Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed

Paul Brennan, PhD [ID],[1],[*],[†] George Davey-Smith, MD, FRS [ID],[2],[†]

[1]International Agency for Research on Cancer (IARC/WHO), Genomic Epidemiology Branch, Lyon, France and [2]Medical Research Council Integrative Epidemiology Unit (IEU), University of Bristol, Clifton, Bristol, UK

[†]The authors contributed equally to this work.

[*]**Correspondence to:** Paul Brennan, PhD, International Agency for Research on Cancer (IARC/WHO), Genomic Epidemiology Branch, 150 Cours Albert Thomas, Lyon 69008, France (e-mail: brennanp@iarc.fr).

Abstract

The burden of cancer from a clinical, societal, and economic viewpoint continues to increase in all parts of the world, along with much debate regarding how to confront this. Projected increases in cancer indicate a 50% increase in the number of cases over the next 2 decades, with the greatest proportional increase in low- and medium-income settings. In contrast to the historic high cancer burden due to viral and bacterial infections in these regions, future increases are expected to be due to cancers linked to westernization including breast, colorectum, lung, and prostate cancer. Identifying the reasons underlying these increases will be paramount to informing prevention efforts. Evidence from epidemiological and laboratory studies conducted in high-income countries over the last 70 years has led to the conclusion that approximately 40% of the cancer burden is explained by known risk factors—the 2 most important being tobacco and obesity in that order—raising the question of what is driving the rest of the cancer burden. International cancer statistics continue to show that approximately 80% of the cancer burden in high-income countries could be preventable in principle, implying that there are important environmental or lifestyle risk factors for cancer that have not yet been discovered. Emerging genomic evidence from population and experimental studies points to an important role for nonmutagenic promoters in driving cancer incidence rates. New research strategies and infrastructures that combine population-based and laboratory research at a global level are required to break this deadlock.

There were 19 million cancer cases and 10 million cancer deaths worldwide in 2020 (1). More than half can be attributed to 6 cancer sites: breast, lung, colorectal, prostate, stomach, and liver cancer. These also comprise more than half of all cancer deaths, although others, including pancreatic and esophageal cancer, also result in a disproportionate number of cancer deaths because of their high case-fatality rate (1). A striking geographical pattern in cancer incidence rates continues to be found between populations. Every cancer type that is common in some population is rare in another, with the difference often being tenfold or more. These differences can occur over relatively small geographical distances, such as esophageal cancer within Northern Iran, central China, or East Africa (2).

In most instances, undercounting or missed diagnoses are unlikely to fully explain this phenomenon thanks to reasonably robust estimates from population-based cancer registries (3). Similarly, rapid changes in age-standardized cancer incidence over time rule out a genetic explanation, as do migrant studies, whereby migrants take on the incidence rate of their new found home within 1 or 2 generations (4). Instead, these large population differences imply that there are tangible (but often unknown) causes for each of these common cancer types. By identifying the specific causes, these could become the target for prevention.

Highlighting the potential for prevention is important given that the cancer burden is destined to increase globally. The cancer burden of 19 million new cases in 2020 is expected to increase to approximately 25 million by 2032 (5) and more than 30 million by 2040 (Figure 1). These numbers take into account the future growth and aging of populations worldwide and assume that age-specific incidence rates will remain constant. The rising cancer burden places increasing pressure on health systems worldwide, most markedly in low- and middle-income countries (LMIC) where the relative increase is greater and the capacity for cancer care is already overstretched (6).
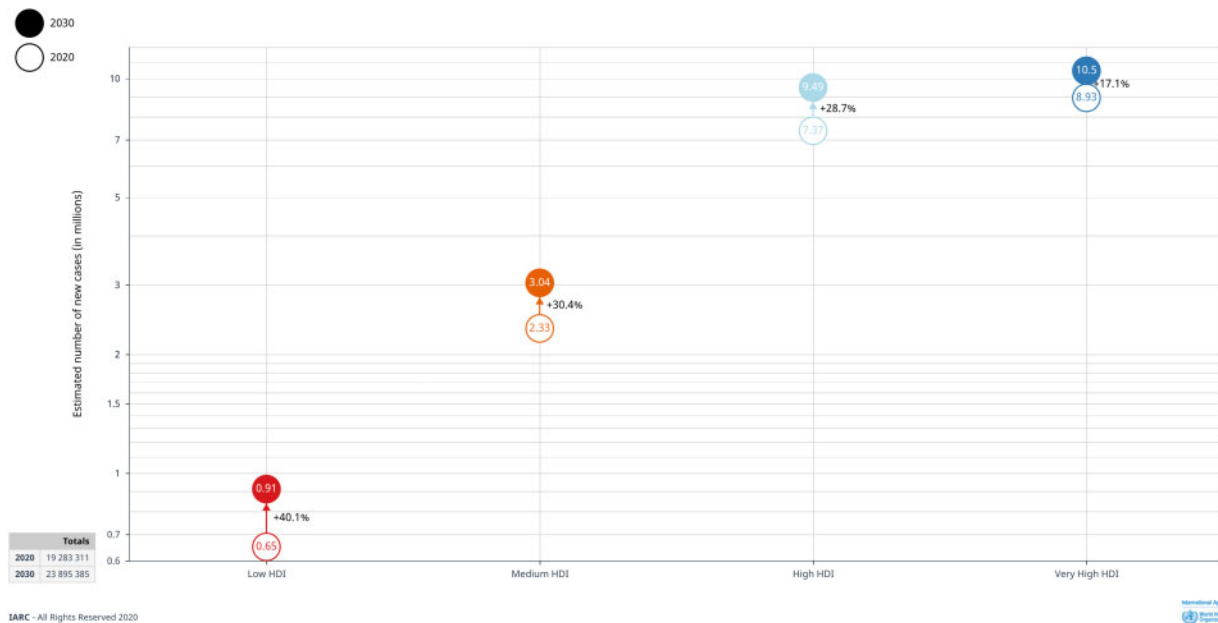
Figure 1. Cancer burden in low, medium, and high human development index (HDI) regions and percent increase in burden projected over the next 10 years. **Solid circles** represent 2030 projected data (in millions of cases), and **empty circles** represent actual 2020 data.

COMMENTARY

Faced with this challenge, the strategic options for reducing this cancer burden fall broadly into 1 of 3 categories: 1) reducing the occurrence of new cases (primary prevention), 2) identifying new cases earlier (secondary prevention), and 3) improving cancer care and cancer survival (tertiary prevention). Based on resources allocated, the emphasis of the cancer community including those who fund cancer research is to develop a greater knowledge of the biology of cancer and translation of this knowledge into treatments (7). This funding is in addition to large investments made by the pharmaceutical industry. Accurate breakdowns of cancer funding by area of research are limited, although globally, cancer prevention research is estimated to receive just between 2% and 9% of total cancer research funding (8). Participation in any large global cancer meeting would leave little doubt about the relatively low priority given to cancer prevention as opposed to new treatments or biological breakthroughs.

In defense of this strategy, the scientific progress in our understanding of cancer biology has been phenomenal. International genome-sequencing efforts have resulted in a comprehensive characterization of driver genes and somatic mutations for all common forms of cancer, and many rare ones (9), and our understanding of the role of epigenetics and the immune system in cancer has led to some remarkable discoveries (10). Yet, although progress in translation to the clinic has been made, it has been stuttering and limited (11). Most approved targeted or immune therapies do not result in cure but instead in an extended remission period that may often be measured in months, with a financial cost that is placing extreme pressures on health budgets in high-income settings, and is simply prohibitive elsewhere (12). Put simply, a strategy based primarily around treatment does not form a coherent response to the current global cancer burden, and a more comprehensive approach is required, involving prevention of cancer where possible, early detection when feasible, and appropriate and cost-effective treatment and care for all.

The Percentage of Cancer That Is Preventable, Both in Practice and in Theory

There is genuine confusion about the proportion of the global cancer burden that is preventable, both in practice and in theory (13). In particular, what proportion is attributable to the environment or lifestyle of a population (including infectious agents), what proportion to genetic inheritance, and what proportion to [bad] luck? An appreciation of this problem is important as any rational division of resources should take into account whether funding allocated to prevention is likely to have an impact on future disease morbidity and mortality. If, as has been argued, the role of primary prevention is limited beyond what we already know, then greater resources should be spent on detecting cancer at earlier and more treatable stages (14). However, if it is apparent that there are important causes that remain to be detected, then future research should clearly be targeted at uncovering these.

Prevention in Practice

Recent evaluations in the United Kingdom, Canada, Australia, Ireland, France, and the United States have come to a similar conclusion that approximately 30%-40% of the cancer burden in these countries can be linked to known risk factors that are largely modifiable. Specifically, it has been estimated that approximately 37% of new cancer cases in the United Kingdom, 41% in Canada and France, 32% in Australia, and 42% in the United States were attributable to potentially modifiable risk factors (15–19). Any difference in the estimates is because of prevalence of known risk factors, as well as slightly different methodological approaches (20). Independent effects from nonmodifiable risk factors such as age at menarche and menopause may also contribute to the cancer burden.

When broken down by exposure, more than half of this known cancer burden is due to 2 prominent risk factors—tobacco

Table 1. Population attributable fraction (PAF) and number of cases attributable to known carcinogenic exposures, overall and separately for ever smokers and never smokers, applied to the UK population[a]

Exposure	UK population (359 547 cancer cases)		Ever smokers (169 210 cancer cases)		Never smokers (190 337 cancer cases)	
	PAF	Attributable cases	PAF	Attributable cases	PAF	Attributable cases
Smoking	15.1	54 271	32.07	54 271	0	0
Overweight/obesity	6.3	22 761	6.3	10 712	6.3	12 049
Ultraviolet radiation	3.8	13 604	3.8	6402	3.8	7202
Occupation	3.8	13 558	3.8	6381	3.8	7177
Infections	3.6	13 086	3.6	6159	3.6	6927
Alcohol	3.3	11 894	3.3	5598	3.3	6296
Insufficient fiber	3.3	11 693	3.3	5503	3.3	6190
Ionizing radiation	1.9	6954	1.9	3273	1.9	3681
Processed meat	1.5	5352	1.5	2519	1.5	2833
Air pollution	1.0	3591	1.0	1690	1.0	1901
Not breastfeeding	0.7	2582	0.7	1215	0.7	1367
Insufficient physical activity	0.5	1917	0.5	902	0.5	1015
Postmenopausal hormones	0.4	1371	0.4	645	0.4	726
Oral contraceptives	0.2	807	0.2	380	0.2	427
All of the above	37.7	135 507	50.1	84 851	26.6	50 647

[a]PAF estimates are from Brown et al. (15). These estimates assume an absence of interaction between tobacco and other carcinogenic exposures. This is likely to be a simplification, especially for the role of alcohol and head and neck cancers. Also, PAF estimates do not include potential cancers prevented by exposures (eg, oral contraceptive use and endometrial or ovarian cancer).

and overweight. For example, in the United Kingdom in 2015, 15% of all cancers could be attributed to smoking and 6.3% to overweight and obesity (15). In a similar exercise conducted for 2010, tobacco was responsible for 19.4% and overweight for 5.5%, reflecting the changing prevalence of these risk factors over just 5 years (21). These estimates are an overview and will differ by characteristics such as age, sex, and background exposures. For example, among smokers in the United Kingdom, approximately 50% of all cancer are attributable to known causes, with smoking explaining about one-third of the cancer burden in this group. Among never smokers however, only about one-quarter of all cases can be attributed to known risk factors (Table 1).

Implementing what is already known has an enormous potential to reduce cancer risk at the population level (22). Workplace carcinogens were among some of the first to be identified, leading to substantial control measures that greatly reduced the burden of occupational cancers, at least in high-income countries (23). Tobacco control measures at local and national levels have also led to sharply decreasing incidence rates for lung and other tobacco-related cancers, and public advice measures on sunlight exposure have had a beneficial effect on rates of melanoma (24). Hepatitis B vaccination programs have an important impact against liver cancers in high-risk settings, and human papillomavirus vaccination offers the possibility of making cervical cancer a rare disease (25).

Prevention in Theory

If approximately 40% of cancer could be prevented in high-income populations, what is causing the other 60%? A small amount is likely to be due to inherited genetic susceptibility that predisposes to a high risk of specific cancers (26). Although this contribution is difficult to measure because of limited data from multiple populations, it is likely to be no more than approximately 5% in most populations (26).

That some part of the cancer burden may simply be due to bad luck or be inherently stochastic is implicit in the multistage model of carcinogenesis, as initially proposed in 1954 (27). As Sir Richard

Doll remarked in a recent commentary on this classic paper, "whether an exposed subject does or does not develop a cancer is largely a matter of luck: bad luck if the several necessary changes all occur in the same stem cell when there are several thousand such cells at risk, good luck if they don't" (28). A recent analysis incorporating stem cell turnover rates in various tissues argued that 50%-60% of the cancer burden was primarily due to chance (14), suggesting that important additional risk factors for cancer were unlikely to be discovered. This much publicized paper, and a subsequent follow-up (29), led to several competing analyses of similar or augmented datasets [eg, (30)] and commentaries attempting to frame the debate within the broader historical context of cancer epidemiology (31). In a comprehensive recent review of the debate from a philosophy of science perspective, Anya Plutynski (32) illustrates how the framing of the questions being asked and the meaning ascribed to fundamental notions such as the nature of chance do not allow for definitive resolution.

A pragmatic approach to estimating the cancer burden that is preventable in principle was initially proposed by Sir Richard Doll in 1977 (33) and with Sir Richard Peto (34) in their landmark work on the causes of cancer in 1981. The approach assumes a background rate of any particular cancer site that is not caused by lifestyle or environmental effects or hereditary effects. This rate must, by definition, be operable in all populations. If a particular cancer is predominantly driven by random sporadic mutations, then one would expect that the incidence of that cancer, after adjusting for age and sex, will not differ substantially between populations, apart from fluctuations because of small numbers. In this situation, our best estimate of the background rate of any cancer will be the lowest rate observed in any population, assuming that the population is large enough to provide an accurate measure of the rate and that it has a reliable cancer registry.

An indication of the quality of any cancer registry is inclusion in the International Agency for Research on Cancer (IARC) *Cancer Incidence in Five Continents*, currently in its 11th edition with data from 343 cancer registries (3). Of these, some could be judged small (ie, based on a population of less than 1 000 000),

leaving 325 large cancer registries in 65 countries. One can adopt a conservative measure of the background rate due to chance for each cancer site by using cancer rates at the bottom fifth percentile among all these 325 cancer registries (Table 2). Using this rate as a background, one can obtain an estimate of the proportion of cancers that are preventable in principle within any country, for men and women separately, by comparing it to these lowest rates seen elsewhere in the world.

Using UK cancer incidence rates as an example, this exercise provides a background rate of 52.2 per 100 000 in men and 58.7 per 100 000 in women, resulting in approximately 17% of cancers among men and 22% of cancers among women being attributed to this background, "chance" rate. Even these numbers are likely to be an overestimate. For example, the rate for lung cancer among men (9.5 per 100 000) is likely to be partially driven by smoking, as rates in never smokers are known to be about 4 per 100 000 (similar to the rate observed among women). These estimates do, however, indicate that approximately 80% of the cancer burden in a high-income country such as the United Kingdom is likely to be due to environmental or lifestyle exposures, both known and unknown, a result remarkably similar to the analysis published 40 years ago by Doll and Peto (34) using cancer incidence data from the Connecticut cancer registry from 1968 to 1972. Furthermore, given that we can account for approximately 40% of the cancer burden in the United Kingdom, these results indicate that approximately an additional 40% of the cancer burden in the United Kingdom remains, in principle, to be elucidated.

For cancers where the greatest differential is observed, cancer registries around the bottom fifth percentile are typically from LMIC countries, although not necessarily populations of extreme poverty or rural lifestyle by contemporary standards, even if the situation 40-50 years ago may have been different. For example, the bottom fifth percentile for rectal cancer among men is found in Pasto, Colombia; for pancreatic cancer, it is Songkhla, Thailand; and for renal cancer, it is Shexian County, China. Although it is impractical to suggest that if high-income countries could mirror the socioeconomic characteristics of these populations, then low cancer rates would follow, understanding the underlying causes between these differentials, be they related to modifiable lifestyle factors or more ubiquitous population-level exposures, could lead to strategies for important cancer reductions.

Where Are the Missing Causes of Cancer?

The gaps in our knowledge on the causes for most cancers that are not predominantly due to tobacco or known infections pose a major challenge for cancer prevention. Large cohort studies of hundreds of thousands of individuals have not provided conclusive evidence for important dietary effects on multiple cancers, with some exceptions such as processed meat (35), and the idea that specific nutrients or food types are driving cancer incidence rates has lost ground.

It is possible that we have underestimated the magnitude of some known causes, given that estimates of their effect may be unreliable or poorly measured. For example, it took more than 50 years of research for the impact of tobacco on multiple cancers to be fully understood, and this is for an exposure that is relatively easy to measure over the life course (36). The role of obesity may be undergoing a similar reevaluation. Obesity is typically measured by body mass index that has important limitations including its inability to differentiate between muscle mass and fat. A one-off measure is also unlikely to characterize obesity at different ages. If obesity over the life course is the more relevant risk factor, then the estimated cancer burden of 6.3% in the United Kingdom associated with overweight and obesity may represent an important underestimate. Recent genetic analyses based on a Mendelian randomization approach provide evidence that this is the case, with the risk when using genetic markers of obesity estimated to be about twice as high for several cancers as opposed to a one-off measure of obesity (37). Genetic data also indicate that overweight and obesity may be contributing to other cancers not thought to be linked to overweight, including lung cancer (38), possibly because

Table 2. Comparison of UK age-adjusted site-specific incidence rates per 100 000 with those from the bottom fifth percentile, among cancer registries included in *Cancer Incidence in Five Continents* by Bray et al. (3) and based on a population of at least 1 000 000

Men			Women		
Cancer site	Worldwide bottom fifth percentile incidence rate per 100 000	UK age-adjusted incidence rate per 100 000	Cancer site	Worldwide bottom fifth percentile incidence rate per 100 000	UK age-adjusted incidence rate per 100 000
Lung[a]	9.53	38.6	Breast	19.12	87.4
Stomach	4.36	7.6	Lung[a]	3.81	27.3
Rectum	3.77	15.5	Cervix uteri	3.78	7.7
Colon	3.64	22.8	Ovary	3.54	10.4
Non-Hodgkin lymphoma	3.26	12.4	Colon	2.88	16.9
Liver	3.03	5	Rectum	2.73	7.9
Bladder	2.63	12.7	Corpus uteri	2.57	13.7
Brain, nervous system	2.27	6.4	Thyroid	2.15	4.6
Prostate	2.25	70.9	Non-Hodgkin lymphoma	2.04	9
Pancreas	1.85	7.3	Stomach	2.02	3.2
Esophagus	1.45	9.8	Brain, nervous system	1.76	4.3
Larynx	1.44	3.6	Pancreas	1.18	5.7
Kidney	1.27	9.8	Liver	1.17	2.2
Myeloid leukemia	1.23	4.2	Myeloid leukemia	1.09	3
Gallbladder, and biliary tract	1.03	1.2	Gallbladder, and biliary tract	0.99	1.3
All sites except C44	52.2	295.1	All sites except C44	58.7	263.5

[a]Including trachea and bronchus. C44 = other malignant neoplasms of skin.

COMMENTARY

of the behavior-modifying effect that obesity has on cigarette smoking (39). Overall, these results suggest that overweight and obesity may in fact be comparable to tobacco smoking in terms of new cases of cancer that are caused each year in a country such as the United Kingdom. A broader understanding of metabolic health beyond the role of obesity, including positive effects of physical activity and the negative effects of diabetes and insulin resistance, may also help clarify important causes of cancer that are currently underappreciated. Inaccurate measurement over the life course of other known causes in Table 1, including physical activity, occupational exposures, dietary fiber, and air pollution, will similarly result in an underestimation of their contribution to the overall cancer burden. The possibility that infectious agents may account for additional cancers also needs continued investigation (40).

Much concern has been raised about the possible carcinogenic effect of environmental factors occurring at low levels although in a ubiquitous fashion across populations. Examples include nutritional contaminants such as pesticides, low-level electromagnetic radiation, or water pollutants such as trihalomethanes or per- and polyfluoroalkyl substances (41,42). Accurate measurements over the life course is challenging for such exposures. For example, air pollution was classified by the IARC Monographs program in 2013 as a group 1 carcinogen (43) and has been estimated to account for approximately 1% of the overall cancer burden (15). There is, however, much uncertainty around this estimate. Similarly, radon is a recognized cause of lung cancer, in particular among never smokers, although evaluating the disease burden is challenging (44). Furthermore, approximately 50 substances at workplaces have been classified as carcinogenic to humans by the IARC Monographs Program (23), with many of these also occurring in the natural environment or through industrial waste disposal, but at much lower exposure levels. The cancer research community has been less successful at identifying important effects for such environmental carcinogens. This may be because such effects are modest and unlikely to be driving an important part of the cancer burden. It could also be because many environmental exposures are extremely difficult to measure (45). Furthermore, epidemiological studies are poor at studying ubiquitous effects that we are all exposed to. For example, if every adult in a population smoked 20 cigarettes a day, then the disease burden from this habit would be immense, but a standard epidemiological study would be unsuccessful in shining any light on this link (31). This illustrates the limits of epidemiology focused on individual exposures and the need for alternative complementary lines of evidence for carcinogen identification and assessment. Formal triangulation of evidence across different domains—all of which may yield biased findings, but where biases would be unrelated across the various approaches—is one approach that could assist causal inference (46,47).

What Strategies Are Likely to Help Identify Additional Causes of Cancer?

Develop National Infrastructures, Increase Their Accessibility, and Incorporate New Technologies

Most epidemiological findings are based on population-based studies supported by laboratory investigations, and emerging study designs and methodologies are likely to provide additional and complementary lines of evidence for cancer causation. Many of these studies rely on the recruitment and follow-up of hundreds of thousands of individuals, with extensive collection of biological samples at baseline, and subsequent clinical follow-up. An example of a mature cohort that has enabled hundreds of collaborations is the European Prospective Investigation into Cancer (EPIC) that includes data on more than 500 000 individuals recruited in the 1990s, with a major focus on the role of nutrition and lifestyle effects on cancer (48). Another example is the UK Biobank, which has been revolutionary in the scope: 500 000 individuals with extensive baseline, demographic, clinical, and biological data, as well as comprehensive follow-up (49). By making this an open access resource to the academic community, it has stimulated a new generation of researchers to investigate links between a broad range of chronic diseases and lifestyle and genetic risk factors.

Beyond the size and accessibility of the UK Biobank, key to its success and utility is the availability of genetic data on all participants. Along with other similar studies with genetic data, this is allowing for a more complete understanding of the genetic architecture of many traits and exposures that are difficult to measure in a conventional observational setting. Using these genetic measures rather than the exposures themselves is providing some surprising results for both novel and established cancer risk factors. This is called a Mendelian randomization (MR) analysis (50,51), and results from such studies are likely to have an increasingly prominent role in assessing causality of nongenetic exposures. For example, MR studies of alcohol consumption and esophageal cancer were instrumental in identifying acetaldehyde as a group 1 carcinogen by the IARC Monographs (52,53) and helped dispel myths about the supposed protective effects of moderate alcohol consumption and cardiovascular disease. MR studies are providing evidence for the causal effects of a range of metabolic features across multiple cancers. For example, studies of colorectal cancer have identified effects for hyperlipidemia and fatty acid profile and provided negative evidence for vitamin D (54). Studies on fasting insulin levels highlight potential causal effects for both renal and pancreatic cancers (55,56). Technological advances that allow metabolomic and proteomic analysis of many thousands of proteins and metabolites are destined to provide information on cancer pathways, and the combination of genetic, proteomic, and metabolomic data will offer additional strategies to identify protein and metabolite markers for cancer that are potentially causal and even druggable (57). MR could also usefully inform which factors to investigate through expensive long-term randomized controlled trials. For example, we suspect that if MR findings had been available, investment may not have been made in a large-scale selenium supplementation trial (based on conventional observational epidemiological findings) aimed at reducing prostate cancer risk (58).

There are, however, important limitations and threats to observational cancer epidemiology. A major limitation is the lack of diversity in large population cohorts in many parts of the world including South America, Africa, and South Asia. Recent cohorts in Iran, Pakistan, and South Africa are notable exceptions (59). There are also major threats and impediments to undertake observational epidemiology across international borders. Recent years have seen a steady progression in the administrative and legal workload that is required before a database or biological sample can be shared between study partners. This is often out of all proportion to the potential risk associated with any inadvertent sharing of data and certainly with respect to the lack of concrete examples of inappropriate use of samples or data, or negative consequences from unauthorized release of data or identification of individuals. Perhaps the most singular example is the encroachment of recent European Union–wide rules related to General Data Protection

Regulation (60) and the negative effect they are having on the potential for European and international research groups to conduct common research projects.

Looking in the Genome—Mutational Signatures

The first 2 cancer genomes were reported in 2009 for lung cancer and melanoma and highlighted the potential for specific causes of cancer to leave their mutational fingerprint or signature in the tumor (61,62). Subsequently, large international cancer sequencing efforts such as The Cancer Genome Atlas and the International Cancer Genome Consortium were undertaken at substantial public expense with a primary aim being to identify cancer genes and somatic mutations that may provide targets for therapy. An additional and extremely beneficial outcome has been the identification of multiple mutational signatures that can be linked to the underlying causes of a cancer, including both exogenous and endogenous factors. The most recent compendium of mutational signatures includes 60 signatures based on single-base or double-base mutations and a further 17 based on specific structural alterations (63). Examples include signatures for varied exposures such as tobacco smoking, ultraviolet (UV) light, and aflatoxin. Such studies can help identify potential carcinogens that were not previously suspected. For example, the International Cancer Genome Consortium study of renal cancer in Europe identified a signature in cases specific for aristolochic acid, a potent carcinogen known to cause upper urothelial tract cancers in parts of East Asia because of the consumption of traditional herbal remedies (64). The exposure to aristolochic acid was subsequently confirmed in studies of normal tissue, although the source of the exposure is still unclear (65). This signature has been detected in a range of other cancers including bladder cancer (66) and liver cancer (67). Another example is the presence of a signature related to UV light among children with B-cell acute lymphoblastic leukemia, suggesting a potential role for UV light in the etiology of these cancers (68). These findings are supported by traditional epidemiological evidence (69). These studies have led to large-scale initiatives that seek to understand novel causes of cancer by revealing mutational signatures through whole-genome sequencing. In particular, the Cancer Research UK Grand Challenge Mutographs project aims to include 5000 individuals across multiple cancer sites (https://www.mutographs.org/). The first results, including 557 whole genomes of squamous esophageal cancer from diverse high-risk populations such as Iran, East Africa, China, Japan, and Brazil, showed a remarkable similarity in the mutational signatures between these genomes, with no specific mutational signature linked to known or suspected risk factors such as consumption of hot drinks. These results point to the importance of nonmutational processes in cancer development (70).

Studies of the mutational landscape of normal tissue have illustrated the unexpected frequency of clonal expansion of cells with high levels of gene mutations normally associated with cancer development. For example, sequencing of normal skin cells identified a mutation burden similar to many skin cancers, including signatures for UV light exposure with more than 30% of cells containing driver mutations (71). Similarly, normal cells from the lung of smokers have been found to have a high rate of mutations, similar to the mutation profile of lung cancers, whereas those of ex-smokers appear to have a much reduced mutation profile, suggesting that normal lung tissue can repopulate epithelial tissue with cells lacking mutations upon smoking cessation (72). Driver mutations have also been detected in up to 50% of normal cells from the esophagus (73) but only about 1% of cells from the colon (74). These results are challenging and argue against the model of cancer development resulting from a simple accumulation of driver mutations that arise because of mutagenic exposures. Animal-based studies of known or suspected carcinogens reveal that most of them do not generate distinct mutational signatures and do not increase the overall tumor burden. These results suggest that most carcinogens do not act as mutagens and that mutations arise from tissue-specific endogenous processes (75). This raises the question of how most carcinogens induce cancers, if they are not mutagenic (76). An alternative is that nonmutagenic-promoting agents are the rate limiting steps for the development of most cancers, with these promoter events occurring via either endogenous or exogenous sources (77), perhaps mediated by the composition and selective activity of tissue microenvironments (78,79). If true, it would suggest a need for better detection of environmental carcinogens that act as promoters as opposed to direct mutagens. It would also imply an important need to understand the processes by which tumor promotion works and whether screening of compounds for tumor promotion activity is in fact feasible.

In summary, the increasing cancer burden that is destined to occur in all parts of the world calls for an increased focus on identifying the causes of cancer and subsequent implementation of this knowledge into cancer prevention. The economic costs of cancer care as well as the effect of lost economic productivity can be measured in trillions of US dollars and, although most apparent in LMICs, are likely to be unsustainable even in high-income countries. As has been said before, we will not be able to treat our way out of this cancer problem (80), and progress in our knowledge of preventable causes of cancer will require ambitious investment in large population-based studies and biorepositories, coupled with appropriate cancer genomics and laboratory studies. All countries in the world are faced with this challenge, calling for an increased level of international collaboration and pooling of talent and resources.

Data Availability

The data underlying this article are available in *Global Cancer Observatory (GLOBOCAN),* at https://gco.iarc.fr/. The datasets were derived from sources in the public domain [see reference (2)].

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Murphy G, McCormack V, Abedi-Ardekani B, et al. International cancer seminars: a focus on esophageal squamous cell carcinoma. *Ann Oncol.* 2017;28(9):2086–2093.
3. Bray FC, Mery L, Piñeros M, Znaor A, Zanetti R, Ferlay J. *Cancer Incidence in Five Continents, Vol. XI.* https://ci5.iarc.fr. Published 2017. Accessed October 1, 2021.
4. Parkin DM, Khlat M. Studies of cancer in migrants: rationale and methodology. *Eur J Cancer.* 1996;32a(5):761–771.
5. Ferlay J, Laversanne M, Ervik M, et al. *Global Cancer Observatory: Cancer Tomorrow.* Lyon, France: International Agency for Research on Cancer; 2020. https://gco.iarc.fr/tomorrow. Accessed October 27, 2021.
6. Mariotto AB, Enewold L, Zhao J, et al. Medical care costs associated with cancer survivorship in the United States. *Cancer Epidemiol Biomarkers Prev.* 2020;29(7):1304–1312.
7. Collins FS, Anderson JM, Austin CP, et al. Basic science: bedrock of progress. *Science.* 2016;351(6280):1405.
8. Drexler M. The Cancer Miracle Isn't a Cure. It's Prevention. *Magazine of the Harvard T.H. Chan School of Public Health.* 2019.
9. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature.* 2020;578(7793):82–93.
10. Sánchez NS, Mills GB, Mills Shaw KR. Precision oncology: neither a silver bullet nor a dream. *Pharmacogenomics.* 2017;18(16):1525–1539.
11. Prasad V. Perspective: the precision-oncology illusion. *Nature.* 2016;537(7619):S63.
12. DeMartino PC, Miljkovic MD, Prasad V. Potential cost implications for all US Food and Drug Administration oncology drug approvals in 2018. *JAMA Intern Med.* 2021;181(2):162–167.
13. Shield KD, Parkin DM, Whiteman DC, et al. Population attributable and preventable fractions: cancer risk factor surveillance, and cancer policy projection. *Curr Epidemiol Rep.* 2016;3(3):201–211.
14. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science.* 2015;347(6217):78–81.
15. Brown KF, Rumgay H, Dunlop C, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *Br J Cancer.* 2018;118(8):1130–1141.
16. Islami F, Goding Sauer A, Miller KD, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin.* 2018;68(1):31–54.
17. Soerjomataram I, Shield K, Marant-Micallef C, et al. Cancers related to lifestyle and environmental factors in France in 2015. *Eur J Cancer.* 2018;105:103–113.
18. Wilson LF, Antonsson A, Green AC, et al. How many cancer cases and deaths are potentially preventable? Estimates for Australia in 2013. *Int J Cancer.* 2018;142(4):691–701.
19. Poirier AE, Ruan Y, Volesky KD, et al.; for the ComPARe Study Team. The current and future burden of cancer attributable to modifiable risk factors in Canada: summary of results. *Prev Med.* 2019;122:140–147.
20. Bray F, Soerjomataram I. Population attributable fractions continue to unmask the power of prevention. *Br J Cancer.* 2018;118(8):1031–1032.
21. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer.* 2011;105(suppl 2):S77–81.
22. Colditz GA, Wolin KY, Gehlert S. Applying what we know to accelerate cancer prevention. *Sci Transl Med.* 2012;4(127):127rv4.
23. Loomis D, Guha N, Hall AL, et al. Identifying occupational carcinogens: an update from the IARC Monographs. *Occup Environ Med.* 2018;75(8):593–603.
24. Schüz J, Espina C, Wild CP. Primary prevention: a need for concerted action. *Mol Oncol.* 2019;13(3):567–578.
25. Finn OJ. The dawn of vaccines for cancer prevention. *Nat Rev Immunol.* 2018;18(3):183–194.
26. Huang KL, Mashl RJ, Wu Y, et al.; for the Cancer Genome Atlas Research Network. Pathogenic germline variants in 10,389 adult cancers. *Cell.* 2018;173(2):355–370.e14.
27. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer.* 1954;8(1):1–12.
28. Doll R. Commentary: the age distribution of cancer and a multistage theory of carcinogenesis. *Int J Epidemiol.* 2004;33(6):1183–1184.
29. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science.* 2017;355(6331):1330–1334.
30. Hochberg ME, Noble RJ. A framework for how environment contributes to cancer risk. *Ecol Lett.* 2017;20(2):117–134.
31. Davey Smith G, Relton CL, Brennan P. Chance, choice and cause in cancer aetiology: individual and population perspectives. *Int J Epidemiol.* 2016;45(3):605–613.
32. Plutynski A. Is cancer a matter of luck? *Biol Philos.* 2021;36(1):3.
33. Doll R. Strategy for detection of cancer hazards to man. *Nature.* 1977;265(5595):589–596.
34. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst.* 1981;66(6):1191–1308.
35. Rohrmann S, Overvad K, Bueno-de-Mesquita HB, et al. Meat consumption and mortality–results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med.* 2013;11:63.
36. Doll R, Peto R, Boreham J, et al. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ.* 2004;328(7455):1519.
37. Mariosa D, Carreras-Torres R, Martin RM, et al. Commentary: What can Mendelian randomization tell us about causes of cancer? *Int J Epidemiol.* 2019;48(3):816–821.
38. Carreras-Torres R, Johansson M, Haycock PC, et al. Obesity, metabolic factors and risk of different histological types of lung cancer: a Mendelian randomization study. *PLoS One.* 2017;12(6):e0177875.
39. Carreras-Torres R, Johansson M, Haycock PC, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ.* 2018;361:k1767.
40. Zur Hausen H. The search for infectious causes of human cancers: where and why (Nobel lecture). *Angew Chem Int Ed Engl.* 2009;48(32):5798–5808.
41. Evlampidou I, Font-Ribera L, Rojas-Rueda D, et al. Trihalomethanes in drinking water and bladder cancer burden in the European Union. *Environ Health Perspect.* 2020;128(1):17001.
42. Shearer JJ, Callahan CL, Calafat AM, et al. Serum concentrations of per- and polyfluoroalkyl substances and risk of renal cell carcinoma. *J Natl Cancer Inst.* 2021;113(5):580–587.
43. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Outdoor Air Pollution.* Lyon, France: International Agency for Research on Cancer; 2016.
44. Torres-Durán M, Barros-Dios JM, Fernández-Villar A, et al. Residential radon and lung cancer in never smokers. A systematic review. *Cancer Lett.* 2014;345(1):21–26.
45. White E, Armstrong BK, Saracci R. Principles of Exposure Measurement in Epidemiology: Collecting, *Evaluating, and Improving Measures of Disease Risk Factors* 2nd ed. Oxford: Oxford University Press; 2008.
46. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol.* 2016;45(6):1866–1886.
47. Munafò MR, Higgins JPT, Smith GD. Triangulating evidence through the inclusion of genetically informed designs. *Cold Spring Harb Perspect Med.* 2021;11(8):a040659.
48. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* 2002;5(6b):1113–1124.
49. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203–209.
50. Adam D. The gene-based hack that is revolutionizing epidemiology. *Nature.* 2019;576(7786):196–199.
51. Davey Smith G, Ebrahim S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.
52. Personal habits and indoor combustions. Volume 100 E. A review of human carcinogens. *IARC Monogr Eval Carcinog Risks Hum.* 2012;100(pt E):1–538.
53. Davey Smith G, Holmes MV, Davies NM, et al. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur J Epidemiol.* 2020;35(2):99–111.
54. Cornish AJ, Tomlinson IPM, Houlston RS. Mendelian randomisation: a powerful and inexpensive method for identifying and excluding non-genetic risk factors for colorectal cancer. *Mol Aspects Med.* 2019;69:41–47.
55. Carreras-Torres R, Johansson M, Gaborieau V, et al. The role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: a Mendelian randomization study. *J Natl Cancer Inst.* 2017;109(9):djx012.
56. Johansson M, Carreras-Torres R, Scelo G, et al. The influence of obesity-related factors in the etiology of renal cell carcinoma–a Mendelian randomization study. *PLoS Med.* 2019;16(1):e1002724.
57. Folkersen L, Gustafsson S, Wang Q, et al. Genomic evaluation of circulating proteins for drug target characterisation and precision medicine. *bioRxiv.* 2020. doi:10.1101/2020.04.03.023804:2020.04.03.023804.
58. Yarmolinsky J, Bonilla C, Haycock PC, et al.; for the PRACTICAL Consortium. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. *J Natl Cancer Inst.* 2018;110(9):1035–1038.
59. Manolio TA, Goodhand P, Ginsburg G. The International Hundred Thousand Plus Cohort Consortium: integrating large-scale cohorts to address global scientific challenges. *Lancet Digit Health.* 2020;2(11):e567–e568.
60. Peloquin D, DiMaio M, Bierer B, et al. Disruptive and avoidable: GDPR challenges to secondary research uses of data. *Eur J Hum Genet.* 2020;28(6):697–705.

COMMENTARY

61. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2010;463(7278): 191–196.

62. Pleasance ED, Stephens PJ, O'Meara S, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*. 2010;463(7278): 184–190.

63. Alexandrov LB, Kim J, Haradhvala NJ, et al.; for the PCAWG Consortium. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793): 94–101.

64. Scelo G, Riazalhosseini Y, Greger L, et al. Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nat Commun*. 2014; 5:5135.

65. Turesky RJ, Yun BH, Brennan P, et al. Aristolochic acid exposure in Romania and implications for renal cell carcinoma. *Br J Cancer*. 2016;114(1):76–80.

66. Poon SL, Huang MN, Choo Y, et al. Mutation signatures implicate aristolochic acid in bladder cancer development. *Genome Med*. 2015;7(1):38.

67. Lu Z-N, Luo Q, Zhao L-N, et al. The mutational features of aristolochic acid-induced mouse and human liver cancers. *Hepatology*. 2020;71(3):929–942.

68. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555(7696): 371–376.

69. Coste A, Hémon D, Orsi L, et al. Residential exposure to ultraviolet light and risk of precursor B-cell acute lymphoblastic leukemia: assessing the role of individual risk factors, the ESCALE and ESTELLE studies. *Cancer Causes Control*. 2017;28(10):1075–1083.

70. Moody S, Senkin S, Islam SMA, et al. Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. *Nat Genet*. 2021;53(11):1553–1563.

71. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348(6237):880–886.

72. Yoshida K, Gowers KHC, Lee-Six H, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*. 2020;578(7794):266–272.

73. Martincorena I, Fowler JC, Wabik A, et al. Somatic mutant clones colonize the human esophagus with age. *Science*. 2018;362(6417):911–917.

74. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*. 2019;574(7779):532–537.

75. Riva L, Pandiri AR, Li YR, et al. The mutational signature profile of known and suspected human carcinogens in mice. *Nat Genet*. 2020;52(11):1189–1197.

76. Lopez-Bigas N, Gonzalez-Perez A. Are carcinogens direct mutagens? *Nat Genet*. 2020;52(11):1137–1138.

77. Balmain A. The critical roles of somatic mutations and environmental tumor-promoting agents in cancer risk. *Nat Genet*. 2020;52(11):1139–1143.

78. Lochhead P, Chan AT, Nishihara R, et al. Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. *Mod Pathol*. 2015;28(1):14–29.

79. Rafaeva M, Erler JT. Framing cancer progression: influence of the organ- and tumour-specific matrisome. *FEBS J*. 2020;287(8):1454–1477.

80. Brennan P, Wild CP. Genomics of cancer and a new era for cancer prevention. *PLoS Genet*. 2015;11(11):e1005522.