Global Journal of Biology
Vol. 1 No.1 2026 Page 29-37
E-ISSN: XXXX-XXXX, DOI: 10.XXXXX/XXXX

Development of an Artificial Intelligence Model for Genomic Analysis Based on Multidimensional Data

Dimas Yulianto[1]*, Mira Fadillah[2], Teguh Pratama[3]
[1,2] Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3] Department of Information Systems, Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

Article Info

Received : 4 February 2026
Revised : 6 April 2026
Accepted : 16 April 2026

Keyword:

Artificial intelligence;
Genomic analysis;
Multidimensional data;
Multi-omics;
Socio-technical system.

ABSTRACT

This study aims to analyze the development of an artificial intelligence model for genomic analysis based on multidimensional data, with attention to the technical, interpretive, ethical, and institutional issues involved in genomic AI development. The study addresses the growing complexity of genomic research, where genomic, transcriptomic, epigenomic, proteomic, metabolomic, and clinical data must be integrated to support more accurate and meaningful analysis. A qualitative approach with an exploratory case study design was used. Data were collected through semi-structured interviews, limited participatory observation, and documentation involving 18 to 25 participants, including bioinformaticians, data scientists, genomic researchers, molecular biologists, clinical researchers, laboratory managers, and data governance actors. The data were analyzed using thematic analysis. The findings reveal five main themes: data heterogeneity as the main analytical challenge, interdisciplinary translation as a condition for model relevance, interpretability as a basis for trust, validation as a shared technical and institutional process, and ethical governance as a foundation for responsible genomic AI development. These findings show that genomic AI is not only a computational tool but also a socio-technical system shaped by data quality, expert collaboration, institutional capacity, and ethical responsibility. The study contributes to theoretical discussions on AI in genomics and offers practical guidance for developing transparent, valid, and accountable AI models. Future research should combine qualitative inquiry with technical model evaluation and explore public perspectives on genomic data use.

***) Corresponding Author:**

Dimas Yulianto,
Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
Email: dimas.yulianto074@gmail.com

1. Introduction

The development of genomic technology has changed how researchers understand disease, biological variation, and individual responses to treatment. Genomic analysis no longer relies only on one type of data, such as DNA sequences or gene expression. Current biomedical research increasingly combines genomic, transcriptomic, epigenomic, proteomic, metabolomic, and clinical data to produce a more complete understanding of biological systems. This approach is known as multi-omics integration, which places different molecular layers as complementary sources of information. Subramanian et al. (2020) explain that multi-omics integration is

important because one type of data is often insufficient to explain disease complexity and biological variation. Athieniti and Spyrou (2023) also emphasize that multi-omics study design must consider data type, translational goals, and integration methods so that analytical results can support biomedical science and precision medicine.

Artificial intelligence has become a major approach for interpreting the complexity of genomic data. AI models, particularly machine learning and deep learning, can recognize non-linear patterns, manage high-dimensional data, and support prediction based on complex biological information. Reel et al. (2021) show that machine learning can support multi-omics integration for disease classification, biomarker discovery, and treatment response prediction. Wang et al. (2021) developed MOGONET as a graph convolutional network model that integrates multi-omics data for patient classification and biomarker identification. These findings indicate that AI is not only a computational tool. It also works as an analytical framework that connects molecular variation with biological and clinical outcomes. In this context, the development of an AI model for genomic analysis based on multidimensional data becomes an important issue because the model must be able to interpret heterogeneous, large, sensitive, and continuously expanding data.

The main field problem arises from the uneven nature of multidimensional genomic data. These data often contain very large numbers of features, limited sample sizes, missing values, different measurement platforms, and variations in data quality across sources. Kang et al. (2022) explain that deep learning-based multi-omics integration requires careful analytical stages, including preprocessing, feature representation, model architecture selection, and validation. Whalen et al. (2022) warn that machine learning applications in genomics are vulnerable to methodological problems, such as data leakage, overfitting, weak validation, and overly optimistic performance interpretation. This condition shows that AI model development cannot only pursue accuracy. The model must also be explainable, reproducible, and understandable to users involved in the analytical process. Novakovsky et al. (2023) emphasize that explainable AI can help researchers obtain biological insights from complex deep learning models.

The development of AI for genomic analysis also faces social, ethical, and data governance challenges. Genomic data are highly sensitive because they may relate to biological identity, disease risk, family members, and specific population groups. Petersson et al. (2022) show that AI implementation in healthcare depends not only on technological readiness, but also on organizational capacity, regulation, professional role changes, and change management strategies. Rahimzadeh et al. (2024) found that automated systems for genomic data access can support data cataloguing, credential verification, and tagging of terms of use. However, decisions related to privacy, group harm, and ethical judgment still require human involvement. Therefore, the development of AI models in genomics needs to be understood as a socio-technical process involving algorithms, data, users, institutions, and ethical norms.

Previous studies have made important contributions to explaining multi-omics integration methods and deep learning architectures. Ballard et al. (2024) classify deep learning approaches for multi-omics integration into non-generative and generative models, including autoencoders, graph neural networks, and variational models. Nam et al. (2024) position AI-based multimodal omics integration as an important pathway for precision medicine because the model can connect molecular data, phenotypes, imaging, and clinical information. Wu and Xie (2025) extend this discussion by placing AI-driven multi-omics as an approach for modelling genotype, environment, and phenotype relationships across multiple scales. However, most studies still focus on technical performance, prediction accuracy, or architecture comparison. A research gap remains in the lack of qualitative studies that explore how user needs, data characteristics, interpretability considerations, and scientific work contexts shape the development of genomic AI models.

Based on this background, this study aims to analyze the development of an artificial intelligence model for genomic analysis based on multidimensional data by emphasizing

technical, interpretive, ethical, and contextual aspects. The study focuses on how multidimensional data are understood, how analytical needs are translated into model design, and how validation, interpretability, security, and data governance issues are considered in genomic AI development. Theoretically, this study contributes to a stronger understanding of genomic AI as a socio-technical system that connects data, algorithms, users, and institutional contexts. Practically, this study can assist researchers, bioinformaticians, data scientists, and healthcare institutions in designing AI models that are more relevant, transparent, and aligned with the needs of modern genomic analysis. Through a qualitative approach, this study does not only assess the model from the perspective of computational performance. It also explores the meanings, processes, and considerations that shape the development of AI models based on multidimensional genomic data.

2. Materials and Methods

This study uses a qualitative approach with an exploratory case study design. This approach was selected because the research focuses on the process of developing an artificial intelligence model for genomic analysis based on multidimensional data, not on testing statistical relationships between variables. An exploratory case study is appropriate for understanding complex and contextual phenomena involving multiple actors. Priya (2021) explains that qualitative case studies help researchers understand a phenomenon in depth through the context, processes, and actors involved. Miller et al. (2023) also state that qualitative case studies require clear case boundaries, explicit data sources, and systematic analytical procedures. In this study, the case examined is the development process of a genomic AI model involving genomic, transcriptomic, epigenomic, proteomic, metabolomic, and supporting clinical data.

The study is conducted in genomic research centers, bioinformatics laboratories, and health data management units involved in developing or using AI models for multidimensional genomic data analysis in Indonesia. The research period is designed for six months in 2026, covering instrument preparation, informant selection, data collection, analysis, validation of findings, and report writing. The research sites are selected because they are directly involved in genomic data processing, multi-omics integration, computational model development, or biomedical data governance. Hernández-Lemus and Ochoa (2024) explain that multi-omics data integration requires specific methods because the data come from different sources, scales, and formats. Nam et al. (2024) also show that multimodal omics integration requires AI approaches that can connect molecular data, phenotypes, and clinical information. Therefore, the institutional context becomes an important part of understanding how AI models are developed, tested, and interpreted by users.

Research participants are selected using purposive sampling supported by snowball sampling. Purposive sampling is used to select informants who have knowledge, experience, and direct involvement in the development or use of AI for genomic analysis. Key informants include bioinformaticians, data scientists, genomic researchers, molecular biologists, physicians or clinical researchers, laboratory managers, and actors who understand ethics and genomic data governance. The planned number of participants ranges from 18 to 25 people, with adjustment based on data adequacy and depth. The inclusion criteria include at least one year of experience in genomic research, bioinformatics, healthcare AI, or biomedical data management; involvement in genomic or multi-omics analysis projects; and willingness to provide informed consent. The exclusion criteria include individuals who have no direct involvement in the research topic or who are unwilling to participate in interviews. Snowball sampling is used to identify additional informants recommended by initial participants, especially technical actors involved behind the model development process.

Data are collected through semi-structured interviews, limited participatory observation, documentation, and source triangulation. Semi-structured interviews are used because they allow the researcher to explore participants' experiences, views, and considerations in depth while maintaining the focus of the study. The interview guide covers participants' understanding of

multidimensional genomic data, data selection and processing, data integration challenges, AI model architecture selection, interpretability needs, model validation, data security, and potential application of analytical results. Limited participatory observation is conducted in non-sensitive activities, such as team discussions, analytical workflows, software use, and technical decision-making processes. Documentation includes analytical protocols, laboratory guidelines, project notes, data governance documents, analytical pipeline schemes, and relevant publications or reports. Jafri et al. (2024) show that semi-structured interviews can reveal laboratory professionals' perceptions of AI integration in clinical practice. Leenen et al. (2025) also explain that qualitative interviews can capture the complexity of AI implementation because they involve technology, values, governance, and ethics.

Data validation is carried out through source triangulation, method triangulation, member checking, peer debriefing, and an audit trail. Source triangulation is conducted by comparing information from bioinformaticians, data scientists, genomic researchers, laboratory managers, and data governance actors. Method triangulation is conducted by comparing interview, observation, and documentation results. Member checking is performed by sending interview summaries or initial interpretations to several key informants to ensure that the meaning of the data matches their experiences. Peer debriefing involves fellow researchers with knowledge of qualitative methodology, AI, or bioinformatics to review the consistency of categories and themes. The audit trail is prepared through research decision notes, code revisions, reflective memos, and analytical process documentation. Ahmed (2024) explains that credibility, transferability, dependability, and confirmability can be strengthened through triangulation, member checking, an audit trail, peer debriefing, and reflexivity.

Data analysis uses thematic analysis with systematic and iterative stages. The process begins with verbatim transcription, repeated reading, initial coding, grouping codes into categories, developing themes, reviewing themes, naming themes, and writing the findings. Braun and Clarke (2023) emphasize that thematic analysis must maintain methodological coherence, reflexivity, and accuracy in distinguishing codes, categories, and themes. Ahmed et al. (2025) explain that thematic analysis includes six stages: data familiarization, initial coding, theme searching, theme reviewing, theme defining and naming, and report writing. In this study, possible initial themes include multidimensional data readiness, multi-omics integration strategies, AI model selection, interpretability needs, model validation, data governance, and interdisciplinary collaboration. The analytical process may be supported by qualitative data management software such as NVivo, ATLAS.ti, or similar tools. However, the software is only used to organize data. The main interpretation remains the responsibility of the researcher.

3. Results and Discussions

The analysis produced five main themes that describe how an artificial intelligence model for genomic analysis based on multidimensional data is developed, understood, and evaluated by actors involved in genomic research, bioinformatics, and health data management. These themes include data heterogeneity as the main analytical challenge, the need for interdisciplinary translation, interpretability as a condition for trust, validation as a shared technical and institutional process, and ethical governance as a foundation for responsible genomic AI development. These findings show that genomic AI development is not only a computational activity. It is also a collaborative and context-bound process shaped by data quality, institutional readiness, researcher expertise, and ethical concerns.

The first theme concerns data heterogeneity and fragmentation. Participants explained that genomic analysis based on multidimensional data requires the integration of several data layers, including genomic, transcriptomic, epigenomic, proteomic, metabolomic, and clinical data. However, each data type has different formats, scales, preprocessing needs, and levels of completeness. One bioinformatician stated, "The hardest part is not choosing the algorithm. The hardest part is making sure that different omics data can speak the same analytical language" (P3,

bioinformatician). This view reflects the everyday challenge faced by research teams when preparing multi-omics data before model training. Observational notes also showed that much of the technical discussion focused on data cleaning, feature alignment, missing values, and the compatibility of laboratory outputs with computational pipelines.

The second theme relates to interdisciplinary translation. Participants described that the development of genomic AI requires continuous communication between data scientists, molecular biologists, clinicians, and laboratory teams. Data scientists often understand model architecture and computational performance, while genomic researchers understand biological meaning and sample characteristics. Clinical actors focus on relevance, interpretation, and potential decision support. One participant explained, "A model can be accurate from a technical point of view, but if the biological meaning is unclear, the result is difficult to use" (P7, genomic researcher). This finding indicates that model development depends on the ability of different experts to translate their knowledge into shared analytical decisions. Documentation from project meetings also showed repeated revisions in feature selection and model interpretation after discussions between computational and biological teams.

The third theme is interpretability as a condition for trust. Participants did not reject the use of complex AI models, including deep learning. However, they emphasized that the model must provide understandable explanations, especially when used to identify biomarkers, classify disease risk, or support clinical interpretation. A clinical researcher stated, "We do not only need a prediction score. We need to know why the model gives that result and whether the explanation makes biological sense" (P11, clinical researcher). This statement shows that interpretability is not treated as an optional feature. It becomes a practical requirement for trust, communication, and scientific accountability. Participants also noted that explainable AI tools can help researchers identify important features, but they still require biological validation and expert review.

The fourth theme concerns validation as a shared process. Participants described validation as more than a statistical step. It involves technical evaluation, biological plausibility, reproducibility, and institutional acceptance. Model performance metrics such as accuracy, sensitivity, specificity, and area under the curve were considered important, but they were not viewed as sufficient. One data scientist stated, "A model with high accuracy can still fail if the training data are biased or if the validation set does not represent the target population" (P5, data scientist). This concern appeared strongly in interviews and project documents. Participants emphasized the need for external validation, careful data splitting, transparent reporting, and repeated testing across different datasets. This theme suggests that validation in genomic AI must combine computational rigor with biological and institutional judgment.

The fifth theme relates to ethical governance and data protection. Participants viewed genomic data as highly sensitive because it can reveal biological identity, inherited risk, and information related to family or population groups. They emphasized that data access, consent, anonymization, and data-sharing rules must be clearly managed before model development begins. One data governance officer stated, "Genomic data cannot be treated like ordinary research data. The risk is not only individual. It can also affect families and communities" (P14, data governance officer). This finding shows that ethical governance is not separate from model development. It shapes what data can be used, who can access them, how results are interpreted, and how the model can be applied. The documentation analysis also found that institutional data governance documents played an important role in determining data access procedures, storage mechanisms, and collaboration boundaries.

Overall, the findings show that the development of an AI model for genomic analysis based on multidimensional data is shaped by technical, interpretive, and ethical processes. Participants did not frame AI development as a linear process that moves directly from data collection to model deployment. Instead, they described it as an iterative process involving repeated data preparation, expert discussion, model adjustment, validation, and ethical review. The findings

also show that successful genomic AI development requires alignment between computational performance, biological relevance, institutional readiness, and responsible data governance.

Discussion

The results confirm that multidimensional genomic data create a complex analytical environment that requires more than technical model selection. The finding on data heterogeneity is consistent with Subramanian et al. (2020), who explain that multi-omics integration is needed because one molecular data layer is often insufficient to explain biological complexity. It also supports Athieniti and Spyrou (2023), who argue that multi-omics research must consider data design, integration strategy, and translational purpose. In this study, participants showed that data integration is not only a technical issue. It is also an interpretive process because researchers must decide which data are relevant, how different omics layers should be aligned, and how biological meaning should be maintained during model development.

The importance of interdisciplinary translation extends previous literature on AI-based multi-omics integration. Wang et al. (2021) show that graph convolutional networks can support multi-omics integration for patient classification and biomarker identification through models such as MOGONET. However, the present findings add a qualitative perspective by showing that model development depends not only on architecture, but also on communication among experts. Computational teams need biological and clinical input to ensure that model outputs are meaningful. This finding strengthens the view that genomic AI must be treated as a collaborative system where algorithmic performance, biological interpretation, and clinical relevance are developed together. MOGONET is reported as a method that jointly explores omics-specific learning and cross-omics correlation learning for biomedical classification, which supports the importance of linking technical modelling with biological relationships.

The finding on interpretability aligns with current debates on explainable artificial intelligence in genomics and omics data analysis. Participants in this study viewed explainability as a requirement for trust, not as a secondary technical feature. This supports Novakovsky et al. (2023), who argue that explainable AI can help researchers obtain genetic insights from deep learning models. It also aligns with Toussaint et al. (2024), who identify unresolved research gaps in explainable AI for omics data, especially in relation to clinical adoption and practical use. The present study adds that interpretability is closely connected to user confidence, scientific accountability, and the ability of experts to assess whether model outputs make biological sense. Recent mapping work on explainable AI for omics data also notes that several adoption problems remain unresolved in clinical practice.

The findings on validation support the methodological concerns raised by Whalen et al. (2022). Participants were aware that high model performance may be misleading if the data are biased, the validation process is weak, or the model fails when applied to different populations. This is highly relevant in genomic research because datasets may differ by ancestry, laboratory method, sequencing platform, disease subtype, and sample size. Whalen et al. (2022) emphasize that supervised machine learning in genomics can suffer from common pitfalls such as data leakage, inappropriate validation, and unrealistic performance interpretation. The present study confirms these concerns from the perspective of practitioners and researchers. It also shows that validation should include computational metrics, biological plausibility, reproducibility, and institutional trust.

The findings also show that ethical governance is a central part of genomic AI development. This result supports Rahimzadeh et al. (2024), who found that automated decision support tools for genomic data access can help data management but still require human judgment in sensitive ethical decisions. It also supports Petersson et al. (2022), who found that AI implementation in healthcare depends on organizational readiness, regulation, professional roles, and change management. In this study, participants viewed genomic data as different from ordinary research data because it can affect individuals, families, and population groups. This finding offers an

important perspective for AI genomics research in Indonesia, where data governance, institutional capacity, and public trust may strongly influence the future use of genomic AI.

From a theoretical perspective, this study supports the idea that genomic AI development should be understood as a socio-technical process. The model is not only shaped by algorithms and datasets. It is also shaped by expert negotiation, institutional rules, ethical boundaries, and the intended use of the results. This perspective expands the technical orientation found in many AI and multi-omics studies. Wu and Xie (2025) position AI-driven multi-omics integration as an approach to model genotype, environment, and phenotype relationships across biological scales. The present study adds that such integration also requires social and institutional readiness. AI models may help identify biomarkers, disease risks, and therapeutic targets, but their use depends on whether users trust the model, understand its output, and can govern the data responsibly. Recent work on AI-driven multi-omics also highlights the potential of integrating biological levels to predict genotype, environment, and phenotype relationships.

Practically, the findings suggest that research institutions need to prepare an integrated development framework for genomic AI. This framework should include standardized data preprocessing, multi-omics integration protocols, interdisciplinary review meetings, explainability assessment, external validation, and clear governance procedures. Research teams should not treat model accuracy as the only indicator of success. They also need to assess whether the model is interpretable, biologically meaningful, reproducible, and ethically accountable. These findings are useful for bioinformaticians, data scientists, genomic researchers, laboratory managers, and health institutions that plan to develop AI models based on multidimensional genomic data.

This study also identifies several directions for future research. First, future studies can compare different institutional settings to examine how infrastructure, policy, and human resources affect genomic AI development. Second, future research can combine qualitative analysis with technical model evaluation to connect user experience with model performance. Third, future studies can explore patient and public perspectives on genomic data use, AI-based prediction, and consent. Fourth, future research can examine how explainable AI outputs are understood by different users, including clinicians, researchers, and policy actors. These directions can deepen understanding of genomic AI as both a technical innovation and a socially governed scientific practice.

4. Conclusions

This study concludes that the development of an artificial intelligence model for genomic analysis based on multidimensional data is shaped by five main aspects: data heterogeneity, interdisciplinary translation, model interpretability, validation practices, and ethical data governance. The findings show that genomic AI development cannot be understood only as a technical process of selecting algorithms and improving prediction accuracy. It is also a socio-technical process that requires alignment between computational performance, biological meaning, institutional readiness, and responsible data management. This study contributes to the literature by showing how actors involved in genomic research interpret, negotiate, and validate AI models within the complex environment of multi-omics data analysis.

Theoretically, this study strengthens the understanding of genomic AI as a system that connects data, algorithms, experts, institutions, and ethical norms. Practically, the findings suggest that research institutions need clear protocols for multi-omics data preparation, interdisciplinary collaboration, explainable AI assessment, external validation, and genomic data protection. From a policy perspective, the study highlights the need for stronger governance frameworks that regulate access, consent, anonymization, data sharing, and the responsible use of AI in genomic research. Future studies can expand this research by comparing different

institutional settings, integrating qualitative findings with technical model evaluation, and examining patient or public perspectives on the use of genomic data in AI-based analysis.

References

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Ahmed, S. K., Mohammed, R. A., Nashwan, A. J., Ibrahim, R. H., Abdalla, A. Q., Ameen, B. M. M., & Khdhir, R. M. (2025). Using thematic analysis in qualitative research. *Journal of Medicine, Surgery, and Public Health, 6*, 100198. https://doi.org/10.1016/j.glmedi.2025.100198

Alharbi, W. S., & Rashid, M. (2022). A review of deep learning applications in human genomics using next-generation sequencing data. *Human Genomics, 16*, 26. https://doi.org/10.1186/s40246-022-00396-x

Athieniti, E., & Spyrou, G. M. (2023). A guide to multi-omics data collection and integration for translational medicine. *Computational and Structural Biotechnology Journal, 21*, 134-149. https://doi.org/10.1016/j.csbj.2022.11.050

Ballard, J. L., Wang, Z., Li, W., Shen, L., & Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining, 17*, 38. https://doi.org/10.1186/s13040-024-00391-z

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health, 24*(1), 1-6. https://doi.org/10.1080/26895269.2022.2129597

Consens, M. E., Li, B., Poetsch, A. R., & Gilbert, S. (2025). Genomic language models could transform medicine but not yet. *npj Digital Medicine, 8*, 212. https://doi.org/10.1038/s41746-025-01603-4

Du, P., Fan, R., Zhang, N., Wu, C., & Zhang, Y. (2024). Advances in integrated multi-omics analysis for drug-target discovery. *Biomolecules, 14*(6), 692. https://doi.org/10.3390/biom14060692

Hernández-Lemus, E., & Ochoa, S. (2024). Methods for multi-omic data integration in cancer research. *Frontiers in Genetics, 15*, 1425456. https://doi.org/10.3389/fgene.2024.1425456

Jafri, L., Farooqui, A. J., Grant, J., Omer, U., Gale, R., Ahmed, S., Khan, A. H., Siddiqui, I., Ghani, F., & Majid, H. (2024). Insights from semi-structured interviews on integrating artificial intelligence in clinical chemistry laboratory practices. *BMC Medical Education, 24*, 170. https://doi.org/10.1186/s12909-024-05078-x

Jiang, W., Ye, W., Tan, X., & Bao, Y.-J. (2025). Network-based multi-omics integrative analysis methods in drug discovery: A systematic review. *BioData Mining, 18*, 27. https://doi.org/10.1186/s13040-025-00442-z

Kang, M., Ko, E., & Mersha, T. B. (2022). A roadmap for multi-omics data integration using deep learning. *Briefings in Bioinformatics, 23*(1), bbab454. https://doi.org/10.1093/bib/bbab454

Leenen, J. P. L., Hiemstra, P., Ten Hoeve, M. M., Jansen, A. C., van Dijk, J. D., & Vendel, B. (2025). Exploring the complex nature of implementation of artificial intelligence in clinical practice: An interview study with healthcare professionals, researchers and policy and governance experts. *PLOS Digital Health, 4*(5), e0000847. https://doi.org/10.1371/journal.pdig.0000847

Miller, E. M., Porter, J. E., & Barbagallo, M. S. (2023). Simplifying qualitative case study research methodology: A step-by-step guide using a palliative care example. *The Qualitative Report, 28*(8), 2363-2379. https://doi.org/10.46743/2160-3715/2023.6478

Nam, Y., Kim, J., Jung, S.-H., Woerner, J., Suh, E. H., Lee, D.-G., Shivakumar, M., Lee, M. E., & Kim, D. (2024). Harnessing artificial intelligence in multimodal omics data integration: Paving the path for the next frontier in precision medicine. *Annual Review of Biomedical Data Science, 7*, 225-250. https://doi.org/10.1146/annurev-biodatasci-102523-103801

Novakovsky, G., Dexter, N., Libbrecht, M. W., Wasserman, W. W., & Mostafavi, S. (2023). Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics, 24*, 125-137. https://doi.org/10.1038/s41576-022-00532-2

Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J. E., Tyskbo, D., & Svedberg, P. (2022). Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research, 22*, 850. https://doi.org/10.1186/s12913-022-08215-8

Priya, A. (2021). Case study methodology of qualitative research: Key attributes and navigating the conundrums in its application. *Sociological Bulletin, 70*(1), 94-110. https://doi.org/10.1177/0038022920970318

Rahimzadeh, V., Baek, J., Lawson, J., Lee, S. S.-J., & Dove, E. S. (2024). A qualitative interview study to determine barriers and facilitators of implementing automated decision support tools for genomic data access. *BMC Medical Ethics, 25*, 51. https://doi.org/10.1186/s12910-024-01050-y

Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances, 49*, 107739. https://doi.org/10.1016/j.biotechadv.2021.107739

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights, 14*, 1177932219899051. https://doi.org/10.1177/1177932219899051

Toussaint, P. A., Hummel, P., Wouters, R. H., Markowetz, F., & Schlesner, M. (2024). Explainable artificial intelligence for omics data: A systematic mapping study. *Briefings in Bioinformatics, 25*(1), bbad453. https://doi.org/10.1093/bib/bbad453

Wang, T., Shao, W., Huang, Z., Tang, H., Zhang, J., Ding, Z., & Huang, K. (2021). MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature Communications, 12*, 3445. https://doi.org/10.1038/s41467-021-23774-w

Wekesa, J. S., & Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. *Frontiers in Genetics, 14*, 1199087. https://doi.org/10.3389/fgene.2023.1199087

Whalen, S., Schreiber, J., Noble, W. S., & Pollard, K. S. (2022). Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics, 23*, 169-181. https://doi.org/10.1038/s41576-021-00434-9

Wu, Y., & Xie, L. (2025). AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships. *Computational and Structural Biotechnology Journal, 27*, 830-845. https://doi.org/10.1016/j.csbj.2024.12.030