

Accelerating Cancer Risk Factor Discovery for Prevention in Younger Generations

Aigerim Dauletkyzy Sarsenova

International Medical School, Almaty 050010, Kazakhstan

Summary

The global incidence of early-onset cancer, diagnosed before age 50, is rising, with pronounced birth cohort effects. This alarming trend underscores an urgent need to accelerate discovery of novel cancer risk factors and their underlying biological networks, and to translate these insights into prevention and interception strategies that protect future generations. In this Perspective, we summarize key milestones and outline conceptual, methodological, and resource-related challenges that are hindering progress. To address these gaps, we propose three interconnected frameworks that extend traditional epidemiologic approaches: (1) a tissue-ecosystem anchored, critical window-aware framework for cancer risk factor discovery; (2) a life course-informed, biological state-based framework for precision cancer risk assessment; and (3) an integrated, dynamic, natural history-based framework to characterize cancer preventability. Together, they offer an initial roadmap for connecting exposures to molecular changes across biological scales, life stages, and populations to reveal the networks of causal risk factors. By beginning to map when and where risk factors disrupt shared hallmarks of cancer, reflected in shifts in (epi)genetic, immune, metabolic, and microbial pathways, these approaches can start to inform biologically grounded, scalable strategies to prevent cancer and other chronic diseases.

INTRODUCTION

Early-onset cancers, commonly defined as cancers diagnosed before age 50, are increasing worldwide (**Figure 1**).¹⁻⁴ Across 42 countries with incidence data from 2003 to 2017, 75% reported rising rates for six cancers (breast, colorectal, thyroid, kidney, endometrial, and leukemia), with average annual percent increase of 0.8% to 3.6%.⁵ Globally, early-onset cancers account for nearly 1 million deaths and 50 million disability-adjusted life years,⁶ imposing substantial personal, societal, and economic burdens. In the United States (US), in contrast to declined cancer mortality among older ages, mortality under age 50 has plateaued since the early 1990s and has risen for colorectal and endometrial cancers; colorectal cancer is now the leading cause of cancer death in men, and breast cancer remains the leading cause in women.¹ In many countries, these increases show strong birth-cohort effects, with Generation X (1965-1980) and Millennials (1981-1996) facing higher risks at younger ages than prior birth cohorts.⁷⁻¹⁰

Confronted with rising early-onset cancers and striking birth-cohort effects, we are compelled to ask whether we are entering a “cancer generation” and whether is time to reconsider how cancer causes have been and should be discovered. For much of the past century, epidemiology has powered this effort, delivering landmark insights on tobacco, alcohol, obesity, and infections through what Sir Richard Peto termed a “black box” strategy¹¹: comparing populations and identifying strong, reproducible associations before underlying mechanisms were known. The successes of this approach have shaped regulation and clinical practice, and contributed to global declines in tobacco-related mortality.^{12,13} Yet only an estimated 30-45% of cancers are currently attributable to established modifiable causes^{14,15}, even though 75-80% maybe preventable in theory¹¹, leaving a substantial proportion arising from uncharacterized or unknown exposures. This gap also reflects the pace of risk factor discovery and translation: even for lung cancer, establishing tobacco and asbestos as causes and translating evidence took decades. With cancer incidence rising in younger generations, such timelines are no longer acceptable.

The central challenge is twofold: accelerate discovery of cancer-causing risk factors and to map how exposures connect to biological pathways to shape susceptibility across the life course, with a unified goal to translate etiology into effective prevention and interception. The urgency is greatest for younger generations, who are experiencing earlier and more complex exposures, including in utero and early-life exposures and stresses,^{16,17} ultra-processed foods,¹⁸ circadian disruption,¹⁹ and ubiquitous environmental chemicals that can act synergistically²⁰ but many are newly introduced²¹ and/or untested for safety.^{22,23} Exposures during gestation or adolescence, periods of rapid tissue development and heightened vulnerability, may promote cancer not only by inducing mutations but also via epigenetic, immune, metabolic, hormonal, and microbial disruptions, with effects that may emerge decades later. The exposome, the totality of exposures from conception to death,²⁴ offer promising conceptual framework to unify exposures, however, because technologies, environments, and policies continually change, it is dynamic, reshaping risk trajectories across generations, yet it remains challenging. Taken together, these realities call for approaches that build on epidemiology’s core advantages while better capturing patterned, time-varying mixtures of exposures and their biological embedding across the life course. The

exposome offers a unifying way to organize exposures from conception to death, yet its dynamic nature, spanning chemical and non-chemical exposures that co-occur, make causal attribution difficult. Together, these challenges call for approaches that build on epidemiology's core strengths while better capturing life-course exposures and their biological embedding across the life course.

In this Perspective, we first revisit the historical trajectory of major risk factor discovery, then highlight key challenges that are particularly salient to early-onset cancers. Finally, we propose three integrative frameworks for novel risk factor discovery, risk assessment, and preventability characterization. These frameworks view cancer risk as an emergent property of evolving tissue ecosystems shaped by patterned exposures across the life course. They offer an initial roadmap for identifying critical exposures, pinpointing windows of heightened susceptibility, and refining estimates of cancer preventability, ultimately enabling more precise, earlier, and accessible approaches to cancer prevention.

DISCOVERY OF CARCINOGENS AND RISK FACTORS: HISTORICAL PERSPECTIVES

In 1981, Doll and Peto's *Causes of Cancer*¹¹ articulated a "black box" strategy for cancer risk factor discovery, alongside a complementary "mechanistic" strategy based on experimental testing of candidate agents. With genomic advances, these approaches shifted thinking from mutation-centric models to multi-stage clonal selection, in which initiation generates dormant initiated cells and subsequent promotion drives clonal expansion.²⁵⁻²⁷ Exogenous factors such as environmental exposures and modifiable behavior are typically identified through population studies and then dissected mechanistically to determine whether they act as initiators, promoters, or both, while endogenous processes encompassing cell-extrinsic (e.g., immunity, hormones, microbiota) and cell-intrinsic factors (e.g., DNA replication and repair) modulate tumor evolution.^{11,28} This convergence of epidemiologic and mechanistic evidence underlies IARC Group 1 carcinogen classifications.²⁹ **Figure 2** summarizes the milestones in uncovering three major preventable exogenous causes of cancer³⁰: tobacco, alcohol, and obesity, alongside advances in cancer genetics that illuminate endogenous pathways.

As Group 1 carcinogens and causes of cancer, tobacco and alcohol exemplify how epidemiology and mechanistic research have worked together, but on different timelines. For tobacco, 18th-century observations linked snuff and pipe use to nasal polyps and lip cancer.³¹⁻³³ By the mid-20th century, case-control studies showed roughly 10–13-fold higher lung cancer risk in heavy smokers.^{34,35} Concerns around biases inherent to retrospective designs prompted creation of large prospective cohorts, including the British Doctors Study of 40,000 physicians, which demonstrated dose-response relationships in lung cancer mortality over decades of follow-up.³⁶ Parallel animal experiments showed cigarette smoke condensates to be carcinogenic.^{37,38} These converging lines of evidence, with strong effect sizes, were synthesized in the 1964 US Surgeon General's Advisory,³⁷ concluding that cigarette smoking causes lung and laryngeal cancers. Subsequent work identified more than 60 carcinogens in tobacco smoke,³⁹ mapped benzo[a]pyrene adducts at *TP53* hotspots and *KRAS* mutations,^{40,41} and defined tobacco-associated mutational signatures⁴².

Alcohol followed a methodologically similar path but with weaker signals and a slower march to consensus. Early 20th-century case series linked absinthe drinking to esophageal cancer⁴³ and excess cancers of upper aerodigestive tract in alcohol-related occupations¹¹, while abstinent religious groups such as Seventh-day Adventists showed lower risks⁴⁴⁻⁴⁶. Subsequent case-control and cohort studies quantified these patterns, and in 1987 IARC⁴⁷ classified alcohol as a Group 1 carcinogen for cancers of the oral cavity, pharynx, larynx, esophagus, and liver with 2-fold higher risk in heavy drinkers.⁴⁸ Later evaluations added colorectum^{49,50} and female breast⁵¹⁻⁵³ to the above list in IARC 2010⁵⁴ with RR around 1.2–1.6.⁴⁸ Mechanistic studies have implicated acetaldehyde toxicity, *ALDH2* polymorphisms, oxidative stress, inflammation, hormonal changes, and synergistic interactions with tobacco.⁵⁵⁻⁵⁹ Reflecting this convergent evidence, current consensus is that no level of drinking is risk-free for cancer,⁶⁰ and a 2025 US Surgeon General's Advisory⁶¹ reinforced alcohol as a cause of 7 cancer types.

Obesity is a major modifiable and leading cancer risk factor contributing to cancer cases and deaths¹⁴. Since the 1970s, rising obesity prevalence^{62,63} prompting investigation into its role in cancer. American Cancer Society's Cancer Prevention Study I of >750,000 participants linked higher body weight to higher cancer mortality.⁶⁴ Subsequent epidemiological studies⁶⁵ and pooled analyses strengthened the evidence, including a large individual participant data meta-analysis of >900,000 adults from 57 prospective cohorts, which reported each 5 kg/m² increase was associated with 10% higher neoplastic mortality.⁶⁶ IARC evaluations in 2002⁶⁷ and 2016⁶⁸ concluded that avoidance of weight gain lowers risk for at least 13 cancer types, with RRs modest (1.1–1.8) but strikingly high for esophageal adenocarcinoma (4.8) and corpus uteri (7.1). Mechanistic work points to inflammation, insulin and IGF-1 signaling, hormonal imbalances, and metabolic reprogramming,^{69,70} with emerging roles for the microbiome and circadian disruption.^{71,72}

In parallel with these exogenous risk factors, advances in cancer genetics have revealed a complementary arc of inherited susceptibility.⁷³ Family-based linkage studies in the early 1990s identified high-penetrance genes such as *BRCA1*

and *BRCA2* for breast and ovarian cancers⁷⁴⁻⁷⁶ and *APC*, *MLH1*, and *MSH2* for colorectal cancer⁷⁷⁻⁸⁰. More than 70 such high-penetrance genes (RR >5) have been described but account for only a minority of familial risk in most cancers.⁸¹ This gap catalyzed genome-wide association studies (GWAS) in the mid-2000s, which agnostically interrogate large case-control populations to identify common, low-penetrance variants (risk allele frequency >5%; RR <1.5), including over 200 for breast cancer⁸² and colorectal cancer^{83,84} and at least 45 for lung cancer.^{85,86} However, approximately one-third of loci show pleiotropy across multiple cancer sites,⁸¹ reinforcing past findings indicating that inherited risk often maps to shared pathways.^{87,88} Aggregated into polygenic risk scores, these variants support population-level risk stratification and personalized screening and prevention.⁸⁹⁻⁹¹

Taken together, these histories, spanning decades to centuries, show why translating observation into causation and then prevention is slow and inherently multidisciplinary. They reaffirm the value of the “black box” strategy in epidemiology, a view Peto famously framed as “need for ignorance in cancer research”.⁹² Tobacco control shows what is possible when strong evidence is matched with decisive implementation, averting >3.8 million lung cancer deaths in the US during 1970–2022.⁹³ Alcohol illustrates slower translation when individual effect sizes are smaller and co-exposures are common. Obesity illustrates a third pattern: even with compelling evidence and mechanistic plausibility, progress is constrained by the difficulty of sustained behavior change in an obesogenic environment,⁹⁴ despite emerging pharmacologic strategies.⁹⁵ With these successes and lessons in view, and as contemporary exposures increasingly co-occur and accumulate across the life course with modest individual effects, the central challenge is to pinpoint emerging risk factors, establish causality more efficiently, and translate evidence into prevention faster.

CHALLENGES IN RISK FACTOR DISCOVERY FOR EARLY-ONSET CANCERS

Cancer risk factor discovery is constrained by fundamental features of carcinogenesis: long induction latency,⁹⁶ ethical limits on experimental exposure,⁹⁷⁻¹⁰⁰ wide inter-individual variation in susceptibility,^{101,102} heterogeneous evolutionary trajectories to malignancy,¹⁰³⁻¹⁰⁵ and complex gene-environment interactions^{106,107}. These challenges are now intensified by rapidly evolving exposome, in which many risk factors are widespread, co-occurring, and vary across generations and life stages, often exerting modest individual effects but potentially meaningful combined influence. In this section, we outline the key conceptual, resource, and methodological challenges that impede risk factor discovery and causal characterization across the multistage continuum of tumor development (**Box 1**), and propose targeted shifts in framing, study design, and analytic approaches that are particularly critical for understanding and preventing early-onset cancers.

Conceptual challenges

Age of onset is a proxy for underlying biology: Much effort has focused on comparing molecular differences between early- and later-onset cancers to identify “unique” early-onset biology, yet evidence to date are limited and often conflicting, with few early-onset specific alterations.^{108,109} It is important to note that age at diagnosis can be a misleading axis for comparison. Statistically, age effects reflect a mixture of birth cohort and period effects.¹¹⁰⁻¹¹² The absence of striking molecular differences does not imply fully shared etiology, because subtle changes in clonal evolution, immune contexture, or developmental timing can be obscured in bulk analyses. Birth cohort comparisons of molecular alterations can also be limited by under-representation of earlier cohorts, differing clinical and laboratory conditions, and survivorship or archival biases. Together, these considerations argue that both age-based and birth cohort-based molecular contrasts are most informative when explicitly interpreted in relation to patterns and timing of exposures, rather than as intrinsic properties of “young” versus “old” tumors.

Risk factors are often treated as one-dimensional and independent: Our understanding of established cancer risk factors and their preventability, is often oversimplified by focusing on one or limited dimensions. Simple questionnaire-based recall of aggregate exposure (e.g. “weekly alcohol intake”) or single measurements at one timepoint (e.g. “current BMI”) miss key dimensions such as intensity, timing and trajectories, and cumulative exposures across life stages. As a result, cancer risk attributable to these factors is likely underestimated. For instance, alcohol consumption between menarche and first pregnancy was associated with increased breast cancer risk independent of drinking later in life,¹¹³ and heavy episodic or binge drinking increased risk of breast cancer especially among moderate lifetime drinkers.¹¹⁴ Emerging behaviors, such as drinking on an empty stomach¹¹⁵, are increasingly common in younger populations but remain poorly characterized with respect to cancer risk. Moving beyond static snapshots toward high-resolution exposure characterization that jointly models intensity, timing, trajectories, and clustering (**Figure 3**) is essential for a more realistic assessment of how established and emerging risk factors shape cancer risk.

The exposome is complex: The concept of “exposome”, first proposed by Christopher Wild in 2005,²⁴ referred to non-genetic environmental exposures that influence health and disease. Humans are exposed to a vast array of chemicals

through food, medications, supplements, consumer products, pollution, and many other sources. Determining which exposures, either individually or in synergy with others, contribute to cancer initiation or progression is conceptually straightforward but experimentally challenging. The universe of potential chemical exposures now numbers in the hundreds of thousands and is rapidly expanding, with many substances incompletely characterized or undisclosed.²¹ Indeed, the number of synthetic industrial chemicals produced globally is projected to exceed one million by 2050.²³ Beyond chemicals, the exposome also includes non-chemical components (e.g., socioeconomic conditions, behaviors, and occupational environments) that are often measured crudely, while biological response to exposures are further shaped by tissue specific context and modulated by host factors such as genetics, microbiome, and overall health.¹¹⁶ Given this complexity, it is not practical to screen all possible exposure combinations in model systems. The path forward will need to leverage emerging measurement technologies capable of profiling large numbers of chemicals in large human cohorts.^{117,118} The application of this approach, called chemical exposomics, at the population scale to samples from longitudinal biobanks has the potential to identify and prioritize the most biologically relevant, ubiquitous, and emerging risk factors for neoplastic disease. Integrated with computational advances and complementary data such as geospatial data, this strategy is poised to discover and ultimately prevent environmental causes of early-onset cancer.

Risk factors are often treated as organ specific: Although cancer is increasingly recognized as a systemic disease,¹¹⁹ risk factor discovery has largely remained organized organ-by-organ. A more tractable and urgent question is whether shared systemic drivers, both exogenous and endogenous, are fueling not only cancers broadly,¹²⁰ but also to emerging epidemiological phenomena such as the rise of early-onset cancers. Many established risk factors, such as smoking, alcohol, and obesity, are implicated across multiple malignancies (**Figure 2**).^{68,120-122} Smoking, for example, affects lung cancer through direct inhalation yet also increases risk for at least 12 other cancers,¹²² illustrating how a single carcinogen can reshape both local tissue and systemic physiology. We now appreciate that many risk factors act predominantly as promoters rather than initiators,²⁸ but have only begun to map their systematic impact. Looking ahead, major advances are likely to come from reframing early-onset cancers as organ-specific manifestations of shared systemic perturbations (e.g., inflammatory alterations and metabolic change),¹¹⁹ and from designing prevention strategies that target shared biological pathways. This will require a paradigm of systemic risk factor discovery that prioritizes emerging, ubiquitous, and modifiable exposures with broad physiological impact, rather than isolating risk factors organ by organ.

Genetics is not driving the increasing cancer incidence: Since allele frequencies change slowly and high-penetrance syndromes have remained stable for decades, the prevailing view is that germline genetics is unlikely to explain the rapid, birth cohort-specific rise in early-onset incidence. However, susceptibility can also arise within a generation through de novo mutations, and it is plausible that delayed parenthood over recent decades, which increases de novo mutation burden,¹²³ could modestly contribute to early-onset cancer risk. Additionally, genetic predisposition can act as a modifier of susceptibility,¹²⁴ shaping inter-individual vulnerability to modern exposures. Consistent with this, UK Biobank analyses suggest unhealthy lifestyle factors confer greater early-onset cancer risk among those with higher genetic risk.¹²⁵ Clarifying these gene-environmental interactions will be essential for understanding why younger generations may be more susceptible.

Humans are humans, mice are mice: The importance of experimental models in cancer risk factor discovery was highlighted by Doll and Peto in *Causes of Cancer*, where the ‘mechanistic’ strategy tries to understand the biology of cancer and exhaustively tests masses of the chemical, infective, or physical agents to which people are likely to be exposed to determine which are likely to be the causes of the cancers of today or of the future”.¹¹ Decades later, animal models remain indispensable for establishing causality, defining dose and timing, and interrogating early tissue responses, while their translational limitations are increasingly apparent. Fundamental differences in physiology, lifespan, immune architecture, and tissue microenvironments create non-trivial gaps between murine systems and human carcinogenesis, gaps that widen when risk factors arise from complex, chronic, co-occurring human exposures. Even when exposure timing and intensities are approximated, experimental models often simplify key components such as adaptive immunity, stromal remodeling, and microbial ecosystems, making it difficult to recapitulate the multifactorial and life-course nature of human cancer.¹²⁶

Resource-related challenges

Reliance on new, prospective cohort studies: Establishing new birth or multi-generational cohorts with longitudinal, repeated exposome data collection throughout the lifespan will be essential for better mapping the fully landscape of contemporary risk factors. Yet a more forward-looking strategy is to treat such cohorts as the next layer, not the starting point. A pressing priority is to harness existing data sources, national databases, cohorts, biobanks, and in-house EHRs, to pinpoint biologically relevant signals and strengthen causal inferences. Insights from these heterogeneous data sources, combined with advances in computational tools, can help simulate alternative cohort designs, prioritize

which exposures and biospecimens to capture, and guide where new cohorts would add the greatest marginal value. Importantly, the growing landscape of longitudinal birth and multigenerational cohorts and linked biobanks already makes early-life and intergenerational exposure assessment feasible at scale, particularly in Europe (e.g., Avon Longitudinal Study of Parents¹²⁷, Danish National Birth Cohort,¹²⁸ Lifelines¹²⁹, Generation R¹³⁰), where pregnancy and perinatal factors are captured prospectively, alongside repeated questionnaires and extensive biospecimen collection.

Reliance on data from developed countries: The global rise in early-onset cancers calls for cross-national, comparative approaches that extend beyond high-income settings. Most established risk factors come from data collected in developed countries, often requiring pooled datasets because of relatively low incidence. In contrast, a forward-looking agenda recognizes that both high- and low-incidence settings are informative: systematically learning from settings without a marked rise in early-onset cancers can reveal protective risk factors, environments, and policies that may buffer younger generations from risk. In parallel, integrating tumor molecular landscapes, including mutational signatures and other multi-omics features, with contextual data across diverse countries, races and ethnicities, and sexes can clarify whether cancers with similar histology share or diverge in etiological pathways. Addressing these questions will require globally coordinated, interdisciplinary efforts that deliberately include countries across the spectrum of incidence to identify both causes and protective factors and translate them into prevention. Such efforts are beginning to emerge (e.g., DISCERN¹³¹, which integrates tumor and normal tissue collection in Europe countries with high and low rates of renal, pancreatic, and colorectal cancer with whole-genome sequencing, multi-omic profiling, and prospective cohorts), but they require substantial sustained investment and infrastructure.

Methodological challenges

Causal inference with multi-modal, longitudinal data: A central methodological barrier to identifying cancer risk factors relevant to younger generations is the misalignment between life-course causality and prevailing analytic paradigms. Most pipelines analyze modalities in isolation or rely on cross-sectional snapshots drawn from limited cohorts, implicitly prioritizing association over causation. Such approaches struggle to resolve time-dependent confounding, mediation, and reverse causality, and are poorly suited to disentangling how early-life, adolescent, and young-adult exposures propagate through molecular and physiological intermediates to influence cancer risk decades later. Although principled causal frameworks for high-dimensional biology are emerging,¹³² their application to integrated omics and exposome data remains limited. A critical gap lies in linking population-scale environmental, lifestyle, and social determinants of health data with longitudinal biospecimens in ways that support explicit causal hypothesis testing across the life-course. Methods such as life-course¹³³ Mendelian randomization,¹³⁴ *g*-methods (including the *g*-formula and marginal structural models),^{135,136} and target trial emulation¹³⁷ provide a rigorous foundation for evaluating time-varying exposures and separating confounding from mediation, yet remain underused at the scale and complexity required for modern cancer prevention research. Emerging artificial intelligence approaches, including large language models, offer opportunities to automate data harmonization, feature extraction, and hypothesis generation across heterogeneous data sources. However, without explicit causal design, such tools risk amplifying spurious correlations rather than accelerating discovery. Realizing their potential will require embedding artificial-intelligence-driven workflows within transparent reproducible causal frameworks that yield interpretable effect estimates relevant to prevention.

Innovative model systems?

Isolated evidence analysis: Research on early-onset cancers has focused on identifying individual risk factors based on single studies, multiple cohorts or meta-analyses. Other studies have used this information to estimate the population attributable fraction (PAF) associated with each risk factor. However, most studies have looked at risk factors in isolation, yielding PAFs that added together reach biologically implausible values. Moreover, these studies failed to relate secular trends in risk factors to the trends in early-onset cancer, leaving uncertainty about how much they contribute to rising incidence even for the most important risk factors. To address this gap, methodology is needed that cannot only capture the impact of risk factors on cancer but can also capture dynamic changes in population and risk factors over time, as well as the delay in exposure to risk factor and actual cancer risk. Validated population simulation models of cancer natural history can provide this capacity, with the most well-known examples are models developed within the Cancer Intervention Surveillance Modeling Network (CISNET).

ACCELERATING CANCER RISK FACTOR DISCOVERY: FOUNDATION AND NEW FRAMEWORKS

Recent reflections on Peto's ideas^{25,138,139} reaffirm that epidemiology remains essential as the foundation of cancer risk factor discovery and that we still need to "embrace ignorance," while also confronting why discovery has stalled.¹⁴⁰ Historically, epidemiology has often identified causal exposures long before mechanistic confirmation, with smoking and lung cancer as the classic example. This approach was effective when exposures were few, extreme, and had large effect

sizes. However, many remaining causes of cancer are likely ubiquitous,¹⁴⁰ poorly measured, or exert their effects early in life, producing little between-person contrast that would empower classical epidemiologic designs. This loss of contrast is particularly consequential for prevention in younger generations, where early-life and cumulative exposures may shape risk long before clinical disease emerges. Insights from evolutionary biology¹⁴¹ and normal-tissue studies show that driver-like mutations and clonal expansions are common in healthy tissues.¹⁴²⁻¹⁴⁵ These observations suggest that mutation acquisition alone is insufficient to explain cancer risk; instead, promoters, tissue ecology, and age- and exposure-dependent selective pressures may determine which initiated clones expand and ultimately progress. Quantifying clonal expansion of drivers in diverse ecological contexts, including differing ages and exposure histories, is vital to the furtherance of our understanding of both somatic genetic drivers and the impact of age and exposure history on the somatic evolution of cancer.

Under these conditions, a strict pipeline from observational association to downstream mechanistic validation will not be efficient or sufficient. Addressing contemporary challenges require a shift from parallel, loosely connected tracks of “epidemiological association” and “mechanism” toward an explicitly iterative, closed-loop framework. In this model, epidemiology identifies and prioritizes candidate risk factors; tissue-based and experimental systems test how these exposures leave durable biological imprints and interact with germline variation to alter tissue vulnerability; and mechanistic insights feedback to refine epidemiologic hypotheses, biomarkers, and prevention strategies. Iterative feedback between human and experimental evidence helps clarify biologically relevant exposures, identify susceptible tissues, prioritize pathways, and sharpen causal inference and further accelerate identification of actionable cancer risk factors relevant to younger generations.

Recent advances illustrate the power of this approach, including integrative programs combining epidemiology with functional models and clinical cohorts to illuminate air pollution driven lung tumor promotion and nominate actionable prevention targets.¹⁴⁴ Consortia such as TaRGET II (Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription II)¹⁴⁶ extend this logic at scale by building longitudinal, multi-omic atlases for known environmental exposures (including PM 2.5, arsenic, lead, BPA, tributyltin, dioxin, and phthalates) across target organs and surrogate blood at weaning and adult time points in mice.¹⁴⁷ These resources reveal persistent, toxicant-specific epigenomic and transcriptomic perturbations, enabling direct comparisons between persistent exposure-linked signatures in human biospecimens and experimental perturbation readouts. Further extending and refining such human-model integrative approaches likely will offer a promising path forward in advancing cancer risk factor discovery and interception strategies.

To achieve this goal and accelerate risk factor discovery and early interception, we propose three complementary, forward-looking frameworks (**Figure 4**) that extend the strengths of traditional epidemiology while advancing it into a more biologically grounded, time-resolved, and integrative science of discovery. Together, these frameworks connect population exposures to the tissue ecosystems they perturb, map tissue states onto dynamic trajectories of susceptibility across the life course, and translate these insights into natural history-anchored estimates of preventability.

Tissue-ecosystem anchored, critical window-aware framework for cancer risk factor discovery

Identifying risk factors for early-onset cancers requires a shift from viewing exposures as isolated variables to understanding how tissues integrate lifelong physiological disturbances. A tissue-anchored framework reframes cancer risk as an emergent property of dynamic tissue ecosystems, shaped by lifelong and intergenerational exposures and germline predisposition. This framework moves beyond exposure-centric and focuses on how cumulative exposures across sensitive life stages create lasting biological imprints that influence tissue vulnerability, somatic evolution, and tumor emergence. Critically, many exposures are transient or poorly captured by conventional measurement, yet they leave durable “tissue memories” encoded in epigenetic programs, immune and stromal set points, metabolic wiring, and microbial community structure.^{148,149}

Mutational epidemiology, an alliance of cancer epidemiology and somatic genomics,¹⁵⁰ shows the value of a tissue-anchored lens. Somatic mutations accumulate throughout life via diverse mutational processes that leave characteristic “mutational signatures”,¹²⁰ and large-scale efforts (e.g., ICGC/TCGA Pan-Cancer Analysis of Whole Genomes) have catalogued many such signatures across >23,000 cancers.¹⁵¹ In one example, sequencing 981 colorectal cancer genomes from 11 countries revealed substantial geographic in mutational signatures, including enrichment of colibactin-associated signatures (SBS88, ID18) in high-incidence countries; these signatures were about 3.3-fold more common in cancers diagnosed before age 40 than after 70 and were imprinted early in tumor evolution, implicating early-life exposure to colibactin-producing bacteria in rising early-onset colorectal cancer and global incidence differences.¹⁵² Complementing tumor-based analyses, the Somatic Mosaicism across Human Tissues (SMaHT) Network is building a

multi-tissue reference catalogue of somatic mutations and clonal expansions in non-diseased human tissues, establishing baseline maps of mutation burden, clonal architecture, and selection across the lifespan.¹⁵³ Together, these underscore a key advantage of the tissue-anchored framework: exposures can leave tissue-specific evolutionary imprints that are difficult or impossible to reconstruct through questionnaires or environmental monitoring alone.

The major challenge in mutational epidemiology is that we increasingly recognize that many cancer risk factors are not mutagenic: they shape tissue fate primarily through systemic circuits. Balmain's reappraisal of Peto's paradox likewise argues that for many established carcinogens the rate-limiting step is not the generation of new mutations but non-mutagenic promotion that drives clonal expansion of pre-existing initiated cells.²⁵ We therefore advocate to expand the concept of "exposure fingerprints" from somatic mutations to the broader tissue ecosystem, by prioritizing the mapping of a small number of tissue-ecosystem axes that are most likely to encode exposure history and modulate multistage carcinogenesis. In this view, epigenetic, immune-stromal-vascular, metabolic-endocrine, epigenetic, and microbial signatures¹¹⁹ are potential candidate causal fingerprints. Insights from developmental plasticity and the developmental origins of health and disease (DOHaD) paradigm demonstrate how exposures in utero, during infancy, or across adolescence and early adulthood can introduce durable epigenetic and physiological reprogramming that shapes long-term tissue function and cancer susceptibility,¹⁵⁴⁻¹⁵⁶ with intergenerational epigenetic, immune, and metabolic imprinting further modifying baseline vulnerability.¹⁵⁷⁻¹⁶⁰ At barrier sites such as the gut, skin, and lungs, epithelial cells, resident immune cells, stromal elements, and commensal microbes form tightly coupled ecosystems that continuously sense diet, pathogens, xenobiotics, and environmental insults, establishing relatively durable patterns of epithelial turnover, cytokine milieu, metabolite profiles, and tissue-resident immunity.¹⁶¹ When disrupted by infections, antibiotics, obesity, high-fat and low-fiber diets, or pollution, these circuits can drive low-grade inflammation, barrier dysfunction, and altered metabolic and microbial states that promote clonal expansion and weaken immune surveillance rather than directly initiating mutations.¹⁶² Mapping how these memories are coordinated across organ systems through shared hormones, metabolites, and inflammatory mediators can reveal common promotive pathways and intervention windows that would otherwise remain hidden. Upstream of these layers, chemical fingerprints such as DNA and protein adducts act as specific sensors of exposure, anchoring suspected agents to their sites of action. Treating proximal adducts and downstream ecosystem axes as structured, exposure-encoded states enables testable hypotheses about causal pathways, and facilitates the discovery of risk factors that would otherwise remain invisible. Animal models that mimic human-relevant exposures across generations provide critical systems for understanding how inherited and early-life insults remodel tissue microenvironments and accelerate oncogenic processes.^{163,164}

This framework also incorporates contemporary principles of cancer evolution,¹⁶⁵⁻¹⁶⁷ recognizing tumorigenesis as a dynamic eco-evolutionary process shaped by germline predisposition, tissue ecology, and cumulative exposures. Cancer emerges within complex adaptive tissue landscapes where systemic regulators such as hormones, growth factors, and immune mediators interact with local features including nutrient and oxygen availability, stromal and matrix architecture, spatial organization, and niche constraints to shape the fitness, competition, and selection of somatic clones.¹⁶⁸ Fitness effects of individual somatic mutations can be quantified.^{169,170} However, early-onset cancers may reflect accelerated or altered evolutionary trajectories across life course.¹⁷¹ During early life, heightened proliferation, plastic differentiation programs, and immature immune surveillance create distinct selective environments that influence how mutant clones are generated, constrained, or expanded. Exposures that occur in utero, childhood, adolescence, and early adulthood can durably reconfigure immune, metabolic, and stromal circuits and reduce the capacity of tissues to police aberrant cells, thereby favoring the emergence of high-risk clones at younger ages. Incorporation of how exposures converge across sensitive life stages and interact with tissue-level biology to accelerate malignancy into multi-hit mechanistic models of carcinogenesis^{96,104,172} is crucial to a complete understanding of how exposures shape tissue fate through systemic circuits and to translational guidance on prevention strategies tailored to younger populations.

Translating this tissue-anchored framework into concrete discoveries of novel risk factors ultimately hinges on how well we can measure and reconstruct exposures across the life course. While progress in identifying new risk factors through questionnaires and conventional epidemiologic instruments has slowed, the discovery of molecular biomarkers of exposure has accelerated rapidly and is poised to expand further with advances in high-throughput molecular profiling technologies. Meeting this challenge will require coordinated, multi-modal strategies that capture exposures across biological and temporal scales, at minimum during critical developmental windows, in both human studies and experimental models. Objective monitoring tools, including personal sensors, wearables, and high-resolution geospatial-temporal models, can quantify environmental, behavioral, and circadian disruptions with far greater precision than questionnaires and geocode linkage solely based on residential addresses. Molecular readouts, including epigenomic marks, chemical exposomics, metabolomics, proteomics, microbiome profiles, adductomics, may provide short- and medium-term fingerprints of complex exposure mixtures that are otherwise difficult to reconstruct; for example, plasma

proteomics has revealed organ-specific aging signatures that precede clinical disease, illustrating how longitudinal profiling can capture latent susceptibility beyond traditional risk factors^{173,174}.

Finally, the field needs systematic “translation layers” that link tissue-ecosystem signatures back to causes by pairing profiling with *in vitro* and *ex vivo* exposure libraries, organoid models, and exposomic perturbation platforms that pinpoint candidate agents and pathways. Looking forward, integrating these outputs with longitudinal cohorts, electronic health records, life-stage-specific questionnaires, and rigorous causal inference can shift exposure science from sparse snapshots to reconstructable trajectories, enabling tissue signatures to be traced to their upstream drivers.

Life-course informed, biological state-based framework for precision cancer risk assessment

The tissue-anchored framework focuses on *where* exposures reshape tissue ecosystems to drive cancer development and emerging backward approach to infer exposures. A complementary question is *when* susceptibility is programmed and how to quantify this temporal dimension for precision prevention. There is broad recognition that cancer risk begins accumulating long before clinical onset. A life course-informed, biological state-based framework for precision cancer risk assessment recognizes that vulnerability to cancer emerges from cumulative exposures beginning *in utero* and extending throughout life, with the pace and impact of accumulation varying by tissue and concentrated in susceptible developmental windows, long before clinical detection or adult-focused models can capture it.

However, current risk prediction models only capture a small portion of this trajectory. Established tools, such as the Gail model for breast cancer,¹⁷⁵ the BOADICEA model for hereditary breast and ovarian cancer,¹⁷⁶ and the Colorectal Cancer Risk Assessment Tool,¹⁷⁷ primarily incorporate age, family history, genetic variants, and limited reproductive or lifestyle factors, were foundational in introducing quantitative risk assessment to clinical practice. Yet these models show only modest discriminative capacity with area under the curve typically between 0.55 and 0.70.¹⁷⁸⁻¹⁸¹ In essence, these tools approximate measures of the exposure at the tissue level, using the epidemiologic measures to drive this approximation.

A life course-informed, biological state-based framework moves beyond static snapshots of risk by treating susceptibility as a biological state that evolves through accumulated exposures, physiological transitions, and intergenerational influences encompassing both inherited genetic risk and non-genetic mechanisms such as epigenetic or metabolic immune programming.^{182,183} It integrates early-life and adolescent exposures alongside adult risk factors to generate dynamic profiles of vulnerability across the lifespan.¹³³ Existing dynamic models, such as the Rosner-Colditz breast cancer model¹⁸⁴⁻¹⁸⁶ and other time-updated risk tools, illustrate how incorporating changes in reproductive, anthropometric, and lifestyle factors over time from early childhood can meaningfully improve prediction. Next-generation cancer-specific and pan-cancer molecular risk scores could integrate not only genomics,¹³⁵ but longitudinal multi-omic inputs, including exposomics, radiomics, epigenomics, metabolomics, proteomics, and microbiome data,^{187,188} capturing both shared and tumor-specific pathways. Such scores could provide a unified platform to stratify risk across multiple malignancies simultaneously, an especially important advance for early-onset cancers where incidence is low and single-cancer strategies are often underpowered.

Recent advances in computational modeling provide practical paths for operationalizing this framework.¹⁸⁹ AI models leveraging longitudinal clinical multi-omic data can detect subtle, preclinical shifts in tissue biology; for example, breast cancer models that incorporate three years of serial screening mammograms achieve a 5-year AUC of 0.80, compared with 0.74 when relying on a single mammogram,^{190,191} and adding traditional risk factors yields little additional improvement in discrimination.¹⁹⁰⁻¹⁹² Emerging transformer-based architectures capable of learning lifetime health trajectories further demonstrate how different data modalities, from diagnoses, lifestyle factors, laboratory values, to imaging, and more, can be integrated to forecast future disease risk with good performance.¹⁹³ Learning latent, evolving biological states that blend trajectories of measured risk factors, rather than simply summing them, these models move beyond traditional risk-factor-based assessment to quantify how individuals transition between lower- and higher-risk tissue/biological states over time. Such approaches may be particularly impactful for younger individuals without strong family histories or well-known risk factors, offering biology-driven risk assessment tools to identify those at elevated risk across multiple malignancies and to guide earlier lifestyle interventions, surveillance, or preventive therapeutics.

Integrated, dynamic, natural history-informed framework to characterize cancer preventability

Risk identification is essential. However, we must ultimately translate knowledge of risks into actionable cancer prevention strategies. Current Population Attributable Fraction (PAF) analyses suggest that roughly 40% of cancers in settings such as the UK are attributable to established modifiable risk factors; less carefully sampled but rigorously analyzed molecular data approximately agree, linking ~40% of cancer causation to putatively preventable mutagenic

sources and a further ~25% to currently unknown processes that may also prove preventable if identified.¹²⁰ In contrast, Peto and colleagues, comparing high and low incidence populations, argued that on the order of 75–80% of cancers are preventable in theory. The discrepancies may come from various sources. First, traditional approaches, such as PAF¹⁹⁴, are anchored to current exposure distributions, assume largely independent effects, rarely incorporate time-varying or life course patterns or interactions, and do not account for the feasibility of changing them in real-world settings.¹⁹⁵ Second, these estimates only weakly reflect contemporary insights from cancer biology and evolution, including multistage carcinogenesis, the distinction between initiators and promoters, and species- and tissue-level adaptations that modulate susceptibility, and often rely on coarse proxies such as BMI¹⁹⁶ that miss etiologically relevant heterogeneity, obscuring preventable risk that more biologically precise phenotyping could reveal. Together, these limitations point to a larger, still untapped space of preventable cancer risk.

To address this gap, we propose individual level based, natural history-informed modeling of cancer preventability that explicitly incorporates population shifts in exposure trajectories across birth cohorts and life stages.¹⁹⁷ This approach differs from ecological analyses that correlate broad risk-factor trends with incidence. An individual-level microsimulation uses each person as the unit of analysis, so multiple risk factors, their clustering and potential interactions, and their time-varying histories can be carried forward through the full disease course, from health to precursor to cancer, and then through screening detection, treatment, and survivorship. Validated population simulation models, such as those developed within the Cancer Intervention and Surveillance Modeling Network (CISNET),¹⁹⁸ provide a powerful foundation for this approach. These models synthesize evidence from cancer registries and clinical and epidemiological studies, to create virtual populations at risk for cancer representative of the population at interest. These models have successfully been used in the past to explain observed trends in cancer incidence and mortality over time, highlighting the contribution of risk factors, screening, and treatment to these trends.¹⁹⁹⁻²⁰¹ The advantage of these models is that they can also simulate counterfactual scenarios, for example in the absence of screening or certain risk factors. This way, they can quantify the individual impact of each of the risk factors to the observed trend. An additional advantage is that these models can also simulate future counterfactual strategies, such as the impact of interventions on risk factors or screening on future cancer incidence and mortality.²⁰² The CISNET consortium is currently focusing on developing multiple risk factor generators for inclusion in the cancer models to accurately mimic lifetime exposure to risk factors and their impact on cancer burden. This way, counterfactual scenarios can also be evaluated specifically aimed at disentangling trends in early-onset cancers. A further advantage of CISNET-style modeling is that population exposure distributions can be calibrated to nationally representative data, enabling real-world estimates of preventability under plausible scenarios of exposure change. For example, projecting obesity prevalence declines from ~40% to 35% or 20% over specified time horizons with widespread GLP-1 uptake, rather than assuming that an exposure can be fully eliminated.

Currently, most models mimic natural history through discrete disease stages. Embedding an explicit evolutionary framework within these models can substantially sharpen estimates of preventability. For example, CISNET has developed biology-driven multistage clonal expansion models that explicitly incorporate cellular processes of initiation and promotion.²⁰³⁻²⁰⁵ Additionally, recent work integrates mutational signatures with gene-specific mutation rates and estimates of selection on somatic variants to partition the relative contributions of endogenous, exogenous, and preventable mutational processes across tumor types.¹²⁰ These evolutionary analyses move beyond counting mutations, instead estimating which mutational processes actually drive oncogenesis and how their influence varies by tissue, exposure context, and stage of disease. This integration of quantitative evolutionary estimates can reveal where modifying particular exposures is most likely to reduce the appearance and expansion of key driver mutations—and, critically, where prevention efforts may yield limited returns despite substantial reduction. In this way, evolutionary modeling helps to define not just what causes cancer, but where prevention is most tractable.

Looking forward, this framework also creates opportunities to move beyond single-cancer models toward integrated, multi-cancer simulation platforms. Growing evidence suggests that many exposures, such as obesity, metabolic dysfunction, inflammation, and microbiome perturbations, affect shared biological pathways and influence risk across multiple cancer types. Modeling multiple cancers within a unified framework would allow estimation of joint preventable burden, identification of shared versus tissue-specific mechanisms, and evaluation of interventions that may reduce risk across several malignancies simultaneously. Such an approach is particularly relevant for early-onset cancers, where incidence is low for individual cancer types but cumulative burden across cancers is substantial.

Importantly, these models should not only estimate whether risk reduction is theoretically possible but also incorporate the time course of prevention. The extent of benefit depends on whether exposures can be sufficiently modified, the timing of intervention along the multistage carcinogenesis process, and the durability of behavioral or

structural changes over time.¹⁹⁵ More realistic estimates of cancer preventability will therefore require data on how quickly and to what extent risk can be reversed when exposures change. Trials that combine behavior change with deep biological profiling, such as weight loss or smoking cessation studies with serial tissue sampling, can quantify the dynamics and partial reversibility of promotion. Incorporating these data into natural history and microsimulation models would allow us to move beyond static PAFs toward dynamic, tissue-informed estimates of preventable burden across cancers and cardiometabolic diseases. Together, these temporal and evolutionary dynamics determine both the pace and ultimate extent of reductions in cancer incidence, and are essential for setting realistic expectations for prevention outcomes. An integrated framework of this kind can synthesize evidence from population, mechanistic, and implementation sciences, and ultimately guide more precise, feasible, and high-impact screening and prevention strategies.

PREVENTION FOR FUTURE GENERATIONS

Early-onset cancer is rising in parallel with earlier diabetes, cardiovascular disease, and other chronic conditions, pointing to shared systemic disruptions in immune, inflammatory, metabolic, epigenetic, and microbial systems that are imprinted by evolving social, physical, and personal environments. This potential convergence creates a chance to design prevention strategies that target common promotive pathways and delays carcinogenesis across multiple organs while also reducing cardiometabolic disease burden. Next-generation prevention should be built around intermediate biomarker trials that quantify modifiable tissue-ecosystem and systemic states within learning health systems, and are explicitly linked to natural history models that convert near-term biological change into projected long-term gains across diseases. Yet two decades of experience in early detection consortia have shown how rarely biomarkers mature into reliable intermediate endpoints,^{206,207} underscoring the need to prioritize markers that are tightly anchored to natural history and meaningful reduction in cancer and cardiometabolic events. To focus these efforts where they will matter most, risk stratification must move beyond age to integrate genetic susceptibility, family history, high risk clinical phenotypes such as early metabolic dysfunction, and structural disadvantage, identifying individuals and communities for whom early, integrated cancer and chronic disease prevention could be most transformative.

CONCLUSION

The rising burden of early-onset cancers is one of the most urgent challenges in modern cancer research, and meeting it will require transforming how we discover and act on cancer risk factors. We advocate for an integrated agenda that dissolves silos between “exposure,” “biology,” and “prevention.” A tissue-ecosystem anchored framework links population exposures to specific tissue circuits and eco-evolutionary constraints, a life course informed, biological state-based framework captures how biological states evolve over time for dynamic risk assessment, and integrated natural history models translate these insights into realistic, time-sensitive estimates of preventability. Together, these frameworks can shift risk factor discovery from primarily cataloguing midlife associations to systematically opening the black box between exposures and disease, reconstructing exposure trajectories, predicting evolving biological states, and guiding upstream interventions that prevent cancers, including early-onset cancers, as well as other chronic diseases that share common pathways.

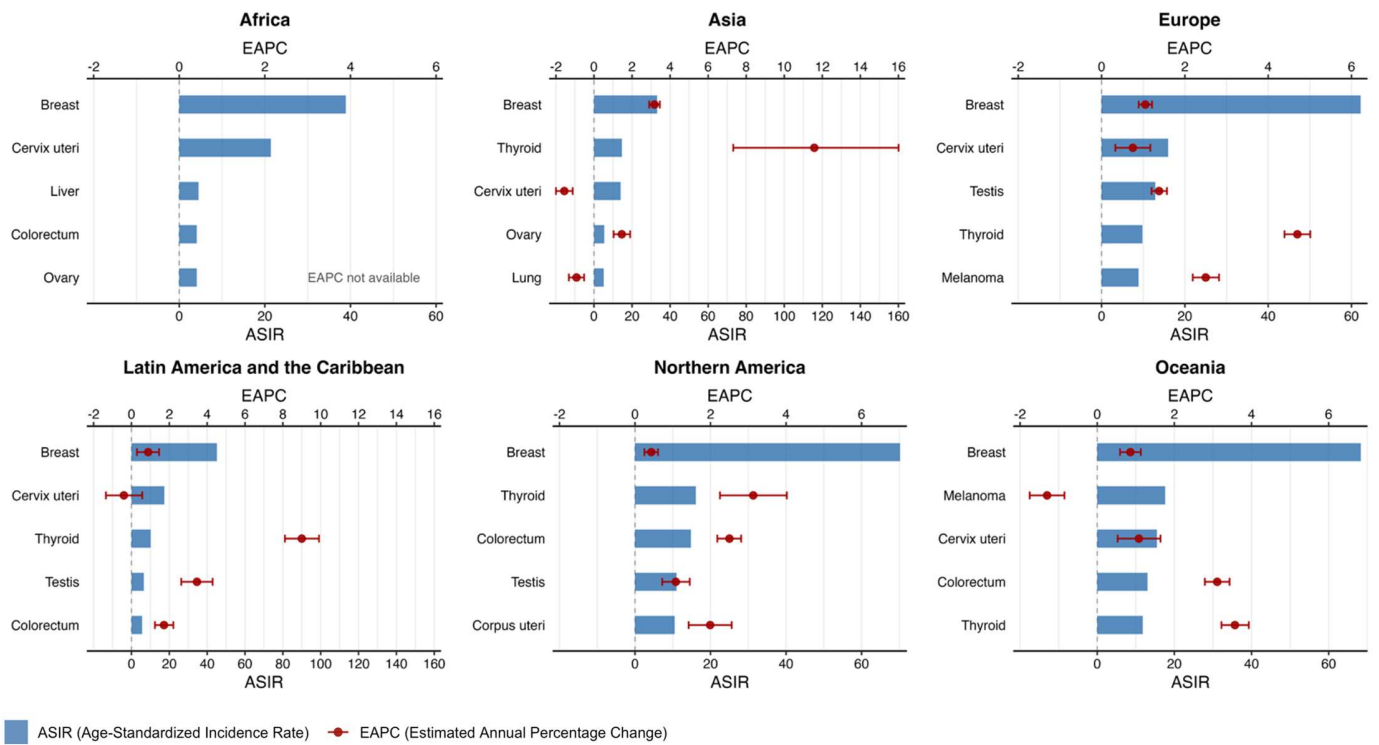
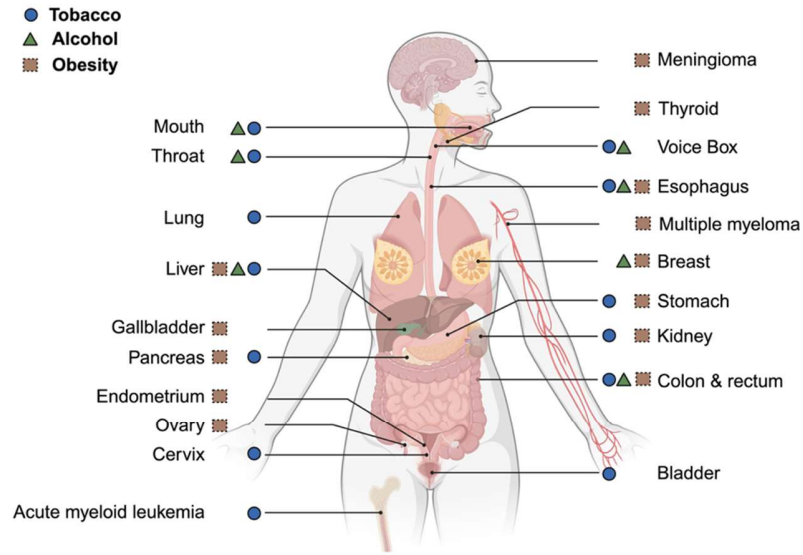


Figure 1. Global burden and trends of early-onset cancer

The five most common early-onset cancers (ages 20–49 years) in each continent are shown by age-standardized incidence rate (ASIR, per 100,000; blue bars) and corresponding estimated annual percentage change (EAPC; red points with 95% confidence intervals). EAPC estimates are available only for a subset of countries or cancer registries within each continent and are frequently derived from subnational data; therefore, EAPC values may not fully represent continent-wide trends. EAPC summarizes the average annual change in ASIR during 2003–2017 and was estimated by fitting a log-linear regression of ASIR on calendar year. Data were obtained from the International Agency for Research on Cancer (IARC, <https://gco.iarc.who.int>), including *Cancer Today* (GLOBOCAN 2022, version 1.1) and *Cancer Over Time* (GLOBOCAN, version 2.1).

A) Cancer types associated with tobacco smoking, alcohol consumption, and obesity



B) Timeline and milestones in the discovery of carcinogens and risk factors for cancer

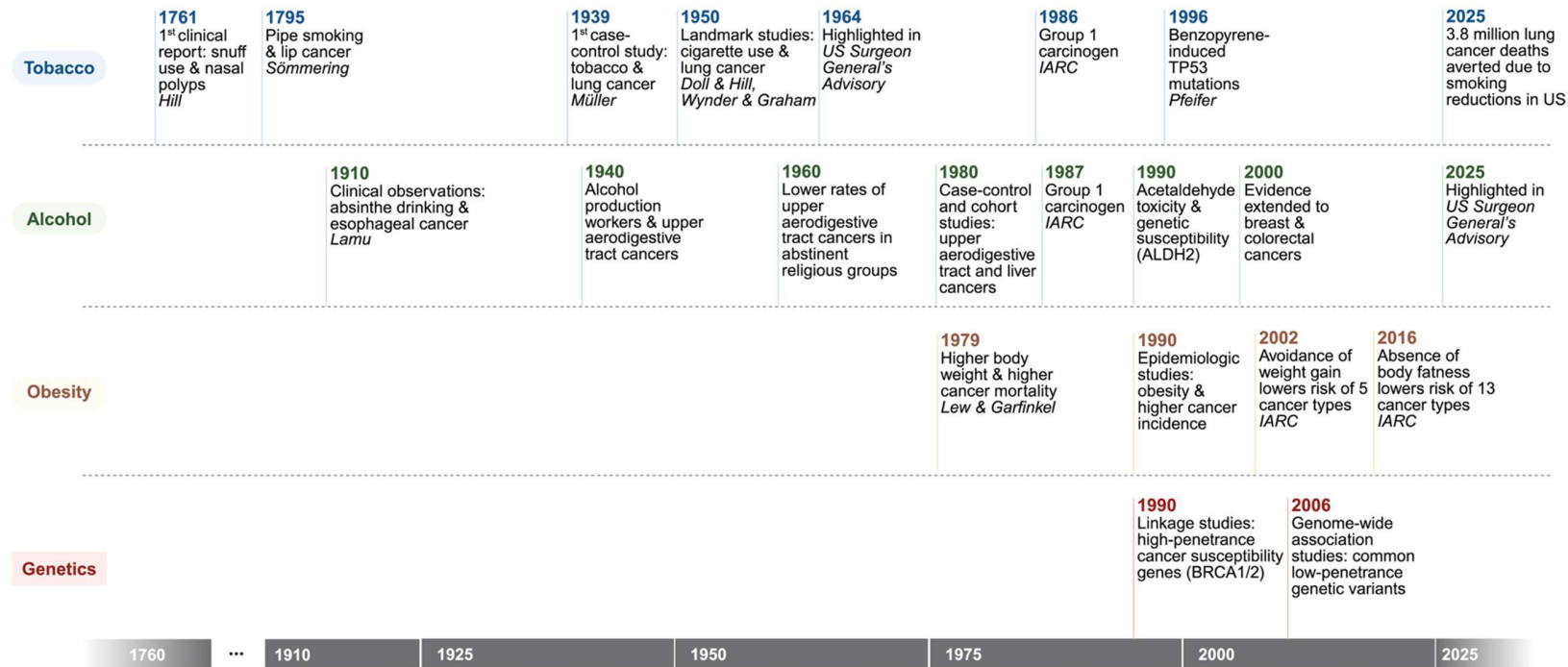


Figure 2. Discovery of major causes of cancer: tobacco, alcohol, obesity, and genetics

(A) Cancers associated with tobacco smoking, alcohol consumption, and obesity. Tobacco smoking and alcohol consumption are IARC Group 1 carcinogens and are causally associated with 13 and 7 cancer types, respectively. Obesity is a major modifiable risk factor; IARC concludes that avoiding excess body fatness reduces risk for 13 cancer types. **(B)** Timeline and milestones in uncovering three major preventable causes of cancer: tobacco smoking, alcohol consumption, and obesity, alongside advances in cancer genetics. The timeline traces the progression from early clinical observations to epidemiologic evidence, mechanistic studies, and genomic discoveries, illustrating the decades-long, multidisciplinary research required to establish causality and to translate evidence into public-health actions for cancer prevention and early detection.

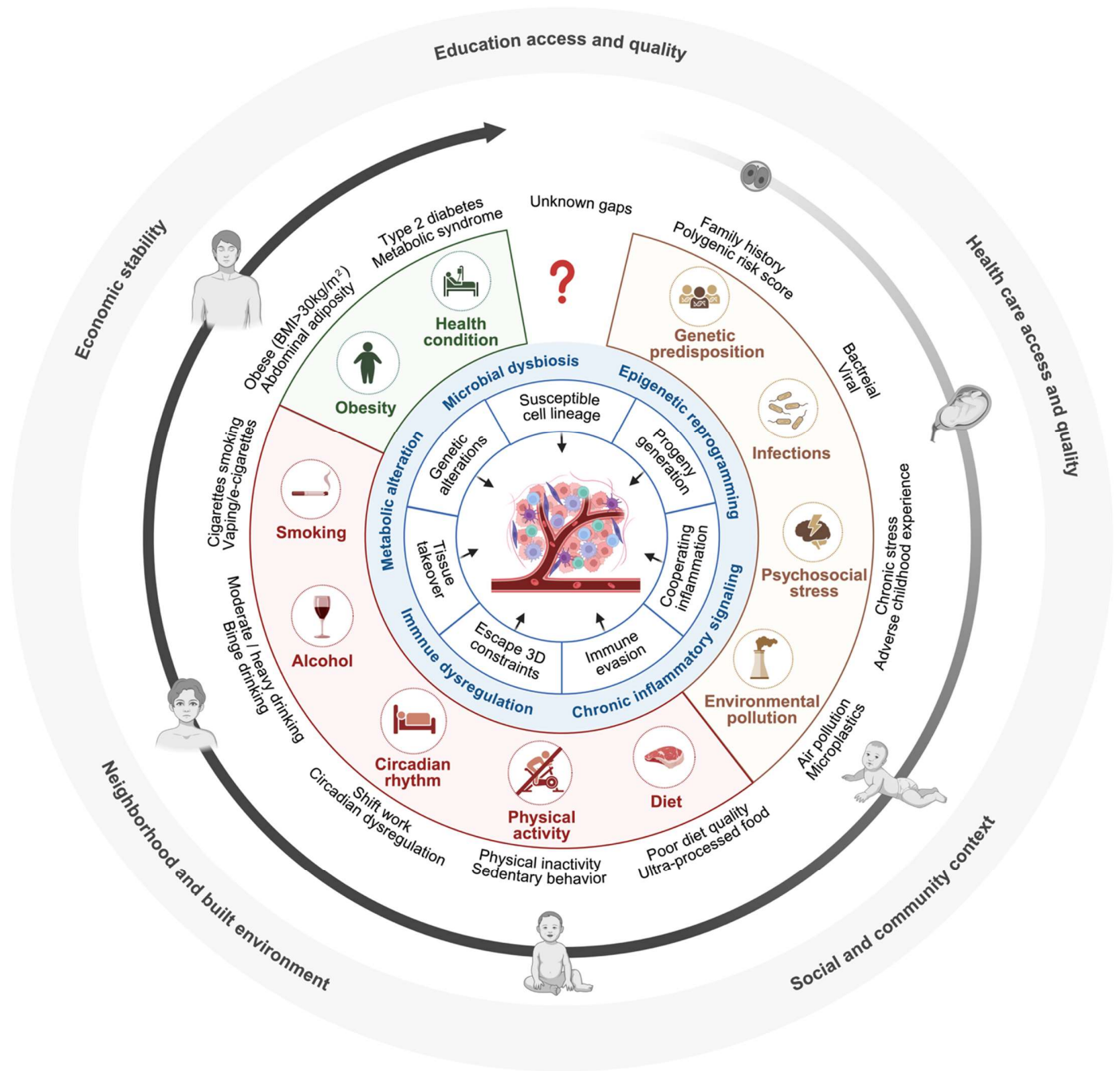


Figure 3. Established and emerging risk factors for main types of early-onset cancers and shared mechanisms. Cancer risk factors (yellow, red, and green) act on the fitness landscape of the tissue through mechanisms (blue) that either initiate tumor progression or promote the expansion of initiated clones. Additional, unidentified exposures (red question mark) may be relevant to younger generations. Exposures accumulate, cluster, and vary across sensitive windows from early life through adulthood (circular arrow and life-stage icons), and social determinants of health (grey outer circle) influence both exposure distributions and their biological impact. These nested layers emphasize that early-onset cancers arise not from isolated risk factors, but from interactions across molecular mechanisms, systemic physiology, individual behaviors, and broader social contexts.

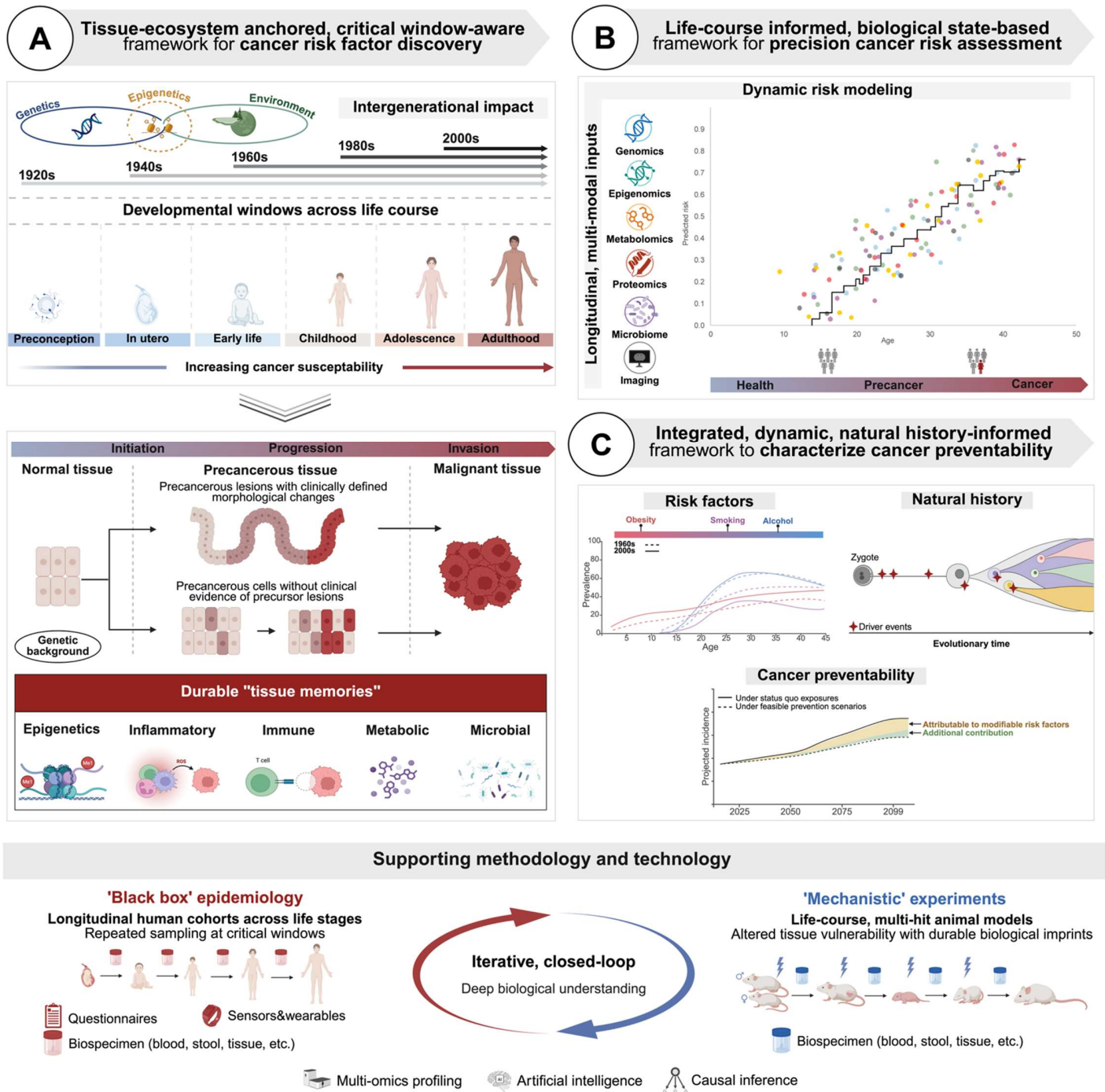


Figure 4. Accelerated frameworks to combat early-onset cancers

A: Tissue-ecosystem anchored, critical window-aware framework: exposures across the life course and intergeneration, together with host genetics, leave molecular imprints that reshape tissue ecosystems and the somatic fitness landscape, influencing initiation, progression, and invasion. **B:** Life-course informed, biological state-based framework: longitudinal multi-omic and imaging data generate dynamic risk states to stratify individuals for precision prevention. **C:** Integrated, dynamic, natural history-based framework: population-based cancer simulation models incorporate secular trends and interventions to estimate incidence and other outcomes. Together, these frameworks bridge biology, prediction, and policy to inform early cancer prevention, supported by enabling methods and technologies.

Box 1. Reimaging risk factors discovery for early-onset cancers.

- **Is age of onset a reliable proxy for “unique” early-onset biology?**
Shift from simple early- vs later-onset comparisons to analyses that disentangle age, birth-cohort, and period effects, and interpret molecular patterns in relation to life-course exposures that may drive apparent age-linked signatures.
- **Are cancer risk factors one-dimensional and independent?**
Shift toward a multidimensional view that captures intensity, timing, patterns, and changes of behavior across the life course, as well as how exposures co-occur and interact.
- **Are risk factors strictly organ-specific?**
Recognize early-onset cancers as systemic diseases potentially driven by shared, cross-organ exposures and mechanisms (e.g. immune, metabolic, epigenetic, microbiome), rather than from isolated organ-level insults.
- **Are findings from animal models inherently limited in relevance to humans?**
Promote integrative, parallel rather than linear pipelines, where experimental models and human observational studies are developed side by side and iteratively inform each other to refine causal hypotheses and mechanistic understanding.
- **Does genetic predisposition explain the rise in early-onset cancers?**
Recognize that high-penetrance germline variants and common low-risk variants account for only a minority of cases; prioritize studies of gene-environment interaction to identify who is most susceptible to modern exposures and why.
- **Is current methodology sufficient for causal inference with complex data?**
Develop, validate, and standardize rigorous pipelines for integrating multimodal, longitudinal, and high-dimensional data, enabling biologically grounded causal discovery at scale.

References

1. Siegel, R.L., Giaquinto, A.N., and Jemal, A. (2024). Cancer statistics, 2024. *CA Cancer J Clin* 74, 12-49. 10.3322/caac.21820.
2. Sung, H., Siegel, R.L., Laversanne, M., Jiang, C., Morgan, E., Zahwe, M., Cao, Y., Bray, F., and Jemal, A. (2025). Colorectal cancer incidence trends in younger versus older adults: an analysis of population-based cancer registry data. *Lancet Oncol* 26, 51-63. 10.1016/s1470-2045(24)00600-4.
3. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 74, 229-263. 10.3322/caac.21834.
4. Jianhui, Z., Liying, X., Jing, S., Mingyang, S., Lijuan, W., Shuai, Y., Yingshuang, Z., Zhengwei, W., Susanna, L., Konstantinos, T., et al. (2023). Global trends in incidence, death, burden and risk factors of early-onset cancer from 1990 to 2019. *BMJ Oncology* 2, e000049. 10.1136/bmjonc-2023-000049.
5. Berrington de Gonzalez, A., Brayley, M., Frost, R., Freedman, N., Gunter, M.J., Jackson, I., Lapitan, P., Shiels, M.S., and García-Closas, M. (2025). Trends in Cancer Incidence in Younger and Older Adults : An International Comparative Analysis. *Ann Intern Med*. 10.7326/annals-24-02718.
6. Yan, W., Liu, M., Jing, W., Kang, L., Zhang, N., Sun, H., He, J., Chen, Z., Liu, J., Liang, W., and Dong, J. (2025). Disparities in the incidence, mortality and disability-adjusted life years of 33 early-onset cancer groups globally, 2012-2021: a systematic analysis. *Exp Hematol Oncol* 14, 38. 10.1186/s40164-025-00634-7.
7. Shiels, M.S., Haque, A.T., Berrington de González, A., Camargo, M.C., Clarke, M.A., Davis Lynn, B.C., Engels, E.A., Freedman, N.D., Gierach, G.L., Hofmann, J.N., et al. (2025). Trends in Cancer Incidence and Mortality Rates in Early-Onset and Older-Onset Age Groups in the United States, 2010-2019. *Cancer Discov* 15, 1363-1376. 10.1158/2159-8290.Cd-24-1678.
8. Sung, H., Jiang, C., Bandi, P., Minihan, A., Fidler-Benaoudia, M., Islami, F., Siegel, R.L., and Jemal, A. (2024). Differences in cancer rates among adults born between 1920 and 1990 in the USA: an analysis of population-based cancer registry data. *Lancet Public Health* 9, e583-e593. 10.1016/s2468-2667(24)00156-7.
9. Sung, H., Siegel, R.L., Rosenberg, P.S., and Jemal, A. (2019). Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. *Lancet Public Health* 4, e137-e147. 10.1016/s2468-2667(18)30267-6.
10. Rosenberg, P.S., and Miranda-Filho, A. (2024). Cancer Incidence Trends in Successive Social Generations in the US. *JAMA Netw Open* 7, e2415731. 10.1001/jamanetworkopen.2024.15731.
11. Doll, R., and Peto, R. (1981). The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *JNCI: Journal of the National Cancer Institute* 66, 1192-1308.
12. Jha, P. (2009). Avoidable global cancer deaths and total deaths from smoking. *Nat Rev Cancer* 9, 655-664. 10.1038/nrc2703.
13. Jha, P., and Peto, R. (2014). Global effects of smoking, of quitting, and of taxing tobacco. *N Engl J Med* 370, 60-68. 10.1056/NEJMra1308383.
14. Islami, F., Marlow, E.C., Thomson, B., McCullough, M.L., Rumgay, H., Gapstur, S.M., Patel, A.V., Soerjomataram, I., and Jemal, A. (2024). Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States, 2019. *CA Cancer J Clin* 74, 405-432. 10.3322/caac.21858.
15. Brennan, P., and Davey-Smith, G. (2022). Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed. *J Natl Cancer Inst* 114, 353-360. 10.1093/jnci/djab204.
16. Cao, Y., Nguyen, L.H., Tica, S., Otegbeye, E., Zong, X., Roelstraete, B., Chan, A.T., Warner, B.B., Stephansson, O., and Ludvigsson, J.F. (2023). Evaluation of Birth by Cesarean Delivery and Development of Early-Onset Colorectal Cancer. *JAMA Netw Open* 6, e2310316. 10.1001/jamanetworkopen.2023.10316.

17. Peeri, N.C., O'Connell, K., Kantor, E.D., Setiawan, V.W., Guo, X., Lipworth, L., and Du, M. (2024). Early-Life Factors and Early-Onset Endometrial Cancer Risk in the UK Biobank. *JAMA Netw Open* 7, e2440181. 10.1001/jamanetworkopen.2024.40181.
18. Wang, C., Du, M., Kim, H., Nguyen, L.H., Wang, Q.L., Drew, D.A., Leeming, E.R., Khandpur, N., Sun, Q., Zong, X., et al. (2025). Ultraprocessed Food Consumption and Risk of Early-Onset Colorectal Cancer Precursors Among Women. *JAMA Oncol*. 10.1001/jamaoncol.2025.4777.
19. Fortin, B.M., Mahieu, A.L., Fellows, R.C., Kang, Y., Lewis, A.N., Ead, A.S., Lamia, K.A., Cao, Y., Pannunzio, N.R., and Masri, S. (2025). The diverse roles of the circadian clock in cancer. *Nat Cancer* 6, 753-767. 10.1038/s43018-025-00981-8.
20. Patel, C.J. (2017). Analytic Complexity and Challenges in Identifying Mixtures of Exposures Associated with Phenotypes in the Exposome Era. *Curr Epidemiol Rep* 4, 22-30. 10.1007/s40471-017-0100-5.
21. Wang, Z., Walker, G.W., Muir, D.C.G., and Nagatani-Yoshida, K. (2020). Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ Sci Technol* 54, 2575-2584. 10.1021/acs.est.9b06379.
22. Parrish, M., and Kuperwasser, C. (2025). Environmental endocrine disruptors: rethinking the origins of early-onset ER(+) breast cancer. *Nat Rev Cancer* 25, 819-820. 10.1038/s41568-025-00854-3.
23. Orešič, M., Karu, N., Zhao, H.N., Moseley, A., Hankemeier, T., Wishart, D.S., Dorrestein, P.C., Fiehn, O., Hyötyläinen, T., and Daouk, R.K. (2025). Metabolome informs about the chemical exposome and links to brain health. *Environ Int* 203, 109741. 10.1016/j.envint.2025.109741.
24. Wild, C.P. (2005). Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14, 1847-1850. 10.1158/1055-9965.Epi-05-0456.
25. Balmain, A. (2023). Peto's paradox revisited: black box vs mechanistic approaches to understanding the roles of mutations and promoting factors in cancer. *Eur J Epidemiol* 38, 1251-1258. 10.1007/s10654-022-00933-x.
26. Jassim, A., Rahrman, E.P., Simons, B.D., and Gilbertson, R.J. (2023). Cancers make their own luck: theories of cancer origins. *Nat Rev Cancer* 23, 710-724. 10.1038/s41568-023-00602-5.
27. Rozhok, A.I., and DeGregori, J. (2017). Chapter 12 - Aging, Somatic Evolution, and Cancer. In *On Human Nature*, M. Tibayrenc, and F.J. Ayala, eds. (Academic Press), pp. 193-209. <https://doi.org/10.1016/B978-0-12-420190-3.00012-0>.
28. Weeden, C.E., Hill, W., Lim, E.L., Grönroos, E., and Swanton, C. (2023). Impact of risk factors on early cancer evolution. *Cell* 186, 1541-1563. 10.1016/j.cell.2023.03.013.
29. Samet, J.M., Chiu, W.A., Coglianò, V., Jinot, J., Kriebel, D., Lunn, R.M., Beland, F.A., Bero, L., Browne, P., Fritschi, L., et al. (2020). The IARC Monographs: Updated Procedures for Modern and Transparent Evidence Synthesis in Cancer Hazard Identification. *J Natl Cancer Inst* 112, 30-37. 10.1093/jnci/djz169.
30. The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019. (2022). *Lancet* 400, 563-591. 10.1016/s0140-6736(22)01438-6.
31. Hill, J. (1761). Cautions against the immoderate use of snuff. Founded on the known qualities of the tobacco plant; and the effects it must produce when this way taken into the body (R. Baldwin; J. Jackson).
32. Redmond, D.E., Jr. (1970). Tobacco and cancer: the first clinical report, 1761. *N Engl J Med* 282, 18-23. 10.1056/nejm197001012820105.
33. Soemmerring, S.T. (1795). *De Morbis Vasorum Absorbentium Corporis Humani Sive Dissertationis Quae Praemium Retulit Societatis Rheno-Traiectinae Anno MDCCCLXXXIV. Pars...; Accedit Index Scriptorum De Systemate Absorbente (Varrentrappius & Wenner)*.
34. Wynder, E.L., and Graham, E.A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. *J Am Med Assoc* 143, 329-336. 10.1001/jama.1950.02910390001001.
35. Doll, R., and Hill, A.B. (1952). A study of the aetiology of carcinoma of the lung. *Br Med J* 2, 1271-1286. 10.1136/bmj.2.4797.1271.

36. Doll, R., and Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *Br med J* 2, 1525-1536.
37. Service, U.P.H. (1964). Surgeon General's advisory committee on smoking and health. Washington, DC: US Government Printing Office.
38. Wynder, E.L., Graham, E.A., and Croninger, A.B. (1953). Experimental production of carcinoma with cigarette tar. *Cancer Res* 13, 855-864.
39. Hecht, S.S. (2003). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer* 3, 733-744. 10.1038/nrc1190.
40. Denissenko, M.F., Pao, A., Tang, M.-s., and Pfeifer, G.P. (1996). Preferential formation of benzo [a] pyrene adducts at lung cancer mutational hotspots in P53. *Science* 274, 430-432.
41. Rodenhuis, S., van de Wetering, M.L., Mooi, W.J., Evers, S.G., van Zandwijk, N., and Bos, J.L. (1987). Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. *N Engl J Med* 317, 929-935. 10.1056/nejm198710083171504.
42. Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. *Science* 354, 618-622. 10.1126/science.aag0299.
43. Lamu, L. (1910). Etude de statistique clinique de 131 cas de cancer de l'oesophage et du cardia. *Arch Mal Appar Dig Mal Nutr* 4, 451-456.
44. Wynder, E.L., Lemon, F.R., and Bross, I.J. (1959). Cancer and coronary artery disease among Seventh-Day Adventists. *Cancer* 12, 1016-1028. 10.1002/1097-0142(195909/10)12:5<1016::aid-cncr2820120523>3.0.co;2-2.
45. Lemon, F.R., Walden, R.T., and Woods, R.W. (1964). CANCER OF THE LUNG AND MOUTH IN SEVENTH-DAY ADVENTISTS. PRELIMINARY REPORT ON A POPULATION STUDY. *Cancer* 17, 486-497. 10.1002/1097-0142(196404)17:4<486::aid-cncr2820170410>3.0.co;2-z.
46. Phillips, R.L., Garfinkel, L., Kuzma, J.W., Beeson, W.L., Lotz, T., and Brin, B. (1980). Mortality among California Seventh-Day Adventists for selected cancer sites. *J Natl Cancer Inst* 65, 1097-1107.
47. Alcohol drinking. (1988). IARC Monogr Eval Carcinog Risks Hum 44, 1-378.
48. Bagnardi, V., Rota, M., Botteri, E., Tramacere, I., Islami, F., Fedirko, V., Scotti, L., Jenab, M., Turati, F., Pasquali, E., et al. (2015). Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer* 112, 580-593. 10.1038/bjc.2014.579.
49. Giovannucci, E., Stampfer, M.J., Colditz, G.A., Rimm, E.B., Trichopoulos, D., Rosner, B.A., Speizer, F.E., and Willett, W.C. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J Natl Cancer Inst* 85, 875-884. 10.1093/jnci/85.11.875.
50. Ferrari, P., Jenab, M., Norat, T., Moskal, A., Slimani, N., Olsen, A., Tjønneland, A., Overvad, K., Jensen, M.K., Boutron-Ruault, M.C., et al. (2007). Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). *Int J Cancer* 121, 2065-2072. 10.1002/ijc.22966.
51. Smith-Warner, S.A., Spiegelman, D., Yaun, S.S., van den Brandt, P.A., Folsom, A.R., Goldbohm, R.A., Graham, S., Holmberg, L., Howe, G.R., Marshall, J.R., et al. (1998). Alcohol and breast cancer in women: a pooled analysis of cohort studies. *Jama* 279, 535-540. 10.1001/jama.279.7.535.
52. Graham, S. (1987). Alcohol and Breast Cancer. *New England Journal of Medicine* 316, 1211-1213. doi:10.1056/NEJM198705073161908.
53. Hamajima, N., Hirose, K., Tajima, K., Rohan, T., Calle, E.E., Heath, C.W., Jr., Coates, R.J., Liff, J.M., Talamini, R., Chantarakul, N., et al. (2002). Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer* 87, 1234-1245. 10.1038/sj.bjc.6600596.
54. Alcohol consumption and ethyl carbamate. (2010). IARC Monogr Eval Carcinog Risks Hum 96, 3-1383.
55. Pöschl, G., and Seitz, H.K. (2004). ALCOHOL AND CANCER. *Alcohol and Alcoholism* 39, 155-165. 10.1093/alcalc/agh057.

56. Druesne-Pecollo, N., Tehard, B., Mallet, Y., Gerber, M., Norat, T., Hercberg, S., and Latino-Martel, P. (2009). Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. *Lancet Oncol* *10*, 173-180. 10.1016/s1470-2045(09)70019-1.
57. Runggay, H., Murphy, N., Ferrari, P., and Soerjomataram, I. (2021). Alcohol and Cancer: Epidemiology and Biological Mechanisms. *Nutrients* *13*. 10.3390/nu13093173.
58. Hashibe, M., Brennan, P., Chuang, S.C., Boccia, S., Castellsague, X., Chen, C., Curado, M.P., Dal Maso, L., Daudt, A.W., Fabianova, E., et al. (2009). Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev* *18*, 541-550. 10.1158/1055-9965.Epi-08-0347.
59. Zeka, A., Gore, R., and Kriebel, D. (2003). Effects of alcohol and tobacco on aerodigestive cancer risks: a meta-regression analysis. *Cancer Causes Control* *14*, 897-906. 10.1023/b:caco.0000003854.34221.a8.
60. Services, U.S.D.o.A.a.U.S.D.o.H.a.H. (December 2020). Dietary Guidelines for Americans, 2020-2025. . <https://www.dietaryguidelines.gov/>.
61. Office of the Surgeon, G. (2025). Publications and Reports of the Surgeon General. In Alcohol and Cancer Risk: The U.S. Surgeon General's Advisory, (US Department of Health and Human Services).
62. Flegal, K.M., Carroll, M.D., Kuczmarski, R.J., and Johnson, C.L. (1998). Overweight and obesity in the United States: prevalence and trends, 1960-1994. *Int J Obes Relat Metab Disord* *22*, 39-47. 10.1038/sj.ijo.0800541.
63. Kuczmarski, R.J., Flegal, K.M., Campbell, S.M., and Johnson, C.L. (1994). Increasing prevalence of overweight among US adults. The National Health and Nutrition Examination Surveys, 1960 to 1991. *Jama* *272*, 205-211. 10.1001/jama.272.3.205.
64. Lew, E.A., and Garfinkel, L. (1979). Variations in mortality by weight among 750,000 men and women. *J Chronic Dis* *32*, 563-576. 10.1016/0021-9681(79)90119-x.
65. Calle, E.E., Rodriguez, C., Walker-Thurmond, K., and Thun, M.J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* *348*, 1625-1638. 10.1056/NEJMoa021423.
66. Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., Qizilbash, N., Collins, R., and Peto, R. (2009). Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* *373*, 1083-1096. 10.1016/s0140-6736(09)60318-4.
67. Cancer, I.A.f.R.o. (2002). Weight control and physical activity. In Weight control and physical activity.
68. Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., and Straif, K. (2016). Body Fatness and Cancer--Viewpoint of the IARC Working Group. *N Engl J Med* *375*, 794-798. 10.1056/NEJMsrl606602.
69. Hopkins, B.D., Goncalves, M.D., and Cantley, L.C. (2016). Obesity and Cancer Mechanisms: Cancer Metabolism. *J Clin Oncol* *34*, 4277-4283. 10.1200/jco.2016.67.9712.
70. Avgerinos, K.I., Spyrou, N., Mantzoros, C.S., and Dalamaga, M. (2019). Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism* *92*, 121-135. 10.1016/j.metabol.2018.11.001.
71. Rakib, A., Mamun, M.A.A., Mandal, M., Sinha, P., and Singh, U.P. (2025). Obesity-cancer axis crosstalk: Molecular insights and therapeutic approaches. *Acta Pharm Sin B* *15*, 2930-2944. 10.1016/j.apsb.2025.04.029.
72. Du, M., Drew, D.A., Goncalves, M.D., Cao, Y., and Chan, A.T. (2025). Early-onset colorectal cancer as an emerging disease of metabolic dysregulation. *Nat Rev Endocrinol* *21*, 686-702. 10.1038/s41574-025-01159-z.
73. Qing, T., Mohsen, H., Marczyk, M., Ye, Y., O'Meara, T., Zhao, H., Townsend, J.P., Gerstein, M., Hatzis, C., Kluger, Y., and Puztai, L. (2020). Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. *Nat Commun* *11*, 2438. 10.1038/s41467-020-16293-7.
74. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., and et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* *266*, 66-71. 10.1126/science.7545954.

75. Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., and et al. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 265, 2088-2090. 10.1126/science.8091231.
76. Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B., and King, M.C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684-1689. 10.1126/science.2270482.
77. Peltomäki, P., Aaltonen, L.A., Sistonen, P., Pylkkänen, L., Mecklin, J.P., Järvinen, H., Green, J.S., Jass, J.R., Weber, J.L., Leach, F.S., and et al. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260, 810-812. 10.1126/science.8484120.
78. Lindblom, A., Tannergård, P., Werelius, B., and Nordenskjöld, M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet* 5, 279-282. 10.1038/ng1193-279.
79. Kinzler, K.W., Nilbert, M.C., Su, L.K., Vogelstein, B., Bryan, T.M., Levy, D.B., Smith, K.J., Preisinger, A.C., Hedge, P., McKechnie, D., and et al. (1991). Identification of FAP locus genes from chromosome 5q21. *Science* 253, 661-665. 10.1126/science.1651562.
80. Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 75, 1027-1038. 10.1016/0092-8674(93)90546-3.
81. Sud, A., Kinnersley, B., and Houlston, R.S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer* 17, 692-704. 10.1038/nrc.2017.82.
82. Zhang, H., Ahearn, T.U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T.A., Zhao, N., Bolla, M.K., et al. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* 52, 572-581. 10.1038/s41588-020-0609-2.
83. Fernandez-Rozadilla, C., Timofeeva, M., Chen, Z., Law, P., Thomas, M., Schmit, S., Díez-Obrero, V., Hsu, L., Fernandez-Tajes, J., Palles, C., et al. (2023). Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat Genet* 55, 89-99. 10.1038/s41588-022-01222-9.
84. Law, P.J., Studd, J., Smith, J., Vijaykrishnan, J., Harris, B.T., Mandelia, M., Mills, C., Dunlop, M.G., and Houlston, R.S. (2024). Systematic prioritization of functional variants and effector genes underlying colorectal cancer risk. *Nat Genet* 56, 2104-2111. 10.1038/s41588-024-01900-w.
85. Bossé, Y., and Amos, C.I. (2018). A Decade of GWAS Results in Lung Cancer. *Cancer Epidemiol Biomarkers Prev* 27, 363-379. 10.1158/1055-9965.Epi-16-0794.
86. Gorman, B.R., Ji, S.G., Francis, M., Sendamarai, A.K., Shi, Y., Devineni, P., Saxena, U., Partan, E., DeVito, A.K., Byun, J., et al. (2024). Multi-ancestry GWAS meta-analyses of lung cancer reveal susceptibility loci and elucidate smoking-independent genetic risk. *Nat Commun* 15, 8629. 10.1038/s41467-024-52129-4.
87. Qian, D.C., Byun, J., Han, Y., Greene, C.S., Field, J.K., Hung, R.J., Brhane, Y., McLaughlin, J.R., Fehringer, G., Landi, M.T., et al. (2015). Identification of shared and unique susceptibility pathways among cancers of the lung, breast, and prostate from genome-wide association studies and tissue-specific protein interactions. *Hum Mol Genet* 24, 7406-7420. 10.1093/hmg/ddv440.
88. Kuenzi, B.M., and Ideker, T. (2020). A census of pathway maps in cancer systems biology. *Nat Rev Cancer* 20, 233-246. 10.1038/s41568-020-0240-7.
89. Huntley, C., Torr, B., Sud, A., Rowlands, C.F., Way, R., Snape, K., Hanson, H., Swanton, C., Broggio, J., Lucassen, A., et al. (2023). Utility of polygenic risk scores in UK cancer screening: a modelling analysis. *Lancet Oncol* 24, 658-668. 10.1016/s1470-2045(23)00156-0.
90. Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.H., Wang, Q., Bolla, M.K., et al. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 104, 21-34. 10.1016/j.ajhg.2018.11.002.

91. Thomas, M., Sakoda, L.C., Hoffmeister, M., Rosenthal, E.A., Lee, J.K., van Duijnhoven, F.J.B., Platz, E.A., Wu, A.H., Dampier, C.H., de la Chapelle, A., et al. (2020). Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am J Hum Genet* 107, 432-444. 10.1016/j.ajhg.2020.07.006.
92. Peto, R. (2023). The need for Ignorance in cancer research. *Eur J Epidemiol* 38, 1239-1243. 10.1007/s10654-023-01085-2.
93. Islami, F., Nargis, N., Liu, Q., Bandi, P., Siegel, R.L., Choudhury, P.P., Freedman, N.D., Warner, K.E., and Jemal, A. (2025). Averted lung cancer deaths due to reductions in cigarette smoking in the United States, 1970-2022. *CA Cancer J Clin* 75, 216-225. 10.3322/caac.70005.
94. Swinburn, B.A., Kraak, V.I., Allender, S., Atkins, V.J., Baker, P.I., Bogard, J.R., Brinsden, H., Calvillo, A., De Schutter, O., Devarajan, R., et al. (2019). The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report. *Lancet* 393, 791-846. 10.1016/s0140-6736(18)32822-8.
95. Dai, H., Li, Y., Lee, Y.A., Lu, Y., George, T.J., Donahoo, W.T., Lee, K.P., Nakshatri, H., Allen, J., Guo, Y., et al. (2025). GLP-1 Receptor Agonists and Cancer Risk in Adults With Obesity. *JAMA Oncol* 11, 1186-1193. 10.1001/jamaoncol.2025.2681.
96. Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 8, 1-12. 10.1038/bjc.1954.1.
97. Mahabir, S., Aagaard, K., Anderson, L.M., Herceg, Z., Hiatt, R.A., Hoover, R.N., Linet, M.S., Medina, D., Potischman, N., Tretli, S., et al. (2012). Challenges and opportunities in research on early-life events/exposures and cancer development later in life. *Cancer Causes Control* 23, 983-990. 10.1007/s10552-012-9962-5.
98. Calabrese, E.J. (2021). Ethical failings: The problematic history of cancer risk assessment. *Environ Res* 193, 110582. 10.1016/j.envres.2020.110582.
99. Akhtar, A. (2015). The flaws and human harms of animal experimentation. *Camb Q Healthc Ethics* 24, 407-419. 10.1017/s0963180115000079.
100. Cheluvappa, R., Scowen, P., and Eri, R. (2017). Ethics of animal research in human disease remediation, its institutional teaching; and alternatives to animal experimentation. *Pharmacol Res Perspect* 5. 10.1002/prp2.332.
101. Patiño-García, A., Gुरुceaga, E., Andueza, M.P., Ocón, M., Fodop Sokoudjou, J.J., de Villalonga Zornoza, N., Alkorta-Aranburu, G., Uria, I.T., Gurrpide, A., Camps, C., et al. (2024). Whole exome sequencing and machine learning germline analysis of individuals presenting with extreme phenotypes of high and low risk of developing tobacco-associated lung adenocarcinoma. *EBioMedicine* 102, 105048. 10.1016/j.ebiom.2024.105048.
102. Foulkes, W.D., Knoppers, B.M., and Turnbull, C. (2016). Population genetic testing for cancer susceptibility: founder mutations to genomes. *Nat Rev Clin Oncol* 13, 41-54. 10.1038/nrclinonc.2015.173.
103. Greaves, M. (2015). Evolutionary determinants of cancer. *Cancer Discov* 5, 806-820. 10.1158/2159-8290.Cd-15-0439.
104. Alfaro-Murillo, J.A., and Townsend, J.P. (2023). Pairwise and higher-order epistatic effects among somatic cancer mutations across oncogenesis. *Math Biosci* 366, 109091. 10.1016/j.mbs.2023.109091.
105. McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* 168, 613-628. 10.1016/j.cell.2017.01.018.
106. Hunter, D.J. (2005). Gene-environment interactions in human diseases. *Nat Rev Genet* 6, 287-298. 10.1038/nrg1578.
107. Hutter, C.M., Mechanic, L.E., Chatterjee, N., Kraft, P., and Gillanders, E.M. (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol* 37, 643-657. 10.1002/gepi.21756.
108. Ahsan, H., Halpern, J., Kibriya, M.G., Pierce, B.L., Tong, L., Gamazon, E., McGuire, V., Felberg, A., Shi, J., Jasmine, F., et al. (2014). A genome-wide association study of early-onset breast cancer identifies PFKM as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. *Cancer Epidemiol Biomarkers Prev* 23, 658-669. 10.1158/1055-9965.Epi-13-0340.

109. Laskar, R.S., Qu, C., Huyghe, J.R., Harrison, T., Hayes, R.B., Cao, Y., Campbell, P.T., Steinfeld, R., Talukdar, F.R., Brenner, H., et al. (2024). Genome-wide association studies and Mendelian randomization analyses provide insights into the causes of early-onset colorectal cancer. *Ann Oncol* 35, 523-536. 10.1016/j.annonc.2024.02.008.
110. Siegel, R.L., Fedewa, S.A., Anderson, W.F., Miller, K.D., Ma, J., Rosenberg, P.S., and Jemal, A. (2017). Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J Natl Cancer Inst* 109. 10.1093/jnci/djw322.
111. Chen, J., and Yang, W. (2024). Comparing cohort and period trends of early-onset colorectal cancer: a global analysis. *JNCI Cancer Spectr* 8. 10.1093/jncics/pkae052.
112. Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics* 39, 311-324.
113. Liu, Y., Colditz, G.A., Rosner, B., Berkey, C.S., Collins, L.C., Schnitt, S.J., Connolly, J.L., Chen, W.Y., Willett, W.C., and Tamimi, R.M. (2013). Alcohol intake between menarche and first pregnancy: a prospective study of breast cancer risk. *J Natl Cancer Inst* 105, 1571-1578. 10.1093/jnci/djt213.
114. White, A.J., DeRoo, L.A., Weinberg, C.R., and Sandler, D.P. (2017). Lifetime Alcohol Intake, Binge Drinking Behaviors, and Breast Cancer Risk. *Am J Epidemiol* 186, 541-549. 10.1093/aje/kwx118.
115. Lupi, M., Martinotti, G., and Di Giannantonio, M. (2017). Drunkorexia: an emerging trend in young adults. *Eat Weight Disord* 22, 619-622. 10.1007/s40519-017-0429-2.
116. Wild, C.P. (2012). The exposome: from concept to utility. *Int J Epidemiol* 41, 24-32. 10.1093/ije/dyr236.
117. Jones, D.P. (2016). Sequencing the exposome: A call to action. *Toxicol Rep* 3, 29-45. 10.1016/j.toxrep.2015.11.009.
118. Stancliffe, E., Schwaiger-Haber, M., Sindelar, M., Murphy, M.J., Soerensen, M., and Patti, G.J. (2022). An Untargeted Metabolomics Workflow that Scales to Thousands of Samples for Population-Based Studies. *Anal Chem* 94, 17370-17378. 10.1021/acs.analchem.2c01270.
119. Swanton, C., Bernard, E., Abbosh, C., André, F., Auwerx, J., Balmain, A., Bar-Sagi, D., Bernards, R., Bullman, S., DeGregori, J., et al. (2024). Embracing cancer complexity: Hallmarks of systemic disease. *Cell* 187, 1589-1616. 10.1016/j.cell.2024.02.009.
120. Cannataro, V.L., Mandell, J.D., and Townsend, J.P. (2022). Attribution of Cancer Origins to Endogenous, Exogenous, and Preventable Mutational Processes. *Mol Biol Evol* 39. 10.1093/molbev/msac084.
121. Gapstur, S.M., Bouvard, V., Nethan, S.T., Freudenheim, J.L., Abnet, C.C., English, D.R., Rehm, J., Balbo, S., Buykx, P., Crabb, D., et al. (2023). The IARC Perspective on Alcohol Reduction or Cessation and Cancer Risk. *N Engl J Med* 389, 2486-2494. 10.1056/NEJMSr2306723.
122. National Center for Chronic Disease, P., Health Promotion Office on, S., and Health (2014). Reports of the Surgeon General. In *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*, (Centers for Disease Control and Prevention (US)).
123. Goldmann, J.M., Veltman, J.A., and Gilissen, C. (2019). De Novo Mutations Reflect Development and Aging of the Human Germline. *Trends Genet* 35, 828-839. 10.1016/j.tig.2019.08.005.
124. Kim, H.S., Xiao, X., Byun, J., Jun, G., DeSantis, S.M., Chen, H., Thrift, A.P., El-Serag, H.B., Kanwal, F., and Amos, C.I. (2022). Synergistic Associations of PNPLA3 I148M Variant, Alcohol Intake, and Obesity With Risk of Cirrhosis, Hepatocellular Carcinoma, and Mortality. *JAMA Netw Open* 5, e2234221. 10.1001/jamanetworkopen.2022.34221.
125. Zhang, Y., Lindström, S., Kraft, P., and Liu, Y. (2025). Genetic risk, health-associated lifestyle, and risk of early-onset total cancer and breast cancer. *J Natl Cancer Inst* 117, 40-48. 10.1093/jnci/djae208.
126. Zhou, R., Tang, X., and Wang, Y. (2024). Emerging strategies to investigate the biology of early cancer. *Nat Rev Cancer* 24, 850-866. 10.1038/s41568-024-00754-y.
127. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., et al. (2013). Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 42, 97-110. 10.1093/ije/dys066.

128. Olsen, J., Melbye, M., Olsen, S.F., Sørensen, T.I., Aaby, P., Andersen, A.M., Taxbøl, D., Hansen, K.D., Juhl, M., Schow, T.B., et al. (2001). The Danish National Birth Cohort--its background, structure and aim. *Scand J Public Health* 29, 300-307. 10.1177/14034948010290040201.
129. Scholtens, S., Smidt, N., Swertz, M.A., Bakker, S.J., Dotinga, A., Vonk, J.M., van Dijk, F., van Zon, S.K., Wijmenga, C., Wolffenbuttel, B.H., and Stolk, R.P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* 44, 1172-1180. 10.1093/ije/dyu229.
130. Hofman, A., Jaddoe, V.W., Mackenbach, J.P., Moll, H.A., Snijders, R.F., Steegers, E.A., Verhulst, F.C., Witteman, J.C., and Büller, H.A. (2004). Growth, development and health from early fetal life until young adulthood: the Generation R Study. *Paediatr Perinat Epidemiol* 18, 61-72. 10.1111/j.1365-3016.2003.00521.x.
131. Cancer, I.A.f.R.o. DISCERN: Discovering the Causes of Three Poorly Understood Cancers in Europe. <https://discern.iarc.who.int/>.
132. Jiang, S., and Colditz, G.A. (2023). Causal mediation analysis using high-dimensional image mediator bounded in irregular domain with an application to breast cancer. *Biometrics* 79, 3728-3738. 10.1111/biom.13847.
133. Wagner, C., Carmeli, C., Jackisch, J., Kivimäki, M., van der Linden, B.W.A., Cullati, S., and Chioloro, A. (2024). Life course epidemiology and public health. *Lancet Public Health* 9, e261-e269. 10.1016/s2468-2667(24)00018-5.
134. Richardson, T.G., Urquijo, H., Holmes, M.V., and Davey Smith, G. (2023). Leveraging family history data to disentangle time-varying effects on disease risk using lifecourse mendelian randomization. *Eur J Epidemiol* 38, 765-769. 10.1007/s10654-023-01001-8.
135. Naimi, A.I., Cole, S.R., and Kennedy, E.H. (2017). An introduction to g methods. *Int J Epidemiol* 46, 756-762. 10.1093/ije/dyw323.
136. De Stavola, B.L., and Daniel, R.M. (2017). Commentary: Incorporating concepts and methods from causal inference into life course epidemiology. *Int J Epidemiol* 46, 771. 10.1093/ije/dyw367.
137. Hernán, M.A., Wang, W., and Leaf, D.E. (2022). Target Trial Emulation: A Framework for Causal Inference From Observational Data. *Jama* 328, 2446-2447. 10.1001/jama.2022.21383.
138. Campbell, P.J. (2023). Demystifying the black box: from ignorance to observation to mechanism in cancer research. *Eur J Epidemiol* 38, 1265-1267. 10.1007/s10654-022-00935-9.
139. Plutynski, A. (2023). On Explaining Peto's Paradox. *Eur J Epidemiol* 38, 1245-1250. 10.1007/s10654-022-00920-2.
140. Davey Smith, G., Hofman, A., and Brennan, P. (2023). Chance, ignorance, and the paradoxes of cancer: Richard Peto on developing preventative strategies under uncertainty. *Eur J Epidemiol* 38, 1227-1237. 10.1007/s10654-023-01090-5.
141. Boddy, A.M. (2023). The need for evolutionary theory in cancer research. *Eur J Epidemiol* 38, 1259-1264. 10.1007/s10654-022-00936-8.
142. Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D., et al. (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 364. 10.1126/science.aaw0726.
143. Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., et al. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 574, 532-537. 10.1038/s41586-019-1672-7.
144. Hill, W., Lim, E.L., Weeden, C.E., Lee, C., Augustine, M., Chen, K., Kuan, F.C., Marongiu, F., Evans, E.J., Jr., Moore, D.A., et al. (2023). Lung adenocarcinoma promotion by air pollutants. *Nature* 616, 159-167. 10.1038/s41586-023-05874-3.
145. Abby, E., Dentre, S.C., Hall, M.W.J., Fowler, J.C., Ong, S.H., Sood, R., Herms, A., Piedrafita, G., Abnizova, I., Siebel, C.W., et al. (2023). Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. *Nat Genet* 55, 232-245. 10.1038/s41588-022-01280-z.
146. Wang, T., Pehrsson, E.C., Purushotham, D., Li, D., Zhuo, X., Zhang, B., Lawson, H.A., Province, M.A., Krapp, C., Lan, Y., et al. (2018). The NIEHS TaRGET II Consortium and environmental epigenomics. *Nat Biotechnol* 36, 225-227. 10.1038/nbt.4099.

147. Zhang, B.A., Miao, B., Fu, S., Coarfa, C., Kumar, R., Kuntala, P.K., Park, B., Grimm, S.L., Jangid, R., Colacino, J.A., et al. (2025). Comprehensive Transcriptomic and Epigenomic Insights into Environmental Toxicant Exposures: The TaRGET II Resource. *bioRxiv*. 10.1101/2025.07.28.667191.
148. Lim, A.I., McFadden, T., Link, V.M., Han, S.J., Karlsson, R.M., Stacy, A., Farley, T.K., Lima-Junior, D.S., Harrison, O.J., Desai, J.V., et al. (2021). Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring. *Science* 373. 10.1126/science.abf3002.
149. Naik, S., and Fuchs, E. (2022). Inflammatory memory and tissue adaptation in sickness and in health. *Nature* 607, 249-255. 10.1038/s41586-022-04919-3.
150. Stratton, M.R., Humphreys, L., Alexandrov, L.B., Balmain, A., Brennan, P., Campbell, P.J., and Phillips, D.H. (2025). Implementing Mutational Epidemiology on a Global Scale: Lessons from Mutographs. *Cancer Discov* 15, 22-27. 10.1158/2159-8290.Cd-24-1687.
151. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94-101. 10.1038/s41586-020-1943-3.
152. Díaz-Gay, M., Dos Santos, W., Moody, S., Kazachkova, M., Abbasi, A., Steele, C.D., Vangara, R., Senkin, S., Wang, J., Fitzgerald, S., et al. (2025). Geographic and age variations in mutational processes in colorectal cancer. *Nature* 643, 230-240. 10.1038/s41586-025-09025-8.
153. Coorens, T.H.H., Oh, J.W., Choi, Y.A., Lim, N.S., Zhao, B., Voshall, A., Abyzov, A., Antonacci-Fulton, L., Aparicio, S., Ardlie, K.G., et al. (2025). The Somatic Mosaicism across Human Tissues Network. *Nature* 643, 47-59. 10.1038/s41586-025-09096-7.
154. Gluckman, P.D., Hanson, M.A., Cooper, C., and Thornburg, K.L. (2008). Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med* 359, 61-73. 10.1056/NEJMra0708473.
155. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13, 97-109. 10.1038/nrg3142.
156. Walker, C.L., and Ho, S.M. (2012). Developmental reprogramming of cancer susceptibility. *Nat Rev Cancer* 12, 479-486. 10.1038/nrc3220.
157. Fitz-James, M.H., and Cavalli, G. (2022). Molecular mechanisms of transgenerational epigenetic inheritance. *Nat Rev Genet* 23, 325-341. 10.1038/s41576-021-00438-5.
158. Lesch, B.J., Tothova, Z., Morgan, E.A., Liao, Z., Bronson, R.T., Ebert, B.L., and Page, D.C. (2019). Intergenerational epigenetic inheritance of cancer susceptibility in mammals. *Elife* 8. 10.7554/eLife.39380.
159. Al Nabhani, Z., and Eberl, G. (2020). Imprinting of the immune system by the microbiota early in life. *Mucosal Immunol* 13, 183-189. 10.1038/s41385-020-0257-y.
160. Patti, M.E. (2013). Intergenerational programming of metabolic disease: evidence from human populations and experimental animal models. *Cell Mol Life Sci* 70, 1597-1608. 10.1007/s00018-013-1298-0.
161. Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121-141. 10.1016/j.cell.2014.03.011.
162. Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. *Cell* 140, 883-899. 10.1016/j.cell.2010.01.025.
163. Sun, Y., Wang, Q., Zhang, Y., Geng, M., Wei, Y., Liu, Y., Liu, S., Petersen, R.B., Yue, J., Huang, K., and Zheng, L. (2020). Multigenerational maternal obesity increases the incidence of HCC in offspring via miR-27a-3p. *J Hepatol* 73, 603-615. 10.1016/j.jhep.2020.03.050.
164. Hilakivi-Clarke, L. (2014). Maternal exposure to diethylstilbestrol during pregnancy and increased breast cancer risk in daughters. *Breast Cancer Res* 16, 208. 10.1186/bcr3649.
165. Merlo, L.M., Pepper, J.W., Reid, B.J., and Maley, C.C. (2006). Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 6, 924-935. 10.1038/nrc2013.
166. Laplane, L., and Maley, C.C. (2024). The evolutionary theory of cancer: challenges and potential solutions. *Nat Rev Cancer* 24, 718-733. 10.1038/s41568-024-00734-2.

167. Somarelli, J.A., Gardner, H., Cannataro, V.L., Gunady, E.F., Boddy, A.M., Johnson, N.A., Fisk, J.N., Gaffney, S.G., Chuang, J.H., Li, S., et al. (2020). Molecular Biology and Evolution of Cancer: From Discovery to Action. *Mol Biol Evol* 37, 320-326. 10.1093/molbev/msz242.
168. Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* 481, 306-313. 10.1038/nature10762.
169. Cannataro, V.L., Gaffney, S.G., and Townsend, J.P. (2018). Effect Sizes of Somatic Mutations in Cancer. *J Natl Cancer Inst* 110, 1171-1177. 10.1093/jnci/djy168.
170. Mandell, J.D., Cannataro, V.L., and Townsend, J.P. (2023). Estimation of Neutral Mutation Rates and Quantification of Somatic Variant Selection Using *cancereffectsizer*. *Cancer Res* 83, 500-505. 10.1158/0008-5472.Can-22-1508.
171. Mauri, G., Patelli, G., Crisafulli, G., Siena, S., and Bardelli, A. (2025). Tumor "age" in early-onset colorectal cancer. *Cell* 188, 589-593. 10.1016/j.cell.2024.12.003.
172. Ashley, D.J. (1969). The two "hit" and multiple "hit" theories of carcinogenesis. *Br J Cancer* 23, 313-328. 10.1038/bjc.1969.41.
173. Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Lee, S.E., Yousef, H., Moran Losada, P., Berdnik, D., Keller, A., Verghese, J., et al. (2019). Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med* 25, 1843-1850. 10.1038/s41591-019-0673-2.
174. Oh, H.S., Rutledge, J., Nachun, D., Pálovics, R., Abiose, O., Moran-Losada, P., Channappa, D., Urey, D.Y., Kim, K., Sung, Y.J., et al. (2023). Organ aging signatures in the plasma proteome track health and disease. *Nature* 624, 164-172. 10.1038/s41586-023-06802-1.
175. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., and Mulvihill, J.J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 81, 1879-1886. 10.1093/jnci/81.24.1879.
176. Antoniou, A.C., Pharoah, P.D., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., and Easton, D.F. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 86, 76-83. 10.1038/sj.bjc.6600008.
177. Freedman, A.N., Slattery, M.L., Ballard-Barbash, R., Willis, G., Cann, B.J., Pee, D., Gail, M.H., and Pfeiffer, R.M. (2009). Colorectal cancer risk prediction tool for white men and women without known susceptibility. *J Clin Oncol* 27, 686-693. 10.1200/jco.2008.17.4797.
178. Kastrinos, F., Kupfer, S.S., and Gupta, S. (2023). Colorectal Cancer Risk Assessment and Precision Approaches to Screening: Brave New World or Worlds Apart? *Gastroenterology* 164, 812-827. 10.1053/j.gastro.2023.02.021.
179. Pashayan, N., Antoniou, A.C., Ivanus, U., Esserman, L.J., Easton, D.F., French, D., Sroczynski, G., Hall, P., Cuzick, J., Evans, D.G., et al. (2020). Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat Rev Clin Oncol* 17, 687-705. 10.1038/s41571-020-0388-9.
180. Louro, J., Posso, M., Hilton Boon, M., Román, M., Domingo, L., Castells, X., and Sala, M. (2019). A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br J Cancer* 121, 76-85. 10.1038/s41416-019-0476-8.
181. Hull, M.A., Rees, C.J., Sharp, L., and Koo, S. (2020). A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nat Rev Gastroenterol Hepatol* 17, 773-780. 10.1038/s41575-020-00368-3.
182. Gluckman, P.D., and Hanson, M.A. (2004). Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. *Pediatr Res* 56, 311-317. 10.1203/01.Pdr.0000135998.08025.Fb.
183. Gluckman, P.D., and Hanson, M.A. (2004). Living with the past: evolution, development, and patterns of disease. *Science* 305, 1733-1736. 10.1126/science.1095292.
184. Rosner, B., and Colditz, G.A. (1996). Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst* 88, 359-364. 10.1093/jnci/88.6.359.
185. Colditz, G.A., and Rosner, B. (2000). Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol* 152, 950-964. 10.1093/aje/152.10.950.

186. Rice, M.S., Tworoger, S.S., Hankinson, S.E., Tamimi, R.M., Eliassen, A.H., Willett, W.C., Colditz, G., and Rosner, B. (2017). Breast cancer risk prediction: an update to the Rosner-Colditz breast cancer incidence model. *Breast Cancer Res Treat* 166, 227-240. 10.1007/s10549-017-4391-5.
187. Baccarelli, A., Dolinoy, D.C., and Walker, C.L. (2023). A precision environmental health approach to prevention of human disease. *Nat Commun* 14, 2449. 10.1038/s41467-023-37626-2.
188. Kachuri, L., Graff, R.E., Smith-Byrne, K., Meyers, T.J., Rashkin, S.R., Ziv, E., Witte, J.S., and Johansson, M. (2020). Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun* 11, 6084. 10.1038/s41467-020-19600-4.
189. Yates, J., and Van Allen, E.M. (2025). New horizons at the interface of artificial intelligence and translational cancer research. *Cancer Cell* 43, 708-727. 10.1016/j.ccell.2025.03.018.
190. Jiang, S., Bennett, D.L., and Colditz, G.A. (2025). Validation of a Dynamic Risk Prediction Model Incorporating Prior Mammograms in a Diverse Population. *JAMA Netw Open* 8, e2512681. 10.1001/jamanetworkopen.2025.12681.
191. Jiang, S., Bennett, D.L., Rosner, B.A., Tamimi, R.M., and Colditz, G.A. (2024). Development and Validation of Dynamic 5-Year Breast Cancer Risk Model Using Repeated Mammograms. *JCO Clin Cancer Inform* 8, e2400200. 10.1200/cci-24-00200.
192. Schopf, C.M., Ramwala, O.A., Lowry, K.P., Hofvind, S., Marinovich, M.L., Houssami, N., Elmore, J.G., Dontchos, B.N., Lee, J.M., and Lee, C.I. (2024). Artificial Intelligence-Driven Mammography-Based Future Breast Cancer Risk Prediction: A Systematic Review. *J Am Coll Radiol* 21, 319-328. 10.1016/j.jacr.2023.10.018.
193. Shmatko, A., Jung, A.W., Gaurav, K., Brunak, S., Mortensen, L.H., Birney, E., Fitzgerald, T., and Gerstung, M. (2025). Learning the natural history of human disease with generative transformers. *Nature* 647, 248-256. 10.1038/s41586-025-09529-3.
194. Rockhill, B., Newman, B., and Weinberg, C. (1998). Use and misuse of population attributable fractions. *Am J Public Health* 88, 15-19. 10.2105/ajph.88.1.15.
195. Colditz, G.A., and Wei, E.K. (2012). Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality. *Annu Rev Public Health* 33, 137-156. 10.1146/annurev-publhealth-031811-124627.
196. Stefan, N., Schick, F., and Häring, H.U. (2017). Causes, Characteristics, and Consequences of Metabolically Unhealthy Normal Weight in Humans. *Cell Metab* 26, 292-300. 10.1016/j.cmet.2017.07.008.
197. Hazelton, W.D., Ni, P., Harlass, M., Hahn, A.I., Tian, R., Zaubner, A.G., Lansdorp-Vogelaar, I., and Cao, Y. (2024). Projected Trends of Obesity throughout the Life Course According to Sex, Race, and Birth Cohorts in the United States. medRxiv, 2024.2010.2015.24315456. 10.1101/2024.10.15.24315456.
198. Institute, N.C. Cancer Intervention and Surveillance Modeling Network. <https://cisnet.cancer.gov/>.
199. Caswell-Jin, J.L., Sun, L.P., Munoz, D., Lu, Y., Li, Y., Huang, H., Hampton, J.M., Song, J., Jayasekera, J., Schechter, C., et al. (2024). Analysis of Breast Cancer Mortality in the US-1975 to 2019. *Jama* 331, 233-241. 10.1001/jama.2023.25881.
200. Moolgavkar, S.H., Holford, T.R., Levy, D.T., Kong, C.Y., Foy, M., Clarke, L., Jeon, J., Hazelton, W.D., Meza, R., Schultz, F., et al. (2012). Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000. *J Natl Cancer Inst* 104, 541-548. 10.1093/jnci/djs136.
201. Edwards, B.K., Ward, E., Kohler, B.A., Ehemann, C., Zaubner, A.G., Anderson, R.N., Jemal, A., Schymura, M.J., Lansdorp-Vogelaar, I., Seeff, L.C., et al. (2010). Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 116, 544-573. 10.1002/cncr.24760.
202. Vogelaar, I., van Ballegooijen, M., Schrag, D., Boer, R., Winawer, S.J., Habbema, J.D., and Zaubner, A.G. (2006). How much can current interventions reduce colorectal cancer mortality in the U.S.? Mortality projections for scenarios of risk-factor modification, screening, and treatment. *Cancer* 107, 1624-1633. 10.1002/cncr.22115.
203. Hazelton, W.D., Jeon, J., Meza, R., and Moolgavkar, S.H. (2012). Chapter 8: The FHCRC lung cancer model. *Risk Anal* 32 Suppl 1, S99-s116. 10.1111/j.1539-6924.2011.01681.x.

204. Jeon, J., Meza, R., Hazelton, W.D., Renehan, A.G., and Luebeck, E.G. (2015). Incremental benefits of screening colonoscopy over sigmoidoscopy in average-risk populations: a model-driven analysis. *Cancer Causes Control* 26, 859-870. 10.1007/s10552-015-0559-7.
205. Curtius, K., Hazelton, W.D., Jeon, J., and Luebeck, E.G. (2015). A Multiscale Model Evaluates Screening for Neoplasia in Barrett's Esophagus. *PLoS Comput Biol* 11, e1004272. 10.1371/journal.pcbi.1004272.
206. Pepe, M.S., Etzioni, R., Feng, Z., Potter, J.D., Thompson, M.L., Thornquist, M., Winget, M., and Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 93, 1054-1061. 10.1093/jnci/93.14.1054.
207. Srivastava, S., and Wagner, P.D. (2020). The Early Detection Research Network: A National Infrastructure to Support the Discovery, Development, and Validation of Cancer Biomarkers. *Cancer Epidemiol Biomarkers Prev* 29, 2401-2410. 10.1158/1055-9965.Epi-20-0237.