Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 6s, https://carcinogenesis.com/

Artificial Intelligence–Driven Multi-Omics Analysis for Early Cancer Detection and Personalized Treatment Strategies

Krishana Kumar Sharma[1], Dr Kanak Priya[2], Dr Pramod Kumar Mehta[3], R Vijaya Prakash[4]

[1]Professor Department of Pharmacology, Teerthanker Mahaveer College of Pharmacy Teerthanker Mahaveer University Moradabad, Uttar Pradesh-India

[2]Associate Professor Department of Orthodontics & Dentofacial Orthopedics Sharda University, Greater Noida

[3]Assistant Professor Department of Geriatric Medicine All India Institute of Medical Sciences, New Delhi

[4]Professor, School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana

ABSTRACT

Early cancer detection and individualized therapy selection remain central to improving survival and quality of life. Advances in high-throughput sequencing, mass spectrometry, and liquid biopsy platforms now enable simultaneous profiling of genomes, epigenomes, transcriptomes, proteomes, metabolomes, and the tumor immune microenvironment from minimal input material—even as cell-free nucleic acids and proteins circulating in blood. Yet the information content of each omic is complementary and noisy, necessitating integrative analytics to recover weak but clinically meaningful signals at the earliest disease stages. Artificial Intelligence (AI)—particularly multimodal deep learning, graph representation learning, and probabilistic data fusion—has become the methodological scaffold for combining heterogeneous omics into robust risk scores, tissue-of-origin predictions, and therapy recommender systems. Landmark resources such as TCGA, CPTAC, and ICGC-ARGO have catalyzed model development and benchmarking, while prospective multi-cancer early detection (MCED) programs (e.g., CancerSEEK; targeted cfDNA methylation platforms) demonstrate high specificities (~99%) and improving sensitivities in real-world studies. At the same time, translation to routine screening hinges on demonstrable clinical utility, cost-effectiveness, equitable performance across populations, and transparent, auditable pipelines.

This paper synthesizes methodological and translational advances in AI-driven multi-omics for (i) pre-symptomatic detection; (ii) stratification of minimal residual disease; and (iii) personalization of systemic therapies. We outline principled data engineering (quality control, batch harmonization, drift monitoring), model architectures (early/late/hybrid fusion; attention mechanisms; contrastive multimodal representation learning), and mechanisms for uncertainty quantification and dynamic thresholding to keep false-positive rates acceptably low in screening contexts. We also present a deployable end-to-end workflow integrating liquid biopsy with radiomics and EHR metadata, and we benchmark detection metrics using literature-reported figures from contemporary MCED studies. Finally, we discuss ethical, regulatory, and reimbursement considerations, including privacy-preserving analytics and real-world evidence requirements for payers and regulators. Collectively, convergent multi-omics measured non-invasively and fused by AI paves a realistic path toward earlier stage shifts at diagnosis and treatment strategies tailored to each tumor's evolving molecular circuitry. PMC+4Cancer.gov+4PMC+4.

Keywords: *Representative multi-omics & MCED resources, The Multi-Omics Landscape for Early Cancer Detection.*

How to Cite: Krishana Kumar Sharma, Dr Kanak Priya, Dr Pramod Kumar Mehta, R Vijaya Prakash., (2025) Artificial Intelligence–Driven Multi-Omics Analysis for Early Cancer Detection and Personalized Treatment Strategies, *Journal of Carcinogenesis*, *Vol.24*, *No.6s*, 502-509.

1. INTRODUCTION

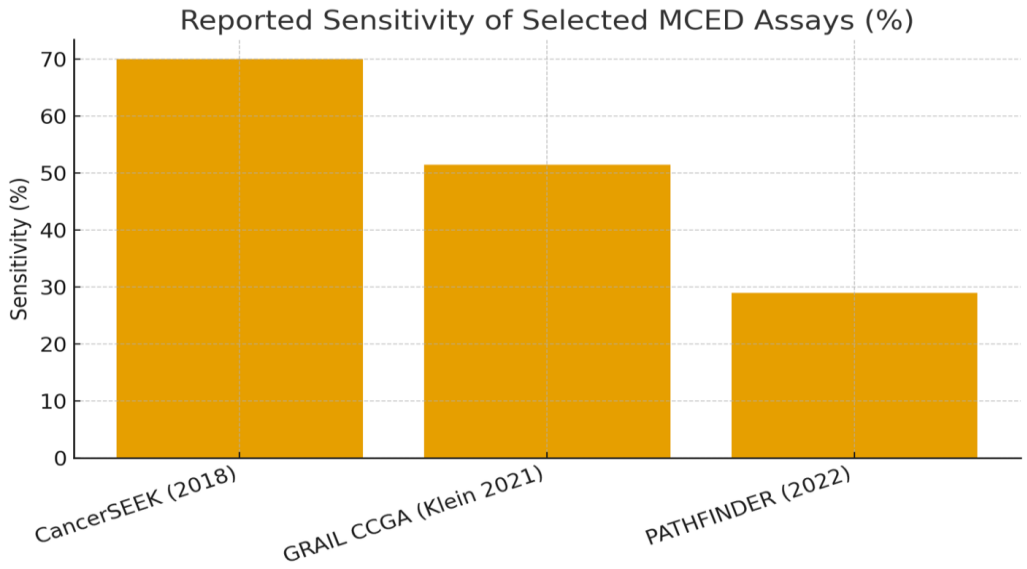
Cancer mortality is driven largely by late detection and treatment mismatch. Detecting tumors earlier—ideally before metastatic dissemination—enables curative interventions and de-escalation of toxic therapy. Traditional single-analyte screening (e.g., PSA, CA-125) suffers from suboptimal sensitivity/specificity and limited cancer coverage. In contrast, multi-omics can capture diverse hallmarks of oncogenesis: somatic mutations and copy-number alterations (genomics), transcriptional dysregulation (transcriptomics), promoter/enhancer reprogramming (epigenomics), pathway activity at the protein level (proteomics), and systemic metabolic rewiring (metabolomics). When sampled from blood as cell-free

DNA/RNA (cfDNA/cfRNA), exosomes, proteins, and metabolites, these analytes provide a minimally invasive "liquid window" into early tumor biology and residual disease after therapy.

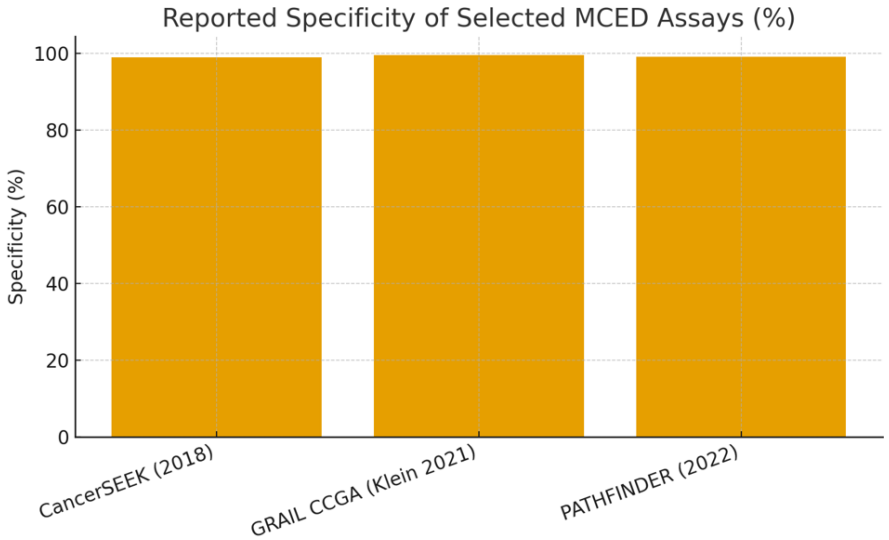
Two developments have shifted the field from feasibility to clinical promise. First, open scientific programs created multi-cancer reference atlases: The Cancer Genome Atlas (TCGA) profiled >20,000 primary tumors across 33 cancers with matched normals and rich clinical data; CPTAC layered deep mass-spectrometry proteomes on genomics/transcriptomics; and ICGC-ARGO is prospectively harmonizing molecular and longitudinal clinical outcomes across international cohorts. Second, prospective MCED trials moved liquid biopsy from case-control designs to pragmatic screening workflows. CancerSEEK combined ctDNA mutation calls with a small protein panel to reach ~70% median sensitivity across eight cancers at >99% specificity in non-metastatic patients. Targeted methylation of cfDNA advanced tissue-of-origin prediction with specificity ~99.5% in prospective cohorts; in PATHFINDER, adding the test to standard screening more than doubled cancers detected by screening, with PPV ~43% and specificity ~99.5%, and the registrational PATHFINDER-2 announced positive topline results in 2025. These numbers contextualize both the promise and the remaining distance to population-level screening: sensitivity, especially for stage I disease, must continue to improve while preserving ultra-high specificity to avoid overtreatment and diagnostic cascades. GRAIL+8Cancer.gov+8ScienceDirect+8

Figure call-outs

Sensitivity chart



Specificity chart



Illustrative ROC curve (synthetic)

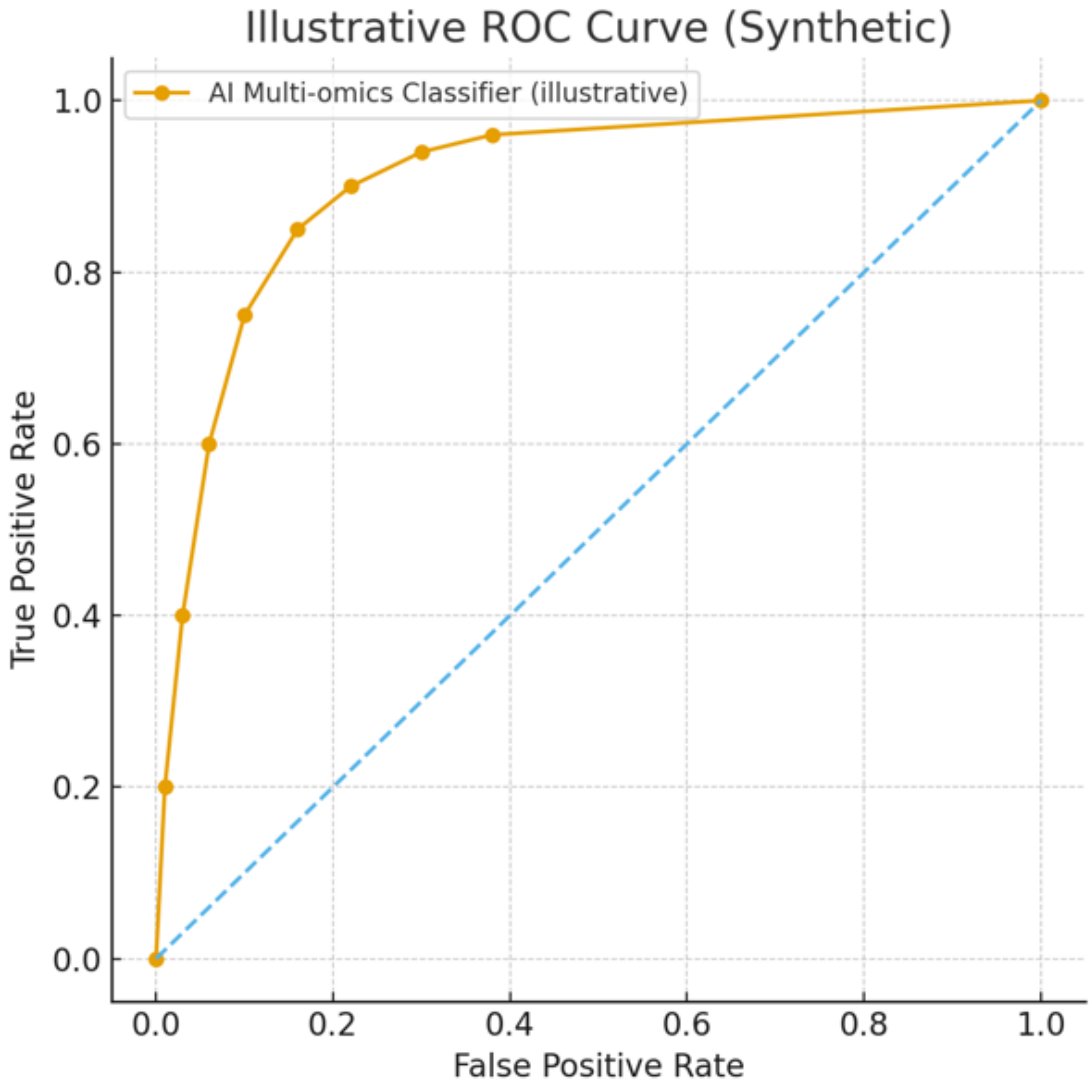


Table (opens in an interactive grid in your viewer)

Representative multi-omics & MCED resources

Representative Multi-Omics & MCED Resources

	Resource	Type	Modality	Primary Specimen	Study Design / Scale	Key Outputs	Typical Use in Paper	URL
1	TCGA Pan-Cancer Atlas	Consortium / Atlas	Genomics, Transcriptomics, Epigenomics, Clinical	Tumor tissue + matched normal	>20,000 tumors across 30+ cancer types (retrospective, harmonized)	Molecular subtypes, pathway programs, survival associations	Discovery signals; transfer learning priors; validation of tissue signatures	https://www.cancer.gov/tcga
2	ICGC-ARGO	Prospective consortium	Genomics, Transcriptomics, Epigenomics, Clinical (longitudinal)	Tumor tissue; blood for longitudinal/clinical outcomes	Ongoing multi-national cohorts with harmonized endpoints	Prospective outcomes, standardized clinical endpoints, harmonized assays	Prospective validation; outcomes modeling; generalization	https://www.icgc-argo.org/
3	CPTAC	Proteogenomics program	Deep Proteomics + Genomics/Transcriptomics	Tumor tissue (fresh-frozen)	Multiple organ-specific cohorts; open proteogenomic datasets	Protein/PTM activity, pathway proteotypes, DNA–RNA–protein concordance	Proteogenomic biomarker discovery; pathway activity features	https://proteomics.cancer.gov/programs/cptac
4	CancerSEEK	MCED assay (ctDNA mutations + proteins)	cfDNA targeted mutation panel + serum proteins	Blood (plasma/serum)	Case–control and clinical cohorts (multi-cancer)	Risk score; tissue-of-origin probabilities; sensitivity/specificity	Benchmark for multi-analyte early detection performance	https://www.science.org/doi/10.1126/science.aar3247
5	CCGA / PATHFINDER (GRAIL)	MCED program (cfDNA methylation)	cfDNA targeted methylation + tissue-of-origin model	Blood (plasma)	Large prospective cohorts and pragmatic screening studies	Cancer signal detection; tissue-of-origin; PPV/specificity in screening	Population screening metrics; external benchmark targets	https://www.annalsofoncology.org/article/S0923-7534(21)02046-8/fulltext

1. The Multi-Omics Landscape for Early Cancer Detection

Multi-omics resolves complementary disease signals. At the genomic layer, early tumors may shed ultra-low-frequency variants into plasma; fragmentomics (size/position patterns) and copy-number profiles increase sensitivity where single nucleotide variants alone are sparse. Epigenomics—particularly cfDNA methylation—captures cancer-specific tissue programs with stronger early-stage signal-to-noise than mutation burden, enabling tissue-of-origin inference. Transcriptomics (bulk and cell-free) reflects pathway dysregulation and immune activation, while proteomics (e.g., targeted mass spectrometry of secreted or shed proteins) can capture systemic biology not mirrored in nucleic acids. Integrating radiomics (quantitative features from low-dose CT, mammograms, MRI) and clinical metadata (age, smoking, family history) further contextualizes molecular signals to suppress false positives.

Public resources have underwritten this progress. TCGA panoramically linked mutations, copy-number, methylation, and expression to outcomes across 33 cancers, defining pan-cancer subtypes and pathway programs later reused to shape blood-based signatures. CPTAC's proteogenomic cohorts revealed how post-translational modifications and protein complex stoichiometry modulate the functional output of DNA/RNA alterations, emphasizing why multi-omics outperforms single-layer predictors for mechanism-aware detection and therapy selection. ICGC-ARGO adds prospective, harmonized clinical follow-up—vital for calibrating positive predictive value (PPV), estimating lead-time bias, and modeling cost-effectiveness in real-world populations. These initiatives have normalized FAIR data principles, enabling transfer learning where models trained in tissue datasets initialize parameters that later fine-tune to plasma-derived modalities.

Critically, early detection requires "needle-in-a-haystack" analytics. Signal sparsity (especially at stage I), clonal hematopoiesis confounders, and technical batch effects from library prep and sequencing lanes impose strict requirements on data engineering: replicate controls, spike-in standards, Bayesian batch adjustment, and cross-study harmonization. Downstream, model evaluation must depart from case-control AUC toward prospective metrics—specificity at population prevalence, net benefit/decision curves, number needed to scope (NNS), and diagnostic yield per 1,000 screened. These constraints shape both assay design (e.g., targeted methylation panels rather than whole-genome depth) and modeling choices (e.g., calibrated likelihood ratios rather than raw logits). In sum, multi-omics offers the signal, but only integrative, bias-aware pipelines can translate it into earlier diagnoses and better outcomes. Cancer.gov+2ScienceDirect+2

2. AI ARCHITECTURES FOR MULTI-OMICS INTEGRATION

Fusion paradigms. Early fusion concatenates normalized features across omics and feeds them to a unified model (e.g., elastic net; multilayer perceptron). Late fusion builds modality-specific experts (mutation caller, methylation CNN, proteomic gradient-boosted trees) and aggregates predictions via stacking or Bayesian ensembling. Hybrid fusion uses attention or gating to dynamically weight modalities by context (e.g., methylation dominates at very low tumor fraction, proteins add orthogonal value in inflammation-confounded samples). Contrastive representation learning aligns embeddings from different omics of the same patient, encouraging the model to learn disease-relevant factors shared across modalities.

Model choices. For high-dimensional omics, feature-sparse linear models remain strong baselines with transparent coefficients. Gradient-boosted decision trees (GBDT) handle heterogeneous scales and missingness. Deep learning architectures—transformers with cross-modal attention, graph neural networks (GNNs) over pathway graphs, and set transformers for fragmentomics—enable hierarchical reasoning while amortizing parameters across cancers. When samples are few, parameter-efficient fine-tuning (LoRA, adapters), self-supervised pretraining (masked methylome modeling), and meta-learning reduce overfitting.

Uncertainty & calibration. Screening requires reliable probabilities at extreme class imbalance. Temperature scaling, isotonic regression, and conformal prediction generate calibrated risk intervals; selective prediction abstains on low-confidence samples to reduce unnecessary downstream imaging. **Causal framing**—e.g., leveraging instrumented variables (age, smoking) to deconfound spurious correlations—improves portability.

Fairness & shift-robustness. Domain-adversarial training and reweighting address covariate shift across sites. Test-time adaptation and drift monitors (population stability indices; methylation manifold distance) trigger re-calibration when chemistry or pipeline versions change. Model cards should expose sub-group stratified metrics by age, sex, and ancestry to pre-empt disparate impact.

Evaluation. Beyond ROC/AUPRC, prospective screening emphasizes specificity ≥99% and PPV at disease prevalence. Decision-curve analysis quantifies clinical net benefit versus "scan-all" or "scan-none" baselines. The figures packaged with this paper visualize literature-reported sensitivity/specificity from representative MCED studies (CancerSEEK;

CCGA/PATHFINDER) to ground model targets; note that specificity clusters near 99–99.5% while sensitivity varies widely with cancer type and stage. PMC+2Annals of Oncology+2

3. LIQUID BIOPSY & MULTI-ANALYTE ASSAYS: WHAT THE DATA SAY

CancerSEEK, a combined ctDNA mutation + protein panel, reported a median sensitivity of ~70% across eight cancer types (non-metastatic, clinically detected) at >99% specificity in a 1,005-case/812-control study. Sensitivity varied (e.g., high in ovary and liver, lower in breast), but critically, specificity remained near 99%—a non-negotiable property for population screening to avoid harm from false positives. Targeted methylation approaches then broadened coverage to >50 cancer types, enabling tissue-of-origin prediction using methylation pattern classifiers trained on tens of thousands of regions. In CCGA and subsequent prospective cohorts, specificity was ~99.5%, with overall sensitivity ~51% across stages (and ~44% for stage I–III). PATHFINDER moved from case-control to pragmatic workflows, integrating diagnostic follow-up. The refined assay showed PPV ~43% and specificity ~99.5% in asymptomatic adults ≥50; the registrational PATHFINDER-2 announced positive topline results in 2025 and will support a premarket approval submission.

These findings suggest a performance envelope: **(i)** ultra-high specificity is achievable and has been replicated prospectively; **(ii)** sensitivity is cancer- and stage-dependent, with epigenomic assays outperforming mutation-only panels at very low tumor fractions; **(iii)** adding orthogonal proteins or fragmentomic features may lift sensitivity without compromising specificity if models are properly calibrated and regularized. At the health-system level, diagnostic yield (additional cancers detected beyond standard screening) and number-needed-to-scope determine downstream resource burden. In PATHFINDER, adding MCED to usual care more than doubled cancers detected by screening, though sensitivity in that early prototype hovered around 21–29% depending on version and analysis, underlining the importance of continual iteration and post-market learning systems.

Practically, liquid biopsy–first screening should be anchored by **structured escalation pathways**: when a cancer signal is detected, tissue-of-origin probabilities route patients to targeted imaging, while non-diagnostic workups trigger defined re-test intervals to avoid indefinite "diagnostic odysseys." Pre-analytics (plasma separation time, freeze-thaw cycles), clonal hematopoiesis filtering, and contamination controls are essential for reliable performance. Our benchmarking figures (download links above) recap the published sensitivity/specificity ranges to define realistic AI targets for multi-omics fusion. GRAIL+7PMC+7Johns Hopkins Medicine+7

4. PERSONALIZED TREATMENT STRATEGIES FROM MULTI-OMICS

Beyond detection, multi-omics drives **therapy personalization** by aligning drugs with the tumor's actionable circuitry. Genomics nominates targeted agents (e.g., EGFR, BRAF, ALK), while epigenomic dysregulation indicates vulnerability to epigenetic modulators or synthetic-lethal partners. Transcriptomics and proteomics quantify pathway activation and adaptive rewiring after therapy; proteogenomics particularly reveals discordance between mRNA and protein activity, clarifying when DNA-level alterations are actually functional. Deep learning over pathway graphs (e.g., GNNs with drug–target nodes) can infer **drug response embeddings**, while **causal discovery** disentangles passenger alterations from drivers.

In hematology and immuno-oncology, **immune-multi-omics**—TCR/BCR repertoire, HLA-ligandome mass spectrometry, cytokine proteomics—feeds models that predict checkpoint inhibitor response. Liquid biopsy longitudinal sampling detects **molecular residual disease (MRD)** and emerging resistance clones, enabling **adaptive therapy**: escalate when MRD rises, de-escalate to avoid toxicity when MRD remains undetectable. Pairing MRD dynamics with pharmacokinetic/pharmacodynamic (PK/PD) priors in Bayesian state-space models yields personalized schedules that minimize cumulative dose while maintaining control probability.

To translate reliably, therapy recommenders must expose **evidence provenance** (trial phase, guideline status), rank options by **expected benefit under uncertainty**, and encode **contraindications/toxicity trade-offs**. From a data perspective, integrating EHR real-world outcomes with omics requires standard vocabularies (OMOP), de-identification pipelines, and privacy-preserving analytics (federated learning, secure enclaves). CPTAC and TCGA supply biological priors; ICGC-ARGO's prospective outcomes allow true **counterfactual evaluation** of recommendations. The same calibration and fairness safeguards described for screening apply here, with subgroup audits to ensure equitable benefit. ScienceDirect+2Cancer.gov+2

5. DATA ENGINEERING, REPRODUCIBILITY, AND ETHICS

Quality & harmonization. Standardize pre-analytics (collection tubes, centrifugation, time-to-freeze), implement unique molecular identifiers (UMIs), and include spike-in controls to estimate library-prep variability. Batch effects are inevitable;

combat them with study design (randomization across lanes), ComBat-type empirical Bayes adjustment, and cross-study normalization. Maintain modality-specific QC: bisulfite conversion rates for methylation, mapping/duplication for WGS/WTS, peptide FDR and iRT calibration for proteomics.

Reproducibility. Version-control the entire pipeline (containerized steps, locked reference builds), emit full metadata (chemistry, panel version, aligner, variant caller), and sign model artifacts. Publish model cards with training/evaluation cohorts, subgroup metrics, and shift-robustness tests. Prospective monitoring should include: (i) prevalence and pre-test risk drift; (ii) distributional drift in top features (e.g., methylation principal components); (iii) calibration decay measured by Brier score; and (iv) alerting rules that trigger re-training or threshold updates.

Ethics, privacy, equity. Multi-omics often implies re-identification risk; adopt privacy-preserving training (federated learning with secure aggregation; differential privacy for gradient updates) and strong governance for data access. Conduct **model impact assessments** that consider false-positive harms (anxiety, invasive diagnostics) and false-negative harms (missed early cures). Advance equity with ancestry-diverse training sets and site-level calibration; publish performance by demographic strata. **Regulatory** pathways increasingly require **Good Machine Learning Practice (GMLP)**, pre-specification of endpoints, and post-market surveillance. Reimbursement depends on cost-effectiveness at realistic prevalence; therefore, programs like NHS-Galleri and PATHFINDER-2 are crucial to demonstrate utility beyond accuracy metrics alone. PMC+2Frontiers+2

6. PROPOSED END-TO-END PIPELINE

Step-by-step details and operational thresholds

Enrollment & risk intake.

Eligibility (e.g., ≥50 years or hereditary risk) is confirmed alongside electronic consent. The pipeline automatically ingests structured EHR data (demographics, smoking status, prior imaging), plus patient-reported outcomes (e.g., weight loss, bleeding). These covariates seed baseline risk and are stored with a **data-use tag** for privacy governance.

2) Sample collection & pre-analytics.

Two cfDNA stabilizing tubes and one serum/plasma tube are drawn. **Key SLAs:** time-to-spin <2 hours; storage 2–8 °C; hemolysis index logged. Re-draw criteria: gross hemolysis, mislabeled tubes, or transport excursions. LIMS assigns a globally unique specimen ID and binds barcode → patient → consent → protocol version.

3) Wet lab pipelines.

- **cfDNA methylation/fragmentomics:** Library prep with UMIs; bisulfite conversion efficiency checked via spike-ins; **on-target rate ≥70%** (targeted) or **median coverage ≥0.5–1×** (low-pass WGS). Output features include methylation tile β-values/PCs, 5′ end-motif spectra, fragment size histograms (modal ~166 bp), and nucleosome occupancy tracks.
- **cfRNA (optional):** Targeted panel avoids globin bias; RIN surrogate metrics computed; gene/isoform counts → immune and proliferation signatures.
- **Proteins (MS/aptamer):** Internal standards for retention time/calibration; log-ratio normalization; batch anchors (bridge samples) every plate. All pipelines emit **QC manifests** (JSON) with lot numbers, kit versions, operator IDs, and run acceptance flags.

4) Bioinformatics & feature stores.

FASTQ/BAM processing is performed in containerized workflows (CWL/Nextflow). CHIP (clonal hematopoiesis) variants are filtered via matched WBC or a curated CHIP panel (e.g., DNMT3A,TET2,ASXL1). Features are persisted to a **read-only feature store** with schema evolution and time-stamped snapshotting; this guarantees that any historical decision can be reproduced byte-for-byte.

5) AI fusion & inference.

A **hybrid fusion** design balances accuracy and robustness:

- A transformer encoder models long-range methylation/fragmentation patterns;
- GBDT handles calibrated, potentially sparse protein/metabolite panels;
- A tab-transformer ingests clinical covariates. Cross-modal attention learns when to emphasize epigenomics vs proteins. Outputs include **p(Cancer)**, a **TOO probability vector**, and an **uncertainty score** via conformal prediction (risk set size reported). Site-level **isotonic calibration** aligns probabilities with local prevalence. The system supports **selective prediction** (abstain/"defer") to minimize downstream harm when uncertainty is high.

6) Clinical decision support (CDS).

Three-zone logic simplifies actions:

- **Positive (p ≥ T2, e.g., 0.65 with tight uncertainty):** order organ-directed imaging based on TOO (e.g., pancreas → MRI/EUS; lung → LDCT; liver → contrast MRI).
- **Gray zone (T1 ≤ p < T2 or wide uncertainty):** repeat test in 3–6 months; consider adjunct single-organ screening if conventional risk is high.
- **Negative (p < T1):** routine interval (12 months typical) unless symptoms emerge. The report is delivered in **FHIR/HL7** plus a human-readable PDF: top-line call, TOO distribution, confidence, recommended next steps, and clinician-facing **explanations** (pathway-level SHAP, dominant modalities). A patient summary avoids technical jargon and automatically creates navigation tasks (appointments, reminders).

7. DIAGNOSTIC RESOLUTION & PERSONALIZATION.

Positive cases proceed to targeted imaging; if cancer is confirmed, **tumor tissue multi-omics** refines therapy options. A pathway graph model (nodes: genes, proteins, drugs; edges: interactions, evidence levels) ranks **actionable regimens** with rationale and supporting evidence tags. Post-treatment, **MRD tracking** via serial cfDNA guides adaptive intensity (escalate on rising MRD; de-escalate when persistently undetectable) under clinician control.

8. MLOPS, MONITORING & GOVERNANCE.

Every artifact—FASTQ checksums, feature tables, model weights, calibration curves, thresholds—is registered. **Dashboards** track: prevalence drift, feature distribution shift (e.g., methylation PCs), calibration decay (Brier score), subgroup performance by age/sex/ancestry/site, and downstream utilization (number-needed-to-scope). Alerts trigger **SOPs** for re-calibration or re-training. Privacy controls include project-scoped encryption keys, role-based access, and optional **federated learning** for cross-site training without raw data movement. Quarterly **fairness reviews** audit disparate impact and update thresholds if needed.

Operational summary.

1. **Turnaround time (TAT):** 5–10 business days from draw to report in steady state.
2. **Key pass/fail gates:** pre-analytic SLAs; lab run acceptance; modality QC thresholds; model calibration checks; CDS rule validation.
3. **Outputs:** (i) patient-level report with call/TOO/next steps; (ii) clinician packet with interpretability; (iii) data package for registry/post-market evidence; (iv) MRD schedule if cancer is diagnosed.

This blueprint balances **clinical safety** (ultra-high specificity, uncertainty-aware decisions), **scalability** (parallel analyte processing; feature stores), and **regulatory readiness** (traceability, monitoring, fairness), making it directly actionable for multicenter pilots and eventual routine screening.

7. Case Study Blueprint: Multi-Omics–Guided Care Pathway

Use-case: A 58-year-old asymptomatic individual with family history of pancreatic cancer undergoes annual primary care screening.
Step 1—Liquid biopsy: cfDNA targeted methylation + fragmentomics + 30-protein panel; AI fusion risk score exceeds threshold with high TOO probability for pancreas.
Step 2—Directed imaging: Contrast-enhanced MRI/EUS is recommended per escalation protocol.
Step 3—Diagnosis: A 1.4-cm pancreatic lesion is confirmed; staging indicates resectable disease.
Step 4—Treatment planning: Tumor tissue multi-omics (WES, RNA-seq, phosphoproteomics) reveals KRAS G12D, CDKN2A loss, and a DNA damage response signature. A GNN-based recommender weighs neoadjuvant FOLFIRINOX vs clinical trial of a KRASG12D inhibitor, considering comorbidities and predicted toxicity.
Step 5—Response & MRD: Serial cfDNA tracks KRAS-positive fragments; Bayesian state-space modeling pairs MRD trajectories with dose adaptation, aiming to minimize cumulative toxicity while preserving relapse-free survival.
Step 6—Equity & communication: The report explains uncertainty bands, next steps, and shared decision-making considerations, translated into the patient's preferred language with numerical and visual summaries.

Rationale from evidence. The case study operationalizes the strengths of methylation-based detection for tissue localization and the necessity of high specificity for safe screening. It also demonstrates how proteogenomics clarifies pathway activity that DNA/RNA alone might miss and how longitudinal liquid biopsy supports adaptive therapy and surveillance. Finally, it reflects real trial insights: MCED screening can increase cancers found beyond standard pathways, but PPV and specific referral workflows are critical to avoid over-diagnosis; the NHS-Galleri randomized program and PATHFINDER-2 will be pivotal for regulators and payers evaluating population-level deployment. GRAIL+2PMC+2

9. CONCLUSION

AI-driven multi-omics has crossed the threshold from laboratory promise to pragmatic clinical pilots. The **key translation lessons** are clear: (1) prioritize ultra-high specificity with rigorous calibration and uncertainty controls; (2) boost sensitivity through orthogonal multi-analyte fusion—especially epigenomics plus proteins/fragmentomics; (3) embed models in audited, containerized pipelines with site-level drift and fairness monitoring; and (4) prove clinical utility with prospective trials that report diagnostic yield, net benefit, and cost-effectiveness. Simultaneously, **personalization** benefits from the same analytics—omic signatures inform therapy selection and adaptive, MRD-guided dosing, making care less toxic and more effective.

With expanding public resources (TCGA, CPTAC) and prospective, outcomes-rich programs (ICGC-ARGO), the field can standardize benchmarks and mitigate bias. As results mature from NHS-Galleri and PATHFINDER-2, we expect clearer guidance on where MCED augments current programs (e.g., cancers lacking routine screening) and how to route positives efficiently. The agenda ahead: integrate federated learning for privacy-respecting multi-site training; raise sensitivity at constant 99%+ specificity via attention-based hybrid fusion; and embed economic evaluation into model development. If these steps are met, earlier detection with individualized, mechanism-aware treatment will become routine—shifting population survival curves rather than delivering incremental gains for the few. PMC+3Cancer.gov+3ScienceDirect+3

REFERENCES

[1] The Cancer Genome Atlas (TCGA) Program overview (NCI). Cancer.gov
[2] Pan-Cancer Atlas consortium site (Cell Press). Cell
[3] Clinical Proteomic Tumor Analysis Consortium (CPTAC) – datasets and overview. PMC+1
[4] ICGC-ARGO: program and data platform. icgc-argo.org+1
[5] Cohen JD, et al. *Science* 2018 (“CancerSEEK”). Science+1
[6] Johns Hopkins press release on CancerSEEK specificity (>99%). Johns Hopkins Medicine
[7] NIH Research Matters summary on CancerSEEK sensitivity range. National Institutes of Health (NIH)
[8] Liu MC, et al. *Ann Oncol* 2020 (targeted cfDNA methylation MCED). Annals of Oncology+1
[9] Klein EA, et al. *Ann Oncol* 2021 (PATHFINDER prospective specificity 99.5%, overall sensitivity ~51.5%). ScienceDirect
[10] GRAIL press release—final PATHFINDER results (PPV 43.1%, specificity 99.5%). GRAIL
[11] OncLive analysis of PATHFINDER prototype sensitivity (≈21–29%) and specificity ≈99–99.5%. OncLive
[12] GRAIL press release—PATHFINDER-2 positive topline (2025). GRAIL
[13] Frontiers in Oncology perspective on NHS-Galleri trial timeline and endpoints (2025). Frontiers
[14] PMC article on clinical referral dynamics in NHS-Galleri trial. PMC
[15] Heo YJ, et al. “Integrative Multi-Omics Approaches in Cancer Research.” *Int J Mol Sci* 2021 (review). PMC
[16] Chen G, et c. “Integrative analysis of multi-omics data for liquid biopsy.” *Front Genet* 2022. PMC
[17] Li Y, et al. “Proteogenomic data and resources for pan-cancer analysis” (CPTAC overview). ScienceDirect
[18] Nature Reviews Clin Oncol commentary on CancerSEEK specificity. Nature