Cell
Leading Edge

CellPress
OPEN ACCESS

Perspective

Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers

Mengyao Shi,[1,2] Gary J. Patti,[3,4,5,6] Marc J. Gunter,[7,8] Ramaswamy Govindan,[9,10] Iris Lansdorp-Vogelaar,[11] Ting Wang,[6,12,13,14] Jeffrey P. Townsend,[15,16,17,18] Graham A. Colditz,[1,10] Yasmine Belkaid,[19] and Yin Cao[1,10,20,21,*]

[1]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[2]Division of Biology and Biomedical Sciences, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[3]Department of Chemistry, Washington University in St. Louis, St. Louis, MO, USA
[4]Center for Mass Spectrometry and Metabolic Tracing, Washington University in St. Louis, St. Louis, MO, USA
[5]Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[6]Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[7]Nutrition and Metabolism Branch, International Agency for Research on Cancer, WHO, Lyon, France
[8]Cancer Epidemiology and Prevention Research Unit, School of Public Health, Imperial College London, London, UK
[9]Division of Oncology, Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[10]Alvin J. Siteman Cancer Center, Barnes-Jewish Hospital and Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[11]Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands
[12]Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[13]McDonnell Genome Institute, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[14]Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO, USA
[15]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[16]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
[17]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA
[18]Program in Genomics, Genetics, and Epigenetics, Yale Cancer Center, New Haven, CT, USA
[19]Metaorganism Unit, Department of Immunology, Institut Pasteur, Paris, France
[20]Division of Gastroenterology, Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
[21]Center for Translational Bioinformatics, Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
*Correspondence: yin.cao@wustl.edu
https://doi.org/10.1016/j.cell.2026.03.019

SUMMARY

Cancer in younger adults is rising globally, with notable birth-cohort effects. This epidemiological shift underscores the urgent need to accelerate the identification of novel causes and underlying biological networks, with the aim of translating these insights into prevention and interception strategies. In this perspective, we revisit the major milestones in the discovery of cancer causes and outline challenges that hinder progress. To address these challenges, we advocate closer integration of epidemiologic and mechanistic studies and propose three interconnected frameworks that extend current epidemiologic approaches: a tissue ecosystem-anchored framework for cancer cause discovery, a biological state-based framework for precision cancer risk assessment, and a dynamic framework to characterize cancer preventability. This roadmap aims to stimulate conceptual, resource, and methodological advances to accelerate cancer etiology research and prevention in the era of rising early-onset cancers.

INTRODUCTION

Cancers are no longer confined to the older population. Early-onset cancers, commonly defined as cancers diagnosed before age 50, are increasing rapidly worldwide (Figure 1).[1–3] Across 42 countries, 75% reported rising incidence in six cancer types (breast, colorectal, thyroid, kidney, endometrial, and leukemia) among adults younger than 50, with average annual increases of 0.8% to 3.6% between 2003 and 2017.[4] Globally, early-onset cancers account for nearly 1 million deaths and 50 million disability-adjusted life years,[5] imposing substantial personal, societal, and economic burdens. In the United States (US), while cancer mortality has declined among older populations, mortality under age 50 has plateaued since the 1990s and has risen for colorectal and endometrial cancers,[6] with colorectal cancer now the leading cause of cancer death in men.[6,7] In many countries, these increases show strong birth-cohort effects, with Generation X (born 1965–1980) and Millennials (born 1981–1996)

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

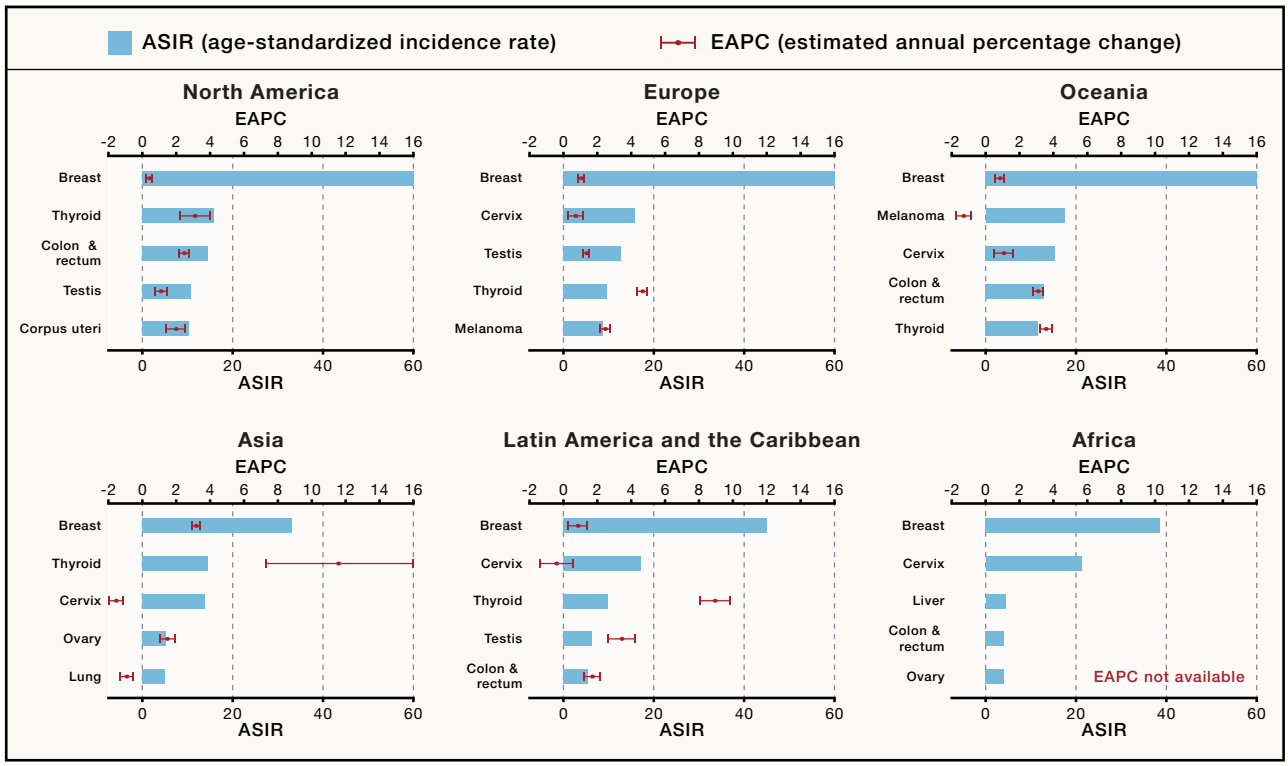

Figure 1. Global incidence (2022) and trends of early-onset cancer (2003–2017)

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 may not fully represent continent-wide trends and is not available for Africa. EAPC summarizes the average annual change in ASIR during 2003–2017 and was estimated by fitting a log-linear regression of ASIR on the calendar year. Data were obtained from the IARC (https://gco.iarc.who.int), including *Cancer Today* (GLOBOCAN 2022, version 1.1) and *Cancer Over Time* (GLOBOCAN, version 2.1).

experiencing higher risk at the same ages than earlier-born cohorts.[8–12]

As we enter "cancer generations," it is timely and critical to review how cancer causes have been discovered and translated into prevention and to pivot on how we should proceed. Although the recent rise in early-onset cancers provides a compelling and urgent context, it reflects broader challenges in how to measure, identify, and intervene on cancer risk factors and susceptibility across the life course. For much of the past century, epidemiologic studies have driven the discovery of cancer causes by identifying robust associations between exposures and disease across populations. As articulated by Sir Richard Peto in 1981,[13] this approach initially "looks at the cancers that people chiefly die of and then looks for populations…which differ in their death rates from these cancers to determine what seem to be the chief manipulable determinants of today's cancers." Over time, prospective cohorts,[14–16] molecular epidemiology with biologically based measures,[17,18] and modern causal-inference approaches[19–23] have strengthened the field's ability to move from associations toward rigorous causal inference. Together, these extensive epidemiologic efforts have led to the discovery of many modifiable cancer causes, such as smoking, alcohol consumption, and obesity, and have informed cancer prevention efforts for past decades, exemplified by tobacco control and declining smoking prevalence that have resulted in significant declines in tobacco-related cancer mortality, averting over 3.8 million lung cancer deaths in the US alone.[24–26] Yet only an estimated 30%–45% of cancers are currently attributable to established, modifiable causes,[27,28] even though 75%–80% may be preventable in theory.[13] In addition to causes yet to be discovered, this gap implies that many causes are likely incompletely characterized, including exposures that are difficult to measure, especially those in early life, or are assessed at inappropriate time points along the etiology pathway.

The central challenge is 2-fold: to accelerate the discovery of cancer causes and to map how they influence biological processes and pathways that shape susceptibility throughout the life course, with the ultimate goal of advancing prevention tailored to individual risk. Addressing these challenges is particularly urgent for younger generations, who are experiencing earlier, more complex, and ubiquitous exposures.[29] These include *in utero* and early-life exposures,[30–32] ultra-processed foods,[33] circadian disruption,[34] infections,[35–37] and environmental chemicals that can act synergistically,[38] many of which are newly introduced[39] or untested for safety.[40,41] This complexity reflects the broader challenges of the exposome, defined as the totality of exposures from conception to death.[42] Exposures may act alone or in combination through

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

mutagenesis and/or through persistent epigenetic, inflammatory, immune, metabolic, and microbial perturbations[43,44] that can shape tissue vulnerability across life stages, especially during windows of susceptibility.[45] However, the exposome is highly dynamic, making both measurement and causal inference difficult, especially for modest effects. These contemporary challenges call for interdisciplinary approaches that build on the core strengths of epidemiology to advance cancer etiology and prevention.

In this perspective, we review the history of cancer cause discovery and highlight the most pressing contemporary challenges, with particular emphasis on those pertinent to early-onset cancers. To address these challenges, we advocate for closer integration between epidemiologic and mechanistic studies and propose three frameworks to accelerate discovery, risk assessment, and preventability. Together, they frame cancer risk as an emergent property of tissue ecosystems across the full continuum, from normal development and homeostasis to invasive tumors. This framing highlights three priorities: characterization of exposures throughout the life course, particularly at windows of susceptibility; a deep understanding of the biological perturbations they induce; and dynamic estimates of risk and preventability. This perspective also aims to stimulate more conceptual, resource, and methodological development to advance cancer etiology research and prevention.

DISCOVERING THE CAUSES OF CANCER: HISTORICAL PERSPECTIVES

In 1981, Doll and Peto's *The Causes of Cancer*[13] articulated two strategies for cancer cause discovery: a "black box" epidemiology approach and a "mechanistic" framework based on experimental testing of candidate agents. Over the past four decades, advances in large prospective cohorts, molecular epidemiology, causal inference, genomics, and mechanistic biology have transformed both discovery and explanation of cancer causes. Conceptually, carcinogenesis is now often framed as a multi-stage clonal selection process, in which initiation generates dormant cells and subsequent promotion drives clonal expansion.[46–48] Exogenous factors such as environmental exposures and modifiable behaviors are typically identified through epidemiologic studies and then dissected mechanistically to determine whether they act as initiators, promoters, or both. Endogenous processes, both cell-extrinsic (e.g., immunity, hormones, and microbiota) and cell-intrinsic (e.g., DNA replication and repair), modulate tumor evolution.[13,49] This convergence of epidemiologic and mechanistic evidence underpins the International Agency for Research on Cancer (IARC) Group 1 carcinogen classifications ("carcinogenic to humans").[50] We also learned that the majority of identified cancer causes act mainly through promotion,[48,51] strengthening the biological rationale for prevention. Figure 2 highlights milestones in uncovering three major exogenous avoidable causes of cancer[52]: tobacco, alcohol, and obesity, alongside advances in cancer genetics that illuminate endogenous pathways.

As Group 1 carcinogens, tobacco and alcohol exemplify how epidemiology and mechanistic research converge, though on different timelines. For tobacco, 18th-century observations linked snuff and pipe use to nasal polyps and lip cancer.[53–55] By the mid-20th century, case-control studies showed 10- to 13-fold higher lung cancer risk in heavy smokers.[56,57] Concerns about retrospective bias prompted large prospective cohorts, including the British Doctors Study of 40,000 physicians, which demonstrated dose-response relationships between smoking and lung cancer mortality.[58] Parallel animal experiments demonstrated that cigarette smoke condensates were carcinogenic.[59,60] These converging lines of evidence were synthesized in the 1964 US Surgeon General's Advisory,[60] concluding that smoking causes lung and laryngeal cancers. Subsequent work identified >60 carcinogens in tobacco smoke,[61] mapped benzo[a]pyrene adducts at *TP53* hotspots and *KRAS* mutations,[62,63] and defined tobacco-associated mutational signatures.[64]

Alcohol as a cause of cancer followed a slower path to consensus with weaker effect sizes. Early 20th-century observations linked alcohol to upper aerodigestive tract cancers,[13,65] while abstinent religious groups showed lower risks.[66–68] In 1987, IARC[69] 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.[70] A subsequent IARC evaluation in 2010[71] expanded this list to include colorectal and female breast cancers with relative risks (RRs) of 1.2–1.6.[70] Mechanistic studies implicated acetaldehyde toxicity, *ALDH2* polymorphisms, oxidative stress, inflammation, hormonal changes, and synergistic interactions with tobacco.[72–76] Reflecting this convergent evidence, the 2025 US Surgeon General's Advisory[77] reinforced alcohol as a cause of seven types of cancer.

Obesity is another major modifiable cancer cause.[28] Since the 1970s, rising obesity prevalence[78] has prompted epidemiologic investigation, including the American Cancer Society's Cancer Prevention Study I, linking higher body weight to cancer mortality.[79] Subsequent studies[80] strengthened the evidence, including an individual participant data meta-analysis of >900,000 adults from 57 prospective cohorts reporting that each 5 kg/m^2 increase was associated with 10% higher cancer mortality.[81] IARC evaluations in 2002[82] and 2016[83] concluded that avoidance of weight gain lowers risk for at least 13 cancer types, with modest RRs (1.1–1.8) for most cancers but substantially higher for esophageal adenocarcinoma (4.8) and corpus uteri (7.1) when comparing the highest body mass index (BMI) category evaluated versus normal BMI. Mechanistic evaluations have implicated inflammation, insulin signaling, steroid hormonal imbalances, metabolic reprogramming,[84] and emerging roles for the microbiome.[85]

In parallel with these exogenous factors, advances in cancer genetics have revealed a complementary arc of inherited susceptibility.[86] Family-based linkage studies in the early 1990s identified high-penetrance genes such as *BRCA1* and *BRCA2* for breast and ovarian cancers[87–89] and *APC*, *MLH1*, and *MSH2* for colorectal cancer.[90–93] More than 70 such high-penetrance genes have been described, but they account for only a minority of familial risk.[94] This gap catalyzed genome-wide association studies (GWASs) in the mid-2000s, which agnostically interrogate large case-control populations to identify common, low-penetrance variants, including over 200 for breast cancer[95]

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

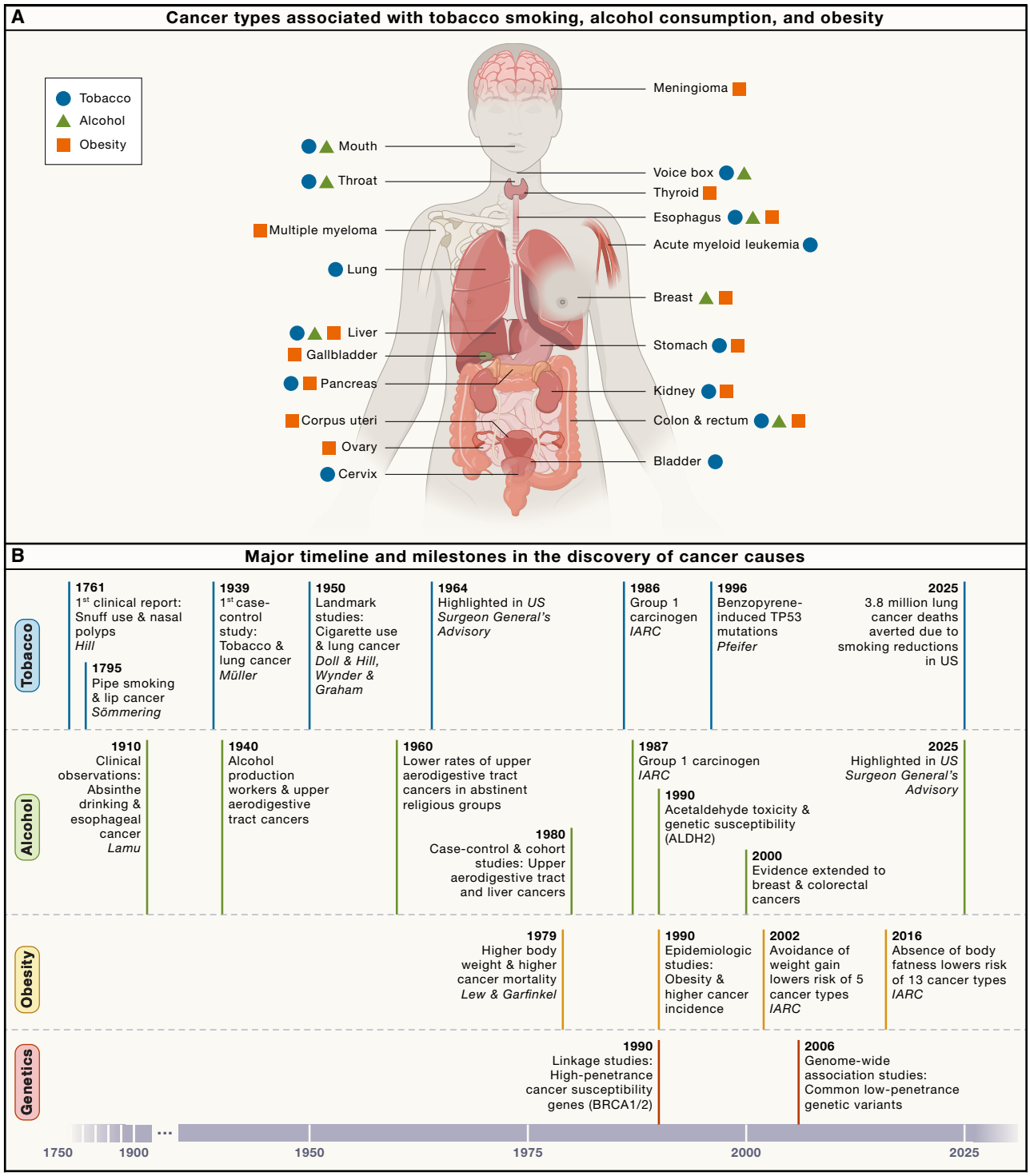

Figure 2. Discovery of major causes of cancer: Tobacco, alcohol, obesity, and genetics

(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 cancer cause, and IARC concludes that avoiding excess body fatness reduces risk for 13 cancer types.

(B) Timeline and milestones in uncovering three major avoidable cancer causes: tobacco smoking, alcohol consumption, and obesity, alongside advances in cancer genetics. The timeline traces the progression encompassing early clinical observations, epidemiologic studies, mechanistic evidence, 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.

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

Box 1. Reimagining the discovery of cancer causes

Is age at diagnosis a reliable proxy for unique early-onset biology? Shift from simple early- versus later-onset comparisons to analyses treating age as a continuous variable, explicitly modeling birth-cohort and period effects, and interpreting molecular patterns alongside exposure patterns.

Are the impacts of cancer causes adequately characterized? Shift from static, single-time point measures toward a comprehensive, longitudinal assessment using innovative, efficient, and potentially objective exposure characterization that captures intensity, timing, trajectories, and clustering during windows of susceptibility.

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 and prioritize studies of gene-environment interactions to identify who is most susceptible to modern exposures and why.

How to leverage mechanistic studies despite translational limits? Leverage mechanistic studies to reinforce causality and embed experimental models as a complementary, hypothesis-testing layer by recapitulating human-relevant exposures and modeling their impact across windows of susceptibility.

How can we leverage epidemiologic resources already in place? Construct a globally connected ecosystem by linking and harmonizing existing cohorts, EHRs, and biobanks with shared standards, enabling rapid evaluation of emerging life-course cancer causes and validation across countries and contexts.

Is the methodology adequate for causal inference with complex data? Develop rigorous methods for causal inference integrating multimodal, correlated, and longitudinal data and improve the estimates of cancer preventability by capturing exposure trends, correlation, feasibility of behavior change, and time course of risk reduction.

and colorectal cancer[96] and at least 45 for lung cancer.[97,98] Approximately one-third of loci show pleiotropy across multiple cancer sites,[94] reinforcing that inherited risk often maps to shared pathways.[99,100] When aggregated into polygenic risk scores, these variants support population-level risk stratification and personalized screening and prevention.[101–103]

Taken together, these extensive efforts, spanning decades to centuries, highlight the importance and success of epidemiologic studies, which successfully identified large, actionable determinants of cancer rates and trends before underlying mechanisms were fully understood.[104] More recently, advances in genomics, molecular epidemiology, and mechanistic biology have begun to clarify how these exposures cause cancer.[105] Looking forward, the strong birth-cohort effects of early-onset cancers underscore an emerging need to identify additional causes that are likely to have small individual effects and arise from widespread, co-occurring exposures that emerge in recent generations and accumulate from early life when exposure measurement in existing epidemiologic cohorts is most limited. At the same time, cancer remains rare in younger adults. The low absolute risk underscores an unmet need for a more precise approach to identify individuals at higher risk for tailored prevention. Together, these contemporary needs require an expanded and interdisciplinary framework to accelerate the discovery of cancer causes for more precision-based prevention.

CONTEMPORARY CHALLENGES IN DISCOVERING CANCER CAUSES

In this section, we outline key contemporary conceptual, resource, and methodological barriers to discovering cancer causes and propose shifts in framing, study design, and analytic approaches to accelerate progress (Box 1).

Conceptual challenges

Age at diagnosis is a limited proxy for underlying biology

Much effort has focused on contrasting early- (<age 50) and later-onset tumors (≥age 50) to identify "unique" early-onset features, but evidence remains limited and inconsistent.[106,107] The lack of striking molecular differences by age at diagnosis does not imply fully shared etiology, as subtle shifts in clonal evolution, immune contexture, or developmental timing can be obscured in bulk analyses. Most importantly, age at diagnosis reflects when the disease is detected and depends on the age of onset, healthcare access, screening, and diagnostic pathways. This process is continuous and varies by people and over time. Statistically, contrasts by age at diagnosis can conflate age-related biology with birth-cohort effects (exposome differences across generations) and period effects (factors that influence all ages concurrently, such as innovations in screening and diagnostic practice).[108] Recent modeling indicates that rising incidence at younger ages may reflect either cohort-specific increases in etiologic exposures beginning around 1970 or period-related shifts toward earlier diagnosis due to improved detection, with their relative contributions varying by cancer type.[109] Analyses will therefore benefit from treating age as a continuous variable, explicitly modeling birth-cohort and period effects, and interpreting molecular patterns alongside exposure patterns.

We have focused on snapshots of cancer causes, not patterns and trajectories

Our initial understanding of established causes is largely based on simplified exposure measures, such as questionnaire recall of aggregate intake (e.g., "weekly alcohol intake") or single-time-point assessments (e.g., "current BMI"). As our understanding of cancer etiology evolved, we realized these snapshots missed intensity, timing, trajectories, and cumulative exposures across the life course, leading to an underestimation of both risk and preventable burden. For instance, alcohol consumption between menarche and first pregnancy was associated with increased breast cancer risk independent of later-life drinking,[110] and heavy episodic or binge drinking increased breast cancer risk even among moderate lifetime drinkers.[111] At the same time, emerging behavioral patterns for these relatively well-established factors, such as drinking on an empty stomach[112] or vaping,[113] are common from early adulthood among younger adults today but are rarely measured or evaluated.[114] Early-life exposures, including *in utero*, childhood, and adolescence, may also contribute to adult cancer risk,[115] including factors such as growth/stature, age at menarche,[116] infections[117] in combination with immunogenetic factors,[116] radiation,[118] diet, obesity (with the exception of premenopausal breast cancer), and putative exposures such as environmental tobacco smoke and physical inactivity.[119] Evidence linking

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

early-life exposures to early-onset cancers is emerging[30–32] but remains largely underexplored. This query is challenging because our evidence base is fragmented: most cohorts begin in midlife, health records rarely capture childhood exposures, exposures such as infections are seldom captured longitudinally, and emerging factors such as circadian rhythm[34] are rarely measured. As a result, the first several decades of life are largely invisible. Looking ahead, innovative, efficient, and potentially objective exposure characterization that captures intensity, timing, trajectories, and clustering[120,121] is needed for a more comprehensive assessment of how established and emerging causes shape cancer risk across the life course (Figure 3).

The exposome is a framework, not an answer

The exposome, proposed by Christopher Wild in 2005, encompasses all non-genetic environmental exposures that influence health and disease.[42] Humans encounter a vast and evolving mixture of chemicals through diet, medications, consumer products, pollution, and a wide range of microbial exposures from symbiotic to pathogenic. Determining which exposures, alone or in combination, contribute to cancer is conceptually straightforward but experimentally challenging. First, the universe of potential chemical exposures already numbers in the hundreds of thousands, many incompletely characterized or undisclosed,[39] with global production of synthetic industrial chemicals projected to exceed one million by 2050.[41] Second, the exposome also includes non-chemical components that are often measured crudely or not at all, such as infections and air pollution. Third, biological responses to exposures are shaped by tissue-specific context and modulated by host factors.[122] Given this complexity, it is not practical to screen all possible combinations of exposures in model systems. The path forward will need to leverage emerging measurement technologies capable of profiling large numbers of chemicals in large, prospective human cohorts.[123,124] When integrated with geospatial data and causal inference, such an approach has the potential to help identify and prioritize biologically relevant and emerging exposures, accelerating discovery and prevention of environmental drivers of early-onset cancer.

Genetics is unlikely to explain the rapid increase, but it shapes susceptibility

Because germline allele frequencies change slowly and high-penetrance cancer 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 cancer incidence. However, susceptibility can also arise within a generation through *de novo* mutations. In recent decades, delayed parenthood has been shown to increase *de novo* mutation burden[125] and may contribute modestly. More broadly, inherited predisposition may act primarily as a modifier of susceptibility, shaping vulnerability to modern lifestyle exposures.[126,127] Finally, pathogens have been a major selective pressure on human immune and inflammatory genes, and variants favored for host defense can involve trade-offs that may increase inflammatory disease risk in modern environments with altered microbial exposures.[128,129] Elucidating emerging gene-environment interactions will be essential for identifying prevention strategies targeted to those at highest risk. Additionally, polygenic risk scores may enable stratification of susceptibility from early life to guide targeted prevention, but this will require better representation of early-onset cancers and addressing limited transferability across populations.[130,131]

Mechanistic studies are helpful for reinforcing causality, but translation limits remain

Mechanistic evidence, including studies in exposed humans, in human cells or tissues, and in other experimental systems, can strengthen hazard evaluation by showing how exposure affects cells and tissues in ways linked to carcinogenesis.[50] Doll and Peto highlighted in *The Causes of Cancer* that mechanistic strategy probes the biology of cancer by exhaustively testing chemical, infectious, or physical agents to identify which are likely to be causes of cancer.[13] Recent advances now allow more direct testing of how human cells and tissues respond to exposures and can link exposures to specific biological pathways with higher resolution than was previously possible.[132,133] However, translation limits from labs to humans remain substantial. Fundamental differences in physiology, lifespan, immune systems,[134] and tissue microenvironments create substantial gaps between murine systems and human carcinogenesis.[135] Even in human-derived cells and tissues, exposures are often tested in isolation, without considering relevant patterns and timing within windows of susceptibility and cumulative and co-occurring exposures. These gaps widen for early-onset cancers, which require a better understanding of life-course insults and even intergenerational exposures.[136] Moving forward, while the discovery of cancer causes will continue to be guided primarily by consistency from epidemiologic studies, experimental models could be embedded as a complementary, hypothesis-testing layer and can likely achieve maximal impact by recapitulating human-relevant exposures and modeling their impact across windows of susceptibility.[137]

Resource-related challenges

Global linkage of epidemiologic resources is the near-term path but remains underbuilt

Most modifiable cancer causes have been identified from epidemiologic cohorts initiated in mid-adulthood. To comprehensively elucidate life-course determinants of early-onset cancers, it will be necessary to establish new birth and multi-generational cohorts with coordinated, longitudinal exposome measurement, leveraging state-of-the-art technologies across windows of susceptibility. However, because cancer incidence is low at young ages, the large sample sizes and extended follow-up required make this approach unlikely to yield timely answers in decades, especially for early life or intergenerational exposures. For instance, with an early-onset colorectal cancer incidence of 14.8 per 100,000 person-years in the US,[2] approximately 200 cases would accrue only after ~14 years of follow-up in a cohort of 100,000 individuals. Accordingly, there is an urgent need to act now by leveraging existing epidemiologic resources rather than waiting for new cohorts to mature. Therefore, the near-term priority is to construct a globally connected ecosystem by linking and harmonizing existing case-control, cohorts, electronic health records (EHRs), and biobanks with shared standards, enabling rapid evaluations of emerging life-course cancer causes and validation across countries and contexts (e.g., sex, race/ethnicity, and birth cohorts). Landmark resources such as

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

CellPress
OPEN ACCESS

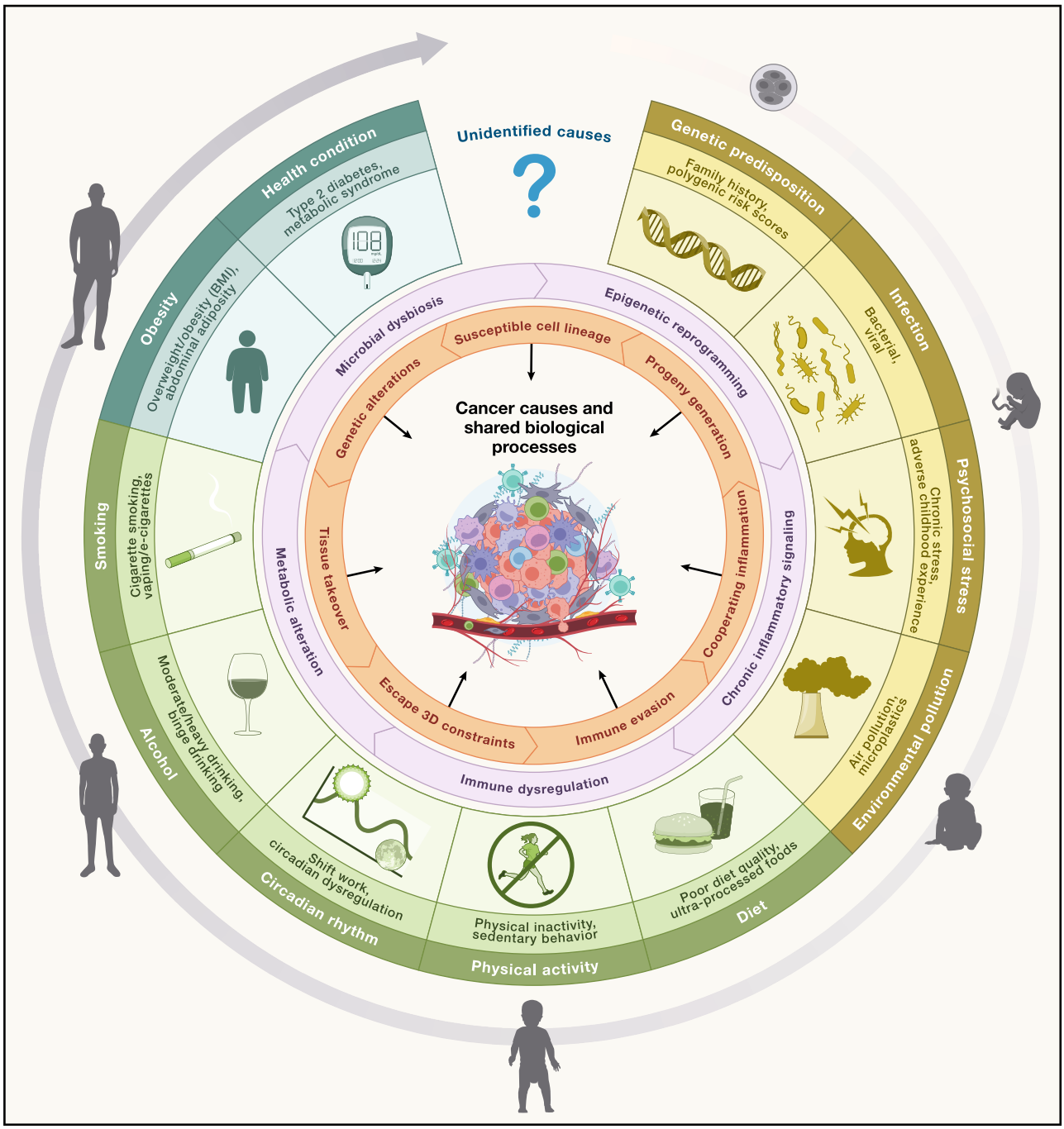

Figure 3. Established and emerging causes of major cancers and shared biological processes

Cancer causes (outer ring) act on the fitness landscape of the tissue through shared biological processes (inner rings) that either initiate tumor progression or promote the expansion of initiated clones. Each exposure category comprises multiple dimensions (e.g., intensity, timing, and pattern of exposure), of which selected examples are illustrated. Additionally, unidentified exposures (blue question mark) may be relevant to younger generations. Exposures accumulate, cluster, and vary across windows of susceptibility from early life through adulthood (circular arrow and life-stage icons), shaping both exposure distributions and their biological impact. These nested layers emphasize that cancers arise not from isolated causes but from interactions across molecular mechanisms, systemic physiology, and individual behaviors.

the Avon Longitudinal Study of Parents and Children,[138] Danish National Birth Cohort,[139] and Generation R[140] already provide prospective early-life data and biospecimens. However, due to privacy and governance considerations, this will require a hybrid model that accommodates federated analyses where data cannot be transferred[141,142] and centrally curated datasets for

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

harmonized data and multi-omics generation. This ecosystem can function as a virtual testbed to identify underexplored domains and to guide targeted recruitment where it offers clear additional value for future cohorts. The Pathways, Risk factors, and mOlecueS to Prevent Early-onset Colorectal Tumors (PROSPECT)[143] consortium of Cancer Grand Challenges is piloting this approach across more than 15 cohorts and EHRs linked to biobanks. Cross-national comparisons, particularly those involving regions with low or declining incidence of early-onset cancers, can reveal protective factors. When combined with molecular profiling of tumor and normal tissues, these comparisons can determine whether cancers with similar histology share or differ in etiological pathways. For instance, Discovering the Causes of Three Poorly Understood Cancers in Europe (DISCERN)[144] illustrates this model by integrating tumor and normal tissue collection in European countries with high and low rates of renal, pancreatic, and colorectal cancer, combining whole-genome sequencing, multi-omics profiling, and prospective cohorts. Achieving these objectives will require robust infrastructure and sustained investment through long-term support for globally coordinated, interdisciplinary initiatives. Additionally, other study designs can be actively leveraged alongside prospective cohorts to address distinct and complementary research questions.[145] Ecological and cross-sectional studies can help generate hypotheses and assess birth-cohort versus period effects. Case-control studies are statistically efficient, particularly the nested case-control design, which offers an immediate path to study past exposures using prediagnostic biospecimens and longitudinal exposure data. Risk-enriched (e.g., family-based) cohorts as well as established genetic consortia[107,146,147] can be powerful for investigating gene-environment interactions.

Methodological challenges

Causal inference methods for multimodal, longitudinal data are just beginning to emerge

Causal inference in epidemiology aims to identify and quantify how modifying an exposure would change cancer risk in a population based on counterfactual reasoning.[148] A central barrier is the misalignment between life-course causality and prevailing analytic paradigms that rely on isolated data modalities or cross-sectional snapshots, which prioritize association over causation. These paradigms struggle with time-dependent confounding, mediation, and reverse causality and are poorly suited to disentangling how life-course exposures propagate through molecular and physiological intermediates to influence cancer risk decades later. Life-course[149] Mendelian randomization,[150] *g*-methods,[151,152] and target trial emulation[22] provide rigorous foundations for evaluating time-varying exposures and separating confounding from mediation, yet they require adaptation to handle multimodal, correlated, and longitudinal data. Causal frameworks for high-dimensional biological data have just begun to emerge.[153] Artificial intelligence (AI) approaches can accelerate data harmonization and hypothesis generation across heterogeneous sources, but without explicit causal design, they risk amplifying spurious correlations. Their value will be realized when embedded within transparent, reproducible causal frameworks that deliver interpretable, prevention-relevant effect estimates.

Evidence synthesis for actionable prevention priorities needs improvement

Synthesizing what we know about cancer causes to estimate the preventability of cancer is critical for setting evidence-based prevention priorities. A cornerstone of this synthesis is comparing and triangulating results across studies, where consistent findings provide better support for identifying causal risk factors.[23] Population attributable fraction (PAF),[154] the estimated fraction of all cases that would not have occurred if there had been no exposure, is the commonly used measurement of the effect of cancer causes. Based on PAF, 30%–45% of cancers are attributable to established modifiable causes.[27,28] Yet, classic PAFs[155] rely on current exposure distributions, assume largely independent effects across exposures, and rarely capture exposure patterns—thus, interactions across exposures can exceed 100% if added together, and the feasibility of sustained behavior can change.[156] In addition, crude exposure proxies such as BMI are often used, which can obscure etiologically relevant adiposity and underestimate the true attributable burden.[157] Moreover, PAF frameworks are minimally integrated with modern cancer biology and evolution. Most PAF estimates do not account for the time course of risk reduction after exposure changes, which depends on exposure history, life stage, latency, and the partial reversibility of tissue-level promotion. Therefore, evidence synthesis should also evolve to integrate population-level exposure trajectories with natural history models, enabling dynamic estimates of preventability and time-varying burden attribution to facilitate the prioritization of prevention strategies. Complementary use of absolute risk estimates can improve interpretability by conveying the magnitude of risk and potential benefit of prevention strategies in clinically meaningful terms for patients and the public.[158]

ACCELERATING DISCOVERY OF CANCER CAUSES: FOUNDATION AND NEW FRAMEWORKS

Recent reflections on Peto's views[48,105,159] reaffirm that epidemiology remains foundational to the discovery of cancer causes.[29] Yet progress has slowed, in part because many remaining causes are likely ubiquitous, poorly measured, or act early in life, yielding limited between-person exposure contrasts to empower classical designs, which are further constrained by underinvestment in studies to measure exposure patterns during windows of susceptibility.[29,160,161] Insights from evolutionary biology[162] and normal-tissue studies show that probable driver mutations and clonal expansions are common in healthy tissues,[51,163–165] implying mutation acquisition alone is necessary but not sufficient to cause cancer. Instead, genetic background and tissue ecology, sculpted by the host immune system, and age- and exposure-dependent selection pressures may initiate clone expansion and, ultimately, progression.[49] Therefore, quantifying clonal expansion of driver mutations across tissue ecologies, ages, and exposure histories is central to understanding somatic evolution and identifying actionable promoters.

These recent advances highlight the need for closer integration between epidemiologic and mechanistic studies to accelerate cancer etiology research. Epidemiologic studies can prioritize candidate exposures and co-occurring patterns and identify

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

CellPress
OPEN ACCESS

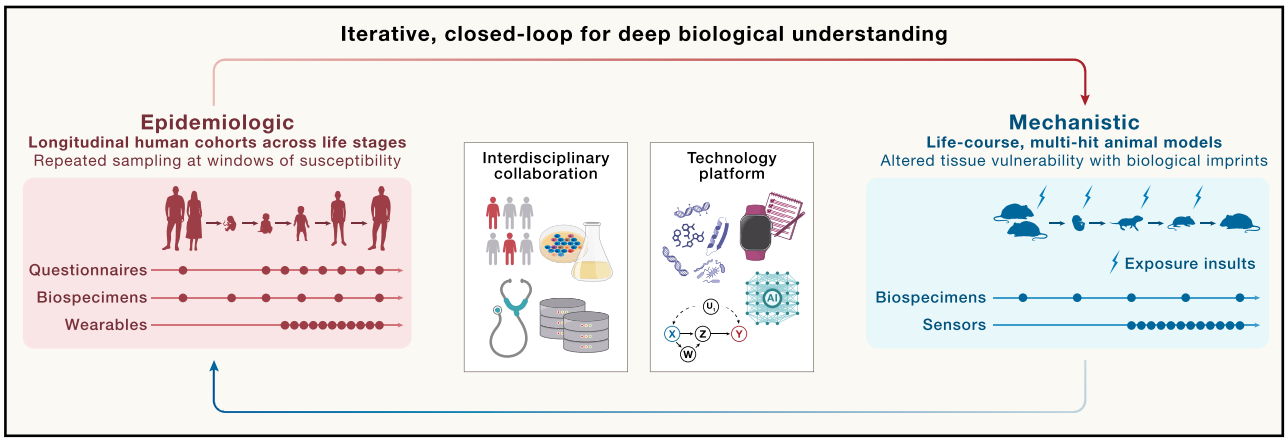

Figure 4. Integrating epidemiologic and mechanistic studies for cancer cause discovery

To accelerate cancer cause discovery research, closer integration between epidemiologic and mechanistic studies is needed. This iterative, closed loop is enabled by interdisciplinary collaboration (epidemiologists, basic scientists, clinicians, data scientists, etc.) and advances in exposure measurements, multi-omic profiling, AI, and causal inference.

susceptible populations, while mechanistic studies test how these exposures induce durable biological changes and reinforce causality by elucidating biological pathways. This process can refine epidemiologic measures, especially molecular markers, and intervention targets (Figure 4). The power of this iterative, closed loop is already evident. For example, a large epidemiologic analysis combined with mouse models and profiling of normal human lung tissue demonstrated that particulate matter measuring $\leq$2.5 μm (PM2.5) promotes EGFR-driven lung adenocarcinoma through an interleukin (IL)-1β-mediated inflammatory response that expands pre-existing mutant clones, thereby motivating molecularly targeted prevention strategies such as IL-1β inhibition.[51] Consortia, such as Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription (TaRGET) II,[166] extend this approach by building longitudinal multi-omic atlases for defined exposures (PM2.5, arsenic, lead, etc.) across target organs and surrogate tissues in mice, enabling direct comparisons between human biospecimens and experimental perturbation readouts.[167] Early cross-tissue analyses suggest that while most toxicant-induced molecular changes were tissue-specific, a subset of co-regulated genes and regulatory elements in liver and blood in response to early-life exposure to toxicants was identified.[168]

To further accelerate cancer cause discovery and interception, we propose three complementary frameworks (Figure 5) that extend the strengths of epidemiology into a more biologically grounded and integrative science. 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 estimates of preventability anchored in natural history.

Tissue ecosystem-anchored framework for cancer cause discovery

A tissue ecosystem-anchored framework for cancer cause discovery (Figure 5A) reframes cancer risk as an emergent property of dynamic tissue ecosystems across the full continuum from normal tissue through precancerous lesions to invasive cancer, shaped by life-course exposures and genetic predisposition. This framework moves beyond exposure-centric to focus on how cumulative exposures across critical life stages create persistent biological signatures that influence tissue vulnerability, somatic evolution, and tumorigenesis. Critically, many exposures, even if transient or poorly captured by conventional measurement, such as infection, leave durable "tissue memories" encoded through epigenetic and metabolic reprogramming, shifts in immune and stromal activities, inflammation, and altered microbiome.[49,169–172]

Mutational epidemiology, linking cancer epidemiology and somatic genomics,[173] shows the promise of a tissue-anchored lens. Somatic mutations accumulate throughout life via diverse mutational processes that leave characteristic "mutational signatures,"[174] now cataloged across >23,000 cancers.[175] In colorectal cancer, analysis of 981 tumor genomes from 11 countries revealed strong geographic heterogeneity. High-incidence regions showed enrichment of signatures associated with colibactin (SBS88/ID18), a genotoxin produced by polyketide synthase-harboring (pks+) bacteria. These signatures were approximately 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.[176] This example illustrates 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.

However, many cancer causes are not directly mutagenic but instead influence tissue homeostasis through systemic perturbations.[177] We therefore advocate expanding the concept of "exposure fingerprints," persistent biological signatures in tissues that encode prior exposures and modulate cancer susceptibility, beyond somatic mutations in tumor and normal tissues[178] to include epigenetic, inflammatory, immune, metabolic, and microbial signatures[43,44] as potential candidate causal fingerprints.

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

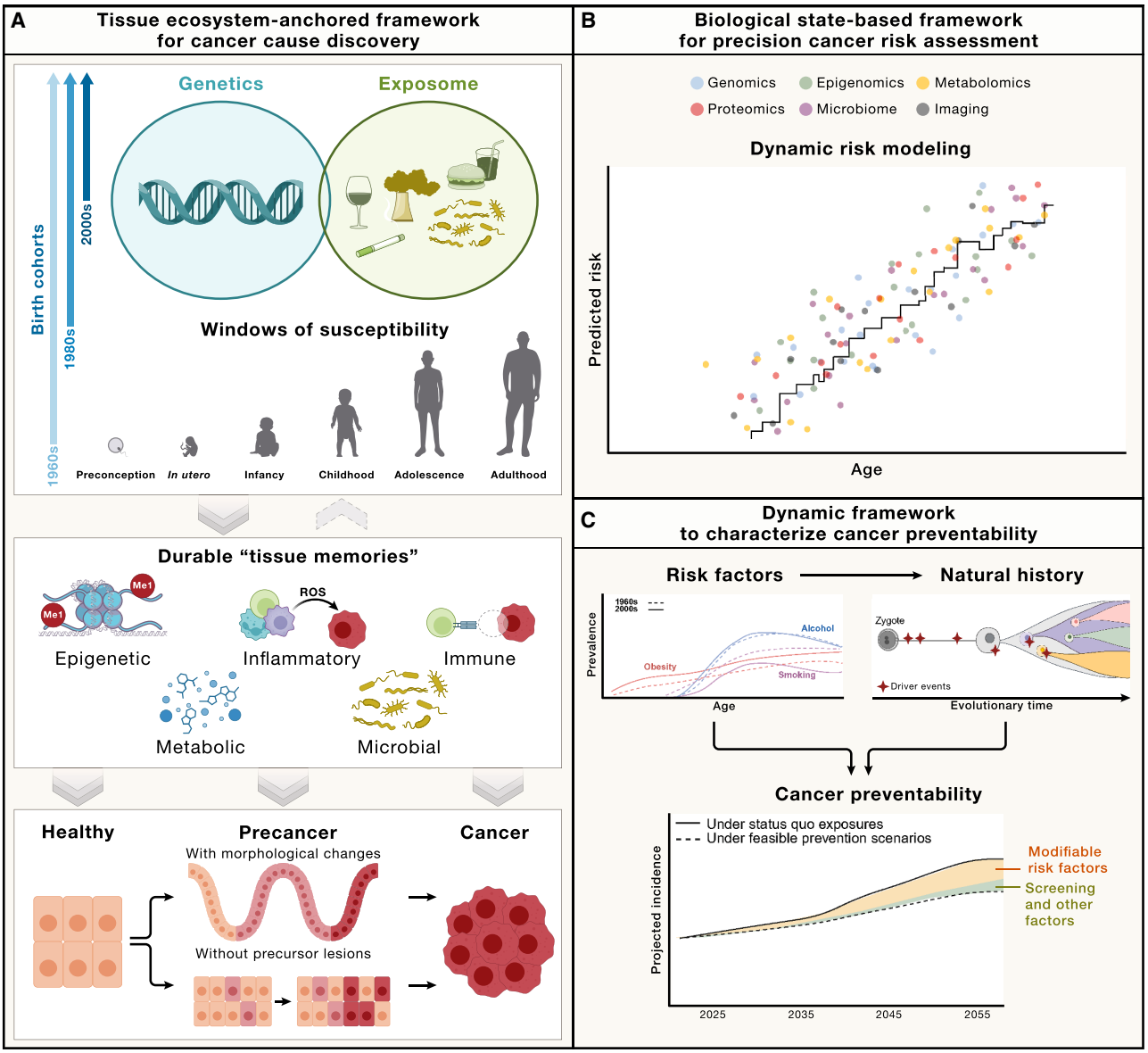

Figure 5. Accelerated frameworks for discovering cancer causes for prevention

(A) Tissue ecosystem-anchored framework for cancer cause discovery: exposures across the life course and intergeneration, together with host genetics, leave molecular imprints that reshape tissue ecosystems and the somatic fitness landscape across the full continuum, from normal development and homeostasis to invasive tumors. Linking these tissue-level signatures back to upstream drivers enables the discovery of cancer causes and the identification of modifiable exposures for prevention.

(B) Biological state-based framework for precision risk assessment: cancer risk evolves as exposures and physiological changes accumulate throughout the life course and can be quantified through longitudinal clinical and multi-omic data to identify high-risk individuals for precision prevention.

(C) Dynamic framework to characterize cancer preventability: simulation models that incorporate secular trends in exposures/interventions, co-occurring exposures, and the natural history of cancer, including knowledge on biological leverage points, can provide dynamic, tissue-informed estimates of preventable burden.

Insights from developmental plasticity and the developmental origins of health and disease paradigm demonstrate how exposures during *in utero*, infancy, adolescence, or early adulthood can introduce durable epigenetic and physiological reprogramming that shapes long-term tissue function and cancer susceptibility.[179–181] Intergenerational epigenetic, immune, and metabolic imprinting further modify baseline vulnerability.[182–185] 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. These interactions establish relatively durable patterns of epithelial turnover, cytokine milieu, metabolite profiles, vascular remodeling, and tissue-resident immunity.[186] When disrupted by infections,

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

antibiotics, obesity, high-fat and low-fiber diets, or pollution, these systems can drive low-grade inflammation, barrier dysfunction, and altered metabolic and microbial states that promote clonal expansion of initiated cancer cells and reshape immune editing, often without additional mutations.[187] Low-grade inflammation can also act systemically by altering hematopoiesis, which can, in turn, reshape immunity within tissues. Moreover, aberrant immune activity itself may be a triggering factor.[188] These microbial-immune-metabolic axes are shaped by early-life exposures and retain durable memories.[189] Mapping how these memories are coordinated across organ systems through shared hormones, metabolites, and inflammatory mediators can reveal common promotive pathways and intervention windows. Chemical fingerprints, such as DNA and protein adducts, serve as upstream sensors of exposure by anchoring suspected agents to their sites of action.[190,191] 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.[192,193]

This framework also incorporates contemporary principles of cancer evolution,[194–196] recognizing tumorigenesis as a dynamic eco-evolutionary process shaped by genetic 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 nutrients and oxygen, stromal and matrix architecture, space, and niche constraints, to shape the fitness, competition, and selection of somatic clones.[197] The fitness effects of individual somatic mutations can be quantified.[198,199] Early-onset cancers may reflect accelerated or altered evolutionary trajectories across the life course.[200] During early life, heightened cell proliferation, developmental plasticity,[201] and immune-microbe interactions[202] create distinct selective environments that influence how mutant clones are generated, constrained, or expanded. Exposures during *in utero*, childhood, adolescence, and early adulthood can durably reconfigure immune, metabolic, and stromal systems, reducing tissue capacity to maintain homeostasis and 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[203–205] is crucial for understanding early-onset cancer and guiding prevention strategies for younger populations.

Translating these tissue-level signatures into prevention requires linking them to upstream drivers and modifiable exposures. One practical path is a closed loop that pairs profiling in longitudinal cohorts and EHR-linked biobanks with controlled experimental systems, including standardized exposure libraries, *ex vivo* models, and organoid perturbation screens that can test dose, timing, and common co-exposures in a human tissue context. Experimental insights can then be mapped onto longitudinal human cohorts and EHRs to strengthen causal inference and prioritize the most plausible, actionable drivers. A key open question is tissue resilience: some exposure-induced states may remain reversible, while others become self-sustaining through dysregulated repair and inflammation, creating a functional "point of no return" that narrows the window for interception.[206] Yet we still have a limited understanding of how such plasticity varies across the life course and across exposure contexts, and advancing this knowledge will be essential for effective prevention.

Biological state-based framework for precision cancer risk assessment

A complementary question centers on prediction: when susceptibility accelerates, who enters a higher-risk trajectory, and how do we quantify an individual's time-varying risk to enable precision prevention? This is especially important for early-onset cancers, where absolute risk is low, and prevention depends on identifying the smaller subset of individuals on a higher-risk path. A biological state-based framework for precision cancer risk assessment (Figure 5B) conceptualizes cancer risk as a continuous process that is updated, ideally in real time, as exposures and physiological changes accumulate throughout the life course. It emphasizes quantifying tissue states that precede clinical detection to improve prediction and to tailor screening and preventive intervention strategies across age groups.

Established cancer prediction tools, such as the Gail model for breast cancer,[207] the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) model for hereditary breast and ovarian cancer,[208,209] and the Colorectal Cancer Risk Assessment Tool,[210] primarily incorporate age, family history, genetic variants, and limited reproductive or lifestyle factors. While foundational in introducing quantitative risk assessment to clinical practice, these models show only modest discriminative capacity, with an area under the curve (AUC) typically between 0.55 and 0.70.[211–214] In essence, these tools approximate tissue-level exposures using epidemiologic exposures. A biological state-based framework moves beyond one-time risk assessment. Dynamic epidemiologic models, such as the Rosner-Colditz breast cancer model,[215–217] illustrate how incorporating changes in reproductive, anthropometric, and lifestyle factors from early life through adulthood can meaningfully improve prediction. However, the detailed longitudinal exposure data these models require can be difficult to collect and maintain in clinical practice.

Recent advances in AI point to a lower-burden path. Longitudinal clinical data collected in routine care, such as diagnoses, medications, laboratory values, vital signs, and imaging when available, are already showing promise for identifying individuals transitioning to higher-risk states. In breast cancer, models that incorporate 3 years of serial screening mammograms achieve a 5-year AUC of 0.80, compared with 0.74 when relying on a single mammogram.[218,219] Interestingly, adding traditional breast cancer risk factors such as age and breast density yields little gain,[218] which supports the promise for breast cancer risk assessment in screening settings. This approach may extend beyond cancer with routine imaging. An AI model trained on longitudinal EHR data from 6 million patients in Denmark predicted pancreatic cancer within 3 years with an AUC of 0.88, demonstrating that even low-incidence cancer can be predicted from patterns in clinical histories using the sequence of disease (International Classification of Diseases [ICD]) codes.[220] More

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

recently, attention-based transformer models were able to predict risk for more than 1,000 diseases simultaneously, leveraging prior diagnoses and additional data, including sex, smoking, alcohol consumption, and BMI.[221] These AI-powered EHR-based and other machine learning risk prediction models[102,222] have achieved strong predictive performance and highlight the potential of scalable, data-driven risk stratification. For diseases with low incidence, such as early-onset cancers, however, strong discrimination can still coincide with modest positive predictive value (PPV), meaning that very few of those classified as high risk will develop the disease over the specific prediction window. Accordingly, the AUC should be interpreted alongside measures of prediction accuracy (e.g., PPV), absolute risk calibration, and decision-relevant metrics (e.g., net benefit)[223] to evaluate whether acting on model predictions improves clinical or public health outcomes.

When risk estimates are intended to guide prevention and interception, integrating biological context may enhance their actionability, transportability, and clinical interpretability. The next-generation risk assessment, both cancer-specific and pan-cancer, could integrate routine clinical care data with genomic and longitudinal multi-omic inputs, including epigenomics, metabolomics, proteomics, and microbiome,[224,225] capturing both tumor-specific and shared pathways. This strategy, likely to achieve better predictive precision, could also enable real-time updating of individual risk trajectories and dynamic stratification for screening and preventive interventions.[226] This is particularly valuable for early-onset cancers, where incidence is low, and single-cancer strategies are often underpowered. As individuals transition into higher-risk biological states, updated risk estimates could inform earlier or more intensive screening and guide preventive strategies, including emerging immune-interception vaccines targeted to shared molecular features rather than germline susceptibility alone.[227] Closing this gap can be strengthened by integrating biological measures that reflect latent disease processes. For instance, functional multistate models that incorporate mammographic imaging features were able to quantify transitions across normal breast tissue, benign lesions, and then to the onset of ductal carcinoma *in situ* (DCIS) or invasive cancer.[228] Similarly, plasma proteomics has revealed systemic and organ-specific aging signatures that precede clinical disease,[229,230] which can improve prediction for several cancers beyond clinical information alone.[231]

Implementation of these biological state-based precision risk assessments may be most appropriate in settings where the potential benefit is high and the burden is low, such as individuals with family history, established genetic susceptibility, or clearly defined high-risk exposures. Implementation should follow data minimization and privacy-by-design, with opt-in consent, transparent governance, safeguards against misuse, and explicit evaluation of psychological harms and equity.

Dynamic framework to characterize cancer preventability

To advance evidence synthesis that complements the traditional PAF approach, we propose a dynamic framework to characterize cancer preventability (Figure 5C) through synthesizing evidence from population, mechanistic, and implementation sciences, with an overarching goal of guiding feasible, high-impact prevention strategies. This framework involves individual-level-based, natural history-informed modeling of cancer preventability that explicitly incorporates exposure shifts across birth cohorts and life stages.[232] Unlike ecological correlations of broad risk factor trends with incidence, microsimulation models use each person as the unit of analysis, allowing multiple cancer causes, their clustering and interactions, and time-varying histories to be carried forward from health to precursor to cancer, and then through screening, treatment, and survivorship.

Validated population simulation models, such as those developed within the Cancer Intervention and Surveillance Modeling Network (CISNET),[233] provide a strong foundation. These models synthesize evidence from clinical, epidemiologic, and registry studies to create virtual populations at risk for cancer representative of the population of interest and have been used to explain observed trends in incidence and mortality by partitioning contributions of cancer causes, screening, and treatment.[234–236] Critically, these models can also simulate past and future counterfactual strategies, such as the impact of interventions on cancer causes or screening on cancer incidence and mortality.[237] For example, CISNET modeling has quantified how tobacco control has reduced mortality over time by incorporating time-varying smoking histories in the US since the 1964 Surgeon General's Report. These analyses estimate lung cancer deaths[235] and tobacco-related deaths[238] averted and dynamically project how birth-cohort-specific smoking trajectories, including initiation, cessation, and intensity, have shaped and will continue to shape lung cancer mortality through 2065.[239] 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 rather than assuming full elimination. In the near term, extending CISNET-style models to incorporate better-characterized life course exposures will allow us to quantify how shifts in exposures in recent birth cohorts contribute to incidence and mortality trends. For example, recent work using nationally representative data spanning more than 60 years provides one of the first birth-cohort (spanning 1910–2005) characterizations of the life-course overweight and obesity burden.[232] Integrating these trajectories into CISNET-style models can estimate how much excess adiposity across the life course has contributed to cancer incidence and mortality across birth cohorts and can project the impact of interventions at different life stages, including the wide adoption of glucagon-like peptide-1 (GLP-1) receptor agonists.

Currently, most models mimic natural history through discrete disease stages. Embedding an explicit evolutionary framework within these models can refine estimates of preventability by linking when mutations arise to which mutations are selectively amplified at each stage of carcinogenesis. For example, CISNET has developed biology-driven multistage clonal expansion models that explicitly incorporate cellular processes of initiation and promotion.[240–242] Additionally, recent work integrates mutational signatures with gene-specific mutation rates and estimates of selection on somatic variants to partition the relative

Please cite this article in press as: Shi et al., Accelerating discovery of cancer causes for prevention in the era of rising early-onset cancers, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.019

contributions of mutational processes across tumor types. In this approach, mutational signatures are first used to estimate the probability that a given mutation arose from a specific endogenous or exogenous process, conditional on local sequence context and tumor-specific mutation rates. These probabilities are then combined with evolutionary models that estimate selection intensity on individual variants by comparing their observed frequencies to expectations under neutrality, enabling inference of which mutational process preferentially generates the variants that confer proliferation and survival advantages to cancer cell lineages. Such analyses demonstrated that the origination of variants in melanoma and lung cancer is predominantly attributable to preventable, exogenous mutational processes (ultraviolet light and tobacco), whereas those in glioma and prostate adenocarcinoma are largely attributable to endogenous processes (aging).[174] By explicitly linking mutation generation to selective expansions, these methods distinguish mutations that occur by chance from those that drive clonal growth and contexts in which reducing an exposure is most likely to prevent the emergence of high-fitness drivers or blunt their subsequent expansion. This context enables calculation of a quantitative bridge needed for dynamic risk prediction: exposure histories and molecular readouts (e.g., signature activity, emerging driver spectra, or indicators of promotion) can be mapped onto time-varying hazards for progression, enabling individualized risk estimates that update as exposures change and population-level risk projections that update as exposure distributions shift. Predictive tools based on underlying biological measures enable both better quantification of risk and interpretability in terms of how to act toward prevention. Together, these tools clarify biological leverage points for prevention.

These leverage points can be translated into dynamic estimates of cancer preventability, because the benefit of risk depends on (1) how far exposures can be modified, (2) when along multistage carcinogenesis the modification occurs, and (3) how durable those changes are.[156] Realistic preventability estimates, therefore, require data on the rate of accumulation of risk and the potential for its reversibility after exposures change, including the kinetics of reduced mutation generation (initiation) and the extent to which promotion can be slowed or reversed. Trials and quasi-experiments that combine intervention with longitudinal biological profiling can estimate these time-dependent parameters and provide the empirical basis for individual-level dynamic prediction, where a person's risk is recalibrated over time as they stop smoking, reduce UV exposure, eradicate a carcinogenic microbe, or as biomarkers of clonal expansion/promotion change. In the longer term, embedding these empirically estimated, time-varying parameters into natural-history-based microsimulation models would move beyond static PAFs toward dynamic stage-aware risk forecasting: at the population level, models can project how shifting exposure distributions, intervention uptake, adherence, and durability alter future incidence; at the individual level, the same framework supports personalized, time-updated estimates of benefit from intervening now versus later, including the identification of windows in which intervention meaningfully changes trajectory versus periods approaching a practical point of no return. Extending from single-cancer models toward integrated, multi-cancer simulation platforms would further enable joint prediction of preventable burden across cancers (and potentially cardiometabolic outcomes) and evaluation of interventions that may reduce risk across several malignancies simultaneously, particularly relevant for early-onset cancers where site-specific incidence is low but cumulative burden across multiple types is substantial.

PREVENTION FOR FUTURE GENERATIONS

Advances within the three proposed frameworks depend on how well we can accurately measure and reconstruct genetic susceptibility and non-genetic exposures across the life course and generations, which will require sustained interdisciplinary collaborations (Figure 4). Cost-effective, scalable questionnaires remain essential for capturing contextual information that passive sensing cannot, such as diet, behaviors, occupations, product use, and early-life histories.[243] Objective measures, including sensors and wearables[244] combined with high-resolution geocoding,[245] can enable more precise and temporally resolved quantification of environmental and behavioral factors, although integration with large population studies remains limited. High-throughput molecular profiling technologies offer complementary approaches for capturing tissue responses and potentially reconstructing exposures (e.g., environmental chemicals through high-resolution mass spectrometry) (Table S1).[246,247] The central constraint is likely the feasibility, especially in countries such as the US, where exposure and outcome data are fragmented across health systems, payers, and research cohorts. Thus, the immediate priority should be to harmonize core measurements across cohorts and to empower EHRs and biobanks to link individual- and community-level exposure data. Over the long term, building low-burden systems that repeatedly capture, harmonize, and link these data streams will require sustained infrastructure support.

In addition to the need for improved exposure assessment and data linkage, the increasing burden of early-onset cancers, as well as the earlier onset of diabetes,[248] cardiovascular disease,[249,250] dementia,[251] and other chronic conditions, highlights the importance of addressing shared systemic disruptions shaped by evolving personal, physical, and social environments. These convergent shifts in chronic disease epidemiology suggest it is time to consider integrative chronic disease prevention that addresses multiple common exposures and promotive pathways to delay the onset of multiple major diseases. Prevention trials that focus on shared intermediate biomarkers, longitudinal tissue-ecosystem profiling, and natural history-based modeling are likely to be promising. However, this will require careful design and significant efforts, drawing from insights from previous decades.[252,253]

CONCLUSIONS

The increasing incidence of early-onset cancers is among the most urgent challenges in modern cancer research. In addition to highlighting the need for closer integration between epidemiologic and mechanistic studies, we propose three frameworks that retain epidemiology's strengths while integrating exposures, biological mechanisms, and prevention strategies. The first is a

tissue ecosystem-anchored framework that links exposures to tissue memories. The second is a biological state-based framework for precision risk assessment. The third is a dynamic framework that translates these insights into realistic, biology-based estimates of preventability. Achieving these objectives requires longitudinal sample collections, technological and methodological advancements in exposure measurement, data integration, and causal inference, as well as interdisciplinary collaboration and sustained infrastructure support. This roadmap outlines an initial strategy to accelerate the discovery of cancer causes, shifting from cataloging midlife associations to systematically elucidating the mechanisms connecting exposures and disease, predicting evolving biological states, and informing upstream interventions to prevent cancers, including early-onset cancers and other chronic diseases with shared etiological processes and pathways.

ACKNOWLEDGMENTS

This work was delivered as part of the PROSPECT team supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (CGCATF-2023/100044 to Y.B. and CGCATF-2023/100037 to G.J.P. and Y.C.), the National Cancer Institute (OT2CA297581 to Y.B. and OT2CA297576 to G.J.P. and Y.C.), the French National Cancer Institute, and the Bowelbabe Fund for Cancer Research UK. I.L.V. is supported by the Dutch Research Council (09150171910047). T.W. is in part supported by NIH/National Human Genome Research Institute R01HG007175, NIH/National Institute of Environmental Health Sciences U24ES026699, NIH/National Institute of Neurological Disorders and Stroke U24NS132103, and NIH/National Institute on Drug Abuse UM1DA058219. G.A.C. is supported by the Breast Cancer Research Foundation (BCRF 25-028). Y.C. is supported by grants from NIH/National Cancer Institute R37CA246175, the Alvin J. Siteman Cancer Center through the Foundation for Barnes-Jewish Hospital, and the NIH/National Institute of Diabetes and Digestive and Kidney Diseases P30DK052574. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.

AUTHOR CONTRIBUTIONS

M.S. and Y.C. drafted the manuscript. All authors contributed substantially to the discussion of the content. All authors reviewed and edited the manuscript before submission.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.03.019.

REFERENCES

1. 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. https://doi.org/10.3322/caac.21834.
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. https://doi.org/10.1016/s1470-2045(24)00600-4.
3. Siegel, R.L., Kratzer, T.B., Wagle, N.S., Sung, H., and Jemal, A. (2026). Cancer statistics, 2026. CA Cancer J. Clin. *76*, e70043. https://doi.org/10.3322/caac.70043.
4. 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. *178*, 1677–1687. https://doi.org/10.7326/annals-24-02718.
5. Yan, W., Liu, M., Jing, W., Kang, L., Zhang, N., Sun, H., He, J., Chen, Z., Liu, J., Liang, W., et al. (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. https://doi.org/10.1186/s40164-025-00634-7.
6. Siegel, R.L., Giaquinto, A.N., and Jemal, A. (2024). Cancer statistics, 2024. CA Cancer J. Clin. *74*, 12–49. https://doi.org/10.3322/caac.21820.
7. Siegel, R.L., Wagle, N.S., and Jemal, A. (2026). Leading Cancer Deaths in People Younger Than 50 Years. JAMA *335*, 632–634. https://doi.org/10.1001/jama.2025.25467.
8. 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. https://doi.org/10.1016/s2468-2667(18)30267-6.
9. 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. https://doi.org/10.1016/s2468-2667(24)00156-7.
10. Rosenberg, P.S., and Miranda-Filho, A. (2024). Cancer Incidence Trends in Successive Social Generations in the US. JAMA Netw. Open *7*, e2415731. https://doi.org/10.1001/jamanetworkopen.2024.15731.
11. 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. https://doi.org/10.1158/2159-8290.Cd-24-1678.
12. Downham, L., Laversanne, M., Perdomo, S., Filho, A.M., Bray, F., and Brennan, P. (2025). Increase of early-onset colorectal cancer: a cohort effect. J. Natl. Cancer Inst. djaf238. https://doi.org/10.1093/jnci/djaf238.
13. Doll, R., and Peto, R. (1981). The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J. Natl. Cancer Inst. *66*, 1191–1308. https://doi.org/10.1093/jnci/66.6.1192.
14. Doll, R., and Hill, A.B. (1954). The mortality of doctors in relation to their smoking habits; a preliminary report. Br. Med. J. *1*, 1451–1455. https://doi.org/10.1136/bmj.1.4877.1451.
15. Hammond, E.C. (1959). Cancer etiology: new prospective epidemiological study. CA Cancer J. Clin. *9*, 177–178. https://doi.org/10.3322/canjclin.9.5.177.
16. Bao, Y., Bertoia, M.L., Lenart, E.B., Stampfer, M.J., Willett, W.C., Speizer, F.E., and Chavarro, J.E. (2016). Origin, Methods, and Evolution of the Three Nurses' Health Studies. Am. J. Public Health *106*, 1573–1581. https://doi.org/10.2105/ajph.2016.303338.
17. Perera, F.P., and Weinstein, I.B. (1982). Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. J. Chronic Dis. *35*, 581–600. https://doi.org/10.1016/0021-9681(82)90078-9.
18. McMichael, A.J. (1994). Invited commentary–"molecular epidemiology": new pathway or new travelling companion? Am. J. Epidemiol. *140*, 1–11. https://doi.org/10.1093/oxfordjournals.aje.a117153.

19. Robins, J.M., Hernán, M.A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology *11*, 550–560. https://doi.org/10.1097/00001648-200009000-00011.

20. Rothman, K.J., and Greenland, S. (2005). Causation and causal inference in epidemiology. Am. J. Public Health *95*, S144–S150. https://doi.org/10.2105/ajph.2004.059204.

21. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. *23*, R89–R98. https://doi.org/10.1093/hmg/ddu328.

22. 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. https://doi.org/10.1001/jama.2022.21383.

23. Gutierrez, S., Glymour, M.M., and Smith, G.D. (2025). Evidence triangulation in health research. Eur. J. Epidemiol. *40*, 743–757. https://doi.org/10.1007/s10654-024-01194-6.

24. Jha, P. (2009). Avoidable global cancer deaths and total deaths from smoking. Nat. Rev. Cancer *9*, 655–664. https://doi.org/10.1038/nrc2703.

25. Jha, P., and Peto, R. (2014). Global effects of smoking, of quitting, and of taxing tobacco. N. Engl. J. Med. *370*, 60–68. https://doi.org/10.1056/NEJMra1308383.

26. 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. https://doi.org/10.3322/caac.70005.

27. 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. https://doi.org/10.1093/jnci/djab204.

28. 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. https://doi.org/10.3322/caac.21858.

29. 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. https://doi.org/10.1007/s10654-023-01090-5.

30. Hur, J., Otegbeye, E., Joh, H.K., Nimptsch, K., Ng, K., Ogino, S., Meyerhardt, J.A., Chan, A.T., Willett, W.C., Wu, K., et al. (2021). Sugar-sweetened beverage intake in adulthood and adolescence and risk of early-onset colorectal cancer among women. Gut *70*, 2330–2336. https://doi.org/10.1136/gutjnl-2020-323450.

31. 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. https://doi.org/10.1001/jamanetworkopen.2023.10316.

32. 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. https://doi.org/10.1001/jamanetworkopen.2024.40181.

33. 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. (2026). Ultraprocessed Food Consumption and Risk of Early-Onset Colorectal Cancer Precursors Among Women. JAMA Oncol. *12*, 49–57. https://doi.org/10.1001/jamaoncol.2025.4777.

34. 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. https://doi.org/10.1038/s43018-025-00981-8.

35. Rugge, M., Busatto, G., Cassaro, M., Shiao, Y.H., Russo, V., Leandro, G., Avellini, C., Fabiano, A., Sidoni, A., and Covacci, A. (1999). Patients younger than 40 years with gastric carcinoma: Helicobacter pylori genotype and associated gastritis phenotype. Cancer *85*, 2506–2511. https://doi.org/10.1002/(SICI)1097-0142(19990615)85:12<2506::AID-CNCR3>3.0.CO;2-I.

36. Liu, J., Tang, W., Budhu, A., Forgues, M., Hernandez, M.O., Candia, J., Kim, Y., Bowman, E.D., Ambs, S., Zhao, Y., et al. (2020). A Viral Exposure Signature Defines Early Onset of Hepatocellular Carcinoma. Cell *182*, 317–328.e10. https://doi.org/10.1016/j.cell.2020.05.038.

37. Dziubańska-Kusibab, P.J., Berger, H., Battistini, F., Bouwman, B.A.M., Iftekhar, A., Katainen, R., Cajuso, T., Crosetto, N., Orozco, M., Aaltonen, L.A., et al. (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat. Med. *26*, 1063–1069. https://doi.org/10.1038/s41591-020-0908-2.

38. 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. https://doi.org/10.1007/s40471-017-0100-5.

39. 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. https://doi.org/10.1021/acs.est.9b06379.

40. Parrish, M., and Kuperwasser, C. (2025). Environmental endocrine disruptors: rethinking the origins of early-onset ER+ breast cancer. Nat. Rev. Cancer *25*, 819–820. https://doi.org/10.1038/s41568-025-00854-3.

41. Orešič, M., Karu, N., Zhao, H.N., Moseley, A., Hankemeier, T., Wishart, D.S., Dorrestein, P.C., Fiehn, O., Hyötyläinen, T., Daouk, R.K., et al. (2025). Metabolome informs about the chemical exposome and links to brain health. Environ. Int. *203*, 109741. https://doi.org/10.1016/j.envint.2025.109741.

42. 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. https://doi.org/10.1158/1055-9965.Epi-05-0456.

43. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46. https://doi.org/10.1158/2159-8290.Cd-21-1059.

44. Hanahan, D. (2026). Hallmarks of cancer-Then and now, and beyond. Cell. https://doi.org/10.1016/j.cell.2025.12.049.

45. Terry, M.B., Michels, K.B., Brody, J.G., Byrne, C., Chen, S., Jerry, D.J., Malecki, K.M.C., Martin, M.B., Miller, R.L., Neuhausen, S.L., et al. (2019). Environmental exposures during windows of susceptibility for breast cancer: a framework for prevention research. Breast Cancer Res. *21*, 96. https://doi.org/10.1186/s13058-019-1168-2.

46. Rozhok, A.I., and DeGregori, J. (2017). Chapter 12. Aging, Somatic Evolution, and Cancer. In On Human Nature, N. On Human, M. Tibayrenc, and F.J. Ayala, eds. (Academic Press), pp. 193–209. https://doi.org/10.1016/B978-0-12-420190-3.00012-0.

47. Jassim, A., Rahrmann, 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. https://doi.org/10.1038/s41568-023-00602-5.

48. 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. https://doi.org/10.1007/s10654-022-00933-x.

49. 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. https://doi.org/10.1016/j.cell.2023.03.013.

50. Samet, J.M., Chiu, W.A., Cogliano, 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. https://doi.org/10.1093/jnci/djz169.

51. 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. https://doi.org/10.1038/s41586-023-05874-3.

52. GBD 2019 Cancer Risk Factors Collaborators (2022). The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. Lancet *400*, 563–591. https://doi.org/10.1016/s0140-6736(22)01438-6.
53. 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).
54. Soemmerring, S.T. (1795). De Morbis Vasorum Absorbentium Corporis Humani Sive Dissertationis Quae Praemium Retulit Societatis Rheno-Traiectinae Anno MDCCLXXXXIV. Pars…; Accedit Index Scriptorum De Systemate Absorbente (Varrentrappius & Wenner).
55. Redmond, D.E., Jr. (1970). Tobacco and cancer: the first clinical report, 1761. N. Engl. J. Med. *282*, 18–23. https://doi.org/10.1056/nejm197001012820105.
56. 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. https://doi.org/10.1001/jama.1950.02910390001001.
57. Doll, R., and Hill, A.B. (1952). A study of the aetiology of carcinoma of the lung. Br. Med. J. *2*, 1271–1286. https://doi.org/10.1136/bmj.2.4797.1271.
58. Doll, R., and Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. Br. Med. J. *2*, 1525–1536. https://doi.org/10.1136/bmj.2.6051.1525.
59. Wynder, E.L., Graham, E.A., and Croninger, A.B. (1953). Experimental production of carcinoma with cigarette tar. Cancer Res. *13*, 855–864.
60. U.S.; Surgeon; General's; Advisory Committee; Smoking; Health (1964). Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service, Washington, D.C (United States Department of Health, Education and Welfare, Public Health Service).
61. Hecht, S.S. (2003). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat. Rev. Cancer *3*, 733–744. https://doi.org/10.1038/nrc1190.
62. 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. https://doi.org/10.1056/nejm198710083171504.
63. 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. https://doi.org/10.1126/science.274.5286.430.
64. 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. https://doi.org/10.1126/science.aag0299.
65. 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.
66. Wynder, E.L., Lemon, F.R., and Bross, I.J. (1959). Cancer and coronary artery disease among Seventh-Day Adventists. Cancer *12*, 1016–1028. https://doi.org/10.1002/1097-0142(195909/10)12:5<1016::aid-cncr2820120523>3.0.co;2-2.
67. 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. https://doi.org/10.1002/1097-0142(196404)17:4<486::aid-cncr2820170410>3.0.co;2-z.
68. 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.
69. IARC (1988). Alcohol drinking. IARC Monogr. Eval. Carcinog. Risks Hum. *44*, 1–378.
70. 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. https://doi.org/10.1038/bjc.2014.579.
71. IARC Working Group; Evaluation; Carcinogenic; Risks; Humans (2010). Alcohol consumption and ethyl carbamate. IARC Monogr. Eval. Carcinog. Risks Hum. *96*, 3–1383.
72. 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. https://doi.org/10.1023/b:caco.0000003854.34221.a8.
73. Pöschl, G., and Seitz, H.K. (2004). ALCOHOL AND CANCER. Alcohol Alcohol. *39*, 155–165. https://doi.org/10.1093/alcalc/agh057.
74. 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. https://doi.org/10.1016/s1470-2045(09)70019-1.
75. 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. https://doi.org/10.1158/1055-9965.Epi-08-0347.
76. Rumgay, H., Murphy, N., Ferrari, P., and Soerjomataram, I. (2021). Alcohol and Cancer: Epidemiology and Biological Mechanisms. Nutrients *13*, 3173. https://doi.org/10.3390/nu13093173.
77. Office; Surgeon General (2025). Publications and Reports of the Surgeon General. In Alcohol and Cancer Risk: the U.S. Surgeon General's Advisory (United States Department of Health and Human Services).
78. 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. https://doi.org/10.1001/jama.272.3.205.
79. Lew, E.A., and Garfinkel, L. (1979). Variations in mortality by weight among 750,000 men and women. J. Chronic Dis. *32*, 563–576. https://doi.org/10.1016/0021-9681(79)90119-x.
80. 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. https://doi.org/10.1056/NEJMoa021423.
81. Prospective; Studies Collaboration, 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. https://doi.org/10.1016/s0140-6736(09)60318-4.
82. International Agency for Research on Cancer (2002). IARC Handbooks of Cancer Prevention: Weight Control and Physical Activity*6* (IARC Press).
83. Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., and Straif, K.; International Agency for Research on Cancer Handbook Working Group (2016). Body Fatness and Cancer–Viewpoint of the IARC Working Group. N. Engl. J. Med. *375*, 794–798. https://doi.org/10.1056/NEJMsr1606602.
84. Hopkins, B.D., Goncalves, M.D., and Cantley, L.C. (2016). Obesity and Cancer Mechanisms: Cancer Metabolism. J. Clin. Oncol. *34*, 4277–4283. https://doi.org/10.1200/jco.2016.67.9712.
85. 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. https://doi.org/10.1038/s41574-025-01159-z.

86. Qing, T., Mohsen, H., Marczyk, M., Ye, Y., O'Meara, T., Zhao, H., Townsend, J.P., Gerstein, M., Hatzis, C., Kluger, Y., et al. (2020). Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. Nat. Commun. *11*, 2438. https://doi.org/10.1038/s41467-020-16293-7.

87. 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. https://doi.org/10.1126/science.2270482.

88. Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., and Averill, D. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science *265*, 2088–2090. https://doi.org/10.1126/science.8091231.

89. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., and Ding, W. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science *266*, 66–71. https://doi.org/10.1126/science.7545954.

90. 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., and McKechnie, D. (1991). Identification of FAP locus genes from chromosome 5q21. Science *253*, 661–665. https://doi.org/10.1126/science.1651562.

91. Peltomäki, P., Aaltonen, L.A., Sistonen, P., Pylkkänen, L., Mecklin, J.P., Järvinen, H., Green, J.S., Jass, J.R., Weber, J.L., and Leach, F.S. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science *260*, 810–812. https://doi.org/10.1126/science.8484120.

92. 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. https://doi.org/10.1038/ng1193-279.

93. 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. https://doi.org/10.1016/0092-8674(93)90546-3.

94. 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. https://doi.org/10.1038/nrc.2017.82.

95. 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. https://doi.org/10.1038/s41588-020-0609-2.

96. 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. https://doi.org/10.1038/s41588-022-01222-9.

97. Bossé, Y., and Amos, C.I. (2018). A Decade of GWAS Results in Lung Cancer. Cancer Epidemiol. Biomarkers Prev. *27*, 363–379. https://doi.org/10.1158/1055-9965.Epi-16-0794.

98. Byun, J., Han, Y., Li, Y., Xia, J., Long, E., Choi, J., Xiao, X., Zhu, M., Zhou, W., Sun, R., et al. (2022). Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer. Nat. Genet. *54*, 1167–1177. https://doi.org/10.1038/s41588-022-01115-x.

99. 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. https://doi.org/10.1093/hmg/ddv440.

100. Kuenzi, B.M., and Ideker, T. (2020). A census of pathway maps in cancer systems biology. Nat. Rev. Cancer *20*, 233–246. https://doi.org/10.1038/s41568-020-0240-7.

101. 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. https://doi.org/10.1016/j.ajhg.2018.11.002.

102. 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. https://doi.org/10.1016/j.ajhg.2020.07.006.

103. 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. https://doi.org/10.1016/s1470-2045(23)00156-0.

104. Peto, R. (1984). The Need for Ignorance in Cancer Research. In The Encyclopaedia of Medical Ignorance: Exploring the Frontiers of Medical Knowledge, M.W.-S.R. Duncan, ed. (Pergamon Press), pp. 129–133.

105. Campbell, P.J. (2023). Demystifying the black box: from ignorance to observation to mechanism in cancer research. Eur. J. Epidemiol. *38*, 1265–1267. https://doi.org/10.1007/s10654-022-00935-9.

106. 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. https://doi.org/10.1158/1055-9965.Epi-13-0340.

107. Laskar, R.S., Qu, C., Huyghe, J.R., Harrison, T., Hayes, R.B., Cao, Y., Campbell, P.T., Steinfelder, 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. https://doi.org/10.1016/j.annonc.2024.02.008.

108. Holford, T.R. (1983). The estimation of age, period and cohort effects for vital rates. Biometrics *39*, 311–324. https://doi.org/10.2307/2531004.

109. Owens, L., Fung, A., Shuhendler, J., Glick, J., Ryser, M.D., Gulati, R., and Etzioni, R. (2025). Trends in young-onset cancer incidence: a modeling perspective. J. Natl. Cancer Inst. *117*, 1350–1359. https://doi.org/10.1093/jnci/djaf050.

110. 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. https://doi.org/10.1093/jnci/djt213.

111. 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. https://doi.org/10.1093/aje/kwx118.

112. Lupi, M., Martinotti, G., and Di Giannantonio, M. (2017). Drunkorexia: an emerging trend in young adults. Eat. Weight Disord. *22*, 619–622. https://doi.org/10.1007/s40519-017-0429-2.

113. Masonbrink, A.R., Bae, D., Cho, J., Miech, R.A., Dai, H.D., Harlow, A.F., Sussman, S., Han, D.-H., Sanchez, L.M., Adjei, A., et al. (2025). Trends in Daily Nicotine Vaping and Unsuccessful Quit Attempts in Youths. JAMA Netw. Open *8*, e2541061. https://doi.org/10.1001/jamanetworkopen.2025.41061.

114. Shi, M., Ahmedin, N.J., Zong, X., Yilmaz, Ö.H., Bishehsari, F., and Cao, Y. (2024). Alcohol Drinking Without Meals Is Associated With Risk of Gastrointestinal Cancers, Including Early-Onset Cases. Gastroenterology *166*, 341–344.e5. https://doi.org/10.1053/j.gastro.2023.11.012.

115. Wild, C.P. (2011). How much of a contribution do exposures experienced between conception and adolescence make to the burden of cancer in

adults? Cancer Epidemiol. Biomarkers Prev. *20*, 580–581. https://doi.org/10.1158/1055-9965.Epi-11-0187.

116. 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. https://doi.org/10.1007/s10552-012-9962-5.

117. Vedham, V., Verma, M., and Mahabir, S. (2015). Early-life exposures to infectious agents and later cancer development. Cancer Med. *4*, 1908–1922. https://doi.org/10.1002/cam4.538.

118. Balk, S.J.; Council; Environmental Health; Section on Dermatology (2011). Ultraviolet radiation: a hazard to children and adolescents. Pediatrics *127*, e791–e817. https://doi.org/10.1542/peds.2010-3502.

119. Clarke, M.A., and Joshu, C.E. (2017). Early Life Exposures and Adult Cancer Risk. Epidemiol. Rev. *39*, 11–27. https://doi.org/10.1093/epirev/mxx004.

120. Mentzer, A.J., Brenner, N., Allen, N., Littlejohns, T.J., Chong, A.Y., Cortes, A., Almond, R., Hill, M., Sheard, S., McVean, G., et al. (2022). Identification of host-pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. Nat. Commun. *13*, 1818. https://doi.org/10.1038/s41467-022-29307-3.

121. Shim, J., Fleisch, E., and Barata, F. (2024). Circadian rhythm analysis using wearable-based accelerometry as a digital biomarker of aging and healthspan. NPJ Digit. Med. *7*, 146. https://doi.org/10.1038/s41746-024-01111-x.

122. Wild, C.P. (2012). The exposome: from concept to utility. Int. J. Epidemiol. *41*, 24–32. https://doi.org/10.1093/ije/dyr236.

123. Jones, D.P. (2016). Sequencing the exposome: A call to action. Toxicol. Rep. *3*, 29–45. https://doi.org/10.1016/j.toxrep.2015.11.009.

124. 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. https://doi.org/10.1021/acs.analchem.2c01270.

125. 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. https://doi.org/10.1016/j.tig.2019.08.005.

126. 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. https://doi.org/10.1001/jamanetworkopen.2022.34221.

127. 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. https://doi.org/10.1093/jnci/djae208.

128. Barreiro, L.B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. *11*, 17–30. https://doi.org/10.1038/nrg2698.

129. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell *167*, 657–669.e21. https://doi.org/10.1016/j.cell.2016.09.025.

130. Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D.J., Martin, I., Kullo, I.J., Kenny, E.E., Pasaniuc, B., Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium Methods Working Group, and Witte, J.S., et al. (2024). Principles and methods for transferring polygenic risk scores across global populations. Nat. Rev. Genet. *25*, 8–25. https://doi.org/10.1038/s41576-023-00637-2.

131. Thomas, C.E., and Peters, U. (2025). Genomic landscape of cancer in racially and ethnically diverse populations. Nat. Rev. Genet. *26*, 336–349. https://doi.org/10.1038/s41576-024-00796-w.

132. Drost, J., and Clevers, H. (2018). Organoids in cancer research. Nat. Rev. Cancer *18*, 407–418. https://doi.org/10.1038/s41568-018-0007-6.

133. Kim, J., Koo, B.K., and Knoblich, J.A. (2020). Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. *21*, 571–584. https://doi.org/10.1038/s41580-020-0259-3.

134. Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A., et al. (2013). Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proc. Natl. Acad. Sci. USA *110*, 2946–2951. https://doi.org/10.1073/pnas.1222738110.

135. de Jong, M., and Maina, T. (2010). Of mice and humans: are they the same?–Implications in cancer translational research. J. Nucl. Med. *51*, 501–504. https://doi.org/10.2967/jnumed.109.065706.

136. Zhou, R., Tang, X., and Wang, Y. (2024). Emerging strategies to investigate the biology of early cancer. Nat. Rev. Cancer *24*, 850–866. https://doi.org/10.1038/s41568-024-00754-y.

137. Pandya, T., Cao, Y., Smith-Byrne, K., and Swanton, C. (2025). Mechanism-enhanced population science: strengthening population studies through functional insights. Ann. Oncol. *36*, 1552–1553. https://doi.org/10.1016/j.annonc.2025.08.006.

138. 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. https://doi.org/10.1093/ije/dys066.

139. 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. https://doi.org/10.1177/14034948010290040201.

140. Hofman, A., Jaddoe, V.W.V., Mackenbach, J.P., Moll, H.A., Snijders, R.F.M., Steegers, E.A.P., Verhulst, F.C., Witteman, J.C.M., 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. https://doi.org/10.1111/j.1365-3016.2003.00521.x.

141. Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J., and Wang, F. (2021). Federated Learning for Healthcare Informatics. J. Healthc. Inform. Res. *5*, 1–19. https://doi.org/10.1007/s41666-020-00082-4.

142. Luo, C., Islam, M.N., Sheils, N.E., Buresh, J., Reps, J., Schuemie, M.J., Ryan, P.B., Edmondson, M., Duan, R., Tong, J., et al. (2022). DLMM as a lossless one-shot algorithm for collaborative multi-site distributed linear mixed models. Nat. Commun. *13*, 1678. https://doi.org/10.1038/s41467-022-29160-4.

143. Cancer Grand Challenges. PROSPECT. https://www.cancergrandchallenges.org/prospect.

144. International Agency for Research on Cancer. DISCERN: Discovering the Causes of Three Poorly Understood Cancers in Europe. https://discern.iarc.who.int/.

145. Terry, M.B., Tehranifar, P., and Kehm, R.D. (2025). Stemming the Rising Tide of Early-Onset Cancers: Research Gaps and Public Health Actions. Annu. Rev. Public Health. https://doi.org/10.1146/annurev-publhealth-082024-063125.

146. Nichols, H.B., Schoemaker, M.J., Wright, L.B., McGowan, C., Brook, M.N., McClain, K.M., Jones, M.E., Adami, H.O., Agnoli, C., Baglietto, L., et al. (2017). The Premenopausal Breast Cancer Collaboration: A Pooling Project of Studies Participating in the National Cancer Institute Cohort Consortium. Cancer Epidemiol. Biomarkers Prev. *26*, 1360–1369. https://doi.org/10.1158/1055-9965.Epi-17-0246.

147. Breast Cancer Association Consortium, Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K.A., Parsons, M.T., Fortuno, C., et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N. Engl. J. Med. *384*, 428–439. https://doi.org/10.1056/NEJMoa1913948.

148. Rubin, D.B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. J. Am. Stat. Assoc. *100*, 322–331. https://doi.org/10.1198/016214504000001880.

149. Wagner, C., Carmeli, C., Jackisch, J., Kivimäki, M., van der Linden, B.W.A., Cullati, S., and Chiolero, A. (2024). Life course epidemiology and public health. Lancet Public Health *9*, e261–e269. https://doi.org/10.1016/s2468-2667(24)00018-5.

150. 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. https://doi.org/10.1007/s10654-023-01001-8.

151. Naimi, A.I., Cole, S.R., and Kennedy, E.H. (2017). An introduction to g methods. Int. J. Epidemiol. *46*, 756–762. https://doi.org/10.1093/ije/dyw323.

152. 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. https://doi.org/10.1093/ije/dyw367.

153. 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. https://doi.org/10.1111/biom.13847.

154. Greenland, S. (2008). Measures of effect and measures of association. In Modern Epidemiology (Wolters Kluwer Health/Lippincott Williams & Wilkins), pp. 51–70.

155. Rockhill, B., Newman, B., and Weinberg, C. (1998). Use and misuse of population attributable fractions. Am. J. Public Health *88*, 15–19. https://doi.org/10.2105/ajph.88.1.15.

156. 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. https://doi.org/10.1146/annurev-publhealth-031811-124627.

157. 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. https://doi.org/10.1016/j.cmet.2017.07.008.

158. Zipkin, D.A., Umscheid, C.A., Keating, N.L., Allen, E., Aung, K., Beyth, R., Kaatz, S., Mann, D.M., Sussman, J.B., Korenstein, D., et al. (2014). Evidence-based risk communication: a systematic review. Ann. Intern. Med. *161*, 270–280. https://doi.org/10.7326/m14-0295.

159. Plutynski, A. (2023). On Explaining Peto's Paradox. Eur. J. Epidemiol. *38*, 1245–1250. https://doi.org/10.1007/s10654-022-00920-2.

160. Rodgers, K.M., Udesky, J.O., Rudel, R.A., and Brody, J.G. (2018). Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. Environ. Res. *160*, 152–182. https://doi.org/10.1016/j.envres.2017.08.045.

161. Terry, M.B., and Colditz, G.A. (2023). Epidemiology and Risk Factors for Breast Cancer: 21st Century Advances, Gaps to Address through Interdisciplinary Science. Cold Spring Harb. Perspect. Med. *13*, a041317. https://doi.org/10.1101/cshperspect.a041317.

162. Boddy, A.M. (2023). The need for evolutionary theory in cancer research. Eur. J. Epidemiol. *38*, 1259–1264. https://doi.org/10.1007/s10654-022-00936-8.

163. 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*, eaaw0726. https://doi.org/10.1126/science.aaw0726.

164. 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. https://doi.org/10.1038/s41586-019-1672-7.

165. Abby, E., Dentro, 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. https://doi.org/10.1038/s41588-022-01280-z.

166. 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. https://doi.org/10.1038/nbt.4099.

167. 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. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.28.667191.

168. Zhang, B., Miao, B., Fu, S., Shao, W., Coarfa, C., Kumar, R., Kuntala, P., Park, B., Grimm, S., Jangid, R., et al. (2025). Cross-tissue molecular responses in the liver and blood after toxicant exposures. https://doi.org/10.21203/rs.3.rs-7610132/v1.

169. Faubert, B., Solmonson, A., and DeBerardinis, R.J. (2020). Metabolic reprogramming and cancer progression. Science *368*, eaaw5473. https://doi.org/10.1126/science.aaw5473.

170. 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*, eabf3002. https://doi.org/10.1126/science.abf3002.

171. Naik, S., and Fuchs, E. (2022). Inflammatory memory and tissue adaptation in sickness and in health. Nature *607*, 249–255. https://doi.org/10.1038/s41586-022-04919-3.

172. Nagaraja, S., Ojeda-Miron, L., Zhang, R., Oreskovic, E., Hock, C., Hu, Y., Zeve, D., Sharma, K., Hyman, R.R., Zhang, Q., et al. (2026). Epigenetic memory of colitis promotes tumour growth. Nature. Published online *25*. https://doi.org/10.1038/s41586-026-10258-4.

173. 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. https://doi.org/10.1158/2159-8290.Cd-24-1687.

174. 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*, msac084. https://doi.org/10.1093/molbev/msac084.

175. 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. https://doi.org/10.1038/s41586-020-1943-3.

176. 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. https://doi.org/10.1038/s41586-025-09025-8.

177. 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. https://doi.org/10.1016/j.cell.2024.02.009.

178. 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. https://doi.org/10.1038/s41586-025-09096-7.

179. 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. https://doi.org/10.1056/NEJMra0708473.

180. Walker, C.L., and Ho, S.M. (2012). Developmental reprogramming of cancer susceptibility. Nat. Rev. Cancer *12*, 479–486. https://doi.org/10.1038/nrc3220.

181. Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat. Rev. Genet. *13*, 97–109. https://doi.org/10.1038/nrg3142.

182. Patti, M.E. (2013). Intergenerational programming of metabolic disease: evidence from human populations and experimental animal models. Cell. Mol. Life Sci. *70*, 1597–1608. https://doi.org/10.1007/s00018-013-1298-0.

183. 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*, e39380. https://doi.org/10.7554/eLife.39380.

184. Al Nabhani, Z., and Eberl, G. (2020). Imprinting of the immune system by the microbiota early in life. Mucosal Immunol. *13*, 183–189. https://doi.org/10.1038/s41385-020-0257-y.

185. Fitz-James, M.H., and Cavalli, G. (2022). Molecular mechanisms of transgenerational epigenetic inheritance. Nat. Rev. Genet. *23*, 325–341. https://doi.org/10.1038/s41576-021-00438-5.

186. Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. Cell *157*, 121–141. https://doi.org/10.1016/j.cell.2014.03.011.

187. Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. Cell *140*, 883–899. https://doi.org/10.1016/j.cell.2010.01.025.

188. de Visser, K.E., and Joyce, J.A. (2023). The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell *41*, 374–403. https://doi.org/10.1016/j.ccell.2023.02.016.

189. Constantinides, M.G., and Belkaid, Y. (2021). Early-life imprinting of unconventional T cells and tissue homeostasis. Science *374*, eabf0095. https://doi.org/10.1126/science.abf0095.

190. Balbo, S., Turesky, R.J., and Villalta, P.W. (2014). DNA adductomics. Chem. Res. Toxicol. *27*, 356–366. https://doi.org/10.1021/tx4004352.

191. Balbo, S., Hecht, S.S., Upadhyaya, P., and Villalta, P.W. (2014). Application of a high-resolution mass-spectrometry-based DNA adductomics approach for identification of DNA adducts in complex mixtures. Anal. Chem. *86*, 1744–1752. https://doi.org/10.1021/ac403565m.

192. Hilakivi-Clarke, L. (2014). Maternal exposure to diethylstilbestrol during pregnancy and increased breast cancer risk in daughters. Breast Cancer Res. *16*, 208. https://doi.org/10.1186/bcr3649.

193. Sun, Y., Wang, Q., Zhang, Y., Geng, M., Wei, Y., Liu, Y., Liu, S., Petersen, R.B., Yue, J., Huang, K., et al. (2020). Multigenerational maternal obesity increases the incidence of HCC in offspring via miR-27a-3p. J. Hepatol. *73*, 603–615. https://doi.org/10.1016/j.jhep.2020.03.050.

194. Merlo, L.M.F., Pepper, J.W., Reid, B.J., and Maley, C.C. (2006). Cancer as an evolutionary and ecological process. Nat. Rev. Cancer *6*, 924–935. https://doi.org/10.1038/nrc2013.

195. 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. https://doi.org/10.1093/molbev/msz242.

196. Laplane, L., and Maley, C.C. (2024). The evolutionary theory of cancer: challenges and potential solutions. Nat. Rev. Cancer *24*, 718–733. https://doi.org/10.1038/s41568-024-00734-2.

197. Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306–313. https://doi.org/10.1038/nature10762.

198. 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. https://doi.org/10.1093/jnci/djy168.

199. 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. https://doi.org/10.1158/0008-5472.Can-22-1508.

200. Mauri, G., Patelli, G., Crisafulli, G., Siena, S., and Bardelli, A. (2025). Tumor "age" in early-onset colorectal cancer. Cell *188*, 589–593. https://doi.org/10.1016/j.cell.2024.12.003.

201. Lea, A.J., Tung, J., Archie, E.A., and Alberts, S.C. (2017). Developmental plasticity: Bridging research in evolution and human health. Evol. Med. Public Health *2017*, 162–175. https://doi.org/10.1093/emph/eox019.

202. Brodin, P. (2022). Immune-microbe interactions early in life: A determinant of health and disease long term. Science *376*, 945–950. https://doi.org/10.1126/science.abk2189.

203. Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. Br. J. Cancer *8*, 1–12. https://doi.org/10.1038/bjc.1954.1.

204. Ashley, D.J. (1969). The two "hit" and multiple "hit" theories of carcinogenesis. Br. J. Cancer *23*, 313–328. https://doi.org/10.1038/bjc.1969.41.

205. 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. https://doi.org/10.1016/j.mbs.2023.109091.

206. Friedrich, M.J., Pham, J., Tian, J., Chen, H., Huang, J., Kehl, N., Liu, S., Lash, B., Chen, F., Wang, X., et al. (2025). Transient hepatic reconstitution of trophic factors enhances aged immunity. Nature *650*, 481–489. https://doi.org/10.1038/s41586-025-09873-4.

207. 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. https://doi.org/10.1093/jnci/81.24.1879.

208. 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. https://doi.org/10.1038/sj.bjc.6600008.

209. Antoniou, A.C., Pharoah, P.P., Smith, P., and Easton, D.F. (2004). The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br. J. Cancer *91*, 1580–1590. https://doi.org/10.1038/sj.bjc.6602175.

210. 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. https://doi.org/10.1200/jco.2008.17.4797.

211. 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. https://doi.org/10.1038/s41416-019-0476-8.

212. 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. https://doi.org/10.1038/s41571-020-0388-9.

213. 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. https://doi.org/10.1038/s41575-020-00368-3.

214. 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. https://doi.org/10.1053/j.gastro.2023.02.021.

215. 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. https://doi.org/10.1093/jnci/88.6.359.

216. 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. https://doi.org/10.1093/aje/152.10.950.

217. 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. https://doi.org/10.1007/s10549-017-4391-5.

218. 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. https://doi.org/10.1200/cci-24-00200.

219. 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. https://doi.org/10.1001/jamanetworkopen.2025.12681.

220. Placido, D., Yuan, B., Hjaltelin, J.X., Zheng, C., Haue, A.D., Chmura, P.J., Yuan, C., Kim, J., Umeton, R., Antell, G., et al. (2023). A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. Nat. Med. *29*, 1113–1122. https://doi.org/10.1038/s41591-023-02332-5.

221. 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. https://doi.org/10.1038/s41586-025-09529-3.

222. Tao, L.R., Ye, Y., and Zhao, H. (2023). Early breast cancer risk detection: a novel framework leveraging polygenic risk scores and machine learning. J. Med. Genet. *60*, 960–964. https://doi.org/10.1136/jmg-2022-108582.

223. Kraft, P., Wacholder, S., Cornelis, M.C., Hu, F.B., Hayes, R.B., Thomas, G., Hoover, R., Hunter, D.J., and Chanock, S. (2009). Beyond odds ratios–communicating disease risk based on genetic profiles. Nat. Rev. Genet. *10*, 264–269. https://doi.org/10.1038/nrg2516.

224. 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. https://doi.org/10.1038/s41467-020-19600-4.

225. Baccarelli, A., Dolinoy, D.C., and Walker, C.L. (2023). A precision environmental health approach to prevention of human disease. Nat. Commun. *14*, 2449. https://doi.org/10.1038/s41467-023-37626-2.

226. 2025). Cancer screening must become more precise. Nat. Med. 31, 3935. 10.1038/s41591-025-04148-x.

227. D'Alise, A.M., Willis, J., Duzagac, F., Hall, M.J., Cruz-Correa, M., Idos, G.E., Thirumurthi, S., Ballester, V., Leoni, G., Garzia, I., et al. (2026). Nous-209 neoantigen vaccine for cancer prevention in Lynch syndrome carriers: a phase 1b/2 trial. Nat. Med. *209*. https://doi.org/10.1038/s41591-025-04182-9.

228. Wang, Y., Colditz, G.A., and Jiang, S. (2025). Score Test for Functional Markov Process With Image Predictor. Stat. Med. *44*, e70231. https://doi.org/10.1002/sim.70231.

229. 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. https://doi.org/10.1038/s41591-019-0673-2.

230. Oh, H.S.H., 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. https://doi.org/10.1038/s41586-023-06802-1.

231. Carrasco-Zanini, J., Pietzner, M., Davitte, J., Surendran, P., Croteau-Chonka, D.C., Robins, C., Torralbo, A., Tomlinson, C., Grünschläger, F., Fitzpatrick, N., et al. (2024). Proteomic signatures improve risk prediction for common and rare diseases. Nat. Med. *30*, 2489–2498. https://doi.org/10.1038/s41591-024-03142-z.

232. Hazelton, W.D., Ni, P., Harlass, M., Hahn, A.I., Tian, R., Zauber, 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. Preprint at medRxiv. https://doi.org/10.1101/2024.10.15.24315456.

233. National Cancer Institute. Cancer Intervention and Surveillance Modeling Network. https://cisnet.cancer.gov/.

234. Edwards, B.K., Ward, E., Kohler, B.A., Eheman, C., Zauber, 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. https://doi.org/10.1002/cncr.24760.

235. 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. https://doi.org/10.1093/jnci/djs136.

236. 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. https://doi.org/10.1001/jama.2023.25881.

237. Vogelaar, I., van Ballegooijen, M., Schrag, D., Boer, R., Winawer, S.J., Habbema, J.D.F., and Zauber, 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. https://doi.org/10.1002/cncr.22115.

238. Holford, T.R., Meza, R., Warner, K.E., Meernik, C., Jeon, J., Moolgavkar, S.H., and Levy, D.T. (2014). Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964–2012. JAMA *311*, 164–171. https://doi.org/10.1001/jama.2013.285112.

239. Jeon, J., Holford, T.R., Levy, D.T., Feuer, E.J., Cao, P., Tam, J., Clarke, L., Clarke, J., Kong, C.Y., and Meza, R. (2018). Smoking and lung cancer mortality in the United States from 2015 to 2065: a comparative modeling approach. Ann. Intern. Med. *169*, 684–693. https://doi.org/10.7326/M18-1250.

240. Hazelton, W.D., Jeon, J., Meza, R., and Moolgavkar, S.H. (2012). Chapter 8. The FHCRC lung cancer model. Risk Anal. *32*, S99–S116. https://doi.org/10.1111/j.1539-6924.2011.01681.x.

241. 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. https://doi.org/10.1007/s10552-015-0559-7.

242. Curtius, K., Hazelton, W.D., Jeon, J., and Luebeck, E.G. (2015). A Multi-scale Model Evaluates Screening for Neoplasia in Barrett's Esophagus. PLoS Comput. Biol. *11*, e1004272. https://doi.org/10.1371/journal.pcbi.1004272.

243. Turner, M.C., Nieuwenhuijsen, M., Anderson, K., Balshaw, D., Cui, Y., Dunton, G., Hoppin, J.A., Koutrakis, P., and Jerrett, M. (2017). Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations. Annu. Rev. Public Health *38*, 215–239. https://doi.org/10.1146/annurev-publhealth-082516-012802.

244. Roos, L.G., and Slavich, G.M. (2023). Wearable technologies for health research: Opportunities, limitations, and practical and conceptual considerations. Brain Behav. Immun. *113*, 444–452. https://doi.org/10.1016/j.bbi.2023.08.008.

245. Clark, L.P., Zilber, D., Schmitt, C., Fargo, D.C., Reif, D.M., Motsinger-Reif, A.A., and Messier, K.P. (2025). A review of geospatial exposure models and approaches for health data integration. J. Expo. Sci. Environ. Epidemiol. *35*, 131–148. https://doi.org/10.1038/s41370-024-00712-8.

246. Wan, M., Simonin, E.M., Johnson, M.M., Zhang, X., Lin, X., Gao, P., Patel, C.J., Yousuf, A., Snyder, M.P., Hong, X., et al. (2025). Exposomics: a review of methodologies, applications, and future directions in molecular medicine. EMBO Mol. Med. *17*, 599–608. https://doi.org/10.1038/s44321-025-00191-w.

247. Rowles, J.L., 3rd, and Patti, G.J. (2026). Decoding cancer across scales with metabolomics. Nat. Rev. Cancer. https://doi.org/10.1038/s41568-026-00908-0.

248. Luk, A., Wild, S.H., Jones, S., Anjana, R.M., Hivert, M.F., McCaffrey, J., Gregg, E.W., and Misra, S. (2025). Early-onset type 2 diabetes: the next major diabetes transition. Lancet *405*, 2313–2326. https://doi.org/10.1016/s0140-6736(25)00830-x.

249. Jain, V., Minhas, A.M.K., Morris, A.A., Greene, S.J., Pandey, A., Khan, S.S., Fonarow, G.C., Mentz, R.J., Butler, J., and Khan, M.S. (2022). Demographic and Regional Trends of Heart Failure-Related Mortality in Young Adults in the US, 1999–2019. JAMA Cardiol. *7*, 900–904. https://doi.org/10.1001/jamacardio.2022.2213.

250. Vasan, R.S., Song, R.J., and van den Heuvel, E.R. (2022). Temporal Trends in Incidence of Premature Cardiovascular Disease Over the Past 7 Decades: The Framingham Heart Study. J. Am. Heart Assoc. *11*, e026497. https://doi.org/10.1161/jaha.122.026497.

251. Hendriks, S., Peetoom, K., Bakker, C., van der Flier, W.M., Papma, J.M., Koopmans, R., Verhey, F.R.J., de Vugt, M., Köhler, S., et al.; Young-Onset Dementia Epidemiology Study Group (2021). Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis. JAMA Neurol. *78*, 1080–1090. https://doi.org/10.1001/jamaneurol.2021.2161.

252. 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. https://doi.org/10.1093/jnci/93.14.1054.

253. 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. https://doi.org/10.1158/1055-9965.Epi-20-0237.