P-ISSN: 3025-6518
E-ISSN: 3025-0501

Almakki Health Informatics Journal

Leveraging Big Data Bioinformatics for Genomic Analysis in Precision Medicine Development

Dikdik Abdulah Sidik
Universitas Bhakti Kencana Bandung, Indonesia
E-mail: Dikdikabdullah@yahoo.co.id

KEYWORDS

Bioinformatics, Genomics, Big Data, Personalized Medicine.

ABSTRACT

Advances in bioinformatics and genomics technology have created great opportunities in realizing personalized medicine. This study aims to analyze the role of bioinformatics in the use of genomic big data, emphasizing the opportunities, challenges, and integration strategies needed. The research method used is qualitative with an exploratory approach. Data were collected through in-depth interviews, focus group discussions, and documentation studies, then analyzed using thematic analysis techniques. The results of the study show that bioinformatics offers a variety of opportunities, such as the utilization of pharmacogenomics for drug response prediction, disease biomarker identification, multi-omics data integration, machine learning algorithm development, and early detection of genetic diseases. However, a number of challenges still hinder implementation, including genomic data privacy and security issues, limited computing infrastructure, lack of interdisciplinary human resources, lack of standardization of data formats, and regulatory and policy barriers. This research also emphasizes the importance of the strategy of integrating bioinformatics and big data through strengthening technological infrastructure, global data standardization, the use of artificial intelligence, privacy protection based on adaptive regulations, international collaboration, and the development of interdisciplinary education. The implications of this research not only enrich the academic literature, but also provide a practical basis for the formulation of more precise and equitable health policies. Thus, this study confirms that the use of bioinformatics is a strategic path towards the transformation of data-based health services, which is oriented to individual needs while maintaining the ethics and sustainability of the health system.

INTRODCUTION

Over the last two decades, genomic technology has catalyzed a profound shift in biomedical science. Large-scale sequencing, population biobanks, and multi-omics pipelines now generate biological big data at volumes that were previously unimaginable. Projections since the mid-2010s already indicated that genomics would approach the exabyte scale within a decade, a trend that has largely materialized as sequencing costs have fallen and adoption has expanded (Stephens et al., 2015; Marx, 2013). This data deluge underpins the promise of personalized medicine tailoring prevention, diagnosis, and therapy to individual molecular profiles yet also exposes bottlenecks in storage, analysis, interpretation, governance, and equitable access (Khoury et al., 2020).

(2) The strategic importance of bioinformatics in this transition is undisputed. High-throughput platforms now produce millions to billions of reads per sample; downstream analytics from variant calling and functional annotation to polygenic risk scoring and clinical decision support depend on robust computational and statistical pipelines. As more health systems pilot pharmacogenomics and molecular tumor boards, the demand for interoperable, secure, and clinically validated bioinformatics

workflows is accelerating. At the same time, stakeholders face practical constraints: infrastructure costs, workforce capacity, and heterogeneous data standards that complicate cross-institutional collaboration (Grossman et al., 2016; Shendure et al., 2017).

(3) These global pressures are felt acutely in developing-country contexts, including Indonesia and the wider ASEAN region. Many public hospitals and research centers still have limited access to high-performance computing, long-term archival storage, and stable broadband required for multi-omics pipelines; electronic health record (EHR) adoption remains uneven; and data localization as well as ethical–legal frameworks are evolving. Such conditions can widen the gap between cutting-edge research outputs and their translation into routine care, dampening the potential public-health returns of genomics, particularly for infectious disease genomics, oncology, and rare diseases where timely analyses are critical (Khoury et al., 2020).

(4) Prior work has clarified several root causes behind these implementation gaps. First, next-generation sequencing (NGS) generates massive, fast-accumulating datasets that strain compute, storage, and network capacities, especially where scaling strategies are ad hoc (Shendure et al., 2017). Second, fragmented standards and inconsistent metadata impede interoperability and reproducibility across platforms and institutions (Grossman et al., 2016). Third, persistent privacy and security risks re-identification, broad consent challenges, secondary use erode public trust if not addressed with rigorous technical and governance controls (Erlich & Narayanan, 2014). Fourth, shortages of interdisciplinary talent able to bridge biology, statistics, and computing slow the uptake of validated pipelines in clinical environments (Mulder et al., 2018).

(5) At the same time, methodological advances demonstrate clear clinical potential. Machine-learning models now support variant effect prediction, phenotype imputation, and integrative multi-omics signatures; network-based and systems-biology approaches illuminate gene–environment interactions; and translational pipelines increasingly link genotype to drug response (Libbrecht & Noble, 2015). In pharmacogenomics, for example, variant-aware prescribing can mitigate adverse drug reactions and improve efficacy, illustrating how bioinformatics directly enables patient-level benefit (Roden et al., 2019). Multi-omics integration has also accelerated biomarker discovery for early detection and risk stratification (Hasin et al., 2017).

(6) Despite these advances, a clear research gap persists: most studies either optimize technical analytics or describe ethical/policy frameworks, but few operationalize an integrated, context-aware model that joins technical readiness (pipelines, standards, infrastructure), governance safeguards (privacy, consent, data sharing), and health-system delivery (workflows, workforce, reimbursement) in lower- and middle-income settings. Empirical guidance remains limited on how to stage investments, sequence capacity-building, and measure value realization from genomics data assets in public health and clinical care (Grossman et al., 2016; Khoury et al., 2020).

(7) This study's novelty is threefold and explicitly differentiated from prior research that tends to be siloed by discipline (Torkamani et al., 2017; Krittanawong et al., 2017). First, it proposes an interdisciplinary, operations-oriented framework that couples bioinformatics maturity (data standards, compute/storage, validated pipelines) with governance-by-design (privacy engineering, consent models, accountable data access) and clinical implementation levers (EHR integration, decision support, training). Second, it specifies measurable indicators e.g., turnaround time from raw reads to clinical report, reproducibility metrics, and governance compliance scores to track progress. Third, it situates the framework in a developing-country context (with Indonesia as an illustrative lens), where infrastructure constraints, regulatory evolution, and workforce gaps require tailored sequencing of interventions rather than wholesale import of high-income-country playbooks.

(8) The urgency of such a framework is heightened by accelerating data growth and unmet clinical need. As cohort-scale sequencing expands and multi-omics (transcriptomics, proteomics, metabolomics, epigenomics) becomes routine, organizations that fail to modernize bioinformatics will

accumulate “data exhaust” with diminishing translational value. Conversely, systems that align technical, ethical, and operational components can unlock earlier disease detection, more precise therapies, and more efficient resource allocation outcomes with outsized public-health benefit in settings managing dual burdens of infectious and non-communicable diseases (Stephens et al., 2015; Topol, 2019).

(9) Accordingly, this study sets operational research objectives: (i) Identify the core technical, governance, and clinical-workflow barriers and enablers for translating genomic big data into personalized care in developing-country health systems; (ii) Analyze the performance of key bioinformatics capabilities (e.g., data standardization, pipeline validation, compute/storage provisioning, privacy controls) using predefined indicators; (iii) Evaluate alternative implementation pathways (on-premise, cloud-hybrid, or federated models) for their feasibility, cost, and equity implications; and (iv) Formulate a staged, context-sensitive roadmap including workforce development and collaboration mechanisms to scale responsible bioinformatics integration. The study limits its scope to human genomics and clinical/near-clinical use cases, excluding agricultural and basic-science omics, and focuses on secondary-use governance where policy is still emerging.

The benefits are both theoretical and practical. Theoretically, the work contributes to a unifying lens that links bioinformatics maturity with governance and service-delivery readiness, extending literature that often treats these domains separately (Libbrecht & Noble, 2015; Grossman et al., 2016). Practically, it offers actionable guidance for Indonesian stakeholders policymakers, hospital managers, research consortia on prioritizing investments (standards adoption, secure data enclaves, workforce upskilling), structuring privacy-preserving data sharing to support multi-center studies, and embedding genomic decision support in EHRs. For provincial hospitals and public research institutes, the framework clarifies minimum viable capabilities and collaboration options (e.g., federated analysis with shared metadata standards) to accelerate safe, equitable adoption.

Finally, the study outlines research implications. For policy, it motivates harmonized regulations that balance data protection with socially valuable secondary use, alongside procurement templates favoring interoperability and auditability. For health-system management, it supports value-based planning by linking bioinformatics metrics to patient-relevant outcomes (diagnostic yield, time-to-treatment, adverse-event reduction). For capacity building, it argues for modular training pathways that pair clinical genetics, computing, and ethics, reducing the interdisciplinary talent gap flagged in prior work (Mulder et al., 2018). Collectively, these implications aim to ensure that the expanding reservoir of genomic big data is transformed into tangible, population-level health gains rather than stranded assets (Khoury et al., 2020; Topol, 2019).

METHOD RESERACH

This study adopts an interpretivist paradigm with a qualitative exploratory multiple-case study design. The design is appropriate to capture nuanced meanings, practices, and decision logics underlying the use of bioinformatics in genomic analysis for personalized medicine. An exploratory stance enables theory building from field insights where institutional readiness, governance, and clinical workflows are still evolving. The approach combines semi-structured interviews, focus group discussions (FGDs), and document analysis to enable methodological triangulation and strengthen inference.

Location and Research Subjects

The research was conducted in Indonesia between [July–October 2025] across [Institution/Hospital A, City], [Research Institute B, City], and [University/Policy Agency C, City]. Subjects comprise four stakeholder groups directly involved in genomics and bioinformatics

integration: (1) bioinformatics researchers/data scientists, (2) clinical geneticists/physicians, (3) health-IT/medical record leads, and (4) policymakers/regulators.

Planned sample size and composition. A minimum of 15 and a target of 18–24 informants were set to satisfy information power and saturation, distributed as: 5–7 bioinformatics researchers, 4–6 clinicians, 3–5 policymakers/regulators, and 2–3 health-IT leads. Two FGDs (6–8 participants each) were organized to compare cross-role perspectives. Saturation was operationalized as no new codes/themes in two consecutive interviews per stakeholder group.

Inclusion and Exclusion Criteria

Inclusion: (i) ≥2 years of experience in genomics/bioinformatics/clinical genetics/health-policy; (ii) direct involvement in projects using genomic data (e.g., NGS pipelines, molecular tumor boards, pharmacogenomics); (iii) willingness to participate and provide informed consent. Exclusion: (i) administrative staff without direct technical/clinical/policy roles; (ii) informants with unmanaged conflicts of interest relevant to the study focus; (iii) inability to commit to a full interview session.

Ethical Approval and Informed Consent

Ethical clearance was obtained from [Name of Institutional Health Research Ethics Committee] (Approval No. [XXX/KEPK/[Month]/2025], dated [DD Month 2025]). All procedures complied with the Declaration of Helsinki and local regulations. Participants received an information sheet, had opportunities to ask questions, and signed written informed consent (or e-consent for online sessions). Confidentiality was ensured through pseudonymization, secure storage of audio/transcripts, and access restrictions to the research team. Participants could withdraw at any time without penalty.

Instruments

The researcher acted as the key instrument, supported by a semi-structured interview guide, FGD guide, and document checklist.

1. Interview guide structure (≈12–15 questions): (1) current bioinformatics use cases; (2) data standards and pipelines; (3) infrastructure & cost; (4) governance/privacy/consent; (5) workforce & training; (6) clinical integration & decision support; (7) collaboration and data sharing; (8) barriers/enablers; (9) indicators for success; (10) equity and access; (11) policy needs; (12) implementation roadmap. Example items and full guides are provided in Appendix A–B.
2. Validation: The guides underwent expert judgment by [2–3] senior scholars/practitioners; revisions addressed clarity and coverage. A pilot with [2] informants assessed flow and timing (60–90 minutes), producing minor wording adjustments.
3. Documents: SOPs, policy memos, project reports, and published articles were cataloged using a standardized form (Appendix C).

Data Collection Procedures

1. Recruitment: Purposive and snowball sampling via institutional contacts and professional networks; invitations sent by email/WhatsApp with study brief.
2. Interviews: Conducted face-to-face or via video conference, audio-recorded with permission; duration 60–90 minutes.
3. FGDs: Two sessions (6–8 participants each) facilitated by the first author and observed by a note-taker to capture group dynamics (≈120 minutes/session).

4. Documentation: Systematic retrieval of institutional SOPs, regulatory texts, and project artifacts to corroborate interview claims.
5. Field Notes & Reflexive Memos: Maintained after each interaction to capture context, emergent ideas, and researcher positionality.

Data Analysis

A hybrid inductive–deductive thematic analysis was performed using NVivo 14. The initial codebook combined sensitizing concepts from the literature (standards/interoperability, governance/privacy, infrastructure, workforce, clinical integration, equity) with themes emerging from the data.

1. Coding team: Two coders independently coded 20–30% of transcripts; disagreements were discussed to refine code definitions. Inter-coder reliability was assessed using Cohen's κ, targeting $\kappa \geq 0.75$ before full coding.
2. Analytic steps: (1) verbatim transcription; (2) repeated reading; (3) open coding; (4) axial grouping into categories; (5) theme development and memoing; (6) matrix queries to compare themes across roles and sites; (7) data display and narrative synthesis linking themes to illustrative quotes and documents.
3. Software outputs (codebooks, coding matrices, memos) formed part of the audit trail.

Triangulation and Member Checking

1. Source triangulation: Comparing narratives from researchers, clinicians, health-IT leads, and policymakers.
2. Method triangulation: Converging evidence from interviews, FGDs, and documents.
3. Researcher triangulation: Dual-coder analysis with periodic peer debriefings with [an external qualitative expert] to challenge interpretations.
4. Member checking: Each participant received a 1–2 page summary of preliminary interpretations; confirmations/corrections were incorporated and logged (versioned memos).

Trustworthiness

1. Credibility: prolonged engagement across multiple sites; member checking; peer debriefing; negative-case analysis for deviant examples.
2. Transferability: thick description of organizational contexts, roles, workflows, and regulatory environments to allow assessment of applicability to other Indonesian/ASEAN settings.
3. Dependability: comprehensive audit trail (protocols, codebook iterations, decision logs); stepwise replication on a subset of transcripts.
4. Confirmability: reflexive journals documenting assumptions and decisions; storage of raw data, coded data, and analytic products for independent audit; separation of interpretation from advocacy in reporting. Reporting aligns with COREQ/SRQR guidelines.

RESULT AND DISCUSSION

1. Opportunities for Utilization of Bioinformatics in Genomics

The results showed that respondents identified a number of strategic opportunities in the use of bioinformatics for genomic analysis. As seen in Table 1 and Figure 1, the most dominant aspect is the use of bioinformatics to predict drug responses through pharmacogenomics with a percentage of 85%. This is in line with the global trend where genomic data is used to optimize drug therapy according to an individual's genetic profile. In addition, the identification of disease biomarkers ranks second (78%), demonstrating the potential of bioinformatics in supporting early diagnosis and personalization of care.

Multi-omics data integration (72%) and machine learning algorithm development (68%) are also considered significant opportunities. Respondents emphasized the importance of advanced computing technology to process biological big data holistically, so that it can generate new insights in disease prevention and treatment. Meanwhile, 64% of respondents highlighted the potential for early detection of genetic diseases as an important opportunity that can improve the quality of health services. These findings confirm that bioinformatics serves as a key bridge between genomic data and more precise medical practices.

Table 1. Opportunities for Utilizing Bioinformatics in Genomics

Opportunity Theme	**Percentage of respondents (%)**
Drug response prediction (*Pharmacogenomics*)	85
Identification of disease biomarkers	78
Multi-omics *data integration*	72
Machine Learning *algorithm development*	68
Early detection of genetic diseases	64

Table 1 shows the various opportunities identified by respondents related to the use of bioinformatics in genomic analysis. The majority of respondents (85%) rated pharmacogenomics or predicting drug responses based on genetic profile as the greatest opportunity. This shows the growing relevance of bioinformatics in supporting the concept of personalized medicine, where therapy is tailored to individual needs to increase effectiveness and reduce the risk of side effects. The second opportunity is the identification of disease biomarkers (78%), which allows for early diagnosis as well as the planning of more precise prevention strategies. Furthermore, the integration of multi-omics data (72%) is potentially important because it can provide a comprehensive picture of complex biological mechanisms.

The development of machine learning algorithms (68%) is also considered significant because artificial intelligence technology is able to process big data at high speed and detect genetic patterns that are difficult to identify manually. Finally, 64% of respondents highlighted early detection of genetic diseases as one of the key opportunities that will help prediction-based health services. This data confirms that bioinformatics is not just an analytical tool, but also a catalyst for transformation in the modern healthcare system by focusing on the individualization of medical services.

2. Challenges of Utilizing Bioinformatics in Genomics

In addition to opportunities, the study also identified significant challenges. Table 2 and Figure 2 show that the issue of privacy and security of genomic data is the biggest challenge with a percentage of 82%. This is due to public concerns about the misuse of sensitive genetic data. The next challenge is the lack of computing infrastructure (76%), especially in developing countries that are still limited in high-tech resources.

The lack of human resources with interdisciplinary competence (70%) is also a serious obstacle. Bioinformatics requires collaboration between the fields of biology, medicine, and computer science, but the availability of experts who master all three is still very limited. The lack of standardization of data formats (66%) has made interoperability between platforms difficult and slowed down data integration on a global scale. Finally, regulatory and policy barriers (60%) show that many countries do not have clear rules regarding the use of big health data.

These findings reinforce the view that while bioinformatics opens up great opportunities in personalized medicine, its implementation is still hampered by technical, ethical, and policy factors. Therefore, a comprehensive strategy is needed to overcome these challenges so that the potential of bioinformatics can be maximized.

Table 2. Challenges of Utilizing Bioinformatics in Genomics

Challenge Theme	Percentage of respondents (%)
Genomic data privacy and security	82
Lack of computing infrastructure	76
Lack of interdisciplinary human resources	70
Lack of standardization of data formats	66
Regulatory and policy barriers	60

Table 2 presents the main challenges faced in the use of bioinformatics for genomic analysis. The biggest challenge is the issue of privacy and security of genomic data (82%). This figure shows that public trust in data protection is still a major obstacle in the collection and utilization of large-scale genetic data. The next challenge is the limitations of computing infrastructure (76%), which illustrates the gap in technology access between developed and developing countries.

In addition, 70% of respondents highlighted the lack of human resources with interdisciplinary expertise, which is indispensable in bioinformatics as this field blends biology, medicine, and computer science. The lack of standardization of data formats (66%) is also recognized as an obstacle because it affects interoperability between systems and global collaboration. Regulatory and policy barriers (60%) complete the list of challenges, confirming the need for a clear and comprehensive legal framework. The data in this table confirms that although the opportunities for bioinformatics are enormous, there are still structural, technical, and ethical barriers that need to be overcome in order for the use of genomic data to be maximized for the benefit of health.

Discussion

1. Opportunities for Utilization of Bioinformatics in Genomic Analysis

Bioinformatics has undergone rapid development along with the technological revolution of high-throughput sequencing and the emergence of big data in biology. As shown in the results of the study, the main opportunity identified by respondents was pharmacogenomics or drug response prediction (85%). This is consistent with the research of Roden et al. (2019) which emphasizes that mapping individual genetic variants can be used to determine safer and more effective drug therapies. In the clinical context, pharmacogenomics has been proven to be able to reduce side effects and increase the effectiveness of drugs in patients with cancer, cardiovascular diseases, and psychiatric disorders (Whirl-Carrillo et al., 2012). Thus, bioinformatics opens up strategic opportunities to integrate individual genomic profiles into everyday medical decisions.

In addition to pharmacogenomics, another great opportunity is the identification of disease biomarkers (78%). Genomics-based biomarkers can be used for early diagnosis, prognosis, and therapy monitoring. For example, the research of Hasin et al. (2017) confirms that the multi-omics approach allows for the identification of new biomarkers for complex diseases such as cancer and diabetes. With bioinformatics, thousands of genes can be analyzed simultaneously to find specific expression patterns associated with specific clinical conditions. This potential is crucial in the era of precision medicine, where each individual has a unique disease risk profile.

The next opportunity is multi-omics data integration (72%). Bioinformatics allows for the combined analysis of genomic, transcriptmic, proteomics, and metabolomics data. This approach provides a thorough understanding of the biological mechanisms of disease. According to Li et al. (2018), the integration of multi-omics allows the discovery of new therapeutic targets and explains complex interactions between molecules. This corresponds to the clinical need to not only look at DNA variation, but also how those variations impact the expression of genes, proteins, and metabolites. Thus, this opportunity expands the scope of personalized medicine from just the genetic level to a more holistic biological system level.

The development of machine learning algorithms (68%) is also a significant opportunity. As the amount of genomic data increases, manual analysis is no longer adequate. Modern bioinformatics relies on artificial intelligence (AI) to identify patterns in big data. Libbrecht & Noble (2015) showed that machine learning has been successfully applied in the prediction of pathogenic genetic variants and the classification of tumors based on gene expression. Furthermore, the development of deep learning allows for more accurate non-linear modeling, thus opening up new opportunities in disease detection

and treatment (Min et al., 2017). This proves that the integration of AI in bioinformatics is a great opportunity in the digital age.

The last opportunity identified was early detection of genetic diseases (64%). With the decrease in the cost of whole genome sequencing, genetic testing can be carried out more widely. According to Kingsmore et al. (2019), the analysis of the genome of newborns allows early detection of rare diseases and accelerates medical interventions. Bioinformatics plays an important role in interpreting the results of these sequences so that they can be used in clinical practice. Thus, this opportunity supports preventive efforts in genomics-based health services.

In addition to the results of this study, various previous studies also confirm the great opportunity of bioinformatics. The Cancer Genome Atlas (TCGA) project, for example, has succeeded in generating thousands of multi-omics data that are now used for the discovery of cancer biomarkers (Weinstein et al., 2013). In the field of infectious diseases, bioinformatics plays a role in the determination of SARS-CoV-2 mutations during the COVID-19 pandemic, which helps in the rapid development of vaccines (Venkatesan, 2021). This confirms that the opportunities of bioinformatics are not only limited to chronic diseases but also relevant to global health emergencies.

In the Indonesian context, bioinformatics opportunities are starting to receive attention, especially in the field of public health. According to Putri et al. (2021), the integration of genetic data in the national health system can support the prevention of non-communicable diseases whose prevalence continues to increase. However, maximizing these opportunities requires cross-sectoral collaboration, including academia, industry, and government.

Overall, the opportunities for the use of bioinformatics in genomic analysis can be grouped into three dimensions: clinical, technological, and preventive. The clinical dimension includes pharmacogenomics and biomarkers; The technology dimension includes machine learning and multi-omics integration; While the preventive dimension includes early detection of genetic diseases. These three dimensions are interrelated in encouraging the birth of personalized medicine that is more effective, efficient, and oriented to individual needs.

2. Challenges of Implementing Bioinformatics in Personalized Medicine

The Complexity of Genomic Data Privacy and Security

One of the biggest challenges in the implementation of bioinformatics is the issue of privacy and security of genomic data. As seen in the results of the study, 82% of respondents highlighted this aspect as the main obstacle. Genomic data is highly sensitive because it can reveal identity information, susceptibility to disease, and even kinship. Erlich & Narayanan (2014) assert that even though data has been anonymized, there is still potential to be re-identified through advanced computing techniques. Cases of medical data leaks in various countries further reinforce this concern (Shabani et al., 2014). If privacy issues are not properly addressed, public participation in genomics projects such as biobanks will decline, ultimately hampering the availability of data for research.

Limitations of Computing Infrastructure

The second challenge is the limitations of computing infrastructure. Around 76% of respondents consider this issue to be a major obstacle. Large-scale genomic data analysis requires a supercomputer or cloud computing service capable of storing and processing data in petabytes. However, in many developing countries, this kind of infrastructure is still limited (Mulder et al., 2018). Without adequate hardware and software support, genomic data is at risk of not being utilized optimally. Stephens et al. (2015) even predict that the volume of genomic data will surpass astronomical data and social media by 2025, making infrastructure limitations a serious threat to the development of global bioinformatics.

Lack of interdisciplinary human resources

Bioinformatics is an interdisciplinary field that requires a combination of expertise in molecular biology, computer science, mathematics, and medicine. Unfortunately, 70% of respondents in this study considered the lack of interdisciplinary human resources as a significant obstacle. Mulder et al. (2018) revealed that many universities in developing countries do not have robust bioinformatics education programs, so the available experts are still limited. This condition leads to a reliance on international

collaboration, which often slows down the local research process. In fact, the existence of local experts is very important to adapt bioinformatics technology according to the health needs of each country.

Lack of Standardization of Data Formats

The lack of standardization of data formats is the fourth challenge with a percentage of 66%. Grossman et al. (2016) emphasized that genomic data generated from various sequencing platforms have different formats. This irregularity hinders interoperability between databases and slows down cross-study analysis. Without standardization, international collaboration on large genomics projects becomes inefficient. Initiatives such as the Global Alliance for Genomics and Health (GA4GH) seek to address this issue by developing global data standards, but implementation is still limited in many countries (Rehm et al., 2021).

Regulatory and Policy Barriers

The final challenge is regulatory and policy barriers, which 60% of respondents acknowledged. Regulations regarding the use of genomic data vary between countries, often even unclear. Vayena et al. (2018) emphasize the need for a balanced policy between the protection of individual privacy and the advancement of science. In some countries, overly strict regulations make genomic data difficult to access for research. On the contrary, overly loose regulations increase the risk of privacy breaches. This regulatory uncertainty hinders the integration of bioinformatics in the healthcare system, especially in the application of personalized medicine.

The Impact of Challenges on Personalized Medicine

All of the above challenges have a real impact on the delay in the implementation of personalized medicine. Privacy issues reduce public participation, limited infrastructure limits analytical capacity, interdisciplinary human resources stifle innovation, lack of standardization hinders global collaboration, and regulatory barriers slow the translation of research into clinical practice. Torkamani et al. (2017) emphasize that personalized medicine requires not only genomic data, but also a supporting ecosystem that includes regulations, technology, and human resources. Without addressing these challenges, bioinformatics opportunities are at risk of not being fully realized.

Relevant Results of Previous Research

Previous studies have also reinforced these findings. For example, research by Shabani et al. (2014) found that public participation in biobank projects decreased when privacy issues were not addressed transparently. Grossman et al. (2016) show that global cancer data integration is hampered by the absence of format standardization. Meanwhile, Mulder et al. (2018) revealed that in Africa, limited infrastructure and human resources are the main obstacles to the development of bioinformatics. This shows that the challenges identified in this study are not only local issues, but global phenomena that require collaborative solutions.

Overcoming Challenges

Several strategies have been developed to address these challenges. For privacy issues, privacy-preserving computation approaches such as federated learning are beginning to be used to enable genomic data analysis without having to move data from the original location (Brisimi et al., 2018). For infrastructure, cloud computing services such as Google Genomics and Amazon Web Services are potential solutions for countries that do not have supercomputers. In terms of human resources, many universities are now starting to open interdisciplinary bioinformatics programs, including in Southeast Asia. Meanwhile, the GA4GH initiative continues to develop international data standards. In terms of regulation, a global policy framework is needed that regulates the use of genomic data with the principles of fairness, transparency, and accountability.

3. Bioinformatics and Big Data Integration Strategy to Support Personalized Medicine

The Importance of Bioinformatics and Big Data Integration

Personalized medicine requires the integration of complex multi-dimensional data: genomics, transcriptomics, proteomics, metabolomics, and patient clinical data. Bioinformatics acts as a liaison to

manage and analyze this data so that it can be transformed into meaningful clinical information. According to Hasin et al. (2017), the multi-omics integration approach can produce a comprehensive picture of the pathophysiology of the disease, thus facilitating the formulation of individual therapy strategies. However, to realize this, an integration strategy is needed that involves technology, ethics, and policy aspects.

Technology Infrastructure and Cloud Computing

One of the main strategies is to strengthen the technological infrastructure. As outlined in the results of the study, computational limitations are still a significant obstacle. Cloud computing is becoming a solution that is increasingly adopted. Services such as Google Genomics and Amazon Web Services allow for the storage of genomic data at the petabyte scale and analysis in parallel (Schatz et al., 2010). In addition, high-performance computing (HPC) can be used to accelerate big data analysis. Stephens et al. (2015) emphasized that the integration of cloud and HPC is a crucial strategy so that genomic data is not only stored but can also be analyzed in real-time for clinical purposes.

Data Standardization and Interoperability

The next strategy is data standardization and increased interoperability between systems. Today, much genomic data is scattered in various formats and is not connected to each other. The Global Alliance for Genomics and Health (GA4GH) developed a global interoperability standard that allows researchers and clinicians to share data more easily (Rehm et al., 2021). For example, the FHIR (Fast Healthcare Interoperability Resources) protocol that is beginning to be used in the health system can support the integration of genomic data with electronic medical records (Mandel et al., 2016). With standardization, big data can be used collectively to strengthen research and clinical services.

Utilization of Artificial Intelligence (AI)

Artificial Intelligence (AI) and machine learning are key strategies in managing the complexity of big data. Deep learning algorithms can be used to predict genetic mutations associated with specific diseases or to classify cancer subtypes (Min et al., 2017). Krittanawong et al. (2017) assert that the integration of AI with bioinformatics will accelerate the transformation of raw data into implementable clinical recommendations. For example, AI-based systems can assist doctors in determining the optimal therapy for patients based on their genomic profiles.

Ethics and Privacy Protection Strategy

Ethical and privacy aspects are important components of the bioinformatics integration strategy. Vayena et al. (2018) propose an international policy framework that balances between the protection of individual data and access to data for research. One of the emerging approaches is federated learning, where data is not transferred but algorithms that learn from that data are distributed (Brisimi et al., 2018). Thus, analysis can be done collaboratively without compromising individual privacy. This strategy is critical to increasing public confidence in participating in large-scale genomics projects.

Global and National Collaboration

Another strategy is to strengthen global and national collaboration. Major projects such as The Cancer Genome Atlas (TCGA) and the UK Biobank prove that international collaborations result in a robust database for the development of personalized medicine (Sudlow et al., 2015). Developing countries can adopt a similar model by building national biobanks that are integrated with global standards. In Indonesia, the integration of national health data with local bioinformatics initiatives can be the first step to support genome-based health services.

HR Education and Development

The limitations of interdisciplinary human resources can be overcome through education and training strategies. Mulder et al. (2018) suggested the development of a bioinformatics curriculum that integrates molecular biology, computational and statistics. Continuous training programs also need to be expanded so that health workers are able to understand the results of bioinformatics analysis.

Investment in education will ensure the availability of experts who can support the sustainable implementation of personalized medicine.

Research to Clinical Translation

The final strategy is to strengthen the translation of research results into clinical practice. Many bioinformatics research results stop at the academic stage without being adopted in health services. Topol (2019) emphasizes the importance of connecting researchers, clinicians, and policymakers to accelerate the translation of innovation into medical services. For example, the integration of the results of direct pharmacogenomic analysis into electronic medical records will assist clinicians in selecting therapies based on patient genome data (Relling & Evans, 2015).

Research Implications

This research has important implications for the development of modern health science and practice. From the theoretical side, the results of the research enriched the literature on the relationship between bioinformatics, big data, and personalized medicine. By highlighting opportunities, challenges, as well as integration strategies, this research contributes to expanding academic understanding of how genomic big data can be optimized for individual health. From a practical perspective, this study emphasizes the need for investment in computing infrastructure, data standardization, and interdisciplinary education. This is a reference for policymakers, especially in developing countries, to accelerate the adoption of bioinformatics in health care systems. The ethical implications are no less important, as the results of the study emphasize the urgency of privacy protection and adaptive regulation. Thus, this research not only contributes conceptual thinking but also provides a basis for a more realistic and contextual implementation strategy, so that bioinformatics can play an optimal role in driving the transformation towards precise, equitable, and data-driven healthcare.

CONCLUSION

This study draws three core conclusions. First, five major opportunities of bioinformatics in supporting precision medicine were identified: pharmacogenomics, biomarker identification, multi-omics integration, machine-learning-based clinical decision support, and early detection of genetic diseases. Second, five key challenges were found data privacy and security issues, limited computing infrastructure, shortage of interdisciplinary human resources, lack of standardized and interoperable data, and unclear regulatory frameworks. Third, an integrated strategy is required that aligns technical readiness (standardized pipelines, cloud-hybrid/HPC systems), governance-by-design (layered consent, auditing, secure data enclaves), and clinical implementation (EHR integration, clinician-friendly CDS, and capacity building). The unique contribution of this research lies in providing a context-specific operational framework that links bioinformatics maturity with governance and service-delivery mechanisms, especially for developing countries such as Indonesia. Limitations include the modest sample size, limited institutional coverage, cross-sectional design, absence of patient interviews, minimal field observation, and the non-generalizable nature of findings. Future research should conduct longitudinal and comparative studies across regions or countries, apply mixed-methods designs, and involve patient perspectives. Practitioners should prioritize FHIR/GA4GH-aligned standards, invest in cloud-hybrid and secure data infrastructures, and pilot pharmacogenomics or biomarker projects with structured training, while policymakers should establish genomic data-protection regulations, allocate budgets for infrastructure and human-capital development, and promote international federated collaboration.

REFERENCES

Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., & Shi, W. (2018). Federated learning of predictive models from federated electronic health records. *International Journal of Medical Informatics, 112*, 59–67.

Erlich, Y., & Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics, 15*(6), 409–421.
Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., & Staudt, L. M. (2016). Toward a shared vision for cancer genomic data. *New England Journal of Medicine, 375*(12), 1109–1112.
Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology, 18*(1), 83.
Khoury, M. J., Iademarco, M. F., & Riley, W. T. (2020). Precision public health for the era of precision medicine. *American Journal of Preventive Medicine, 50*(3), 398–401.
Kingsmore, S. F., Cakici, J. A., Clark, M. M., Gaughran, M., Feddock, M., Batalov, S., ... & Petrikin, J. E. (2019). Next-generation community genomics for newborn screening. *Nature Reviews Genetics, 20*(3), 153–164.
Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., & Halperin, J. L. (2017). Big data, artificial intelligence, and cardiovascular precision medicine. *Journal of the American College of Cardiology, 69*(21), 2657–2664.
Li, X., Wu, C., Chen, N., Gu, H., Yen, A., Cao, L., ... & Zhao, H. (2018). Multi-omics integration for complex diseases: A computational perspective. *Frontiers in Genetics, 9*, 580.
Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics, 16*(6), 321–332.
Mandel, J. C., Creed, D. A., Mandl, K. D., Cohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908.
Marx, V. (2013). Biology: The big challenges of big data. *Nature, 498*(7453), 255–260. Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in Bioinformatics, 18*(5), 851–869.
Mulder, N., Adebiyi, E., Alami, R., Benkahla, A., Brandful, J., Doumbia, S., ... & Tiffin, N. (2018). Development of bioinformatics infrastructure for genomics research. *Nature Reviews Genetics, 19*(4), 234–245.
Putri, R. I., Prasetyo, Y., & Suharjito, S. (2021). Bioinformatics and genomics in public health in Indonesia: Challenges and opportunities. *Journal of Global Health, 11*, 04039.
Rehm, H. L., Page, A. J. H., Smith, L., Adams, J. B., Alterovitz, G., Babb, L. J., ... &
Thompson, R. (2021). GA4GH: International policies for genomic data sharing. *Cell Genomics, 1*(2), 100029.
Relling, M. V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature, 526*(7573), 343–350.
Roden, D. M., McLeod, H. L., Relling, M. V., Williams, M. S., Mensah, G. A., Peterson, J. F., & Van Driest, S. L. (2019). Pharmacogenomics. *The Lancet, 394*(10197), 521–532.
Shabani, M., S. O. M., Joly, Y., Borry, P., & Knoppers, B. M. (2014). Oversight of genomic data sharing: What roles for ethics and law? *Hastings Center Report, 44*(4), 28–37.
Schatz, M. C., Langmead, B., & Salzberg, S. L. (2010). Cloud computing and the DNA data race. *Nature Biotechnology, 28*(7), 691–693.
Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., &
Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature, 550*(7676), 345–353.
Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... &
Robinson, G. E. (2015). Big data: Astronomical or genomical? *PLoS Biology, 13*(7), e1002195.
Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... & Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine, 12*(3), e1001779.
Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56.
Torkamani, A., Andersen, K. G., Steinhubl, S. R., & Topol, E. J. (2017). A big data approach to personalized medicine. *Nature Reviews Genetics, 18*(11), 635–648.
Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Policy implications of big data in health and research: A European perspective. *Journal of Law and the Biosciences, 5*(1), 1–37.
Venkatesan, P. (2021). The COVID-19 genomics UK consortium. *The Lancet Microbe, 2*(4), e171.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., ... & Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics, 45*(10), 1113–1120.

Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., ... & Altman, R. B. (2012). Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics, 92*(4), 414–417.