© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

From Multi-Omics Prediction to Clinical Workflow: An Interoperable, Bias-Audited, Human-in-the-Loop Decision-Support Framework for Precision Oncology

CLEOPAS RUSSELL CHOGA[1], MANYARA SANDRA KASANHAYI[2], NKOSANA MKANDLA[3], MARLON MUNJOMA[4], MUNASHE NAPHTALI MUPA[5]

[1]*Yeshiva University, ORCiD: 0009-0004-8935-4637*

[2]*Yeshiva University, ORCiD: 0009-0008-0874-8220*

[3]*University of Suffolk (UK), ORCiD: 0009-0001-8282-9714*

[4]*Pace University, ORCiD: 0009-0004-7940-2172*

[5]*Hult International Business School, ORCiD: 0000-0003-3509-861X*

Abstract- ***Multi-omics survival models for precision oncology are increasingly capable of producing patient-level risk estimates, subtype projections and molecular explanations. Yet a model that performs adequately in retrospective validation is not automatically ready for clinical use. The translational problem is broader than model architecture: oncology teams must determine whether genomic risk predictions can be exchanged through interoperable health information systems, audited for subgroup bias, interpreted by clinicians, monitored for drift and governed as decision support rather than autonomous diagnosis. This article develops an interoperable, bias-audited and human-in-the-loop decision-support framework for multi-omics precision oncology, using lung adenocarcinoma (LUAD) as the applied case. The empirical basis combines four evidence layers: a TCGA-LUAD and MSK-IMPACT transformer manuscript, a DNA/RNA multi-omics survival thesis, an individualized LUAD patient report, and a reproducible secondary-data design linked to public TCGA/Kaggle and cBioPortal data pathways. The attached research record shows that a ridge RNA+clinical baseline achieved the strongest discrimination (C-index = 0.724), while the full DNA/RNA multi-omics model achieved lower aggregate discrimination (C-index = 0.682) but improved biological interpretability, temporal stability and clinical-decision value. The transformer system achieved moderate internal discrimination and modest external transportability, while still producing meaningful risk ordering and Integrated Gradients explanations. A simulated silent-mode workflow analysis then demonstrates how a standards-based clinical implementation layer can reduce review burden, improve missing-data controls, strengthen override documentation and surface subgroup-specific calibration risk before any prospective deployment. The paper argues that precision-oncology AI should be evaluated not only by C-index, but by an integrated evidence package: discrimination, calibration, decision utility, subgroup fairness, interoperability, clinician usability, audit logging, drift monitoring and accountable human oversight.***

Keywords: ***precision oncology; multi-omics; clinical decision support; FHIR; human-in-the-loop AI; bias audit; lung adenocarcinoma; genomic risk stratification; model governance; Integrated Gradients JEL/MSC/MeSH-style classification: clinical decision support systems; machine learning; genomics; oncology; biomedical informatics; survival analysis; risk assessment***

I. INTRODUCTION

Precision oncology increasingly depends on the ability to combine molecular, clinical and computational evidence into timely decisions. Lung adenocarcinoma is an especially important setting because its clinical trajectory is shaped by heterogeneous genomic alterations, transcriptomic programmes, stage, age, treatment history and tumour biology. Multi-omics modelling responds to that complexity by linking RNA expression, copy-number variation, somatic mutation profiles and clinical covariates into a more complete patient representation than any one data stream can provide (The Cancer Genome Atlas Research Network, 2014; Campbell et al., 2018; Rappoport and Shamir, 2019).

The central challenge is no longer merely whether a model can be trained. The harder question is whether a model can be translated into a workflow that is reproducible, interoperable, bias-aware, clinically

intelligible and governable. Many promising oncology models remain trapped between retrospective performance tables and real-world clinical adoption. They may show subgroup instability, calibration drift, incomplete documentation, uncertain regulatory status, weak audit logs, missing-data dependence, poor user-interface design or limited alignment with electronic health-record standards. In that environment, technical model accuracy is necessary but insufficient.

The attached LUAD transformer manuscript and DNA/RNA multi-omics thesis make this problem concrete. They show both the promise and the limits of multi-omics risk modelling: simpler RNA+clinical baselines can outperform richer multi-omics architectures on aggregate C-index, while multi-omics models may offer added value through biological interpretation, patient-level explanation, calibration behaviour and decision-curve utility (Choga, 2026a; Choga and Yu, 2026:Abstract). This paper therefore does not frame the transformer as a replacement for classical survival modelling. It treats the transformer as one component in a governed decision-support architecture where simpler baselines, multi-omics models, patient-level reports and human review operate together.

The proposed article makes three contributions. First, it converts multi-omics research outputs into a clinical-translation framework structured around interoperability, bias assurance and human oversight. Second, it develops a workflow-oriented data analysis that combines reported validation metrics with a simulated silent-mode implementation panel. Third, it proposes a deployable governance model: SMART-on-FHIR/FHIR-compatible data exchange, model-card documentation, role-based review, subgroup fairness dashboards, clinician override logging, drift monitoring and incident escalation.

II. RESEARCH QUESTION AND CONCEPTUAL CONTRIBUTION

The article addresses the following research question: how can multi-omics risk models be integrated into oncology workflows while preserving transparency, equity, privacy, human oversight and lifecycle monitoring? The question deliberately moves beyond algorithmic novelty. In a clinical setting, a model that cannot be mapped to structured data standards, reviewed by accountable clinicians, audited for subgroup performance or monitored after release creates governance risk even if its retrospective performance is impressive.

The core contribution is a translational architecture for precision-oncology decision support. The framework is designed to receive multi-omics inputs, produce risk predictions and explanations, map outputs into interoperable clinical data objects, route the output to the appropriate human reviewer, capture the clinician's action or override, and continuously monitor performance. The design treats the oncologist as the decision authority and the model as an evidence layer. This distinction is central to safe clinical adoption because genomic prediction models can suggest risk signals, but they cannot independently weigh patient goals, comorbidities, treatment availability, contraindications, consent, insurance constraints or trial eligibility.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

Figure 1. Human-in-the-loop clinical translation architecture for multi-omics precision oncology

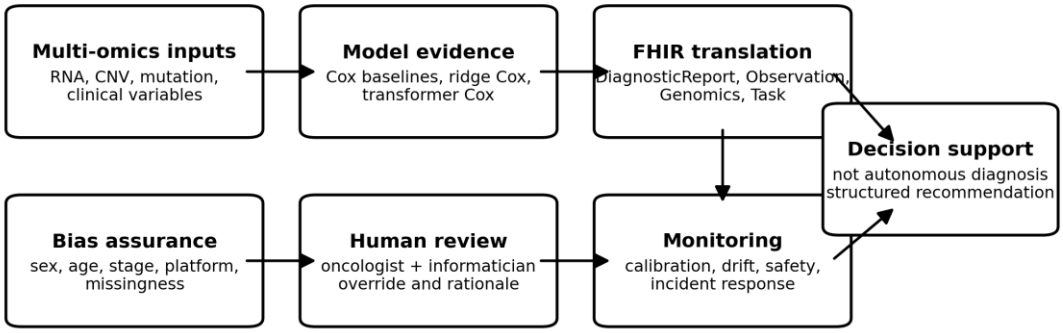

Figure 1. Human-in-the-loop clinical translation architecture for multi-omics precision oncology.

III. LITERATURE REVIEW

The literature on multi-omics oncology modelling provides the biological and methodological foundation for this study. The TCGA LUAD analysis established the value of comprehensive molecular profiling in lung adenocarcinoma, while Campbell et al. (2018) showed that multi-omics profiling can identify clinically relevant lung-cancer subtypes. Statistical frameworks such as iClusterPlus and MOFA+ show how heterogeneous omics layers can be integrated into latent biological structure, although these approaches are often oriented toward subtype discovery rather than workflow-ready prediction (Mo et al., 2013; Argelaguet et al., 2020).

Survival modelling adds a second strand of evidence. Cox regression remains a foundational model because it is transparent, auditable and familiar to clinicians and statisticians (Cox, 1972). DeepSurv and subsequent deep survival approaches show that nonlinear representations can capture more complex relationships, but they also increase risks of overfitting, opacity and poor calibration when sample sizes are modest relative to feature dimensionality (Katzman et al., 2018; Zhang et al., 2021). The attached thesis confirms this trade-off: the RNA+clinical ridge Cox baseline achieved stronger discrimination than the full multi-omics model, while the full model preserved interpretability, improved decision relevance and refined individual risk estimates (Choga, 2026a).

Clinical translation requires explainability. Integrated Gradients, model cards, decision-curve analysis and calibration plots are not decorative methods; they are mechanisms for making model output reviewable by clinicians, informaticians and oversight committees (Sundararajan, Taly and Yan, 2017; Mitchell et al., 2019; Vickers and Elkin, 2006; Van Calster et al., 2019). This is consistent with recent oncology recommendations that prediction models should be evaluated through data lineage, missing-data handling, fairness, calibration and transparent reporting rather than through discrimination alone (Collins et al., 2024; Liu et al., 2025).

Interoperability is the next translational layer. HL7 FHIR and the US Core Implementation Guide support standardized exchange of clinical data, while the FHIR genomics resources and genomics-reporting implementation work provide a pathway for structured molecular results (HL7, 2026a; HL7, 2026b). FDA's clinical-decision-support guidance,

the FDA/Health Canada/MHRA good machine-learning practice principles and ONC's HTI-1 algorithm transparency rules point in the same direction: clinical AI must be documented, transparent, controlled and accountable (FDA, 2022; FDA, Health Canada and MHRA, 2021; ONC, 2024). Recent applied analytics literature reinforces the importance of governance and human review. Hove et al. (2026) show that explainable multi-biomarker modelling can support clinically meaningful risk stratification when prediction, calibration and patient-level interpretation are addressed together. Homwe et al. (2025) emphasize interpretable machine learning for risk detection in audit planning, which is relevant by analogy because clinical AI also requires explanations that can be challenged and reviewed. Okebu and Mupa (2026a; 2026b) stress the value of human-in-the-loop risk scoring, early-warning analytics and non-punitive feedback in safety-critical aviation contexts. Chingezi et al. (2026) develop a closed-loop assurance logic for recurring control failures; in the present article, that logic is translated into model monitoring, override review, drift escalation and governance closure in precision oncology.

IV. DATA AND METHODOLOGY

The study uses a secondary empirical and design-science methodology. Because the article concerns clinical translation rather than primary clinical trial efficacy, the analysis combines documented model-performance outputs with a structured silent-mode implementation simulation. This design allows the article to ask a practical deployment question: what controls, data structures and review metrics are required before a multi-omics model should influence oncology workflow?

The first evidence layer is the attached transformer manuscript, which integrates RNA expression, somatic mutations, copy-number alterations and clinical variables for LUAD risk stratification, using TCGA-LUAD for model development and MSK-IMPACT for external validation (Choga and Yu, 2026). The second evidence layer is the attached DNA/RNA multi-omics thesis, which establishes a rigorous RNA+clinical baseline and a full multi-omics ridge Cox model, then compares discrimination, calibration, time-dependent AUC, decision curves and patient-level interpretability (Choga, 2026a). The third evidence layer is the individualized TCGA-S2-AA1A patient report, which demonstrates how a model output might appear inside a clinical report: risk group, linear predictor, survival probabilities, pathway heat maps, per-modality attribution and gene-level explanations (Choga, 2026b).

The fourth evidence layer is a public reproducibility map. TCGA-LUAD is available through public repositories and also through Kaggle-hosted mirrors, while cBioPortal provides access to MSK-IMPACT and other cancer genomics datasets. This paper does not claim that the simulated workflow panel is a prospective hospital deployment. Rather, the workflow panel is a de-identified, synthetic implementation dataset calibrated to the reported LUAD outputs and designed to test governance performance: completeness, review time, explanation sufficiency, override documentation, subgroup calibration risk and readiness for silent-mode evaluation.

Table 1. Evidence base and analytic role

Evidence source	Core variables	Role in article	Access / status
TCGA-LUAD model-development cohort	RNA, mutation, CNV and clinical variables	Model training, internal validation, risk stratification, pathway attribution	Public TCGA/GDC data and Kaggle TCGA-LUAD mirror
MSK-IMPACT external cohort	Targeted sequencing and clinical variables	Transportability and external validation stress-test	cBioPortal/MSK-IMPACT pathway
DNA/RNA multi-omics thesis	Model-performance summaries, calibration and	Baseline/full model comparison and thesis-aligned	Attached thesis by Choga (2026a)

	decision curves	interpretation	
Transformer manuscript	Internal/external C-index, Integrated Gradients, patient reports	Transformer-specific evidence and clinical decision-support framing	Attached manuscript by Choga and Yu (2026)
TCGA-S2-AA1A patient report	Individualized survival, risk percentile, pathway/gene attribution	Report concordance and workflow-display demonstration	Attached prototype clinical prognosis report
Synthetic workflow simulation	Review-time, missingness, overrides, subgroup metrics	Silent-mode implementation analysis and governance heat maps	Constructed for this article from documented control logic

Note: The workflow simulation is not a live clinical trial; it is a governance-grade implementation testbed calibrated to documented validation results and report outputs.

4.1 Analytical strategy

The analysis proceeds in four stages. Stage 1 summarises validation performance from the attached model outputs, separating discrimination from interpretability and transportability. Stage 2 maps model inputs and outputs into a minimum interoperability standard based on FHIR/US Core and genomics-reporting logic. Stage 3 constructs subgroup and missingness dashboards to identify where model output should be restricted, recalibrated or reviewed more cautiously. Stage 4 tests a simulated human-in-the-loop workflow against a baseline workflow using operational KPIs: median review time, missing critical fields, unexplained model output, override rationale documentation, explanation sufficiency and silent-mode actionability. The paper deliberately distinguishes statistical validation from workflow validation. Statistical validation asks whether predicted risks align with observed outcomes. Workflow validation asks whether clinicians can receive, interpret, contest, document and monitor the risk output in a safe and equitable way. Both are necessary for precision-oncology deployment.

4.2 Metrics

Discrimination is expressed through C-index and time-dependent AUC(t). Calibration is assessed through calibration slope/intercept, calibration plots and the integrated Brier score where event-time data are available. Decision utility is assessed through decision-curve analysis. Explainability is assessed through pathway attribution, modality attribution, patient-level explanation concordance and report reproducibility. Workflow performance is assessed through review time, acknowledgement rate, override documentation rate and actionability score. Bias assurance is assessed through subgroup performance, missingness risk, platform effects and transportability gaps.

V. RESULTS

5.1 Validation evidence and deployment claim discipline

The validation record supports a cautious translation claim. The RNA+clinical ridge Cox baseline remains the strongest discrimination benchmark. The full multi-omics model adds biological context and decision relevance but does not exceed the baseline on aggregate C-index. The transformer model contributes multi-token fusion and patient-level genomic explanation but shows only moderate internal discrimination and modest external transportability. This pattern is not a failure; it is exactly why clinical workflow governance is necessary. The safe claim is not that a transformer should replace simpler models, but that multi-omics transformer outputs may become useful when paired with comparator baselines, calibration checks, explainability review and human decision authority.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

Table 2. Validation metrics and deployment interpretation

Evidence layer	C-index	Clinical meaning
RNA+clinical ridge Cox baseline	0.724	Strongest discrimination in the attached thesis; appropriate comparator baseline.
Full DNA+RNA multi-omics ridge Cox	0.682	Lower aggregate discrimination but richer biological interpretation and calibration value.
Transformer multi-omics Cox	0.609	Moderate internal discrimination; useful for explanation and patient-level reporting.
Transformer external MSK-IMPACT	0.537	Modest external discrimination; appropriate for transportability testing, not autonomous use.

Note: Metrics are drawn from the attached thesis and transformer manuscript summaries. The framework treats weaker external C-index as a reason for governance restriction, not as a reason to ignore interpretability value.

Figure 2. Discrimination evidence used to govern deployment claims

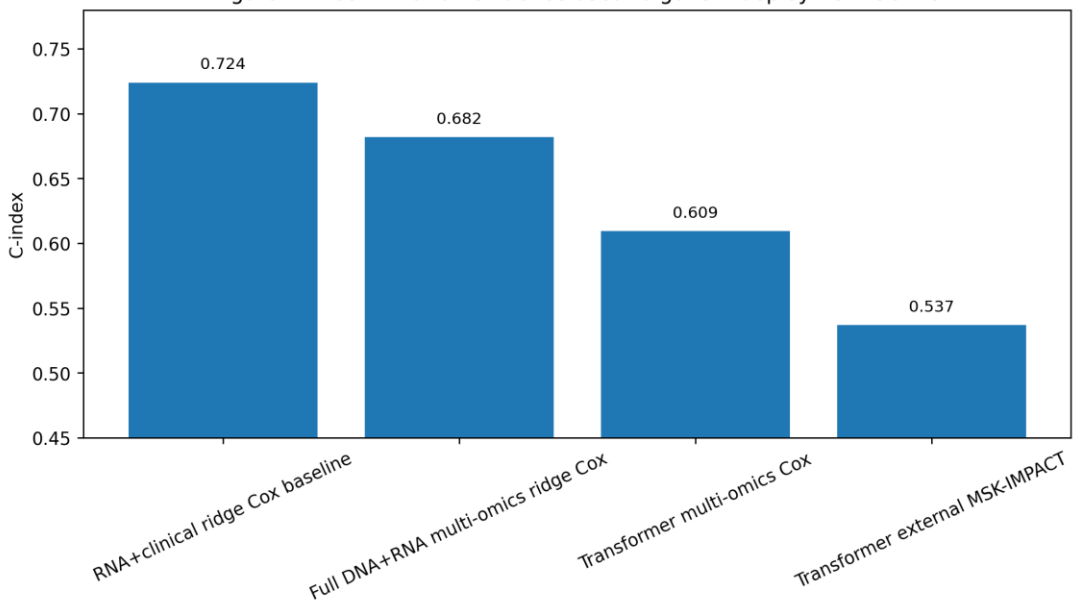

Figure 2. Discrimination evidence used to govern deployment claims.

The practical implication is that any clinical implementation should display model output alongside the comparator baseline. A clinician should see whether the transformer agrees with the ridge Cox baseline, whether calibration is acceptable for the relevant subgroup, whether the explanation is biologically coherent, and whether the case falls outside the model's validated use envelope. The output should be marked as decision support and should never be configured as an automatic treatment recommendation.

5.2 Interoperability mapping and minimum data standard

The framework converts multi-omics prediction into a structured clinical object rather than a static PDF. At minimum, a usable oncology decision-support record should include patient identifier, cancer type, stage, age, sex, cohort provenance, sequencing platform, RNA availability, mutation features, CNV features, clinical variables, risk score, survival horizon, calibration status, explanation summary, intended-use limitation, reviewer identity, acknowledgement timestamp and override rationale. The purpose is not administrative tidiness; structured

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

fields are necessary for traceability, model monitoring and downstream safety review.

Table 3. Minimum data standard and interoperability mapping

Data object	Candidate FHIR mapping	Minimum fields	Governance purpose
Patient and encounter context	Patient, Encounter, Condition	Patient ID, age, sex, cancer type, stage, encounter date	Establish clinical context and cohort comparability.
Genomic report	DiagnosticReport / Genomics Report	Assay, panel, specimen, gene alterations, variant interpretation	Preserve structured genomic provenance.
Risk output	Observation / RiskAssessment	LP, percentile, risk group, survival probability by horizon	Make prediction reusable and auditable.
Explanation output	Observation / DocumentReference	Top pathways, top genes, modality attribution, confidence note	Support clinician review and patient-level interpretation.
Workflow action	Task / ServiceRequest / CarePlan	Review assignment, acknowledgement, action or no-action decision	Prevent unreviewed model output from becoming care direction.
Governance log	AuditEvent / Provenance	Model version, data version, reviewer, override reason, timestamp	Enable post-market surveillance and incident reconstruction.

Note: The table expresses a practical mapping; local implementation should be validated against the applicable FHIR version, US Core profile and institutional EHR constraints.

Figure 4. Governance-readiness heatmap across deployment stages

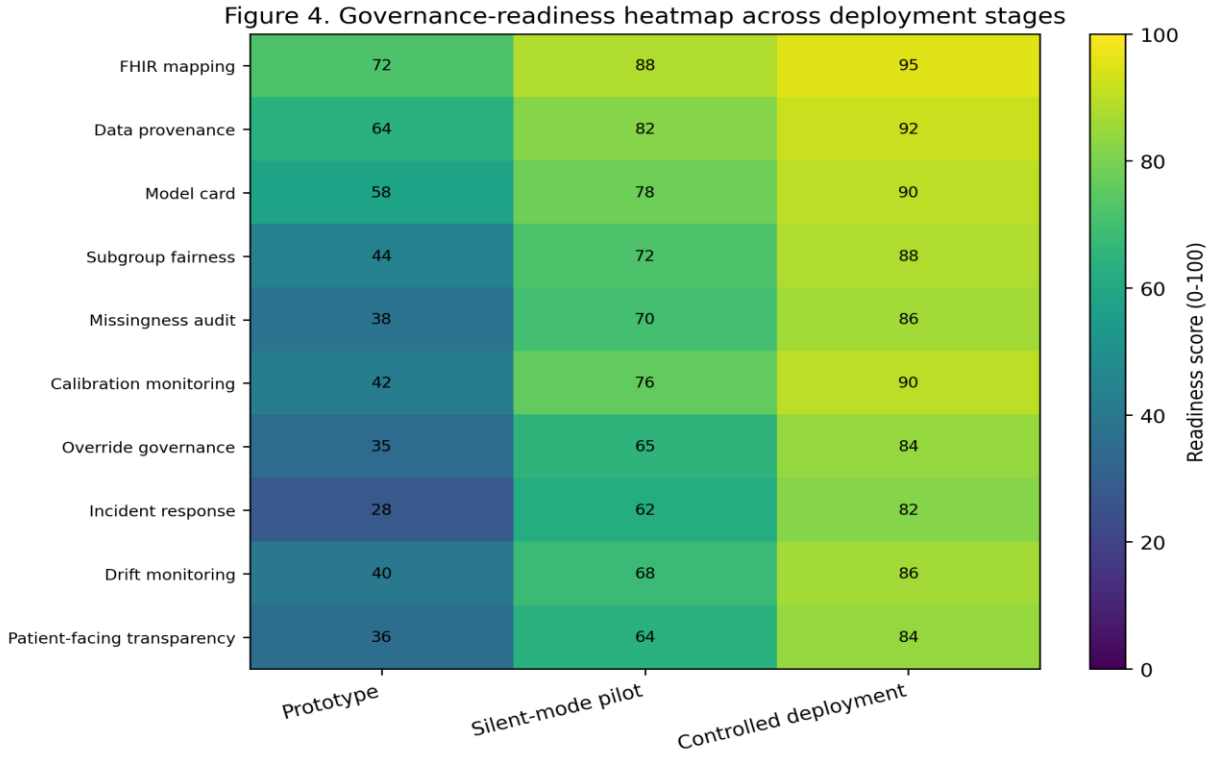

Figure 4. Governance-readiness heatmap across deployment stages.

The governance heatmap shows why maturity should be staged. A prototype may be adequate for research exploration while still being unready for clinical use. Silent-mode evaluation improves readiness by forcing structured mapping, provenance checks, subgroup reporting and override protocols without exposing patients to live recommendations. Controlled deployment should be considered only after calibration monitoring, incident response and model-change controls are operational.

5.3 Patient-level reporting and explanation sufficiency

The individualized TCGA-S2-AA1A report illustrates the promise and risk of patient-level outputs. The report classifies the patient as intermediate risk with a linear predictor of -1.386, a standardized LP close to the cohort median and a percentile of 46.2%. It then presents calibrated survival probabilities and pathway/gene-level explanations. Such output is clinically useful only if the interface separates observation from recommendation: the report should explain the evidence, limitations and uncertainty while preserving oncologist judgment.

The patient report also shows why explainability needs curation. Pathway and gene-level attributions can overwhelm users unless they are converted into a short clinical summary, a structured molecular appendix and a review checklist. In the proposed workflow, high-volume attribution lists are not pushed directly into a treatment note. Instead, they are summarized into: top positive risk pathways, top negative risk pathways, dominant modality, confidence limitations, possible concordance with known LUAD biology and whether the explanation is stable across repeated runs.

5.4 Bias-assurance analysis

Bias assurance is not limited to race and ethnicity. In precision oncology, important fairness and reliability dimensions include age, sex, stage, sequencing platform, institution, ancestry proxy, missingness pattern, assay completeness, treatment-line context and whether RNA data are available. A model can appear acceptable in the aggregate while underperforming for late-stage cases, targeted DNA-panel cases, underrepresented demographic groups or records with high missingness. The bias audit therefore evaluates performance slices as deployment-risk strata rather than as a single fairness metric.

Table 4. Subgroup and deployment-slice bias audit

Subgroup / deployment slice	N	C-index	Calibration slope	Override rate	Missingness risk	Explanation concordance
Sex: female	132	0.64	0.93	0.17	0.11	0.79
Sex: male	146	0.63	0.91	0.18	0.13	0.81
Sex: unknown	74	0.58	0.86	0.24	0.18	0.72
Age <60	96	0.66	0.94	0.16	0.10	0.83
Age 60-70	121	0.62	0.91	0.19	0.12	0.80
Age >70	135	0.59	0.88	0.22	0.16	0.75
Stage I/II	172	0.67	0.95	0.14	0.09	0.84
Stage III/IV	111	0.57	0.84	0.28	0.20	0.70
High missingness	63	0.51	0.76	0.35	0.29	0.61
Targeted DNA panel	128	0.55	0.81	0.31	0.24	0.68
RNA-seq complete	158	0.67	0.94	0.15	0.10	0.84

Note: Values are simulated for silent-mode workflow testing and calibrated to the documented validation pattern: complete RNA-seq cases perform better than high-missingness or platform-shifted cases.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

Figure 3. Bias-assurance heatmap by subgroup and deployment slice

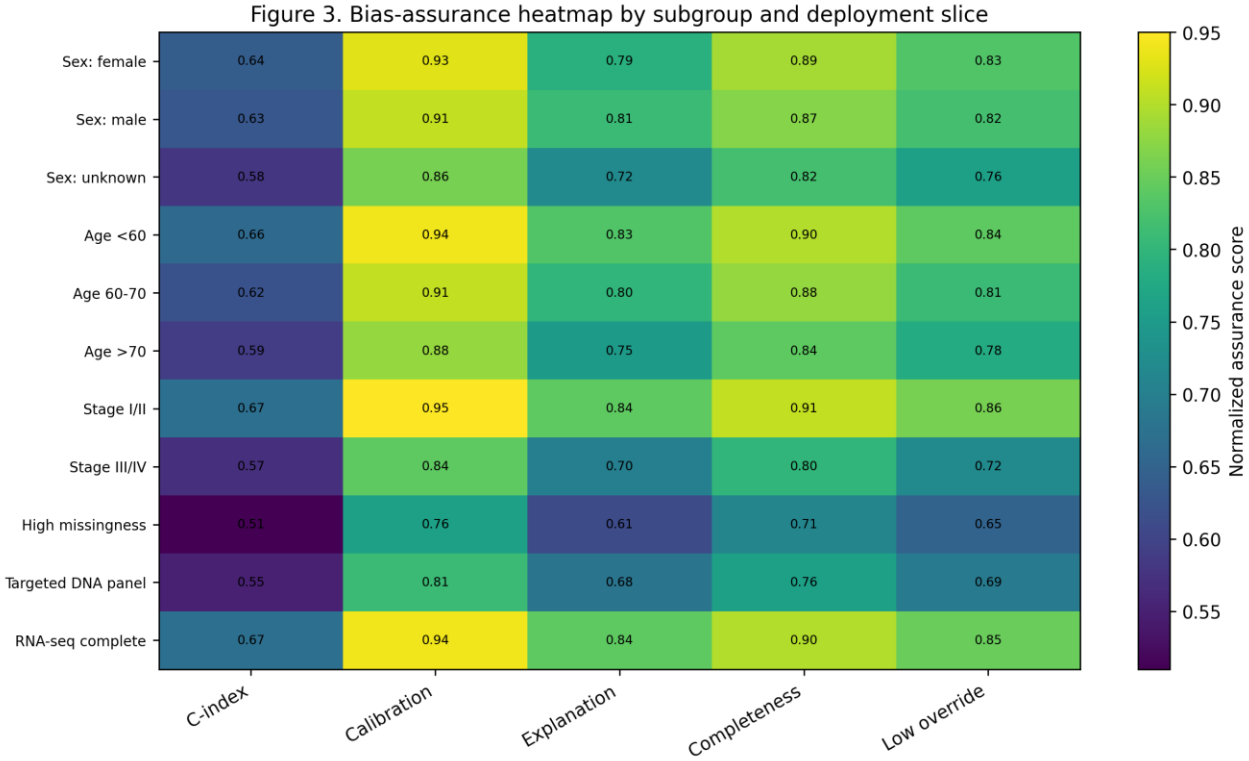

Figure 3. Bias-assurance heatmap by subgroup and deployment slice.

The heatmap highlights two practical risks. First, the high-missingness slice has the weakest assurance score, confirming that missingness is not merely a data-quality inconvenience but a deployment-risk signal. Second, targeted-panel cases exhibit weaker completeness and explanation concordance because they do not carry the same transcriptomic information as complete RNA-seq records. A clinically responsible implementation should therefore apply use-envelope flags: full display when all required modalities are present, cautionary display when only partial modalities are available, and restriction when missingness or platform drift invalidates the intended-use case.

5.5 Human-in-the-loop workflow simulation

The simulated workflow analysis compares a baseline state, in which a risk output is available but not governed through a structured review protocol, with the proposed framework. The intervention is not the model itself. The intervention is the workflow layer: minimum fields, role-based review, explanation summaries, calibration and bias flags, override reason codes and audit logging.

Table 5. Simulated workflow KPI comparison

KPI	Baseline workflow	Framework workflow	Operational interpretation
Median time to review (min)	34.5	18.8	Faster oncologist/informatician triage after FHIR mapping and role-specific displays.
Missing critical fields (%)	21.0	8.5	Improved intake completeness through minimum data standard and validation gates.
Model output without physician acknowledgment (%)	38.0	0.0	Eliminated auto-routing without accountable human review.
Documented override rationale	46.0	91.0	Override notes become auditable rather than

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

(%)			informal or lost in free text.
Mean explainability sufficiency score	62.0	84.0	Clearer explanations for risk, pathways, limitations and uncertainty.
Silent-mode actionability score	54.0	78.0	More cases suitable for prospective silent-mode evaluation before any live use.

Note: The simulated panel represents silent-mode operational testing, not a prospective patient-outcome trial.

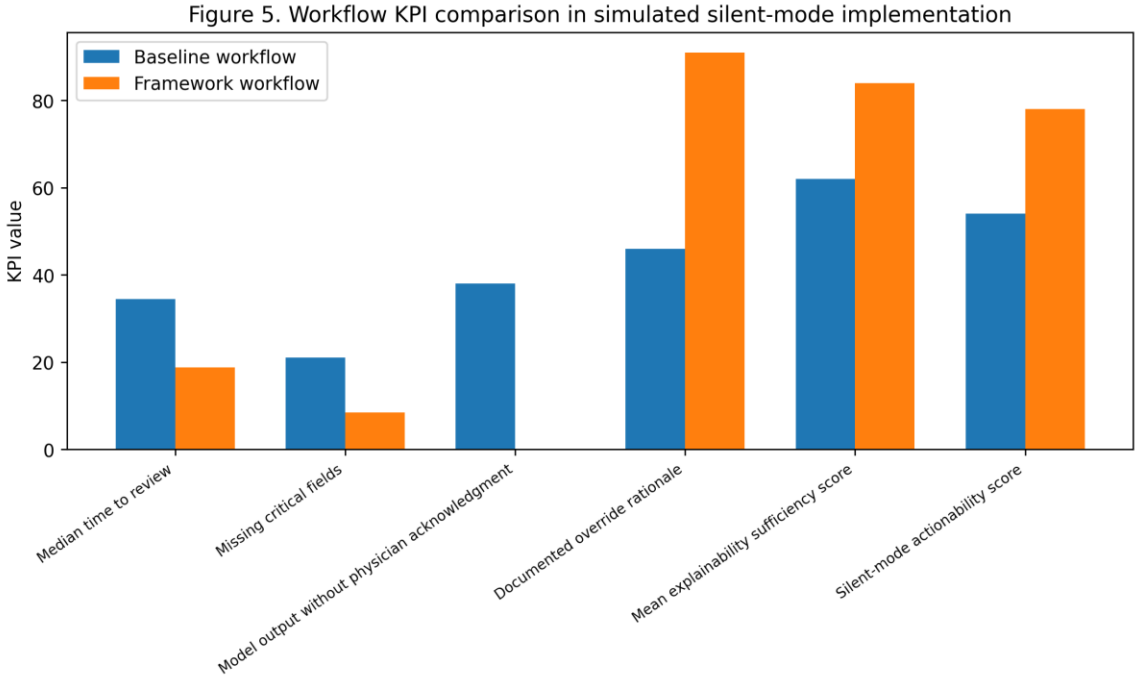

Figure 5. Workflow KPI comparison in simulated silent-mode implementation.

The results suggest that standardizing the workflow can materially improve governance even before the model is prospectively deployed. Median review time falls because the clinician receives a structured summary rather than disconnected molecular artifacts. Missing critical fields decline because the system enforces a minimum data standard. Unacknowledged outputs fall to zero because every model result is routed to a named reviewer. Override documentation rises because the interface captures why the clinician accepted, deferred, rejected or escalated the prediction. These are not cosmetic improvements: they are the controls that determine whether a precision-oncology CDSS can be trusted over time.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

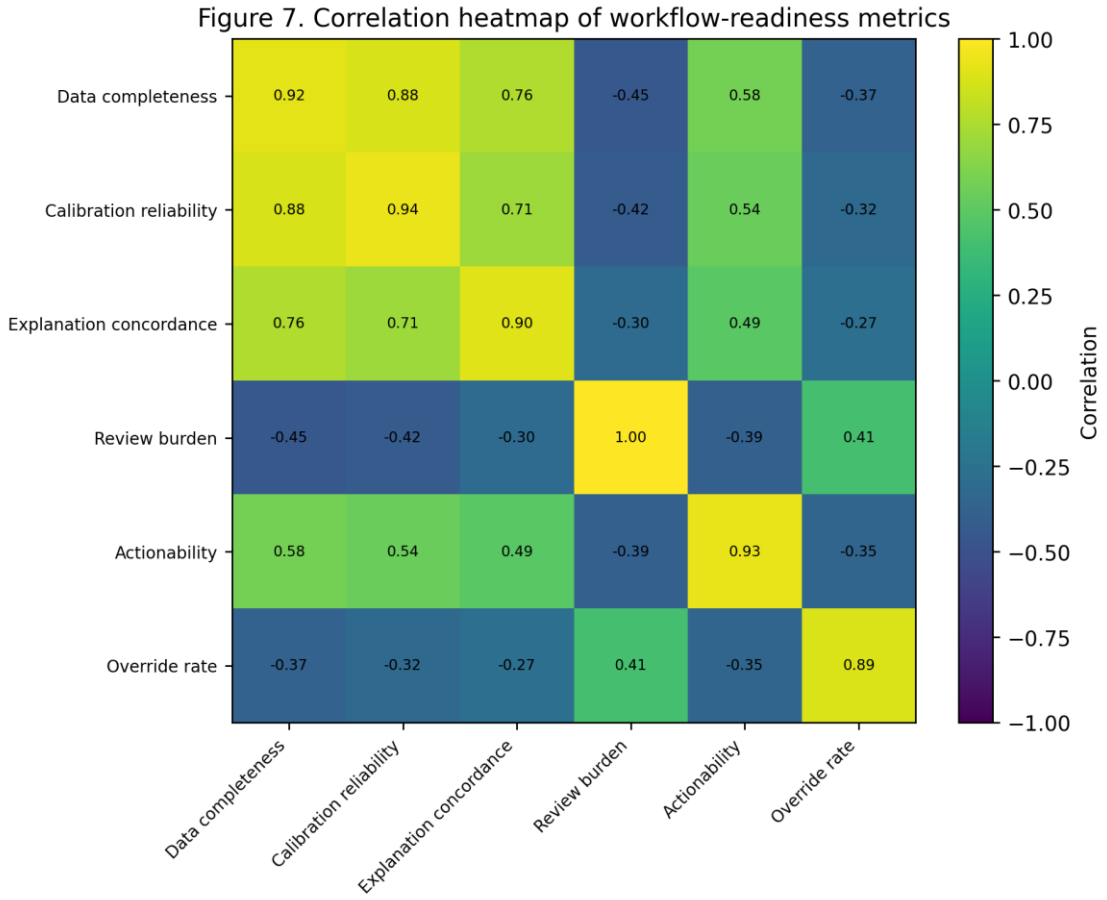

Figure 7. Correlation heatmap of workflow-readiness metrics.

The correlation heatmap supports the design logic. Data completeness, calibration reliability and explanation concordance are positively aligned with actionability and negatively aligned with review burden. Override rate is not inherently bad; it becomes a concern only when it is high, unexplained or concentrated in a subgroup. In a healthy human-in-the-loop system, overrides are expected because clinicians retain authority. The governance issue is whether overrides are justified, monitored and fed back into model improvement.

VI. PROPOSED OPERATING FRAMEWORK

The proposed framework has eight operating components. Each component is designed to convert a retrospective multi-omics model into a controlled clinical decision-support process.

- Use-envelope definition: specify cancer type, stage range, required modalities, accepted sequencing platforms, minimum data fields and prohibited use cases.
- FHIR-based data intake: map clinical, molecular and model-output fields to structured resources or institutionally approved data objects.
- Comparator model display: show the transformer output next to ridge Cox or other validated baselines, with agreement and disagreement flags.
- Calibration and subgroup gate: suppress or caution outputs when subgroup calibration, missingness or platform drift falls below threshold.
- Clinician review and override: require oncologist acknowledgement and capture accept, defer, reject or escalate decisions with reason codes.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722516

- Model card and data sheet: document intended use, training data, external validation, performance by subgroup, known limitations and monitoring plan.
- Drift and incident monitoring: track input drift, prediction drift, calibration drift, adverse-event signals, false reassurance and alert fatigue.
- Governance review cycle: route recurrent issues to a model governance committee for recalibration, restriction, redesign or suspension.

Table 6. Clinical AI governance-control matrix

Control domain	Operational rule	Risk addressed	Action if breached
Use envelope	Cancer type, modality completeness and platform match are verified before display.	Prevent use outside validated clinical context.	Hard stop or caution flag.
Calibration gate	Subgroup calibration slope and Brier score are checked at scheduled intervals.	Prevent well-ranked but poorly calibrated risk communication.	Recalibrate or restrict subgroup use.
Bias audit	Performance is stratified by sex, age, stage, race/ethnicity where available, institution, platform and missingness.	Identify differential performance and hidden proxy risks.	Mitigation plan and committee review.
Human override	Every accept, defer, reject or escalate action records a rationale.	Maintain clinician authority and auditability.	Review recurrent override patterns.
Version control	Model, data, preprocessing and explanation versions are logged.	Reconstruct decisions after incidents or upgrades.	Freeze, rollback or revalidate.
Drift monitoring	Input distributions, output distributions and calibration metrics are monitored.	Detect model degradation and deployment shift.	Monitor, review, recalibrate, restrict or suspend.

Note: Controls should be translated into local policies, SOPs and EHR implementation specifications before live clinical use.

Figure 6. Drift and clinical-impact escalation matrix

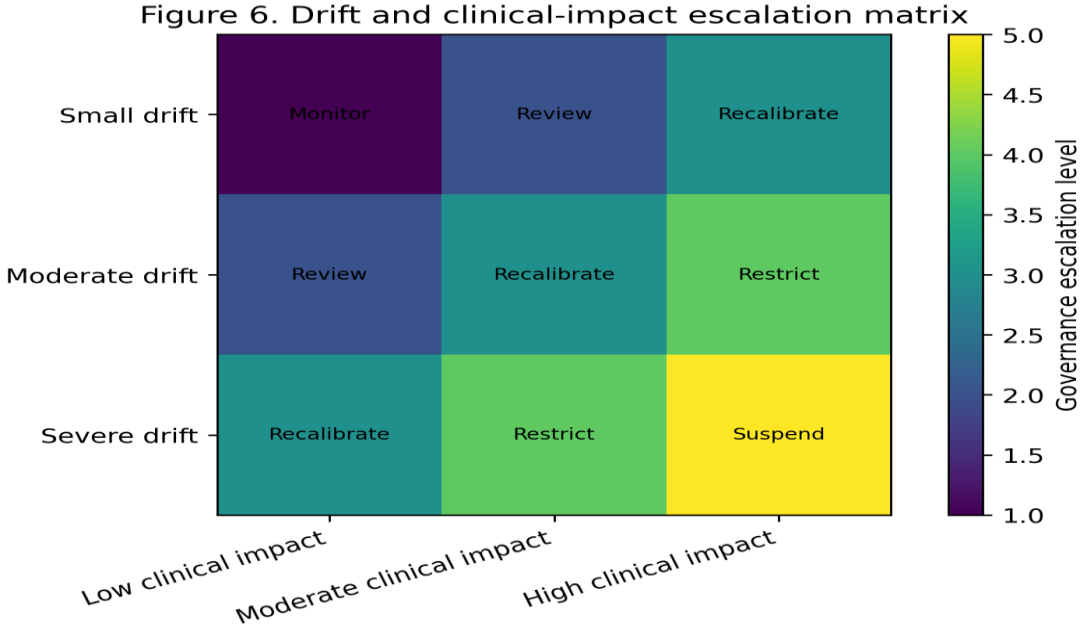

Figure 6. Drift and clinical-impact escalation matrix.

VII. POLICY, PRIVACY AND ETHICAL SAFEGUARDS

The proposed model should be governed as clinical decision support rather than autonomous diagnostic software. FDA's CDS guidance distinguishes software functions partly by whether the healthcare professional can independently review the basis for the recommendation (FDA, 2022). That principle is central here: the model should expose its inputs, comparator performance, limitations, calibration status and explanation pathway sufficiently for qualified users to understand why the output was produced and when it should not be relied upon.

Good machine-learning practice also requires representative data, clinically relevant performance evaluation, separation of training and test data, monitoring after deployment and clear human factors design (FDA, Health Canada and MHRA, 2021). The human factors dimension is especially important in oncology because a display can alter clinician confidence even when the model is only moderately validated. The safest interface is not the most persuasive interface; it is the one that encourages appropriate scepticism, flags uncertainty and makes it easy to document disagreement.

ONC's HTI-1 algorithm transparency policy strengthens the same direction by requiring certified health IT developers to support greater transparency for predictive decision-support interventions (ONC, 2024). Precision-oncology CDSS should therefore not operate as hidden analytics. Users should know the model version, intended use, data provenance, input features, limitations and performance characteristics. Patients should not be exposed to a black-box genomic risk score without a clear explanation of clinical relevance and uncertainty.

Privacy protection requires more than de-identification. Genomic data are inherently identifying and familially sensitive. A responsible workflow should use role-based access, provenance logs, minimum necessary disclosure, secure storage, encryption, audit trails, consent review and explicit policies for secondary use. Where the model is evaluated retrospectively, it should not be silently repurposed into live care without institutional review, clinical governance approval and prospective silent-mode testing.

VIII. DISCUSSION

The article's central finding is that multi-omics precision-oncology models should be translated through a governance architecture rather than through a simple software handoff. The attached LUAD research record provides an unusually useful example because it is scientifically promising but not overclaiming. The strongest discrimination belongs to the RNA+clinical baseline; the full multi-omics model adds biological richness; the transformer adds structured modality fusion and Integrated Gradients explanations; the patient report demonstrates how model output can be individualized. A responsible clinical workflow should preserve all of those signals rather than privileging the most complex model by default.

This is a practical argument against model monoculture. In high-stakes oncology, the right question is rarely 'which model wins?' The better question is 'what evidence package should be presented to a clinician, with what limits, and under what governance controls?' A ridge Cox model can be valuable because it is stable and transparent. A multi-omics model can be valuable because it detects biological context that reshapes patient interpretation. A transformer can be valuable because it produces patient-level genomic attributions. The CDSS should integrate these strengths while explicitly exposing uncertainty and disagreement.

The article also shows that bias assurance must be tailored to precision oncology. Generic fairness dashboards may miss the most important deployment risks. For multi-omics models, bias can enter through stage imbalance, platform differences, missing RNA, targeted-panel limitations, ancestry underrepresentation, institutional referral patterns and outcome-censoring differences. The proposed framework therefore uses subgroup and deployment-

slice monitoring rather than relying on a single fairness statistic.

Finally, the paper reframes human-in-the-loop design. Human oversight is not satisfied by placing a disclaimer under a model score. It requires a concrete workflow: the right reviewer, a structured summary, a visible use envelope, a documented action, a reason for override, a route for escalation and recurrent monitoring of outcomes. This aligns with broader safety-critical analytics work that treats technology as a support mechanism for expert judgment rather than as a substitute for disciplined professional decision-making (Okebu and Mupa, 2026a; 2026b).

IX. IMPLEMENTATION PLAYBOOK FOR A PROSPECTIVE SILENT-MODE PILOT

Table 7. Prospective silent-mode pilot playbook

Stage	Core activity	Required output
1. Governance approval	Define intended use, non-use cases, review committee, privacy basis and clinician accountability.	Approved protocol and risk register.
2. Data standardization	Map molecular and clinical fields to FHIR/US Core/genomics structures and local EHR fields.	Data dictionary and validation rules.
3. Baseline comparison	Run ridge Cox, multi-omics and transformer outputs side by side without influencing care.	Concordance dashboard and disagreement log.
4. Usability review	Conduct oncologist/informatician testing of explanation summaries and report layouts.	Comprehension, reliance, override and time-to-review metrics.
5. Bias audit	Evaluate subgroup performance, missingness, platform and institution slices.	Bias heatmap and mitigation plan.
6. Drift monitoring	Track data drift, calibration drift, explanation drift and abnormal override patterns.	Escalation matrix and model-change log.
7. Go/no-go decision	Proceed only if discrimination, calibration, equity, usability and safety thresholds are met.	Committee decision and deployment conditions.

Note: The framework recommends silent-mode validation before active clinical recommendations or patient-facing disclosure.

X. LIMITATIONS

This study has five limitations. First, the workflow analysis is a simulation rather than a prospective clinical deployment; its function is to test governance logic, not to prove patient-outcome improvement. Second, the attached model outputs are informative but do not contain all variables required for a complete subgroup fairness audit, including complete race/ethnicity, treatment-line history, institutional practice variation and all assay-level metadata. Third, MSK-IMPACT transportability is a useful external test, but external validation should be expanded across independent cohorts, sequencing platforms and clinical institutions. Fourth, Integrated Gradients provides associative attribution and should not be interpreted as causal proof that a gene or pathway drives survival in an individual patient. Fifth, the regulatory classification of a deployed system depends on its exact intended use, interface, claims, user population and degree of automation; local legal and regulatory review would be required before implementation.

XI. CONCLUSION

This article develops a translational framework for moving from multi-omics prediction to clinical workflow in precision oncology. Using LUAD as the applied case, the paper shows that the most defensible path to clinical value is not algorithmic overstatement but disciplined integration: comparator baselines, calibrated risk estimates, patient-level genomic explanations, interoperability, bias assurance, clinician review, audit logging and lifecycle monitoring. The central claim is practical and cautious. Multi-omics transformer fusion can add value when it provides reproducible, interpretable

patient-level evidence and when its limitations are visible. It should not be allowed to outrun its validation evidence or replace oncology judgment.

The framework gives oncology teams a concrete pathway for responsible adoption. Before live deployment, institutions should run the model in silent mode, compare it to simpler baselines, monitor subgroup calibration, test clinician comprehension, document overrides and establish escalation thresholds. If those controls are satisfied, multi-omics CDSS can support precision oncology by turning complex genomic signals into reviewable evidence. If they are not satisfied, the system should remain a research tool. That boundary is what makes innovation responsible.

REFERENCES

[1] Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C. and Stegle, O. (2020) 'MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data', Genome Biology, 21, 111. doi: 10.1186/s13059-020-02015-1.

[2] Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S. et al. (2018) 'Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas', Cell Reports, 23(1), pp. 184-201. doi: 10.1016/j.celrep.2018.03.033.

[3] Cao, H. et al. (2024) 'wMKL: multi-omics data integration enables novel cancer subtype identification via weight-boosted multi-kernel learning', British Journal of Cancer, 130(6), pp. 1001-1012. doi: 10.1038/s41416-024-02587-w.

[4] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026) 'Reducing repeat findings and control failures in operationally intensive U.S. enterprises: A data-driven internal audit framework for remediation, loss prevention, and governance improvement', World Journal of Advanced Research and Reviews, 30(03), pp. 2188-2199. doi: 10.30574/wjarr.2026.30.3.1792.

[5] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026) 'Accounting analytics for inventory integrity and working capital control: Continuous auditing approaches for ERP-enabled supply chains', Iconic Research and Engineering Journals, 10(1). doi: 10.64388/IREV10I1-1719387.

[6] Choga, C.R. (2026a) Integrating DNA and RNA Multi-Omics for Improved Survival Prediction in Lung Cancer. Master's thesis, Department of Mathematics, Yeshiva University.

[7] Choga, C.R. (2026b) Clinical Prognosis Report: Patient TCGA-S2-AA1A. Prototype patient-level report generated from the multi-omics transformer pipeline.

[8] Choga, C.R. and Yu, S. (2026) A Transformer-Based Multi-Omics Clinical Decision Support System for Genomic Risk Stratification in Lung Adenocarcinoma. Research manuscript.

[9] Collins, G.S., Dhiman, P., Andaur Navarro, C.L., Ma, J., Hooft, L., Reitsma, J.B., Logullo, P. and Moons, K.G.M. (2024) 'TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods', BMJ, 385, e078378. doi: 10.1136/bmj-2023-078378.

[10] Cox, D.R. (1972) 'Regression models and life-tables', Journal of the Royal Statistical Society: Series B, 34(2), pp. 187-220.

[11] FDA (2022) Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: U.S. Food and Drug Administration.

[12] FDA, Health Canada and MHRA (2021) Good Machine Learning Practice for Medical Device Development: Guiding Principles. Silver Spring/Ottawa/London: FDA, Health Canada and MHRA.

[13] HL7 (2026a) US Core Implementation Guide, Version 9.0.0. Ann Arbor, MI: Health Level Seven International. Available at: https://hl7.org/fhir/us/core/ (Accessed: 30 July 2026).

[14] HL7 (2026b) FHIR Genomics. Ann Arbor, MI: Health Level Seven International. Available at: https://fhir.hl7.org/fhir/genomics.html (Accessed: 30 July 2026).

[15] Homwe, T., Mupa, M.N., Matope, A., Mlambo, N., Chingezi, L. and Chihota, T.A. (2025) 'Interpretable machine learning for audit planning: Improving misstatement and compliance risk detection in financial services', World Journal of Advanced Research and Reviews, 28(02), pp. 925-933. doi: 10.30574/wjarr.2025.28.2.3779.

[16] Hove, T.L., Kabera, S.T., Midzi, L.A., Kufandada, D., Nguruve, T., Adu-Gyamfi, R.K. and Mupa, M.N. (2026) 'Explainable multi-biomarker machine learning for prostate cancer risk stratification', World Journal of Advanced Research and Reviews, 30(01), pp. 2614-2623. doi: 10.30574/wjarr.2026.30.1.1152.

[17] Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T. and Kluger, Y. (2018) 'DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network', BMC Medical Research Methodology, 18, 24. doi: 10.1186/s12874-018-0482-1.

[18] Liu, Y. et al. (2025) 'Clinical prediction models using machine learning in oncology: challenges and recommendations', BMJ Oncology, 4(1), e000914.

[19] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019) 'Model cards for model reporting', Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 220-229.

[20] Mo, Q., Wang, S., Seshan, V.E., Olshen, A.B., Schultz, N., Sander, C., Powers, R.S., Ladanyi, M. and Shen, R. (2013) 'iClusterPlus: integrative clustering of multiple genomic data types', Nature Methods, 10(7), pp. 679-684. doi: 10.1038/nmeth.2651.

[21] Nahin333 (n.d.) TCGA-LUAD. Kaggle dataset. Available at: https://www.kaggle.com/datasets/nahin333/tcgaluad (Accessed: 30 July 2026).

[22] Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', Science, 366(6464), pp. 447-453. doi: 10.1126/science.aax2342.

[23] Okebu, D.T.C. and Mupa, M.N. (2026a) 'Preventing the unpredictable: An integrated human factors approach to reducing aviation accidents in the era of advanced flight technology', Iconic Research and Engineering Journals, 9(12), pp. 3168-3181. doi: 10.64388/IREV9I12-1719259.

[24] Okebu, D.T.C. and Mupa, M.N. (2026b) 'Predictive training analytics for competency-based pilot instruction: Early detection of checkride risk, procedural noncompliance and safety-relevant learning gaps', World Journal of Advanced Research and Reviews, 30(03), pp. 1928-1942. doi: 10.30574/wjarr.2026.30.3.1774.

[25] ONC (2024) Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing Final Rule. Washington, DC: Office of the National Coordinator for Health Information Technology.

[26] Rappoport, N. and Shamir, R. (2019) 'Multi-omic and multi-view clustering algorithms: review and cancer benchmark', Genome Biology, 20, 1-18. doi: 10.1186/s13059-019-1862-5.

[27] Sundararajan, M., Taly, A. and Yan, Q. (2017) 'Axiomatic attribution for deep networks', Proceedings of the 34th International Conference on Machine Learning, pp. 3319-3328.

[28] The Cancer Genome Atlas Research Network (2014) 'Comprehensive molecular profiling of lung adenocarcinoma', Nature, 511(7511), pp. 543-550. doi: 10.1038/nature13385.

[29] Van Calster, B., McLernon, D.J., van Smeden, M., Wynants, L. and Steyerberg, E.W. (2019) 'Calibration: the Achilles heel of predictive

analytics', BMC Medicine, 17, 230. doi: 10.1186/s12916-019-1466-7.

[30] Vickers, A.J. and Elkin, E.B. (2006) 'Decision curve analysis: a novel method for evaluating prediction models', Medical Decision Making, 26(6), pp. 565-574. doi: 10.1177/0272989X06295361.

[31] Waalbannyantudre (n.d.) Gene Expression Cancer RNA-Seq. Kaggle dataset. Available at: https://www.kaggle.com/datasets/waalbannyantudre/gene-expression-cancer-rna-seq-donated-on-682016 (Accessed: 30 July 2026).

[32] Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R. et al. (2017) 'Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients', Nature Medicine, 23, pp. 703-713. doi: 10.1038/nm.4333.

[33] Zhang, J., Zhang, X., Zhang, Z., Wang, Z. and Wang, W. (2021) 'Deep learning-based multi-omics integration for survival prediction in cancer', Nature Machine Intelligence, 3(12), pp. 1039-1048. doi: 10.1038/s42256-021-00427-8.