data types using the same inference framework so as to broaden the scope of investigations on microbes and hosts. ‘Holo-omics’ investigations incorporate information from many omic levels in the host and microbial domains (Nyholm et al., 2020). Expanding the scope of biological interpretation and examining biological pathways in greater detail are made possible by the ability to integrate many meta-omics levels, like meta-genomes, meta-transcriptomes, meta-proteomes, and meta-metabolomes. Epigenomic and exposomic profiling is made possible by similar technologies, and this can help to further disentangle the biochemical interactions between host-microbiota and environment and their impact on host phenotypes (Kumar et al., 2014; Rogler and Vavricka, 2015). GEMs offer a better comprehension of how intracellular infections make use of the host’s existing milieu. The host-cell nutritional environment and gene expression data from *S. Typhimurium* grown inside macrophage cell lines were used to study *Salmonella* metabolism during infection (Raghunathan et al., 2009). *Salmonella*’s metabolic changes proceeding from the early stages of infection until chronic infection was predicted by simulations of the GEM (iRR1083) (Raghunathan et al., 2009). Their data reveal occurrence of a minimal set of metabolic pathways that is necessary for *Salmonella* to successfully replicate inside the host cell. Additionally, this model provides a framework for the identification of networked metabolic pathways, incorporation of high-throughput data to produce hypotheses regarding metabolism during infection, and the logical development of new antibiotics. Another work (Ding et al., 2016) used *in silico* metabolic modeling to predict the crucial genes of enterobacterial human pathogens (*E. coli* and *Salmonella* strains) in different host habitats, including the human bloodstream, urinary tract, and macrophage for understanding the pathogen’s survival and infection mechanisms. It is possible to explore condition-specific pathogenicity by mapping multi-omics data to GEMs. Although the technology to produce

huge amounts of data for use in a holo-omics environment is currently available, the data integration methods to uncover and detect host-microbe interactions are still limited, thus opening new avenues towards applied research.

Way forward: Integrative data analytics

Data integration approaches broadly fall into two distinct categories depending upon the assumption as to whether the biological variation is unidirectional or multidirectional i.e., multi-staged analysis and meta-dimensional analysis (Ritchie et al., 2015; Jendoubi, 2021).

A multi-staged analysis refers to the integration of data in a hierarchical or stepwise manner wherein only two different data types are combined at once to investigate the relationship between them. In contrast, meta-dimensional analysis refers to simultaneous integration of multiple variables from different data types (Ritchie et al., 2015). Though meta-dimensional analysis is statistically more robust as compared to multi-staged analysis, it also increases the dimensionality of the data while combining many data types, making it more complex to interpret. The choice of data integration approach primarily depends on the aim of the study along with other factors such as sampling, omics platforms, and quality of the data (Graw et al., 2021). Recently, such a multi-dimensional approach has been used for drug target prioritization in MDR *K. pneumoniae* (Ramos et al., 2018).

Further, meta-dimensional analyses could be categorized into three different methods depending upon the stage of data integration i.e., concatenation-based (early integration), transformation-based (intermediate integration), and model-based (late integration) (Ritchie et al., 2015). In concatenation-

based methods, data gathered across various omics platforms could be combined to create a joint matrix that serves as an input dataset for machine learning algorithms. This approach has been used to study stress response in *E. coli* wherein a combined dataset of transcriptomics and metabolomics was used as an input for machine learning algorithms. K-means clustering and canonical clustering analysis (CCA) were used to understand the coordinated changes in transcripts and metabolites under different stress conditions (Jozefczuk et al., 2010).

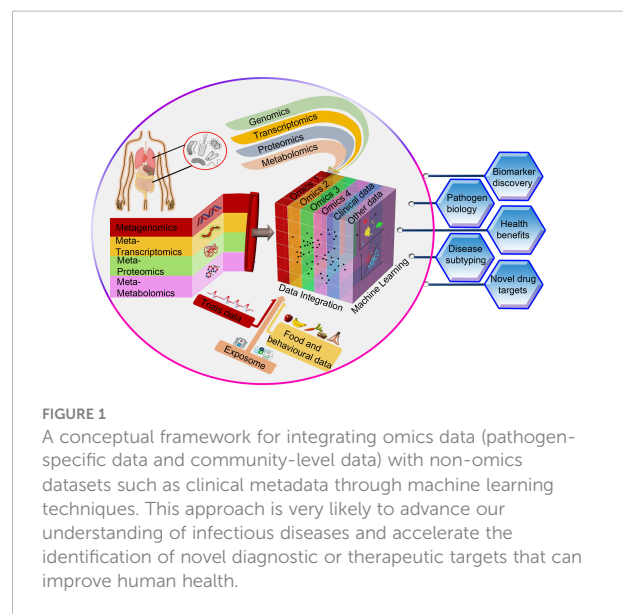
In transformation-based methods, data are first transformed into intermediate forms such as graph and kernel matrix followed by integration into a combined matrix and data analysis. As the data are transformed into intermediate form, this particular strategy of integration preserves the characteristics of each unique data type. Different machine learning frameworks have been developed to learn from transformed datasets. DeepDRK is one such deep learning model which involves kernel-based integration of multi-omics data to predict drug response of cancer cell lines (Wang et al., 2021).

In model-based approach, individual omics datasets are first used as training datasets to build respective models and finally, multiple models are integrated to mine biological processes. MOMA (Multi-Omics Model and Analytics) (Kim et al., 2016) is one such platform wherein model-based data integration was used to study the cellular states of *E. coli* under unexplored conditions.

Although multi-omics data integration techniques have lately gained popularity in a number of scientific domains, this area of study is still in its infancy in case of bacterial species. Given the exponential increase in multi-omics data, integrated analytics may prove to be one of the most effective methods to comprehend both the basic as well as stress physiology of bacteria. This strategy can assist biomedical researchers in discovering strain-specific biomarkers thereby elucidating cellular mechanisms of pathogenesis and developing novel therapeutic approaches.

Open challenges and future directions

The advantage of panomics data integration to get a holistic understanding of biological processes and infection mechanisms has its own inherent challenges. Multi-omics analyses present additional obstacles such as methods to be used for integration, clustering, visualization, and functional characterization on top of the difficulties that single-omics analyses entail (Pinu et al., 2019). For instance, researchers may encounter difficulties with data harmonization (data scaling, data normalization, and data transformation methods pertaining to individual omics datasets) prior to combining two or more omics datasets. Furthermore, the computational resources and storage space needs can be



prohibitive for a given study due to dimensionality limits when integrating huge datasets. Our ability to integrate pathogen-specific omics data, community-level omics data and the non-omics datasets such as clinical metadata will improve the understanding of infectious diseases and hasten the discovery of new diagnostic or therapeutic targets (Figure 1). Despite inevitable practical, financial, and computational challenges, the incorporation of various multi-omics data types from both the microbe and the host sides could revolutionize the understanding of infections caused by AMR bacteria. Given this, in-depth analyses of the disease coordinates, both at the levels of pathogens and hosts, would be beneficial in devising personalized treatments.

Author contributions

AS and BA contributed equally to the writing of the review. VB prepared table. NA conceptualized the opinionated content, and contributed to editing and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We thank the Indian Council of Medical Research (ICMR) for a grant awarded to NA (AMR/257/2021/ECD-II). We also acknowledge the University of Hyderabad for providing facilities and institutional support (Institution of Eminence -PDF; financial support to UoH-IoE by MHRD [F11/9/2019-U3A]). AS acknowledges the Junior Research fellowship from the Department

of Biotechnology, Government of India and the Prime Minister's Research Fellowship (PMRF) from the Indian Government.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ahmed, N., Dobrindt, U., Hacker, J., and Hasnain, S. E. (2008). Genomic fluidity and pathogenic bacteria: Applications in diagnostics, epidemiology and intervention. *Nat. Rev. Microbiol.* 6, 387–394. doi: 10.1038/nrmicro1889
- Ali, S., Alam, M., Hasan, G. M., and Hassan, M. I. (2022). Potential therapeutic targets of klebsiella pneumoniae: A multi-omics review perspective. *Brief. Funct. Genomics* 21, 63–77. doi: 10.1093/bfpg/elab038
- Baddam, R., Kumar, N., Shaik, S., Lankapalli, A. K., and Ahmed, N. (2014). Genome dynamics and evolution of salmonella typhi strains from the typhoid-endemic zones. *Sci. Rep.* 4, 7457. doi: 10.1038/srep07457
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2007). GenBank. *Nucleic Acids Res.* 35, D21–D25. doi: 10.1093/nar/gkl986
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Cesur, M. F., Siraj, B., Uddin, R., Durmuş, S., and Çakır, T. (2019). Network-based metabolism-centered screening of potential drug targets in at genome scale. *Front. Cell. Infect. Microbiol.* 9, 447. doi: 10.3389/fcimb.2019.00447
- Chen, Y., Pi, B., Zhou, H., Yu, Y., and Li, L. (2009). Triclosan resistance in clinical isolates of acinetobacter baumannii. *J. Med. Microbiol.* 58, 1086–1091. doi: 10.1099/jmm.0.008524-0
- Chen, L., Zhao, X., Li, R., and Yang, H. (2022). Integrated metabolomics and transcriptomics reveal the adaptive responses of salmonella enterica serovar typhimurium to thyme and cinnamon oils. *Food Res. Int.* 157, 111241. doi: 10.1016/j.foodres.2022.111241
- Choi, C., Münch, R., Leupold, S., Klein, J., Siegel, I., Thielen, B., et al. (2007). SYSTOMONAS—an integrated database for systems biology analysis of pseudomonas. *Nucleic Acids Res.* 35, D533–D537. doi: 10.1093/nar/gkl823
- Cho, S.-H., Lee, K. M., Kim, C.-H., and Kim, S. S. (2020). Construction of a lectin–glycan interaction network from enterohemorrhagic escherichia coli strains by multi-omics analysis. *Int. J. Mol. Sci.* 21, 2681. doi: 10.3390/ijms21082681
- Crouse, A., Schramm, C., Emond-Rheault, J.-G., Herod, A., Kerhoas, M., Rohde, J., et al. (2020). Combining whole-genome sequencing and multimodel phenotyping to identify genetic predictors of virulence. *mSphere* 5, e00293–20. doi: 10.1128/mSphere.00293-20
- Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., et al. (2020). The PATRIC bioinformatics resource center: Expanding data and analysis capabilities. *Nucleic Acids Res.* 48, D606–D612. doi: 10.1093/nar/gkz943
- Deatherage Kaiser, B. L., Li, J., Sanford, J. A., Kim, Y.-M., Kronewitter, S. R., Jones, M. B., et al. (2013). A multi-omic view of host-Pathogen-Commensal interplay in salmonella-mediated intestinal infection. *PLoS One* 8, e67155. doi: 10.1371/journal.pone.0067155
- Ding, T., Case, K. A., Omolo, M. A., Reiland, H. A., Metz, Z. P., Diao, X., et al. (2016). Predicting essential metabolic genome content of niche-specific enterobacterial human pathogens during simulation of host environments. *PLoS One* 11, e0149423. doi: 10.1371/journal.pone.0149423
- Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids* (Cambridge University Press). Available at: https://books.google.com/books/about/Biological_Sequence_Analysis.html?hl=&id=R5P2GJJvlgQC.
- Durinx, C., McEntyre, J., Appel, R., Apweiler, R., Barlow, M., Blomberg, N., et al. (2016). Identifying ELIXIR core data resources. *F1000Res* 5, ELIXIR-2422. doi: 10.12688/f1000research.9656.2
- Fernando, D. M., Chong, P., Singh, M., Spicer, V., Unger, M., Loewen, P. C., et al. (2017). Multi-omics approach to study global changes in a triclosan-resistant mutant strain of acinetobacter baumannii ATCC 17978. *Int. J. Antimicrobial Agents* 49, 74–80. doi: 10.1016/j.ijantimicag.2016.10.014
- Filho, F. M., do Nascimento, A. P. B., de Oliveira Cerqueira e Costa, M., Merigueti, T. C., de Menezes, M. A., Nicolás, M. F., et al. (2021). A systematic strategy to find potential therapeutic targets for pseudomonas aeruginosa using integrated computational models. *Front. Mol. Biosci.* 8, 728129. doi: 10.3389/fmolb.2021.728129
- Fong, S. S., Nanchen, A., Palsson, B. O., and Sauer, U. (2006). Latent pathway activation and increased pathway capacity enable escherichia coli adaptation to loss of key metabolic enzymes. *J. Biol. Chem.* 281, 8024–8033. doi: 10.1074/jbc.M510016200
- Forde, B. M., Roberts, L. W., Phan, M.-D., Peters, K. M., Fleming, B. A., Russell, C. W., et al. (2019). Population dynamics of an escherichia coli ST131 lineage during recurrent urinary tract infection. *Nat. Commun.* 10, 3643. doi: 10.1038/s41467-019-11571-5
- García-Alcalde, F., García-López, F., Dopazo, J., and Conesa, A. (2011). Paintomics: A web based tool for the joint visualization of transcriptomics and metabolomics data. *Bioinformatics* 27, 137–139. doi: 10.1093/bioinformatics/btq594
- Grady, S. L., Malfatti, S. A., Gunasekera, T. S., Dalley, B. K., Lyman, M. G., Striebig, R. C., et al. (2017). A comprehensive multi-omics approach uncovers adaptations for growth and survival of pseudomonas aeruginosa on n-alkanes. *BMC Genomics* 18, 334. doi: 10.1186/s12864-017-3708-4
- Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S.2nd, et al. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Mol. Omics* 17, 170–185. doi: 10.1039/d0mo00041h
- Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., et al. (2020). CNSA: a data repository for archiving omics data. *Database* 2020, baaa055. doi: 10.1093/database/baaa055
- Han, M.-L., Zhu, Y., Creek, D. J., Lin, Y.-W., Gutu, A. D., Hertzog, P., et al. (2019). Comparative metabolomics and transcriptomics reveal multiple pathways associated with polymyxin killing in pseudomonas aeruginosa. *mSystems* 4, e00149–18. doi: 10.1128/mSystems.00149-18
- Hawken, S. E., and Snitkin, E. S. (2019). Genomic epidemiology of multidrug-resistant gram-negative organisms. *Ann. N. Y. Acad. Sci.* 1435, 39–56. doi: 10.1111/nyas.13672
- Hazen, T. H., Michalski, J., Luo, Q., Shetty, A. C., Daugherty, S. C., Fleckenstein, J. M., et al. (2017). Comparative genomics and transcriptomics of escherichia coli isolates carrying virulence factors of both enteropathogenic and enterotoxigenic e. coli. *Sci. Rep.* 7, 3513. doi: 10.1038/s41598-017-03489-z
- Hossain, T., Kamruzzaman, M., Choudhury, T. Z., Mahmood, H. N., Nabi, A. H. M. N., and Hosen, M. I. (2017). Application of the subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against subsp. poona. *BioMed. Res. Int.* 2017, 3783714. doi: 10.1155/2017/3783714
- Hua, X., Liu, L., Fang, Y., Shi, Q., Li, X., Chen, Q., et al. (2017). Colistin resistance in acinetobacter baumannii MDR-ZJ06 revealed by a multiomics approach. *Front. Cell. Infect. Microbiol.* 7. doi: 10.3389/fcimb.2017.00045
- Janda, J. M., and Abbott, S. L. (2021). The changing face of the family (Order):: New members, taxonomic issues, geographic expansion, and new diseases and disease syndromes. *Clin. Microbiol. Rev.* 34, e00174–20. doi: 10.1128/CMR.00174-20

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Jendoubi, T. (2021). Approaches to integrating metabolomics and multi-omics data: A primer. *Metabolites* 11, 84. doi: 10.3390/metabo11030184
- Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhilber, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* 6, 364. doi: 10.1038/msb.2010.18
- Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., et al. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Res.* 33, D29–D33. doi: 10.1093/nar/gki098
- Kaper, J., Nataro, J., and Mobley, H. (2004). Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2, 123–140. doi: 10.1038/nrmicro818
- Karash, S., Liyanage, R., Qassab, A., Lay, J. O. J. R., and Kwon, Y. M. (2017). A comprehensive assessment of the genetic determinants in salmonella typhimurium for resistance to hydrogen peroxide using proteogenomics. *Sci. Rep.* 7, 17073. doi: 10.1038/s41598-017-17149-9
- Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 20, 1085–1093. doi: 10.1093/bib/bbx085
- Keegan, K. P., Glass, E. M., and Meyer, F. (2016). MG-RAST, a metagenomics service for analysis of microbial community structure and function. *Methods Mol. Biol.* 1399, 207–233. doi: 10.1007/978-1-4939-3369-3_13
- Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martinez, C., Caspi, R., et al. (2021). The EcoCyc database in 2021. *Front. Microbiol.* 12, 711077. doi: 10.3389/fmicb.2021.711077
- Khan, K., Jalal, K., and Uddin, R. (2022). An integrated in silico based subtractive genomics and reverse vaccinology approach for the identification of novel vaccine candidate and chimeric vaccine against XDR salmonella typhi H58. *Genomics* 114, 110301. doi: 10.1016/j.ygeno.2022.110301
- Kim, M., Rai, N., Zorraquino, V., and Tagkopoulos, I. (2016). Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7, 13090. doi: 10.1038/ncomms13090
- Kumar, H., Lund, R., Laiho, A., Lundelin, K., Ley, R. E., Isolauri, E., et al. (2014). Gut microbiota as an epigenetic regulator: Pilot study based on whole-genome methylation analysis. *MBio* 5, e02113–14. doi: 10.1128/mBio.02113-14
- Laxminarayan, R., Matsoso, P., Pant, S., Brower, C., Røttingen, J.-A., Klugman, K., et al. (2016). Access to effective antimicrobials: A worldwide challenge. *Lancet* 387, 168–175. doi: 10.1016/S0140-6736(15)00474-2
- Lê Cao, K.-A., González, I., and Déjean, S. (2009). integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* 25, 2855–2856. doi: 10.1093/bioinformatics/btp515
- Lee, M., Pinto, N. A., Kim, C. Y., Yang, S., D'Souza, R., Yong, D., et al. (2019). Network integrative genomic and transcriptomic analysis of carbapenem-resistant *Klebsiella pneumoniae* strains identifies genes for antibiotic resistance and virulence. *mSystems* 4, e00202–19. doi: 10.1128/mSystems.00202-19
- Markowitz, V. M., Chen, L.-M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, D568–D573. doi: 10.1093/nar/gkt919
- Métris, A., Sudhakar, P., Fazekas, D., Demeter, A., Ari, E., Olbei, M., et al. (2017). SalmoNet, an integrated network of ten strains reveals common and distinct pathways to host adaptation. *NPJ Syst. Biol. Appl.* 3, 31. doi: 10.1038/s41540-017-0034-z
- Molina-Mora, J. A., and García, F. (2021). Molecular determinants of antibiotic resistance in the Costa Rican AG1 by a multi-omics approach: A review of 10 years of study. *Phenomics* 1, 129–142. doi: 10.1007/s43657-021-00016-z
- Mooradian, A. D., van der Post, S., Naegle, K. M., and Held, J. M. (2020). ProteoClade: A taxonomic toolkit for multi-species and metaproteomic analysis. *PLoS Comput. Biol.* 16, e1007741. doi: 10.1371/journal.pcbi.1007741
- Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., and Gray, A. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* 399, 629–655. doi: 10.1016/S0140-6736(21)02724-0
- Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., et al. (2020). Holo-omics: Integrated host-microbiota multi-omics for basic and applied biological research. *iScience* 23, 101414. doi: 10.1016/j.isci.2020.101414
- Olbei, M., Bohar, B., Fazekas, D., Madgwick, M., Sudhakar, P., Hautefort, I., et al. (2022). Multilayered networks of SalmoNet2 enable strain comparisons of the salmonella genus on a molecular level. *mSystems* 7, e0149321. doi: 10.1128/mSystems.01493-21
- Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., et al. (2022). Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat. Protoc.* 17, 1735–1761. doi: 10.1038/s41596-022-00710-w
- Perez-Riverol, Y., Bai, M., da Veiga Leprevost, F., Squizzato, S., Park, Y. M., Haug, K., et al. (2017). Discovering and linking public “Omics” datasets using the omics discovery index. *Nat. Biotechnol.* 35, 406. doi: 10.1038/nbt.3790
- Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., et al. (2018). A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* 172, 358–372.e23. doi: 10.1016/j.cell.2017.12.006
- Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J., et al. (2019). Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. *Metabolites* 9, 76. doi: 10.3390/metabo9040076
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, D61–D65. doi: 10.1093/nar/gkl842
- Raghunathan, A., Reed, J., Shin, S., Palsson, B., and Daefler, S. (2009). Constraint-based analysis of metabolic capacity of salmonella typhimurium during host-pathogen interaction. *BMC Syst. Biol.* 3, 38. doi: 10.1186/1752-0509-3-38
- Ramos, P. I. P., Fernández Do Porto, D., Lanzarotti, E., Sosa, E. J., Burguener, G., Pardo, A. M., et al. (2018). An integrative, multi-omics approach towards the prioritization of *Klebsiella pneumoniae* drug targets. *Sci. Rep.* 8, 10755. doi: 10.1038/s41598-018-28916-7
- Rashid, M. I., Naz, A., Ali, A., and Andleeb, S. (2017). Prediction of vaccine candidates against *Pseudomonas aeruginosa*: An integrated genomics and proteomics approach. *Genomics* 109, 274–283. doi: 10.1016/j.ygeno.2017.05.001
- Rigden, D. J., and Fernández, X. M. (2022). The 2022 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 50, D1–D10. doi: 10.1093/nar/gkab1195
- Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* 16, 85–97. doi: 10.1038/nrg3868
- Rogler, G., and Vavricka, S. (2015). Exposome in IBD: recent insights in environmental factors that influence the onset and course of IBD. *Inflamm. Bowel Dis.* 21, 400–408. doi: 10.1097/MIB.0000000000000229
- Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13, e1005752. doi: 10.1371/journal.pcbi.1005752
- Rueggsegger, L., Xiao, J., Naziripour, A., Kanumuambidi, T., Brown, D., Williams, F., et al. (2022). Multidrug-resistant gram-negative bacteria in burn patients. *Antimicrob. Agents Chemother.* 66, e0068822. doi: 10.1128/aac.00688-22
- Schuler, G. D., Epstein, J. A., Ohkawa, H., and Kans, J. A. (1996). Entrez: molecular biology database and retrieval system. *Methods Enzymol.* 266, 141–162. doi: 10.1016/S0076-6879(96)66012-1
- Shaik, S., Singh, A., Suresh, A., and Ahmed, N. (2022). Genome informatics and machine learning-based identification of antimicrobial resistance-encoding features and virulence attributes in *Escherichia coli* genomes representing globally prevalent lineages, including high-risk clonal complexes. *MBio* 12, e0379621. doi: 10.1128/mbio.03796-21
- Suresh, A., Shaik, S., Baddam, R., Ranjan, A., Kumar, S., Jadhav, S., et al. (2021). Evolutionary dynamics based on comparative genomics of pathogenic *Escherichia coli* lineages harboring polyketide synthase (*pkS*) island. *MBio* 12, e03634–20. doi: 10.1128/mbio.03634-20
- Suzuki, S., Horinouchi, T., and Furusawa, C. (2014). Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* 5, 5792. doi: 10.1038/ncomms6792
- Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., Sugawara, H., et al. (2002). DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Res.* 30, 27–30. doi: 10.1093/nar/30.1.27
- Tian, S., Wang, C., Li, Y., Bao, X., Zhang, Y., and Tang, T. (2021). The impact of SlyA on cell metabolism of *A. baumannii*: A joint study of transcriptomics and metabolomics. *J. Proteome Res.* 20, 184–190. doi: 10.1021/acs.jproteome.0c00281
- Tuncbag, N., Gosline, S. J. C., Kedaigle, A., Soltis, A. R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. *PLoS Comput. Biol.* 12, e1004879. doi: 10.1371/journal.pcbi.1004879
- Veeraraghavan, B., and Walia, K. (2019). Antimicrobial susceptibility profile & resistance mechanisms of global antimicrobial resistance surveillance system (GLASS) priority pathogens from India. *Indian J. Med. Res.* 149, 87–96. doi: 10.4103/ijmr.IJMR_214_18
- Wang, Y., Yang, Y., Chen, S., and Wang, J. (2021). DeepDRK: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Brief. Bioinform.* 22, bbab048. doi: 10.1093/bib/bbab048
- Xie, R., Shao, N., and Zheng, J. (2020). Integrated Co-functional network analysis on the resistance and virulence features in. *Front. Microbiol.* 11, 598380. doi: 10.3389/fmicb.2020.598380
- Yoon, S. H., Han, M.-J., Jeong, H., Lee, C. H., Xia, X.-X., Lee, D.-H., et al. (2012). Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol.* 13, R37. doi: 10.1186/gb-2012-13-5-r37