

Review

Not peer-reviewed version

---

# Software Applications in Biomedicine: A Narrative Review of Translational Pathways from Data to Decision

---

[Gabriela Georgieva Panayotova](#) \*

Posted Date: 6 January 2026

doi: 10.20944/preprints202601.0462.v1

Keywords: bioinformatics; molecular modeling; epidemiology; workflows; reproducibility; machine learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Software Applications in Biomedicine: A Narrative Review of Translational Pathways from Data to Decision

Gabriela Georgieva Panayotova

Division of Physiology, Department of Physiology and Pathophysiology, Medical University of Varna, 9002 Varna, Bulgaria; gabrielapanayotova95@gmail.com

## Abstract

**Background/Objectives:** Software has become core infrastructure in biomedical science; however, tools and workflows remain fragmented across subfields, limiting reproducibility and slowing translation from data generation to actionable decisions. This narrative review synthesizes representative software ecosystems across three major pillars—bioinformatics, molecular modeling and simulations, and epidemiology/public health—and evaluates their translational readiness using a shared cross-domain framework focused on reproducibility, validation, interoperability, usability, and decision relevance. **Methods:** A narrative review of articles indexed in PubMed/NCBI, Web of Science, and Scopus between 2000 and 2025 was conducted. Domain-specific terms related to bioinformatics, molecular modeling, docking, molecular dynamics, epidemiology, public health, and workflow management were combined with software- and algorithm-focused keywords. Studies describing, validating, or applying documented tools with biomedical relevance were included. **Results:** Across domains, mature data standards and reference resources (e.g., FASTQ, BAM/CRAM, VCF, mzML), widely adopted platforms (e.g., BLAST/BLAST+, Bioconductor, AutoDock Vina, GROMACS, Epi Info, QGIS), and increasing use of workflow engines were identified. Software pipelines routinely transform molecular and surveillance data into interpretable features supporting hypothesis generation. **Conclusions:** Integrated, standards-based, and validated software pipelines can shorten the path from measurement to decision in biomedicine and public health. Future progress depends on reproducibility practices, benchmarking, user-centered design, portable implementations, and responsible deployment of machine learning.

**Keywords:** bioinformatics; molecular modeling; epidemiology; workflows; reproducibility; machine learning

---

## 1. Introduction

Over the past three decades, the biomedical sciences have undergone a profound turning point: highly specialized instruments—from sequencers and mass spectrometers to microscopes and portable biosensors—now generate data of such volume and complexity that manual analysis is no longer feasible [1–3]. As a result, software has moved from a historically supporting role to a central driver of discovery, translation, and decision-making [4–6]. Today, fundamental questions in biology and medicine—how genes and proteins function, how molecules interact, how diseases spread, and how interventions should be implemented—are routinely addressed using diverse programs that combine statistics, machine learning, and data storage and analysis [7–9].

Three classes of software have been particularly important historically. First, bioinformatics tools now convert raw biological signals into interpretable features: short reads become assembled genomes; expression matrices become differential expression profiles; clinical records become systematic cohorts [10–12]. These workflows depend on standardized formats (e.g. FASTQ, BAM/CRAM, VCF for genomics; mzML for proteomics), reference databases, and algorithmic

building blocks such as sequence alignment, multiple sequence analysis, clustering, dimensionality reduction, and predictive modeling [13–15]. The result is a relatively defined path from asynchronous measurements to hypotheses about function, mechanism, and phenotype [16].

Second, molecular modeling and simulation software connects structure with function. Visualization tools enable researchers to inspect macromolecular architecture and binding interfaces; docking programs propose plausible protein–ligand poses and rank them using scoring functions; molecular dynamics simulations probe conformational landscapes and temporal stability [17,18]. In combination with experimental restraints (e.g. cryo-EM density maps and NMR-derived constraints), these software systems support structure-based design, shrink the search space, and accelerate the optimization of lead compounds in drug discovery [9,19].

Third, epidemiological and public health software provides situational awareness and guides population-level interventions [20]. Statistical packages and specialized tools allow analysts to ingest clinical case reports, laboratory confirmations, mobility data, and environmental signals, then fit compartmental or agent-based models, estimate reproduction numbers, forecast incidence, and analyze counterfactual scenarios [9,21]. Geographic information systems and interactive dashboards translate these data into spatially explicit, time-sensitive visualizations for decision-makers—capabilities that proved essential during recent years of pandemics and uncertainty in disease spread [22,23].

The convergence of these domains is transforming translational pipelines. For example, a pathogen genome assembled and annotated by bioinformatics tools can immediately inform molecular modeling software about potential inhibitor candidates, while epidemiological platforms assess how, where, and in whom these interventions would have the greatest impact [24]. Likewise, the integration of multi-omics (genomics, transcriptomics, proteomics, metabolomics) with electronic health records and imaging data requires software that can link heterogeneous sources, ensure traceability, and provide explainable outputs suitable for clinical review [25,26].

Despite this progress, several challenges demand a critical view of the current landscape. Data heterogeneity and missingness complicate integration [27]. Many tools are powerful but difficult for practitioners to learn, creating barriers for smaller laboratories and health services with limited staff [28]. Data analysis itself can be inconsistent—varying datasets, metrics, and preprocessing steps can create an impression of progress without clear real-world benefit [29]. In addition, the sustainability of open-source projects—on which the community heavily relies—depends on funding, governance, and data stewardship, raising concerns about long-term maintenance and the responsible use and sharing of sensitive information [30,31].

Accordingly, this review maps representative software ecosystems across bioinformatics, molecular modeling, and epidemiology/public health, and critically evaluates their strengths, limitations, and translational readiness from measurement to decision.

Priority is given to widely adopted, well documented tools with proven value, while also noting emerging approaches that substantially extend capabilities (e.g. machine learning for structure prediction and graph-based methods for multimodal integration). With an emphasis on reproducibility and usability, the article summarizes domains, tasks, licensing, computational requirements, and validation, in order to align tools with scientific or public health objectives. The goal is to provide a framework that orients newcomers and supports practitioners in planning robust and efficient workflows.

In this review, “translational readiness” is defined as the capacity of software systems to reliably transform primary measurements into interpretable outputs that inform downstream decisions. In bioinformatics, this pathway begins with raw sequencing or omics data and culminates in annotated features, biomarkers, or stratified cohorts. In molecular modeling, structural data and chemical libraries are converted into ranked hypotheses about molecular interactions and stability. In epidemiology and public health, heterogeneous surveillance data are transformed into forecasts, risk maps, and intervention scenarios. Although these processes differ technically, they share common requirements for reproducibility, validation, interoperability, and usability. By examining these

shared requirements across domains, the review highlights where translation succeeds and where friction remains.

## 2. Materials and Methods

### 2.1. Study design

This narrative review synthesizes information on software applications used in three pillars of the biomedical sciences: bioinformatics, molecular modeling and simulations, and epidemiological/public health analytics. This approach was chosen because of the methodological heterogeneity of the current literature (algorithms, benchmarks, case studies), and integrates peer-reviewed evidence from authoritative sources.

A narrative review design was selected because the included software spans heterogeneous methodological paradigms, including algorithm development, simulation frameworks, statistical modeling, and decision-support systems, which are not amenable to uniform outcome measures required for systematic synthesis. The objective was not to quantify pooled effect sizes, but to evaluate translational readiness across domains with respect to reproducibility, validation, interoperability, and usability.

### 2.2. Information sources and search strategy

The search covered the period 2000–2025. The primary databases were PubMed/NCBI, Web of Science, and Scopus. Domain-specific terms were combined with software/algorithm keywords: “bioinformatics”, “molecular modeling”, “docking”, “molecular dynamics”, “epidemiology”, “public health”, as well as phrases such as “multiple sequence alignment software”, “docking program validation”, and “compartmental model software”. The reference lists of included articles were screened manually for additional sources (“snowballing”).

### 2.3. Inclusion criteria

Sources were included if they:

- described, validated, or applied software central to one of the three pillars;
- provided sufficient methodological detail on core algorithms/approaches;
- demonstrated scientific or translational relevance.

Purely conceptual papers without implemented software, tools lacking minimal documentation, and applications without clear significance for the biomedical sciences were excluded. Where multiple versions existed, the most recent release with clear traceability was prioritized.

### 2.4. Study selection

Titles and abstracts were screened for domain relevance. Full texts were then assessed against the inclusion/exclusion criteria. For widely used tools with extensive literature, key methodological papers plus representative benchmark and application studies were selected to avoid overrepresentation while preserving breadth. Selection prioritized tools with sustained adoption, documented validation, and evidence of real-world use in biomedical research or public health practice.

### 2.5. Data extraction

For each tool, the following were extracted: domain and main tasks; typical inputs/outputs and supported data standards; core algorithms/approaches; validation evidence; usability characteristics; interoperability; computational profile; licensing and sustainability; and known limitations. Validation evidence was assessed with attention to benchmark design, dataset representativeness, evaluation metrics, and independent replication, recognizing that reported performance is context-dependent.

## 2.6. Synthesis methods

Evidence was summarized narratively and organized by pillar, with overview tables comparing tools by functionality, validation, usability, and implementation characteristics. Where possible, quantitative performance metrics were reported alongside dataset characteristics and evaluation protocols to contextualize results. End-to-end workflows from raw data to decisions were mapped to illustrate how tools interact in practice.

## 3. Results

### 3.1. Bioinformatics

In the reviewed sources, bioinformatics software supports sequence search/alignment, genome/transcriptome analysis, variant detection and interpretation, and multi-omics statistics. Representative examples of widely used bioinformatics tools, their typical inputs/outputs, and core characteristics are summarized in Table 1.

**Table 1.** Major bioinformatics software systems and pipelines and use cases.

Tool	Main role	Typical inputs → outputs	Key features
<b>BLAST / BLAST+</b>	Local similarity search, support for annotation	Sequence(s) → ranked alignments, statistics	Web interface and command line; supports local databases; integrates into automated pipelines.
<b>Clustal Omega</b>	Multiple sequence alignment at scale	Unaligned sequences → multiple sequence alignment	Designed for large data sets; fast and accurate alignment.
<b>Bioconductor</b>	Omics analyses (RNA-sequencing, methylation, proteomics), statistics	Count/intensity matrices → differential expression, enrichment, reports	Open source; strict packaging, training materials, and reproducible workflows.

Common inputs include FASTQ (raw reads), BAM/CRAM (alignments), VCF (variants), and count/intensity matrices; outputs include annotated genomes, prioritized variants, and differential expression or pathway signatures [32–37]. Tools such as BLAST/BLAST+ remain fundamental for similarity search and functional inference, while local pipelines enable large-scale, automated analyses [38].

Large-scale multiple sequence alignment is routinely handled by Clustal Omega, which emphasizes speed and accuracy for large protein sets, whereas Bioconductor provides a broad, open ecosystem of R packages for omics workflows, statistics, and reproducibility [39,40].

Key strengths include mature conventions for file formats and reference resources; rich ecosystems of packages; and support for containerization/workflows for deployment on HPC or cloud environments—critical for reproducibility and research scalability [41,42].

Limitations include sensitivity to preprocessing choices and reference biases; batch effects when integrating results; and variable usability, including limited language/geographic support. This creates a need for skills in command-line interfaces and workflow management [43].

Collectively, these outputs represent the first translational step, converting high-dimensional molecular measurements into structured features suitable for downstream mechanistic modeling and population-level interpretation.

### 3.2. Molecular Modeling and Simulations

This pillar encompasses molecular visualization, docking/virtual screening, and molecular dynamics simulations. Inputs include 3D structures, force fields, and ligand libraries; outputs include predicted binding poses and/or scores, as well as trajectories and derived observables [44,45].

Key examples include AutoDock Vina, a widely used tool for receptor–ligand docking that improves speed and pose accuracy over earlier AutoDock versions; PyMOL, a user-sponsored and globally adopted visualization system; and GROMACS, a high-performance, actively maintained package for molecular dynamics with extensive user documentation [46].

Strengths include diverse structure-based design workflows that integrate visualization, docking, and molecular dynamics; and acceleration via GPU support and parallelization, which makes demanding simulations more accessible to a broader community, especially when paired with abundant training materials [47].

Limitations include biases in scoring functions, sensitivity to protonation states and other preparation steps, and the computational cost of exhaustive sampling, which remains only partially mitigated by advanced sampling schemes [48]. Table 2 outlines key molecular modeling platforms, detailing for each tool its primary function, the types of data it processes and produces, and the main features that support its widespread use.

**Table 2.** Representative tools for molecular modeling and their roles.

Tool	Main role	Typical inputs → outputs	Key features
<b>AutoDock Vina</b>	Docking / virtual screening	Protein and ligand structures → binding poses and scores	Multithreaded; substantial speed-up with competitive pose accuracy.
<b>PyMOL</b>	Molecular visualization	Structures/trajectories → annotated figures/videos	Widely used in academia and industry; actively maintained documentation and regular releases.
<b>GROMACS</b>	Molecular dynamics	Topologies, coordinates, force fields → trajectories, observables	High performance; up-to-date releases and extensive user manuals.

These tools translate molecular-scale hypotheses into ranked and testable candidates, providing a mechanistic bridge between omics-derived targets and experimental or clinical validation.

### 3.3. Epidemiology and Public Health Analytics

Software in this pillar supports surveillance data management, situational analysis, statistical/cluster modeling, mapping, and monitoring dashboards. Processed data include case counts, laboratory confirmations, mobility/environmental signals, and metadata. Outputs comprise estimates, forecasts, maps, and scenario analyses to support decisions [4,49,50].

Epi Info is a prominent example: it provides public-domain tools for questionnaire design, data entry and analysis, visualization, and mapping, with extensions such as mobile apps for vector surveillance. QGIS is a community-driven geographic platform for spatial analysis and map production, widely used in public health [51,52]. Projects such as Nextstrain demonstrate real-time phylogenetics and visualization for pathogens [53].

Strengths include rapid situational awareness through integrated data capture, analytics, and geographic visualization, as well as the availability of public-domain and open-source platforms that lower barriers for health professionals [54].

Weaknesses include limitations related to data quality/latency; interoperability issues with laboratory information systems and electronic health records; and wide variation in statistical capacity across institutions [55].

Overall, the reviewed software supports reproducible pipelines from raw biological signals to actionable outputs: bioinformatics pipelines that convert sequences/omics into features and

hypotheses; modeling environments that translate structural hypotheses into ranked ligand poses and evidence of dynamic stability; and epidemiological platforms that transform case and contextual data into forecasts and spatially explicit guidance [10,56,57]. Table 3 summarizes representative software platforms used in epidemiology and public health, detailing their main roles, typical data flows from input to output, and key distinguishing features.

**Table 3.** Representative tools for epidemiology/public health and their roles.

Tool	Main role	Typical inputs → outputs	Key features
<b>Epi Info</b>	Capture and analysis of surveillance data	Case/questionnaire data → statistics, graphs, maps	Public domain; widely used for investigating current epidemiological situations.
<b>Epi Info (Vector App)</b>	Mobile vector surveillance	Field observations → datasets ready for analysis	Android application for vector surveillance and analysis.
<b>QGIS</b>	Spatial analysis and cartography	Spatial layers → analytical maps, geoprocessing	Free and open source; extensive plugin ecosystem and comprehensive documentation.

In practice, results from one pillar “lock into” the others—for example, an annotated pathogen genome produced by bioinformatics tools can initiate docking campaigns against key proteins, while public health software quantifies the expected population-level impact and targeting of interventions [58,59]. To make this linkage more explicit, Table 4 summarises the typical software outputs in each domain and illustrates the kinds of downstream research and public health decisions they inform.

Here, software completes the translational pathway by contextualizing biological and clinical signals into population-level risk estimates and decision-support outputs.

**Table 4.** Software outputs across domains and examples of downstream research and public health decisions enabled.

Domain	Main outputs/products	Examples of downstream decisions enabled
<b>Bioinformatics</b>	Annotated genomes and gene sets; differential expression and other multi-omics signatures; variant/biomarker panels with pathway and network annotations.	Prioritisation of candidate targets and biomarkers for functional studies or clinical trials; cohort stratification and patient subtyping; design of diagnostic panels and validation experiments; selection of follow-up omics or imaging assays.
<b>Molecular modeling</b>	Ranked docking poses and scores; molecular-dynamics-derived stability and flexibility profiles; interaction energy decompositions; high-quality 3D structure visualisations.	Triage and optimisation of lead compounds or protein variants; formulation of mechanistic hypotheses about binding and allostery; selection of candidates for synthesis, biophysical testing, and in vivo studies; generation of publication- and submission-ready structural figures.

<b>Epidemiology &amp; public health</b>	Time-series estimates of incidence and reproduction numbers; short- and medium-term forecasts under alternative scenarios; spatial risk maps at different administrative levels; quantitative evaluations of past or ongoing interventions.	Allocation of clinical, laboratory, and public health resources; timing and geographic targeting of interventions (e.g., vaccination, screening, NPIs); setting trigger thresholds for escalation or de-escalation of measures; design and optimisation of surveillance and monitoring systems.
---	---	---

The broad adoption of domain standards, common file formats, and evidence-based workflow solutions enables integration of software components across the full chain from research to translation.

## 4. Discussion

This review highlights how software has become an indispensable substrate of modern biomedical science, enabling large-scale studies, structure-informed design, and timely situational awareness in epidemiology. Across all pillars, standardized file formats, widely adopted toolchains, and growing workflow support enhance interoperability and reproducibility. At the same time, persistent challenges—data heterogeneity, uneven usability, variable validation practices, and sensitive data governance—continue to shape what is practically achievable in research and translation.

### 4.1. Reproducibility and Alignment with Good Data/Software Management Principles

Workflow engines (e.g. CWL/WDL/Nextflow/Snakemake) and containerization (Docker/Singularity) now underpin many production pipelines. These practices reduce the classic “works on my machine” failure mode and render analyses auditable [42]. However, reproducibility still depends on complete reporting: exact versions, reference databases, parameters, and traceability of intermediate files [60]. Common gaps include incomplete documentation of preprocessing steps (e.g. read trimming, normalization, batch effect correction) and of environment details that substantially affect results. Community templates and checklists can raise the baseline without stifling innovation [61,62].

### 4.2. Validation and Benchmarking

Performance claims are most useful when linked to transparent datasets, well-justified metrics, and baselines reflecting current practice [63]. In bioinformatics, reference biases, class imbalance, and data leakage can inflate reported accuracy [59]. In docking and molecular dynamics, scoring functions may favor certain chemotypes, and insufficient conformational sampling limits generalizability [47]. In epidemiology, forecast quality must be interpreted alongside data latency and revision effects [64]. Validation across multiple datasets, blinded or hold-out evaluation, and replication by independent groups are practical safeguards [65].

Across domains, insufficiently transparent benchmarking remains a key translational bottleneck. Performance metrics reported without access to datasets, preprocessing pipelines, or baseline comparisons limit reproducibility and impede independent verification. From a translational perspective, validation quality is often more decisive than marginal algorithmic improvements, particularly when outputs inform experimental prioritization or public health action.

#### 4.3. Artificial Intelligence and Machine Learning

Machine learning is increasingly central—from variant effect prediction and multi-omics integration, to structure prediction and epidemic “nowcasting” [7]. The advantages are clear: enhanced sensitivity/specificity and automation of labor-intensive steps [66]. The risks are equally concrete: opaque decision boundaries, domain shifts, and propagation of biases from historical data. Priority directions include interpretable features or post-hoc explanations suitable for scientific and clinical scrutiny; quantitative uncertainty estimation (calibrated probabilities, interval forecasts); and rigorous external validation across laboratories, instruments, and populations [3,67].

A recurring theme across all three pillars is that limitations in reproducibility, usability, and validation—not algorithmic capability—now represent the dominant barriers to translation. In many contexts, simpler and well-validated tools outperform more complex models when deployed under real-world constraints. Consequently, investment in benchmarking, documentation, and user-centered design is likely to yield greater translational impact than incremental gains in model accuracy alone.

#### 4.4 Usability, Training, and Human Capacity

Powerful tools will not be used to their full potential if interfaces are opaque or documentation is sparse. Teams with mixed skills—bench scientists, analysts, software engineers—benefit from graphical user interfaces for routine tasks, and from command-line interfaces and APIs for advanced use [68]. Interactive notebooks improve transparency but must be combined with environment capture to remain reproducible [60]. Investment in training (short courses, tutorials, example datasets) often yields disproportionate returns, especially for health services and small laboratories [69].

#### 4.5. Integration Across Pillars

Interoperability is no longer “nice to have” but essential: bioinformatics outputs (e.g. pathogen genomes, differential expression signatures) routinely feed molecular modeling campaigns, while public health platforms contextualize candidate interventions through risk maps and forecasts [24,57,58]. Practical friction points include harmonization of identifiers (genes, proteins, compounds), alignment of coordinate systems and force fields, and linking tabular analyses with spatial layers [70]. “Data contracts”—explicit expectations about schema, units, and metadata—help teams move faster with fewer surprises throughout the lifecycle [60].

#### 4.6. Data Governance, Privacy, and Security

Biomedical and surveillance data are inherently sensitive. Encryption at rest and in transit, role-based access control, auditable logs, and data minimization are baseline requirements [71]. When using cloud resources, organizations must plan for jurisdictional constraints and vendor lock-in by adopting portable workflows and clear exit strategies [72].

#### 4.7. Limitations

As a review spanning diverse sub-disciplines, this article prioritizes widely used, well documented tools and workflows rather than exhaustively cataloguing every niche package. Heterogeneous evaluation practices across domains complicate direct “metric-to-metric” comparisons; where possible, the emphasis is on patterns and decision criteria rather than absolute performance values.

### 5. Conclusions

Software now constitutes core translational infrastructure in biomedicine, enabling the conversion of heterogeneous measurements into decisions at molecular, clinical, and population

scales. This review demonstrates that, despite differences in data types and algorithms, bioinformatics, molecular modeling, and epidemiological software share common dependencies on standards, reproducibility practices, and transparent validation.

Near-term gains are most likely to come from disciplined engineering and stewardship: transparent benchmarking, clear “data contracts” and traceability, user-centered interfaces built atop robust command-line/API foundations, and secure, portable deployments. Machine learning enhances sensitivity, specificity, and speed, and is poised to remain a key enabler across multiple domains.

Future progress depends on integrated, multimodal pipelines (omics, imaging, electronic health records), standardized evaluation with open datasets, sustained investment in training and equitable access to compute, durable open-source ecosystems, and strong safeguards for security, privacy, and ethical data use. Under these conditions, software will continue to shorten the distance between measurement and decision, offering the clearest and most compelling value proposition for biomedical science and public health.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The author declares no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

3D	three-dimensional
API	application programming interface
BAM	Binary Alignment/Map (alignment file format)
BLAST	Basic Local Alignment Search Tool
CRAM	Compressed Reference-oriented Alignment Map
CWL	Common Workflow Language
Epi Info	epidemiological information software package for public health surveillance
FASTQ	text format for nucleotide sequences with quality scores
GPU	graphics processing unit
GROMACS	GROningen MACHine for Chemical Simulations (molecular dynamics package)
HPC	high-performance computing
mzML	mass spectrometry markup language
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NPI	non-pharmaceutical intervention
PubMed	public interface to the MEDLINE biomedical literature database
QGIS	Quantum Geographic Information System (open-source GIS platform)
RNA	ribonucleic acid
VCF	Variant Call Format
WDL	Workflow Description Language

## References

1. Roukos, V.; Misteli, T.; Schmidt, C.K. Descriptive No More: The Dawn of High-Throughput Microscopy. *Trends Cell Biol* **2010**, *20*, 503–506, doi:10.1016/j.tcb.2010.06.008.
2. Buggenthin, F.; Marr, C.; Schwarzfischer, M.; Hoppe, P.S.; Hilsenbeck, O.; Schroeder, T.; Theis, F.J. An Automatic Method for Robust and Fast Cell Detection in Bright Field Images from High-Throughput Microscopy. *BMC Bioinformatics* **2013**, *14*, 297, doi:10.1186/1471-2105-14-297.

3. Pegoraro, G.; Misteli, T. High-Throughput Imaging for the Discovery of Cellular Mechanisms of Disease. *Trends Genet* **2017**, *33*, 604–615, doi:10.1016/j.tig.2017.06.005.
4. Berkhout, M.; Smit, K.; Versendaal, J. Decision Discovery Using Clinical Decision Support System Decision Log Data for Supporting the Nurse Decision-Making Process. *BMC Med Inform Decis Mak* **2024**, *24*, 100, doi:10.1186/s12911-024-02486-3.
5. Pastorino, R.; De Vito, C.; Migliara, G.; Glocker, K.; Binenbaum, I.; Ricciardi, W.; Boccia, S. Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives. *Eur J Public Health* **2019**, *29*, 23–27, doi:10.1093/eurpub/ckz168.
6. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of Machine Learning in Drug Discovery and Development. *Nat Rev Drug Discov* **2019**, *18*, 463–477, doi:10.1038/s41573-019-0024-5.
7. Zitnik, M.; Nguyen, F.; Wang, B.; Leskovec, J.; Goldenberg, A.; Hoffman, M.M. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf Fusion* **2019**, *50*, 71–91, doi:10.1016/j.inffus.2018.09.012.
8. Sevimoglu, T.; Arga, K.Y. The Role of Protein Interaction Networks in Systems Biomedicine. *Comput Struct Biotechnol J* **2014**, *11*, 22–27, doi:10.1016/j.csbj.2014.08.008.
9. Dwivedi, S.; Purohit, P.; Misra, R.; Pareek, P.; Goel, A.; Khattri, S.; Pant, K.K.; Misra, S.; Sharma, P. Diseases and Molecular Diagnostics: A Step Closer to Precision Medicine. *Indian J Clin Biochem* **2017**, *32*, 374–398, doi:10.1007/s12291-017-0688-8.
10. Clark, A.J.; Lillard, J.W. A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology. *Genes (Basel)* **2024**, *15*, 1036, doi:10.3390/genes15081036.
11. Jamialahmadi, H.; Khalili-Tanha, G.; Nazari, E.; Rezaei-Tavirani, M. Artificial Intelligence and Bioinformatics: A Journey from Traditional Techniques to Smart Approaches. *Gastroenterol Hepatol Bed Bench* **2024**, *17*, 241–252, doi:10.22037/ghfbb.v17i3.2977.
12. Rosati, D.; Palmieri, M.; Brunelli, G.; Morrione, A.; Iannelli, F.; Frullanti, E.; Giordano, A. Differential Gene Expression Analysis Pipelines and Bioinformatic Tools for the Identification of Specific Biomarkers: A Review. *Comput Struct Biotechnol J* **2024**, *23*, 1154–1168, doi:10.1016/j.csbj.2024.02.018.
13. Larson, N.B.; Oberg, A.L.; Adjei, A.A.; Wang, L. A Clinician’s Guide to Bioinformatics for next-Generation Sequencing. *J Thorac Oncol* **2023**, *18*, 143–157, doi:10.1016/j.jtho.2022.11.006.
14. Lubin, I.M.; Aziz, N.; Babb, L.J.; Ballinger, D.; Bisht, H.; Church, D.M.; Cordes, S.; Eilbeck, K.; Hyland, F.; Kalman, L.; et al. Principles and Recommendations for Standardizing the Use of the Next-Generation Sequencing Variant File in Clinical Settings. *J Mol Diagn* **2017**, *19*, 417–426, doi:10.1016/j.jmoldx.2016.12.001.
15. Ismail, F.N.; Amarasoma, S. Mars: Simplifying Bioinformatics Workflows through a Containerized Approach to Tool Integration and Management. *Bioinform Adv* **2025**, *5*, vbaf074, doi:10.1093/bioadv/vbaf074.
16. Luo, Y.; Zhao, C.; Chen, F. Multiomics Research: Principles and Challenges in Integrated Analysis. *Biores Res* **2024**, *6*, 0059, doi:10.34133/bdr.0059.
17. Ferreira, L.G.; dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* **2015**, *20*, 13384–13421, doi:10.3390/molecules200713384.
18. O’Donoghue, S.I.; Goodsell, D.S.; Frangakis, A.S.; Jossinet, F.; Laskowski, R.A.; Nilges, M.; Saibil, H.R.; Schafferhans, A.; Wade, R.C.; Westhof, E.; et al. Visualization of Macromolecular Structures. *Nature Methods* **2010**, *7*, S42, doi:10.1038/nmeth.1427.
19. Leelananda, S.P.; Lindert, S. Using NMR Chemical Shifts and Cryo-EM Density Restraints in Iterative Rosetta-MD Protein Structure Refinement. *J Chem Inf Model* **2020**, *60*, 2522–2532, doi:10.1021/acs.jcim.9b00932.
20. Deshpande, A.; Margevicius, K.; Generous, E.; Taylor-McCabe, K.; Castro, L.; Longo, J.; Priedhorsky, R. Tools and Apps to Enhance Situational Awareness for Global Disease Surveillance. *Online J Public Health Inform* **2014**, *6*, e111, doi:10.5210/ojphi.v6i1.5183.
21. Deo, V.; Ranganathan, P. Statistical Tools and Packages for Data Collection, Management, and Analysis - A Brief Guide for Health and Biomedical Researchers. *Perspect Clin Res* **2024**, *15*, 209–212, doi:10.4103/picr.picr\_160\_24.

22. Shaw, N.T. Geographical Information Systems and Health: Current State and Future Directions. *Healthc Inform Res* **2012**, *18*, 88–96, doi:10.4258/hir.2012.18.2.88.
23. Opiyo, S.O.; Nalunkuma, R.; Nanyonga, S.M.; Mugenyi, N.; Kanyike, A.M. Empowering Global AMR Research Community: Interactive GIS Dashboards for AMR Data Analysis and Informed Decision-Making. *Wellcome Open Res* **2024**, *9*, 234, doi:10.12688/wellcomeopenres.21010.4.
24. Sintchenko, V.; Roper, M.P.V. Pathogen Genome Bioinformatics. *Methods Mol Biol* **2014**, *1168*, 173–193, doi:10.1007/978-1-4939-0847-9\_10.
25. Sathasivam, S.; Rajendran, K.; Logeswaran, K.; Periasamy, Dr.K.; Sharmila, V.; Sangeetha, M. Integration of Multi-Omics Data: Genomics, Proteomics, Metabolomics. In: 2025; pp. 149–184 ISBN 979-8-3693-9521-9.
26. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-Omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights* **2020**, *14*, 1177932219899051, doi:10.1177/1177932219899051.
27. Putrama, I.M.; Martinek, P. Heterogeneous Data Integration: Challenges and Opportunities. *Data Brief* **2024**, *56*, 110853, doi:10.1016/j.dib.2024.110853.
28. Bessen, J.L.; Alexander, M.; Foroughi, O.; Brathwaite, R.; Baser, E.; Lee, L.C.; Perez, O.; Gustavsen, G. Perspectives on Reducing Barriers to the Adoption of Digital and Computational Pathology Technology by Clinical Labs. *Diagnostics (Basel)* **2025**, *15*, 794, doi:10.3390/diagnostics15070794.
29. Weber, L.M.; Saelens, W.; Cannoodt, R.; Sonesson, C.; Hapfelmeier, A.; Gardner, P.P.; Boulesteix, A.-L.; Saeys, Y.; Robinson, M.D. Essential Guidelines for Computational Method Benchmarking. *Genome Biol* **2019**, *20*, 125, doi:10.1186/s13059-019-1738-8.
30. Ye, Y.; Barapatre, S.; Davis, M.K.; Elliston, K.O.; Davatzikos, C.; Fedorov, A.; Fillion-Robin, J.-C.; Foster, I.; Gilbertson, J.R.; Lasso, A.; et al. Open-Source Software Sustainability Models: Initial White Paper From the Informatics Technology for Cancer Research Sustainability and Industry Partnership Working Group. *J Med Internet Res* **2021**, *23*, e20028, doi:10.2196/20028.
31. Scheirer, M.A.; Dearing, J.W. An Agenda for Research on the Sustainability of Public Health Programs. *Am J Public Health* **2011**, *101*, 2059–2067, doi:10.2105/AJPH.2011.300193.
32. Cotto, K.C.; Feng, Y.-Y.; Ramu, A.; Richters, M.; Freshour, S.L.; Skidmore, Z.L.; Xia, H.; McMichael, J.F.; Kunisaki, J.; Campbell, K.M.; et al. Integrated Analysis of Genomic and Transcriptomic Data for the Discovery of Splice-Associated Variants in Cancer. *Nat Commun* **2023**, *14*, 1589, doi:10.1038/s41467-023-37266-6.
33. Bravo, A.M.; Typas, A.; Veening, J.-W. 2FAST2Q: A General-Purpose Sequence Search and Counting Program for FASTQ Files. *PeerJ* **2022**, *10*, e14041, doi:10.7717/peerj.14041.
34. Sadedin, S.P.; Oshlack, A. Bazam: A Rapid Method for Read Extraction and Realignment of High-Throughput Sequencing Data. *Genome Biol* **2019**, *20*, 78, doi:10.1186/s13059-019-1688-1.
35. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158, doi:10.1093/bioinformatics/btr330.
36. Althagafi, A.; Zhapa-Camacho, F.; Hoehndorf, R. Prioritizing Genomic Variants through Neuro-Symbolic Knowledge-Enhanced Learning. *Bioinformatics* **2024**, *40*, btae301, doi:10.1093/bioinformatics/btae301.
37. Reimand, J.; Isser, R.; Voisin, V.; Kucera, M.; Tannus-Lopes, C.; Rostamianfar, A.; Wadi, L.; Meyer, M.; Wong, J.; Xu, C.; et al. Pathway Enrichment Analysis and Visualization of Omics Data Using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **2019**, *14*, 482–517, doi:10.1038/s41596-018-0103-9.
38. Schmid, S.; Jeevannavar, A.; Julian, T.R.; Tamminen, M. Portable BLAST-like Algorithm Library and Its Implementations for Command Line, Python, and R. *PLoS One* **2023**, *18*, e0289693, doi:10.1371/journal.pone.0289693.
39. Sievers, F.; Higgins, D.G. Clustal Omega for Making Accurate Alignments of Many Protein Sequences. *Protein Sci* **2018**, *27*, 135–145, doi:10.1002/pro.3290.
40. Lloyd, G.R.; Jankevics, A.; Weber, R.J.M. Struct: An R/Bioconductor-Based Framework for Standardized Metabolomics Data Analysis and Beyond. *Bioinformatics* **2020**, *36*, 5551–5552, doi:10.1093/bioinformatics/btaa1031.

41. Le Piane, F.; Vozza, M.; Baldoni, M.; Mercuri, F. Integrating High-Performance Computing, Machine Learning, Data Management Workflows, and Infrastructures for Multiscale Simulations and Nanomaterials Technologies. *Beilstein J Nanotechnol* **2024**, *15*, 1498–1521, doi:10.3762/bjnano.15.119.
42. Ahmed, A.E.; Allen, J.M.; Bhat, T.; Burra, P.; Fliege, C.E.; Hart, S.N.; Heldenbrand, J.R.; Hudson, M.E.; Istanto, D.D.; Kalmbach, M.T.; et al. Design Considerations for Workflow Management Systems Use in Production Genomics Research and the Clinic. *Sci Rep* **2021**, *11*, 21680, doi:10.1038/s41598-021-99288-8.
43. Yu, Y.; Mai, Y.; Zheng, Y.; Shi, L. Assessing and Mitigating Batch Effects in Large-Scale Omics Studies. *Genome Biol* **2024**, *25*, 254, doi:10.1186/s13059-024-03401-9.
44. Terefe, E.M.; Ghosh, A. Molecular Docking, Validation, Dynamics Simulations, and Pharmacokinetic Prediction of Phytochemicals Isolated From Croton Dichogamus Against the HIV-1 Reverse Transcriptase. *Bioinform Biol Insights* **2022**, *16*, 11779322221125605, doi:10.1177/11779322221125605.
45. Varghese, A.; Liu, J.; Patterson, T.A.; Hong, H. Integrating Molecular Dynamics, Molecular Docking, and Machine Learning for Predicting SARS-CoV-2 Papain-like Protease Binders. *Molecules* **2025**, *30*, 2985, doi:10.3390/molecules30142985.
46. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J Comput Chem* **2010**, *31*, 455–461, doi:10.1002/jcc.21334.
47. Blanes-Mira, C.; Fernández-Aguado, P.; de Andrés-López, J.; Fernández-Carvajal, A.; Ferrer-Montiel, A.; Fernández-Ballester, G. Comprehensive Survey of Consensus Docking for High-Throughput Virtual Screening. *Molecules* **2022**, *28*, 175, doi:10.3390/molecules28010175.
48. Kamenik, A.S.; Linker, S.M.; Riniker, S. Enhanced Sampling without Borders: On Global Biasing Functions and How to Reweight Them. *Phys Chem Chem Phys* **24**, 1225–1236, doi:10.1039/d1cp04809k.
49. Sutton, R.T.; Pincock, D.; Baumgart, D.C.; Sadowski, D.C.; Fedorak, R.N.; Kroeker, K.I. An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success. *NPJ Digit Med* **2020**, *3*, 17, doi:10.1038/s41746-020-0221-y.
50. Ogwu, M.; Izah, S. Innovations in Disease Surveillance and Monitoring. In; 2025; pp. 83–108 ISBN 978-3-031-82621-4.
51. Su, Y.; Yoon, S.S. Epi Info – Present and Future. *AMIA Annu Symp Proc* **2003**, *2003*, 1023.
52. Graser, A.; Sutton, T.; Bernasocchi, M. The QGIS Project: Spatial without Compromise. *Patterns (N Y)* **2025**, *6*, 101265, doi:10.1016/j.patter.2025.101265.
53. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* **2018**, *34*, 4121–4123, doi:10.1093/bioinformatics/bty407.
54. Kolivand, P.; Azari, S.; Bakhtiari, A.; Namdar, P.; Ayyoubzadeh, S.M.; Rajaie, S.; Ramezani, M. AI Applications in Disaster Governance with Health Approach: A Scoping Review. *Arch Public Health* **2025**, *83*, 218, doi:10.1186/s13690-025-01712-2.
55. Syed, R.; Eden, R.; Makasi, T.; Chukwudi, I.; Mamudu, A.; Kamalpour, M.; Kapugama Geeganage, D.; Sadeghianasl, S.; Leemans, S.J.J.; Goel, K.; et al. Digital Health Data Quality Issues: Systematic Review. *J Med Internet Res* **2023**, *25*, e42615, doi:10.2196/42615.
56. Vitorino, R. Transforming Clinical Research: The Power of High-Throughput Omics Integration. *Proteomes* **2024**, *12*, 25, doi:10.3390/proteomes12030025.
57. Wratten, L.; Wilm, A.; Göke, J. Reproducible, Scalable, and Shareable Analysis Pipelines with Bioinformatics Workflow Managers. *Nat Methods* **2021**, *18*, 1161–1168, doi:10.1038/s41592-021-01254-9.
58. D’Auria, G.; Schneider, M.V.; Moya, A. Live Genomics for Pathogen Monitoring in Public Health. *Pathogens* **2014**, *3*, 93–108, doi:10.3390/pathogens3010093.
59. Zhang, S.; Liu, K.; Liu, Y.; Hu, X.; Gu, X. The Role and Application of Bioinformatics Techniques and Tools in Drug Discovery. *Front Pharmacol* **2025**, *16*, 1547131, doi:10.3389/fphar.2025.1547131.
60. Leipzig, J.; Nüst, D.; Hoyt, C.T.; Ram, K.; Greenberg, J. The Role of Metadata in Reproducible Computational Research. *Patterns (N Y)* **2021**, *2*, 100322, doi:10.1016/j.patter.2021.100322.

61. Čuklina, J.; Lee, C.H.; Williams, E.G.; Sajic, T.; Collins, B.C.; Rodríguez Martínez, M.; Sharma, V.S.; Wendt, F.; Goetze, S.; Keele, G.R.; et al. Diagnostics and Correction of Batch Effects in Large-scale Proteomic Studies: A Tutorial. *Mol Syst Biol* **2021**, *17*, e10240, doi:10.15252/msb.202110240.
62. Huang, L.; Kim, Y.; Balluff, B.; Cillero-Pastor, B. Quality Control Standards for Batch Effect Evaluation and Correction in Mass Spectrometry Imaging. *Anal Chem* **97**, 10919–10928, doi:10.1021/acs.analchem.5c02020.
63. Welzel, C.; Brückner, S.; Brightwell, C.; Fenech, M.; Gilbert, S. A Transparent and Standardized Performance Measurement Platform Is Needed for On-Prescription Digital Health Apps to Enable Ongoing Performance Monitoring. *PLOS Digit Health* **2024**, *3*, e0000656, doi:10.1371/journal.pdig.0000656.
64. Tomov, L.; Chervenkov, L.; Miteva, D.G.; Batselova, H.; Velikova, T. Applications of Time Series Analysis in Epidemiology: Literature Review and Our Experience during COVID-19 Pandemic. *World J Clin Cases* **2023**, *11*, 6974–6983, doi:10.12998/wjcc.v11.i29.6974.
65. Suetake, H.; Fukusato, T.; Igarashi, T.; Ohta, T. Workflow Sharing with Automated Metadata Validation and Test Execution to Improve the Reusability of Published Workflows. *Gigascience* **2023**, *12*, giad006, doi:10.1093/gigascience/giad006.
66. Josten, C.; Lordan, G. Automation and the Changing Nature of Work. *PLoS One* **2022**, *17*, e0266326, doi:10.1371/journal.pone.0266326.
67. Mehdiyev, N.; Majlatow, M.; Fettke, P. Interpretable and Explainable Machine Learning Methods for Predictive Process Monitoring: A Systematic Literature Review. *Artif Intell Rev* **2025**, *58*, 378, doi:10.1007/s10462-025-11399-0.
68. Faulkner, M.; Wells, M. A Conversation with Research Software Engineers at the International Brain Laboratory. *Patterns (N Y)* **2025**, *6*, 101315, doi:10.1016/j.patter.2025.101315.
69. Arora, A.; Alderman, J.E.; Palmer, J.; Ganapathi, S.; Laws, E.; McCradden, M.D.; Oakden-Rayner, L.; Pfohl, S.R.; Ghassemi, M.; McKay, F.; et al. The Value of Standards for Health Datasets in Artificial Intelligence-Based Applications. *Nat Med* **2023**, *29*, 2929–2938, doi:10.1038/s41591-023-02608-w.
70. Diamant, I.; Clarke, D.J.B.; Evangelista, J.E.; Lingam, N.; Ma'ayan, A. Harmonizome 3.0: Integrated Knowledge about Genes and Proteins from Diverse Multi-Omics Resources. *Nucleic Acids Res* **2024**, *53*, D1016–D1028, doi:10.1093/nar/gkae1080.
71. Malin, B.A.; Emam, K.E.; O'Keefe, C.M. Biomedical Data Privacy: Problems, Perspectives, and Recent Advances. *J Am Med Inform Assoc* **2013**, *20*, 2–6, doi:10.1136/amiajnl-2012-001509.
72. Salih, S.; Hamdan, M.; Abdelmaboud, A.; Abdelaziz, A.; Abdelsalam, S.; Althobaiti, M.M.; Cheikhrouhou, O.; Hamam, H.; Alotaibi, F. Prioritising Organisational Factors Impacting Cloud ERP Adoption and the Critical Issues Related to Security, Usability, and Vendors: A Systematic Literature Review. *Sensors (Basel)* **2021**, *21*, 8391, doi:10.3390/s21248391.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.