



Review

Integrative Genomic and AI Approaches to Lung Cancer and Implications for Disease Prevention in Former Smokers

Katya H. Bénard ^{1,2,*}, Vanessa G. P. Souza ^{1,*}, Greg L. Stewart ^{1,2}, Katey S. S. Enfield ^{1,2,3} and Wan L. Lam ^{1,2,3}

¹ British Columbia Cancer Research Institute, Vancouver, BC V5Z 1L3, Canada

² Interdisciplinary Oncology Program, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

³ Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC V6T 1Z7, Canada

* Correspondence: kbenard@bccrc.ca (K.H.B.); vsouza@bccrc.ca (V.G.P.S.)

† These authors contributed equally to this work.

Abstract

Tobacco smoking accounts for nearly 90% of lung cancer deaths worldwide, yet the mechanisms underlying persistent cancer risk in former smokers are not fully understood. Epidemiological evidence shows that more than 40% of lung cancers develop over 15 years after cessation, demonstrating that while some smoking-induced molecular alterations resolve rapidly, others remain as long-lasting scars that promote carcinogenesis. This review synthesizes longitudinal and cross-sectional genomic, epigenomic, and transcriptomic studies of airway and lung tissues to distinguish persistent from nonpersistent smoking-induced molecular alterations. Persistent alterations include somatic mutations in *TP53* and *KRAS*, DNA methylation at tumor suppressor loci, dysregulated noncoding RNAs, chromosomal instability, and epigenetic age acceleration. Nonpersistent changes, such as acute inflammatory responses and detoxification pathways, generally normalize within months to several years following cessation. Multi-omics profiling reveals coordinated patterns of dysregulation consistent with field cancerization in former smokers. In addition, the integration of multi-omics data with artificial intelligence may enable composite molecular signatures for stratifying high-risk former smokers, link molecular persistence to clinical outcomes, and inform chemoprevention strategies. Collectively, these observations clarify which molecular alterations sustain long-term cancer risk despite smoking cessation and highlight opportunities for precision prevention and earlier detection in high-risk populations.

Keywords: smoking cessation; genomic profiling; persistent molecular alterations; lung cancer; DNA methylation; artificial intelligence; precision prevention; field cancerization



Academic Editor: Magdalena Orzechowska

Received: 10 November 2025

Revised: 30 December 2025

Accepted: 1 January 2026

Published: 4 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Lung cancer remains the leading cause of cancer death worldwide, accounting for approximately 1.8 million deaths annually and representing 18% of all cancer deaths [1]. Despite progress in early detection and treatment, mortality rates remain high due to late-stage diagnoses, persistent global smoking prevalence, limited specificity and sensitivity of screening methods, and treatment resistance. Non-small-cell lung cancer (NSCLC) comprises 87% of lung cancer cases, with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) as the primary histologic subtypes [2]. Small-cell lung cancer (SCLC) accounts for the remaining 13% of cases [2]. Among lung cancers, LUSC and SCLC have the

strongest association with tobacco exposure, with more than 95% of SCLC cases occurring in individuals with a history of tobacco use [3,4]. Smoking is the predominant risk factor for lung cancer and is responsible for approximately 85% of cases globally [5]. Risk increases with both duration and quantity of smoking [2]. Although global tobacco use has declined from 32.7% of adults in 2000 to 21.7% in 2020, with projections to fall below 20% by 2025, former smokers continue to face persistently elevated risk of lung cancer that can last for decades after quitting [6–8]. This enduring susceptibility, combined with the large population of individuals with a history of smoking, represents a substantial and ongoing health burden from tobacco-related lung cancers.

Tobacco smoke induces cytotoxic damage to airway epithelial cells, resulting in oxidative stress, DNA damage, and chronic inflammation, which are key factors that promote smoking-related lung disease [9,10]. Chronic exposure to tobacco carcinogens leads to the formation of DNA adducts, resulting in oncogenic mutations in critical genes, including *TP53* and *KRAS* [11–13]. Tobacco smoke is also associated with aberrant DNA methylation of promoter regions in tumor suppressor genes such as *CDKN2A/p16*, silencing their expression [14,15]. These processes promote basal cell hyperplasia and squamous metaplasia in the airway epithelium, contributing to epithelial remodeling, barrier dysfunction, and other early histopathologic changes that represent or precede premalignant lesions [16,17]. For example, polycyclic aromatic hydrocarbons (PAHs), compounds that primarily act as local carcinogens in the bronchial epithelium, form PAH-DNA adducts that generate characteristic mutations such as the excess G to T transversions that are commonly observed in smoking-related squamous tumors [15]. These PAH-associated mutational patterns, including site-specific damage at *TP53* hotspots, are consistent with genomic alterations underlying LUSC arising from chronically smoke-exposed airway epithelium. The tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) acts as a systemic lung carcinogen that mainly induces adenocarcinoma in experimental models. This finding is consistent with the prevalence of LUAD among smokers [15]. The accumulation of these molecular insults across the respiratory tract creates a “field cancerization” effect in which large areas of cells harbor pre-neoplastic changes [18–20]. Epigenetic studies have reported that tobacco smoke is associated with reproducible alterations in DNA methylation and dysregulation of miRNAs. Histone modifications associated with tobacco smoke exposure are also well documented [21].

Not all molecular changes revert after smoking cessation. Some remain after decades of abstinence and are therefore described as “persistent” or “irreversible.” Other molecular alterations gradually normalize to levels of never-smokers and are referred to as “reversible” or “nonpersistent” [17,22,23]. These terms are not defined with uniform temporal thresholds across studies. Follow-up durations vary, ranging from months to several years, and in some instances, more than a decade after smoking cessation. Therefore, persistence is more appropriately viewed as a spectrum rather than a fixed, universally applicable cutoff, reflecting differences in molecular class, genomic context, tissue type, and study design. In this review, these alterations are referred to as persistent when they remain significantly altered in long-term former smokers or are reproducibly observed across independent cohorts and tissue types, even when assessed years after cessation. Nonpersistent alterations show partial or complete reversion toward never-smoker levels in longitudinal or cross-sectional cessation studies, typically within months or several years. This framework emphasizes reproducibility and signal durability rather than a fixed cessation interval. Certain features, such as DNA methylation marks, reduced DNA repair capacity, and self-sustaining signaling loops, may persist long after smoke exposure ends and may contribute to elevated long-term cancer risk despite cessation. Epidemiological

data demonstrate that excess cancer risk declines slowly after cessation, plateauing at about 20% above baseline only 20 years after quitting [7].

Persistent changes can be leveraged for risk stratification and allow opportunities for early detection strategies and personalized surveillance among high-risk former smokers. Identifying nonpersistent alterations may reveal reversible molecular targets for chemoprevention, allowing intervention before persistent changes establish malignancy. Early detection is especially important in lung cancer, as most cases are detected in advanced stages, when curative options are limited [24,25].

Distinguishing between persistent and nonpersistent molecular changes is critical for precision prevention. However, these alterations span multiple molecular classes (e.g., genetic, epigenetic, transcriptomic, proteomic, metabolomic), exhibit high inter-patient variability, and are captured in vast, complex datasets that are difficult to interpret using conventional analytic approaches [17,21]. Cohorts are often relatively small and diverse and have limited labeling of persistence, which is typically determined by longitudinal assessments. The process of molecular recovery also takes time, making it difficult for current methods to capture risk trajectories. In addition, models created from one population frequently fail to generalize to others. Advanced computational methods, including artificial intelligence (AI) and machine learning (ML), are emerging as valuable approaches for integrating multi-omics data, distinguishing short-lived from enduring molecular scars, and translating the findings into actionable biomarkers for monitoring and intervention [26–28].

Although extensive research has characterized smoking-related molecular alterations and AI-based cancer risk prediction has advanced substantially, an important gap remains in how these areas are integrated to distinguish persistent from nonpersistent changes and to inform prevention strategies in former smokers. To address this gap, relevant literature was identified through targeted searches of biomedical databases using terms related to smoking cessation, persistent versus nonpersistent molecular alterations, airway and lung epithelium, field cancerization, and longitudinal or cross-sectional study designs. Studies were prioritized based on relevance to persistence dynamics, tissue context, methodological rigor, and reproducibility across independent cohorts. This review examines the molecular landscape of persistent and nonpersistent alterations induced by tobacco exposure, discusses implications for clinical risk reduction strategies, and appraises the potential and limitations of AI-based approaches in advancing precision prevention and personalized approaches to lung cancer control.

2. Smoking-Induced Molecular Changes

Tobacco smoke is a complex mixture of toxic constituents, containing over 7000 chemical compounds, at least 69–80 of which are recognized carcinogens by the International Agency for Research on Cancer (IARC) and U.S. National Cancer Institute [15,29–32]. These include PAHs, tobacco-specific nitrosamines, heavy metals like arsenic and cadmium, and radioactive elements such as polonium-210 [33]. Components of combustible cigarette smoke also include aromatic amines, reactive aldehydes, and benzene, which collectively contribute to DNA damage, oxidative stress, and genomic instability [15]. Nicotine and certain nitrosamines can aberrantly activate cell signaling pathways that act as tumor-promoting rather than directly mutagenic agents, promoting the survival and clonal expansion of damaged epithelial cells [15,18]. These mechanisms reflect exposure to combustion-derived toxicants in cigarette smoke, and therefore, the molecular alterations discussed in this review should not be extrapolated to non-combustible nicotine delivery systems, for which long-term genomic and cancer risk data remain limited. Despite these well-characterized carcinogenic mechanisms, the temporal dynamics of smoking-induced molecular alterations present challenges for risk assessment and intervention.

The clinical consequences of these carcinogenic mechanisms are evident in persistent cancer risk following cessation. Even after 25 years of cessation, lung cancer risk in former smokers remains over three times higher than in never-smokers [34,35]. Analysis of the Framingham Heart Study, encompassing the Original ($n = 3905$) and Offspring ($n = 5002$) cohorts with longitudinal follow-up for smoking exposure and lung cancer incidence from 1954 to 2013, found that 40.8% of lung cancers in former smokers occurred after more than 15 years since quitting, demonstrating the long temporal window where molecular alterations continue to affect cancer risk [35]. Although quitting at any age reduces risk, earlier cessation greatly improves outcomes, as quitting before age 40 lowers the risk of death from tobacco-related disease by ~90% [34]. Cessation at ages 60, 50, 40, or 30 years extends life expectancy by approximately 3, 6, 9, or 10 years, respectively, with cumulative lung cancer incidence declining progressively with longer cessation periods [36].

The persistence of smoking-induced molecular changes is contingent on the nature of the alteration, whether epigenetic, transcriptional, or genetic, and on the genomic loci involved [17,37,38]. While many expression changes normalize after cessation, aberrant DNA methylation at regulatory sites may persist for decades [37,38]. Persistence of smoking-related epigenetic changes appears to be site-specific and not directly dependent on exposure intensity or duration, with certain loci remaining altered for decades after cessation [37,38]. However, cumulative lifetime exposure has been associated with accelerated epigenetic aging, reflecting the broader effects of long-term cigarette smoke exposure [39]. Genetic variation and cigarette smoking independently influence DNA methylation through primarily distinct sets of loci. This differential impact may contribute to interindividual differences in susceptibility to smoking-related disease [40,41]. Collectively, these accumulated epigenetic and genetic alterations in the bronchial epithelium may represent early molecular events in the pathogenesis of smoking-related lung cancers [42,43].

The sensitivity of airway epithelium to tobacco exposure shows a dose-response relationship with no clear threshold for molecular alterations. Low levels of exposure (0.1 ± 0.3 pack-years) produce detectable transcriptomic changes in small airway epithelium, with 34% of differentially expressed genes observed between never-smokers and low-level exposed individuals [44]. The most sensitive genes to tobacco metabolites include *PLA2G10* and *CXCL6*, which respond to nicotine urine levels below 2 ng/mL. Cotinine-responsive genes, such as *CYP2E1* and *GAD1*, show alterations at urine concentrations of approximately 6.2–7.3 ng/mL, demonstrating that molecular damage begins with minimal exposure [44].

2.1. Nonpersistent Molecular Changes

2.1.1. Transcriptomic and Functional Recovery

Many smoking-induced changes are nonpersistent, particularly those in xenobiotic metabolism and acute stress pathways. A core set of nine genes (*CYP1B1*, *ALDH3A1*, *AKR1B10*, *AKR1C1*, *AKR1C2*, *AKR1C3*, *MUC5AC*, *NQO1*, and *SCGB1A1*), identified through a cross-sectional comparison of bronchial epithelium in current and former smokers, consistently return to normal expression levels [45]. Longitudinal studies found that genes related to xenobiotic metabolism (*CYP1A1*, *CYP1B1*, *ALDH3A1*) and homeostasis (*MUC2*, *MUC13*) in the nasal epithelium are among the most rapidly reversible, with expression levels reverting toward baseline within 4 weeks after cessation [46]. 88.2% of smoking-upregulated gene expression changes showed downregulation by 8 weeks, with 11.8% beginning to decrease within 4 weeks, indicating an early reversal trend following smoking cessation [46]. The earliest molecular responses to smoking cessation include rapid epigenetic recovery within months, characterized by widespread changes in DNA methylation at CpG sites

and alterations in cellular stress and metabolic pathways. Short-term cessation studies (3–6 months) reveal global decreases in DNA methylation affecting 3878 CpG sites, with 694 sites showing increased methylation and 3,184 showing decreased methylation [47]. These methylation changes correlate with improved lung function and reduced inflammatory biomarkers, indicating that molecular recovery begins immediately upon cessation.

Inflammation and stress response genes also demonstrate recovery. Examples include *MMP10* in human airway epithelial cells and cytokines such as IL-1 α , TNF- α , CCL2, and CCL3, which normalize in animal studies, alongside immune cell counts [22,48]. Broader metabolic recovery takes longer, as metabolic and antioxidant expression profiles of former smokers resemble never-smokers after ~2 years [22]. Genes involved in nucleotide metabolism, xenobiotic metabolism, and mucus secretion (e.g., *TFF3*, *CABYR*, *ENTPD8*) recover, with partial reversal of *MUC5AC* [49]. The PI3K pathway, an early smoke-responsive signaling axis, has also been found to normalize following months of cessation, enhanced by targeted intervention with myo-inositol treatment [50].

2.1.2. Epigenetic and microRNA Recovery

While most studies focus on airway epithelial changes, systemic effects are also evident in blood-derived markers, demonstrating broader epigenetic recovery after smoking cessation. Although many methylation scars are persistent, a larger fraction revert to baseline levels. Analysis of whole-blood DNA methylation data identified 602 nonpersistent versus 149 persistently differentially methylated CpG sites [38]. Time-dependent reversal patterns of CpGs have also been observed: 32 CpG sites showed significant change within 4 years of cessation and 30 within 5–14 years (10 sites shared between these groups), with only *AHRR* cg26703534 persisting after 14 years [37]. Key sentinel sites, including *AHRR* and *F2RL3*, demonstrate robust reversion toward never-smoker levels in long-term blood-based studies [51]. Circulating gene expression biomarkers complement tissue-based findings, with blood-based analysis identifying 94 nonpersistent genes that normalize to never-smoker levels and 31 genes that revert more slowly out of the 132 smoking-related genes analyzed [52]. Similarly, ~65% of smoking-altered miRNAs in small airway epithelium returned to baseline within 3 months of quitting smoking [53]. Strulovici-Barel et al. reported that 67% of smoking-dysregulated genes reversed within 12 months, while persistent apoptosis and growth-related genes were more resistant [17]. These findings demonstrate that while many molecular alterations regress after cessation, a subset of alterations persist and likely sustain risk. Blood-based and airway-based biomarkers may serve as non-invasive tools for the surveillance of cessation success and long-term molecular damage, with potential applications in population-level risk assessment and screening.

2.2. Persistent Molecular Changes

Smoking leaves behind a range of persistent molecular changes that continue to influence airway biology long after cessation. These include irreversible DNA mutations, sustained shifts in gene expression and regulation, epigenetic reprogramming, and immune or structural remodeling, each described in the subsections below. Figure 1 outlines the timeline of these processes, highlighting nonpersistent changes that recover within months to years versus persistent changes that remain for decades.

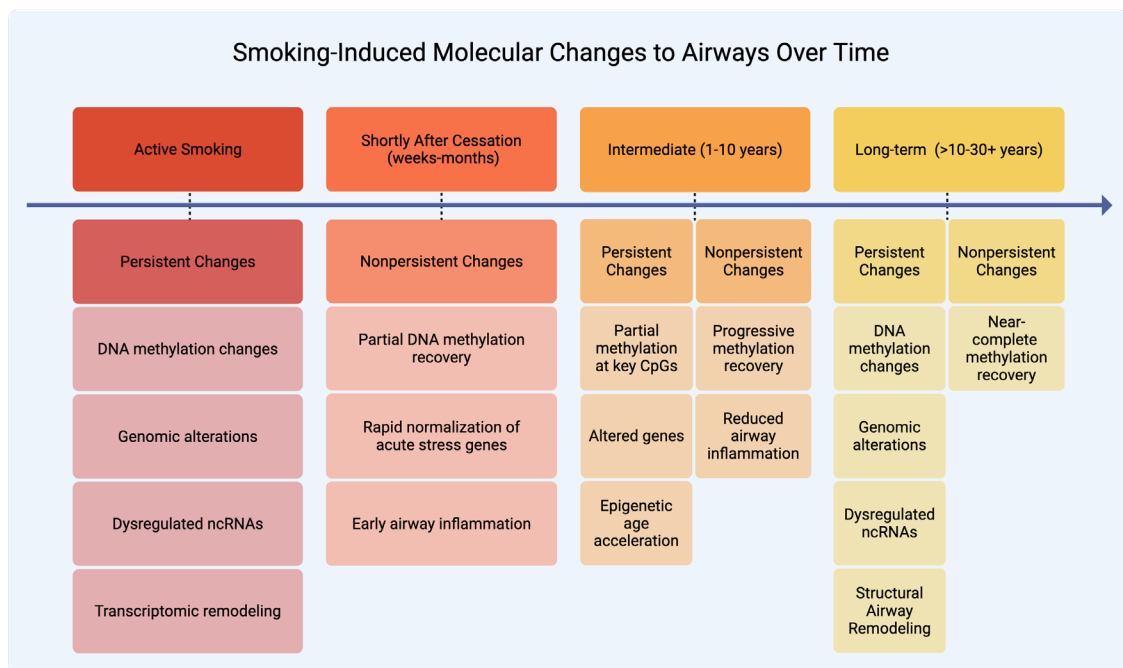


Figure 1. Smoking-Induced Molecular Changes to Airways Over Time. This figure depicts how active smoking and cessation affect airway molecular and cellular changes over time. The red column indicates active smoking, dark orange represents the period shortly after cessation (weeks-months), light orange shows the intermediate period after cessation (1–10 years), and yellow indicates the long-term period after cessation (>10–30+ years). Within each time period, the segments are divided into persistent and nonpersistent changes where applicable. Shortly after smoking cessation, some changes show rapid or partial recovery (nonpersistent), such as the normalization of acute stress gene expression and reduced airway inflammation. Other changes remain for years to decades (persistent), including DNA methylation abnormalities at key CpGs, genomic scars, dysregulated ncRNAs, epigenetic age acceleration, and structural airway remodeling. Figure 1 illustrates how persistent molecular alterations leave lasting molecular marks that contribute to lung cancer risk long after exposure ends.

2.2.1. Genetic Alterations

Structural genetic lesions represent the most permanent consequences of smoking. Studies have found that about 62% of former smokers, with an average cessation period of 27 months, harbor clonal genetic alterations in histologically normal lung tissue [43]. These include loss of heterozygosity at 3p14 (*FHIT*, observed in 75% of informative smokers overall), 9p21 (*CDKN2A*, 57%), and 17p13 (*TP53*, 18%). Among former smokers, LOH at 3p14 was detected in 45%, compared with 88% of current smokers ($p = 0.01$).

Unlike some partially reversible epigenetic and transcriptomic changes, DNA lesions are fundamentally irreversible. Once the mutations occur, they last for the lifetime of that cell. Whole-genome sequencing of 632 single-cell-derived bronchial epithelial colonies from current, former, and never-smokers shows that tobacco exposure adds thousands to tens of thousands of mutations per cell, and that these alterations persist in affected cell lineages [54]. The clonal patches harboring these mutations remain as permanent genomic scars in former smokers.

2.2.2. Gene Expression and Regulatory Changes

Longitudinal small airway epithelium studies show that a subset of smoking-dysregulated genes remains abnormally expressed after cessation. In one 12-month study, 53 (11%) of 475 genes did not normalize, including *CYP1B1*, *PIR*, *ME1*, *TRIM16*, with apoptosis and proliferation pathways most resistant [17]. Spira et al. identified 13 persistently altered genes detectable even 20–30 years post-cessation [22]. These included decreased expression of potential tumor suppressor genes such as *TU3A* and *CX3CL1*, and increased expression of the oncogenes *HN1* and *CEACAM6*. In addition, three metallothionein genes located at 16q13 remained persistently downregulated, suggesting a fragile site for DNA damage in smokers. Beane et al. identified 28 persistently dysregulated genes in large airway epithelium [23]. The persistent down-regulation of genes such as *SULF1*, *UPK1B*, and metallothioneins suggest the clonal selection of altered epithelial cells that maintain smoke-induced molecular changes. MiRNAs also contribute to persistent remodeling. Of 34 small airway epithelium miRNAs altered by smoking, 12 remained dysregulated after 3 months of cessation, including miR-218, miR-133a/b, miR-487b, and miR-1246 [53]. The target genes of these miRNAs are primarily enriched for the Wnt/ β -catenin signaling pathway. In the airway epithelium of current smokers, a self-amplifying *EGFR*–amphiregulin autocrine loop was identified that is absent in never-smokers and drives basal-cell hyperplasia and squamous metaplasia [55]. This smoke-induced feedback maintains *EGFR* activation and may contribute to persistent epithelial remodeling and increased susceptibility to smoking-related lung disease. Distinct miRNA expression patterns in LUAD based on smoking history include 66 miRNAs showing differential alterations: 25 in current smokers, 14 in former smokers, and 27 in never-smokers [56]. These smoking status-specific miRNA networks show prognostic significance and suggest that the molecular impact of smoking influences treatment response and survival outcomes. Tissue-based analysis identified six distinct persistently dysregulated genes—*LEF1*, *ADAMTS1*, *SFXN1*, *CST7*, *CCR7*, and *GNB2L1*—as markers of lasting gene expression changes in former smokers, highlighting differences in biomarker signatures associated with smoking cessation [52].

2.2.3. Epigenetic Modifications

Tobacco smoke induces long-lasting epigenetic alterations, with DNA methylation changes persisting for decades after cessation and influencing key regulatory pathways across the genome. Epigenome-wide association studies using whole-blood DNA have identified 149 CpG sites that remain differentially methylated >35 years post-cessation [38]. Key smoking-associated methylation sites include cg05575921 in the *AHRR* gene, methylation changes in *F2RL3* and *GFI1*, and broader differentially methylated regions such as 6p21.33 on chromosome 6 and 2q37.1 on chromosome 2 [38]. Genome-wide methylation analysis has confirmed the extensive epigenetic impact of smoking across multiple loci, with 972 CpG sites showing significant methylation differences (>5%), and 187 of these CpG sites were replicated in an additional cohort [57]. The sentinel site cg05575921 in *AHRR* demonstrates the highest level of detectable DNA methylation changes, with ~24% hypomethylation in current smokers. The widespread nature of these changes, detected across all autosomes in whole blood, includes altered protein binding at the sentinel site cg05575921 in *AHRR*, suggesting potential effects on transcription factor binding and gene expression regulation. These findings demonstrate that smoking induces broad epigenetic remodeling in blood-derived immune cells and that these alterations may extend beyond traditional cancer-associated genes [57]. Single-cell methylation profiling of bronchial basal progenitors isolated via bronchial brushing reveals persistent genome-wide hypomethylation affecting loci such as *KRAS*, *ROS1*, *CDKN1A*, *CHRNA4*, *CADM1* [42]. Persistent marks

also overlap age-associated CpGs and Polycomb targets, implicating developmental and immune-related pathways consistent with aging-associated epigenetic remodeling [58].

While most persistent alterations are characterized in airway and lung tissue, blood-based profiling may be a less invasive method of assessing the systemic and long-term molecular impacts of smoking. The functional consequences of persistent methylation changes have been demonstrated in prospective cohort studies that examine pre-diagnostic blood samples. Analysis of 796 case-control pairs throughout four independent cohorts showed that hypomethylation at *AHRR* cg05575921 and *F2RL3* cg03636183 was highly associated with future lung cancer risk, with odds ratios of 0.37 (95% CI: 0.31–0.54) and 0.40 (95% CI: 0.31–0.56) per standard deviation increase in methylation, respectively [59]. These associations remain strong after adjusting for smoking status, indicating that methylation alterations may have independent predictive value outside of smoking history alone. Mediation analysis of methylation at these two specific CpG sites in *AHRR* and *F2RL3* estimated that approximately 37% (95% CI: 19–66%) of the total effect of tobacco smoking on lung cancer odds is mediated by methylation at these loci [59]. This suggests that these epigenetic changes may play a causal role rather than serving as exposure biomarkers alone. The authors note that this observation could partly reflect chance or residual confounding, warranting cautious interpretation. On average, lung cancer cases were diagnosed about 3.88–9.6 years after blood collection in NOWAC, MCCS, and NSHDS cohorts, exemplifying the long-term predictive capacity of the persistent methylation changes [59]. Cross-sectional analysis of former smokers at differing time points post-cessation shows that *AHRR* and *F2RL3* methylation levels gradually approach never-smoker levels. The most substantial recovery took place within the first 10 years after quitting, although complete normalization was not achieved in even long-term former smokers.

Epigenetic age acceleration is another component of persistent smoking-induced alterations. Smoking has been found to increase the epigenetic age of airway cells by an average of 4.9 years and of lung tissue by 4.3 years [60]. After cessation, epigenetic age acceleration reversed in airway cells to never-smoker levels, but not in lung tissue. This incomplete reversal suggests that long-lived or slowly renewing cells in the lung retain smoking-induced molecular damage, potentially maintaining a pro-oncogenic tissue environment even after cessation. The clinical relevance of epigenetic changes is further supported by airway-specific methylation patterns in the lung. Former smokers, who had quit at least two years before study inclusion, displayed chronic mucus hypersecretion and increased promoter methylation of lung cancer risk genes, such as *SULF2*, when compared to asymptomatic former smokers [61]. Therefore, persistent respiratory symptoms may be good indicators of lasting epigenetic dysregulation, even after cessation.

2.2.4. Immune and Structural Consequences

Persistent immune dysfunction includes dysregulation of neutrophil-mediated immunity and interferon- γ -related pathways, contributing to elevated lung cancer risk lasting over 10 years after smoking cessation [62]. Animal models confirm structural irreversibility, as elevated IL-12, reduced IL-10, alveolar enlargement, right ventricular hypertrophy, and ongoing inflammation persisted 8 weeks after smoke exposure ended in A/J mice [48]. In longer-term mouse models, neutrophilic inflammation, macrophage accumulation, and destructive changes consistent with lung remodeling persisted six months after smoking cessation [63]. Chronic smoke-exposed mice also showed progressive alveolar damage and inflammation that lasted longer than exposure periods, suggesting ongoing structural deterioration after cessation [63]. In humans, adaptive immune changes appear more persistent than innate alterations. While innate responses normalize, cytokine response patterns from T cells stay altered in former smokers, potentially linked to epigenetic memory [64].

3. Role of AI in Advancing Strategies for Prevention and Intervention

AI has rapidly evolved from a niche computational tool into a widely used approach across science and medicine. Over the past decade, continuous advances in algorithms, computing power, and data accessibility have driven rapid growth in AI capabilities and applications. In oncology, AI has shown promising proof-of-concept success in improving early cancer detection, predicting treatment responses, and identifying molecular or imaging-based biomarkers, offering potential to accelerate research and support clinical decision-making [26,27,65,66].

ML and deep learning (DL) models provide computational frameworks to capture complex biological signals beyond single-gene markers [27,65]. Rather than relying on a handful of marker genes, these AI-driven approaches analyze integrated multi-omics datasets, including transcriptomic, epigenomic, proteomic, and metabolomic profiles, to identify composite molecular signatures that reflect coordinated changes across numerous features [67,68]. Such signatures may strengthen the ability to distinguish persistent from nonpersistent smoking-induced alterations. These methods aim to improve interpretability and generalizability by modeling coordinated biological signals rather than isolated features [26]. By applying these approaches to smoking-related airway and lung datasets, AI may support early detection by identifying individuals at elevated risk, enabling personalized risk stratification in former smokers, and revealing nonpersistent molecular pathways that could be targeted with early interventions [8,16,65]. By doing so, AI offers an outline for operationalizing the persistent and nonpersistent framework to improve prevention, monitoring, and treatment strategies [20,24,26].

Within the context of tobacco-associated lung carcinogenesis, one major challenge lies in distinguishing persistent from nonpersistent molecular alterations induced by smoking exposure [17,23,42,54]. This distinction is essential for understanding why some molecular changes revert after smoking cessation while others persist as durable molecular alterations that sustain cancer risk, and for developing effective, personalized prevention and intervention strategies [8,16]. However, efforts to characterize persistent and nonpersistent alterations generate extensive and complex datasets that are difficult to interpret with conventional analytical methods [26]. Current approaches are limited by fragmented multi-omics signals that lack integration across datasets, heterogeneous cohorts, incomplete labeling of persistence (generally labeled via longitudinal data showing temporal stability of alterations), and the barriers of modeling molecular recovery as a dynamic process [23,37,40,69]. The interplay of genetic susceptibility, cumulative exposure, and time since cessation further complicates the identification of truly causal, persistent alterations within a landscape of reversible changes [8,35,54].

To address these challenges, AI-driven approaches can be organized into a structured workflow that links molecular data with clinical application. Figure 2 provides an overview of this framework, beginning with data integration and multi-omics analysis, progressing through molecular signature identification, validation and model refinement, and extending into clinical decision support. Table 1 complements this figure by summarizing representative datasets, sequencing platforms, AI models, and software tools corresponding to each stage [70–122]. The example tools and models listed in Table 1 represent commonly used approaches across lung cancer risk prediction and multimodal analysis; persistence-specific applications remain comparatively limited and are discussed further in Section 4. Together, Table 1 and Figure 2 illustrate how AI may serve as a bridge between complex biological data and actionable strategies for lung-cancer prevention and intervention.

Table 1. Representative datasets, sequencing platforms, AI models, and bioinformatics tools across the lung-cancer prevention and precision-oncology workflow. These resources correspond to the stages illustrated in Figure 2 and exemplify how genomic, imaging, and clinical data can be integrated through AI-driven pipelines for biomarker discovery, risk stratification, and clinical decision support [70–122].

Workflow Step	Datasets, Models, or Platforms	Primary Role
Data Acquisition & Integration	TCGA/Genomic Data Commons	Large-scale multimodal genomic and clinical public datasets [70,71].
	LIDC-IDRI/LUNA16/EMRs	Annotated lung nodule CT datasets and electronic medical records for training and validation [72,73].
	PacBio HiFi, Oxford Nanopore Duplex, Illumina platforms	Long- and short-read sequencing for variant and methylation multi-omic profiling [74–77].
Preprocessing & QC	GATK, SAMtools	Standard pipelines for variant calling and quality-control analysis [78,79].
	PLINK	Genotype data management and association analysis tool [80].
Prevention & Risk Stratification	Sybil	DL model predicting lung-cancer risk from LDCT [81,82].
	XGBoost, PLCOm2012	Gradient-boosting and statistical models for clinical risk prediction [83,84].
AI/ML Frameworks & Core Methods	DL architectures, Transformer Networks, Graph Neural Networks	Neural-network approaches for pattern recognition and modeling of molecular networks and pathways across multimodal data [85–87].
	Radiomics pipelines	Quantitative feature extraction from medical imaging to characterize tumor phenotypes [88].
Multi-Omics Integration & Network Analysis	WGCNA, MOFA+/mixOmics	Co-expression network and multi-omic factor analysis for integrated profiling [89–91].
	PANDAOmics	AI-driven commercial platform for drug target discovery integrating multi-omic data [92,93].
Segmentation & Image Processing	U-Net, ConvPath	Automated CT and whole-slide image segmentation with CNNs [94,95].
	3D Slicer, PyRadiomics	Extraction of quantitative radiomic features from segmented regions [96,97].
Diagnosis & Lesion Detection	Lunit INSIGHT CXR	AI-based detection of pulmonary nodules and lesions in chest radiographs [98,99].
	Paige.AI, CLAM, TIAToolbox	AI platforms and frameworks for automated histopathological image analysis [100–102].
Biomarker Discovery & Pathway Analysis	QIAGEN IPA, GSEA	Pathway enrichment and functional annotation of gene signatures [103,104].
	clusterProfiler, Cytoscape	Functional enrichment and network visualization of molecular interactions [105,106].
	DeepNovo, PepNet, PepFormer	DL tools for peptide sequencing and neoantigen discovery in immunotherapy [107–110].

Table 1. Cont.

Workflow Step	Datasets, Models, or Platforms	Primary Role
Clinical Decision Support	Tempus Lens, FoundationOne CDx, Caris MI	AI-enabled decision-support platforms integrating genomic and clinical data for treatment selection [111–113].
	cBioPortal	Interactive platform for exploring multidimensional cancer-genomics data [114].
Validation & Model Refinement	PrecisionFDA	Regulatory benchmarking and reproducibility testing for genomic pipelines [115].
	IGV (Integrative Genomics Viewer)	Visualization tool for validation of variants and expression patterns [116].
General AI/Development Frameworks	QuPath, MONAI, OpenSlide	Open-source libraries for digital pathology and scalable image analysis [117–119].
	scikit-learn, TensorFlow, PyTorch	Core ML libraries for model development and deployment [120–122].

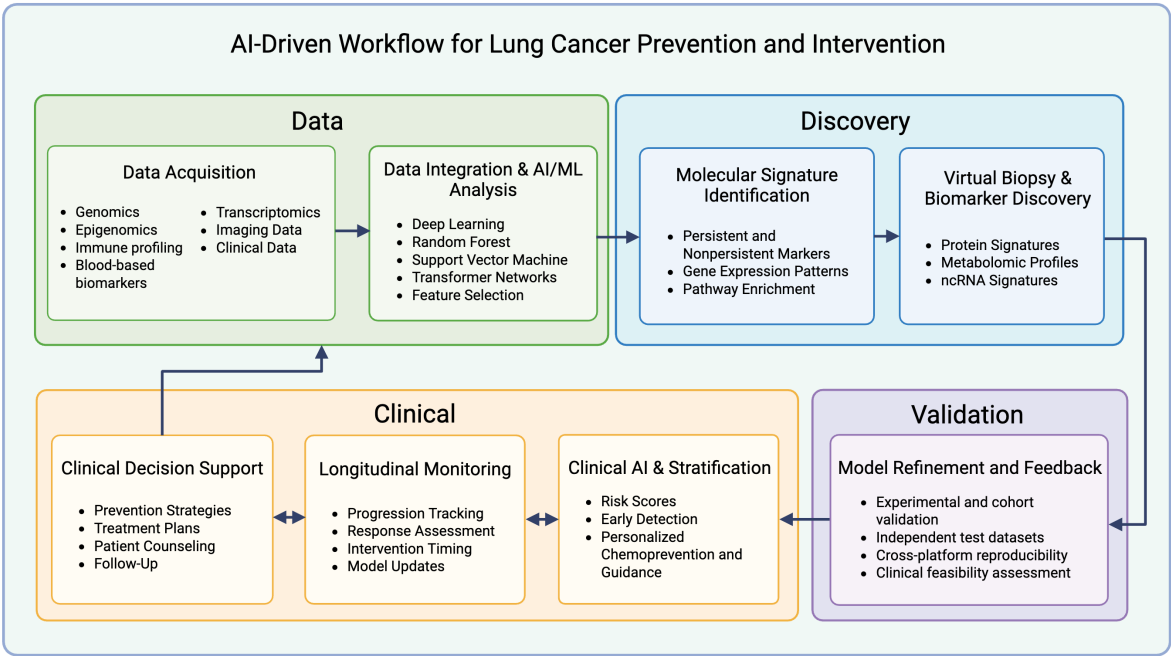


Figure 2. AI-Driven Workflow for Lung Cancer Prevention and Intervention. This schematic illustrates an eight-step workflow linking multi-source data acquisition, AI/ML analysis, molecular signature identification, validation and model refinement, and virtual biomarker discovery. Data inputs span genomics, epigenomics, transcriptomics, imaging, immune profiling, and blood-based biomarkers. Clinical insights from these analyses support risk stratification, personalized prevention, and treatment planning. Validated discovery outputs transition into clinical application, while longitudinal monitoring and bidirectional feedback between data integration, validation, monitoring, and decision support enable continuous model refinement and improved clinical relevance, highlighting approaches to distinguish persistent from reversible tobacco-related molecular changes in lung cancer.

3.1. Identifying Molecular Signatures

AI-based ML and DL models are increasingly used to identify novel biomarkers and analyze complex biological datasets by leveraging coordinated changes across omics patterns rather than relying on single markers [27,123]. DL performs well at discerning complicated patterns in large datasets, which is optimal for the interpretation of smoking-induced molecular heterogeneity and for cohorts with incomplete or indirect labeling

of persistence [26,124]. While the direct application of these models to differentiate between persistent and nonpersistent smoking-induced changes remains an emerging area of research, previous foundational work in oncology demonstrates the feasibility of this approach [125,126].

An important capability of AI is the ability to reclassify tumors with ambiguous features into well-defined molecular categories. For example, a DL model was successfully able to reclassify combined hepatocellular-cholangiocarcinoma (cHCC-CCA), which is a rare and biphenotypic cancer, into more distinct hepatocellular carcinoma (HCC) or intrahepatic cholangiocarcinoma (ICCA) categories [126]. This morphological reclassification was validated by its strong correlation with distinct spatial gene expression profiles and genetic alterations (e.g., TERT, CTNNB1, FGFR2) establishing that AI can connect histological patterns with functionally distinct molecular states [126]. This application may provide conceptual precedent for distinguishing persistent, high-risk molecular states from nonpersistent states in the airways of former smokers.

AI has also been applied to identify subtle but functionally important genetic alterations within broad genomic patterns [27,125]. For example, the DL tool “Dig” maps genome-wide somatic mutation rates and identifies driver mutations under positive selection by comparing the observed mutation counts to predicted neutral rates [125]. This tool has been used to reveal significant mutations in splice sites and 5′ untranslated regions that are linked to altered gene expression and frequently overlooked by traditional methods. This research sets a precedent for using AI to uncover specific, functionally significant and persistent mutations within a landscape of neutral or transient variation [125].

Together, these advances demonstrate that AI can identify biomarkers based on complex omics signatures that reflect coordinated alterations across molecular features. By extension, an adequately trained deep neural network could help elucidate the molecular signatures that differentiate persistent from nonpersistent changes in the airway epithelium of smokers. Numerous ML algorithms, such as Support Vector Machines (SVM), Random Forests (RF), and the Least Absolute Shrinkage and Selection Operator (LASSO), are often used for feature selection and classification, enabling the identification of salient molecular alterations from large-scale datasets [123,127,128].

3.2. Integration of Multi-Omics Data

AI is also a powerful tool for integrating multi-omics data, which is necessary for a more holistic understanding of smoking-induced damage. Conventional approaches face difficulty when interpreting signals across genomics, transcriptomics, imaging, and clinical layers, highlighting the need for models that can discern results from fragmented data [129,130]. Transformer neural networks and advanced DL models are adept at synthesizing multimodal data such as imaging with pathology or genomics data [26,131]. These networks may be valuable for linking persistent molecular alterations with early neoplastic changes observed in histopathologic or radiologic images. Research has shown that AI-driven integration of multi-omics data can refine molecular subtypes, predict prognosis, and identify therapeutic responses in lung cancer [67,124,132]. A key advantage of these models is their ability to identify biomarkers as shifts in coordinated omics patterns across multiple layers, rather than changes in a few isolated genes [65,123]. These integrative models may help reveal how persistent epigenetic marks, such as DNA methylation at AHRR and F2RL3 loci, interact with transcriptomic changes to sustain a pro-tumorigenic microenvironment long after an individual quits smoking [21,38].

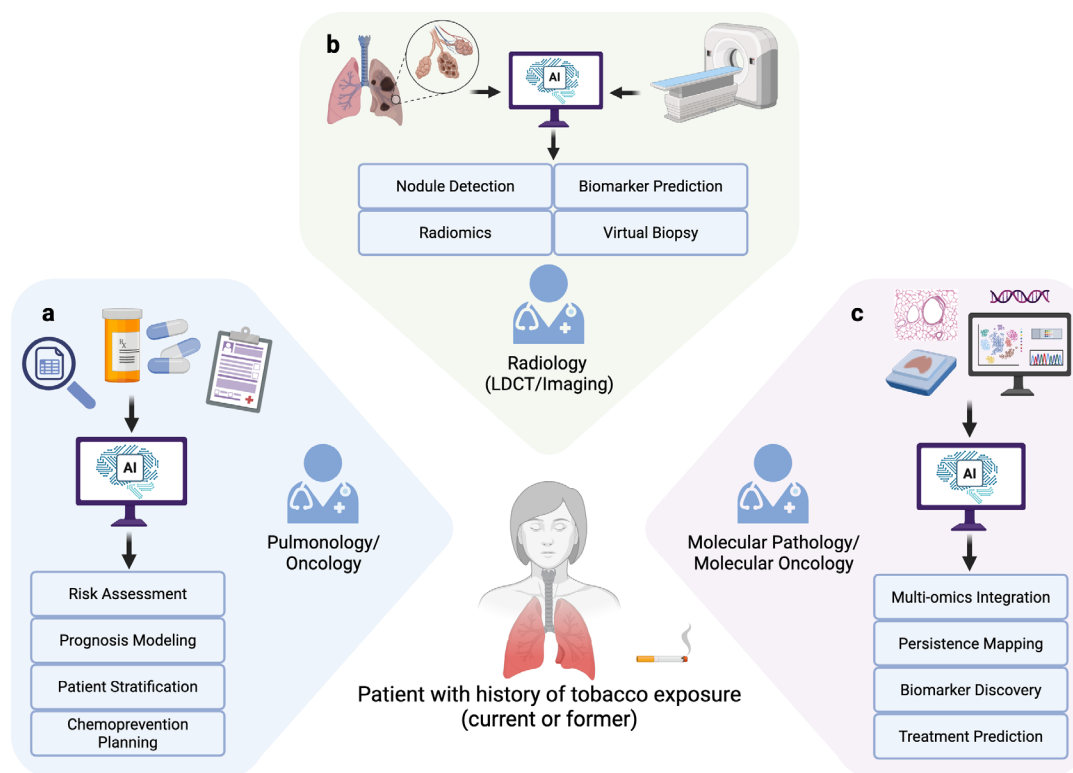
Traditional bisulfite sequencing remains a gold standard for DNA methylation analysis but involves DNA degradation and limited ability to discriminate between different epigenetic modifications [133,134]. The recent availability of 5-base sequencing technolo-

gies, such as PacBio HiFi, Oxford Nanopore duplex, and Illumina's 5-Base Solution further strengthens these integrative approaches [77,135,136]. Long-read platforms enable the direct detection of multiple base modifications (e.g., 5mC, 5hmC) at single-molecule resolution, while Illumina's short-read approach presents a parallel method for simultaneous detection of genomic variants and cytosine methylation in one workflow [77,135,136]. Incorporating this additional layer of epigenetic information into multi-omics frameworks enhances the ability of AI-driven models to identify biomarkers as coordinated omics patterns, providing a more comprehensive view of persistent versus nonpersistent smoking-induced alterations.

3.3. Acceleration of Biomarker Development and "Virtual Biopsies"

AI can also be used to accelerate the development of biomarkers for risk stratification and chemoprevention [27,137,138]. By analyzing longitudinal molecular and imaging data from former smokers, ML models may identify molecular signatures associated with long-term cancer risk [26,137]. For instance, a DL model could be trained on sequential multi-omics and imaging profiles to predict which individuals are likely to follow a persistence-prone molecular trajectory and would therefore benefit from targeted chemopreventive interventions [27,138].

These advanced AI-based models can non-invasively predict molecular features from routine clinical data or images, providing a "virtual biopsy" that may be used to monitor molecular changes over time [26,139]. Virtual biopsies typically involve training AI models on paired radiologic and molecular data; once trained, these models can predict molecular alterations solely from radiological or histopathological images [139–142]. In practice, these models use paired datasets where imaging-derived features are matched to molecular readouts (e.g., mutation status or expression-related biomarkers), enabling prediction of molecular states from images in independent samples [140–142]. When used as adjunctive tools rather than replacements for histopathology, such approaches could enhance low-dose computed tomography (LDCT) screening by linking radiographic findings with molecular signatures and potentially reducing the need for unnecessary invasive procedures [81,82]. Previous research has demonstrated that DL models can predict numerous clinically relevant mutations, such as EGFR, STK11, and KRAS, from H&E-stained pathology slides in lung cancer [141]. This function has been extended to radiologic images, where AI models analyze CT scans to predict driver mutations, like EGFR, and the expression of immunotherapy biomarkers, such as PD-L1 [142,143]. Beane et al. demonstrated the potential for identifying transcriptomic biomarkers by developing a highly accurate classifier from 28 persistently dysregulated genes that could classify former and current smokers [23]. These findings indicate that airway gene-expression patterns may serve as sensitive indicators of prior exposure and long-term risk, and that AI-based models could refine such signatures to distinguish transient from enduring molecular damage. However, the clinical deployment of these approaches requires careful validation and should be viewed as complementary to established diagnostic standards. Together, these applications demonstrate how AI could enhance lung cancer patient care across clinical, imaging, and molecular domains. This multidisciplinary framework is summarized in Figure 3, illustrating AI uses for (a) risk prediction in smokers without cancer, (b) radiologic assessment and virtual biopsies, and (c) mapping persistent molecular alterations in diagnostic or surveillance contexts.



AI-assisted roles in lung cancer prevention and management

Figure 3. AI-assisted roles in lung cancer prevention and management. A patient with a history of tobacco use can be evaluated across several disciplines with the use of AI. (a) For individuals at risk, pulmonologists and oncologists can integrate smoking history, comorbidities, and molecular profiles with AI to predict long-term risk, stratify susceptibility, and guide personalized prevention strategies. (b) For screening and early detection, radiologists can apply AI to LDCT and PET imaging for automated nodule detection, segmentation, and radiomic analysis, enabling prediction of persistent structural patterns and non-invasive “virtual biopsies.” (c) For confirmed lung cancer cases, molecular pathologists can employ AI to integrate multi-omics, single-cell, and spatial data to distinguish persistent from reversible (persistence mapping) smoking-related changes. These approaches also support the prediction of treatment-relevant biomarkers that may inform future precision-oncology or chemoprevention strategies, potentially accelerating early detection and prevention efforts.

4. Limitations, Generalizability, and Future Directions

Despite rapid technological advancements, translating AI-based multi-omics approaches for smoking-related persistence into clinical applications remains limited by biological heterogeneity, data availability, imperfect persistence labels, and methodological constraints [26,28,70]. This section summarizes the primary sources of bias and uncertainty that affect generalizability, interpretability, and clinical relevance, and outlines priorities for future research.

4.1. Cohort Heterogeneity and Generalizability

Most datasets used to assess smoking-related molecular persistence are ancestry- and geography-biased, and baseline methylation or expression profiles differ across populations, limiting model portability [37,41,70]. Differences in smoking intensity, cumulative exposure, and time since cessation also contribute to heterogeneity and may shift molecular persistence trajectories in ways that are inconsistently captured across studies [8,37]. In addition, biospecimen type strongly influences the biology being measured; persistence signatures derived from blood, airway brushings, or lung tissue may reflect distinct cellular

processes and may not replicate across sample types [41,70]. Genetic background can also affect smoking-associated epigenetic responses, introducing inter-individual variability that may challenge transferability when AI models are trained on a single cohort [41]. To improve generalizability, AI models should be developed on sufficiently large and diverse cohorts and evaluated in independent populations. The use of publicly available resources and reproducible workflows is also critical so that findings can be validated across studies [26,28,70].

4.2. Interpretation, Causality, and Clinical Relevance

A key limitation of current persistence frameworks is that most proposed molecular signatures remain observational, and strong statistical associations do not establish whether a marker is mechanistically involved in carcinogenesis or reflects long-lasting exposure history [8,26]. This limitation is especially relevant for persistent epigenetic markers and composite AI-derived signatures that integrate numerous correlated features [26,70]. As a result, AI-based approaches should be framed primarily as tools for risk stratification and hypothesis generation, with causal claims requiring independent functional validation [26,28]. Interpretability of AI models should also be prioritized to link predictions back to underlying biology and clinical context, especially in prevention and early-detection settings [28,139]. Clinical use cases should also be defined conservatively, as histopathology remains the diagnostic gold standard. Virtual biopsy approaches are best positioned as adjunctive decision-support tools rather than replacements, particularly in screening contexts where false positives and downstream harms are clinically meaningful [26,28].

4.3. Technical and Methodological Constraints of AI Models

Many AI models applied to lung cancer genomics are underpowered, lack external validation, and rely on performance metrics that may not directly translate to clinical decision-making [26,28,139]. These challenges are amplified in research of smoking-related molecular persistence, as labels often depend on longitudinal follow-up, reducing effective sample sizes and increasing missingness across omics layers [70]. Managing missing data points remains a major technical challenge, since variable assay availability and quality control issues across multi-omics datasets can degrade model performance. Using robust missing data handling techniques, such as imputation or model architectures resilient to missingness, is essential to maintain accuracy and generalizability of AI models [70]. In addition, persistent smoking-associated molecular alterations may arise from clonal expansion of long-lived altered cell populations, so “ground truth” labels often reflect complex mixtures of cell states instead of discrete molecular categories. This biological complexity may reduce model stability when training data are weakly labeled or heterogeneous [55].

4.4. Future Directions

Future progress will require larger, more diverse longitudinal cohorts, standardized operational definitions of molecular persistence, integrations with functional and experimental validation, and prospective evaluation prior to clinical integration [8,26,28,70]. Longitudinal sampling is particularly important for modeling molecular recovery as a dynamic process and for distinguishing transient from persistent smoking-induced alterations [8,23]. When possible, considering genetic susceptibility, exposure history, and tissue context will be necessary for creating generalizable models across populations and biospecimens [37,41]. These priorities have been highlighted in recent reviews, which emphasize the need for explainable AI, robust validation, and cautious clinical framing before adoption [26,28,138]. With these priorities in place, the rapid emergence of multimodal frameworks, explainable architectures, and large collaborative datasets establish a strong

foundation for the use of AI tools in advancing prevention, monitoring, and early detection in tobacco-related lung cancers.

5. Conclusions

Tobacco exposure causes widespread molecular alterations across the airway and lung tissues. While some of these changes undergo partial or complete recovery after smoking cessation, others last for decades and continue to influence disease risk. Persistent alterations, such as aberrant DNA methylation, impaired DNA repair responses, and immune dysregulation, may help explain the lasting vulnerability of former smokers, while ongoing exposure sustains elevated risk in current smokers. Nonpersistent alterations demonstrate the rapid biological benefits of smoking cessation and identify potential avenues for chemoprevention prior to malignant transformation.

Recognizing which molecular changes are persistent and nonpersistent is essential for refining early detection, risk stratification, and prevention strategies. AI-based methods that combine multi-omics, radiologic, and pathologic data can identify persistence signatures, accelerate biomarker discoveries, and improve individual risk profiling beyond the capabilities of traditional analyses. However, the clinical translation of these approaches remains constrained by cohort heterogeneity, limited longitudinal validation, and challenges related to causal inference and model interpretability. The evidence synthesized in this review supports that persistent smoking-induced molecular alterations represent a distinct biological state that is not solely captured by smoking status. Clarifying which alterations endure across tissues and over time helps define the limits of current prevention strategies, signifying where improved longitudinal and mechanistic studies are most needed.

By integrating longitudinal, single-cell, and spatial multi-omics datasets along with environmental, genetic, and immune factors, AI can model recovery as a dynamic process and improve predictions of smoking-related molecular persistence. Addressing these limitations through diverse cohorts, standardized definitions of persistence, and prospective validation will be essential prior to the integration of AI frameworks in routine lung cancer prevention or screening strategies. With continued progress, biological and computational insights can be translated into practical tools that guide screening, surveillance, and chemopreventive strategies, ultimately reducing the burden of tobacco-related lung cancer and improving outcomes for the millions of current and former smokers worldwide.

Author Contributions: K.H.B., V.G.P.S. and W.L.L. initiated the project. K.H.B., V.G.P.S., G.L.S., K.S.S.E. and W.L.L. designed, researched, analyzed, and wrote about the topics covered in the article. K.H.B. designed the figures. W.L.L. and K.S.S.E. are the principal investigators. K.H.B., V.G.P.S., G.L.S., K.S.S.E. and W.L.L. revised the manuscript. All authors contributed to data interpretation and manuscript preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funds from the Canadian Institutes of Health Research (CIHR, FRN-143345, 183775, and FRN-204063), the Terry Fox Foundation, the Lotte and John Hecht Memorial Foundation, and the BC Cancer Foundation. V.G.P.S. was supported by the BC Cancer Rising Stars Fellowship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: Figures were generated using Biorender.com.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study, the collection, analyses, or interpretation of the data, the writing of the manuscript, or the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
AI	Artificial Intelligence
AHRR	Aryl Hydrocarbon Receptor Repressor
ALDH	Aldehyde Dehydrogenase
CDKN2A/p16	Cyclin Dependent Kinase Inhibitor 2A
cHCC-CCA	Combined Hepatocellular-Cholangiocarcinoma
CI	Confidence Interval
CNN	Convolutional Neural Network
CpG	Cytosine–Phosphate–Guanine Dinucleotide
CT	Computed Tomography
CYP	Cytochrome P450
DL	Deep Learning
DNA	Deoxyribonucleic Acid
EMR	Electronic Medical Record
EGFR	Epidermal Growth Factor Receptor
EWAS	Epigenome-Wide Association Study
F2RL3	Coagulation Factor II (Thrombin) Receptor-Like 3
FHIT	Fragile Histidine Triad
GATK	Genome Analysis Toolkit
GDC	Genomic Data Commons
GSEA	Gene Set Enrichment Analysis
H&E	Hematoxylin and Eosin
IARC	International Agency for Research on Cancer
IGV	Integrative Genomics Viewer
IPA	Ingenuity Pathway Analysis
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LDCT	Low-Dose Computed Tomography
LIDC-IDRI	Lung Image Database Consortium – Image Database Resource Initiative
LOH	Loss of Heterozygosity
LUNA16	Lung Nodule Analysis 2016 Dataset
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MCCS	Melbourne Collaborative Cohort Study
miRNA	MicroRNA
ML	Machine Learning
MOFA+	Multi-Omics Factor Analysis Plus
MONAI	Medical Open Network for Artificial Intelligence
ncRNA	Noncoding RNA
NNK	4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone
NOWAC	Norwegian Women and Cancer Study
NSCLC	Non-Small-Cell Lung Cancer
NSHDS	Northern Sweden Health and Disease Study
OR	Odds Ratio
PAH	Polycyclic Aromatic Hydrocarbon
PD-L1	Programmed Death-Ligand 1

PET	Positron Emission Tomography
PI3K	Phosphoinositide 3-Kinase
PLCOm2012	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial 2012 Risk Prediction Model
QC	Quality Control
RF	Random Forest
RNA	Ribonucleic Acid
SCLC	Small-Cell Lung Cancer
SVM	Support Vector Machine
TCGA	The Cancer Genome Atlas
TP53	Tumor Protein p53
U-Net	Convolutional Neural Network Architecture for Biomedical Image Segmentation
WGCNA	Weighted Gene Co-expression Network Analysis
XGBoost	eXtreme Gradient Boosting

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
2. American Cancer Society. *Cancer Facts & Figures 2025*; American Cancer Society: Atlanta, GA, USA, 2025. Available online: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2025/2025-cancer-facts-and-figures-acf.pdf> (accessed on 15 September 2025).
3. Pesch, B.; Kendzia, B.; Gustavsson, P.; Jöckel, K.; Johnen, G.; Pohlabein, H.; Olsson, A.; Ahrens, W.; Gross, I.M.; Bröske, I.; et al. Cigarette smoking and lung cancer—Relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int. J. Cancer* **2012**, *131*, 1210–1219. [CrossRef]
4. Kim, S.Y.; Park, H.S.; Chiang, A.C. Small Cell Lung Cancer: A Review. *JAMA* **2025**, *333*, 1906. [CrossRef]
5. World Health Organization. Lung Cancer. 2023. Available online: <https://www.who.int/news-room/fact-sheets/detail/lung-cancer> (accessed on 25 October 2025).
6. World Health Organization. *WHO Global Report on Trends in Prevalence of Tobacco Use 2000–2030*; Report No.: 978-92-4-008828-3; World Health Organization: Geneva, Switzerland, 2024; pp. 1–128. Available online: <https://www.who.int/publications/i/item/9789240088283> (accessed on 15 September 2025).
7. Reitsma, M.; Kendrick, P.; Anderson, J.; Arian, N.; Feldman, R.; Gakidou, E.; Gupta, V. Reexamining Rates of Decline in Lung Cancer Risk after Smoking Cessation. A Meta-analysis. *Ann. Am. Thorac. Soc.* **2020**, *17*, 1126–1132. [CrossRef] [PubMed]
8. Kondo, K.K.; Rahman, B.; Ayers, C.K.; Relevo, R.; Griffin, J.C.; Halpern, M.T. Lung cancer diagnosis and mortality beyond 15 years since quit in individuals with a 20+ pack-year history: A systematic review. *CA Cancer J. Clin.* **2024**, *74*, 84–114. [CrossRef]
9. Cipollina, C.; Bruno, A.; Fasola, S.; Cristaldi, M.; Patella, B.; Inguanta, R.; Vilasi, A.; Aiello, G.; La Grutta, S.; Torino, C.; et al. Cellular and Molecular Signatures of Oxidative Stress in Bronchial Epithelial Cell Models Injured by Cigarette Smoke Extract. *Int. J. Mol. Sci.* **2022**, *23*, 1770. [CrossRef]
10. Kode, A.; Yang, S.-R.; Rahman, I. Differential effects of cigarette smoke on oxidative stress and proinflammatory cytokine release in primary human airway epithelial cells and in a variety of transformed alveolar epithelial cells. *Respir. Res.* **2006**, *7*, 132. [CrossRef]
11. Halvorsen, A.R.; Silwal-Pandit, L.; Meza-Zepeda, L.A.; Vodak, D.; Vu, P.; Sagerup, C.; Hovig, E.; Myklebost, O.; Børresen-Dale, A.-L.; Brustugun, O.T.; et al. TP53 Mutation Spectrum in Smokers and Never Smoking Lung Cancer Patients. *Front. Genet.* **2016**, *7*, 85. [CrossRef] [PubMed]
12. Pfeifer, G.P.; Denissenko, M.F.; Olivier, M.; Tretyakova, N.; Hecht, S.S.; Hainaut, P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **2002**, *21*, 7435–7451. [CrossRef] [PubMed]
13. Gibbons, D.L.; Byers, L.A.; Kurie, J.M. Smoking, p53 mutation, and lung cancer. *Mol. Cancer Res. MCR* **2014**, *12*, 3–13. [CrossRef]
14. Jarmalaite, S.; Kannio, A.; Anttila, S.; Lazutka, J.R.; Husgafvel-Pursiainen, K. Aberrant p16 promoter methylation in smokers and former smokers with nonsmall cell lung cancer. *Int. J. Cancer* **2003**, *106*, 913–918. [CrossRef]
15. Centers for Disease Control and Prevention (US); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US). *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease*; Publications and Reports of the Surgeon General; Centers for Disease Control and Prevention (US): Atlanta, GA, USA, 2010; ISBN 978-0-16-084078-4. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK53017/> (accessed on 5 November 2025).

16. Moghaddam, S.J.; Savai, R.; Salehi-Rad, R.; Sengupta, S.; Kammer, M.N.; Massion, P.; Beane, J.E.; Ostrin, E.J.; Priolo, C.; Tennis, M.A.; et al. Premalignant Progression in the Lung: Knowledge Gaps and Novel Opportunities for Interception of Non-Small Cell Lung Cancer. An Official American Thoracic Society Research Statement. *Am. J. Respir. Crit. Care Med.* **2024**, *210*, 548–571. [CrossRef]
17. Strulovici-Barel, Y.; Rostami, M.R.; Kaner, R.J.; Mezey, J.G.; Crystal, R.G. Serial Sampling of the Small Airway Epithelium to Identify Persistent Smoking-dysregulated Genes. *Am. J. Respir. Crit. Care Med.* **2023**, *208*, 780–790. [CrossRef]
18. Du, B.; Leung, H.; Khan, K.M.F.; Miller, C.G.; Subbaramaiah, K.; Falcone, D.J.; Dannenberg, A.J. Tobacco Smoke Induces Urokinase-Type Plasminogen Activator and Cell Invasiveness: Evidence for an Epidermal Growth Factor Receptor–Dependent Mechanism. *Cancer Res.* **2007**, *67*, 8966–8972. [CrossRef] [PubMed]
19. Kadara, H.; Wistuba, I.I. Field Cancerization in Non–Small Cell Lung Cancer: Implications in Disease Pathogenesis. *Proc. Am. Thorac. Soc.* **2012**, *9*, 38–42. [CrossRef]
20. Korde, A.; Ramaswamy, A.; Anderson, S.; Jin, L.; Zhang, J.; Hu, B.; Velasco, W.V.; Diao, L.; Wang, J.; Pisani, M.A.; et al. Cigarette smoke induces angiogenic activation in the cancer field through dysregulation of an endothelial microRNA. *Commun. Biol.* **2025**, *8*, 511. [CrossRef]
21. Kaur, G.; Begum, R.; Thota, S.; Batra, S. A systematic review of smoking-related epigenetic alterations. *Arch. Toxicol.* **2019**, *93*, 2715–2740. [CrossRef] [PubMed]
22. Spira, A.; Beane, J.; Shah, V.; Liu, G.; Schembri, F.; Yang, X.; Palma, J.; Brody, J.S. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10143–10148. [CrossRef] [PubMed]
23. Beane, J.; Sebastiani, P.; Liu, G.; Brody, J.S.; Lenburg, M.E.; Spira, A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.* **2007**, *8*, R201. [CrossRef]
24. Keith, R.L.; Miller, Y.E. Lung cancer chemoprevention: Current status and future prospects. *Nat. Rev. Clin. Oncol.* **2013**, *10*, 334–343. [CrossRef]
25. Daneshkhan, A.; Prabhala, S.; Viswanathan, P.; Subramanian, H.; Lin, J.; Chang, A.S.; Bharat, A.; Roy, H.K.; Backman, V. Early detection of lung cancer using artificial intelligence-enhanced optical nanosensing of chromatin alterations in field carcinogenesis. *Sci. Rep.* **2023**, *13*, 13702. [CrossRef]
26. Zhu, E.; Muneer, A.; Zhang, J.; Xia, Y.; Li, X.; Zhou, C.; Heymach, J.V.; Wu, J.; Le, X. Progress and challenges of artificial intelligence in lung cancer clinical translation. *npj Precis. Oncol.* **2025**, *9*, 210. [CrossRef] [PubMed]
27. Çalışkan, M.; Tazaki, K. AI/ML advances in non-small cell lung cancer biomarker discovery. *Front. Oncol.* **2023**, *13*, 1260374. [CrossRef]
28. Prelaj, A.; Miskovic, V.; Zanitti, M.; Trovo, F.; Genova, C.; Viscardi, G.; Rebuzzi, S.E.; Mazzeo, L.; Provenzano, L.; Kosta, S.; et al. Artificial intelligence for predictive biomarker discovery in immuno-oncology: A systematic review. *Ann. Oncol.* **2024**, *35*, 29–65. [CrossRef] [PubMed]
29. National Toxicology Program, U.S. Department of Health and Human Services. *Report on Carcinogens*, 15th ed.; Public Health Service; National Toxicology Program: Research Triangle Park, NC, USA, 2021. Available online: <https://ntp.niehs.nih.gov/research/assessments/cancer/roc> (accessed on 20 September 2025).
30. Office on Smoking and Health (US). *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General*; Publications and Reports of the Surgeon General; Centers for Disease Control and Prevention (US): Atlanta, GA, USA, 2006. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK44324/> (accessed on 20 September 2025).
31. Office of the Surgeon General (US); Office on Smoking and Health (US). *The Health Consequences of Smoking: A Report of the Surgeon General*; Reports of the Surgeon General; Centers for Disease Control and Prevention (US): Atlanta, GA, USA, 2004. Available online: <http://www.ncbi.nlm.nih.gov/books/NBK44695/> (accessed on 20 September 2025).
32. U.S. Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*; Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health: Atlanta, GA, USA, 2014. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK179276/> (accessed on 20 September 2025).
33. National Cancer Institute (U.S.). *Harms of Cigarette Smoking and Health Benefits of Quitting*. 2017. Available online: <https://www.cancer.gov/about-cancer/causes-prevention/risk/tobacco/cessation-fact-sheet#:~:text=,Formaldehyde> (accessed on 20 September 2025).
34. Jha, P.; Ramasundarahettige, C.; Landsman, V.; Rostron, B.; Thun, M.; Anderson, R.N.; McAfee, T.; Peto, R. 21st-Century Hazards of Smoking and Benefits of Cessation in the United States. *N. Engl. J. Med.* **2013**, *368*, 341–350. [CrossRef] [PubMed]
35. Tindle, H.A.; Stevenson Duncan, M.; Greevy, R.A.; Vasan, R.S.; Kundu, S.; Massion, P.P.; Freiberg, M.S. Lifetime Smoking History and Risk of Lung Cancer: Results from the Framingham Heart Study. *J. Natl. Cancer Inst.* **2018**, *110*, 1201–1207, Erratum in *J. Natl. Cancer Inst.* **2018**, *110*, 1153. [CrossRef]
36. Doll, R.; Peto, R.; Boreham, J.; Sutherland, I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* **2004**, *328*, 1519. [CrossRef]

37. Wilson, R.; Wahl, S.; Pfeiffer, L.; Ward-Caviness, C.K.; Kunze, S.; Kretschmer, A.; Reischl, E.; Peters, A.; Gieger, C.; Waldenberger, M. The dynamics of smoking-related disturbed methylation: A two time-point study of methylation change in smokers, non-smokers and former smokers. *BMC Genom.* **2017**, *18*, 805. [\[CrossRef\]](#)
38. Guida, F.; Sandanger, T.M.; Castagné, R.; Campanella, G.; Polidoro, S.; Palli, D.; Krogh, V.; Tumino, R.; Sacerdote, C.; Panico, S.; et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.* **2015**, *24*, 2349–2359. [\[CrossRef\]](#)
39. Klopach, E.T.; Carroll, J.E.; Cole, S.W.; Seeman, T.E.; Crimmins, E.M. Lifetime exposure to smoking, epigenetic aging, and morbidity and mortality in older adults. *Clin. Epigenet.* **2022**, *14*, 72. [\[CrossRef\]](#)
40. Qiu, W.; Wan, E.; Morrow, J.; Cho, M.H.; Crapo, J.D.; Silverman, E.K.; DeMeo, D.L. The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. *Epigenetics* **2015**, *10*, 1064–1073. [\[CrossRef\]](#)
41. Lee, K.W.K.; Pausova, Z. Cigarette smoking and DNA methylation. *Front. Genet.* **2013**, *4*, 132. [\[CrossRef\]](#)
42. Khulan, B.; Ye, K.; Shi, M.K.; Waldman, S.; Marsh, A.; Siddiqui, T.; Okorozo, A.; Desai, A.; Patel, D.; Dobkin, J.; et al. Normal bronchial field basal cells show persistent methylome-wide impact of tobacco smoking, including in known cancer genes. *Epigenetics* **2025**, *20*, 2466382. [\[CrossRef\]](#)
43. Mao, L.; Lee, J.S.; Kurie, J.M.; Fan, Y.H.; Lippman, S.M.; Broxson, A.; Khuri, F.R.; Hong, W.K.; Lee, J.J.; Yu, R.; et al. Clonal Genetic Alterations in the Lungs of Current and Former Smokers. *JNCI J. Natl. Cancer Inst.* **1997**, *89*, 857–862. [\[CrossRef\]](#)
44. Strulovici-Barel, Y.; Omberg, L.; O'Mahony, M.; Gordon, C.; Hollmann, C.; Tilley, A.E.; Salit, J.; Mezey, J.; Harvey, B.-G.; Crystal, R.G. Threshold of biologic responses of the small airway epithelium to low levels of tobacco smoke. *Am. J. Respir. Crit. Care Med.* **2010**, *182*, 1524–1532. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Zhang, L.; Lee, J.J.; Tang, H.; Fan, Y.-H.; Xiao, L.; Ren, H.; Kurie, J.; Morice, R.C.; Hong, W.K.; Mao, L. Impact of smoking cessation on global gene expression in the bronchial epithelium of chronic smokers. *Cancer Prev. Res.* **2008**, *1*, 112–118. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Hijazi, K.; Malyszko, B.; Steiling, K.; Xiao, X.; Liu, G.; Alekseyev, Y.O.; Dumas, Y.-M.; Hertsgaard, L.; Jensen, J.; Hatsukami, D.; et al. Tobacco-Related Alterations in Airway Gene Expression are Rapidly Reversed Within Weeks Following Smoking-Cessation. *Sci. Rep.* **2019**, *9*, 6978. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Shang, J.; Nie, X.; Qi, Y.; Zhou, J.; Qi, Y. Short-term smoking cessation leads to a universal decrease in whole blood genomic DNA methylation in patients with a smoking history. *World J. Surg. Oncol.* **2023**, *21*, 227. [\[CrossRef\]](#)
48. Braber, S.; Henricks, P.A.J.; Nijkamp, F.P.; Kraneveld, A.D.; Folkerts, G. Inflammatory changes in the airways of mice caused by cigarette smoke exposure are only partially reversed after smoking cessation. *Respir. Res.* **2010**, *11*, 99. [\[CrossRef\]](#)
49. Chari, R.; Lonergan, K.M.; Ng, R.T.; MacAulay, C.; Lam, W.L.; Lam, S. Effect of active smoking on the human bronchial epithelium transcriptome. *BMC Genom.* **2007**, *8*, 297. [\[CrossRef\]](#)
50. Gustafson, A.M.; Soldi, R.; Anderlind, C.; Scholand, M.B.; Qian, J.; Zhang, X.; Cooper, K.; Walker, D.; McWilliams, A.; Liu, G.; et al. Airway PI3K Pathway Activation Is an Early and Reversible Event in Lung Cancer Development. *Sci. Transl. Med.* **2010**, *2*, 26ra25. [\[CrossRef\]](#)
51. Keshawarz, A.; Joehanes, R.; Guan, W.; Huan, T.; DeMeo, D.L.; Grove, M.L.; Fornage, M.; Levy, D.; O'Connor, G. Longitudinal change in blood DNA epigenetic signature after smoking cessation. *Epigenetics* **2022**, *17*, 1098–1109. [\[CrossRef\]](#)
52. Vink, J.M.; Jansen, R.; Brooks, A.; Willemsen, G.; van Grootheest, G.; de Geus, E.; Smit, J.H.; Penninx, B.W.; Boomsma, D.I. Differential gene expression patterns between smokers and non-smokers: Cause or consequence? *Addict. Biol.* **2017**, *22*, 550–560. [\[CrossRef\]](#)
53. Wang, G.; Wang, R.; Strulovici-Barel, Y.; Salit, J.; Staudt, M.R.; Ahmed, J.; Tilley, A.E.; Yee-Levin, J.; Hollmann, C.; Harvey, B.-G.; et al. Persistence of smoking-induced dysregulation of miRNA expression in the small airway epithelium despite smoking cessation. *PLoS ONE* **2015**, *10*, e0120824. [\[CrossRef\]](#)
54. Yoshida, K.; Gowers, K.H.C.; Lee-Six, H.; Chandrasekharan, D.P.; Coorens, T.; Maughan, E.F.; Beal, K.; Menzies, A.; Millar, F.R.; Anderson, E.; et al. Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **2020**, *578*, 266–272. [\[CrossRef\]](#)
55. Zuo, W.-L.; Yang, J.; Gomi, K.; Chao, I.; Crystal, R.G.; Shaykhiev, R. EGF-Amphiregulin Interplay in Airway Stem/Progenitor Cells Links the Pathogenesis of Smoking-Induced Lesions in the Human Airway Epithelium. *Stem Cells* **2017**, *35*, 824–837. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Vucic, E.A.; Thu, K.L.; Pikor, L.A.; Enfield, K.S.S.; Yee, J.; English, J.C.; MacAulay, C.E.; Lam, S.; Jurisica, I.; Lam, W.L. Smoking status impacts microRNA mediated prognosis and lung adenocarcinoma biology. *BMC Cancer* **2014**, *14*, 778. [\[CrossRef\]](#)
57. Zeilinger, S.; Kühnel, B.; Klopp, N.; Baurecht, H.; Kleinschmidt, A.; Gieger, C.; Weidinger, S.; Lattka, E.; Adamski, J.; Peters, A.; et al. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS ONE* **2013**, *8*, e63812. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Ramirez, J.M.; Ribeiro, R.; Soldatkina, O.; Moraes, A.; García-Pérez, R.; Oliveros, W.; Ferreira, P.G.; Melé, M. The molecular impact of cigarette smoking resembles aging across tissues. *Genome Med.* **2025**, *17*, 66. [\[CrossRef\]](#) [\[PubMed\]](#)

59. Fasanelli, F.; Baglietto, L.; Ponzi, E.; Guida, F.; Campanella, G.; Johansson, M.; Grankvist, K.; Johansson, M.; Assumma, M.B.; Naccarati, A.; et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat. Commun.* **2015**, *6*, 10192. [CrossRef]
60. Wu, X.; Huang, Q.; Javed, R.; Zhong, J.; Gao, H.; Liang, H. Effect of tobacco smoking on the epigenetic age of human respiratory organs. *Clin. Epigenet.* **2019**, *11*, 183. [CrossRef] [PubMed]
61. Bruse, S.; Petersen, H.; Weissfeld, J.; Picchi, M.; Willink, R.; Do, K.; Siegfried, J.; Belinsky, S.A.; Tesfaigzi, Y. Increased methylation of lung cancer-associated genes in sputum DNA of former smokers with chronic mucous hypersecretion. *Respir. Res.* **2014**, *15*, 2. [CrossRef]
62. de Biase, M.S.; Massip, F.; Wei, T.-T.; Giorgi, F.M.; Stark, R.; Stone, A.; Gladwell, A.; O'Reilly, M.; Schütte, D.; de Santiago, I.; et al. Smoking-associated gene expression alterations in nasal epithelium reveal immune impairment linked to lung cancer risk. *Genome Med.* **2024**, *16*, 54. [CrossRef]
63. De Cunto, G.; De Meo, S.; Bartalesi, B.; Cavarra, E.; Lungarella, G.; Lucattelli, M. Smoking Cessation in Mice Does Not Switch off Persistent Lung Inflammation and Does Not Restore the Expression of HDAC2 and SIRT1. *Int. J. Mol. Sci.* **2022**, *23*, 9104. [CrossRef]
64. Saint-André, V.; Charbit, B.; Biton, A.; Rouilly, V.; Possémé, C.; Bertrand, A.; Rotival, M.; Bergstedt, J.; Patin, E.; Albert, M.L.; et al. Smoking changes adaptive immunity with persistent effects. *Nature* **2024**, *626*, 827–835. [CrossRef]
65. Chapla, D.; Chorya, H.P.; Ishfaq, L.; Khan, A.; Vr, S.; Garg, S. An Artificial Intelligence (AI)-Integrated Approach to Enhance Early Detection and Personalized Treatment Strategies in Lung Cancer Among Smokers: A Literature Review. *Cureus* **2024**, *16*, e66688. [CrossRef]
66. Ladbury, C.; Amini, A.; Govindarajan, A.; Mambetsariev, I.; Raz, D.J.; Massarelli, E.; Williams, T.; Rodin, A.; Salgia, R. Integration of artificial intelligence in lung cancer: Rise of the machine. *Cell Rep. Med.* **2023**, *4*, 100933. [CrossRef] [PubMed]
67. Picard, M.; Scott-Boyer, M.-P.; Bodein, A.; Périn, O.; Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3735–3746. [CrossRef]
68. Zhang, J.; Che, Y.; Liu, R.; Wang, Z.; Liu, W. Deep learning-driven multi-omics analysis: Enhancing cancer diagnostics and therapeutics. *Brief. Bioinform.* **2025**, *26*, bbaf440. [CrossRef]
69. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinforma. Biol. Insights* **2020**, *14*, 1177932219899051. [CrossRef] [PubMed]
70. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [CrossRef]
71. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. Review The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Współczesna Onkol.* **2015**, *1A*, 68–77. [CrossRef] [PubMed]
72. Setio, A.A.A.; Traverso, A.; De Bel, T.; Berens, M.S.N.; Bogaard, C.V.D.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B.; et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med. Image Anal.* **2017**, *42*, 1–13. [CrossRef] [PubMed]
73. Armato, S.G.; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A.; et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. *Med. Phys.* **2011**, *38*, 915–931. [CrossRef]
74. Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **2008**, *456*, 53–59. [CrossRef]
75. Payne, A.; Holmes, N.; Clarke, T.; Munro, R.; Debebe, B.J.; Loose, M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* **2021**, *39*, 442–450. [CrossRef]
76. Hon, T.; Mars, K.; Young, G.; Tsai, Y.-C.; Karalius, J.W.; Landolin, J.M.; Maurer, N.; Kudrna, D.; Hardigan, M.A.; Steiner, C.C.; et al. Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci. Data* **2020**, *7*, 399. [CrossRef] [PubMed]
77. Illumina, Inc. An Introduction to the Illumina 5-Base Solution. 2025. Available online: <https://www.illumina.com/science/genomics-research/articles/5-base-solution.html> (accessed on 20 October 2025).
78. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
79. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
80. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]

81. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **2019**, *25*, 954–961, Erratum in *Nat. Med.* **2019**, *25*, 1319. [CrossRef]
82. Mikhael, P.G.; Wohlgend, J.; Yala, A.; Karstens, L.; Xiang, J.; Takigami, A.K.; Bourgouin, P.P.; Chan, P.; Mrah, S.; Amayri, W.; et al. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk from a Single Low-Dose Chest Computed Tomography. *J. Clin. Oncol.* **2023**, *41*, 2191–2200. [CrossRef]
83. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: San Francisco, CA, USA, 2016. Available online: <https://dl.acm.org/doi/10.1145/2939672.2939785> (accessed on 6 November 2025).
84. Tammemägi, M.C.; Katki, H.A.; Hocking, W.G.; Church, T.R.; Caporaso, N.; Kvale, P.A.; Chaturvedi, A.K.; Silvestri, G.A.; Riley, T.L.; Commins, J.; et al. Selection Criteria for Lung-Cancer Screening. *N. Engl. J. Med.* **2013**, *368*, 728–736, Erratum in *N. Engl. J. Med.* **2013**, *369*, 394. [CrossRef]
85. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2021**, *1*, 57–81, Erratum in *AI Open* **2025**, *6*, 331–332. [CrossRef]
86. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Long Beach, CA, USA, 2017; pp. 5998–6008. Available online: <https://arxiv.org/abs/1706.03762> (accessed on 8 November 2025).
87. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
88. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef] [PubMed]
89. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]
90. Argelaguet, R.; Arnol, D.; Bredikhin, D.; Deloro, Y.; Velten, B.; Marioni, J.C.; Stegle, O. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **2020**, *21*, 111. [CrossRef] [PubMed]
91. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **2017**, *13*, e1005752. [CrossRef]
92. Pun, F.W.; Ozerov, I.V.; Zhavoronkov, A. AI-powered therapeutic target discovery. *Trends Pharmacol. Sci.* **2023**, *44*, 561–572. [CrossRef]
93. Kamya, P.; Ozerov, I.V.; Pun, F.W.; Tretina, K.; Fokina, T.; Chen, S.; Naumov, V.; Long, X.; Lin, S.; Korzinkin, M.; et al. PandaOmics: An AI-Driven Platform for Therapeutic Target and Biomarker Discovery. *J. Chem. Inf. Model.* **2024**, *64*, 3961–3969. [CrossRef]
94. Wang, S.; Wang, T.; Yang, L.; Yang, D.M.; Fujimoto, J.; Yi, F.; Luo, X.; Yang, Y.; Yao, B.; Lin, S.; et al. ConvPath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. *eBioMedicine* **2019**, *50*, 103–110. [CrossRef]
95. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Springer International Publishing: Munich, Germany, 2015; pp. 234–241. Available online: <https://arxiv.org/abs/1505.04597> (accessed on 6 November 2025).
96. Van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef]
97. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef]
98. Gao, C.; Wu, L.; Wu, W.; Huang, Y.; Wang, X.; Sun, Z.; Xu, M.; Gao, C. Deep learning in pulmonary nodule detection and segmentation: A systematic review. *Eur. Radiol.* **2025**, *35*, 255–266. [CrossRef]
99. Yoo, H.; Kim, K.H.; Singh, R.; Digumarthy, S.R.; Kalra, M.K. Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs. *JAMA Netw. Open* **2020**, *3*, e2017135. [CrossRef]
100. Lu, M.Y.; Williamson, D.F.K.; Chen, T.Y.; Chen, R.J.; Barbieri, M.; Mahmood, F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **2021**, *5*, 555–570. [CrossRef]
101. Pocock, J.; Graham, S.; Vu, Q.D.; Jahanifar, M.; Deshpande, S.; Hadjigeorgiou, G.; Shephard, A.; Bashir, R.M.S.; Bilal, M.; Lu, W.; et al. TIAToolbox as an end-to-end library for advanced tissue image analytics. *Commun. Med.* **2022**, *2*, 120. [CrossRef]
102. Vorontsov, E.; Bozkurt, A.; Casson, A.; Shaikovski, G.; Zelechowski, M.; Severson, K.; Zimmermann, E.; Hall, J.; Tenenholtz, N.; Fusi, N.; et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat. Med.* **2024**, *30*, 2924–2935. [CrossRef]

103. Krämer, A.; Green, J.; Pollard, J.; Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523–530. [\[CrossRef\]](#)
104. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [\[CrossRef\]](#)
105. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [\[CrossRef\]](#)
106. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [\[CrossRef\]](#)
107. Chu, Y.; Zhang, Y.; Wang, Q.; Zhang, L.; Wang, X.; Wang, Y.; Salahub, D.R.; Xu, Q.; Wang, J.; Jiang, X.; et al. A transformer-based model to predict peptide–HLA class I binding and optimize mutated peptides for vaccine design. *Nat. Mach. Intell.* **2022**, *4*, 300–311. [\[CrossRef\]](#)
108. Cheng, H.; Rao, B.; Liu, L.; Cui, L.; Xiao, G.; Su, R.; Wei, L. PepFormer: End-to-End Transformer-Based Siamese Network to Predict and Enhance Peptide Detectability Based on Sequence Only. *Anal. Chem.* **2021**, *93*, 6481–6490. [\[CrossRef\]](#)
109. Tran, N.H.; Zhang, X.; Xin, L.; Shan, B.; Li, M. De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8247–8252. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Liu, K.; Ye, Y.; Li, S.; Tang, H. Accurate de novo peptide sequencing using fully convolutional neural networks. *Nat. Commun.* **2023**, *14*, 7974. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Beaubier, N.; Bontrager, M.; Huether, R.; Igartua, C.; Lau, D.; Tell, R.; Bobe, A.M.; Bush, S.; Chang, A.L.; Hoskinson, D.C.; et al. Integrated genomic profiling expands clinical options for patients with cancer. *Nat. Biotechnol.* **2019**, *37*, 1351–1360. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Chalmers, Z.R.; Connelly, C.F.; Fabrizio, D.; Gay, L.; Ali, S.M.; Ennis, R.; Schrock, A.; Campbell, B.; Shlien, A.; Chmielecki, J.; et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **2017**, *9*, 34. [\[CrossRef\]](#)
113. Domyenyuk, V.; Benson, K.; Carter, P.; Magee, D.; Zhang, J.; Bhardwaj, N.; Tae, H.; Wacker, J.; Rath, F.; Miick, S.; et al. Clinical and analytical validation of MI Cancer Seek[®], a companion diagnostic whole exome and whole transcriptome sequencing-based comprehensive molecular profiling assay. *Oncotarget* **2025**, *16*, 642–659. [\[CrossRef\]](#)
114. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2*, 401–404, Erratum in *Cancer Discov.* **2012**, *2*, 960. [\[CrossRef\]](#) [\[PubMed\]](#)
115. U.S. Food and Drug Administration. precisionFDA. Available online: <https://precision.fda.gov/> (accessed on 15 October 2025).
116. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Bankhead, P.; Loughrey, M.B.; Fernández, J.A.; Dombrowski, Y.; McArt, D.G.; Dunne, P.D.; McQuaid, S.; Gray, R.T.; Murray, L.J.; Coleman, H.G.; et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **2017**, *7*, 16878. [\[CrossRef\]](#)
118. Goode, A.; Gilbert, B.; Harkes, J.; Jukic, D.; Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Inform.* **2013**, *4*, 27. [\[CrossRef\]](#)
119. Cardoso, M.J.; Li, W.; Brown, R.; Ma, N.; Kerfoot, E.; Wang, Y.; Murrey, B.; Myronenko, A.; Zhao, C.; Yang, D.; et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv* **2022**. [\[CrossRef\]](#)
120. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. *arXiv* **2016**. [\[CrossRef\]](#)
121. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
122. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035. [\[CrossRef\]](#)
123. Vaziri-Moghadam, A.; Foroughmand-Araabi, M.-H. Integrating machine learning and bioinformatics approaches for identifying novel diagnostic gene biomarkers in colorectal cancer. *Sci. Rep.* **2024**, *14*, 24786. [\[CrossRef\]](#) [\[PubMed\]](#)
124. Huang, D.; Li, Z.; Jiang, T.; Yang, C.; Li, N. Artificial intelligence in lung cancer: Current applications, future perspectives, and challenges. *Front. Oncol.* **2024**, *14*, 1486310. [\[CrossRef\]](#)
125. Sherman, M.A.; Yaari, A.U.; Priebe, O.; Dietlein, F.; Loh, P.-R.; Berger, B. Genome-wide mapping of somatic mutation rates uncovers drivers of cancer. *Nat. Biotechnol.* **2022**, *40*, 1634–1643. [\[CrossRef\]](#)
126. Calderaro, J.; Ghaffari Laleh, N.; Zeng, Q.; Maille, P.; Favre, L.; Pujals, A.; Klein, C.; Bazille, C.; Heij, L.R.; Uguen, A.; et al. Deep learning-based phenotyping reclassifies combined hepatocellular-cholangiocarcinoma. *Nat. Commun.* **2023**, *14*, 8290. [\[CrossRef\]](#)

127. Alruily, M.; Elbashir, M.K.; Ezz, M.; Aldughayfiq, B.; Alrowaily, M.A.; Allahem, H.; Mohammed, M.; Mostafa, E.; Mostafa, A.M. Comprehensive Network Analysis of Lung Cancer Biomarkers Identifying Key Genes Through RNA-Seq Data and PPI Networks. *Int. J. Intell. Syst.* **2025**, *2025*, 9994758. [\[CrossRef\]](#)
128. Joo, M.S.; Pyo, K.-H.; Chung, J.-M.; Cho, B.C. Artificial intelligence-based non-small cell lung cancer transcriptome RNA-sequence analysis technology selection guide. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1081950. [\[CrossRef\]](#)
129. Simon, B.D.; Ozyoruk, K.B.; Gelikman, D.G.; Harmon, S.A.; Türkbe, B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: A narrative review. *Diagn. Interv. Radiol. Ank. Turk.* **2025**, *31*, 303–312. [\[CrossRef\]](#)
130. Baião, A.R.; Cai, Z.; Poulos, R.C.; Robinson, P.J.; Reddel, R.R.; Zhong, Q.; Vinga, S.; Gonçalves, E. A technical review of multi-omics data integration methods: From classical statistical to deep generative approaches. *Brief. Bioinform.* **2025**, *26*, bbaf355. [\[CrossRef\]](#) [\[PubMed\]](#)
131. Marra, A.; Morganti, S.; Pareja, F.; Campanella, G.; Bibeau, F.; Fuchs, T.; Loda, M.; Parwani, A.; Scarpa, A.; Reis-Filho, J.S.; et al. Artificial intelligence entering the pathology arena in oncology: Current applications and future perspectives. *Ann. Oncol.* **2025**, *36*, 712–725. [\[CrossRef\]](#)
132. Ruan, L.-J.; Weng, K.-Q.; Zhang, W.-Y.; Zhuang, Y.-N.; Li, J.; Lin, L.-M.; Chen, Y.-T.; Zeng, Y.-M. Machine learning integration with multi-omics data constructs a robust prognostic model and identifies PTGES3 as a therapeutic target for precision oncology in lung adenocarcinoma. *Front. Immunol.* **2025**, *16*, 1651270. [\[CrossRef\]](#)
133. Dai, Q.; Ye, C.; Irklyenko, I.; Wang, Y.; Sun, H.-L.; Gao, Y.; Liu, Y.; Beadell, A.; Perea, J.; Goel, A.; et al. Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nat. Biotechnol.* **2024**, *42*, 1559–1570. [\[CrossRef\]](#)
134. Liu, Y.; Rosikiewicz, W.; Pan, Z.; Jillette, N.; Wang, P.; Taghbalout, A.; Foox, J.; Mason, C.; Carroll, M.; Cheng, A.; et al. DNA methylation-calling tools for Oxford Nanopore sequencing: A survey and human epigenome-wide evaluation. *Genome Biol.* **2021**, *22*, 295. [\[CrossRef\]](#) [\[PubMed\]](#)
135. Halliwell, D.O.; Honig, F.; Bagby, S.; Roy, S.; Murrell, A. Double and single stranded detection of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore sequencing. *Commun. Biol.* **2025**, *8*, 243. [\[CrossRef\]](#) [\[PubMed\]](#)
136. Cheung, W.A.; Johnson, A.F.; Rowell, W.J.; Farrow, E.; Hall, R.; Cohen, A.S.A.; Means, J.C.; Zion, T.N.; Portik, D.M.; Saunders, C.T.; et al. Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort. *Nat. Commun.* **2023**, *14*, 3090. [\[CrossRef\]](#) [\[PubMed\]](#)
137. Choi, H.; Na, K.J. A Risk Stratification Model for Lung Cancer Based on Gene Coexpression Network and Deep Learning. *BioMed Res. Int.* **2018**, *2018*, 2914280. [\[CrossRef\]](#)
138. Alum, E.U. AI-driven biomarker discovery: Enhancing precision in cancer diagnosis and prognosis. *Discov. Oncol.* **2025**, *16*, 313. [\[CrossRef\]](#)
139. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; De Jong, E.E.C.; Van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [\[CrossRef\]](#) [\[PubMed\]](#)
140. Muti, H.S.; Heij, L.R.; Keller, G.; Kohlruss, M.; Langer, R.; Dislich, B.; Cheong, J.-H.; Kim, Y.-W.; Kim, H.; Kook, M.-C.; et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: A retrospective multicentre cohort study. *Lancet Digit. Health* **2021**, *3*, e654–e664, Erratum in *Lancet Digit. Health* **2021**, *3*, e622. [\[CrossRef\]](#)
141. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirogas, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [\[CrossRef\]](#)
142. Wen, Q.; Yang, Z.; Dai, H.; Feng, A.; Li, Q. Radiomics Study for Predicting the Expression of PD-L1 and Tumor Mutation Burden in Non-Small Cell Lung Cancer Based on CT Images and Clinicopathological Features. *Front. Oncol.* **2021**, *11*, 620246. [\[CrossRef\]](#)
143. Ayasa, Y.; Alajrami, D.; Idkedek, M.; Tahayneh, K.; Akar, F.A. The Impact of Artificial Intelligence on Lung Cancer Diagnosis and Personalized Treatment. *Int. J. Mol. Sci.* **2025**, *26*, 8472. [\[CrossRef\]](#) [\[PubMed\]](#)

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