GAmBBa: A Platform for Automated Laboratory Workflow Management and Molecular Surveillance Biobanking

Andrêza Leite de Alencar[1], **Saulo Gomes**[4], **Karina Moura** [5], **Jessica Silva** [2], **Nivison Nery Jr** [2], **Ricardo Khouri**[2,3]

[1]Departamento de Ciência da Computação,
Universidade Federal Rural de Pernambuco (UFRPE) Recife, PE, Brazil

[2]Laboratório de Medicina e Saúde Pública de Precisão, Instituto Gonçalo Moniz,
Fundação Oswaldo Cruz, Salvador, BA, Brazil

[3]Instituto Nacional de Ciência e Tecnologia em Saúde Digital (DigiSaude)
Salvador, BA, Brazil

[4]Universidade Federal de Pernambuco (UFPE), Recife, PE, Brazil

[5]Instituto Federal de Pernambuco (IFPE), Recife, PE, Brazil

andreza.leite@ufrpe.br

Abstract. *Molecular surveillance and the diagnosis of infectious and neglected diseases require a robust and traceable data flow, spanning from sample collection to the interpretation and reporting of genomic results. However, in many laboratory settings, these processes still depend on manual procedures and spreadsheet-based workflows, which increase the risk of transcription errors, compromise traceability, and limit scalability. At the Molecular Surveillance Platform (PVM) at Fiocruz, these challenges became particularly evident during high-demand scenarios such as the COVID-19 pandemic, when 47,220 samples from 16 municipalities in Bahia, Brazil, were processed to support SARS-CoV-2 diagnosis. In this study, we present GAmBBa (Management of Biological Samples by Barcode), a web-based computational platform designed to automate laboratory workflow management, enable end-to-end sample tracking through barcode identification, integrate communication between laboratory instruments and information systems, and support biobank storage and retrieval. GAmBBa structures the entire sample lifecycle digitally, reducing manual data entry and improving interoperability across laboratory processes. The implementation of GAmBBa has enhanced operational efficiency by minimizing errors, increasing processing capacity, and reducing turnaround time for diagnostic results. The platform proved particularly valuable in high-throughput environments, supporting large-scale molecular surveillance and strengthening the reliability of laboratory outputs. Our findings demonstrate that integrated digital solutions such as GAmBBa are essential to modernize laboratory infrastructures and to support timely and scalable responses to public health demands.*

1. Introduction

Advanced molecular surveillance is essential for monitoring pathogens, supporting timely public health responses, and improving the diagnosis of neglected diseases. However, the reliability and scalability of these activities depend on a well

structured laboratory workflow that ensures traceability, data integrity, and efficient communication across the different stages of sample processing [Mrazek et al. 2020, Office of the National Coordinator for Health Information Technology 2025]. In practice, many laboratories still rely on manual procedures, spreadsheets, and fragmented systems to register samples, transfer results, and reconcile clinical information, which increases the risk of transcription errors and delays in diagnosis [Mrazek et al. 2020, Bennett et al. 2018].

The literature on total laboratory testing shows that most errors occur in the pre analytical phase, where patient and sample identification, labeling, aliquoting, and data transcription are highly vulnerable to manual failures [Mrazek et al. 2020, Akingbade et al. 2024]. These weaknesses are especially problematic in molecular surveillance settings, where each sample must be tracked through multiple analytical steps and linked to metadata such as collection site, clinical context, and genomic result. In this scenario, automation becomes a key strategy to reduce human intervention, improve turnaround time, and standardize laboratory operations [Berg et al. 2020].

Biobanking adds another layer of complexity to laboratory workflows because biological samples must be stored, retrieved, and documented under strict quality requirements. The ISO 20387 standard defines general requirements for biobanking and emphasizes competence, quality management, and the traceability of biological materials and associated data [ISO 2018]. Automated storage and retrieval systems have been shown to improve sample integrity, reduce handling time, and support large-scale translational research, particularly when large biospecimen collections must be managed efficiently [Berg et al. 2020]. Therefore, a digital platform that integrates sample tracking and biobank management is not only desirable but necessary for laboratories that handle high volumes of biospecimens [ISO 2018, Berg et al. 2020].

Interoperability is another central requirement for modern laboratory environments. Laboratory information systems, public health databases, and analytical instruments often operate as isolated components, creating information silos that hinder workflow continuity and require repeated manual data entry [O'Carroll et al. 2016, Bennett et al. 2018]. Standards based integration improves the exchange of laboratory data and supports more reliable reporting across clinical and public health systems [Office of the National Coordinator for Health Information Technology 2025]. In this context, software platforms that mediate communication between systems and machines can substantially improve operational efficiency and data consistency [O'Carroll et al. 2016].

Within the Molecular Surveillance Platform (PVM) at Fiocruz, these challenges are amplified by the need to handle samples from external public health laboratories, manage aliquoting and internal tracking, process molecular results, and store institutional interest specimens in a biobank. These demands are not only conceptual but operational, involving a high-throughput environment with continuous inflow of samples, daily laboratory processing routines, and substantial monthly volumes of molecular analyses and nucleic acid extractions. During the COVID-19 pandemic, for instance, the PVM processed 47,220 samples originating from 16 municipalities in the state of Bahia, Brazil, to support SARS-CoV-2 diagnosis. This large-scale operation highlighted critical bottlenecks in manual workflows, including sample registration, tracking across multiple labo-

ratory steps, and integration of results generated by different instruments. The reliance on spreadsheet-based routines and manual data entry limited scalability, increased the likelihood of transcription errors, and delayed the release of diagnostic results at a time when rapid response was essential.

The experience gained in this context reinforced the need for an integrated, automated, and scalable solution capable of supporting both routine laboratory operations and emergency scenarios. To address these limitations, we developed GAmBBa, a web-based platform designed to automate laboratory workflow management, support biobank organization, and improve interoperability across the molecular surveillance ecosystem. By structuring the sample life cycle digitally, enabling barcode-based tracking, and facilitating system-to-system and system-to-machine communication, GAmBBa reduces manual effort, strengthens traceability, and improves the efficiency and reliability of laboratory processes. The implementation of this platform has contributed to optimizing sample processing, increasing analytical capacity, and accelerating the turnaround time for diagnostic results, particularly in high-demand settings such as public health emergencies.

2. The Molecular Diagnostic Workflow: Problem Description and Motivation

The current molecular diagnostic process is a multi stage pipeline. It begins with the **(1)** Reception of samples at the laboratory; these samples are sourced from various origins, in general LACENs (Central Public Health Laboratories) [1], and are already registered in the GAL system [2], which provides their primary identifier (ID). Upon reception, the workflow proceeds to **(2)** Aliquoting, where samples are divided into aliquots and need the assignment of a secondary IDs to ensure precise tracking throughout the internal laboratory pipeline. This is followed by **(3)** Sample Processing and Extraction, where samples undergo chemical or mechanical processing to isolate genetic material, such as RNA extraction.

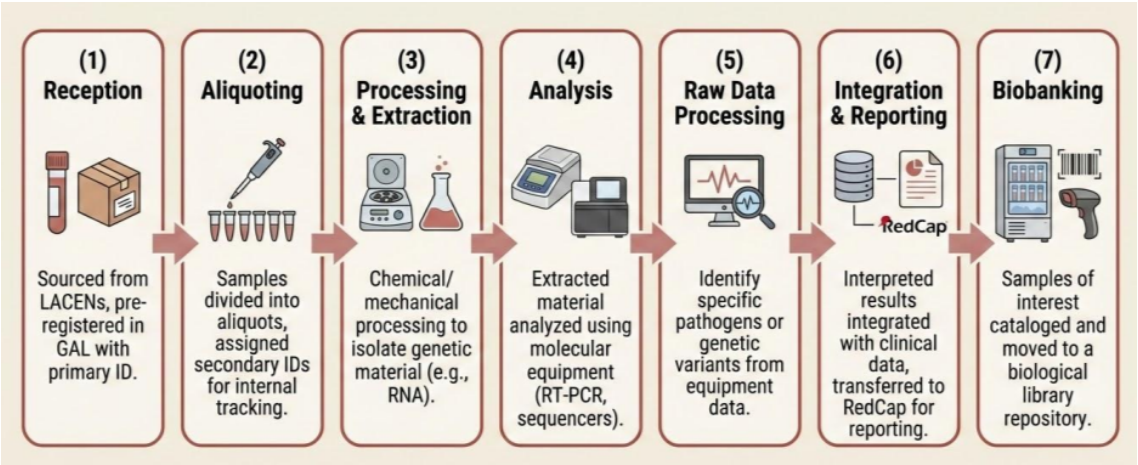

Figure 1. The Molecular Diagnostic Process. A multi-stage pipeline

(4) The extracted material is analyzed using molecular equipment, primarily thermal cyclers (e.g., for RT-PCR) or sequencers. Following the analysis **(5)** the raw data

[1] https://www.gov.br/saude/pt-br/composicao/svsa/sislab/lacen
[2] http://gal.datasus.gov.br/GALL/index.php

generated by the equipment is processed to identify specific pathogens or genetic variants. **(6)** The interpreted results need to be integrated with the patient's clinical and sociodemographic information. The diagnostics results are transferred manually to RedCap[3], where researchers can integrate, explore and generate customizable reports.

(7) Post analysis, samples of institutional interest are cataloged and moved to a biobank. This repository functions as a large scale biological library.

Despite this structured sequence, the workflow faces significant bottlenecks due to a lack of systemic integration. Currently, the PVM relies heavily on manual operations and extensive use of spreadsheets for archiving results, interpreting data, and transferring information between platforms. These gaps result in:

- **Excessive Human Intervention**: Data must undergo exhaustive manual conversion and treatment routines before it can be sent to other systems.
- **Operational Risks**: The high level of human intervention increases the probability of errors during the exchange of information between systems and machines.
- **Reduced Productivity and Quality**: Manual workflows lead to increased time expenditure across the entire process, from initial archiving to the final delivery of diagnostics.
- **Complex Interpretation Challenges**: The need to associate the simultaneous detection of different genes manually can lead to inconsistent interpretations and delayed response times.

The demand for a solution that automates the communication between system-to-system, system-to-machine, and machine-to-system is therefore critical to reducing processing time and increasing the quality of molecular surveillance in public health laboratories

3. Related Work: Automation, Biobanking, Traceability and Interoperability

The scaling of molecular diagnostics and public health surveillance relies on the convergence of automation, robust biobanking, and interoperable informatics. Automation in medical laboratories has been proven to increase throughput while minimizing human error and shortening turnaround times [Islam et al. 2023]. In scenarios of limited resources, such as the surveillance of neglected tropical diseases, versatile and modular workflows—including open-source robotics and isothermal amplification—offer cost-effective alternatives to manual processing [Iturritza et al. 2024].

Regarding sample management, automated preparation improves precision and analyst safety [More et al. 2024]. A critical factor for public health laboratories is traceability; monitoring individual process steps through automated systems ensures the integrity of sample data from reception to storage [Thurow 2023]. This level of control is a cornerstone of modern laboratory quality, as established by the ISO 20387 standard for biobanking, which emphasizes the need for consistent monitoring of biological materials and their associated metadata [ISO 2018].

Furthermore, laboratory information management systems (LIMS) and informatics are now inseparable from automation technologies [Islam et al. 2023]. Interoperability remains a critical challenge, as public health systems often operate as isolated

[3]https://redcapbrasil.com.br

silos. The implementation of ”lean” workflows, supported by real-time visibility and automated reporting, has shown significant clinical benefits in reducing the time to results [Mencacci et al. 2023]. For complex workflows like Next-Generation Sequencing (NGS), informatics integration is the primary strategy to overcome throughput and standardization barriers in public health labs [Socea et al. 2023]. Despite challenges such as platform costs, the integration of collaborative robotics and real-time quality control defines the next phase of innovation in automated diagnostics [Brown and Badrick 2022, Thurow 2023].

4. The GAmBBa Solution

The GAmBBa system was developed to serve as the management core for the molecular surveillance ecosystem.

4.1. System Architecture and Technical Implementation

The GAmBBa platform is built upon a robust multi-tier architecture designed to ensure scalability, data integrity, and cross-platform accessibility. The system's core is developed using the Laravel 8 framework[4] and , which serves as the backend engine, managing the application's business logic, authentication protocols, and API integrations. This backend communicates directly with a MySQL database management system[5], where all laboratory metadata, including sample IDs, storage locations, and genomic results, are stored in a structured and relational format. The system provides two primary user interfaces: a Web interface, optimized for administrative tasks and complex laboratory workflows (such as plate mapping and biobank management), and a Mobile interface, for quick data entry during sample reception and manipulation. The interaction between these tiers is mediated through a secure service layer, ensuring that clinical and sociodemographic data are handled according to the anonymization requirements of the LGPD.

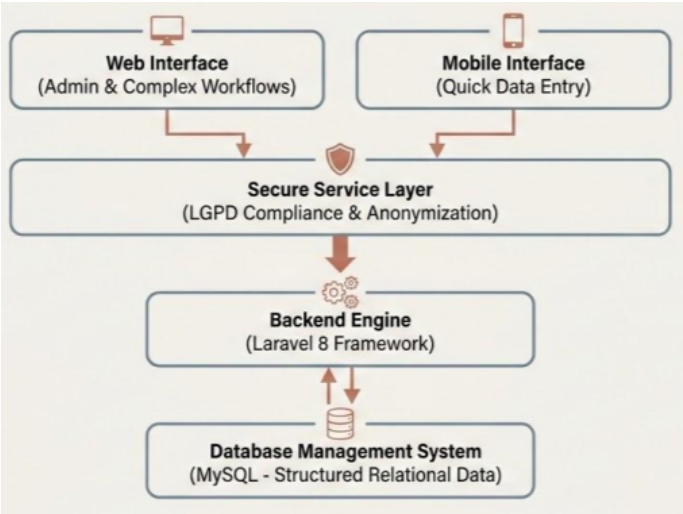

Figure 2. GAmBBa system Architecture

[4]https://laravel.com
[5]https://www.mysql.com

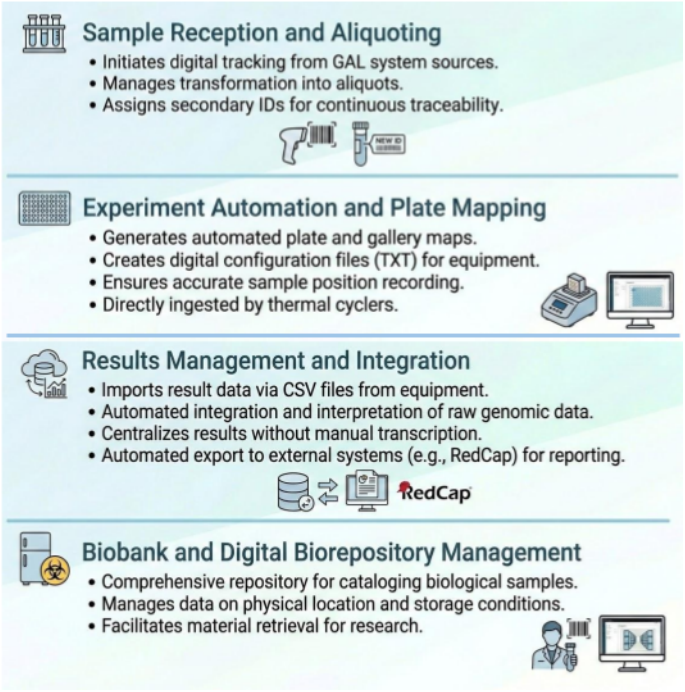

Figure 3. GAmBBa's main functionalities

4.2. GAmBBa Main Functionalities

The platform is structured into modular components that mirror the physical workflow of a molecular biology laboratory as described before. These modules are designed to replace manual spreadsheet based tasks with automated data management.

- **Sample Reception and Aliquoting**: This module initiates the digital tracking process by integrating information received from the sources through the GAL system. Upon reception, the system manages the transformation of these samples into aliquots, assigning secondary IDs that ensure continuous traceability across the entire pipeline.
- **Experiment Automation and Plate Mapping**: To streamline the preparation for molecular tests, the system generates automated plate and gallery maps. These maps serve as the primary digital input for generating configuration files for molecular equipment, such as thermal cyclers, ensuring that sample positions are accurately recorded before the analysis phase. GAmBBa bridges the gap between digital management and physical laboratory analysis through specific file based interoperability. The system generates configuration files in TXT format, tailored for direct ingestion by thermal cyclers to automate plate setup.
- **Results Management and Integration**: Once the molecular analysis is complete, the platform imports result data via CSV files exported by the equipment, allowing for the automated integration and interpretation of raw genomic data without manual transcription. By centralizing these results, the platform also performs the automated transfer (exportation) to an external management system (RedCap) that allows for the generation of customizable reports linking the molecular findings directly to the patient's clinical and sociodemographic information.
- **Biobank and Digital Biorepository Management**: This module acts as a comprehensive repository for cataloging biological samples of institutional interest.

GAmBBa manages the precise data regarding the physical location and storage conditions of each sample, facilitating the retrieval of materials for subsequent research and longitudinal studies.

It is important to notice that GAmBBa implements strict security protocols for handling sensitive patient information to ensure compliance with the Brazilian General Data Protection Law (LGPD). The system utilizes data anonymization techniques for clinical and sociodemographic records, ensuring that personal identifiers are decoupled from molecular results during large-scale analysis. Access is controlled through a multi-level authentication module, as shown in the mobile interface 4, which restricts data visibility based on user roles and institutional mandates.

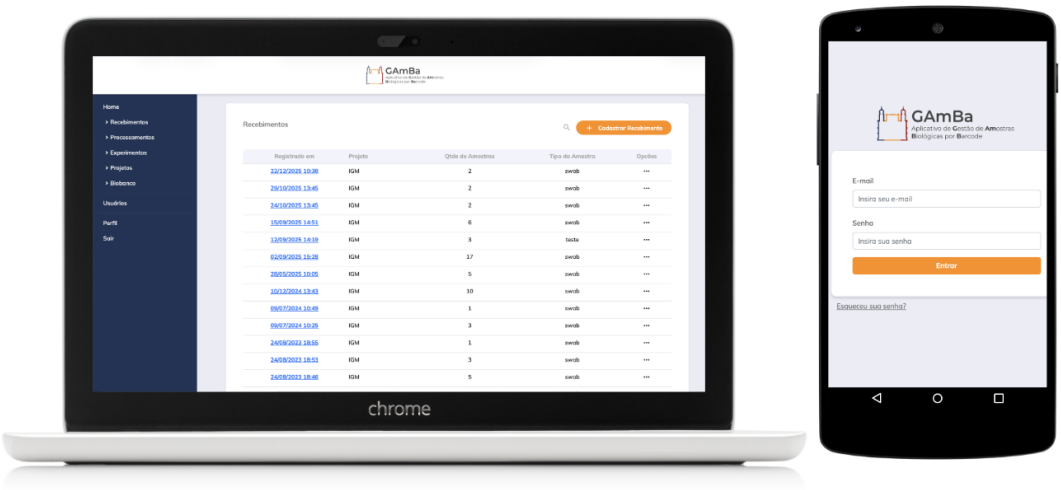

Figure 4. GAmBBa system interface exemple: desktop version displaying the molecular surveillance pipeline with sample reception and mobile version for the authentication module.

4.3. Technical Differential and Expected Results

Unlike generic LIMS, GAmBBa focuses on *system-to-machine (S2M)* communication, specifically through the automated generation of plate maps that configure thermal cyclers. This eliminates manual gallery mapping, a major source of operational risk. Additionally, the platform enables seamless data exportation to RedCap, facilitating the integration of clinical and molecular data for rapid reporting. The ongoing development of a decision making algorithm for results interpretation aims to automate complex interpretations, further reducing human intervention in the final diagnostic stage.

The implementation of the platform aims to drastically reduce the response time between collection and the availability of the diagnosis in LACENs. Expected outcomes include:

- **Increased Data Quality**: Reduction of errors caused by human intervention.
- **Productivity Gains**: Automation of communication between systems and actors in the ecosystem.
- **Enhanced Reporting**: Integration with data management systems (e.g., RedCap) to generate customizable reports that include sociodemographic and clinical information.

4.4. Preliminary Evaluation and Usability Testing

To evaluate the effectiveness and user experience of the GAmBBa platform, a controlled usability test was conducted with four participants (n=4) from the laboratory team. The group consisted of professionals familiar with molecular analysis and sample management processes.

Methodology and Protocol: The evaluation was performed using the Maze usability testing platform[6], which allowed for remote monitoring through screen captures and mission-tracking. The testing protocol was strictly restricted to desktop environments to ensure a standardized assessment of the web interface. The session protocol included:

- Pre-test Phase: Collection of demographic data and assessment of prior experience with sample management systems.
- Active Testing Phase: Participants were integrated into a simulated project and tasked with executing the end-to-end laboratory workflow.
- Post-test Phase: A survey focused on qualitative feedback, general impressions, and suggestions for improvement.

Evaluated Tasks: Participants were required to perform three core missions using real or simulated IDs from the GAL system:

- Sample Reception: Registering incoming samples and generating internal secondary IDs.
- Sample Processing: Managing the transformation of biological material and generating aliquots.
- Experiments/Analysis: Executing the final analysis and managing tests results.

Results and Findings: The system demonstrated high efficacy across all evaluated procedures. Quantitative data retrieved from Maze revealed a 100% mission completion rate for all participants across the three main workflows. Furthermore, 100% of the participants followed the expected interaction paths defined by the developers. Average task durations were recorded as follows: Sample Reception: 208.5 seconds; Processing: 1,172.0 seconds; and Experiments: 2,007.0 seconds. Qualitative feedback indicated that the platform provides essential resources to optimize the experimentation process. Participants highlighted the automated mapping(system-to-lab and system-to-machines) and tracking features as the most useful tools. No significant difficulties were reported in navigating the system or understanding the laboratory stages, confirming that the tool effectively supports the molecular surveillance workflow.

5. Conclusion and Future Work

GAmBBa represents a strategic advancement in the digital transformation of molecular surveillance at Fiocruz. By replacing fragmented, manual spreadsheet based processes with an integrated web based architecture, the platform establishes a robust framework for data integrity and laboratory traceability. The automation of the workflow, from sample reception and automated plate mapping to biobank cataloging, mitigates operational risks and addresses the urgent need for scalable diagnostic solutions in the context of tropical and neglected diseases in Brazil. Thus, GAmBBa is not merely a management tool, but

[6]https://app.maze.co/home

a critical infrastructure for accelerating public health response times and enhancing the quality of molecular data within the national surveillance ecosystem.

The next phase of the platform involves the implementation of an innovative algorithm designed for the automated interpretation of molecular results. This module aims to address specific technical and conceptual challenges that arise when laboratory results must be translated into clinical diagnostics. It will implement decision making logic to manage cases where multiple detections lead to varying or complex interpretations of the same sample. This feature will be integrated directly into the data flow, utilizing the state of the art methodologies (such as RT-PCR) to provide rapid, standardized diagnostic outputs. The current version of this diagnostic interpretation module is based on standardized logic from commercial and institutional testing kits. This rule based approach handles the detection of multiple genes simultaneously to provide consistent diagnostic outputs. Future iterations of the algorithm will incorporate Machine Learning (ML) models to identify complex genomic patterns and enhance the predictive accuracy of the surveillance system as the database of molecular results expands. By automating this final analytical step, the platform will support another routine currently performed manually and ensure higher consistency in the results delivered to the public health ecosystem.

Furthermore, future research will explore the semantic enrichment of the persisted data by integrating the platform with established biomedical ontologies. By leveraging semantic web technologies, the system aims to standardize metadata representation, enabling advanced data discovery and seamless interoperability with international public health repositories. This semantic layer will provide a formal structure to the genomic and clinical relationships stored in GAmBa, facilitating more complex cross-platform data integration and knowledge extraction in the field of molecular surveillance

AI Usage Declaration

Gemini 3 was utilized for text editing, grammatical refinement, image generation and enhancing clarity. All suggestions generated by the tool were critically analyzed and entirely reviewed by the authors, who assume full responsibility for the final content presented.

References

Akingbade, A. et al. (2024). Preanalytical errors in clinical laboratory testing at a glance. *Cureus*, 16(3).

Bennett, E. et al. (2018). Evaluation of data exchange process for interoperability and impact on electronic laboratory reporting quality to a state public health agency. *Online Journal of Public Health Informatics*, 10(2):e216.

Berg, M. et al. (2020). Automated sample storage in biobanking to enhance translational research. *Biopreservation and Biobanking*, 18(1):1–10.

Brown, A. and Badrick, T. (2022). The next wave of innovation in laboratory automation: systems for auto-verification, quality control and specimen quality assurance. *Clinical Chemistry and Laboratory Medicine (CCLM)*, 61:37 – 43.

Islam, S. U., Kamboj, K., and Kumari, A. (2023). Laboratory automation and its effects on workflow efficiency in medical laboratories. *Middle East Journal of Applied Science amp; Technology*.

ISO (2018). Biotechnology – biobanking – general requirements for biobanking.

Iturritza, M. U., Mlotshwa, P., Gantelius, J., Alfvén, T., Loh, E., Karlsson, J., Hadjineophytou, C., Langer, K., Mitsakakis, K., Russom, A., Jönsson, H. N., and Gaudenzi, G. (2024). An automated versatile diagnostic workflow for infectious disease detection in low-resource settings. *Micromachines*, 15.

Mencacci, A., Socio, G. D. D., Pirelli, E., Bondi, P., and Cenci, E. (2023). Laboratory automation, informatics, and artificial intelligence: current and future perspectives in clinical microbiology. *Frontiers in Cellular and Infection Microbiology*, 13.

More, D., Khan, N., Tekade, R., and Sengupta, P. (2024). An update on current trend in sample preparation automation in bioanalysis: strategies, challenges and future direction. *Critical Reviews in Analytical Chemistry*, 55:1461 – 1485.

Mrazek, C. et al. (2020). Errors within the total laboratory testing process, from test selection to medical decision-making – a review. *Biochemia Medica*, 30(2):020502.

O'Carroll, P. W. et al. (2016). Interoperability of information systems managed and used by the public health system. *BMC Public Health*, 16:1–12.

Office of the National Coordinator for Health Information Technology (2025). Laboratory data standards for interoperability.

Socea, J. N., Stone, V. N., Qian, X., Gibbs, P. L., and Levinson, K. J. (2023). Implementing laboratory automation for next-generation sequencing: benefits and challenges for library preparation. *Frontiers in Public Health*, 11.

Thurow, K. (2023). Strategies for automating analytical and bioanalytical laboratories. *Analytical and Bioanalytical Chemistry*, pages 1 – 10.