Causality and Chance in the Development of Cancer

Lucio Luzzatto, M.D., and Pier Paolo Pandolfi, M.D., Ph.D.

The notion that, in addition to heredity (HER) and the environment (ENV), chance (C) is also a factor in oncogenesis is not new.1, 2 However, it has received new impetus from the recent work by Tomasetti and Vogelstein,3 who reported a strong correlation between the frequency of tumors of individual organs or tissue types and the estimated number of stem-cell divisions in those organs or tissues. Thus, the wide variation between common tumors (e.g., colorectal cancer, which will develop in nearly 1 in 20 persons) and very rare tumors (e.g., osteosarcoma, which will develop in fewer than 1 in 10,000 persons) could be explained in large part by how many stem-cell divisions — each of which entails the risk of random mutations — have accumulated by a certain age or over a lifetime. This work has had resounding echoes in the general press, and some journalists have conveyed the misleading take-home message that the cause of cancer is not lifestyle but, rather, bad luck. Here we discuss briefly how chance, heredity, and the environment overlap and interact.

The current model of tumor formation, or oncogenesis, is based on the notion that each of a succession of somatic mutations (Fig. 1) confers onto the mutant cell a growth advantage over normal cells in a particular microenvironment. The neoplastic cell, having accumulated a number \(n\) of causal (or driver) mutations, is now capable of aberrant growth in a given microenvironment.4 Thus, at the somatic-cell level, oncogenesis recapitulates a Darwinian process in which mutations are the innovative force, whereas the environment selects the mutations that are advantageous. An articulate formulation of this model was first given by Cairns in 1975,5 before any specific oncogenic mutations were known at all. An estimate of \(n\) was first given by Armitage and Doll in 19546; later, Knudson7 surmised, from epidemiologic data, that for sporadic retinoblastoma, \(n=2\). It now appears likely that \(n\) varies in different tumors, but it is probably a small number, perhaps 3 to 6 in most cases.8 Today the landscape of somatic mutations in individual types of cancer has been characterized extensively.9

Somatic mutations are by their nature stochastic events10: they fall under C, chance. The simplest type of mutation, a single base change, is a tax that must be paid every time a cell replicates. By any standards, the tax is small — estimated to be of the order of \(10^{-7}\) per gene per cell division11 — but considering the number of cell divisions it takes to make an adult human, one can estimate that in an adult virtually every gene will be mutated in at least one cell. This type of spontaneous mutation is stochastic, because it results from mispairing, which in turn results from the equilibrium that exists in solution between tautomeric forms of the purine and pyrimidine bases12; this equilibrium is dictated by quantum-mechanical principles.

The number \(M\) of somatic mutations that spontaneously accumulate in any set of cells (e.g., one tissue or its stem cells, or the whole body) is proportional to the number \(D\) of cell divisions:

\[
M = \mu \times D.
\] (1)

The proportionality factor \(\mu\) is the mutation rate. The majority of mutations that make up \(M\) are irrelevant to cancer: we are all riddled with somatic mutations in nonstem cells, which cannot therefore be oncogenic (except for those that are powerful enough to make a normal nonstem cell become a tumor stem cell13, 14). In normal tissues (even in stem cells), we must have mutations that are neutral with respect to oncogenesis. However, a very small minority of the somatic mutations that make up \(M\) are oncogenic; therefore, \(M\) measures, in first approximation, the risk of cancer. Equation 1 accounts well for the major effect of age on the incidence of cancer, because somatic mutations accumulate with time.
Figure 1. Discrete Stochastic Events in the Change from a Normal Cell to a Cancer Cell.

Panel A depicts the process of neoplastic transformation. In this particular case, the first oncogenic event is a somatic mutation (red star symbol), the second event is epigenetic (a change in chromatin conformation, symbolized by a yellow star), and the third event is again a somatic mutation. The probability that three oncogenic events will accumulate in the progeny of a single cell depends on the number \(D\) of cell divisions — because each round of DNA replication carries a risk of mutation — and on the mutation rate \(\mu\). Panel B illustrates the path for a person in whom both \(D\) and \(\mu\) are low; therefore, during the entire lifetime of this person, it might happen that no cell has accumulated a sufficient number of oncogenic events to produce a cancer cell. Panel C illustrates the path for a person in whom both \(D\) and \(\mu\) are higher; therefore, it is more likely that in a single cell, the number of oncogenic events that have accumulated is sufficient to produce cancer. The figure is not meant to imply that in any particular person \(D\) and \(\mu\) must be both high or both low; any combination is possible. In addition, \(D\) and \(\mu\) may change over time in the same person. Because somatic mutations and epigenetic events have a stochastic nature, chance always plays a role in tumorigenesis; however, hereditary and environmental factors, through their effects on \(\mu\) and \(D\), can exploit chance so strongly as to become predominant.
Tomasetti and Vogelstein\textsuperscript{3} have shown how extensively $D$ varies from one tissue to another, and this may be the main determinant of the highly variable frequency of cancer in different tissues. Some findings are surprising. For instance, the small bowel is a very large organ that produces enormous numbers of cells daily, yet its rate of cancer development is very low. The estimate that Tomasetti and Vogelstein have extracted from the literature for the number of stem-cell divisions ($D$) in the small intestine is lower than the number of such divisions in the colon; however, the method for assessing tissue stem-cell divisions is not standardized, and plasticity of stem-cell hierarchies may have to be taken into account.\textsuperscript{15} There are other possibilities. First, after absorption of up to 2 liters of water per day, the concentration of mutagens from food may be higher in the colon; even in mice, amino-$\alpha$-carboline causes more tumors in the colon than in the small intestine.\textsuperscript{16} With reference to equation 1, this means that $\mu_{\text{ENV}}$ (see below) may be higher in the colon than in the small intestine. Second, different segments of the intestine have very different floras — now termed the microbiome — and it has been suggested that this may affect carcinogenesis.\textsuperscript{17}

It also seems strange at first sight that leukemia is so much more rare than, say, breast cancer or prostate cancer, given that the rate of cell division in the hematopoietic tissue is very high. However, the vast majority of these cell divisions are in differentiating cells and in differentiated cells, whereas hematopoietic stem cells are very few (perhaps only about 400 contribute to hematopoiesis at any one time\textsuperscript{18}). On the other hand, for the breast and prostate, there seems to be a dire lack of information on stem cells,\textsuperscript{19} and both of these organs stand out in that their epithelial cells are subject to strong hormonal regulation. Unlike $D$, $\mu$ in the absence of exogenous agents is probably relatively uniform in different tissues of one individual\textsuperscript{20}; thus, in each tissue of each person, the risk of cancer (proportional to $M$) depends on the $D$ value of that tissue and on the $\mu$ value of that person. Mutations occur by chance, and what genes they hit is also largely due to chance. However, their numbers ($M$) vary a great deal because neither $\mu$ nor $D$ is fixed; equation 1 helps to explain how HER and ENV, by influencing $\mu$ and $D$, interact with $C$ (chance).

The $\mu$ value must be genetically determined because it is much higher than normal in persons with a defect in DNA repair (e.g., those with Fanconi’s anemia\textsuperscript{21}). In addition, it varies considerably in different individuals; indeed, log $\mu$ has a normal gaussian distribution in healthy people,\textsuperscript{21} like other quantitative traits (e.g., stature) that depend on the action of several genes. To this extent, HER determines in each person a constitutional value of $\mu$. On the other hand, mutagens will, by definition, affect $\mu$,\textsuperscript{22,23} and they may do so to a different extent in different tissues, depending on the mode of entry and on the metabolism of each individual mutagen. In other words, one might regard $\mu$ as consisting of two components: $\mu = \mu_{\text{ENV}} + \mu_{\text{GEN}}$, whereby the former is intrinsic or constitutional, and the latter may change depending on exposure to mutagens and on lifestyle. As for $D$, we know little about its genetic determinants: perhaps it has little genetic variation, because it is tightly tied to organogenesis in the embryo and to stem-cell renewal in the adult. However, $D$ will be increased — sometimes greatly so — by ENV effects, such as those that cause or influence inflammation and tissue regeneration,\textsuperscript{24} and $D$ may be decreased by calorie restriction.\textsuperscript{25} Thus, HER and ENV can have major effects on $M$ and therefore on the risk of cancer; however, it will be still by chance (the C factor) that an oncogenic set of mutations will accumulate in an individual cell.

Here are a few examples of how ENV, HER, or both can increase the risk of cancer by affecting $\mu$, $D$, or both. Cigarette smoking causes lung cancer through mutagens (increase in $\mu$)\textsuperscript{26} and also because inflammation will increase $D$ in the bronchial epithelium.\textsuperscript{27} Hepatitis C causes death of hepatocytes and inflammatory changes that tend to subvert the liver architecture: both will increase $D$ in liver stem cells, until the development of cirrhosis and eventually liver cancer.\textsuperscript{28} The hepatitis B virus does the same, but it might also affect $\mu$ on integration in the host-cell genome.\textsuperscript{29} In familial adenomatous polyposis, through a mutation in the gene APC, the $D$ value must be massively increased in intestinal epithelial stem cells\textsuperscript{30}; thus, through a HER effect, the hazard of colon cancer shifts from a risk to a certainty. In Fanconi’s anemia, there is a strong HER effect because an inherited mutation of one of the FANC genes markedly increases $\mu$ (see
above), and either leukemia or a solid tumor is almost inevitably produced. Ultraviolet light has a strong ENV effect on μ, which is highly pertinent to the pathogenesis of skin cancer. The risk of breast cancer is markedly increased with BRCA1 or BRCA2 mutations (a HER effect that is probably mediated by increased μ) but also, and much more frequently, by the action of hormones, which are influenced by pregnancy, lactation, and the use of contraceptives. In other words, there is a strong ENV effect when we include lifestyle as part of ENV.

Since the seminal work on the hypomethylation of certain genes in colon cancer, a large body of data has accumulated on epigenetic changes in relation to cancer, leading to the notion of cancer as a dysregulated epigenome. Epigenetic events that modify a gene and may thus modify the cell in which they occur have three important features in common with somatic mutations: they are discrete events that to a large extent occur at random; they are faithfully perpetuated in the progeny of a somatic cell; and their occurrence can be influenced by ENV and, probably to a lesser extent, by HER. With these factors taken into account, equation 1 becomes

\[ M' = (\mu + \mu_e) \times D, \]

where \(M'\) is the sum total of somatic mutations and epigenetic changes, and \(\mu\) is the rate of occurrence of epigenetic changes. At least one somatic mutation is almost certainly present in every cancer; however, at the somatic cell level (i.e., in the domain of oncogenesis), an epigenetic change is essentially equivalent to a somatic mutation and is therefore susceptible to Darwinian selection to the extent that it confers a growth advantage. The cooperation in oncogenesis of somatic mutations and of epigenetic events is made more poignant by the fact that not infrequently in cancer one finds mutations of one of the many genes that influence epigenetics.

In our view, there is no conflict between the stochastic component in oncogenesis and the causes of cancer that can be classified under HER and ENV. The most important point in the article by Tomasetti and Vogelstein is that they have pinpointed in \(D\), which depends on the biology of human development, a likely basis for the enormously different incidence of tumors arising in the different tissues of the body. At the same time, these authors have considered deliberately an “extra risk score” that takes into account the effects of HER and ENV.

It is clear that \(M\) will be lowest when someone is born with a low \(\mu\), is not exposed to mutagens, and does not have disorders or lifestyle factors that increase \(D\). Under these conditions, the risk of cancer is lowest (although increasing with age); if cancer does develop in such a person, the role of \(C\) emerges as prominent and predominant. A mistake that has surfaced in the lay press is not to realize that when the “extra risk” is huge, as in the case of cigarette smoking, an ENV factor can dwarf \(C\). Indeed, HER and ENV have effects (mediated by \(\mu\), \(\mu_e\), and \(D\) that at least in principle can be quantified for each type of cancer, whereas \(C\) is a factor in every case of cancer (Fig. 1).

Because we all have a heritage and we cannot live except in an environment, it is meaningless to ask how much cancer would occur if HER and ENV played no role. But no matter how powerful these factors may be, they still require somatic mutations — stochastic events — before a tumor develops: hence, the notion of cancer due to bad luck. Of course, one could reverse the argument: indeed, only a minority of heavy smokers end up with lung cancer, and the majority of inherited cancer-prone genes have incomplete penetrance. Thus, even someone who has smoked throughout life or who has a BRCA2 frameshift mutation may, through good luck, not get cancer.

Epicurus (4th century BCE) tried to visualize chance by saying that occasionally the normally straight paths of atoms in the universe bend a little, and the atoms “swerve.” If one considers mispairing during DNA replication, perhaps he was not far off. In order to prevent cancer, we must use our ingenuity to minimize the effect of HER on its causation; we must use our willpower to stop smoking (ENV); and we must reduce industrial and air pollution to prevent physical and chemical carcinogenesis (ENV). Rather than adopting the heredity-centered view that we ought to blame cancer on our ancestors or the guilt-ridden view that we ought to blame it always on the way we behave, we must also recognize the role of chance in oncogenesis, because it is always there.