Transcriptomic Signatures of Chemoprevention: RNA-seq Insights from Carcinogen-Induced 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.18734796**
Received: 18 January 2026
Accepted: 27 January 2026

Abstract

Cancer chemoprevention represents a critical strategy for reducing cancer incidence through the prevention, delay, or suppression of tumor development using bioactive agents. High-throughput RNA sequencing (RNA-seq) technology has revolutionized our ability to characterize transcriptomic signatures associated with carcinogen-induced transformation and chemoprevention in animal models. This review synthesizes current knowledge on how RNA-seq-derived transcriptomic profiling from carcinogen-induced animal models reveals molecular mechanisms of chemoprevention, identifies predictive biomarkers, and enables the discovery of novel therapeutic targets. We examine established animal models including the DMBA/TPA mouse skin carcinogenesis model, 4NQO-induced oral carcinogenesis, and benzo(a)pyrene exposure systems, alongside state-of-the-art RNA-seq methodologies for differential gene expression analysis, pathway enrichment, and functional annotation. Particular emphasis is placed on phytochemical chemopreventive agents including sulforaphane, curcumin, and ursolic acid, their mechanisms of action at the transcriptomic level, and the identification of key molecular pathways including NRF2-ARE antioxidant signaling, NF-κB inflammatory pathways, MAPK/ERK cascades, and apoptotic regulatory mechanisms. We discuss advanced computational approaches including weighted gene co-expression network analysis (WGCNA), machine learning classifiers, and single-cell RNA-seq for unraveling tumor microenvironment dynamics. Finally, we address challenges in translating transcriptomic discoveries from animal models to clinical biomarker development, quality control considerations, and future perspectives for precision cancer prevention.

Keywords: RNA-seq, chemoprevention, carcinogen-induced animal models, transcriptomic signatures, biomarkers, molecular pathways, phytochemicals

1. Introduction

Cancer development is a multistep process involving the progressive accumulation of genetic and epigenetic alterations that transform normal cells into malignant entities (Tan & Jimeno, 2011; Kemp, 2015; Kotecha et al., 2016). The hallmark features of cancer including self-sufficient growth signaling, evasion of growth inhibition, resistance to apoptosis, unlimited replicative potential, sustained angiogenesis, and tissue invasion emerge through disruption of critical cellular pathways governing proliferation, differentiation, and cell death (Kotecha et al., 2016). While conventional cancer treatment focuses on eliminating established tumors, cancer chemoprevention adopts a fundamentally different paradigm: preventing, delaying, or suppressing tumor initiation and progression through intervention at early disease stages (Gu et al., 2020; Swetha & Ramakrishna, 2022).

The concept of cancer chemoprevention has gained substantial momentum following the FDA approval of tamoxifen and raloxifene for breast cancer prevention (Lee et al., 2020), as well as emerging evidence supporting the role of nonsteroidal anti-inflammatory drugs (NSAIDs) in colorectal cancer reduction. However, the identification of new chemopreventive agents and the determination of their molecular mechanisms remain constrained by incomplete understanding of the transcriptional rewiring that accompanies carcinogenic transformation and its reversal by chemopreventive interventions.

The advent of RNA sequencing (RNA-seq) technology has fundamentally transformed cancer research by providing an unbiased, genome-wide assessment of transcriptional changes with single-nucleotide resolution (Goeman & Naccari, 2016). Unlike microarray-based approaches, RNA-seq offers superior sensitivity for detecting novel transcripts, splice variants, and non-coding RNAs, while providing dynamic quantification of gene expression across entire transcriptomes (Hong et al., 2020). This comprehensive transcriptomic profiling capability, when coupled with carefully designed carcinogen-induced animal models, enables the identification of cancer-specific molecular signatures, validation of mechanistic pathways, and discovery of novel biomarkers predictive of chemoprevention efficacy (Tan & Jimeno, 2011; Gu et al., 2020; Swetha & Ramakrishna, 2022).

Carcinogen-induced animal models including DMBA/TPA-induced skin carcinogenesis, 4NQO-induced oral carcinogenesis, and benzo(a)pyrene exposure systems provide experimentally tractable systems for studying the molecular basis of transformation and chemopreventive intervention (Kemp, 2015; Abel et al., 2009). These models offer critical advantages including temporal control of carcinogen exposure, histopathologically verified disease progression, and accessibility to tissues at sequential stages of transformation. When paired with RNA-seq analysis, these models reveal dynamic gene expression changes characterizing initiation, promotion, and progression phases of carcinogenesis (Foy et al., 2016).

This review synthesizes current knowledge on transcriptomic signatures of chemoprevention derived from carcinogen-induced animal models, integrating mechanistic insights, analytical methodologies, and translational implications. We examine how phytochemical and synthetic chemopreventive agents modulate gene expression networks at critical nodes of carcinogenic pathways, discuss advanced computational approaches for extracting biological meaning from high-dimensional transcriptomic data, and address key challenges in translating animal-derived findings to clinical biomarker discovery and precision cancer prevention.

2. Carcinogen-Induced Animal Models: Molecular Basis and Transcriptomic Characterization

2.1 DMBA/TPA Two-Stage Skin Carcinogenesis Model

The DMBA/TPA (7,12-dimethylbenz[a]anthracene/12-O-tetradecanoylphorbol-13-acetate) two-stage mouse skin carcinogenesis model represents one of the most extensively characterized chemical carcinogenesis systems (Abel et al., 2009; Kemp, 2015). This protocol involves a single topical application of DMBA to initiate DNA damage and transformation, followed by repeated TPA applications that promote proliferation and malignant progression, eventually leading to papilloma and squamous cell carcinoma (SCC) development (Li et al., 2021).

DMBA acts as a complete carcinogen by requiring metabolic activation through cytochrome P450 (CYP) enzymes, particularly CYP1A1 and CYP1B1, which convert the parent compound into electrophilic metabolites capable of forming bulky DNA adducts (Kemp, 2015; Gelhaus et al., 2011). These adducts, if not repaired by nucleotide excision repair mechanisms, lead to mutation accumulation (Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015). Remarkably, DMBA/TPA-induced tumors exhibit mutations in Hras codon 61 and p53 genes with remarkable consistency, making this model exceptionally valuable for studying the functional consequences of these common human cancer mutations (Abel et al., 2009).

Transcriptomic profiling of DMBA/TPA carcinogenesis across temporal stages reveals dynamic rewiring of gene expression networks (Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015; Kemp, 2015). Early transcriptional responses, occurring within hours of DMBA exposure, involve upregulation of xenobiotic metabolism genes (Cyp2a5, Aldh1a3) and oxidative stress response pathways, reflecting the cellular recognition of electrophilic insult (Loss of endogenous Nfatc1 reduces the rate of

DMBA/TPA-induced skin tumorigenesis, 2015). Progressive stages demonstrate sequential activation of inflammatory pathways, particularly NF-κB signaling, followed by sustained changes in cell cycle regulation genes (CDK1, CDKN2C, CDC20) (Tan & Jimeno, 2011).

TPA-induced promotion independently triggers transcriptional changes distinct from DMBA initiation, including pronounced activation of MAPK/ERK signaling cascades and enhanced expression of cyclins and cyclin-dependent kinases (Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015; Foy et al., 2016). The accessibility of tissue for non-invasive monitoring and straightforward application of topical agents makes DMBA/TPA carcinogenesis particularly suitable for evaluating chemopreventive compounds that can be applied directly to skin (Li et al., 2021; Kemp, 2015).

2.2 4NQO-Induced Oral Carcinogenesis Model

Four-nitroquinoline-1-oxide (4NQO) carcinogenesis represents the predominant murine model for studying molecular mechanisms of oral cancer development, producing histopathologically authentic oral squamous cell carcinoma (OSCC) with high frequency and reproducibility (Lee et al., 2023; Foy et al., 2016). Unlike DMBA/TPA, which relies on topical application to skin, 4NQO is administered orally, making this model particularly relevant for understanding carcinogenesis in the oral epithelium and adjacent structures.

4NQO requires metabolic activation, and its mechanism parallels tobacco-induced carcinogenesis through generation of DNA adducts and reactive oxygen species (Lee et al., 2023). Chronic 4NQO administration produces a predictable sequence of histopathological changes: hyperplasia within 2-4 weeks, dysplasia at 8-16 weeks, and invasive carcinoma by 20-24 weeks, recapitulating the multi-step progression observed in human oral cancer (Lee et al., 2023; Foy et al., 2016).

RNA-seq analysis of 4NQO-induced tissues reveals critical temporal dynamics in transcriptomic alterations (Foy et al., 2016). Early gene expression signatures (EGS), corresponding to hyperplastic and early dysplastic stages, demonstrate significant enrichment of NF-κB pathway genes and altered immune infiltration signatures (Foy et al., 2016). Progressive gene expression signatures (PCS), observed at late dysplastic and carcinoma stages, show activation of MAPK/ERK signaling and altered expression of genes controlling cell-cell adhesion, epithelial-mesenchymal transition (EMT), and invasion (Foy et al., 2016).

Genomic analysis of 4NQO-induced tumors reveals recurrent mutations in Trp53, Notch1, Fat1, and Pik3ca, creating molecular parallels with human HNSCC (Lee et al., 2023). Integration of genomic and transcriptomic data demonstrates that gene expression changes at early stages predict sensitivity to downstream inhibitors of the MAPK pathway, providing mechanistic rationale for precision prevention strategies (Foy et al., 2016).

2.3 Benzo(a)pyrene (BaP) Exposure Models

Benzo(a)pyrene, a prototype polycyclic aromatic hydrocarbon (PAH), induces carcinogenesis through a well-characterized metabolic pathway requiring activation by CYP1A1/1B1 to the ultimate carcinogen BaP-7,8-diol-9,10-epoxide (BaPDE) (Gelhaus et al., 2011; DNA adduct formation in DNA repair-deficient mice, 2012; Tung et al., 2014). Systemic BaP exposure produces DNA adducts predominantly in the esophagus, liver, and lung, with tissue-specific patterns reflecting differential CYP expression and DNA repair capacity (DNA adduct formation in DNA repair-deficient mice, 2012; Tung et al., 2014).

The formation of BaP-DNA adducts, particularly the tumorigenic N2-deoxyguanosine lesion (BPdG), leads to transversion mutations and p53 inactivation through covalent modification of DNA-binding residues (Gelhaus et al., 2011). Importantly, BaP-induced carcinogenesis demonstrates pronounced genotype dependence: Xpa−/−p53+/− mice deficient in nucleotide excision repair and haploinsufficient for p53 develop tumors at accelerated rates with higher burden of DNA adducts (DNA adduct formation in DNA repair-deficient mice, 2012), illustrating the critical interplay between genetic susceptibility and environmental carcinogen exposure.

Transcriptomic studies of BaP-exposed tissues reveal dynamic regulation of DNA damage response pathways, with significant upregulation of both homologous recombination and non-homologous end-joining repair genes (Tung et al., 2014). However, excessive DNA repair pathway activation paradoxically increases carcinogenic potential through facilitation of aberrant recombination events (Tung et al., 2014), underscoring the nuanced role of stress response pathways in transformation.

3. Molecular Basis of Carcinogenesis: Transcriptomic Mechanisms

3.1 Multistage Carcinogenesis and Phase-Specific Gene Expression

The classical three-stage model of carcinogenesis initiation, promotion, and progression finds molecular expression in phase-specific transcriptomic signatures (Kemp, 2015; Kotecha et al., 2016; Liu et al., 2015). Initiation by carcinogens generates DNA damage, which, if unrepaired, becomes fixed as mutations in critical proto-oncogenes and tumor suppressors. Promotion involves selection and clonal expansion of initiated cells through proliferative signals and suppression of apoptosis. Progression encompasses further oncogenic alterations and loss of differentiation capacity, culminating in invasive and metastatic disease (Kemp, 2015; Liu et al., 2015).

RNA-seq profiling across carcinogenesis stages reveals coordinated transcriptional rewiring at each phase (Tan & Jimeno, 2011; Foy et al., 2016). Early initiation involves transient activation of p53-dependent stress responses (TP53, BAX, CDKN1A) and xenobiotic metabolism genes (CYPs, GSTs, NQO1) (Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015; Lan et al., 2016). If cells survive this checkpoint, transition to promotion involves downregulation of p53 activity or acquisition of p53 mutations, coupled with sustained activation of growth signals (MYC, cyclin family members) (Tan & Jimeno, 2011; Foy et al., 2016).

The promotion-progression transition demonstrates enhanced complexity, with accumulation of additional driver mutations and loss of tumor suppressor function (RB, p16, p27) (Abel et al., 2009). Late-stage progression exhibits dramatic transcriptomic reorganization including marked downregulation of differentiation genes, upregulation of EMT factors (SNAIL, SLUG, TWIST), and activation of angiogenic programs (Foy et al., 2016).

3.2 Critical Oncogenic Pathways in Carcinogen-Induced Transformation

3.2.1 KRAS and p53 Mutations

KRAS (Kirsten rat sarcoma viral oncogene homolog) activation represents one of the most frequent oncogenic alterations in human cancers and is consistently found in DMBA/TPA-induced skin tumors (Abel et al., 2009; Wang et al., 2023; Rachmawati et al., 2019). KRAS mutations, particularly at codon 61 (A→T transversion characteristic of DMBA mutagenesis), activate downstream RAF/MEK/ERK signaling cascades (Wang et al., 2023; Kealey et al., 2022). At the transcriptomic level, KRAS activation promotes expression of proliferation genes (cyclins, CDKs, E2F targets) while suppressing differentiation and apoptotic pathways (Wang et al., 2023; Kealey et al., 2022).

p53, the "guardian of the genome," undergoes mutation in DMBA/TPA tumors through formation of BaP-DNA adducts at critical DNA-binding domain residues (Gelhaus et al., 2011). Unlike wild-type p53, which activates stress response genes (BAX, CDKN1A, PUMA), mutant p53 loses transcriptional capacity for apoptotic targets while acquiring gain-of-function properties that promote proliferation and metastasis (Wang et al., 2023; Rachmawati et al., 2019; Kealey et al., 2022). The p53-KRAS co-mutation significantly enhances transcriptional signatures of EMT, angiogenesis, and metabolic reprogramming (Wang et al., 2023; Kealey et al., 2022).

3.2.2 Cell Cycle Regulatory Networks

Dysregulation of cell cycle checkpoints represents a hallmark of carcinogenesis, prominently reflected in transcriptomic changes during DMBA/TPA promotion (Tan & Jimeno, 2011). Cyclin-CDK complexes (Cyclin E-CDK2, Cyclin A-CDK2, Cyclin B-CDK1) exhibit progressive upregulation across promotion and early progression phases (Tan & Jimeno, 2011; Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced

skin tumorigenesis, 2015). Concurrently, CDK inhibitors (p16/CDKN2A, p21/CDKN1A, p27/CDKN1B) are downregulated through both transcriptional suppression and p16 promoter hypermethylation (Chou et al., 2024; Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015).

Retinoblastoma (RB) pathway inactivation, either through RB mutations or enhanced CDK-mediated RB phosphorylation, releases E2F transcription factors, leading to amplification of proliferation gene expression (Tan & Jimeno, 2011). RNA-seq reveals coordinated upregulation of E2F target genes (DNA polymerase α, thymidine kinase, cyclin A, cyclin E) as carcinogenesis progresses (Tan & Jimeno, 2011; Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015; Foy et al., 2016).

4. RNA-seq Methodologies for Transcriptomic Analysis of Carcinogenesis

4.1 Sample Preparation and Quality Control

Accurate transcriptomic profiling depends critically on high-quality RNA extraction and preservation of RNA integrity (Clinical value of RNA sequencing, 2018). Tissue procurement protocols must minimize post-harvest degradation by immediate snap-freezing in liquid nitrogen or preservation in RNAlater solution (Clinical value of RNA sequencing, 2018). For formalin-fixed, paraffin-embedded (FFPE) tissues, deparaffinization, rehydration, and proteinase K-mediated cross-link reversal are essential preprocessing steps (Clinical value of RNA sequencing, 2018).

RNA quality assessment utilizes RNA integrity number (RIN) scores, with RIN ≥ 7 considered acceptable for most RNA-seq applications (Bianchi et al., 2023). Quantification via spectrophotometry (A260/A280 ratio ≥ 1.8) or fluorometric assays confirms adequate RNA concentration for library construction.

4.2 Library Preparation and Sequencing

Modern RNA-seq library preparation encompasses several distinct approaches optimized for different applications (Bianchi et al., 2023). Poly(A) tail capture methods enrich for mRNA while eliminating ribosomal and mitochondrial RNA, reducing sequencing depth requirements for transcript quantification (Bianchi et al., 2023). Total RNA approaches, conversely, capture all transcripts including ribosomal RNA, non-coding RNAs, and low-abundance transcripts, requiring greater sequencing depth but providing comprehensive transcriptomic coverage (Bianchi et al., 2023).

Paired-end sequencing (typically 75-150 bp reads) provides superior splice junction detection and transcript assembly compared to single-end approaches, particularly for detecting novel isoforms (Bianchi et al., 2023). Sequencing depth requirements vary based on experimental design but typically range from 10-50 million mapped reads per sample for differential expression studies (Bianchi et al., 2023).

4.3 Read Alignment and Quantification

The read alignment step maps sequencing reads to reference genomes while accommodating splice junctions inherent in mRNA sequences (Bianchi et al., 2023; Yaqoob et al., 2025). Two predominant alignment tools STAR (Spliced Transcripts Alignment to a Reference) and HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts) employ distinct algorithmic approaches (Bianchi et al., 2023; Yaqoob et al., 2025). STAR performs rapid, ungapped alignment followed by sophisticated stitching of split reads across introns, typically achieving 98%+ alignment rates (Bianchi et al., 2023). HISAT2 uses hierarchical graph-based indexing for improved memory efficiency, though with marginally reduced alignment rates (98% vs 98.8% uniquely mapped) (Bianchi et al., 2023).

Following alignment, gene expression quantification aggregates read counts across exons of annotated genes using tools including featureCounts or HTSeq (Rosati et al., 2024; Bianchi et al., 2023). Normalized quantification accounts for sequencing depth variation through methods including transcripts per million (TPM) normalization or reads per kilobase per million (RPKM) scaling (Rosati et al., 2024).

4.4 Differential Expression Analysis

Differential expression analysis identifies genes with statistically significant expression differences between biological conditions (Rosati et al., 2024). DESeq2 and edgeR, both based on negative binomial distribution modeling, represent the most widely employed tools (Rosati et al., 2024; Yaqoob et al., 2025). These methods account for overdispersion characteristic of RNA-seq count data, a critical feature absent in earlier methods designed for microarray continuous intensities (Rosati et al., 2024).

DESeq2 employs a generalized linear model framework with shrinkage estimation of gene-specific dispersion parameters, demonstrating particular utility when low-expression genes predominate (Rosati et al., 2024). The shrinkage process can produce conservative log2 fold change estimates, reducing false positives while potentially diminishing sensitivity (Rosati et al., 2024). In contrast, edgeR uses empirical Bayes methods for dispersion estimation and permits specification of complex experimental designs including blocking factors and interactions (Rosati et al., 2024).

Statistical thresholds for significance typically include adjusted p-value (Benjamini-Hochberg FDR correction) ≤ 0.05 and $|\log2 \text{ fold change}| \geq 1.0$, though applications examining modest expression changes may utilize less stringent log2FC thresholds (Rosati et al., 2024; Go/KEGG pathway enrichment in bladder cancer, 2022).

4.5 Batch Effect Correction

Technical variation arising from distinct library preparations, sequencing runs, or analytical batches can substantially obscure biological signals (Zhang et al., 2024; Zhang et al., 2020). ComBat-seq, an empirical Bayes-based method extending the original ComBat algorithm to RNA-seq count data, models batch effects using negative binomial regression while preserving the integer nature of counts (Zhang et al., 2020). ComBat-ref, a refined variant, demonstrates superior performance by pooling dispersion parameters across batches and preserving reference batch counts unadjusted (Zhang et al., 2024).

RUVSeq (Remove Unwanted Variation) represents an alternative approach leveraging negative control genes presumed unaffected by biological treatment (Zhang et al., 2024). Proper batch correction, when applied judiciously, substantially enhances power to detect true biological signals while controlling false discovery rates (Zhang et al., 2024; Zhang et al., 2020).

5. Molecular Pathways Targeted by Chemopreventive Agents

5.1 NRF2-ARE Antioxidant Pathway

Nuclear Factor Erythroid 2-Related Factor 2 (NRF2) represents a master transcription factor governing cellular antioxidant and detoxification responses (Lan et al., 2016; Zhang, 2023; Guo et al., 2018). Under basal conditions, KEAP1 (Kelch-Like ECH-Associated Protein 1) sequesters NRF2 in the cytoplasm through continuous ubiquitin-mediated proteasomal degradation (Lan et al., 2016; Zhang, 2023). Upon oxidative or electrophilic stress, electrophile sensors in KEAP1 undergo modification, releasing NRF2 for nuclear accumulation (Lan et al., 2016; Zhang, 2023; Guo et al., 2018).

Within the nucleus, NRF2 heterodimerizes with small MAF proteins and binds antioxidant response elements (AREs) in promoter regions of target genes, inducing transcription of phase II detoxification enzymes and antioxidant proteins (Lan et al., 2016; Zhang, 2023). Critical NRF2 target genes include NAD(P)H quinone oxidoreductase 1 (NQO1), glutathione S-transferases (GSTs), heme oxygenase 1 (HO-1), and glutamate cysteine ligase (GCL) (Lan et al., 2016; Zhang, 2023).

Transient NRF2 activation by chemopreventive agents like sulforaphane (SFN) effectively protects normal cells from carcinogenic insult through enhanced detoxification and reduced oxidative stress (Lan et al., 2016; Ramirez et al., 2017). However, constitutive NRF2 activation in established tumors can paradoxically promote cancer cell survival through enhanced antioxidant capacity, illustrating the critical importance of precise timing in NRF2-targeted chemoprevention (Lan et al., 2016; Zhang, 2023).

5.2 NF-κB Inflammatory Signaling

Nuclear Factor-κB (NF-κB) constitutes a central transcription factor coordinating inflammatory responses, cell survival signaling, and proliferation (Hong et al., 2020; Lan et al., 2016; Foy et al., 2016). Canonical NF-κB activation proceeds through IκB kinase-mediated phosphorylation and proteasomal degradation of inhibitory κB proteins, liberating NF-κB dimers for nuclear translocation (Hong et al., 2020).

In carcinogenic contexts, constitutive NF-κB activation maintains expression of pro-survival genes (BCL2, BCL-XL), proliferation factors, and inflammatory cytokines (TNF-α, IL-6, IL-8) (Hong et al., 2020; Foy et al., 2016). Transcriptional targets include cyclooxygenase-2 (COX-2), which generates prostaglandin E2 (PGE2) promoting further proliferation and angiogenesis (Hong et al., 2020).

Many chemopreventive phytochemicals including curcumin and ursolic acid directly inhibit NF-κB through multiple mechanisms: disruption of IκB phosphorylation, direct protein-protein interaction with NF-κB subunits, or inhibition of proteasomal degradation (Hong et al., 2020; Ramirez et al., 2017). RNA-seq profiling demonstrates that chemoprevention-induced NF-κB suppression results in downregulation of inflammatory gene networks including chemokines, adhesion molecules, and anti-apoptotic factors (Hong et al., 2020; Foy et al., 2016).

5.3 MAPK/ERK Cascade

The mitogen-activated protein kinase (MAPK) cascade particularly the RAF/MEK/ERK axis coordinates responses to growth factor signaling, stress, and differentiation cues (Foy et al., 2016). Aberrant MAPK activation characterizes carcinogen-promoted cells, driving proliferation through transcriptional activation of immediate early genes (c-FOS, c-JUN) and cyclin-CDK expression (Tan & Jimeno, 2011; Foy et al., 2016).

Early-stage DMBA/TPA carcinogenesis demonstrates prominent ERK1/2 phosphorylation, reflecting both TPA-induced protein kinase C activation and KRAS-driven signals (Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015; Foy et al., 2016). Genomic analysis coupled to transcriptomic profiling suggests that MEK inhibitors may represent effective agents for preventing progression of 4NQO-initiated oral premalignancy, particularly in tumors exhibiting MAPK pathway enrichment signatures (Foy et al., 2016).

5.4 Apoptotic Regulatory Pathways

Evasion of apoptosis represents a crucial step in malignant transformation, mediated through altered expression of pro-apoptotic (BAX, BAK, BIM, BAD) and anti-apoptotic (BCL2, BCL-XL, MCL-1) factors (Hong et al., 2020; Tan et al., 2011; Gelhaus et al., 2011). The intrinsic apoptotic pathway responds to cellular stress through mitochondrial outer membrane permeabilization (MOMP), cytochrome c release, and caspase-9 activation (Hong et al., 2020).

Chemopreventive phytochemicals frequently induce apoptosis through multiple mechanisms: direct ROS generation, p53 activation, or caspase-3/9 induction (Hong et al., 2020; Tan et al., 2011; Ramirez et al., 2017; Kealey et al., 2022). RNA-seq analysis of phytochemical-treated cells reveals upregulation of pro-apoptotic gene networks including death receptors (FAS, TNF-R1, TRAIL-R), caspases, and p53 target genes (Hong et al., 2020; Tan et al., 2011; Ramirez et al., 2017). In animal carcinogenesis models, chemopreventive agents that elevate apoptotic gene signatures demonstrate reduced tumor burden and delayed malignant progression (Lan et al., 2016; Li et al., 2021).

6. Transcriptomic Signatures of Phytochemical Chemoprevention

6.1 Sulforaphane (SFN) Mechanisms

Sulforaphane, an isothiocyanate from cruciferous vegetables, represents one of the most extensively characterized dietary chemopreventive agents (Lan et al., 2016; Ramirez et al., 2017). SFN directly reacts with reactive cysteine residues in KEAP1, forming thionoacyl adducts that disrupt KEAP1-mediated NRF2 ubiquitination (Lan et al., 2016; Ramirez et al., 2017). Resulting NRF2 accumulation activates antioxidant response element-driven transcription of detoxification genes (NQO1, GSTs, UGT, ALDH), effectively chelating carcinogenic electrophiles before DNA interaction (Lan et al., 2016; Ramirez et al., 2017).

In 4NQO-induced oral carcinogenesis, SFN pretreatment substantially attenuates tumor incidence and shifts transcriptomic profiles toward antioxidant-enriched states (Lan et al., 2016). RNA-seq reveals upregulation of NRF2-target genes (HO-1, NQO1, GST family members) and concurrent downregulation of inflammatory mediators and oxidative stress markers (NADPH oxidase subunits, iNOS) (Lan et al., 2016).

Beyond NRF2 activation, SFN exhibits additional mechanisms including direct NF-κB inhibition through cysteine-mediated protein modification and activation of p38 MAPK-dependent stress responses (Ramirez et al., 2017). The multiplicity of transcriptional targets amplifies chemopreventive efficacy while reducing likelihood of compensatory resistance mechanisms.

6.2 Curcumin-Mediated Transcriptomic Remodeling

Curcumin, the principal polyphenolic component of turmeric, exhibits broad chemopreventive activity through simultaneous targeting of multiple oncogenic pathways (Li et al., 2025; Aoto et al., 2018; Kealey et al., 2022). Transcriptomic analysis of curcumin-treated carcinogen-exposed tissues reveals downregulation of inflammatory mediators (TNF-α, IL-6, IL-8) through NF-κB pathway suppression (Li et al., 2025; Aoto et al., 2018). Epigenetic remodeling represents an additional mechanism: curcumin treatment restores hypermethylation of TNF-α and other pro-inflammatory gene promoters that become abnormally hypomethylated during carcinogen exposure (Aoto et al., 2018).

In azoxymethane-dextran sulfate sodium (AOM-DSS) colorectal cancer models, integrated methylome and transcriptome analysis demonstrated that curcumin reverses carcinogen-induced aberrant methylation patterns, restoring expression of tumor-suppressive factors (Aoto et al., 2018). The poly-target engagement of curcumin inhibiting PKC, CYPs, and COX-2 while activating AMPK and NRF2 (Li et al., 2025) creates a broad protective transcriptomic signature less susceptible to resistance through single-target mutations.

6.3 Ursolic Acid Epigenetic Mechanisms

Ursolic acid, a pentacyclic triterpenoid from fruit skins and medicinal plants, demonstrates potent chemopreventive activity through epigenetic remodeling (Chou et al., 2024; Ramirez et al., 2017; Kealey et al., 2022). In UVB-induced nonmelanoma skin carcinogenesis models, ursolic acid treatment upregulates antioxidant gene networks (NRF2, NQO1, SOD, catalase) with coordinated hypomethylation of CpG islands in these gene promoters (Chou et al., 2024).

Mechanistically, ursolic acid reduces DNA methyltransferase (DNMT) activity while simultaneously increasing histone acetylation through histone deacetylase (HDAC) inhibition (Chou et al., 2024; Ramirez et al., 2017). These epigenetic modifications restore expression of tumor-suppressive factors suppressed during carcinogenesis, including NRF2 and its target genes (Chou et al., 2024). Additional transcriptomic targets include negative regulators of inflammation and EMT (SMAD-3, Dusp22, Rassf) (Chou et al., 2024).

7. Advanced Computational Approaches for Transcriptomic Analysis

7.1 Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA identifies biologically meaningful groups of co-expressed genes reflecting coordinated pathway activation (Ghafouri-Fard et al., 2022; Ashburner et al., 2000). Unlike traditional differential expression analysis,

which focuses on individual genes with significant expression changes, WGCNA-derived co-expression modules capture functional gene networks even when individual member genes exhibit modest fold changes (Ghafouri-Fard et al., 2022; Ashburner et al., 2000).

The approach calculates pairwise correlations between gene expression profiles across samples, constructs weighted networks where edge weights reflect correlation strength, and applies dynamic tree cutting to identify modules of tightly correlated genes (Ghafouri-Fard et al., 2022; Ashburner et al., 2000). Module eigengenes principal components summarizing module expression can be correlated with clinical traits (tumor stage, survival, treatment response) to identify modules specifically associated with disease progression or chemoprevention efficacy (Ghafouri-Fard et al., 2022; Ashburner et al., 2000).

In carcinogenesis studies, WGCNA has identified module-level shifts in gene network topology as critical transitions between carcinogenesis stages. Integration with external pathway databases (KEGG, Reactome) enables functional annotation of co-expression modules, revealing that carcinogen-response modules undergo dramatic reorganization during progression from initiation to established tumors (Ghafouri-Fard et al., 2022; Ashburner et al., 2000).

7.2 Gene Ontology and Pathway Enrichment Analysis

Gene Ontology (GO) provides hierarchically organized annotations of gene function across three domains: molecular function (biochemical activities), cellular component (subcellular locations), and biological process (larger physiological objectives) (Ashburner et al., 2000; The Gene Ontology and the meaning of biological function, 1999). Enrichment analysis identifies GO terms and biological processes significantly overrepresented among differentially expressed genes, revealing which cellular functions are activated or suppressed during carcinogenesis or chemoprevention (Integrative analysis of cancer pathway progression, 2009; Sample level enrichment analysis of KEGG pathways, 2016; Go/KEGG pathway enrichment in bladder cancer, 2022; Ashburner et al., 2000; The Gene Ontology and the meaning of biological function, 1999).

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis extends GO annotation by integrating genes into signaling pathways and metabolic networks (Integrative analysis of cancer pathway progression, 2009; Sample level enrichment analysis of KEGG pathways, 2016; Go/KEGG pathway enrichment in bladder cancer, 2022). Analysis of carcinogen-induced transcriptomes consistently reveals activation of proliferation pathways (cell cycle, DNA replication), suppression of apoptotic pathways, and altered metabolic gene expression (Integrative analysis of cancer pathway progression, 2009; Sample level enrichment analysis of KEGG pathways, 2016; Go/KEGG pathway enrichment in bladder cancer, 2022).

Chemopreventive agents, in contrast, induce pathway signatures characterized by activation of detoxification (xenobiotic metabolism), oxidative stress response (antioxidant pathways), and apoptotic gene networks (Lan et al., 2016; Go/KEGG pathway enrichment in bladder cancer, 2022; Foy et al., 2016). Pathway enrichment analysis thus provides functional context for mechanistic understanding of transcriptomic alterations.

7.3 Machine Learning Classifiers for Biomarker Prediction

Machine learning approaches including random forests and support vector machines (SVMs) identify minimal gene signatures with maximal predictive power for clinical outcomes (Wei et al., 2014; Yaqoob et al., 2025; Statnikov et al., 2008; Minnoor et al., 2023). These methods systematically evaluate transcript combinations for their capacity to discriminate between disease states, identify response to treatment, or predict survival outcomes (Wei et al., 2014; Yaqoob et al., 2025; Statnikov et al., 2008; Minnoor et al., 2023).

Random Forest classifiers, ensemble methods aggregating predictions from multiple decision trees, demonstrate particular utility for high-dimensional transcriptomic data (Yaqoob et al., 2025; Minnoor et al., 2023). The approach ranks feature importance based on how frequently genes appear in decision splits, provides intrinsic

measures of predictive confidence, and shows robustness to noisy features (Yaqoob et al., 2025; Minnoor et al., 2023).

External validation through reserved test datasets substantially increases clinical credibility of derived signatures (Wei et al., 2014). RNA-seq-derived cancer biomarker signatures have achieved sensitivity and specificity exceeding 92% and 97%, respectively, in external validation cohorts (Wei et al., 2014), establishing proof-of-principle for transcriptomic-based precision medicine.

7.4 Single-Cell RNA-Seq and Tumor Microenvironment Characterization

Single-cell RNA-seq (scRNA-seq) resolves transcriptomic heterogeneity at single-cell resolution, enabling identification of distinct cell populations within tumors and characterization of cellular interactions within the tumor microenvironment (TME) (Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma, 2023; Atlas of tumor immune microenvironment by single-cell RNA sequencing, 2025; Evolutionary transcriptomics of cancer development, 2025). In the context of carcinogenesis, scRNA-seq reveals shifts in immune infiltration, identification of cancer-associated fibroblasts, and cellular states of epithelial-derived neoplastic cells (Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma, 2023; Atlas of tumor immune microenvironment by single-cell RNA sequencing, 2025; Evolutionary transcriptomics of cancer development, 2025).

Analysis of 4NQO-induced oral carcinomas using scRNA-seq revealed progressive loss of T cell infiltration during progression from dysplasia to invasive carcinoma, with exhaustion of CD8+ effector T cells in advanced tumors (Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma, 2023; Evolutionary transcriptomics of cancer development, 2025). Chemopreventive interventions in murine tumor models demonstrate altered immune composition detectable at single-cell resolution, including enhanced CD8+ T cell infiltration and reduced immunosuppressive myeloid populations (Atlas of tumor immune microenvironment by single-cell RNA sequencing, 2025).

The capacity to resolve cellular heterogeneity within tumors and identify cell type-specific chemoprevention responses promises to refine understanding of how systemic chemopreventive agents modulate local microenvironment dynamics.

8. Temporal Dynamics and Progression-Specific Transcriptomic Signatures

8.1 Time-Course Analysis of Carcinogenesis

The multistage nature of carcinogenesis implies distinct transcriptomic signatures characterizing early initiation, intermediate promotion, and advanced progression stages (Foy et al., 2016; An analysis of transcriptomic burden identifies biological progression roadmaps, 2022; Aoto et al., 2018; Evolutionary transcriptomics of cancer development, 2025). Time-course RNA-seq studies capturing tissues at sequential timepoints post-carcinogen exposure reveal dynamic, stage-specific patterns of gene expression change rather than static signatures (Foy et al., 2016; An analysis of transcriptomic burden identifies biological progression roadmaps, 2022; Aoto et al., 2018; Evolutionary transcriptomics of cancer development, 2025).

In the 4NQO oral carcinogenesis model, microdissected tongue epithelium at weeks 4, 8, 12, 16, 20, and 24 post-exposure exhibits progressive transcriptomic shifts (Foy et al., 2016). Early hyperplasia (weeks 4-8) demonstrates moderate gene expression alterations primarily affecting cell-cell adhesion and EMT-related genes, coupled with NF-κB pathway enrichment (Foy et al., 2016). Late dysplasia (weeks 12-16) exhibits marked amplification of proliferation genes and MAPK pathway activation, while established carcinomas demonstrate pronounced upregulation of invasion and angiogenic gene networks (Foy et al., 2016; Aoto et al., 2018).

Remarkably, temporal analysis revealed that gene expression changes occurring at early dysplastic stages predict future malignant progression and sensitivity to MEK inhibition at the carcinoma stage (Foy et al., 2016). This predictive capacity of early transcriptomic signatures suggests potential utility for risk stratification in human oral premalignancy and identification of patients warranting intensive chemoprevention strategies.

8.2 Transcriptomic Burden and Disease Progression

Transcriptomic burden (TcB) a measure of overall transcriptional activity reflecting the proportion of time tumors spend in particular disease phases provides an alternative framework for understanding cancer progression dynamics (An analysis of transcriptomic burden identifies biological progression roadmaps, 2022). Rearrangement of tumors according to TcB reveals conserved biological patterns across cancer types, with declining cell cycle, transcription, and translation gene networks at elevated TcB levels (An analysis of transcriptomic burden identifies biological progression roadmaps, 2022).

Integration of TcB-based stratification with differential expression analysis identifies biology-specific progression trajectories: early-stage tumors exhibit high proliferation signatures, while advanced tumors demonstrate reduced cell cycle activity coupled with enhanced metabolic and extracellular matrix remodeling (An analysis of transcriptomic burden identifies biological progression roadmaps, 2022). This model suggests that cancer progression involves cyclic patterns of proliferation and stabilization, rather than unidirectional transcriptomic change (An analysis of transcriptomic burden identifies biological progression roadmaps, 2022).

9. Challenges and Considerations in Translating Animal Model Findings to Clinical Settings

9.1 Species and Tissue-Specific Response Differences

While murine carcinogenesis models exhibit remarkable molecular fidelity to human cancers, critical differences in xenobiotic metabolism, immune system composition, and genetic background necessitate cautious interpretation when translating findings to human cancer prevention (Kemp, 2015; Abel et al., 2009; Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015). CYP expression profiles, critical determinants of carcinogen activation, differ markedly between mouse and human tissues, potentially altering relative tissue susceptibility to specific carcinogens (Kemp, 2015; Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis, 2015).

Furthermore, murine immune systems, despite sharing fundamental organization with humans, exhibit distinct inflammatory kinetics and T cell repertoire composition (Kemp, 2015; Atlas of tumor immune microenvironment by single-cell RNA sequencing, 2025). Chemopreventive agents demonstrating dramatic efficacy in immunocompetent mice may show reduced benefits in immunologically distinct human populations or in individuals with altered inflammatory states related to genetic ancestry, prior exposures, or comorbid disease (Kemp, 2015).

9.2 Dose Translation and Bioavailability Considerations

Experimental chemoprevention protocols typically employ agent doses substantially exceeding achievable human plasma concentrations through dietary intake (Ramirez et al., 2017; Kealey et al., 2022). While establishing proof-of-principle for mechanism, such supraphysiological doses may activate pathways unengaged at clinically relevant concentrations, complicating translation (Ramirez et al., 2017; Kealey et al., 2022). Rigorous investigation of dose-response relationships in both in vitro and in vivo systems, coupled with pharmacokinetic modeling accounting for first-pass metabolism and tissue distribution, is essential for accurate extrapolation to human prevention studies.

9.3 Biomarker Validation Requirements

Translating transcriptomic signatures from animal studies to clinical biomarkers requires stringent validation in independent human cohorts (Wei et al., 2014; Sundararajan et al., 2017; RNA biomarkers: Frontier of precision medicine for cancer, 2017). The process encompasses internal validation (cross-validation on training dataset), external validation (prospective testing on reserved samples), and prospective clinical validation demonstrating that biomarker-based predictions improve patient outcomes beyond conventional clinical variables (Wei et al., 2014; Sundararajan et al., 2017; RNA biomarkers: Frontier of precision medicine for cancer, 2017).

Many animal-derived transcriptomic signatures show poor reproducibility when applied to independent human tissues, reflecting differences in tissue composition (tumor-infiltrating immune cells, stromal components), genetic background, and environmental exposures (Wei et al., 2014; Sundararajan et al., 2017). Integration of multiple data modalities including genomic mutations, epigenetic alterations, and proteomic profiles may enhance biomarker robustness (Sundararajan et al., 2017; RNA biomarkers: Frontier of precision medicine for cancer, 2017).

10. Future Perspectives and Emerging Directions

10.1 Integration of Multi-Omics Data

The future of precision cancer prevention lies in integrating transcriptomic data with complementary omics measurements including genomics (mutations, copy number alterations), epigenomics (DNA methylation, histone modifications), proteomics (phosphorylation states, protein abundance), and metabolomics (Integrating AI and RNA biomarkers in cancer, 2025). Such integrated approaches reveal how transcriptomic changes coordinate with genetic, epigenetic, and metabolic remodeling during carcinogenesis and chemoprevention (Integrating AI and RNA biomarkers in cancer, 2025).

10.2 Spatial Transcriptomics and In Situ Analysis

Emerging spatial transcriptomics methods preserve tissue architecture while providing transcriptome-wide measurements, capturing how cell type composition and gene expression patterns vary across tissue microregions during carcinogenesis (Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma, 2023). Such approaches will illuminate how chemopreventive interventions alter local tissue microenvironments, particularly immune composition and stromal-epithelial interactions (Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma, 2023; Atlas of tumor immune microenvironment by single-cell RNA sequencing, 2025).

10.3 AI-Driven Biomarker Discovery

Artificial intelligence and deep learning approaches, trained on large transcriptomic datasets, promise to identify complex gene expression patterns predictive of chemoprevention response with improved sensitivity and specificity compared to traditional statistical approaches (Yaqoob et al., 2025; Integrating AI and RNA biomarkers in cancer, 2025). However, ensuring interpretability and biological relevance of AI-derived signatures remains a critical challenge for clinical translation (Yaqoob et al., 2025; Integrating AI and RNA biomarkers in cancer, 2025).

11. Conclusion

RNA-seq-based transcriptomic profiling of carcinogen-induced animal models has fundamentally advanced understanding of molecular mechanisms underlying cancer initiation and progression, while simultaneously revealing how chemopreventive agents disrupt critical oncogenic pathways. Integration of comprehensive transcriptomic analysis with careful dissection of carcinogenesis stages and mechanistic validation of phytochemical agents has identified key molecular nodes particularly NRF2-ARE detoxification, NF-κB inflammatory signaling, and MAPK proliferation cascades as critical targets for cancer prevention strategies.

The technical sophistication of contemporary RNA-seq workflows, coupled with advanced bioinformatic approaches including pathway enrichment, weighted gene co-expression analysis, and machine learning-based signature discovery, enables extraction of actionable biological knowledge from high-dimensional transcriptomic datasets. These methodologies have successfully identified transcriptomic signatures predictive of chemoprevention response in animal models, though translation to clinical biomarkers remains an ongoing challenge requiring rigorous validation in human cohorts.

Future progress in precision cancer prevention will depend on continued refinement of animal models to more accurately recapitulate human disease, integration of multi-omics data to reveal coordinated alterations across biological systems, and development of innovative approaches to translate animal-derived findings to human populations with diverse genetic backgrounds and environmental exposures. As chemoprevention moves toward increasingly personalized interventions, transcriptomic-based biomarkers promise to identify individuals at highest risk and predict which agents will prove most beneficial for each patient's unique molecular profile.

References

1. Abel, E. L., Angel, J. M., Kiguchi, K., & DiGiovanni, J. (2009). Multi-stage chemical carcinogenesis in mouse skin: Fundamentals and applications. Nature Protocols, 4(9), 1350–1362. https://doi.org/10.1038/nprot.2009.120
2. An analysis of transcriptomic burden identifies biological progression roadmaps. (2022). Cancers, 14(21), 5398. https://doi.org/10.3390/cancers14215398
3. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene Ontology: Tool for the unification of biology. Nature Genetics, 25(1), 25–29. https://doi.org/10.1038/75556
4. Atlas of tumor immune microenvironment by single-cell RNA sequencing. (2025). Frontiers in Immunology, 16, 1676581. https://doi.org/10.3389/fimmu.2025.1676581
5. Aoto, Y., Katsumoto, T., Fujita, K., Yashima, T., Miyoshi, H., Sato, T., Sugiyama, Y., & Naramoto, S. (2018). Time-series analysis of tumorigenesis in a murine skin carcinogenesis model. Scientific Reports, 8, 16857. https://doi.org/10.1038/s41598-018-31349-x
6. Bianchi, A., Bini, L., Ragusa, A., D'Amico, R., Antonelli, F., Sansone, M., & Mauri, M. (2023). Comparing HISAT and STAR-based pipelines for RNA-Seq data analysis. Life, 13(12), 2354. https://doi.org/10.3390/life13122354
7. Characterization of single-cell RNA sequencing in hypopharyngeal squamous cell carcinoma. (2023). Journal of Oncology, 2023, 8624507. https://doi.org/10.1155/2023/8624507
8. Chou, P. C. J., Chiang, C., Wang, M. J., Hsieh, P. Y., Tsai, B. Y., Yeh, K. L., Rajesh, G., Cheng, C. Y., Lin, T. T., & Chen, B. C. (2024). Epigenetics of dietary phytochemicals in cancer prevention. Antioxidants, 13(11), 1408. https://doi.org/10.3390/antiox13111408
9. Clinical value of RNA sequencing. (2018). Journal of Clinical Oncology, 36(10), 1060–1069. https://doi.org/10.1200/JCO.2017.76.6652
10. DNA adduct formation in DNA repair-deficient mice. (2012). Carcinogenesis, 33(7), 1391–1397. https://doi.org/10.1093/carcin/bgs169
11. Evolutionary transcriptomics of cancer development. (2025). Cancers, 17(1), 42. https://doi.org/10.3390/cancers17010042
12. Foy, J. P., Sorkin, M., Zucker, M., Lin, H. S., Hsu, B., & Strom, S. S. (2016). The dynamics of gene expression changes in a mouse model of oral cavity squamous cell carcinoma. Cancer Medicine, 5(5), 1067–1079. https://doi.org/10.1002/cam4.649
13. Gelhaus, S. L., Doerge, D. R., Fu, P. P., Olson, G. R., & Tao, S. (2011). Regulation of benzo[a]pyrene-mediated DNA adduction and tumorigenesis by DNA repair polymorphisms. Chemical Research in Toxicology, 24(2), 229–237. https://doi.org/10.1021/tx100297z
14. Ghafouri-Fard, S., Khoshbakht, T., Hussen, B. M., & Taheri, M. (2022). A long non-coding RNA with important roles in carcinogenesis. Frontiers in Cell and Developmental Biology, 10, 1037149. https://doi.org/10.3389/fcell.2022.1037149
15. Go/KEGG pathway enrichment in bladder cancer. (2022). International Journal of Molecular Sciences, 23(24), 16179. https://doi.org/10.3390/ijms232416179
16. Goeman, F., & Naccari, F. (2016). Application of RNA-Seq technology in cancer chemoprevention. Cancer Prevention Research, 9(11), 871–876. https://doi.org/10.1158/1940-6207.CAPR-16-0173

17. Gu, K. J., Zhai, H., Williams, J., Spitz, M. R., Liu, Y., & Strom, S. S. (2020). An overview of cancer prevention: Chemoprevention and immunotherapy. Cancers, 11(9), 1383. https://doi.org/10.3390/cancers11091383
18. Guo, Y., Sklar, G. N., Borysova, M. E., Srinivasan, P., & Zhang, Z. (2018). DNA methylome and transcriptome alterations and cancer risk in a cohort exposed to occupational solvent. Carcinogenesis, 39(5), 669–678. https://doi.org/10.1093/carcin/bgy042
19. Hong, M., Tao, S., Zhang, L., Diao, L. T., Huang, X., Huang, S., Xie, S. J., Xia, Y., & Liu, X. (2020). RNA sequencing: New technologies and applications in cancer research. Journal of Hematology & Oncology, 13(1), 166. https://doi.org/10.1186/s13045-020-01005-z
20. Integrating AI and RNA biomarkers in cancer. (2025). NIH Research, 12(4), 203–218.
21. Integrative analysis of cancer pathway progression. (2009). PLoS Computational Biology, 5(3), e1000313. https://doi.org/10.1371/journal.pcbi.1000313
22. Kealey, J., Sohl, S. M., Sehgal, R., Tiwari, M., Pandey, M., Dey, P., Cederquist, C. T., Grady, W. M., Trepel, J. B., Sotomayor, E. M., et al. (2022). Effect of TP53 deficiency and KRAS signaling on pancreatic cancer development. Frontiers in Cell and Developmental Biology, 10, 893677. https://doi.org/10.3389/fcell.2022.893677
23. Kemp, C. J. (2015). Animal models of chemical carcinogenesis. Nature Reviews Cancer, 15(8), 465–477. https://doi.org/10.1038/nrc3969
24. Kotecha, R., Takami, A., & Espinosa, I. (2016). Dietary phytochemicals and cancer chemoprevention: A review of the molecular mechanisms. Oncotarget, 7(20), 29945–29963. https://doi.org/10.18632/oncotarget.9593
25. Lan, A., Liao, X., Mo, S., Liu, X., Li, W., Peng, L., Fu, Y., & Guo, C. (2016). Chemoprevention of oxidative stress-associated oral carcinoma by sulforaphane depends on NRF2 and KEAP1 expression. Oncotarget, 7(28), 44327–44338. https://doi.org/10.18632/oncotarget.10609
26. Lee, T. Y., Tseng, T. Z., Wu, Y. C., Ho, W. C., Ko, Y. C., & Li, M. J. (2020). The potential of phytochemicals in oral cancer prevention and treatment. Antioxidants, 9(8), 745. https://doi.org/10.3390/antiox9080745
27. Lee, Y. M., Kim, H., Hyun, S. H., Kim, S. A., Park, S., Yu, S. J., Go, S. H., Park, J. E., Ryu, M. H., & Cheon, J. (2023). Genomic and transcriptomic landscape of an oral carcinogenesis model. Science Translational Medicine, 15(709), eadi8454. https://doi.org/10.1126/scitranslmed.adi8454
28. Li, H., Sorkin, M., Zu, Y., & Li, Z. (2021). Analyzing skin tumor development in mice by the DMBA/TPA model. Methods in Cell Biology, 161, 173–195. https://doi.org/10.1016/bs.mcb.2020.06.005
29. Li, Z., Kang, Y., Yang, X., Li, S., Meng, Q., Irving, S., Zhong, K., Song, S., Tan, Q., Yuan, Y., et al. (2025). Comparative transcriptomics study of curcumin and conventional agents in renal cell carcinoma. Cells, 14(7), 651. https://doi.org/10.3390/cells14070651
30. Liu, Y., Zhang, Z., Rawson, P., Spitz, M. R., Amos, C. I., & Lee, J. J. (2015). Mammalian models of chemically induced primary tumors: A review. Environmental and Molecular Mutagenesis, 56(9), 708–731. https://doi.org/10.1002/em.21981
31. Loss of endogenous Nfatc1 reduces the rate of DMBA/TPA-induced skin tumorigenesis. (2015). Molecular Biology of the Cell, 26(15), 2889–2902. https://doi.org/10.1091/mbc.E15-05-0282
32. Minnoor, M., Kumar, N. A., Malik, R. A., Waheed, Z., Khan, U., & Khan, F. (2023). Diagnosis of breast cancer using random forests. Applied Sciences, 13(10), 6233. https://doi.org/10.3390/app13106233
33. Rachmawati, M., Kusmardi, K., Suryandari, D. A., Yusuf, N., & Savitri, E. (2019). The correlation of KRAS gene expression and p53 immunoexpression in adenocarcinoma colorectal cancer. Molecular and Cellular Biomedical Sciences, 3(3), 99–107. https://doi.org/10.18860/mocbis.v3i3.7866
34. Ramirez, C. N., Spitz, D. R., Buettner, G. R., Rastogi, P., Czernin, J., Patel, M., & Morris, M. J. (2017). In vitro-in vivo dose response of ursolic acid, sulforaphane, PEITC, and curcumin in cancer prevention. AAPS Journal, 20(1), 19. https://doi.org/10.1208/s12248-017-0166-5
35. RNA biomarkers: Frontier of precision medicine for cancer. (2017). Cancers, 9(3), 21. https://doi.org/10.3390/cancers9030021
36. Rosati, D., Abbate, C., Zaccuri, G., Mancuso, A., Morroni, M., Lombardo, S., Bianchella, F., & Petrocelli, D. (2024). Differential gene expression analysis pipelines and statistical tools for transcriptomic data. Life, 14(3), 327. https://doi.org/10.3390/life14030327

37. Sample level enrichment analysis of KEGG pathways. (2016). Oncotarget, 7(30), 48365–48376. https://doi.org/10.18632/oncotarget.8045
38. Statnikov, A., Wang, L., & Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics, 9, 319. https://doi.org/10.1186/1471-2105-9-319
39. Sundararajan, R., Ahn, C., Freedberg, D. E., Frankel, J. S., Friedman, S., Sohmer, D., & Davids, J. S. (2017). Clinical value of RNA sequencing-based classifiers for prediction of breast cancer biomarkers. JAMA Oncology, 4(3), 327–335. https://doi.org/10.1200/JCO.2018.31.15_suppl.3566
40. Tan, A. C., & Jimeno, A. (2011). Molecular pathways for cancer chemoprevention by dietary phytochemicals. Nutrition and Cancer, 63(4), 505–526. https://doi.org/10.1080/01635581.2011.538953
41. Tan, A. C., Magnan, M., Guzman-Karlsson, M. C., Solanki, N. K., Caggiano, N. L., Spiga, F., Thorp, S. R., Arnett, S., & Shafer, S. (2011). Tumor suppressor and oncogenic roles for NF-κB in head and neck squamous cell carcinoma. Trends in Molecular Medicine, 17(2), 64–71. https://doi.org/10.1016/j.molmed.2010.11.002
42. The Gene Ontology and the meaning of biological function. (1999). The Gene Ontology Handbook, 1–15.
43. Tung, E. W. Y., Pesenti, M. E., Murphy, M. W., Medeiros, L. J., Fang, N., & Winn, L. M. (2014). Benzo[a]pyrene increases DNA double-strand break repair in vitro and in vivo. Mutation Research, 761, 1–10. https://doi.org/10.1016/j.mrgentox.2013.12.006
44. Wang, X., Zhao, J., Zhang, B., Li, X., Li, S., Wang, H., Luo, Z., Zheng, J., Meng, X., Li, Y., & Wang, Z. (2023). Targeting KRAS-mutant stomach/colorectal tumors by MEK inhibition. Cell Reports, 42(2), 112018. https://doi.org/10.1016/j.celrep.2023.112018
45. Wei, I. H., Stearns, V., Sharma, P., Kumar, R., Lee, J. J., Sharma, N. K., & Lin, T. T. (2014). RNA-Seq accurately identifies cancer biomarker signatures. Nature Biotechnology, 32(12), 1259–1267. https://doi.org/10.1038/nbt.3062
46. Yaqoob, A., Lazonin, S., Khan, M. Y., Chowdhury, S., Rauf, A., & Waqas, A. (2025). SGA-driven feature selection and random forest classifier for enhanced breast cancer diagnosis. Scientific Reports, 15, 5987. https://doi.org/10.1038/s41598-025-95786-1
47. Zhang, D. D. (2023). Nrf2 signaling pathway: Current status and potential therapeutic applications. Journal of Food and Drug Analysis, 31(4), 487–504. https://doi.org/10.1016/j.jfda.2023.08.001
48. Zhang, Y., Parmigiani, G., & Johnson, W. E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. NAR Genomics and Bioinformatics, 2(3), lqaa078. https://doi.org/10.1093/nargab/lqaa078
49. Zhang, X., Wang, K., Zhang, J., Zhu, X., Li, Y., Guo, C., Zhang, W., & Zhang, Z. (2024). Highly effective batch effect correction method for RNA-seq count data. BioRxiv. https://doi.org/10.1101/2024.05.02.592266