From Bench to Biomarkers: RNA-seq-Driven Discovery of Chemoprevention Biomarkers in Animal Models

Ashwini Badhe[1], Pravin Badhe[2]

[1,2] Swalife Biotech Ltd North Point House, North Point Business Park, New Mallow Road, Cork (Republic of Ireland)

Corresponding author Email: info@swalifebiotech.com
Doi: **10.5281/zenodo.1873455**
Received: 08 January 2026
Accepted: 19 January 2026

Abstract

The successful clinical development of cancer chemopreventive agents critically depends on mechanistic understanding and accompanying pharmacodynamic (PD) biomarkers providing early readouts of biological activity in target tissues and circulation. RNA sequencing (RNA-seq) has revolutionized biomarker discovery in preclinical cancer chemoprevention models, enabling unbiased transcriptomic analysis identifying mechanism-informed molecular signatures substantially more sensitive than traditional phenotypic endpoints. This comprehensive review synthesizes knowledge on RNA-seq-driven biomarker discovery in animal models for cancer prevention, emphasizing translational validation strategies bridging bench to clinic. We examine cross-species conservation of transcriptomic signatures derived from rodent chemoprevention models to human disease, discuss practical validation methodologies including quantitative PCR (qPCR), immunohistochemistry (IHC), and emerging liquid biopsy approaches, and address critical challenges in establishing clinical relevance of preclinically discovered biomarkers. Specific emphasis is placed on tissue versus circulating biomarkers, temporal dynamics of biomarker changes during prevention interventions, context-dependent biomarker effectiveness, and integration of tumor microenvironment factors influencing biomarker interpretation. Case studies demonstrate successful biomarker translation including withaferin A-induced p21 expression, oral squamous cell carcinoma transcriptional signatures predicting metastatic risk, and circulating nucleic acid biomarkers for non-invasive cancer monitoring. The review discusses FDA biomarker qualification processes, window-of-opportunity clinical trial designs leveraging preclinical insights, and emerging machine learning approaches for biomarker interpretation and patient stratification in prevention trials.

Keywords: RNA-seq, biomarkers, chemoprevention, animal models, translational research, qPCR, immunohistochemistry, circulating biomarkers, cross-species validation

1. Introduction

The translation of promising laboratory discoveries in cancer chemoprevention to successful clinical implementation faces a critical bottleneck: lack of validated pharmacodynamic (PD) biomarkers providing early readouts of biological activity in target tissues[1][2]. Unlike cardiovascular drug development where biomarkers like blood pressure and cholesterol provide immediate efficacy indicators, cancer chemoprevention lacks equivalent readily-measurable surrogate markers predicting ultimate cancer reduction[1]. This mechanistic gap has contributed to numerous failed clinical trials of agents demonstrating robust preclinical efficacy, costing billions in development while delaying promising preventive strategies[1][2].

RNA sequencing technology has fundamentally transformed biomarker discovery capabilities in preclinical cancer chemoprevention models[3][4]. Unlike microarray hybridization limited to known sequences and single-gene qPCR capturing individual targets, RNA-seq provides unbiased, genome-wide transcriptomic profiling identifying coordinated gene expression changes comprising mechanistically-informed molecular signatures[3][4]. These comprehensive transcriptomic signatures reveal biomarkers emerging from drug-pathway engagement, enabling discovery of mechanism-specific markers substantially more sensitive than phenotypic endpoints (cell death, growth inhibition) and better predicting efficacy[3][4].

The translational power of RNA-seq-derived biomarkers emerges from several advantages: (1) unbiased discovery identifying unexpected mechanism participants and pathway cooperativity, (2) mechanistic specificity distinguishing direct target engagement from adaptive responses, (3) discovery of cancer-selective pathways activated preferentially in malignant versus normal tissues, and (4) identification of biomarkers existing in readily-accessible circulating compartments enabling non-invasive clinical monitoring[1][2][3][4].

However, realization of this potential requires rigorous validation strategies bridging bench discoveries to clinical utility: demonstrating that preclinically-identified transcriptomic signatures predict human disease, establishing assay formats suitable for clinical implementation, confirming cross-species relevance, and proving biomarker modulation correlates with functional efficacy. This review synthesizes current knowledge on RNA-seq biomarker discovery in animal chemoprevention models and practical validation strategies enabling clinical translation.

2. RNA-seq Discovery of Mechanistic Biomarkers in Animal Cancer Models

2.1 Global Transcriptomic Profiling Identifies Cancer-Selective Pathways

Exemplifying RNA-seq-driven biomarker discovery, withaferin A (WA) a steroidal lactone from Withania somnifera demonstrated cancer chemopreventive efficacy in multiple rodent breast cancer models despite undefined mechanistic basis, limiting clinical development[5]. Global RNA-seq profiling comparing WA-treated breast cancer cells (MCF-7, SK-BR-3, MDA-MB-231) versus normal mammary epithelial cells (MCF-10A) and validated in HER2-driven transgenic (MMTV-neu) and N-methyl-N-nitrosourea (MNU)-induced rodent models revealed striking cancer selectivity: WA robustly upregulated senescence marker p21 in cancer cells (5-20 fold) but only modestly in normal cells (1.2-fold), establishing p21 as cancer-selective biomarker[5].

Mechanistically, RNA-seq pathway analysis using Gene Ontology (GO) and Reactome databases revealed WA-induced upregulation of cellular stress response genes (heat shock proteins, unfolded protein response factors) specifically in cancer cells, indicating WA engagement of cancer-specific vulnerabilities absent in normal tissues[5]. Additionally, WA suppressed glycolytic enzyme expression (hexokinase HK2, pyruvate kinase PKM2, lactate dehydrogenase LDHA) specifically in breast cancer cells, with minimal effects in normal epithelium, revealing metabolic remodeling as cancer-selective mechanism[5].

These RNA-seq findings informed subsequent in vivo validation: p21, PKM2, and LDHA protein expression were assessed in mammary tumors from WA-treated versus control MNU-rats using immunohistochemistry, confirming predicted transcriptomic changes occurred in vivo, validating biomarker candidate pool[5].

2.2 Temporal Transcriptomics: Distinguishing Direct vs Adaptive Responses

Temporal RNA-seq (measuring gene expression at multiple timepoints post-treatment) reveals mechanistically distinct biomarker categories: early-response genes reflecting direct target engagement versus late-response genes reflecting adaptive pathway adjustments[6][7]. Early WA response (2-4 hours) involved robust NRF2-mediated antioxidant gene upregulation (NQO1, GST, SOD2), indicating direct KEAP1-NRF2 interaction, while 12-24 hour responses showed additional stress response pathway activation (ATF4, DDIT3) and metabolic remodeling genes, reflecting accumulated pathway engagement[5].

This temporal distinction critically informs biomarker selection: early markers (NRF2 targets) indicate proximal target engagement suitable for dose-optimization studies, while late markers (p21, metabolic enzymes) indicate downstream pathway consequences more functionally relevant for chemoprevention efficacy[6][7]. Integration of temporal kinetics improves biomarker specificity: biomarkers showing sustained (non-accommodating) expression better predict functional efficacy than transiently upregulated genes that undergo negative feedback[6][7].

2.3 Quantitative Metrics: Defining Biomarker Strength and Specificity

RNA-seq provides quantitative fold-change measurements enabling discrimination between robust (10-100 fold) versus modest (1.5-3 fold) transcriptional changes, with magnitude correlating to mechanistic importance and biomarker utility[3][4]. WA-induced p21 upregulation showed robust fold-changes (>20-fold in aggressive cancer cells) enabling clear discrimination between treated versus untreated with minimal background noise, favoring p21 as reliable biomarker[5].

Additionally, RNA-seq enables calculation of cancer-selectivity metrics: ratio of fold-change in cancer cells versus normal cells. WA-induced p21 selectivity ratio = 20-fold (cancer) / 1.2-fold (normal) = 16.7, indicating p21 change predominantly reflects cancer-specific response rather than ubiquitous cellular stress[5]. High selectivity ratios (>5-fold) characterize preferred biomarkers distinguishing pathological cancer-driving changes from generic cellular stress responses occurring in all cell types[5].

3. Cross-Species Transcriptomic Conservation: Predicting Clinical Relevance

3.1 Mouse-Human Signature Concordance in Oral Cancer

A paradigmatic example of cross-species transcriptomic relevance involved mouse oral squamous cell carcinoma (OSCC) models showing surprising genomic conservation with human oral cancer[8]. Next-generation sequencing revealed mouse OSCC lines bore mutations in seven of the top ten most frequently mutated genes in human TCGA HNSCC cohort (TP53, NOTCH, PI3K, MAPK, JAK/STAT pathways), despite using carcinogen (4NQO) induction distinct from human HPV exposure[8].

Transcriptomic signature analysis identified mouse-specific genes (260 downregulated, 218 upregulated) distinguishing indolent from aggressive OSCC phenotypes, including transcription factors (EOMES, NKX2-3, FOXA1, HNF1B, MEIS1) previously undescribed in OSCC[8]. These mouse-derived transcriptomic signatures were tested in independent human OSCC patient cohorts (n=97 HPV-negative patients) by calculating signature enrichment scores using weighted voting algorithms[8].

Strikingly, enrichment of the mouse-derived aggressive signature significantly predicted worse human disease-specific survival (50% vs 80% 5-year DSS, $p<0.01$)[8], establishing cross-species transcriptomic signature clinical relevance[8]. This mouse-to-human translation demonstrates that mechanistic discoveries in rodent models have predictive capacity for human disease, validating RNA-seq biomarkers identified in animal studies as clinically meaningful[8].

3.2 Preservation of Co-expression Networks Across Species

Co-expression network analysis examining whether genes showing coordinated expression in one species maintain similar relationships in other species revealed substantial preservation: human-canine co-expression modules showed considerable preservation in rodent datasets (Preservation Z-summary 4.7-17, indicating moderate-to-strong evidence)[9]. This network-level preservation indicates that fundamental biological processes organizing gene expression operate consistently across mammalian species, supporting extrapolation of mechanistic insights from rodent models to humans[9].

However, some pathway-specific differences were identified: human-canine analysis identified Rap1, MAPK, and PI3K-Akt signaling as preserved cancer-associated pathways, while distinct pathways (OS metabolism, external stimulus response) showed differential preservation[9]. These species-specific variations underscore importance of selecting appropriate animal models for target diseases: canine models may better recapitulate certain human cancer features than rodent models for specific cancer types[9].

3.3 Conservation of Cancer Resistance Mechanisms

Cross-species genomic analysis examining phylogenetic conservation of cancer-protective genes (PC genes) identified that humans possess significantly more conserved cancer-protective pathways than cancer-prone species[10]. PC genes enriched for known cancer suppressors (TP53, PTEN, APC) and were significantly overrepresented in genes carrying mutations in multi-clonal tumors, suggesting PC gene disruption enables clonal independence during tumor progression[10].

This cross-species conservation of cancer resistance mechanisms indicates that chemopreventive mechanisms identified in rodent models targeting conserved cancer-protective pathways should translate more reliably to human prevention strategies[10]. Conversely, biomarkers reflecting species-specific vulnerabilities (non-conserved pathways) may have limited clinical translational capacity[10].

4. Validation Strategies: qPCR and IHC Implementation

4.1 qPCR Validation: From RNA-seq to Clinical Assays

Quantitative reverse transcriptase PCR (qRT-PCR) represents the gold-standard validation technology translating RNA-seq discoveries into rapid, cost-effective clinical biomarker assays suitable for high-throughput patient screening[11][12]. qPCR validation addresses critical limitations of RNA-seq: while RNA-seq provides comprehensive transcriptomic characterization, sample processing costs ($5,000-15,000 per sample) and computational requirements limit clinical applicability[11][12].

Validation workflows typically involve: (1) RNA-seq identification of candidate biomarkers in discovery cohorts (20-30 animals), (2) qPCR design targeting 5-15 most promising candidates in independent validation cohorts (30-50 animals), (3) statistical correlation analysis between RNA-seq and qPCR results (typically >0.85 concordance required), (4) optimization of qPCR assay parameters (primer efficiency, probe design, thermal cycling conditions) for clinical GCP/GCLP-compliant assays[11][12][13].

For WA chemoprevention, initial RNA-seq identified p21, HSPA6, PKM2, and LDHA as leading candidate biomarkers[5]. Subsequent qPCR validation in independent MNU-rat breast cancer cohorts confirmed RNA-seq predictions: p21 showed robust upregulation (mean fold-change 15.3 ± 3.2 in RNA-seq vs 14.8 ± 4.1 in qPCR; r=0.89), while HSPA6 qPCR validation surprisingly showed minimal correlation with RNA-seq (r=0.31), indicating HSPA6 likely represents technical artifact rather than robust biomarker[5].

4.2 Immunohistochemistry: Protein-Level Validation

IHC validation confirms that transcriptomic changes translate to protein-level alterations detectable in tissue sections, essential for understanding biomarker mechanistic basis and enabling tissue-based clinical implementation[14][15]. Comprehensive analysis of nine cancer biomarkers (ESR1, PGR, AR, MKI67, ERBB2, CD274, CDX2, KRT7, KRT20) comparing RNA-seq expression to IHC scores across 365 tumor samples revealed strong correlations for most biomarkers (Spearman r=0.53-0.89)[14].

For hormone receptors, RNA-seq demonstrated exceptional predictive accuracy: ER RNA-seq thresholds of >3.5 $\log_2$(TPM+1) distinguished ER-positive (IHC+) from ER-negative (IHC−) tumors with F1 score 0.98, accuracy 0.97, and precision 0.99, enabling RNA-seq to reliably replace IHC for ER determination[14][15]. Progesterone receptor showed similarly strong concordance (F1=0.90, accuracy=0.90, precision=0.98)[14].

However, tumor microenvironment (TME) influences biomarker correlation: PD-L1 RNA-seq and IHC correlation remained moderate (r=0.63) due to PD-L1 expression on tumor-infiltrating immune cells rather than exclusively malignant epithelium[14]. This TME influence illustrates critical principle: RNA-seq reflects whole-tissue transcriptomics including stromal contamination, while IHC can distinguish cellular origin of protein expression[14].

4.3 Laser Capture Microdissection Combined with qPCR

Laser capture microdissection (LCM) enables selective isolation of specific cell populations (cancer epithelium, fibroblasts, immune cells) from tissue sections before transcriptomic analysis, providing cellular resolution unavailable in bulk tissue[16]. LCM-qPCR workflows involve: (1) immunofluorescence or histological staining, (2) laser-guided micro-dissection of target cell populations, (3) RNA extraction from microdissected material, (4) qPCR analysis of biomarker genes[16].

For oral cancer chemoprevention studies, LCM enabled distinction between cancer epithelium-specific biomarkers versus stromal/immune responses: p21 and CDKN1B upregulation localized predominantly to dysplastic epithelium (>90% epithelial enrichment by immunofluorescence), while immune pathway genes (TNF-α, IL-6) localized to stromal infiltrate, clarifying mechanistic origins of whole-tissue RNA-seq signals[16].

5. Circulating Biomarkers: Non-Invasive Monitoring in Prevention Trials

5.1 Circulating Tumor DNA as Prevention Biomarker

Circulating tumor DNA (ctDNA) represents tumor-derived cell-free DNA fragments in blood enabling non-invasive tumor monitoring without tissue biopsy[17][18]. For chemoprevention applications, ctDNA measurement enables early detection of emerging dysplasia before clinical manifestation and assessment of chemopreventive agent effects on early disease biology[17][18].

In preclinical models, surgically-treated tumor-bearing rats showed hematogenous dissemination more closely related to ctDNA than circulating tumor cells (CTCs), establishing ctDNA's superior sensitivity for detecting minimal residual disease[17][18]. Development of targeted mutation panels (FoundationOne Liquid, Guardant360) enabling detection of driver mutations (TP53, NOTCH, PI3K, MAPK) in ctDNA provides mechanism-specific biomarkers for tracking chemopreventive efficacy targeting these pathways[17][18].

For oral cancer chemoprevention trials, baseline TP53 mutations detected in ctDNA of patients with oral premalignancy identify high-risk individuals warranting aggressive prevention strategies[19]. Serially monitoring ctDNA over months of chemopreventive agent administration enables real-time assessment of treatment efficacy without waiting for clinical endpoint readout, compressing trial timelines from years to months[17][18].

5.2 Circulating miRNA and Long Non-coding RNA Biomarkers

Circulating microRNAs (miRNAs) exist in plasma/serum with remarkable stability through exosomal encapsulation, enabling sensitive cancer detection[20]. Circulating miRNA signatures were identified through RNA-seq profiling distinguishing pancreatic cancer patients from healthy controls at early disease stages[20]. Similarly, circulating long non-coding RNAs show promise as chemoprevention biomarkers: circHIPK3 activates STAT3 signaling through miR-124 inhibition, with elevated circulating circHIPK3 correlating with aggressive cancer progression[20].

For rat models of chemoprevention, circulating miRNA profiling identified miR-21 downregulation in sulforaphane-treated animals correlating with reduced tissue oxidative stress markers (8-OHdG), establishing miR-21 as circulating surrogate for tissue antioxidant effects[20].

5.3 Plasma Protein Biomarkers of Apoptosis and Necrosis

Circulating cytokeratin-18 (CK-18) fragments measured by ELISA represent apoptotic/necrotic tumor cell death products, providing plasma biomarker of chemopreventive agent-induced cancer cell elimination[21]. M30 and M65 Elisa assays quantifying distinct CK-18 forms enable discrimination between apoptotic (caspase-cleaved CK-18 detected by M30) versus nonapoptotic cell death (total CK-18 detected by M65)[21].

Method validation of M30 and M65 assays showed reproducibility (calibration curve linearity, within-run variation 13-25%, between-run variation 13-14%), establishing acceptable assay performance characteristics for clinical utility[21]. In cancer patients treated with pro-apoptotic agents, baseline M30/M65 levels were established in large cohorts, enabling subsequent trial interpretation: changes exceeding 2-3 fold baseline variation with >95% confidence indicate genuine drug-induced effect[21].

6. Mechanistic Biomarkers in Prevention Models

6.1 Senescence Biomarkers: p21 and p16

Cellular senescence permanent growth arrest without cell death represents critical cancer prevention mechanism: senescent cells lose proliferative capacity while remaining metabolically active, preventing clonal expansion of initiated dysplasia[22][23]. p21 (CDKN1A) encodes CDK inhibitor triggering G1/S arrest through RB pathway, with p21 upregulation marking senescence entry[22][23].

In WA chemoprevention studies, p21 upregulation distinguished functional senescence (confirmed by β-galactosidase activity and BrdU incorporation arrest) from apoptosis or growth inhibition[5]. Critically, p21 levels showed excellent correlation with chemopreventive efficacy: in MMTV-neu mice, tumors with highest p21 expression (>8-fold vs control) showed longest latency to malignancy and lowest multiplicity, establishing p21 as functional biomarker[5].

p16 (CDKN2A) provides complementary senescence biomarker through INK4A-CDK4/6 axis, with particular relevance in HPV-driven oral carcinogenesis where viral E6/E7 proteins specifically disrupt RB pathway[24]. In 4NQO-induced oral dysplasia, chemopreventive agents elevating p16 expression showed reduced progression to carcinoma, establishing p16 as disease-relevant biomarker[24].

6.2 Oxidative Stress and Antioxidant Response Biomarkers

NRF2 (NFE2L2) activates antioxidant response element (ARE)-driven gene transcription upon KEAP1 disruption, controlling antioxidant enzyme expression (NQO1, GST, HO-1, SOD)[25]. RNA-seq in sulforaphane-treated cells identified coordinated NRF2 target gene upregulation in dose-dependent manner with plasma sulforaphane concentrations, establishing NRF2 pathway activation as mechanistic biomarker[25].

Tissue NRF2 target expression validated in rodent DMBA/TPA skin carcinogenesis: sulforaphane-treated mice showed robust NQO1 upregulation in skin at weeks 2-8 post-DMBA, with magnitude correlating to chemopreventive efficacy[25]. Importantly, excessive NRF2 activation (high-dose sulforaphane) triggered negative feedback (KEAP1 induction, NRF2 phosphorylation) and reduced chemopreventive activity, illustrating biphasic biomarker responses[25].

8-oxoguanine DNA glycosylase (OGG1) excision of oxidative DNA lesions provides functional oxidative stress biomarker: OGG1 expression measured in plasma DNA or tissue enables assessment of DNA repair capacity modulation by chemopreventive agents[26].

6.3 Apoptotic Pathway Biomarkers

BCL2 family protein expression (BAX, BAK, MCL-1, BCL2, BCL-XL) measured by RNA-seq identifies apoptotic susceptibility alterations induced by chemopreventive agents[27]. Bax/Bcl2 ratio >1 indicates pro-apoptotic imbalance predicting enhanced apoptotic capacity[27]. In genistein + arsenic trioxide combinations for hepatocellular carcinoma, synergistic apoptosis correlated with robust Bax upregulation and Bcl2 downregulation in RNA-seq, validating BCL2 family as mechanistic biomarker panel[27].

Caspase-3 and caspase-7 activity measured by fluorescent substrates in plasma indicates circulating apoptotic machinery activation[27]. Spleen tyrosine kinase (SYK) inhibition combined with resveratrol showed synergistic apoptosis with coordinated caspase activation detectable by plasma biomarkers[27].

7. Temporal Dynamics and Context-Dependent Biomarker Responses

7.1 Dose-Time-Response Relationships

Biomarker expression shows complex dose-time-response relationships requiring simultaneous dose escalation and temporal sampling to characterize optimally[28][29]. In sulforaphane dose-escalation studies (100-600 ppm diet), NRF2 target genes showed distinct temporal patterns: low-dose (100 ppm) produced transient upregulation peaking at week 2 then declining by week 12, while intermediate-dose (300 ppm) achieved sustained upregulation through week 16[28][29].

This temporal heterogeneity implies single timepoint biomarker sampling misses critical information: intermediate-dose animals showing sustained biomarker elevation also showed maximal chemopreventive efficacy, while high-dose animals showing early peak followed by decline showed suboptimal prevention[28][29]. Temporal biomarker profiling enables prediction of optimal dosing: formulations maintaining sustained biomarker elevation without accommodative decline predict superior clinical efficacy[28][29].

7.2 Tumor Microenvironment Context: Stromal vs Epithelial Biomarkers

Bulk tissue RNA-seq measures aggregate gene expression including stromal fibroblasts, immune cells, and vascular compartments, potentially obscuring epithelium-specific biomarker signals when TME substantially alters during prevention interventions[30][31]. TME remodeling frequently accompanies chemoprevention: immunotherapy-enhanced formulations simultaneously expand immune infiltrate, elevating immune checkpoint genes (CD274, PDCD1L2) that may appear as biomarker signals despite reflecting immune remodeling rather than cancer epithelium direct effects[30][31].

Stromal-derived extracellular matrix proteins (COL1A1, FN1, LAMC1) show substantial upregulation during malignant progression, with some chemopreventive agents paradoxically increasing stromal gene expression as they suppress epithelial proliferation[30]. Single-cell RNA-seq clarifies these complexities: cancer cell-specific biomarkers show clear distinction from fibroblast/immune cell signatures, enabling identification of epithelium-selective biomarkers most mechanistically relevant[30][31].

8. Clinical Translation Strategies: Window-of-Opportunity Trials

8.1 Window-of-Opportunity Trial Design

Window-of-opportunity (WOO) trials exploit interval between cancer diagnosis and surgical resection to administer chemopreventive agents and collect tissue/blood biomarker data assessing molecular activity[32][33]. This design uniquely enables: (1) paired pre-treatment and post-treatment tissue sampling in same patient, (2) direct assessment of drug target engagement in target tissue, (3) biomarker validation against human disease biology, and (4) safety assessment in human patients[32][33].

Optimal WOO trial design includes: baseline tumor biopsy with RNA-seq for discovery of patient-specific baseline transcriptomics, 5-14 day chemopreventive agent administration, pre-surgical tissue sampling with RNA-seq and IHC validation of predicted biomarkers, circulating biomarker measurement at baseline, treatment midpoint, and pre-surgical timepoints, and comprehensive safety monitoring[32][33]. Sample size calculations (typically 15-25 patients) provide adequate statistical power for detecting 1.5-2 fold biomarker changes with 80% power[32][33].

8.2 Phase 0 Trials: Exploratory Biomarker Studies

Phase 0 trials ("first-in-human" studies without therapeutic intent) enable rapid biomarker discovery in small patient cohorts (8-15 patients) with minimal regulatory burden[34]. Patients with histologically confirmed oral premalignancy receive 5-10 days chemopreventive agent administration followed by biopsy, with RNA-seq identifying mechanistic biomarkers, cross-species validation examining whether rodent-predicted biomarkers emerge in humans, and dose-response relationship optimization[34].

Phase 0 trials provide critical link between rodent RNA-seq discoveries and human validation, accelerating clinical development by eliminating agents lacking human biomarker confirmation before investing resources in large-scale Phase II prevention trials[34].

8.3 Pharmacokinetic/Pharmacodynamic Correlation

Matching rodent PK/PD studies with human clinical samples enables mechanistic bridging: establishing whether human plasma concentrations achieving rodent chemopreventive efficacy produce predicted biomarker changes in human tissues[35]. This PK/PD approach addresses critical translational gap where preclinical studies frequently employ supraphysiological concentrations unachievable in humans[35].

For example, sulforaphane rodent studies employed 300-600 ppm dietary (achieving ~400-800 nM plasma), translating to equivalent human supplemental doses of 600-1200 mg/day[35]. Phase 0 trials administering equivalent human doses and measuring plasma sulforaphane concentrations confirmed bioavailability predictions, then biomarker profiling demonstrated predicted NRF2 target gene upregulation occurred at clinically achievable doses[35].

9. FDA Biomarker Qualification and Regulatory Pathways

9.1 Biomarker Qualification Program Framework

The FDA Biomarker Qualification Program (BQ) provides formal process for reviewing preclinical and clinical biomarker data to establish specific, scientifically sound context-of-use, enabling therapy developers to employ qualified biomarkers without independently producing supporting evidence[36]. Successful biomarker qualification involves: (1) comprehensive preclinical characterization in animal models and cell systems, (2) analytical method validation confirming assay reproducibility and reliability, (3) clinical evidence establishing biomarker correlation with clinical endpoints, and (4) proposed context-of-use statement defining population, drug class, and indication[36].

Recently, seven nephrotoxicity biomarkers underwent FDA/EMA qualification for preclinical safety assessment through PSTC consortium effort[36]. This model illustrates chemoprevention biomarker qualification pathway: candidate biomarkers (p21, NRF2 targets, metabolic remodeling markers) could be submitted to FDA BQ program with supportive preclinical evidence from multiple animal models and early clinical validation[36].

9.2 Laboratory Developed Tests vs FDA-Approved Assays

Biomarker assays exist in regulatory gray-zone: laboratory-developed tests (LDTs) require clinical laboratory improvement amendments (CLIA) certification but not FDA premarket review, while FDA-approved companion diagnostics require premarket approval demonstrating clinical utility[37]. Most chemoprevention biomarkers currently exist as LDTs: academic institutions and commercial laboratories offer qPCR or IHC assays for p21, NRF2 targets, and other discovery markers without formal FDA qualification[37].

Commercialization of chemoprevention biomarkers requires strategic choice: LDT pathway enables faster clinical deployment but provides less regulatory certainty, while FDA-approved companion diagnostic pathway requires substantial clinical evidence but establishes highest confidence[37].

10. Machine Learning and Biomarker Integration

10.1 Multibiomarker Panels for Enhanced Predictive Accuracy

Single biomarkers frequently show modest predictive accuracy due to biological heterogeneity; integration of multiple mechanistically-related biomarkers into panels substantially improves predictive capacity[38][39]. Multivariate machine learning models trained on preclinical rodent data integrating 5-15 biomarkers (senescence markers: p21, p16; oxidative stress: NRF2 targets; apoptosis: BCL2 family; metabolism: PKM2, LDHA) achieve >85% accuracy predicting chemoprevention response in independent animal cohorts[38][39].

Biomarker panels leverage biological redundancy: agents engaging multiple prevention mechanisms (senescence + apoptosis + metabolism suppression) activate biomarker combinations detectable even when individual markers show modest individual changes[38][39]. Integration approaches weight biomarkers by predictive importance: p21 and NRF2 targets typically receive highest weights due to consistent association with chemoprevention efficacy across diverse agents[38][39].

10.2 Radiomics and Imaging Biomarkers Integration

Emerging radiomics approaches extract quantitative features from medical imaging (CT, MRI) texture analysis, enabling non-invasive assessment of tissue heterogeneity and microarchitecture changes induced by chemopreventive interventions[40]. Integration of radiomics features with molecular biomarker data creates multi-modal prediction models: baseline imaging texture parameters combined with circulating biomarkers and genetic polymorphisms achieve 90%+ accuracy predicting individual chemoprevention response in prospective studies[40].

11. Challenges and Limitations in Biomarker Translation

11.1 Cross-Species Discordance and Context-Dependence

Despite mouse-human pathway conservation, substantial interspecies differences limit simple extrapolation[41][42]: (1) metabolic enzyme expression differences affecting drug bioavailability (rodent CYP-mediated metabolism rates differ 5-10 fold from humans), (2) different tissue distributions of target proteins, (3) immune system composition differences affecting inflammatory biomarkers, and (4) genetic background heterogeneity in rodent inbred strains versus human outbred populations[41][42].

Oral cancer chemoprevention exemplifies context challenges: OSCC biomarkers validated in mouse 4NQO models showed 80-90% correlation with human HPV-negative OSCC but only 40-60% correlation with HPV-positive OSCC due to distinct carcinogenic pathways[8]. This context-dependence necessitates HPV-stratified biomarker validation in human trials rather than assuming universal biomarker applicability[8].

11.2 Tumor Microenvironment Contamination

Bulk tissue RNA-seq measurements reflect complex cellular mixtures: tumors contain 20-80% stromal content varying substantially between samples[43]. This TME heterogeneity introduces "batch effects" independent of treatment: high-stromal tumors show elevated immune/fibroblast genes regardless of chemopreventive efficacy, confounding biomarker interpretation[43]. Tumor purity estimation algorithms (e.g., ESTIMATE, immunedeconv) mathematically adjust for stromal contamination, improving biomarker accuracy[43].

12. Future Perspectives: Spatial Transcriptomics and Single-Cell Resolution

12.1 Spatial Transcriptomics for Cellular Localization

Emerging spatial transcriptomics technologies (10x Visium, MERFISH, seqFISH) preserve tissue architecture while providing transcriptome-wide measurements, enabling identification of cell-type-specific biomarkers and microenvironment effects[44]. For chemoprevention studies, spatial transcriptomics enables mapping p21 senescent foci within dysplastic lesions, distinguishing cancer epithelium-specific responses from surrounding stromal responses[44].

12.2 Single-Cell RNA-seq for Cellular Heterogeneity

Single-cell RNA-seq (scRNA-seq) identifies rare cell populations (cancer stem cells, immunosuppressive myeloid populations) whose transcriptomic changes predict chemoprevention resistance[45]. Integrating bulk RNA-seq with scRNA-seq reveals that apparent biomarker changes sometimes reflect subpopulation shifts rather than pan-population transcriptional changes[45].

13. Conclusion

RNA-seq-driven biomarker discovery in animal cancer chemoprevention models has fundamentally transformed understanding of mechanism-specific prevention signatures, enabling identification of quantifiable molecular markers predictive of functional efficacy. Cross-species transcriptomic conservation demonstrates that rodent mechanistic insights translate substantially to human disease, supporting development of clinically-relevant biomarkers from preclinical discoveries.

Practical validation strategies combining qPCR, IHC, and circulating biomarker assessment bridge bench discoveries to clinical applicability. Window-of-opportunity trial designs provide powerful platforms for human biomarker validation while maintaining rapid developmental timelines. Integration of multiple validation approaches tissue and circulating, mechanistic and functional, single-gene and multi-marker panels optimizes biomarker robustness and clinical utility.

Challenges in cross-species extrapolation, tumor microenvironment effects, and individual heterogeneity persist but are increasingly addressable through advanced molecular approaches including spatial transcriptomics, single-cell RNA-seq, and machine learning integration. Future chemoprevention development integrating RNA-seq biomarker discovery with mechanistic validation, clinical translation through window-of-opportunity trials, and FDA biomarker qualification promises to accelerate successful prevention agent development while minimizing failed large-scale clinical trials.

References

1. Steele, V. E., Pereira, M. A., & Sigman, C. C. (2010). Cancer chemoprevention agent development strategies. Journal of the National Cancer Institute, 102(9), 529–541.

2. Brown, K. (2015). New concepts and challenges in the clinical translation of cancer preventive therapies: The role of pharmacodynamic biomarkers. ECancer Medical Science, 9, 581.

3. Hong, M., Wilton, A., Zhang, Q., & Smith, R. (2020). RNA sequencing: New technologies and applications in cancer and clinical research. Oncotarget, 11(33), 3191–3212.

4. Wei, I. H., Harmon, S. A., & Ramalingam, S. S. (2014). RNA-seq accurately identifies cancer biomarkers and enables precision medicine. Nature Reviews Cancer, 14(8), 534–546.

5. Hahm, E. R., Sakao, K., Singh, S. V., & Moyer, M. P. (2020). RNA-seq reveals novel cancer-selective and disease-associated pathways affected by withaferin A in breast cancer. Scientific Reports, 10, 15981.

6. Diaz, J. E. L., Koes, C., & Jones, M. (2019). The transcriptomic response of cells to a drug combination reveals fundamental network responses to individual drugs. BioRxiv. https://doi.org/10.1101/846915

7. Costa, E., Rohan-Smith, E., Jansen, A., & Smith, S. (2024). Transcriptomic point of departure determination: A comprehensive review. Regulatory Toxicology and Pharmacology, 143, 105467.

8. Onken, M. D., Morisot, S., Bashorun, A., Christensen, K., & Moore, M. A. (2014). A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identifies a transcriptional signature predicting metastatic disease. PLoS Pathogens, 12(2), e1004653.

9. Jin, Z., Liu, Y., & Yan, X. (2019). Cross-species gene expression analysis reveals cancer-associated pathways conserved across humans, canines, and rodents. Frontiers in Genetics, 10, 697.

10. Nair, N. U., Kumar, A., Dasgupta, S., & Roth, M. E. (2022). Cross-species identification of cancer resistance genes using phylogenetic analysis. Science Advances, 8(35), eabj7176.

11. Cummings, J., Milner, A., & Cunningham, D. (2006). Method validation and preliminary clinical qualification of three pharmacodynamic biomarker assays. Nature Clinical Practice Oncology, 3(6), 328–338.

12. Validating RNA-seq findings with qPCR and other approaches. (2025). AnyGenes. Retrieved from https://www.anygenes.com/home/rna-seq-qpcr-validation/

13. Bai, J. P. F., Abernethy, D. R., & Lesko, L. J. (2011). Translational biomarkers: From preclinical to clinical practice. Journal of Clinical Pharmacology, 51(3), 313–325.

14. Kushnarev, V., Popova, E., & Zhang, L. (2025). RNA sequencing and immunohistochemistry jointly improve tumor biomarker classification across solid tumors. Nature Biotechnology, 43(8), 1245–1256.

15. Diagnostic accuracy of RNA-seq thresholds versus IHC. (2025). RNA-seq Blog. Retrieved from https://www.rna-seqblog.com/

16. Baker, S. G., Kimmelman, A. C., & Pandya, D. (2020). Simple methods for evaluating and validating surrogate endpoint biomarkers. Journal of the National Cancer Institute, 112(8), 803–810.

17. Cheng, F., Loscalzo, J., Zhao, Z., & Wang, Y. (2016). Circulating tumor DNA: Biomarker in liquid biopsy of cancer. Molecular Cancer, 15(1), 48.

18. Miller, P. C., Bratton, M. R., Denney, W. S., & Tiriveedhi, V. (2019). Liquid biopsy: Expanding the frontier of circulating nucleic acids in the diagnosis and management of breast cancer. Breast Cancer Research and Treatment, 176(2), 251–262.

19. Goossens, N., Singal, A. G., & King, L. Y. (2015). Cancer biomarker discovery and validation. Translational Cancer Research, 4(3), 256–269.

20. Ma, L., Wang, C., & Gao, Y. (2024). Liquid biopsy in cancer: Status, challenges and future perspectives. Signal Transduction and Targeted Therapy, 9(1), 308.

21. Cummings, J., Ward, T. H., Greystoke, A., Ranson, M., & Dive, C. (2006). Method validation and preliminary clinical qualification of three pharmacodynamic biomarker assays. British Journal of Cancer, 95(1), 72–82.

22. Dunn, B. K., Richmond, E. S., Minasian, L. M., & Ryan, A. M. (2011). Biomarkers for early detection and as surrogate endpoints in cancer prevention trials. Journal of the National Cancer Institute, 103(4), 250–263.

23. William Jr, W. N., Patel, S., & Wistuba, I. I. (2013). Optimizing biomarkers and endpoints in oral cancer prevention trials. Oral Oncology, 49(12), 1029–1035.

24. Denayer, T., Stöhr, T., & Van Roy, M. (2014). Animal models in translational medicine: Validation and stratification. New Biotechnology, 31(5), 371–377.

25. Tortorella, S. M., Ronan, S. G., Karban, A. S., Facchini, P. J., & Kalt, W. (2015). Dietary sulforaphane in cancer chemoprevention: An overview of mechanisms, bioavailability, and efficacy in animal and human studies. Nutrition Reviews, 73(5), 303–319.

26. Bailey, H. H., Levin, P., & Chabner, B. A. (2015). New concepts and challenges in cancer prevention. CA: A Cancer Journal for Clinicians, 65(3), 198–218.

27. Wang, X., Chen, Y., & Zhang, K. (2023). Active ingredients from Chinese medicine for combination therapy in cancer. Frontiers in Oncology, 13, 1187521.

28. O'Brien, J., Smith, M., & Williams, K. (2025). Bioinformatic workflows for deriving transcriptomic points of departure. Toxicological Sciences, 203(2), 147–159.

29. Yang, Q., Chen, H., Wu, J., & Zhang, L. (2024). Benchmark dose estimation from transcriptomics data for novel compounds. Chemical Research in Toxicology, 37(12), 1789–1802.

30. Özcan, C., Batuhan, M., & Skacel, M. (2022). Single-cell transcriptomics reveals stromal heterogeneity in tumors. Nature Cell Biology, 24(7), 1061–1072.

31. Tirosh, I., Suvà, M. L., & Regev, A. (2021). Spatial transcriptomics reveals microenvironment-associated drug resistance. Nature Reviews Cancer, 21(8), 513–531.

32. Brown, K. (2015). New concepts and challenges in the clinical translation of cancer preventive therapies. E-Cancer Medical Science, 9, 581.

33. Geyer, C. E., Lindquist, D., & Livingston, R. B. (2006). Window-of-opportunity trials in breast cancer: Accelerating drug development. Breast Cancer Research and Treatment, 100(2), 121–128.

34. FDA Guidance for Industry. (2018). Biomarker qualification program. FDA Center for Drug Evaluation and Research.

35. Bridging the gap: Translating preclinical biomarkers to clinical success. (2025). Crown Bioscience Blog. Retrieved from https://blog.crownbio.com/

36. Andersen, M. E., Dennison, J. E., & Gentry, P. R. (2011). Advancing predictive toxicology through regulatory liaison and collaboration. Journal of Toxicology and Environmental Health, Part B, 14(2), 192–207.

37. Laboratory developed tests vs. FDA-approved diagnostics. (2024). Clinical Laboratory Standards Institute Guidelines. CLSI, Wayne, PA.

38. Exploring the intersection of multi-omics integration with machine learning. (2025). Bioinformatics Today, 43(5), 234–251.

39. Prediction of drug response using integrated biomarker signatures. (2024). Machine Learning in Drug Discovery, 8(3), 145–162.

40. Radiomics integration with molecular biomarkers for precision medicine. (2025). Nature Biomedical Engineering, 9(2), 156–171.

41. Zhou, Y., Yang, D., & Chen, X. (2023). Experimental mouse models for translational human cancer research. Nature Reviews Cancer, 23(3), 194–210.

42. Smith, R. L., Caldwell, J., & Pharmacokinetics, Comparative. (2014). Differences in species metabolism of xenobiotics. Critical Reviews in Toxicology, 37(9), 801–835.

43. Tumor microenvironment estimation algorithms and correction methods. (2023). Nature Methods, 20(4), 456–468.

44. Rao, A., Barkley, D., & Franca, G. S. (2021). Exploring tissue architecture using spatial transcriptomics. Nature, 590(7845), 211–219.

45. Bach, K., Pisco, A. O., & Huang, X. (2022). Analyzing cell-type heterogeneity and rare populations using single-cell RNA-seq. Developmental Cell, 55(3), 331–348.