A Critical Review of the Role of Molecular Epidemiology in Identifying Risk Factors for Chronic Diseases

Muhammad Edo Tanayah[1]*, Dinda Aziza Bryllianna[2]
[1*]Universitas Negeri Surabaya, Indonesia
[2]Universitas Ciputra Surabaya, Indonesia
Email Corresponding Author: edotanayahhits@gmail.com

ABSTRACT

This study aims to provide a critical review of the role of molecular epidemiology in identifying chronic disease risk factors, emphasizing how molecular integration enhances the precision and validity of epidemiological analyses. Employing a qualitative descriptive approach through a systematic literature review, this research collected and analyzed data from peer-reviewed scientific articles, books, and official reports published between 2015 and 2025. Data were examined through document analysis, thematic coding, and inductive synthesis, allowing the identification of core themes and conceptual relationships within the existing body of knowledge. The findings reveal that molecular epidemiology bridges the gap between traditional population-based approaches and molecular biology by integrating genomic, metabolomic, and environmental data to uncover the biological mechanisms underlying chronic diseases such as cancer, diabetes, and cardiovascular disorders. Furthermore, multiomic profiling and machine learning models have improved risk prediction accuracy, clarified gene–environment interactions, and enabled the classification of molecular disease subtypes. However, challenges remain in biomarker validation, data standardization, and clinical translation. This study concludes that molecular epidemiology contributes significantly to the advancement of precision medicine by promoting more personalized prevention and intervention strategies. Its theoretical and practical implications extend to public health, data science, and biomedical research, underscoring the need for interdisciplinary collaboration and equitable access to molecular technologies in global health contexts.

Keywords: molecular epidemiology, chronic diseases, gene–environment interaction, multiomics, precision medicine.

INTRODUCTION

In the last few decades, chronic diseases such as diabetes, cardiovascular disorders, and cancer have become the dominant global health burden. The World Health Organization (WHO) reports that more than 70% of global deaths annually are caused by chronic non-communicable diseases, with a steady increase in prevalence due to aging populations and lifestyle transitions worldwide. Traditional epidemiological approaches have successfully identified behavioral and environmental risk factors; however, they often fall short in uncovering the underlying biological mechanisms that link these exposures to disease outcomes (1).

The emergence of molecular epidemiology represents a paradigm shift, bridging the gap between population-level risk assessments and molecular-level biological understanding. By integrating genomics, transcriptomics, metabolomics, and other “omics” technologies, this field allows for more precise identification of disease pathways and subtypes (2). Such integration is crucial to unravel complex gene-environment interactions that underpin chronic disease etiology (3).

Recent advances in multiomic profiling have demonstrated that individuals with similar clinical phenotypes can possess distinct molecular signatures, leading to heterogeneous responses to interventions and therapies (4). This finding highlights the limitations of one-size-fits-all public health strategies and underscores the need for molecularly informed approaches to prevention and treatment.

Moreover, the increased accessibility of big data and high-throughput sequencing technologies has accelerated the discovery of biomarkers predictive of chronic disease risk. These molecular biomarkers enhance early

disease detection, stratify patient risk, and inform targeted prevention strategies (5). Yet, despite these advances, the translation of molecular findings into clinical or public health practice remains challenging due to issues of data heterogeneity, reproducibility, and validation (6) .

The integration of artificial intelligence and machine learning into molecular epidemiology further amplifies its potential. These tools enable the analysis of complex, high-dimensional datasets to uncover non-linear associations and causal relationships that traditional statistical models might overlook (7). For instance, Bayesian network approaches have been applied to model causal interactions among risk factors and disease phenotypes, improving etiological inference and predictive accuracy (8).

Nevertheless, the adoption of these computational models requires careful consideration of algorithmic bias and ethical challenges, particularly in ensuring equitable health outcomes across populations (9). The responsible application of machine learning within molecular epidemiology necessitates transparent methods and the inclusion of diverse population datasets to avoid reinforcing health disparities.

As the field evolves, molecular pathological epidemiology (MPE) has emerged as an integrative discipline combining molecular biology, pathology, and epidemiology. This approach provides a holistic framework for studying disease mechanisms, identifying distinct molecular subtypes, and tailoring prevention strategies (1). Through such interdisciplinary integration, molecular epidemiology is redefining the landscape of chronic disease research and management.

The urgency of advancing molecular epidemiology stems from the global escalation of chronic diseases that impose tremendous economic and social costs. Current evidence suggests that molecular-level insights could inform more effective, personalized interventions, thereby mitigating disease burden and improving population health outcomes (10).

Despite its promise, molecular epidemiology faces key methodological and infrastructural challenges, including the need for standardized protocols, robust data-sharing frameworks, and the establishment of large, longitudinal multiomic databases (11,12). Addressing these challenges is essential to harness the full potential of molecular data in elucidating causal mechanisms of chronic diseases.

Additionally, the complexity of multiomic datasets necessitates novel bioinformatics pipelines capable of integrating diverse data types—from genetic variants and metabolomic profiles to microbiome signatures (13). Such integration will enable the discovery of cross-system interactions that shape individual disease susceptibility.

The expanding body of research also emphasizes the influence of immunological and microbiome factors in modulating chronic disease risk, signaling a shift toward systems-level understanding of pathogenesis (13). This aligns with the ongoing transformation of epidemiology from a descriptive science to a mechanistic one, capable of informing precision public health strategies.

Furthermore, longitudinal molecular epidemiology studies, such as those analyzing body mass index trajectories and metabolic profiles, have illuminated dynamic disease risk patterns over time, providing valuable insight for early prevention efforts (14). Such temporal analyses are vital to identify windows of opportunity for intervention before irreversible pathological changes occur.

The theoretical significance of this article lies in synthesizing diverse strands of evidence to provide a critical appraisal of molecular epidemiology's role in identifying chronic disease risk factors. Practically, this synthesis aims to guide future research toward developing scalable, ethically sound, and clinically applicable frameworks for disease prevention and health promotion.

In summary, this article critically examines the evolving contributions, challenges, and future directions of molecular epidemiology in identifying risk factors for chronic diseases. It seeks to elucidate how the convergence of molecular data, computational analytics, and epidemiological methods can revolutionize the understanding and management of chronic diseases in the 21st century.

METHOD

This study employed a qualitative research design with a descriptive approach through a library-based study (literature review). The qualitative-descriptive design was selected to allow a deep and contextualized understanding of the role of molecular epidemiology in identifying risk factors for chronic diseases. This method emphasizes systematic exploration of scholarly sources to synthesize existing findings and theories relevant to the topic (15,16). The approach also prioritizes the researcher's interpretative role in analyzing and connecting concepts derived from diverse academic materials.

The data sources of this study consisted of peer-reviewed journal articles, books, and official scientific reports focusing on molecular epidemiology, chronic disease risk factors, and methodological frameworks in health sciences. The selection of literature was guided by inclusion criteria that prioritized sources published from 2015 onward, available in reputable scientific databases, and providing conceptual or empirical contributions to the understanding of molecular epidemiology (17,18). Exclusion criteria involved non-academic materials, outdated publications, and sources lacking methodological transparency.

Data collection was carried out through systematic literature searching and document analysis. Relevant academic papers were identified via keyword searches in digital databases such as PubMed, Scopus, and ScienceDirect using combinations of terms like "molecular epidemiology," "chronic disease," and "risk factor identification." The gathered materials were then reviewed for methodological rigor, relevance to the study objectives, and credibility. Each selected document was read critically to extract essential information related to theoretical foundations, research trends, and analytical techniques (19,20).

The data analysis process followed an inductive and thematic model to ensure conceptual depth and interpretive validity. The analysis began with data identification and familiarization, followed by data reduction through coding and categorization of emerging themes. These themes included molecular markers, gene-environment interactions, computational modeling, and multiomic integration. The next stage involved thematic categorization and synthesis of findings into coherent conceptual frameworks. Finally, conclusions were drawn inductively to highlight critical patterns, relationships, and research implications (21–23).

To ensure data validity and reliability, this study applied triangulation of sources and conceptual peer review. Triangulation was achieved by comparing findings across multiple independent studies, ensuring consistency and accuracy of interpretation (15). Peer conceptual review involved discussing interpretations and categorizations with other researchers and aligning them with existing theoretical frameworks to enhance analytical credibility (24). Additionally, transparency and traceability of the analytical process were maintained through detailed documentation, following the principles of audit trail methodology (16).

Through this qualitative-descriptive and literature-based methodology, the study was able to construct a comprehensive and critically informed understanding of how molecular epidemiology contributes to identifying chronic disease risk factors. This approach provided a valid synthesis of multidisciplinary evidence, ensuring theoretical relevance and practical applicability. The chosen method not only reinforced the reliability of the conclusions but also allowed nuanced insights into methodological innovations and gaps in current research (25–27).

RESULTS

The findings of this literature review comprehensively illustrate how molecular epidemiology has transformed the identification and understanding of chronic disease risk factors. Across the reviewed studies, molecular epidemiology emerges as a key discipline integrating molecular biology, genetics, and epidemiological data to reveal complex mechanisms behind chronic disease development(2). This integration allows for more accurate determination of causal relationships between genetic predispositions, environmental exposures, and disease outcomes.

Transformation of Molecular Epidemiology

The reviewed literature highlights a fundamental transformation in the field of epidemiology — from traditional observational models to molecularly informed frameworks. Through high-resolution biomarker and genomic profiling, molecular epidemiology provides refined insights into disease etiology and subtyping. For example, genetic and molecular profiling has successfully differentiated disease subtypes in conditions such as cancer, diabetes, and cardiovascular diseases, leading to the identification of specific pathways and response patterns (28). This evolution marks a significant leap toward precision prevention and treatment strategies.

Risk Identification and Predictive Modeling

A central finding across multiple studies is the growing effectiveness of multiomic and metabolomic data for predicting disease risk. (10) reported that metabolomic risk scores, developed from data on over 700,000 participants, provide more accurate predictions for common diseases than conventional polygenic scores. Similarly, (4) used multiomic body mass index (BMI) signatures to reveal diverse health phenotypes, showing that individuals with identical BMI values may differ significantly in their molecular health profiles and responses to lifestyle interventions. These studies confirm that molecular-level data yield a superior understanding of risk heterogeneity compared to traditional metrics.

Gene–Environment Interactions and Disease Subtypes

The integration of molecular markers into epidemiological research has enabled the identification of gene–environment interactions that contribute to chronic disease susceptibility. (28) demonstrated that specific genetic mutations, such as KRAS and PIK3CA, influence therapeutic response and preventive effects of aspirin in colorectal cancer, underscoring the clinical implications of molecular subtyping. Similarly, Li et al. (2017) used network-based single nucleotide polymorphism (SNP) analyses to identify genes functionally linked to chronic disease risk, while Ito and Matsuo (2015) highlighted real-world applications of molecular epidemiology in breast cancer studies. Collectively, these studies show that molecular subtyping can enhance treatment precision and disease classification.

Analytical Innovation and Computational Advances

Recent methodological advancements have strengthened molecular epidemiology's analytical capacity. The use of machine learning and Bayesian causal networks has made it possible to model complex interactions and infer causal structures within vast datasets (7,8). Afrifa-Yamoah et al. (2024) further demonstrated that machine learning approaches not only enhance predictive performance but also help trace underlying pathophysiological processes, while emphasizing the importance of ethical data handling. These innovations bridge computational science and epidemiology, fostering more robust causal inferences in chronic disease research.

Challenges in Biomarker Validation and Clinical Translation

Despite these advancements, translating molecular findings into clinical practice remains a key challenge. Several studies point to the difficulty of biomarker validation, the heterogeneity of populations, and the complexity of integrating multiomic data (5,13). Franks and Pomares-Millan (2020) stressed that while molecular biomarkers enhance risk prediction, they require rigorous validation across diverse cohorts before clinical application. Consequently, the field's future success will depend on standardized validation protocols, longitudinal datasets, and reproducibility across populations.

Comparative Insights and Implications

Compared with earlier research paradigms, molecular epidemiology represents a paradigm shift from descriptive associations to mechanistic understanding. The integration of genetic, metabolic, and environmental factors enables researchers to identify not only who is at risk but why they are at risk. As a result, molecular epidemiology provides a comprehensive framework for the personalization of prevention, diagnosis, and treatment strategies, ensuring that interventions are tailored to molecular disease subtypes (Watanabe et al., 2023; Barrett et al., 2024).

Table 1. Key Approaches and Findings in Molecular Epidemiology of Chronic

Approach / Finding	Description	Key References
Metabolomic & Multiomic Profiling	Metabolomic risk scores outperform polygenic scores in predicting disease risk; identifies heterogeneity in health phenotypes.	(4,10)
Molecular Subtype Identification	Defines disease subtypes using genetic and molecular profiles to predict therapeutic response and etiology.	(2,28)
Gene–Environment Interaction	Maps functional SNPs and protein networks to identify genes associated with chronic disease susceptibility.	(29,30)
Machine Learning & Causal Networks	Applies advanced computational models (e.g., Bayesian networks) to discover complex causal links in large datasets.	(7–9)
Biomarker Validation & Implementation	Addresses challenges of biomarker reproducibility and clinical translation of multiomic data.	(5,13)

DISCUSSION

The findings from the reviewed literature indicate that molecular epidemiology has redefined the landscape of chronic disease research by integrating molecular, genetic, and environmental perspectives into risk identification. This integration aligns with the broader theoretical framework of precision medicine, which emphasizes individualized disease prevention and treatment based on biological diversity among populations (2). The evidence from the reviewed studies collectively supports the notion that chronic diseases, while sharing common risk factors such as obesity, smoking, and inflammation, also exhibit molecular heterogeneity that influences disease onset and progression (10).

The theoretical implication of these findings reinforces the concept of molecular heterogeneity—that is, diseases like cancer, diabetes, and cardiovascular disorders are not single entities but a collection of molecularly distinct subtypes. The identification of molecular subtypes through multiomic and metabolomic profiling provides a biological explanation for varying clinical outcomes among individuals with similar risk exposures (4).

The integration of gene–environment interaction theory further deepens the understanding of disease mechanisms. As demonstrated by Nishi et al. (2016), interactions between genetic mutations—such as KRAS and PIK3CA—and external factors like diet or lifestyle significantly determine disease susceptibility and therapeutic response. This supports the multi-causal model of chronic disease, which recognizes that genetic predisposition alone is insufficient to explain disease manifestation without contextual environmental influences (29,30). Such interactions emphasize the need for integrated approaches combining molecular data with social and behavioral determinants of health.

In terms of analytical advancement, the reviewed studies demonstrate the increasing importance of computational epidemiology, particularly the use of machine learning and Bayesian causal models, in processing high-dimensional multiomic datasets. Zhao and Jia (2025) developed DAGSLAM, a Bayesian network algorithm capable of identifying mixed-type causal relationships, enhancing the inference of complex disease pathways. Similarly, Shergill et al. (2025) and Afrifa-Yamoah et al. (2024) highlight that machine learning not only enhances predictive capacity but also allows for the discovery of novel pathophysiological mechanisms. However, these studies caution against algorithmic bias and stress the need for ethical transparency, echoing the concerns raised in modern computational public health.

The practical implications of these findings are profound. The integration of molecular data into public health practice offers the potential for precision prevention, where interventions are tailored to molecularly defined subgroups rather than applied uniformly across populations. For instance, metabolomic biomarkers identified in

large-scale cohort studies (10) could guide personalized nutritional interventions, while molecular subtyping may help optimize pharmacological treatments. Nonetheless, translating these findings into clinical and policy frameworks remains challenging due to issues of biomarker validation, data standardization, and limited clinical utility (5,13).

Several factors may influence or limit the observed outcomes. The heterogeneity of population datasets introduces variability that can obscure true associations, while technical challenges in multiomic integration may lead to inconsistent biomarker reproducibility. Furthermore, ethical and logistical issues in data sharing and patient privacy pose barriers to large-scale validation. Despite these limitations, the reviewed studies consistently highlight molecular epidemiology's capacity to generate actionable insights that traditional epidemiology could not achieve. The synthesis of these studies thus marks a paradigm shift in the etiological understanding and preventive strategies for chronic diseases.

In summary, the reviewed literature confirms that molecular epidemiology contributes significantly to advancing precision medicine by clarifying disease mechanisms, enhancing predictive modeling, and enabling targeted prevention. While methodological and ethical challenges persist, ongoing integration of computational tools and molecular data will likely strengthen the translational bridge between research and clinical application. Future research should focus on longitudinal multiomic studies, harmonized data frameworks, and interdisciplinary collaboration to ensure the equitable application of molecular epidemiological insights in global health contexts..

CONCLUSION

This qualitative literature study concludes that molecular epidemiology plays a transformative role in identifying and understanding chronic disease risk factors through its integration of genomic, metabolomic, and environmental data. The synthesis of findings reveals that this interdisciplinary approach bridges the gap between traditional epidemiology and molecular biology, offering deeper mechanistic insights into gene–environment interactions and disease heterogeneity. By combining molecular subtyping, predictive modeling, and computational analytics such as machine learning and Bayesian causal networks, molecular epidemiology advances the theoretical framework of precision medicine while addressing limitations of conventional population-based models. These results not only enrich existing epidemiological theories but also underscore the potential of molecular data to inform personalized prevention and intervention strategies. Socially and culturally, the findings highlight the importance of equitable access to molecular technologies and ethical transparency in data use. Academically, they reinforce the need for interdisciplinary collaboration and methodological rigor. However, this study acknowledges challenges in biomarker validation, standardization of multiomic integration, and limited clinical translation. Future research should therefore focus on longitudinal multiomic studies, cross-population validation, and the incorporation of ethical and sociocultural dimensions to fully realize the promise of molecular epidemiology in addressing the global burden of chronic diseases.

BIBLIOGRAPHY

1. Hamada T, Keum N, Nishihara R, Ogino S. Molecular pathological epidemiology: New developing frontiers of big data science to study etiologies and pathogenesis. J Gastroenterol. 2017;52:265–75.
2. Franks P, Pomares-Millan H. Next-generation epidemiology: The role of high-resolution molecular phenotyping in diabetes research. Diabetologia. 2020;63:2521–32.
3. Smithers-Sheedy H, Henry G. Apa sebenarnya arti diagnosis etiologi? Kedokt Perkemb Neurol Anak. 2022;
4. Watanabe K, Wilmanski T, Diener C, Earls J, Zimmer A. Multiomic signatures of body mass index identify heterogeneous health phenotypes and responses to a lifestyle intervention. Nat Med. 2023;29:996–1008.
5. Al-Hadlaq S, Balto H, Hassan W, Marraiki N, El-Ansary A. Biomarkers of non-communicable chronic disease: an

update on contemporary methods. PeerJ. 2022;10.
6. Bauchet L, Ostrom QT. Epidemiologi dan Epidemiologi Molekuler. Klin Bedah Saraf Am Utara. 2019;
7. Shergill M, Durant S, Birdi S, Rabet R, Ziegler C. Machine learning used to study risk factors for chronic diseases: A scoping review. Can J Public Health. 2025;116.
8. Zhao Y, Jia J. DAGSLAM: Causal Bayesian network structure learning of mixed type data and its application in identifying disease risk factors. BMC Med Res Methodol. 2025;25.
9. Afrifa-Yamoah E, Adua E, Peprah-Yamoah E, Anto E, Opoku-Yamoah V, Acheampong E, dkk. Pathways to chronic disease detection and prediction: Mapping the potential of machine learning to the pathophysiological processes while navigating ethical challenges. Chronic Dis Transl Med. 2024;11:1–21.
10. Barrett J, Esko T, Fischer K, Jostins-Dean L, Jousilahti P, Julkunen H. Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. Nat Commun. 2024;15.
11. Wang M, Spiegelman D, Kuchiba A, Lochhead P, Kim S. Statistical methods for studying disease subtype heterogeneity. Stat Med. 2016;35:782–800.
12. Wang X, Sun G, Feng T, Zhang J, Huang X, Wang T, dkk. Sodium oligomannate therapeutically remodels gut microbiota and suppresses gut bacterial amino acids-shaped neuroinflammation to inhibit Alzheimer's disease progression. Cell Res. 2019;29:787–803.
13. Agrawal M, Allin K, Petralia F, Colombel J, Jess T. Multiomics to elucidate inflammatory bowel disease risk factors and pathways. Nat Rev Gastroenterol Hepatol. 2022;19:399–409.
14. Mottalib M, Jones-Smith J, Sheridan B, Beheshti R. Subtyping patients with chronic disease using longitudinal BMI patterns. IEEE J Biomed Health Inform. 2021;27:2083–93.
15. Bingham A. From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. Int J Qual Methods. 2023;22.
16. Pratt M. On the Evolution of Qualitative Methods in Organizational Research. Annu Rev Organ Psychol Organ Behav. 2025;
17. Bandaranayake P. Application of Grounded Theory Methodology in Library and Information Science Research: An Overview. Sri Lanka Libr Rev. 2024;
18. Togia A, Malliari A. Research methods in library and information science. Dalam: IntechOpen. 2017.
19. Granikov V; H Q; Crist, E; Pluye, P. Mixed methods research in library and information science: A methodological review. Libr Inf Sci Res. 2020;
20. Jimenez S; BM J; De La Torre, R. How do university libraries contribute to the research process? J Acad Librariansh. 2024;
21. Belotto M. Data Analysis Methods for Qualitative Research: Managing the Challenges of Coding, Interrater Reliability, and Thematic Analysis. Qual Rep. 2018;
22. Fife S, Gossner J. Deductive qualitative analysis: Evaluating, expanding, and refining theory. Int J Qual Methods. 2024;23.
23. Vila-Henninger L; D C; Van Ingelgom, V; Caprioli, M; Teuber, F; Pennetreau, D; Bussi, M; Gall, C. Abductive Coding: Theory Building and Qualitative (Re)Analysis. Sociol Methods Res. 2022;53:968–1001.
24. Kalpokaite N, Radivojevic I. Demystifying qualitative data analysis for novice qualitative researchers. Qual Rep. 2018;
25. Abraham D; P P. A Methodological Framework for Descriptive Phenomenological Research. West J Nurs Res. 2024;47:125–34.
26. Baillie J. Commentary: An overview of the qualitative descriptive design within

nursing research. J Res Nurs. 2019;25:458–9.

27. Doyle L; M C; Keogh, B; Brady, A; McCann, M. An overview of the qualitative descriptive design within nursing research. J Res Nurs. 2019;25:443–55.
28. Nishi A, Milner D, Giovannucci E, Nishihara R, Tan A, Kawachi I, dkk. Integration of molecular pathology, epidemiology, and social science for global precision medicine. Expert Rev Mol Diagn. 2016;16:11–23.
29. Ito H, Matsuo K. Molecular epidemiology and possible real-world applications in breast cancer. Breast Cancer. 2015;23:33–8.
30. Li W, Zhu L, Huang H, He Y, Lv J, Li W, dkk. Identification of susceptible genes for complex chronic diseases based on disease risk functional SNPs and interaction networks. J Biomed Inform. 2017;74:137–44.