
Organoids Multi Omics Emerging Pollutants Toxicity Assessment and Environmental Health: A Survey

www.surveyx.cn

Abstract

This survey paper explores the interdisciplinary approach combining organoids, multi-omics, and systems biology to assess the impact of emerging pollutants on environmental health. Emerging pollutants, including pharmaceuticals and engineered nanoparticles, present significant risks due to their persistence and ability to evade conventional water treatment. Advanced biological models like organoids offer a more accurate representation of human tissues, crucial for toxicity assessment. The integration of multi-omics techniques provides a comprehensive examination of biological responses, enhancing our understanding of pollutant-induced toxicity mechanisms. This paper highlights the challenges in data integration, such as high dimensionality and heterogeneity, and the need for advanced computational methods, including machine learning and network analysis, to overcome these obstacles. Organoids, enhanced by engineering strategies and multi-omics data, serve as sophisticated models for studying human physiology and disease, offering novel insights into environmental health. The survey underscores the necessity of interdisciplinary collaboration to address the complexities of emerging pollutants, emphasizing the potential of these integrated methodologies to improve risk assessment and develop targeted interventions. Future research directions include refining organoid models, enhancing data integration techniques, and exploring innovative computational approaches to advance environmental health outcomes.

1 Introduction

1.1 Significance of Studying Emerging Pollutants

Studying emerging pollutants is essential due to their widespread presence and potential adverse effects on human health and ecosystems. These pollutants, which include pharmaceuticals, personal care products, and engineered nanoparticles, often evade conventional water treatment systems, posing significant risks to aquatic environments and human populations [1]. Their complex nature and diverse chemical properties necessitate a thorough understanding of their interactions within biological systems, particularly regarding endocrine disruption, as exemplified by compounds like Bisphenol A [2].

Advanced biological models, such as organoids, are pivotal for toxicity assessment, as they more accurately replicate human tissues compared to traditional models, offering crucial insights into the toxicological impacts of these pollutants [2]. Organoids modeling neurological tissues have significantly enhanced our understanding of human brain development and disease [2]. The integration of multi-omics technologies is vital for elucidating the effects of emerging pollutants, enabling comprehensive examinations of the exposome and aiding in identifying potential disease etiological factors [3].

Developing effective wastewater treatment technologies is imperative to mitigate the risks posed by these pollutants to human health and the environment [1]. The inherent complexity of emerging pollutants demands robust methodologies to unravel their biological effects, underscoring the

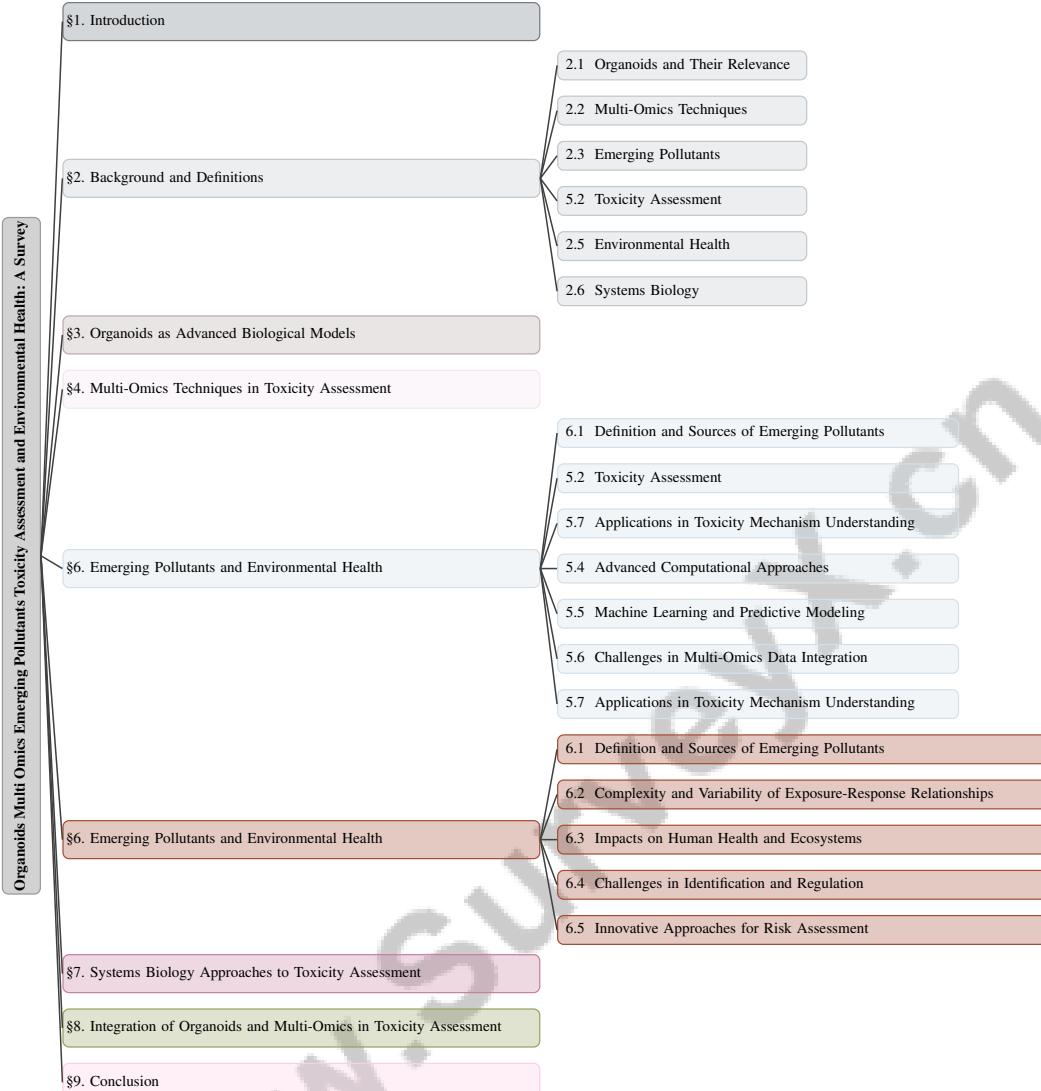


Figure 1: chapter structure

importance of interdisciplinary research [3]. By utilizing advanced biological models, multi-omics techniques, and innovative treatment technologies, researchers can better predict, assess, and mitigate risks associated with these contaminants, thereby safeguarding environmental and public health [4].

1.2 Interdisciplinary Nature of the Research

The interdisciplinary nature of research that combines organoids, multi-omics, and systems biology is crucial for advancing our understanding of environmental health. This convergence enables a thorough exploration of complex biological systems and their interactions with environmental pollutants. Organoids, derived from human pluripotent stem cells (hPSCs) and adult stem cells (AdSCs), exemplify this interdisciplinary approach by serving as advanced models for studying human physiology and disease in the context of environmental health [5]. The integration of various omics technologies, including genomics, proteomics, metabolomics, and microbiomes, enhances the depth of this research, facilitating a nuanced understanding of biological responses to environmental stimuli [6].

The application of Explainable Artificial Intelligence (xAI) methods in multi-omics research addresses interpretability challenges associated with deep learning models, thereby improving the transparency and reliability of predictive analyses [7]. This highlights the necessity of integrating computational

and biological sciences to enhance the interpretive power of complex datasets. Furthermore, the development and characterization of human organoids illustrate the collaborative efforts required in organoid technology, bridging gaps between biology, engineering, and computational modeling [2].

Systems biology plays a vital role in this interdisciplinary framework by integrating experimental and computational methods to model biological processes such as apoptosis, providing insights into cellular responses to environmental factors [4]. This approach not only enhances our understanding of fundamental biological processes but also informs the development of novel therapeutic strategies and environmental health interventions. Collectively, these interdisciplinary strategies emphasize the importance of collaborative research in addressing the complexities of environmental health and emerging pollutants.

1.3 Structure of the Survey

This survey is meticulously structured to provide a comprehensive examination of the interdisciplinary approaches employed in studying the impact of emerging pollutants on environmental health. The paper begins with an **Introduction** that underscores the significance of understanding emerging pollutants and their implications for human health and ecosystems through advanced methodologies such as organoids, multi-omics, and systems biology, setting the stage for detailed exploration.

The next section, **Background and Definitions**, offers foundational knowledge by defining key concepts such as organoids, multi-omics techniques, emerging pollutants, toxicity assessment, environmental health, and systems biology, equipping readers with the necessary background to appreciate the complexities involved in subsequent discussions.

Following this, the survey delves into **Organoids as Advanced Biological Models**, highlighting their role in replicating human tissues and their applications in disease modeling and therapeutic testing. This section also addresses the engineering advancements, challenges, and limitations associated with organoid systems, along with recent developments in imaging and analysis techniques.

The paper then transitions to **Multi-Omics Techniques in Toxicity Assessment**, exploring the integration of diverse omics data, advanced computational approaches, and the role of machine learning in predictive modeling. It also identifies challenges faced in multi-omics data integration and its applications in understanding toxicity mechanisms.

In the section on **Emerging Pollutants and Environmental Health**, the survey defines emerging pollutants, discusses their sources, and examines their impact on human health and ecosystems. It addresses the complexities of exposure-response relationships and the challenges in identifying and regulating these pollutants while exploring innovative risk assessment approaches.

The subsequent section, **Systems Biology Approaches to Toxicity Assessment**, elaborates on the application of systems biology in understanding toxicity mechanisms through data integration, computational modeling, and network analysis, highlighting the importance of dynamic systems and predictive modeling.

The penultimate section, **Integration of Organoids and Multi-Omics in Toxicity Assessment**, explores how these methodologies can be combined to enhance environmental health research, particularly regarding emerging pollutants.

In the **Conclusion**, we synthesize key insights derived from the survey, emphasizing the critical significance of adopting an interdisciplinary approach that integrates diverse fields such as genomics, proteomics, and machine learning. We outline prospective directions for future research, focusing on advancements in explainable artificial intelligence (xAI) methods that can enhance the interpretability of complex multi-omics data, thereby facilitating more effective clinical applications and improving our understanding of intricate biological systems [8, 7]. This structured approach ensures a holistic understanding of the complex interplay between emerging pollutants and environmental health, fostering advancements in both research and practical applications. The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Organoids and Their Relevance

Organoids, derived from stem cells like human embryonic and induced pluripotent stem cells, are sophisticated three-dimensional structures that mimic human tissue architecture and functionality. These models are crucial in environmental health research, providing more accurate representations of human biology than traditional models, particularly in studying organ development and disease mechanisms, including vascularization and structural precision [9]. Human cerebral organoids, for instance, cultivated from iPSCs, serve as innovative in vitro models, bridging traditional studies and disease modeling by closely replicating the cellular architecture of the human brain. This allows for deeper insights into brain development and neurological disorder pathogenesis [10, 5, 11]. In environmental health, organoids offer physiologically relevant systems for examining the effects of pollutants, facilitating the integration of systems biology approaches to understand how environmental factors influence organ development and function, ultimately aiding in developing effective interventions and regulatory strategies [5].

2.2 Multi-Omics Techniques

Multi-omics techniques, which integrate genomics, transcriptomics, proteomics, metabolomics, and epigenomics, provide a comprehensive understanding of biological systems and their responses to environmental stimuli. These methodologies are essential in toxicity assessment, offering a multidimensional view of organism responses to pollutants [6]. By combining diverse omics data, researchers can dissect complex biological interactions and reveal regulatory networks influenced by toxic exposures [3]. Advanced computational methods, including systems biology frameworks and machine learning algorithms, are pivotal in analyzing multi-omics data, enabling key pathway identification in toxicity mechanisms. Techniques like Harmonic Persistent Homology aid in biomarker and disease subtype discovery within multi-omics datasets [12]. The limitations of single omics studies underscore the necessity of multi-omics approaches [6]. Single-cell multi-omics technologies further enhance this field by providing unprecedented resolution in understanding cellular heterogeneity and its implications for toxicity, facilitating the identification of specific cell types affected by toxic exposures and improving risk assessment and targeted interventions [9, 3].

2.3 Emerging Pollutants

Emerging pollutants, or contaminants of emerging concern, include pharmaceuticals, personal care products, and engineered nanoparticles, posing potential risks to health and ecosystems due to their persistence and bioaccumulation potential, often evading conventional wastewater treatment processes [13]. Their complexity is compounded by interactions within environmental matrices and biological systems, leading to unpredictable toxicological outcomes [14]. Sources include domestic and industrial waste, agricultural runoff, and urban effluents, highlighting the need for comprehensive monitoring and regulatory frameworks [14]. Pharmaceuticals and personal care products frequently detected in water bodies are linked to adverse ecological and health effects [1]. Despite their technological benefits, engineered nanoparticles pose significant health and environmental risks [15]. The study of emerging pollutants is crucial as they can disrupt biological systems, causing chronic and acute toxicity in aquatic life and humans [16]. Pollution is implicated in diseases like inflammatory bowel disease, emphasizing broader health implications [17]. Integrating multi-omics data at single-cell resolution is essential for elucidating biological responses to these pollutants [13]. Innovative methodologies are needed to evaluate the long-term impacts of multivariate exposures on health [14]. Research on emerging pollutants is vital for developing effective risk assessment and management strategies to protect environmental and public health [1]. Advanced scientific approaches, such as chemical systems biology and multi-omics, enhance understanding of pollutant interactions with biological systems, ultimately improving environmental health outcomes.

2.4 Toxicity Assessment

Toxicity assessment evaluates adverse effects of chemical substances, including emerging pollutants, on organisms and ecosystems, focusing on their interaction mechanisms with biological systems [18]. Traditional methodologies relied on animal models and cell lines, often falling short in accu-

rately representing human tissue responses. This has led to interest in advanced biological models like organoids, which offer precise investigation of pollutants' effects on health [18]. Integrating multi-omics techniques into toxicity assessment enhances the understanding of biological responses to pollutants, facilitating comprehensive exposome examinations and identifying potential disease etiological factors [19]. However, the diverse nature of omics data presents analysis challenges. Innovative methodologies, such as the Mediation External Summary Statistic, are crucial for understanding environmental exposure pathways [20]. Bayesian methods for modeling cumulative exposure provide nuanced insights into exposure-response relationships [13]. These approaches, combined with datasets measuring toxicity and estrogenic responses, enhance understanding of pollutant-induced toxicity mechanisms and their implications for health [19]. Assessing emerging pollutants' toxicity is complicated by their diverse chemical properties, necessitating novel methodologies and advanced approaches like chemical systems biology and multi-omics [18]. Interdisciplinary strategies provide insights into pollutant-induced toxicity mechanisms and develop effective health and ecosystem mitigation strategies.

2.5 Environmental Health

Environmental health examines the interactions between environmental factors and human health, emphasizing their impact on well-being and disease etiology. The dynamic nature of environmental health challenges necessitates a comprehensive approach, as environmental factors significantly modify disease risk and natural history, such as in inflammatory bowel disease [17]. A systems-level approach is essential to navigate interactions between pollutants and biological systems. Advanced analytical frameworks, such as the Bayesian multiple index model (BMIM), enhance decision-making in environmental health policy [21, 22]. Integrating diverse disciplines, including systems biology, multi-omics, and computational modeling, is necessary to unravel toxicity mechanisms and develop effective interventions. Advanced biological models, particularly organoids, provide a promising avenue for investigating pollutants' effects on health and ecosystems, enabling studies of intricate interactions between environmental factors and biological systems [17]. Integrating multi-omics techniques with organoid models facilitates a comprehensive examination of the exposome, aiding in identifying potential etiological factors for diseases [17]. Innovative methodologies for risk assessment and management, through advanced statistical models, system dynamics, and multi-omics data analysis, aim to improve understanding of complex interactions of environmental exposures and their cumulative health effects, addressing high-dimensional data analysis challenges and the need for holistic assessment frameworks [21, 22]. System-based approaches, as employed in systems biology, are crucial for addressing multifaceted challenges posed by emerging pollutants, enabling researchers to consider interactions between various environmental factors and biological systems for a more comprehensive understanding of environmental health issues and the development of targeted interventions.

2.6 Systems Biology

Systems biology seeks to understand complex interactions within biological systems by integrating experimental and computational approaches, emphasizing dynamic regulation critical for elucidating toxicity mechanisms [23]. Network science and emerging computational technologies, such as quantum computing, enhance systems biology's capacity to analyze disease mechanisms, offering novel insights into environmental health [24]. In understanding toxicity mechanisms, systems biology involves formal representation of expert knowledge, advancing precision medicine [25]. This approach categorizes existing research methodologies and insights into receptor-induced apoptosis, deepening the understanding of molecular underpinnings of toxicity [4]. Computational methods like Computational Fluid Dynamics (CFD) integrate mass transfer dynamics with reaction kinetics, providing a holistic understanding of biological processes and their responses to toxic exposures [26]. Feedback mechanisms explored through systems theory are integral to understanding regulatory processes governing biological responses to environmental stressors, significantly influencing treatment design and patient monitoring. Bayesian uncertainty analysis facilitates efficient global parameter searches, accounting for variability and uncertainty from multiple sources when modeling complex biological interactions [27]. Systems biology provides a comprehensive framework for modeling signaling pathways and cellular responses to environmental pollutants, informing strategies to mitigate adverse health effects and improving environmental health outcomes.

3 Organoids as Advanced Biological Models

3.1 Overview of Organoid Technology

Organoid technology marks a pivotal advancement in biomedical research, enabling the *in vitro* replication of human tissues and organs with high fidelity. Derived from stem cells, such as human pluripotent stem cells (hPSCs) and adult stem cells (AdSCs), organoids mimic the architectural and functional characteristics of their *in vivo* counterparts, addressing limitations of traditional two-dimensional cultures that often fail to capture complex cellular interactions and microenvironmental conditions [28]. Engineered biomaterial microsystems further enhance their physiological relevance by providing structural support and biochemical signals that guide cell differentiation and organization [23]. Vascularized organoids, for instance, are engineered to reflect early prenatal brain development through human ETS variant 2 (hETV2) expression, illustrating this innovation [29]. Despite these advancements, challenges such as vascularization, scalability, and reproducibility persist. Strategies like agent-based modeling frameworks are being explored to simulate organoid growth and analyze cellular behaviors, enhancing their clinical applicability and paving the way for personalized medicine and therapeutic testing [30, 31].

In environmental health research, organoids serve as physiologically relevant systems for studying the effects of environmental pollutants on human tissues [5]. By integrating systems biology approaches, researchers can elucidate how environmental factors influence organ development and function, contributing to improved health outcomes [5]. Organoids have been pivotal in generating diverse cell types, such as those related to the cerebral cortex and retina, offering a versatile platform for studying human development and disease. Systems biology approaches, including Computational Fluid Dynamics (CFD), can simulate reactant transport and kinetics, enhancing the physiological relevance of organoid models in toxicity assessment [26].

3.2 Engineering and Enhancements in Organoid Systems

Advancements in organoid systems have been driven by innovative engineering strategies that enhance their functionality and applicability in environmental health research. Stem cell niche engineering recreates the microenvironment necessary for stem cell growth and differentiation, increasing organoid models' fidelity to their *in vivo* counterparts [31]. Microfluidic approaches enable the study of dynamic interactions between organoids and environmental factors, maintaining physiological relevance through precise control of fluid flow and nutrient gradients. These advancements facilitate high-throughput screening of environmental pollutants, aiding in the identification of toxicological effects and informing targeted interventions.

Integrating these engineering strategies with systems biology approaches enhances our understanding of the complex mechanisms by which environmental pollutants affect human health. As organoid technology evolves, it holds promise for personalized medicine and targeted interventions based on individual susceptibility to environmental stressors [32, 2, 33, 28].

3.3 Organoids in Disease Modeling and Therapeutic Testing

Organoids have become transformative tools in disease modeling and therapeutic testing, offering significant advantages over traditional *in vitro* models by accurately replicating human tissue complexity [28]. In neurological research, brain organoids have demonstrated potential in modeling human brain development and forming spontaneously active neuronal networks [34]. They have been used to study key regulatory gene expression, such as ZEB2, and associated morphological changes, advancing our understanding of neurological diseases and therapeutic interventions. Transplantation of human brain organoids into host brains shows their ability to mature and integrate into the host's neural networks, forming functional neuronal connections [35].

Despite their promise, challenges such as the reproducibility and maturation of organoid cultures remain, which are critical for ensuring consistent and reliable results across studies [32]. Addressing these issues is essential for maximizing organoids' potential in disease modeling and therapeutic testing.

3.4 Challenges and Limitations of Organoid Models

While organoid technology offers transformative potential in replicating human tissues, several challenges and limitations affect their full applicability in environmental health research. The reliance on mouse-derived extracellular matrices can impact the reproducibility and clinical relevance of organoids, especially when modeling human-specific diseases [5]. Incomplete maturation of organoids poses challenges, as it affects their ability to replicate native tissue complexity [5]. Current models often lack essential microenvironmental features, such as blood vessels and immune cells, which are crucial for replicating *in vivo* environments and understanding tissue interactions with environmental pollutants [33]. The absence of fully functional vasculature can lead to necrosis, affecting experimental reliability.

Replicating the full cellular diversity and interactions present in native tissues is another significant challenge. For instance, liver organoids frequently lack the diverse cell types and interactions necessary for a comprehensive understanding of liver diseases [36]. The scalability of organoid production remains a concern, as existing methodologies often lack the robustness required for high-throughput applications essential for large-scale toxicity assessments [9]. Challenges in multi-omics integration necessitate sustained interdisciplinary collaboration and innovative engineering and computational strategies, such as customizable deep learning models and interaction-based frameworks. These approaches manage high-dimensional data, capture interrelationships between molecular features, and facilitate research findings' sharing and reproducibility through established data repositories and open platforms [37, 38, 39].

3.5 Advancements in Imaging and Analysis of Organoids

Recent advancements in imaging and analysis techniques have significantly enhanced organoids' utility as models for studying human biology. As illustrated in Figure 2, these advancements highlight key categories such as engineered organoid systems, the integration of SM-Omics for molecular profiling, and advanced imaging techniques like the SegmentAnything approach. Engineered organoid systems offer increased reproducibility and functional relevance, critical for effective drug screening and disease modeling [31]. The integration of SM-Omics, which combines high-throughput automation with simultaneous RNA and protein level measurement, provides a versatile tool for detailed molecular profiling of organoids [40]. This integration is crucial for understanding the biological processes underpinning organ development and disease mechanisms.

Advanced computational methods, like the SegmentAnything approach, improve the accuracy of organoid imaging and analysis by leveraging state-of-the-art segmentation techniques to delineate organoid structures [41]. Accurate segmentation is essential for quantifying organoids' structural and functional properties, critical for disease modeling and therapeutic testing. Datasets such as MultiOrg allow for a comprehensive examination of label noise and its impact on model performance, enhancing organoid detection and analysis reliability in biomedical research [42].

The creation of functional and vascularized models, such as human brain tissue organoids, represents a significant advancement, facilitating a deeper understanding of brain development and disease mechanisms [33]. Integration of advanced imaging techniques with these models maximizes their potential for disease modeling and therapeutic applications [33]. These advancements enable unprecedented insights into organ development, disease modeling, and potential therapeutic interventions [32]. Continued refinement of these methodologies will further unlock organoids' potential as powerful tools for advancing our understanding of human biology and improving health outcomes.

4 Multi-Omics Techniques in Toxicity Assessment

4.1 Integration of Diverse Omics Data

The integration of diverse omics data is fundamental in toxicity assessment, offering comprehensive insights into biological responses to environmental pollutants. Multi-omics approaches encompass genomics, transcriptomics, proteomics, metabolomics, and epigenomics, each providing unique perspectives on the molecular mechanisms of toxicity. By combining these data layers, researchers can achieve a holistic understanding of the intricate interactions within biological systems and their responses to toxic exposures [3].

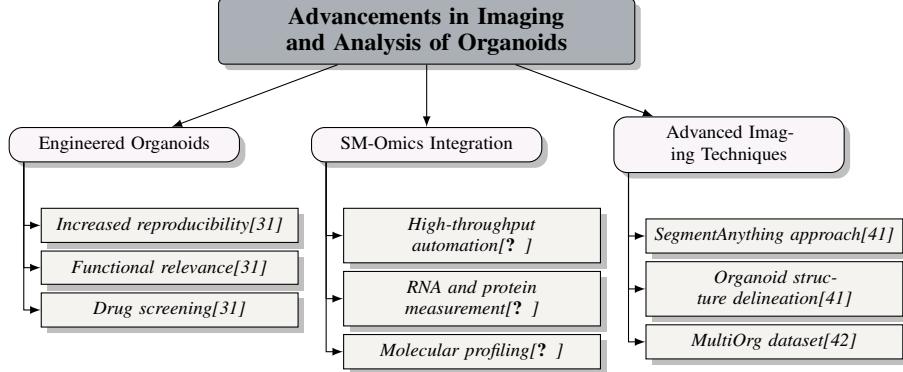


Figure 2: This figure illustrates the recent advancements in imaging and analysis of organoids, highlighting key categories such as engineered organoid systems, integration of SM-Omics for molecular profiling, and advanced imaging techniques like the SegmentAnything approach. These advancements enhance reproducibility, functional relevance, and accuracy in organoid models, facilitating drug screening, disease modeling, and therapeutic applications.

As depicted in Figure 3, this figure illustrates the integration of diverse omics data, highlighting key multi-omics approaches, computational methods, and challenges along with their respective solutions for toxicity assessment. The visual representation emphasizes how the integration enriches the understanding of molecular mechanisms and improves risk assessment strategies.

A primary challenge in multi-omics integration is the heterogeneity and complexity of the data, which necessitates advanced computational methods. The self-omics framework utilizes self-supervised learning techniques to enhance data integration, thereby improving the identification of critical molecular interactions and pathways affected by pollutants [43]. Additionally, the Machine Learning for Pathway Analysis (MLP-DPI) method merges machine learning with network comparison to uncover discriminant pathways and key gene interactions influenced by disease conditions [28].

Network-based methods, such as MOD-Finder, further advance the field by identifying key molecular interactions and pathways impacted by toxic exposures. These approaches employ topological data analysis techniques, including harmonic persistent homology, to pinpoint essential simplices in multi-omics data, facilitating the discovery of critical biomarkers and disease subtypes [18, 12]. The integration of various omics types enriches the understanding of molecular mechanisms underlying toxicity, contributing to improved risk assessment and targeted interventions for mitigating the adverse effects of emerging pollutants [3].

Advanced computational methods, including machine learning and network analysis, are vital in interpreting multi-omics data for toxicity assessment [43]. Machine learning algorithms, particularly the MLP-DPI method, are instrumental in identifying key gene interactions and pathways perturbed by environmental pollutants, providing novel insights into the molecular basis of toxicity. Moreover, Bayesian uncertainty analysis allows researchers to account for variability from multiple sources in modeling complex biological interactions, enhancing the reliability of toxicity assessments [27].

4.2 Advanced Computational Approaches

Advanced computational approaches are essential for analyzing multi-omics data and deciphering complex biological interactions. Leveraging high-throughput technologies, these methods enable researchers to uncover insights into the molecular mechanisms of toxicity induced by emerging pollutants [12]. Techniques such as machine learning, network analysis, and topological data analysis are powerful tools for integrating and interpreting multi-omics data, thus providing a comprehensive understanding of interactions between different omics layers [43].

As illustrated in Figure 5, the figure highlights these advanced computational approaches in multi-omics data analysis, showcasing the significance of machine learning, network-based methods, and quantum computing, along with specific frameworks and models pertinent to each category.

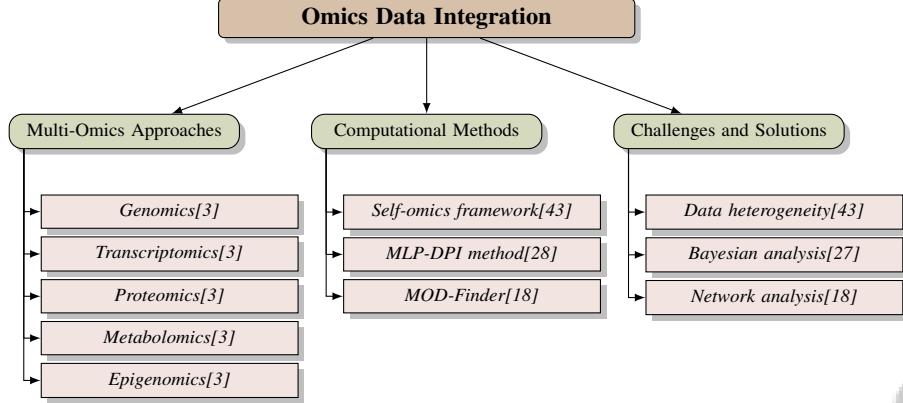


Figure 3: This figure illustrates the integration of diverse omics data, highlighting key multi-omics approaches, computational methods, and challenges with their respective solutions for toxicity assessment. The integration enriches understanding of molecular mechanisms and improves risk assessment strategies.

Machine learning has emerged as a valuable asset in multi-omics analysis, facilitating the identification of complex patterns within large datasets [43]. The self-omics framework addresses challenges in data integration, enabling the discovery of novel molecular interactions and pathways involved in toxicity mechanisms through pre-training techniques that enhance interpretability and predictive power [43].

Network-based methods, such as MOD-Finder, enhance the identification of key molecular interactions and pathways affected by toxic exposures, employing topological data analysis to model biological interactions and identify crucial nodes and edges influenced by pollutants [18, 44]. By integrating multi-omics data, researchers can gain a comprehensive understanding of toxicity mechanisms and develop targeted interventions to mitigate adverse effects on human health and ecosystems.

Despite advancements, challenges remain in integrating multi-omics data for toxicity assessment. The heterogeneity and complexity of biological systems, along with the diverse nature of omics data types, pose significant obstacles to effective integration and interpretation [3]. Developing advanced computational tools, including self-supervised learning frameworks like self-omics, is crucial for overcoming these challenges and maximizing the potential of multi-omics techniques in toxicity assessment [43]. Through these methodologies, researchers can deepen their understanding of molecular mechanisms underlying pollutant-induced toxicity, ultimately contributing to more effective risk assessment and management strategies for emerging pollutants.

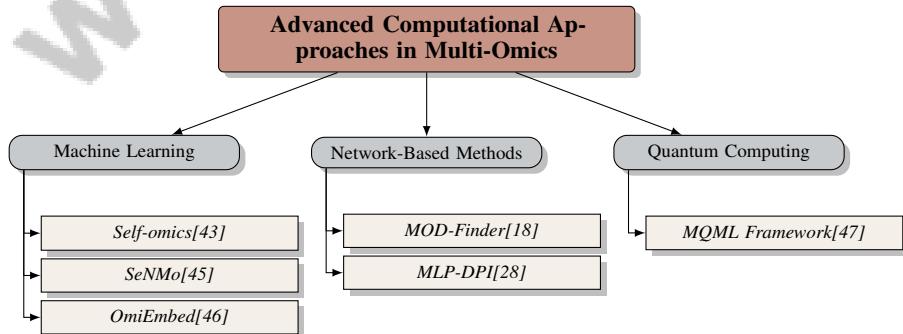


Figure 4: This figure illustrates the advanced computational approaches in multi-omics data analysis, highlighting machine learning, network-based methods, and quantum computing, with specific frameworks and models for each category.

5 Emerging Pollutants and Environmental Health

As environmental concerns continue to mount, the identification and understanding of emerging pollutants have become paramount in the field of environmental health. These substances, which include a range of synthetic and naturally occurring chemicals, pose significant risks to ecosystems and human health due to their persistence and potential for bioaccumulation. To effectively address these challenges, it is essential to first define what constitutes emerging pollutants and explore their various sources. This foundational knowledge will set the stage for a deeper examination of their implications on environmental health, as discussed in the following subsection.

5.1 Definition and Sources of Emerging Pollutants

Emerging pollutants, also referred to as contaminants of emerging concern (CECs), represent a diverse array of synthetic or naturally occurring chemicals that are not commonly monitored in the environment but have the potential to enter the environment and cause adverse ecological and human health effects [15]. These substances encompass a wide range of compounds, including pharmaceuticals, personal care products, endocrine-disrupting chemicals, and nanomaterials, which are often resistant to conventional water treatment processes and can persist in the environment. The widespread occurrence of emerging pollutants in aquatic environments is attributed to various anthropogenic activities, such as industrial discharges, agricultural runoff, and domestic wastewater effluents [14].

The persistence and bioaccumulative nature of emerging pollutants pose significant challenges to environmental health, as these substances can accumulate in the food chain and ultimately impact human health and ecosystems [15]. The potential risks associated with emerging pollutants are further compounded by their ability to interact with multiple biological pathways, leading to complex exposure-response relationships and potential adverse health effects [13]. For instance, pharmaceuticals and personal care products (PPCPs) are frequently detected in aquatic environments and have been associated with various ecological and health effects, including endocrine disruption and antibiotic resistance [1].

The identification and regulation of emerging pollutants require a comprehensive understanding of their sources, pathways, and potential impacts on human health and ecosystems. This necessitates the development of innovative monitoring and analytical techniques, as well as the integration of multi-omics approaches to elucidate the complex interactions between these pollutants and biological systems [3]. By leveraging advanced scientific methodologies, researchers can better understand the potential risks associated with emerging pollutants and develop effective strategies for mitigating their impacts on human health and the environment .

5.2 Toxicity Assessment

Toxicity assessment is a fundamental aspect of environmental health research, aiming to evaluate the potential adverse effects of chemical substances, including emerging pollutants, on living organisms and ecosystems. This process involves a comprehensive analysis of the interactions between these substances and biological systems, focusing on the mechanisms through which they exert their toxic effects [18]. The complexity of biological systems and the diverse chemical properties of emerging pollutants present significant challenges in toxicity assessment, as they can lead to unpredictable toxicological outcomes .

Traditional toxicity assessment methodologies have predominantly relied on animal models and cell lines, which often fail to capture the full complexity of human-specific biological responses and the dynamics of epithelial-immune interactions . This has led to a growing interest in utilizing advanced biological models, such as organoids, which offer a more accurate representation of human tissues and organs *in vitro*. Organoids, derived from human pluripotent stem cells (hPSCs) and adult stem cells (AdSCs), provide a promising alternative for studying human-specific toxicological responses, allowing researchers to investigate the effects of environmental pollutants on human health with greater precision [18].

The integration of multi-omics techniques into toxicity assessment has further enhanced our understanding of the complex biological responses to environmental pollutants. These methodologies enable a comprehensive examination of the exposome, facilitating the identification of potential

etiological factors for diseases [19]. However, the integration of diverse omics data types presents challenges due to the varying nature of each dataset, which can complicate the analysis and interpretation of results .

Innovative approaches, such as the Mediation External Summary Statistic-based Inference (MESSI) method, have been proposed to improve the estimation of natural direct and indirect effects in mediation analysis, which is crucial for understanding the pathways through which environmental exposures impact health [20]. Additionally, Bayesian methods have been developed to model cumulative exposure and its impacts, providing a more nuanced understanding of exposure-response relationships [13]. These approaches, combined with datasets measuring toxicity and estrogenic responses across multiple exposure concentrations, contribute to a more detailed understanding of the mechanisms underlying pollutant-induced toxicity and their potential implications for human health [19].

The challenge of accurately assessing the toxicity of emerging pollutants is compounded by their diverse chemical properties and interactions within biological systems [3]. For instance, nanoparticles present unique challenges due to their distinct physicochemical properties, which can lead to varied toxicological effects during their lifecycle, including production, use, and disposal [48]. This complexity necessitates the development of novel methodologies and the integration of advanced scientific approaches, such as chemical systems biology and multi-omics, to unravel the intricate biological effects of these pollutants [18]. By employing such interdisciplinary approaches, researchers can better understand the mechanisms underlying pollutant-induced toxicity and develop effective strategies for mitigating their impacts on human health and ecosystems .

5.3 Applications in Toxicity Mechanism Understanding

The application of multi-omics techniques in toxicology has significantly enhanced our understanding of the complex mechanisms by which emerging pollutants impact biological systems. By integrating diverse omics data—such as transcriptomics, proteomics, and metabolomics—researchers can capture a comprehensive view of the molecular responses and regulatory pathways activated by chemical exposure. This holistic approach not only improves the identification of specific biomarkers associated with toxicity but also facilitates a systemic understanding of adverse outcome pathways, ultimately advancing risk assessment and informing regulatory decisions. [49, 50]. These integrative approaches enable researchers to dissect the complex interactions between various omics layers, providing a comprehensive view of the molecular pathways and networks that are perturbed by toxic exposures . By leveraging multi-omics data, researchers can identify key molecular signatures associated with toxicity, facilitating the development of targeted interventions and risk assessment strategies .

The use of multi-omics techniques has proven particularly valuable in elucidating the mechanisms through which environmental pollutants exert their toxic effects. For example, the integration of genomics, transcriptomics, and proteomics data has enabled the identification of novel gene interactions and pathways that are disrupted by toxic exposures, providing insights into the complex biological processes underlying toxicity [28]. Machine learning algorithms, such as those employed in the Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI), have been instrumental in identifying key gene interactions and pathways affected by disease conditions, offering a more comprehensive understanding of the molecular mechanisms underlying toxicity [51].

The development of advanced computational methods, such as those that incorporate self-supervised learning techniques, has further enhanced the capacity of multi-omics approaches to analyze complex datasets and identify critical molecular interactions [43]. Self-omics, for instance, employs a pre-training paradigm that integrates multiple Self-Supervised Learning (SSL) techniques, enabling the analysis of multi-omics cancer data with improved accuracy and efficiency [43].

Despite these advancements, integrating diverse omics data for toxicity assessment remains challenging due to the heterogeneity of data types and the need for advanced computational tools to analyze and interpret these complex datasets [3]. Techniques such as regression/association-based methods, clustering-based methods, and network-based methods have been developed to address these challenges, offering a framework for the systematic integration of multi-omics data in toxicity assessment . These methodologies enable researchers to overcome the limitations of traditional single-omics approaches, providing a more holistic understanding of the biological mechanisms

underlying pollutant-induced toxicity and informing the development of targeted interventions to mitigate the adverse effects of emerging pollutants on human health and ecosystems .

5.4 Advanced Computational Approaches

In the rapidly evolving field of multi-omics data analysis, advanced computational approaches have become indispensable for deciphering the complex interactions inherent in biological systems. These approaches leverage high-throughput technologies to analyze large-scale datasets, enabling researchers to uncover intricate patterns and relationships that underpin the molecular mechanisms of toxicity induced by emerging pollutants. Machine learning, in particular, has emerged as a powerful tool in multi-omics data analysis, offering the capability to identify complex patterns and relationships that may not be apparent using traditional methods [43].

One of the significant advancements in this domain is the development of the Self-Normalizing Foundation Model (SeNMo), which employs self-normalizing neural networks to facilitate robust predictions from complex multi-omics datasets [45]. This approach enhances the interpretability and predictive power of multi-omics analyses, allowing researchers to better understand the molecular interactions and pathways involved in toxicity mechanisms. Moreover, the OmiEmbed framework processes high-dimensional omics data through a deep embedding module and multiple downstream task modules, providing a unified multi-task deep learning approach that improves the integration and analysis of multi-omics data [46].

Network-based methods have further advanced the integration of multi-omics data, offering novel insights into the complex interactions within biological systems. The MOD-Finder approach, for instance, enables the identification of key molecular interactions and pathways affected by toxic exposures by utilizing topological data analysis techniques [18]. Additionally, the Machine Learning for Pathway Analysis (MLP-DPI) method combines machine learning with network comparison techniques to identify discriminant pathways and key gene interactions influenced by disease conditions [28].

Furthermore, the application of quantum computing in multi-omics data analysis, as demonstrated by the Multi-Omic Quantum Machine Learning (MQML) framework, provides a novel approach for the classification of lung cancer subtypes, showcasing the potential of quantum technologies to enhance the analysis of complex biological datasets [47]. The integration of these advanced computational methods with multi-omics data enables researchers to gain a more comprehensive understanding of the molecular mechanisms underlying toxicity, ultimately contributing to improved risk assessment and the development of targeted interventions to mitigate the adverse effects of emerging pollutants on human health and ecosystems.

This is illustrated in Figure 5, which highlights the advanced computational approaches in multi-omics data analysis, including machine learning, network-based methods, and quantum computing, with specific frameworks and models for each category.

Despite these advancements, challenges remain in the integration of multi-omics data for toxicity assessment. The heterogeneity and complexity of biological systems, coupled with the diverse nature of omics data types, present significant obstacles to effective data integration and interpretation [3]. To address these challenges, the Pathway Dysregulation Score (PDS) has been proposed as an input feature in a Cox-PH model for predicting patient survival, demonstrating the integration of multi-omics data for toxicity assessment [52]. By leveraging these advanced methodologies, researchers can gain a more comprehensive understanding of the molecular mechanisms underlying pollutant-induced toxicity, ultimately contributing to the development of more effective risk assessment and management strategies for emerging pollutants. This interdisciplinary approach holds great promise for advancing our understanding of environmental health and improving public health outcomes.

5.5 Machine Learning and Predictive Modeling

Machine Learning and Predictive Modeling have become integral components of multi-omics data analysis, offering transformative capabilities in the understanding and prediction of toxicity mechanisms associated with emerging pollutants. The Image2Omics approach exemplifies the innovative application of deep learning models to predict multi-omics outcomes from high-content imaging, thereby bridging the gap between imaging data and molecular insights [53]. This methodology

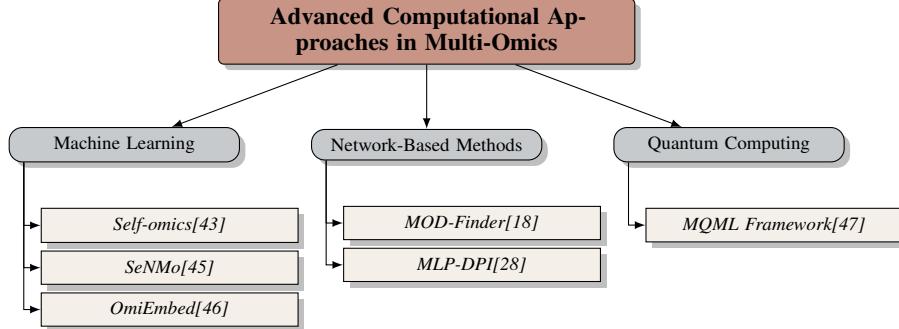


Figure 5: This figure illustrates the advanced computational approaches in multi-omics data analysis, highlighting machine learning, network-based methods, and quantum computing, with specific frameworks and models for each category.

highlights the potential of machine learning to enhance the predictive accuracy and interpretability of multi-omics analyses, facilitating the identification of key molecular interactions and pathways influenced by toxic exposures.

In the context of multi-omics integration, AI-driven approaches have demonstrated significant promise in overcoming the challenges associated with data heterogeneity and complexity. The AI-driven multi-omics integration framework offers a multiscale approach that adaptively fine-tunes segmentation parameters in real-time, enhancing the accuracy and efficiency of multi-omics data analysis [54]. This capability is critical for identifying complex patterns and relationships within large-scale datasets, ultimately contributing to a more comprehensive understanding of the molecular mechanisms underlying toxicity.

The development of specialized machine learning pipelines, such as SystemMatch, further underscores the role of predictive modeling in optimizing preclinical model systems. This pipeline is designed to assess and optimize these systems based on their similarity to human target populations, thereby ensuring the relevance and applicability of preclinical studies to human health outcomes . Revised Sentence: "Innovative strategies, such as multi-omics data integration and advanced membrane-based separation processes, are crucial for effectively translating laboratory findings to real-world human health contexts. These approaches enhance our understanding of the complex interactions between emerging pollutants—such as hazardous chemicals and toxic metals—and biological systems, ultimately enabling the development of targeted interventions that mitigate the adverse health effects associated with these pollutants." [49, 55, 56, 15, 16]

Neural network-based methods, including models like scMVAE and totalVI, have been employed for data integration in single-cell multi-omics studies, utilizing autoencoders to enhance the interpretability and integration of complex datasets [57]. These methodologies are instrumental in unraveling the intricate biological processes underlying toxicity mechanisms, providing novel insights into the cellular and molecular responses to environmental pollutants.

The application of Explainable Artificial Intelligence (xAI) methods in multi-omics research further addresses the interpretability challenges associated with deep learning models. Model-agnostic approaches, such as SHAP and LIME, alongside model-specific techniques like attention mechanisms and Grad-CAM, improve the transparency and reliability of predictive analyses, enabling researchers to gain a deeper understanding of the molecular interactions driving toxicity mechanisms [7].

5.6 Challenges in Multi-Omics Data Integration

The integration of multi-omics data for toxicity assessment presents a series of challenges primarily due to the inherent complexity, high dimensionality, and heterogeneity of the data involved. One of the most significant difficulties lies in the high dimensionality of omics data, where the number of variables often exceeds the number of samples. This imbalance can lead to challenges in network inference and the risk of spurious correlations, complicating the analysis and interpretation of the data [58]. The heterogeneity of datasets, which includes genomics, transcriptomics, proteomics,

metabolomics, and epigenomics, further complicates the integration process due to differences in data size, complexity, and noise.

As illustrated in Figure 6, the primary challenges in multi-omics data integration can be categorized into three key areas: data complexity challenges, analytical method limitations, and integration approaches. Another significant challenge in multi-omics data integration is the presence of unpaired data, which can result in substantial information loss and decreased analytical accuracy [59]. The variability in nanoparticle characteristics, coupled with the lack of standardized toxicity assessment methods, further complicates the integration and analysis of multi-omics data for toxicity assessment [48]. Existing machine learning methods often struggle with data integration due to the diverse nature of omics data types and the black-box nature of deep learning models, which can hinder the interpretability and reliability of predictive analyses [7].

The underdetermined nature of network inference in multi-omics data analysis is another significant challenge, as the number of variables often exceeds the number of samples, increasing the risk of spurious correlations. This issue is compounded by inconsistencies in hyperparameter settings, data distributions, and test protocols, which can lead to ambiguous results regarding the performance of different integration methods. To address these challenges, advanced computational approaches, such as self-supervised learning frameworks like Self-omics, have been developed to enhance the integration and interpretation of multi-omics data [43].

Despite the various challenges associated with data integration, the synthesis of diverse omics datasets—such as transcriptomics, proteomics, and metabolomics—is crucial for enhancing our comprehension of the molecular mechanisms that drive toxicity. This integrative approach not only allows for a more comprehensive understanding of the complex biological pathways involved in toxic responses but also aids in the development of robust risk assessment and management strategies for emerging pollutants. By leveraging multi-omics data, researchers can identify key biomarkers and elucidate the intricate interactions among different biological molecules, ultimately leading to improved predictive models for assessing the impact of chemical exposures on human health and the environment [60, 49, 50]. By leveraging advanced computational approaches, such as machine learning, network analysis, and topological data analysis, researchers can gain a more comprehensive understanding of the complex interactions between environmental pollutants and biological systems, ultimately contributing to improved environmental health outcomes and public health protection. This interdisciplinary approach holds great promise for addressing the challenges posed by emerging pollutants and advancing our understanding of their impact on human health and ecosystems.

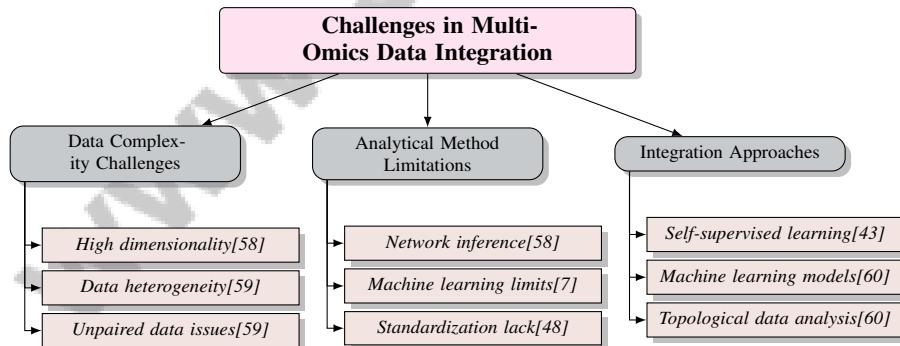


Figure 6: This figure illustrates the primary challenges in multi-omics data integration, categorizing them into data complexity challenges, analytical method limitations, and integration approaches.

5.7 Applications in Toxicity Mechanism Understanding

The application of multi-omics techniques has significantly advanced the understanding of toxicity mechanisms associated with emerging pollutants, offering a comprehensive framework for analyzing the complex molecular interactions and pathways influenced by toxic exposures. By integrating various omics layers—such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics—researchers can achieve a holistic view of biological systems and their responses to environmental stressors. This integrative approach allows for the dissection of complex interactions and the identification of molecular signatures that are critical for understanding toxicity mechanisms.

The utility of multi-omics in toxicity mechanism understanding is significantly enhanced by advanced computational methodologies. The Self-omics framework, which employs self-supervised learning techniques, has emerged as a powerful tool for integrating diverse omics data, facilitating the identification of key molecular interactions and pathways affected by toxic exposures [43]. These methods, combined with network-based approaches like MOD-Finder, enable the identification of critical molecular interactions and pathways that are disrupted by emerging pollutants [18].

Moreover, machine learning algorithms have been instrumental in advancing the analysis of multi-omics data. Approaches such as the Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI) have been developed to identify key gene interactions and pathways perturbed by environmental pollutants, offering novel insights into the molecular mechanisms underlying toxicity. The integration of computational techniques with multi-omics data creates a comprehensive framework for elucidating the intricate biological responses to toxic chemical exposures. This approach enhances our understanding of the interconnected regulatory pathways and molecular interactions involved in toxicity, which single-omics analyses often overlook. By combining diverse omics layers—such as transcriptomics, proteomics, and metabolomics—researchers can identify key biomarkers and pathways that contribute to adverse health effects and ecological impacts. Furthermore, this multi-omics strategy facilitates the development of targeted interventions aimed at mitigating the detrimental consequences of toxic exposures on both human health and ecosystems, ultimately advancing risk assessment and prevention efforts in toxicology. [60, 49]

The application of quantum computing in multi-omics data analysis, as demonstrated by the Multi-Omic Quantum Machine Learning (MQML) framework, has shown promise in enhancing classification accuracy through the integration of diverse omics data types [47]. This novel approach highlights the potential of quantum technologies to enhance the analysis of complex biological datasets, offering new avenues for understanding the molecular mechanisms underlying toxicity [24].

Despite these advancements, challenges remain in the integration of diverse omics data for toxicity assessment, primarily due to the heterogeneity and complexity of biological systems and the diverse nature of omics data types [3]. The development of advanced computational tools and methodologies, such as self-supervised learning frameworks like Self-omics, is essential for overcoming these challenges and maximizing the potential of multi-omics techniques in toxicity assessment [43]. By leveraging these advanced methodologies, researchers can gain a more comprehensive understanding of the molecular mechanisms underlying pollutant-induced toxicity, ultimately contributing to the development of more effective risk assessment and management strategies for emerging pollutants. This interdisciplinary approach holds great promise for advancing our understanding of environmental health and improving public health outcomes.

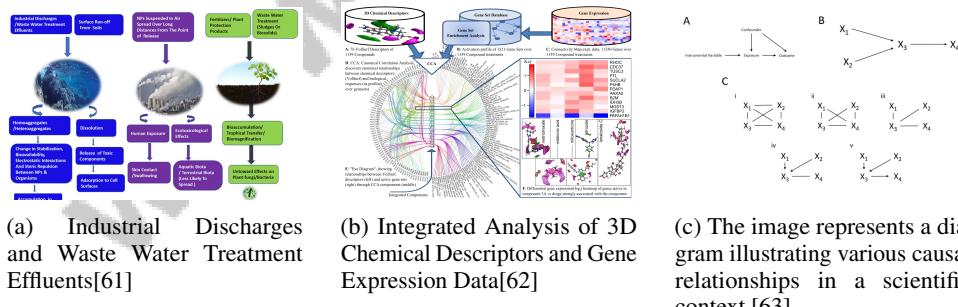


Figure 7: Examples of Applications in Toxicity Mechanism Understanding

As shown in Figure 7, The study of emerging pollutants and their implications for environmental health is a critical area of research, particularly in understanding the mechanisms of toxicity. This example explores various applications in this field, as illustrated by three distinct images. The first image presents a flowchart detailing the pathways and environmental impacts of industrial discharges and wastewater treatment effluents. It highlights the sources and potential ecological consequences of these pollutants, emphasizing the need for effective management and treatment strategies. The second image focuses on an integrated analysis of 3D chemical descriptors and gene expression data, showcasing the complex interactions between chemical compounds and biological systems. This analysis involves a comprehensive dataset of 76 VolSurf descriptors and gene activation

profiles, providing insights into the molecular mechanisms underlying toxicity. The third image depicts a diagram of causal relationships in a scientific context, illustrating how various variables interact to influence exposure and outcomes. This diagram is divided into sections that demonstrate both simple and complex causal pathways, underscoring the multifaceted nature of toxicity mechanisms. Together, these examples underscore the importance of multidisciplinary approaches in understanding and mitigating the effects of emerging pollutants on environmental health. [? lgupta2018nanoparticles,khan2011comprehensivedatadrivenanalysisimpact,kelly2022discoverymethodssystematicanalysis)

In recent years, the study of emerging pollutants has gained significant attention due to their complex nature and potential risks to environmental health. To elucidate the intricate relationships between these pollutants and their effects, Figure 8 illustrates the hierarchical structure of the key concepts related to emerging pollutants and their impact on environmental health. This figure categorizes the information into three main areas: the definition and complexity of emerging pollutants, their impacts on human health and ecosystems, and the challenges and innovative approaches for identification, regulation, and risk assessment. By providing a visual representation of these interconnected themes, the figure enhances our understanding of the multifaceted issues surrounding emerging pollutants and underscores the need for comprehensive strategies in addressing their risks.

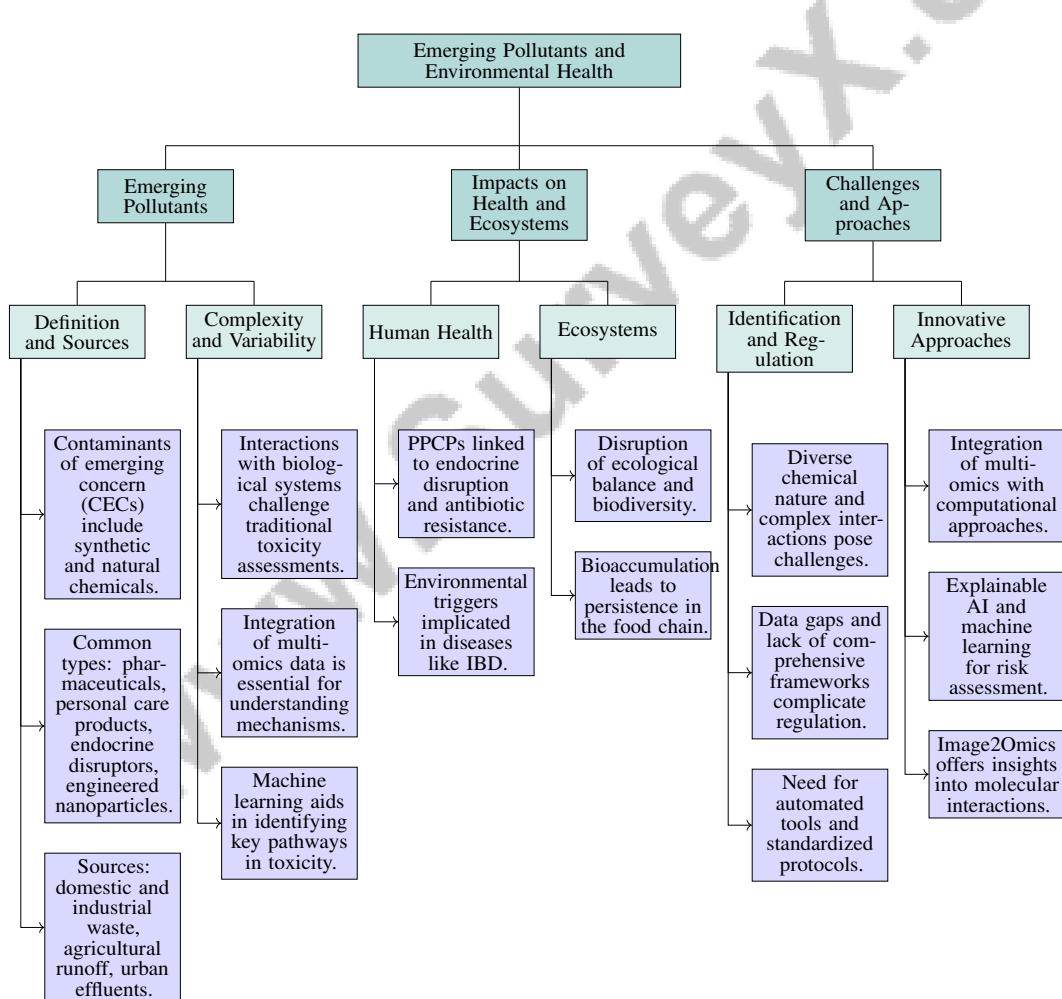


Figure 8: This figure illustrates the hierarchical structure of the key concepts related to emerging pollutants and their impact on environmental health. It categorizes the information into three main areas: the definition and complexity of emerging pollutants, their impacts on human health and ecosystems, and the challenges and innovative approaches for identification, regulation, and risk assessment.

6 Emerging Pollutants and Environmental Health

6.1 Definition and Sources of Emerging Pollutants

Emerging pollutants, known as contaminants of emerging concern (CECs), include synthetic and naturally occurring chemicals not routinely monitored but capable of entering ecosystems and posing risks to human health and ecological systems [15]. These include pharmaceuticals, personal care products, endocrine-disrupting chemicals, and engineered nanoparticles, often found in aquatic environments due to their persistence, bioaccumulation, and resistance to conventional wastewater treatment. They originate from domestic and industrial waste, agricultural runoff, and urban effluents, introduced by anthropogenic activities like industrial discharges and domestic wastewater [14]. Pharmaceuticals and personal care products (PPCPs), such as analgesics and antibiotics, are linked to adverse ecological and human health effects [1]. Despite their technological benefits, engineered nanoparticles pose significant health and environmental risks due to unique physicochemical properties and potential toxicity [15]. The complexity of emerging pollutants is heightened by their interactions within environmental matrices and biological systems, leading to unpredictable toxicological outcomes [13]. Advanced approaches, like chemical systems biology and multi-omics, are essential to unravel these intricate effects [64]. Integrating multi-omics data—genomics, transcriptomics, proteomics, metabolomics, and epigenomics—is crucial for elucidating the complex biological mechanisms underlying pollutant-induced toxicity [43]. Understanding sources, pathways, and impacts on human health and ecosystems is vital for formulating effective mitigation strategies [3].

6.2 Complexity and Variability of Exposure-Response Relationships

The study of emerging pollutants and their effects on environmental health is complicated by the inherent complexity and variability of exposure-response relationships. These pollutants, including pharmaceuticals and engineered nanoparticles, interact with biological systems in ways that challenge traditional toxicity assessment methods [13]. The heterogeneity of biological systems and the diverse nature of omics data complicate the integration and interpretation of multi-omics data for toxicity assessment [52]. Integrating various omics data types is essential for elucidating complex biological mechanisms, yet advanced computational tools are necessary to overcome challenges in data integration and interpretation [43]. Dynamic biological systems contribute to variable toxicological responses, influenced by factors like genetic diversity and environmental conditions [12]. Machine learning algorithms, such as the Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI), aid in identifying key pathways involved in toxicity mechanisms [43].

6.3 Impacts on Human Health and Ecosystems

Emerging pollutants, including pharmaceuticals, personal care products, endocrine disruptors, and engineered nanoparticles, are prevalent in environmental matrices such as water, soil, and air due to their persistence and bioaccumulation potential [65, 16]. They can profoundly affect human health by interacting with multiple biological pathways, resulting in complex exposure-response relationships and adverse health outcomes [13]. PPCPs, for instance, are linked to endocrine disruption and antibiotic resistance [1]. Environmental triggers, like pollution, have been implicated in diseases such as inflammatory bowel disease (IBD), emphasizing broader health implications [17].

This is further illustrated in Figure 9, which depicts the impacts of emerging pollutants on human health and ecosystems, highlighting key categories such as pharmaceuticals, personal care products, and endocrine disruptors. The figure also explores the health impacts, including endocrine disruption and antibiotic resistance, alongside ecosystem disruptions that affect ecological balance and biodiversity. In ecosystems, these pollutants disrupt ecological balance and biodiversity, affecting the growth, reproduction, and survival of aquatic organisms [15]. Their bioaccumulative nature allows them to persist in the food chain, leading to unpredictable toxicological outcomes [13]. The integration of multi-omics techniques has enhanced understanding of toxicity mechanisms, providing a comprehensive framework for analyzing complex molecular interactions and pathways influenced by toxic exposures [19]. By incorporating diverse omics data types, researchers can achieve a holistic view of biological systems, improving disease subtype prediction accuracy and highlighting therapeutic implications for complex diseases [12].

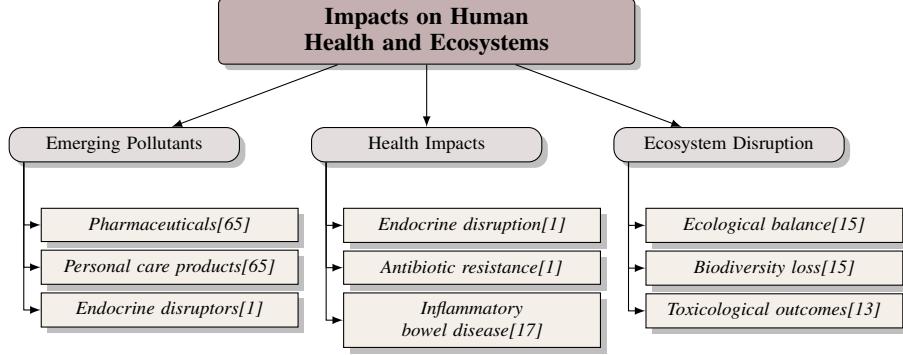


Figure 9: This figure illustrates the impacts of emerging pollutants on human health and ecosystems, highlighting key categories such as pharmaceuticals, personal care products, and endocrine disruptors. It further explores the health impacts like endocrine disruption and antibiotic resistance, and ecosystem disruptions including ecological balance and biodiversity loss.

6.4 Challenges in Identification and Regulation

Identifying and regulating emerging pollutants is challenging due to their diverse chemical nature, complex environmental interactions, and unpredictable biological effects. A primary hurdle is the noise and sparsity inherent in multi-omics studies, leading to information loss and decreased analytical accuracy, which hinders understanding multifaceted interactions between pollutants and biological systems [42, 42]. This complexity is compounded by variability in statistical distributions across different data modalities [42]. The regulatory landscape is further complicated by data gaps and the absence of comprehensive frameworks addressing the environmental presence and potential risks to human health and ecosystems [65]. The dynamic nature of emerging pollutants necessitates integrating advanced scientific approaches, such as chemical systems biology and multi-omics, to unravel their intricate biological effects and inform regulatory decision-making [20]. Developing automated tools for identifying corresponding multi-omics datasets related to defined chemicals of interest is essential for overcoming challenges associated with naming variability and data complexity [50]. Standardized protocols and methodologies are needed to ensure the reliability and reproducibility of findings, addressing the challenges posed by unpaired data and ensuring comprehensive understanding of the multifaceted interactions between emerging pollutants and biological systems [42].

6.5 Innovative Approaches for Risk Assessment

Benchmark	Size	Domain	Task Format	Metric
COMET[66]	1,000,000	Biological Data Integration	Multi-omics Evaluation	R2, F1
BioPreDyn-bench[67]	1,000,000	Metabolic Modeling	Parameter Estimation	NRMSE
MOB[68]	1,000	Survival Analysis	Survival Prediction	C-index, Integrated Brier Score
BPA-Benchmark[69]	1,000	Environmental Toxicology	Toxicity Assessment	LC50, EC50
MultiOrg[42]	63,042	Lung Organoid Detection	Object Detection	mAP@0.5IoU, F1-score

Table 1: The table presents a comprehensive overview of representative benchmarks used in the domain of biological data integration and risk assessment. It details the size, domain, task format, and evaluation metrics for each benchmark, providing a foundation for understanding their application in multi-omics evaluation, metabolic modeling, survival analysis, environmental toxicology, and lung organoid detection.

Addressing the challenges posed by emerging pollutants requires innovative approaches to risk assessment, as they often evade traditional monitoring and regulatory frameworks. Table 1 provides a detailed overview of representative benchmarks utilized in innovative approaches for risk assessment, highlighting their significance in the integration of multi-omics techniques and advanced computational methods. The integration of multi-omics techniques with advanced computational approaches, such as machine learning and network analysis, provides a powerful framework for understanding complex interactions between emerging pollutants and biological systems. Explainable

Artificial Intelligence (xAI) methods have emerged as a promising approach to analyze and interpret multi-omics data, enabling deeper insights into molecular interactions driving toxicity mechanisms and contributing to more accurate risk assessment and management strategies [43, 7]. The Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI) facilitates identifying key pathways involved in toxicity mechanisms [43]. The integration of high-content imaging with multi-omics analyses, as seen in the Image2Oomics approach, offers a comprehensive view of molecular interactions and pathways affected by toxic exposures, aiding in the identification of critical biomarkers and disease mechanisms [53]. Despite advancements, challenges remain in integrating diverse omics data for risk assessment due to the heterogeneity and complexity of biological systems and diverse omics data types [3]. Addressing these challenges requires advanced computational tools and methodologies to maximize the potential of multi-omics approaches in understanding toxicity mechanisms and developing effective strategies for mitigating impacts on human health and ecosystems.

7 Systems Biology Approaches to Toxicity Assessment

7.1 Data Integration and Computational Modeling

Systems biology employs data integration and computational modeling to explore complex biological interactions and toxicity mechanisms of emerging pollutants. This approach synthesizes genomic, transcriptomic, proteomic, metabolomic, and epigenomic data, offering a comprehensive view of biological processes and their regulation. Integrative methods enhance the analysis of biological interactions and adaptive responses to environmental stressors, refining risk assessments and intervention strategies. Multi-omics data integration elucidates toxicity pathways and adverse outcome pathways, improving biomarker identification and detection of toxicant responses, which is crucial for evaluating chemical mixtures and associated health risks [21, 49].

Advanced computational techniques, including machine learning and network-based methods, facilitate multi-omics data analysis, clarifying molecular interactions in toxicity mechanisms. The self-supervised learning framework, Self-omics, enhances data integration and interpretation, identifying critical molecular interactions affected by pollutants [43]. However, challenges such as non-standardized data formats and lack of robust verification processes remain, which can affect simulation accuracy. Future research should focus on standardizing organoid protocols and integrating these data with human tissue datasets to improve model reliability in environmental health research [2].

Innovative methodologies like the Collaborative Graphical Lasso (CoGLasso) and Multi-Omic Quantum Machine Learning (MQML-LungSC) are essential for overcoming integration challenges. These approaches enhance the accuracy and interpretability of toxicity assessments by combining diverse data sources with advanced topological data analysis techniques, such as harmonic persistent homology. This integration aids in identifying biomarkers and classifying disease subtypes, advancing our understanding of molecular mechanisms underlying toxic responses and informing therapeutic strategies [49, 12].

Explainable Artificial Intelligence (xAI) methods are incorporated into multi-omics research to address interpretability challenges in deep learning models, enhancing predictive analyses' transparency and reliability [7]. This integration of computational and biological sciences strengthens the interpretive power of complex datasets, advancing the understanding of pollutant-induced toxicity mechanisms and informing effective risk assessment and management strategies.

7.2 Network and Pathway Analysis

Network and pathway analysis are pivotal in systems biology, enhancing our understanding of molecular mechanisms underlying toxicity and environmental health responses. By integrating multi-omics data, researchers gain a comprehensive view of biological responses to chemical exposures, revealing complex biomolecular interactions and regulatory pathways. This approach facilitates biomarker identification and elucidates causal relationships and time-dependent responses, leading to effective risk assessments and targeted toxicology interventions [70, 49, 71].

The convergence of network science and systems biology has fostered innovative computational tools that deepen our understanding of biological processes and their responses to environmental stressors

[24]. Network-based methods, such as MOD-Finder, utilize topological data analysis techniques to identify key molecular interactions and pathways disrupted by toxic exposures, offering insights into toxicity's molecular foundations [18]. These methods identify critical nodes and edges impacted by pollutants, supporting the development of targeted interventions to mitigate adverse effects on human health and ecosystems [12].

Integrating multi-omics data into network and pathway analysis enhances systems biology's capacity to elucidate toxicity mechanisms [3]. By incorporating diverse omics data types, researchers gain a comprehensive understanding of molecular pathways and networks affected by toxic exposures, revealing key molecular interactions disrupted by emerging pollutants. Advanced computational methods, including network-based techniques and topological data analysis, capture the dynamic regulation of biological processes, considering single-cell heterogeneity and tissue spatial distribution, which are vital for understanding molecular underpinnings of toxicity [12, 23].

7.3 Dynamic Systems and Predictive Modeling

Dynamic systems and predictive modeling are integral to systems biology, offering frameworks for understanding complex, time-dependent biological interactions and toxicity mechanisms. These models simulate biological processes, predicting cellular responses to environmental stressors and integrating diverse data types to enhance toxicity predictions and mechanistic interpretations. Techniques like Time Course Network Enrichment (TiCoNE) facilitate time-dependent biological response analysis, while Bayesian uncertainty analysis addresses challenges associated with high-dimensional parameter spaces in model evaluation [27, 25, 4].

Advanced computational approaches, including machine learning and network analysis, enhance dynamic systems and predictive modeling's capacity to analyze complex biological interactions and predict toxic exposure outcomes. Machine learning algorithms, such as the Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI), enrich our understanding of disease mechanisms by identifying key molecular pathways and networks affected by toxic exposures. Integrating Bayesian uncertainty analysis into dynamic systems modeling accounts for variability and uncertainty from multiple sources, bolstering predictive model reliability [27].

Systems biology approaches, integrating experimental and computational methods, have advanced dynamic systems and predictive modeling in toxicity assessment [25]. Computational Fluid Dynamics (CFD) techniques, for instance, integrate mass transfer models to simulate complex biological interactions under toxic exposures [26]. These approaches provide a comprehensive understanding of molecular mechanisms underlying toxicity, capturing dynamic interactions within biological systems and identifying key pathways affected by toxic exposures.

7.4 Technological Integration and Methodologies

Advanced technologies and methodologies are pivotal for advancing systems biology, particularly in toxicity assessment for emerging pollutants. Systems biology employs computational and experimental techniques, including multi-omics integration, to enhance our understanding of molecular mechanisms underlying pollutant-induced toxicity. This comprehensive approach identifies regulatory pathway responses and facilitates personalized toxicity predictions based on patient-specific data [49, 71, 72].

Network science and topological data analysis-based computational tools significantly enhance systems biology's ability to elucidate toxicity's molecular mechanisms. For instance, the Collaborative Graphical Lasso (CoGLasso) method provides insights into complex biological interactions by integrating local and global structural information, improving predictive power for synthetic data and facilitating biomarker discovery [73]. The Multi-Omics Sampling-Based Graph Transformer (MSGT-SL) enhances interpretive power by combining diverse omics data, while the TiCoNE framework identifies biologically relevant regions in networks based on temporal expression patterns [71, 74].

Machine learning and predictive modeling applications in systems biology further enhance multi-omics data interpretation, identifying critical molecular interactions and pathways involved in toxicity mechanisms [43]. The development of self-supervised learning frameworks, such as Self-omics, enhances multi-omics data integration and interpretation, identifying critical molecular interactions and pathways involved in toxicity mechanisms [43].

The integration of advanced imaging and analysis techniques, such as SM-Omics and the SegmentAnything approach, enhances organoids' utility as models for studying human disease and drug interactions. These methodologies empower researchers to gain insights into organogenesis, disease mechanisms, and potential therapeutic interventions, contributing to improved environmental health outcomes [32, 21, 20, 11]. Continued refinement of these methodologies and their integration with systems biology approaches are essential for advancing toxicity assessment and management strategies.

8 Integration of Organoids and Multi-Omics in Toxicity Assessment

8.1 Enhancing Organoid Models with Multi-Omics Data

Integrating multi-omics data with organoid technology has significantly advanced environmental health research by elucidating cellular and molecular toxicity mechanisms. Organoids, derived from stem cells, serve as sophisticated three-dimensional models that closely mimic human tissues, providing invaluable insights into the effects of emerging pollutants on human health [5]. Multi-omics techniques enhance organoids' capacity to replicate complex biological processes, offering a multidimensional perspective on how environmental pollutants impact human health. By incorporating genomics, transcriptomics, proteomics, metabolomics, and epigenomics, researchers gain a comprehensive understanding of disrupted molecular interactions and pathways, crucial for clarifying pollutant-induced toxicity mechanisms and improving environmental health outcomes [3].

Recent computational advancements have streamlined the integration of multi-omics data with organoid models. Network-based approaches and machine learning algorithms facilitate the analysis of large-scale datasets, identifying critical molecular interactions and pathways influenced by environmental pollutants. For instance, the Machine Learning Pipeline for Discriminant Pathways Identification (MLP-DPI) has been instrumental in pinpointing key gene interactions affected by toxic exposures, offering novel insights into toxicity mechanisms [51]. Standardized multi-omics datasets, such as gene expression profiles, enhance organoid model analysis in environmental health research, allowing precise examinations of disrupted molecular pathways [75]. Innovative methodologies like the Self-omics framework, which employs self-supervised learning, further enhance data integration and interpretation, identifying critical molecular interactions in toxicity mechanisms [43]. This interdisciplinary approach deepens our understanding of the complex interactions between pollutants and biological systems, leading to improved environmental health outcomes.

8.2 Applications in Environmental Health and Emerging Pollutants

The integration of organoids with multi-omics approaches represents a significant advancement in environmental health research, providing innovative methods to investigate the impacts of emerging pollutants on human health and ecosystems. Organoids, derived from human pluripotent stem cells (hPSCs) and adult stem cells (AdSCs), accurately replicate human tissue architecture and function [5]. Multi-omics techniques integrated with organoid models enhance our ability to elucidate the molecular mechanisms driving pollutant toxicity. By harnessing transcriptomics, proteomics, and metabolomics, researchers gain a comprehensive understanding of affected biological pathways, facilitating the identification of critical biomolecular interactions and feedback loops. This enhances pathway detection and enables more accurate toxic effect predictions, enriching insights into pollutant-induced toxicity and supporting more effective risk assessment strategies [6, 49, 50, 72].

Recent computational advancements have augmented the integration of multi-omics data with organoid models, providing unprecedented insights into pollutant-induced toxicity mechanisms [43]. Machine learning algorithms and network-based approaches facilitate extensive dataset interpretation, yielding novel insights into disrupted pathways and interactions [12]. Self-supervised learning frameworks like Self-omics enhance data integration and interpretation, identifying crucial molecular interactions and pathways involved in toxicity mechanisms [43]. These integrated methodologies in environmental health research hold significant potential for advancing our understanding of interactions between pollutants and biological systems. Leveraging organoids' physiological relevance with multi-omics insights allows researchers to devise more effective risk assessment and management strategies to mitigate emerging pollutants' adverse effects on human health and ecosystems. Future research may focus on improving model efficiency and exploring further applications of multi-omics data integration, ultimately contributing to enhanced environmental health outcomes [76].

The availability of standardized multi-omics datasets, including gene expression profiles, has been critical for refining computational models and enhancing organoid study accuracy in environmental health research [75]. This interdisciplinary approach underscores the necessity for collaboration among researchers across diverse fields to effectively address environmental pollutants' complexities and their implications for human health and ecosystems. By refining these methodologies and integrating them with systems biology approaches, researchers can further advance our understanding of the molecular mechanisms underlying pollutant-induced toxicity, ultimately contributing to improved environmental health outcomes and public health protection.

9 Conclusion

9.1 Challenges and Future Directions

The complex interdisciplinary domain involving organoids, multi-omics, emerging pollutants, toxicity assessment, environmental health, and systems biology presents numerous challenges that necessitate innovative strategies moving forward. A key challenge is the effective integration and standardization of multi-omics data, which are inherently high-dimensional and heterogeneous. The various omics layers, such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics, require advanced data integration algorithms and computational tools to manage these intricate datasets effectively. The issue of unpaired data further complicates this task, often resulting in information loss and diminished analytical precision.

Future research should focus on improving model interpretability and integrating clinical phenotypes with single-cell multi-omics technologies to enhance analytical precision and address the inherent complexities of this field. Developing more robust computational models for multi-omic integration is crucial for understanding the molecular mechanisms driving pollutant-induced toxicity, which involves optimizing models for larger datasets, exploring additional downstream tasks, and enhancing the interpretability of learned representations.

In the realm of organoid research, optimizing these sophisticated biological models is essential for their application in environmental health studies. Despite their capability for high-fidelity replication of human tissues, organoids face challenges such as incomplete maturation and a lack of essential microenvironmental features, limiting their effectiveness in modeling adult-onset diseases and understanding the complexities of human tissues. Future efforts should focus on standardizing toxicity assessment methods and exploring the long-term impacts of emerging pollutants on organoid models. Moreover, developing co-culture systems that incorporate immune cells and vascular structures, along with improvements in *in vivo* models, is crucial for enhancing the physiological relevance of organoid models in environmental health research.

The application of systems biology approaches, which combine experimental and computational methods, has been pivotal in advancing the understanding of dynamic systems and predictive modeling in toxicity assessment. Incorporating network-based approaches and topological data analysis techniques can aid in identifying critical molecular interactions and pathways disrupted by emerging pollutants, thereby deepening our understanding of the molecular mechanisms underlying toxicity.

Furthermore, integrating multi-omics data with systems biology approaches offers a comprehensive framework for modeling complex biological interactions and identifying critical molecular pathways and networks affected by toxic exposures. Future research will also explore various graph neural network architectures to improve scalability and investigate unsupervised learning approaches for cancer subtype detection.

References

- [1] Nadia Morin-Crini, Eric Lichtfouse, Marc Fourmentin, Ana Rita Lado Ribeiro, Constantinos Noutsopoulos, Francesca Mapelli, Éva Fenyvesi, Melissa Gurgel Adeodato Vieira, Lorenzo A Picos-Corrales, Juan Carlos Moreno-Piraján, et al. Removal of emerging contaminants from wastewater using advanced treatments. a review. *Environmental Chemistry Letters*, 20(2):1333–1375, 2022.
- [2] Christoph Bock, Michael Boutros, J Gray Camp, Laura Clarke, Hans Clevers, Juergen A Knoblich, Prisca Liberali, Aviv Regev, Anne C Rios, Oliver Stegle, et al. The organoid cell atlas. *Nature biotechnology*, 39(1):13–17, 2021.
- [3] Guillermo de Anda-Jáuregui and Enrique Hernández-Lemus. Computational oncology in the multi-omics era: state of the art. *Frontiers in oncology*, 10:423, 2020.
- [4] François Bertaux, Dirk Drasdo, and Gregory Batt. System modeling of receptor-induced apoptosis, 2017.
- [5] Xiao-Yan Tang, Shanshan Wu, Da Wang, Chu Chu, Yuan Hong, Mengdan Tao, Hao Hu, Min Xu, Xing Guo, and Yan Liu. Human organoids in basic research and clinical applications. *Signal transduction and targeted therapy*, 7(1):168, 2022.
- [6] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome biology*, 18:1–15, 2017.
- [7] Ahmad Hussein, Mukesh Prasad, and Ali Braytee. Explainable ai methods for multi-omics analysis: A survey, 2024.
- [8] SeoWon Gabriel Choi and Sungkyu Jung. Integrative decomposition of multi-source data by identifying partially-joint score subspaces, 2023.
- [9] Ya Ren, Xue Yang, Zhengjiang Ma, Xin Sun, Yuxin Zhang, Wentao Li, Han Yang, Lei Qiang, Zezheng Yang, Yihao Liu, et al. Developments and opportunities for 3d bioprinted organoids. *International Journal of Bioprinting*, 7(3):364, 2021.
- [10] Oliver L Eichmüller and Juergen A Knoblich. Human cerebral organoids—a new tool for clinical neurology research. *Nature Reviews Neurology*, 18(11):661–680, 2022.
- [11] Elizabeth Di Lullo and Arnold R Kriegstein. The use of brain organoids to investigate neural development and disease. *Nature Reviews Neuroscience*, 18(10):573–584, 2017.
- [12] Davide Gurnari, Aldo Guzmán-Sáenz, Filippo Utro, Aritra Bose, Saugata Basu, and Laxmi Parida. Probing omics data via harmonic persistent homology, 2024.
- [13] Rob Trangucci, Jesse Contreras, Jon Zelner, Joseph N. S. Eisenberg, and Yang Chen. Bayesian methods for modeling cumulative exposure to extensive environmental health hazards, 2024.
- [14] Elizabeth A. Gibson, Junhui Zhang, Jingkai Yan, Lawrence Chillrud, Jaime Benavides, Yanelli Nunez, Julie B. Herbstman, Jeff Goldsmith, John Wright, and Marianthi-Anna Kioumourtzoglou. Principal component pursuit for pattern identification in environmental mixtures, 2021.
- [15] Suhas P Dharupaneedi, Sanna Kotrappanavar Nataraj, Mallikarjuna Nadagouda, Kakarla Raghava Reddy, Shyam S Shukla, and Tejraj M Aminabhavi. Membrane-based separation of potential emerging pollutants. *Separation and Purification Technology*, 210:850–866, 2019.
- [16] Ionela Cătălina Vasilachi, Dana Mihaela Asiminicesei, Daniela Ionela Fertu, and Maria Gavrilescu. Occurrence and fate of emerging pollutants in water environment and options for their removal. *Water*, 13(2):181, 2021.
- [17] Ashwin N Ananthakrishnan, Charles N Bernstein, Dimitrios Iliopoulos, Andrew Macpherson, Markus F Neurath, Raja A Raja Ali, Stephan R Vavricka, and Claudio Fiocchi. Environmental triggers in ibd: a review of progress and evidence. *Nature reviews Gastroenterology & hepatology*, 15(1):39–49, 2018.

-
- [18] A Murat Eren, Evan Kiefl, Alon Shaiber, Iva Veseli, Samuel E Miller, Matthew S Schechter, Isaac Fink, Jessica N Pan, Mahmoud Yousef, Emily C Fogarty, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nature microbiology*, 6(1):3–6, 2021.
- [19] Xiaojing Chu, Bowen Zhang, Valerie ACM Koeken, Manoj Kumar Gupta, and Yang Li. Multi-omics approaches in immunological research. *Frontiers in Immunology*, 12:668045, 2021.
- [20] Jonathan Boss, Wei Hao, Amber Cathey, Barrett M. Welch, Kelly K. Ferguson, John D. Meeker, Jian Kang, and Bhramar Mukherjee. Mediation with external summary statistic information (messi), 2023.
- [21] Glen McGee, Ander Wilson, Brent A Coull, and Thomas F Webster. Integrating biological knowledge in kernel-based analyses of environmental mixtures and health, 2022.
- [22] Danielle J Currie, Carl Smith, and Paul Jagals. The application of system dynamics modelling to environmental health decision-making and policy-a scoping review. *BMC public health*, 18:1–11, 2018.
- [23] Matthew S Hall, Joseph T Decker, and Lonnie D Shea. Towards systems tissue engineering: elucidating the dynamics, spatial coordination, and individual cells driving emergent behaviors, 2019.
- [24] Sabrina Maniscalco, Elsi-Mari Borrelli, Daniel Cavalcanti, Caterina Foti, Adam Glos, Mark Goldsmith, Stefan Knecht, Keijo Korhonen, Joonas Malmi, Anton Nykänen, Matteo A. C. Rossi, Harri Saarinen, Boris Sokolov, N. Walter Talarico, Jussi Westergren, Zoltán Zimborás, and Guillermo García-Pérez. Quantum network medicine: rethinking medicine with network science and quantum algorithms, 2022.
- [25] Pedro Zuidberg Dos Martires, Vincent Derkinderen, Luc De Raedt, and Marcus Krantz. Automated reasoning in systems biology: a necessity for precision medicine, 2024.
- [26] Weidong Huang, Chundu Wu, Bingjia Xiao, and Weidong Xia. Computational fluid dynamic approach for biological system modeling, 2005.
- [27] Ian Vernon, Junli Liu, Michael Goldstein, James Rowe, Jen Topping, and Keith Lindsey. Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions, 2018.
- [28] Yotam E Bar-Ephraim, Kai Kretzschmar, and Hans Clevers. Organoids in immunological research. *Nature Reviews Immunology*, 20(5):279–293, 2020.
- [29] Tao Zhang, Sarthak Gupta, Madeline A. Lancaster, and J. M. Schwarz. How human-derived brain organoids are built differently from brain organoids derived from genetically-close relatives: A multi-scale hypothesis, 2025.
- [30] Helena Coggan, Clare E. Weeden, Philip Pearce, Mohit P. Dalwadi, Alastair Magness, Charles Swanton, and Karen M. Page. An agent-based modelling framework to study growth mechanisms in egfr-l858r mutant alveolar type ii cells, 2024.
- [31] Moritz Hofer and Matthias P Lutolf. Engineering organoids. *Nature Reviews Materials*, 6(5):402–420, 2021.
- [32] Giuliana Rossi, Andrea Manfrin, and Matthias P Lutolf. Progress and potential in organoid research. *Nature Reviews Genetics*, 19(11):671–687, 2018.
- [33] Anne C Rios and Hans Clevers. Imaging organoids: a bright future ahead. *Nature methods*, 15(1):24–26, 2018.
- [34] Giorgia Quadrato, Tuan Nguyen, Evan Z Macosko, John L Sherwood, Sung Min Yang, Daniel R Berger, Natalie Maria, Jorg Scholvin, Melissa Goldman, Justin P Kinney, et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*, 545(7652):48–53, 2017.

-
- [35] Abed AlFatah Mansour, J Tiago Gonçalves, Cooper W Bloyd, Hao Li, Sarah Fernandes, Daphne Quang, Stephen Johnston, Sarah L Parylak, Xin Jin, and Fred H Gage. An in vivo model of functional and vascularized human brain organoids. *Nature biotechnology*, 36(5):432–441, 2018.
- [36] Nicole Prior, Patricia Inacio, and Meritxell Huch. Liver organoids: from basic research to therapeutic applications. *Gut*, 68(12):2228–2237, 2019.
- [37] Michal Krassowski, Vivek Das, Sangram K Sahu, and Biswapriya B Misra. State of the field in multi-omics research: from computational needs to data mining and sharing. *Frontiers in Genetics*, 11:610798, 2020.
- [38] Hakim Benkirane, Yoann Pradat, Stefan Michiels, and Paul-Henry Cournède. Customics: A versatile deep-learning based strategy for multi-omics integration, 2022.
- [39] Sina Tabakhi, Mohammad Naimul Islam Suvon, Pegah Ahadian, and Haiping Lu. Multimodal learning for multi-omics: A survey, 2022.
- [40] Youjin Hu, Qin An, Katherine Sheu, Brandon Trejo, Shuxin Fan, and Ying Guo. Single cell multi-omics technology: methodology and application. *Frontiers in cell and developmental biology*, 6:28, 2018.
- [41] Xiaodan Xing, Chunling Tang, Yunzhe Guo, Nicholas Kurniawan, and Guang Yang. Segmentation anything helps microscopy images based automatic and quantitative organoid detection and analysis, 2024.
- [42] Christina Bukas, Harshavardhan Subramanian, Fenja See, Carina Steinchen, Ivan Ezhov, Gowtham Boosarpu, Sara Asgharpour, Gerald Burgstaller, Mareike Lehmann, Florian Kofler, and Marie Piraud. Multiorg: A multi-rater organoid-detection dataset, 2024.
- [43] Sayed Hashim, Karthik Nandakumar, and Mohammad Yaqub. Self-omics: A self-supervised learning framework for multi-omics cancer data, 2022.
- [44] Pengcheng Zeng and Zhixiang Lin. sciml: Information-theoretic co-clustering-based multi-view learning for the integrative analysis of single-cell multi-omics data, 2022.
- [45] Asim Waqas, Aakash Tripathi, Sabeen Ahmed, Ashwin Mukund, Hamza Farooq, Matthew B. Schabath, Paul Stewart, Mia Naeini, and Ghulam Rasool. Self-normalizing foundation model for enhanced multi-omics data analysis in oncology, 2024.
- [46] Xiaoyu Zhang, Yuting Xing, Kai Sun, and Yike Guo. Omiembed: a unified multi-task deep learning framework for multi-omics data, 2021.
- [47] Mandeep Kaur Saggi, Amadeep Singh Bhatia, Mensah Isaiah, Humaira Gowher, and Sabre Kais. Multi-omic and quantum machine learning integration for lung subtypes classification, 2024.
- [48] Vinay Kumar, Neha Sharma, and SS Maitra. In vitro and in vivo toxicity assessment of nanoparticles. *International Nano Letters*, 7(4):243–256, 2017.
- [49] Sebastian Canzler, Jana Schor, Wibke Busch, Kristin Schubert, Ulrike E Rolle-Kampczyk, Hervé Seitz, Hennicke Kamp, Martin von Bergen, Roland Buesen, and Jörg Hackermüller. Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology*, 94:371–388, 2020.
- [50] S. Canzler, J. Hackermüller, and J. Schor. Mod-finder: Identify multi-omics data sets related to defined chemical exposure, 2019.
- [51] Annalisa Barla, Giuseppe Jurman, Roberto Visintainer, Margherita Squillario, Michele Filosi, Samantha Riccadonna, and Cesare Furlanello. A machine learning pipeline for discriminant pathways identification, 2011.
- [52] Lana X Garmire. Strategies to integrate multi-omics data for patient survival prediction, 2020.

-
- [53] Rahil Mehrizi, Arash Mehrjou, Maryana Alegro, Yi Zhao, Benedetta Carbone, Carl Fishwick, Johanna Vappiani, Jing Bi, Siobhan Sanford, Hakan Keles, Marcus Bantscheff, Cuong Nguyen, and Patrick Schwab. Multi-omics prediction from high-content cellular imaging with deep learning, 2024.
- [54] You Wu and Lei Xie. Ai-driven multi-omics integration for multi-scale predictive modeling of causal genotype-environment-phenotype relationships, 2024.
- [55] Léa Maitre, Jean-Baptiste Guimbaud, Charline Warembourg, Nuria Güil-Oumrait, The Exposome Data Challenge Participant Consortium, Paula Marcela Petrone, Marc Chadeau-Hyam, Martine Vrijheid, Juan R. Gonzalez, and Xavier Basagaña. State-of-the-art methods for exposure-health studies: results from the exposome data challenge event, 2022.
- [56] Valeria Dulio, Bert van Bavel, Eva Brorström-Lundén, Joop Harmsen, Juliane Hollender, Martin Schlabach, Jaroslav Slobodník, Kevin Thomas, and Jan Koschorreck. Emerging pollutants in the eu: 10 years of norman in support of environmental policies and regulations. *Environmental Sciences Europe*, 30:1–13, 2018.
- [57] Stefan Stanojevic, Yijun Li, and Lana X. Garmire. Computational methods for single-cell multi-omics integration and alignment, 2022.
- [58] Katherine H. Shutta, Deborah Weighill, Rebekka Burkholz, Marouen Ben Guebila, Dawn L. DeMeo, Helena U. Zacharias, John Quackenbush, and Michael Altenbuchinger. Dragon: Determining regulatory associations using graphical models on multi-omic networks, 2022.
- [59] Chen Zhao, Anqi Liu, Xiao Zhang, Xuewei Cao, Zhengming Ding, Qiuying Sha, Hui Shen, Hong-Wen Deng, and Weihua Zhou. Clcls: Cross-omics linked embedding with contrastive learning and self attention for multi-omics integration with incomplete multi-omics data, 2023.
- [60] Stefan Graw, Kevin Chappell, Charity L Washam, Allen Gies, Jordan Bird, Michael S Robeson, and Stephanie D Byrum. Multi-omics data integration considerations and study design for biological systems and disease. *Molecular omics*, 17(2):170–185, 2021.
- [61] Ritu Gupta and Huan Xie. Nanoparticles in daily life: applications, toxicity and regulations. *Journal of Environmental Pathology, Toxicology and Oncology*, 37(3), 2018.
- [62] Suleiman A. Khan, Ali Faisal, John Patric Mpindi, Juuso A. Parkkinen, Tuomo Kalliokoski, Antti Poso, Olli P. Kallioniemi, Krister Wennerberg, and Samuel Kaski. Comprehensive data-driven analysis of the impact of chemoinformatic structure on the genome-wide biological response profiles of cancer cells to 1159 drugs, 2011.
- [63] Jack Kelly, Carlo Berzuini, Bernard Keavney, Maciej Tomaszewski, and Hui Guo. Discovery methods for systematic analysis of causal molecular networks in modern omics datasets, 2022.
- [64] Alessandra Cabassi, Denis Seyres, Mattia Frontini, and Paul D. W. Kirk. Two-step penalised logistic regression for multi-omic data with an application to cardiometabolic syndrome, 2020.
- [65] John Wilkinson, Peter S Hooda, James Barker, Stephen Barton, and Julian Swinden. Occurrence, fate and transformation of emerging contaminants in water: An overarching review of the field. *Environmental Pollution*, 231:954–970, 2017.
- [66] Yuchen Ren, Wenwei Han, Qianyuan Zhang, Yining Tang, Weiqiang Bai, Yuchen Cai, Lifeng Qiao, Hao Jiang, Dong Yuan, Tao Chen, Siqi Sun, Pan Tan, Wanli Ouyang, Nanqing Dong, Xinzhu Ma, and Peng Ye. Comet: Benchmark for comprehensive biological multi-omics evaluation tasks and language models, 2024.
- [67] Alejandro F Villaverde, David Henriques, Kieran Smallbone, Sophia Bongard, Joachim Schmid, Damjan Cicin-Sain, Anton Crombach, Julio Saez-Rodriguez, Klaus Mauch, Eva Balsa-Canto, Pedro Mendes, Johannes Jaeger, and Julio R Banga. Biopredyn-bench: benchmark problems for kinetic modelling in systems biology, 2014.
- [68] Moritz Herrmann, Philipp Probst, Roman Hornung, Vindi Jurinovic, and Anne-Laure Boulesteix. Large-scale benchmark study of survival prediction methods using multi-omics data, 2020.

-
- [69] John Moreman, Okhyun Lee, Maciej Trznadel, Arthur David, Tetsuhiro Kudoh, and Charles R Tyler. Acute toxicity, teratogenic, and estrogenic effects of bisphenol a and its alternative replacements bisphenol s, bisphenol f, and bisphenol af in zebrafish embryo-larvae. *Environmental science & technology*, 51(21):12796–12805, 2017.
 - [70] Azam Yazdani, Akram Yazdani, Philip L. Lorenzi, and Ahmad Samiei. Integrated systems approach identifies pathways from the genome to triglycerides through a metabolomic causal network, 2018.
 - [71] Christian Wiwie, Alexander Rauch, Anders Haakonsson, Inigo Barrio-Hernandez, Blagoy Blagoev, Susanne Mandrup, Richard Röttger, and Jan Baumbach. Elucidation of time-dependent systems biology cell response patterns with time course network enrichment, 2017.
 - [72] Narsis A. Kiani, Ming-Mei Shang, Hector Zenil, and Jesper Tegnér. Predictive systems toxicology, 2018.
 - [73] Alessio Albanese, Wouter Kohlen, and Pariya Behrouzi. Collaborative graphical lasso, 2024.
 - [74] Xusheng Zhao, Hao Liu, Qiong Dai, Hao Peng, Xu Bai, and Huailiang Peng. Multi-omics sampling-based graph transformer for synthetic lethality prediction, 2023.
 - [75] Tony Hauptmann and Stefan Kramer. A fair experimental comparison of neural network architectures for latent representations of multi-omics for drug response prediction, 2022.
 - [76] Arman Hasanzadeh, Ehsan Hajiramezanali, Nick Duffield, and Xiaoning Qian. Morel: Multi-omics relational learning, 2022.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

www.SurveyX.Cn