

Systems biology and its potential role in radiobiology

Ludwig Feinendegen · Philip Hahnfeldt ·
Eric E. Schadt · Michael Stumpf · Eberhard O. Voit

Received: 4 June 2007 / Accepted: 21 November 2007 / Published online: 18 December 2007
© Springer-Verlag 2007

Abstract About a century ago, Conrad Röntgen discovered X-rays, and Henri Becquerel discovered a new phenomenon, which Marie and Pierre Curie later coined as *radio-activity*. Since their seminal work, we have learned much about the physical properties of radiation and its effects on living matter. Alas, the more we discover, the more we appreciate the complexity of the biological processes that are triggered by radiation exposure and eventually lead (or do not lead) to disease. Equipped with modern biological methods of high-throughput experimentation, imaging, and vastly increased computational

prowess, we are now entering an era where we can piece some of the multifold aspects of radiation exposure and its sequelae together, and develop a more systemic understanding of radiogenic effects such as radio-carcinogenesis than has been possible in the past. It is evident from the complexity of even the known processes that such an understanding can only be gained if it is supported by mathematical models. At this point, the construction of comprehensive models is hampered both by technical inadequacies and a paucity of appropriate data. Nonetheless, some initial steps have been taken already and the generally increased interest in systems biology may be expected to speed up future progress. In this context, we discuss in this article examples of relatively small, yet very useful models that elucidate selected aspects of the effects of exposure to ionizing radiation and may shine a light on the path before us.

Overview on talks presented at the First International Workshop on Systems Radiation Biology, February 14–16, 2007, GSF-Research Centre, Neuherberg, Germany.

L. Feinendegen
Department of Nuclear Medicine, University Hospital,
Heinrich-Heine-University, Düsseldorf, Germany

P. Hahnfeldt
Center for Cancer Systems Biology,
Caritas St. Elizabeth's Medical Center,
Tufts University School of Medicine,
Boston, MA, USA

E. E. Schadt
Department of Genetics, Rosetta Inpharmatics, LLC,
Merck & Co., Inc., Seattle, WA, USA

M. Stumpf
Centre for Bioinformatics, Imperial College London,
London, UK

E. O. Voit (✉)
Department of Biomedical Engineering,
Georgia Institute of Technology,
313 Ferst Drive, Suite 4103, Atlanta,
GA 30332-0535, USA
e-mail: eberhard.voit@bme.gatech.edu

Introduction

In late 1895, Wilhelm Conrad Röntgen of the University of Würzburg in Germany demonstrated an “unknown source of rays” by placing his wife’s hand between a cathode tube and a fluorescent screen [1]. The unknown electromagnetic “X-rays” penetrated her hand, clearly showing her bones and her wedding band, and Röntgen received for the discovery of these invisible rays, the first Nobel Prize in physics in 1901. Fascinated by Röntgen’s discovery the French physicist Antoine Henri Becquerel investigated the similarities and differences between X-rays, fluorescence, and phosphorescence and discovered that the energy in all three cases had to come from an emanating source object and penetrated through interspersed matter [2]. While this

insight was important, it was his famous student couple, Pierre and Marie Curie, who realized that the observed phenomenon is an atomic property of matter. They coined the term “radio-activity” and ultimately figured out how to measure emanations from various elements, such as uranium, polonium, and radium to cause electrical charges in an appropriate metal, i.e., altered conductivity. A deeper understanding of atoms and nuclei came with Ernest Rutherford’s famous experiments of 1911, in which he used radioactive material that emitted alpha particles to bombard a piece of gold foil. The fact that only relatively small numbers of the particles were stopped by the foil or bounced back led him to conclude that matter consisted primarily of empty space surrounding atomic nuclei. Almost 20 years passed before the *Drosophila* geneticist Hermann Müller demonstrated that ionizing radiation not just passes through matter but that it is able to cause mutations in living organisms [3, 4]. His work and that of Timofejew-Ressowski, Zimmer, and Delbrück may be considered the foundation for research on DNA damage and repair, which began in the 1930s [5] and continues to the present day [6].

By now we have come to realize that radiation may affect living tissue in multifold ways. We understand in some detail how high doses of ionizing radiation may cause physical damage to chromosomes in the form of single or double strand breaks, which may be sometimes followed by faulty rejoining or even translocations. It is clear that radiation, even in smaller doses, can lead to mutations, which in stem cells may or may not be direct precursors to carcinogenesis. We also know now that much subtler modifications may result in the form of altered gene expression, that signal transduction systems and transcription factor networks may be activated in unique ways, that the intracellular oxidative state may be affected, and that metabolic pathways may directly or indirectly respond to radiation exposure (see Fig. 1). We appreciate that inter-individual variations in the body’s defenses may cause one person to fall ill from radiation, while another person may remain apparently unharmed. Indeed, Zimmer, a pioneer in radiation research, wrote, “one cannot use radiations for elucidating the normal state of affairs without considering the mechanisms of their actions, nor can one find out much about radiation-induced changes without being interested in the normal state of the material under investigation” [7].

The discoveries of the diverse biological effects of ionizing radiation were followed, with some delay, by specific mathematical modeling efforts that aimed at quantification, prediction, and, importantly, cancer risk estimation. Consistent with the tradition of reductionism in biology, each of these models focused on one particular effect, and there is no doubt that we have learned quite a bit from these types of studies. Some of the investigations

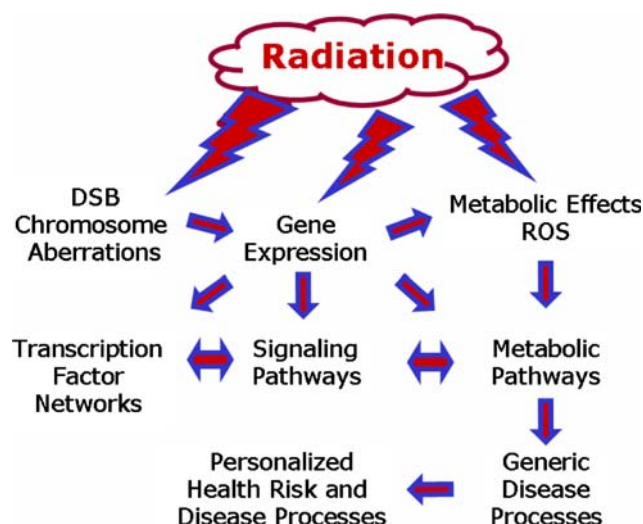


Fig. 1 Radiation exposure triggers a variety of responses at numerous levels of biological organization. These responses form a complex system that requires new approaches of computational systems biology

focused on mechanistic details, others on the generic process of carcinogenesis, and yet others on inferences from large datasets, including those collated from the atomic bomb survivors in Hiroshima and Nagasaki.

To appreciate the diversity of these approaches, it is useful to look at some of the recent literature. At the biophysical end, for instance, Friedland et al. [8] characterized some of the details of DNA damage and fragmentation, and Ballarini and Ottolenghi [9] described how chromosome aberrations may be induced even by low doses of radiation. Moolgavkar et al. [10, 11] studied how biological insights into DNA damage may be formalized, and proposed the concept of several hits needed to transform a normal cell into a cancer cell. This translation of mechanistic ideas into mathematical models of staged carcinogenesis has been widely used to assess radiation and environmental exposure risks. In yet another set of approaches, mechanistic modeling was applied to epidemiological data in an attempt to determine risks of disease more specifically than is possible with statistics alone [12].

While biophysical and mathematical models of radiation exposure and effects have been very important for our understanding of the responses of living organisms, each of them usually focuses on one or two aspects only, thereby ignoring the tightly connected and integrative nature of biological systems. Thus, while strand breaks may clearly lead to aberrant gene organization, which in itself is important to study, they also alter gene expression directly or through structural or regulatory effects on transcription factor networks. These combine with direct metabolic effects, such as changes in reactive oxygen or nitrogen species, ultimately resulting in altered physiological

responses that are still very difficult to predict. Moreover, it is well possible that ionizing radiation shows a variety of epigenetic effects, which may persist for a long time. In this context one should also keep in mind that exposure to ubiquitous primordial natural radiation leads to a life-long dose rate of about 0.1 mSv/h, which, in our body, results in about 100 radioactive decays per kg body weight per second. All these responses not only involve a large number of components and processes, but these processes are complex and usually nonlinear. The nonlinearity prevents us from using simple cause-and-effect argumentation or even the principle of superposition that allows us to study system components in isolation. In fact, many of these processes involve threshold-like phenomena, where responses to small and large inputs are qualitatively and not only quantitatively different. A premier example for the non-trivial nature of combined effects is the study of low doses of radiation, which reveals that extrapolation from high doses is not necessarily valid, because the high and low exposures involve vastly different response patterns, including repair and an array of defense and protective mechanisms commonly known as stress responses (see later sections).

The availability of molecular high-throughput experiments that are increasing in quality while decreasing in price and complication, combined with greatly improved access and power of computation, suggests that the time is ripe to begin assessing responses to radiation in a more comprehensive, integrative fashion that draws from the principles of the new field of systems biology. It is evident that we are still far away from being able to accomplish global analyses of radiation effects, but it is nevertheless feasible to take the first steps toward a fully integrative analysis on the various systems levels of a whole organism. These steps could consist of simultaneous analyses of many diverse data in some relatively coarse, global manner, but they might also merge two, three, or more formerly isolated modeling approaches into models that have more latitude.

In this article we provide a flavor of emerging, integrative approaches to radiation systems biology in the form of three vignettes. The first presents a concrete example of merging two traditionally isolated modeling ventures, namely the mathematical formulation of the molecular-level events of initiation, promotion, and transformation, which are considered key events in carcinogenesis, with the physical limitations that control the growth of tumors and critically affect cell–cell interactions. In this way, it is appreciated that the driving forces of cancer extend beyond the tumor cell. On the other hand, models of this type presume that cancer risks increase more or less monotonically with the extent of radiation exposure. This is not necessarily true. Thus, the second vignette describes some of the enormous complications that any extrapolation

efforts between high- to low-dose radiation assessments face. This discussion leaves no doubt that the simultaneous accounting for metabolic and signaling pathways, genetic alterations, epigenetic effects, and defense and repair mechanisms at several physiological levels requires formulations of radiation phenomena as dynamic systems that quickly overwhelm any purely intuitive approaches and mandate the development of mathematical and computational modeling approaches that address the general and specific nature of networks and systems. The third vignette discusses several strategies of network analysis to handle this complexity. In the simplest case, this is accomplished by simply abstracting systems components and processes into nodes and edges of a graph. Such a graph is then translated into linear or nonlinear models. Depending on how this translation is implemented, different insights into biological systems may be gained. Not surprisingly, if one increases the complexity of the chosen model, the model becomes more realistic and predictive, but the technical challenges of its design and analysis grow commensurately.

Combining organizational levels to estimate cancer risk

A comprehensive, biologically based model of radiogenic cancer risk should ideally involve all levels of biological organization, from damaged DNA to clinical manifestations of cancer. If such a model could be constructed, it would certainly be superior to the phenomenological alternatives currently employed, because it could, on one hand, rely on mechanistic rules to draw upon a broad range of data, and, on the other hand, explain a large portion of the details along the path from radiation exposure to cancer. The model would integrate all types of supporting data, including biophysical measurements on cell cultures, data obtained from experimental animals, clinical observations in specifically exposed human populations, such as uranium miners, or data from initially diverse populations such as the atomic bomb survivors. If the key biology tying these data together could be successfully incorporated in a model, this model could be expected to exhibit high predictive value. Alas, we are far from being able to collect and integrate all data types and to construct sufficiently flexible and efficient mathematical models to cover the entire process of carcinogenesis. As an intermediate goal, it is therefore useful to combine models on one or two levels at a time, in the hope of constructing sets of modules that will eventually be merged into a comprehensive model of radiogenic carcinogenesis. In this section, we connect, in a rudimentary fashion and as a proof of concept, the levels of molecular-level events of initiation and promotion with the higher-level scale of cell–cell interactions and the limitation of growth in maturing tumors.

Biologically based carcinogenesis modeling

Progress on mathematical cancer prediction has for the most part been through the modeling of the molecular steps leading up to the first cancer cell. Most of these models presume that cancer evolves through a series of mutations and clonal expansions, the end result being the creation of the first tumor cell. The simplest of these is the “initiation–promotion–transformation” timeline paradigm for radiogenic cancer development. By this mechanism, normal cells are randomly initiated (i.e., acquire a single growth-facilitating mutation) at some dose-dependent rate ω . As these clones expand during the promotion phase, cells in these populations proliferate at a rate λ and randomly undergo a second transforming event at a rate μ , where $\mu < \lambda$, to become tumor cells. The rate of tumor cell generation is identified by comparison of model predictions with epidemiologic data on carcinogenesis risk, after an adjustment (usually in the form of a lag time) is made to account for the time delay between tumor cell creation and clinical detection.

In recent years, however, it has become clear that many factors can change the course of carcinogenesis even after a tumor cell has been created. Because many of these factors are either unpredictable or conditional in their influence, the current practice of bridging theoretical predictions of tumor cell creation to final clinical incidence through a simple lag time appears over-simplistic in that it overlooks important progression-level biology controlling damage propagation in complex systems including carcinogenesis. Black and Welch [13] support this possibility in a study that was originally intended to explore how advances in diagnostic imaging have been able to reduce the size at which internal abnormalities can be detected. These authors examined histological sections of tissues taken at autopsy from middle-aged people who died of non-cancer causes. The findings were startling—for a wide range of cancers, actual incidence far exceeded lifetime clinical (epidemiologic) incidence. They found that 35% of women aged 40–50 had in situ cancers of the breast, even though only 12% of women are ever diagnosed with the disease [14]. Similar disparities were found for prostate cancer in men. For thyroid cancer, observed differentials were even more extreme: While less than 1% of the population will ever be diagnosed with thyroid cancer, 99.9% of individuals autopsied were found to have thyroid cancer lesions. In the laboratory, demonstrations that stroma may play a permissive role in cancer development augment our expanding understanding of the angiogenesis dependence of tumor growth. This and other evidence suggests that carcinogenesis is actually a multi-scale process, and that a systems-level treatment of the problem that links molecular, cell–cell, and inter-tissue contributions will be required to explain cancer occurrence.

The two-stage logistic (TSL) model [15] takes a step in this direction by adapting the molecular-level initiation–promotion–transformation timeline paradigm for tumor cell creation to include the natural Gompertz-like growth limitation (Fig. 2) expected for all cell populations, including the initiated cell population being modeled here. The TSL model is otherwise a deterministic variant of the stochastic two-stage clonal expansion (TSCE) model commonly used today [16–18]. The TSL model is able to explain the essential features of important data sets for cancer incidence, e.g., for atomic bomb survivors [15]. By interfacing molecular-level limitations exhibited by pre-initiated cells and initiated cells with the population-level limitation of proliferation slowdown, the TSL model may be capturing important multi-scale effects pertinent to early tumor formation.

The TSL model actually describes how the process of carcinogenesis advances to cancer cell creation once the sizes of the normal and initiated cell populations are known. In this sense, it is describing baseline carcinogenesis. Irradiation is handled as a local perturbation to this baseline. In addition to initiating new cells, radiation kills cells, the consequence of which, is a short-lived deviation from the quasi equilibrium that forms between the normal and initiated cell compartments. Depending on how the initiated population $m(t)$ and the normal population $N(t)$ re-grow to re-establish their original quasi-equilibrium during the short recovery period Δt following irradiation, the value of $m(t + \Delta t)$ will be altered over what it otherwise would have been. The amount of this alteration could depend on the differential effects of radiation on the two cell populations as follows [19]: (a) the difference between the probabilities that an initiated cell and a normal cell will survive the dose, $\Delta 1$; (b) the factor increase in initiated cells expected due to radiation action on normal cells, $\Delta 2$;

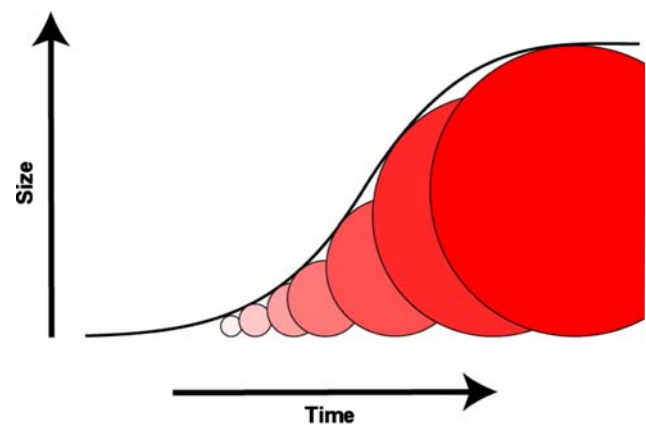


Fig. 2 The general shape of the Gompertz ($dV/dt = -V \log(V/K)$) and related Logistic ($dV/dt = \beta(V(1 - (V/K)^2)$) functions. These capture a pervasive property of cell populations—the tendency for them to grow more slowly as their size increases

and (c) the extent to which the initiated population will proportionally increase in size within the normal tissue during the rebound period due to any re-growth advantage it has over the normal population while the two populations are out of equilibrium, $\Delta 3$. If an acute radiation dose is delivered at time t_0 , then $m(t_0 + \Delta t)$ will be increased due to the radiation by a factor $(1 + \Delta 1 + \Delta 2 + \Delta 3)$.

The TSL model, then, describes the baseline, equilibrium-phase advancement in $m(t)$ over time, as adjusted by any prior radiation insults. In brief, it is assumed that the large normal cell population of constant equilibrium size N produces initiated cells at a slow, constant rate of $v = \omega N$ cells per unit time, and that initiated cells transform (become cancerous) at a slow rate of μ per cell per unit time. Initiated cells proliferate at a rate λ per unit time. The incorporation of cell–cell effects into this otherwise molecular model is through the imposition of a Gompertz-like growth slowdown that approaches a theoretical limiting population size K . This is accomplished by a reduction in proliferation by the factor $(1 - m/K)$. The equation governing $m(t)$, then, becomes

$$dm/dt = (v - \mu m) + \lambda m(1 - m/K). \tag{1}$$

The hazard function (risk), $H(t)$, for occurrence of a cancer cell is just the rate of exit of cells from the initiated cell compartment due to transformation. This is $\mu m(t)$. The equation obeyed by $H(t)$, then, is:

$$dH/dt = \mu v + (\lambda - \mu)H - \lambda H^2/\mu K. \tag{2}$$

As mentioned, the TSL model, like the TSCE model, can explain excess relative risk data for the atomic bomb survivors. The excess relative risk, $ERR(t)$, due to a radiation dose delivered at time $t = 0$ is defined as the fractional increase in the baseline hazard at time t resulting from that dose. Because $H(t)$ is a monotonically increasing function of time, the post-radiation hazard at time $t = 0$ equates to the pre-irradiation hazard evaluated at some time $t_0 > 0$. Further, because Eq. 2 does not depend explicitly on t , knowing H at a particular time determines the value of H for all later times. The excess relative risk due to any prior dose experience (assumed to be complete by $t = \tau$) may thus be written as

$$ERR(t) = H(t_0 + t)/H(t) - 1 \text{ for some } t_0 > 0, \tag{3}$$

for all $t > \tau$, where $\tau > 0$, $H(0) = 0$

Because $d/dt[H'(t)/H(t)] < 0$, it can be shown that $ERR(t)$ is a decreasing function of t . Its generally downward sloping behavior accords with the data obtained on the atomic bomb survivors. This turns out to be true whether or not the major effect of radiation is to kill fewer initiated than normal cells or initiate new cells [a factor $\Delta 1$ or $\Delta 2$ contribution to $m(t)$] or to promote existing initiated cells [a factor $\Delta 3$ contribution to $m(t)$]. In the TSCE model, the

effect of radiation is more complicated. If radiation acts more to initiate new cells, it can be shown that the shape of the ERR curve generated would not accord with the data. As yet, the effect of radiation in the timeline paradigm is uncertain. As more is known, however, such comparisons among alternative theories should prove invaluable for sorting out the basic mechanistic processes underlying carcinogenesis.

Inter-tissue effects in carcinogenesis

As the nascent tumor cell begins to expand in the progression phase, new levels of effect on carcinogenesis come into play. At the inter-tissue scale, these include tumor/stromal interactions and tumor/endothelial interactions. Either can determine whether a cancer will rise to clinical detection. It has been shown, for example, that fibroblasts in the adjacent stroma can directly modulate tumor evolution [20–22], and that the acquisition of angiogenic capacity is essential for the nourishment of tumors too large to be accommodated by simple diffusion of nutrients into its interior. Until the tumor overcomes issues of insufficient blood supply and substrate diffusion, its growth will again be asymptotically limited. The dormant tumors observed in the Black and Welch study [13] may fall into this category. Once the tumor acquires invasive properties and becomes angiogenic, it can escape this bottleneck to expand once again.

Interestingly, it has been shown that a tumor can produce both stimulators and inhibitors of angiogenesis, suggesting that tumor growth control may be a vestige of a broader tissue mass control process, and that vessel formation may be the key to this control. The finding that the inhibitor half-lives tend to be long while the stimulator half-lives tend to be short bolstered this hypothesis, because it assures that an inhibitor would always overtake a stimulator as the tumor grows. This in turn assures, as would be expected in controlled tissue growth, that a theoretical limit to tumor size always exists, but is adjustable by micro-environmental conditions.

We explored this possibility by modeling tumor growth $V(t)$ in terms of a Gompertz [23] formulation where growth slowdown to a plateau value is expected according to

$$V' = -\beta V \log \left[\frac{V}{V_{\max}} \right] \tag{4}$$

with β being the rate of incremental diminution in tumor growth, and where the plateau value V_{\max} may be variable and reflect the induction of supporting microvasculature by the tumor [24]. To calculate what the contributions due to angiogenesis stimulation and inhibition should look like, solutions to the diffusion–consumption equation, Eq. 5,

were considered in the limiting cases where stimulator clearance is instantaneous while inhibitor clearance is zero:

$$D^2 \nabla^2 n - cn + s = 0, \quad (5)$$

where n is the concentration of stimulator/inhibitor, D is the diffusion rate, c is the clearance rate, and s is the source term (s_0 inside tumor, 0 outside tumor).

Assuming radial symmetry, the equation was solved for the two limiting cases. For stimulator concentrations inside and outside the tumor (Fig. 3), one obtains

$$n_{\text{stimulator, inside}}(r) \approx \frac{s_0}{c}; \quad n_{\text{stimulator, outside}}(r) \approx 0, \quad (6a)$$

and for inhibitor concentrations inside and outside the tumor (r_0 is the tumor radius),

$$n_{\text{inhibitor, inside}}(r) \approx \frac{s_0}{6D^2} (3r_0^2 - r^2); \quad (6b)$$

$$n_{\text{inhibitor, outside}}(r) \approx \frac{s_0 r_0^3}{3D^2 r}.$$

The concentration of stimulator at the tumor periphery thus has a zero-power dependence on r_0 (i.e., is constant irrespective of r_0) whereas the concentration of inhibitor at the tumor periphery has a squared-order dependence on r_0 . Translating these results dimensionally into terms of V or V_{max} and observing the predicted curves, the following general form for V_{max} can be deduced:

$$(\partial V_{\text{max}} / \partial t) / V_{\text{max}} = aV / V_{\text{max}} - bV^{2/3} \quad (7)$$

By including an extra term $-e^*g(t)$ (where $g(t) = \int_0^t r(t') \exp(-c(t-t')) dt'$ and $r(t)$ is the rate of inhibitor injection per unit time) on the right hand side of Eq. 7 to account for exogenous administration of inhibitors, it is

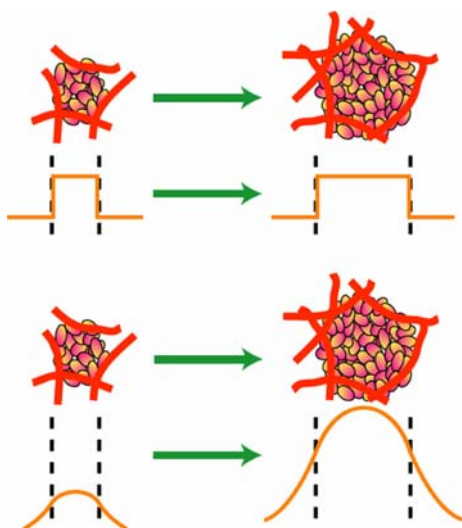


Fig. 3 Concentration of stimulator and inhibitor across the tumor volume. The concentrations within and outside the tumor obey Eqs. 6a (upper panel) and 6b (lower panel)

possible to test this deduction in tumor-bearing animals that were administered the angiogenesis inhibitors angiostatin, endostatin, and TNP-470 [24]. The results confirmed this general form and supported the central principle that tumors, and perhaps organs, have a theoretical set point to growth that can be adjusted according to the angiogenic state of the micro-environment. This, in turn, lends mechanistic support to the notion that dormant tumors may be commonplace, awaiting a pro-angiogenic stimulus, perhaps such as ionizing radiation, to trigger their advance towards clinical presentation. This is a vital augment to existing theories, which have equated tumor cell creation with clinical disease.

Radiation damage propagation in complex and adaptive biological systems

Many carcinogenesis models implicitly assume a smooth, monotonic relationship between carcinogen exposure and the risk of the ultimate formation of tumors. However, this assumption is not necessarily true in radiation biology, especially for very low doses, which some studies actually determined as protective. More generally, research on cells and animals over the past few decades has increasingly been providing evidence that low doses of ionizing radiation initiate biological responses that were totally unpredictable from responses per unit dose at higher exposure levels [25–27]. In fact, entirely new phenomena have been uncovered such as low-dose induced, delayed appearing and mostly temporarily lasting cellular signaling changes affecting intracellular enzyme activities, reactions to reactive oxygen species, DNA synthesis and repair, apoptosis, cell differentiation, and immune competence [28]. These adaptive responses occur in conjunction with altered gene expression by up- and down-regulation of such genes that respond only to low doses. Many of these genes also respond to metabolic stress such as from elevated levels of reactive oxygen species (ROS). The low-dose specific cell responses are to be seen in the context of other newly recognized phenomena which, however, may also arise after high dose irradiation to individual cells and become observable again with a delay of hours and may have long lasting effects in the affected cells. The category of such responses comprises both the so-called bystander effects [29], as well as genomic instability [30] and epigenetic effects, which may befall cellular progeny over many cell generations [31].

Perturbations of homeostasis by ionizing radiation and protective barriers

Low-level ionizing radiation nearly always affects multiple sites of biological systems at the molecular level, by interacting with atoms and causing energy deposition along

particle tracks [32]. These events are more or less stochastically distributed throughout the irradiated tissue and may affect genetic, structural, functional, and signaling components of the cell. They may damage molecules directly on site or secondarily, e.g., by means of ROS from radiogenic hydrolysis [33]. Severe damage to signaling components may become especially disastrous, as they are usually involved in amplification cascades and control cell communication, thereby bridging between cells, tissues, organs, and the functioning of the entire organism. The indirect damages to non-irradiated cells are referred to as bystander effects [29]. These and other types of effects may ultimately cause homeostatic perturbations anywhere in the system and, if strong enough or amplified to affect larger cell communities, as is common with higher values of absorbed dose, may become observable at higher organizational levels such as tissues and organs [34]. The probability of higher-level effects depends on the type, quality, and extent of initial homeostatic perturbation at the molecular level and on the tolerance by the body's various homeostatic control mechanisms [35].

Healthy organisms command an array of physiological barriers, such as the skin, that protect against immediate and late consequences of potentially life threatening exogenous impacts. Operating at different levels of organization, the barriers actually form a sequence of protections that prevent the propagation of damage into clinical disease. They may be classified as: (a) defenses by scavenging mechanisms at the atomic-molecular level; (b) molecular repair, especially of DNA, with reconstitution of essential cell constituents and functions; and (c) removal of damaged cells from tissue either by signal induced cell death, i.e., apoptosis, by cell differentiation, or by a stimulated immune response that is often associated with a replacement of lost cells [28, 36–39].

Immediate responses to low and high doses of ionizing radiation are quite well understood with respect to primary DNA damage [33]. Within minutes after irradiation, DNA double strand breaks can be observed with immuno-histochemical methods [40]. These are broadly distributed in their numbers per cell, reflecting again the stochasticity of the process. Usually well within 24 h, the fluorescent foci indicative of double strand breaks can be reduced to a low number that is close to that of a “spontaneous”, i.e., non-radiogenic, incidence, and homeostasis is restored in the affected cell if the initial structural damage is non-lethal. It is known that DNA repair is species specific and in mammals involves more than 150 genes (Winters, T.A., personal communication).

One should note that the various barriers may be physical, preventing a potentially damaging, yet relatively mild impact from disrupting a biological structure with a concomitant disturbance of homeostasis, or biochemical–

cellular, responding to a sudden disturbance of homeostasis with rapid signaling for reconstitution of structure and function. Both types of protections are known to operate non-linearly with the degree of impact and, thus, individually exhibit a threshold-like behavior on the cellular level [41]. Since increasing doses of ionizing radiation eventually paralyze barriers at all levels, higher doses may allow damage at basic levels of organization to propagate with minimal or no inhibition to evolve into clinical disease. As a very important consequence, dose–response functions then expectedly tend to be linear at higher doses, but not so at low doses. Because all physiological barriers are under some type of genetic control, certain defects in the involved genes, which control these physiological barriers, may change individual radiation sensitivity drastically [42].

The physiological barriers also operate against the development of some clinical cancer. An illustrative example is the relationship between the extent of DNA damage caused in hemopoietic stem cells by radiation and the probability of leukemia induction in the exposed individual. Experimental and epidemiological observations suggest that, at high doses, the ratio of radiation-induced DNA double strand breaks, including those of the multi-damage site-type and lethal leukemia, has been estimated to be close to 10^{12} [36]. The claim that even a single DNA double strand break, however grave, in a potentially carcinogenic stem cell may cause cancer thus must be considered scientifically unjustified. In addition, extrapolation to low doses and low dose rates implicitly assumes that the biological targets respond at proportionally less but in qualitatively the same way. Today, we know definitely that this is not the case (see [43] and later discussion).

In addition to the immediate physiological barriers, the body exhibits various adaptive protections especially in instances when homeostatic perturbations are below the level of barrier destruction, i.e., in stress situations. Adaptive responses are well known, for instance, against oxidative stress [44, 45]. Similarly, low-level exposure to ionizing radiation does change cellular signaling through temporary modifications in enzyme and hormone activities that are involved in protecting against ROS and in DNA synthesis [46, 47], DNA repair [48, 49], and damage removal by various routes [28, 36–39, 50–52]. These responses are currently understood to be the consequences of delayed and temporary up-regulation of existing physiological barriers at increasingly higher levels of biological organization and generally begin to operate within a few hours after exposure and may last from hours to months or even longer. In the context of radiation, delayed up- or down-regulation of the physiological barriers may be observed at very low doses in the range of mGy and on average show a maximum effectiveness at a dose of about 100 mGy. They disappear as doses increase beyond

200 mGy and are hardly or not seen anymore beyond about 500 mGy [28, 53, 54]. By contrast, the probability of apoptosis induction apparently increases linearly beyond 500 mGy over a certain dose region. Figure 4 presents a schematic summary presentation of average dose response values from all available observations of low-dose induced up-regulations of various barriers (for reviews see [26, 28, 36–39]).

A number of genes respond to both low- and high-level irradiation. However, low-level radiation exposure also modulates a set of genes, mostly involved in stress responses that do not respond to high-level exposure, and vice versa [43, 55]. Out of 10,500 genes analyzed in human keratinocytes, 853 genes were modulated between 3 and 74 h after irradiation, and of these, only 214 (mainly stress response genes) appeared to change in expression after 10 mGy irradiation. A high dose of 2 Gy modulated 639 genes specifically; both doses changed expression in 269 genes.

It is worthwhile to put low-level radiation damage into perspective of normal cell metabolism. For instance, in a normal cell, about 10^9 ROS per second arise in the cytoplasm by leakage from mitochondria, and from metabolic reactions; ROS bursts may occur from various responses to external cell signaling [56, 57]. When an average electron, for instance produced by 100 kV X-rays, hits a cell, about 150 ROS occur in that cell within a fraction of a millisecond. In addition to metabolically produced ROS bursts, once to several times a year, supra-basal bursts of ROS occur within that cell from normal background radiation and may trigger reactions commonly addressed as cellular oxidative stress responses that invoke cell signaling of many kinds with potential consequences of cellular damage

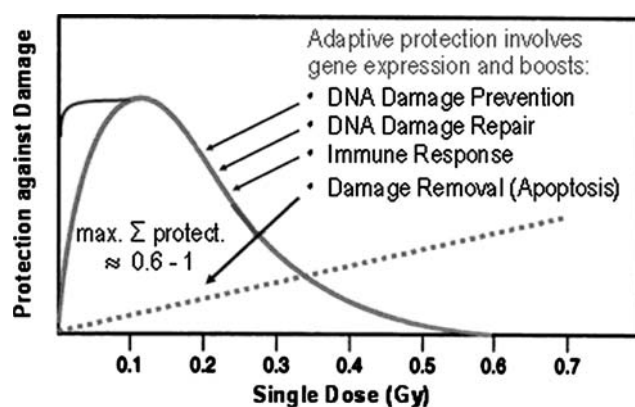


Fig. 4 Schematic presentation of adaptive responses to radiation exposure as a function of dose. These may be observed at very low doses in the range of mGy and show, on average, a maximum effectiveness at a dose of about 0.1 Gy. They tend to disappear as doses increase beyond 0.2 Gy and are hardly seen anymore beyond about 0.5 Gy. However, the probability of apoptosis apparently increases linearly beyond 0.5 Gy over a certain dose region

as well as benefit, in terms of the up-regulation of defense mechanisms, repair of DNA and cell structures, as well as changes of cell cycle times, induction of apoptosis and immunogenic alterations leading, for instance, to immune stimulation [45, 58].

Low-level radiation exposure and cancer risk assessment

To assess cancer risk from low-level exposure in a coherent quantitative fashion, a model is needed; various options are available [36–39, 59–63]. We will focus here on an approach that was introduced in 1995 [36] but has become more sophisticated over the years. The basis of the approach is a distinction between the risk R_1 of introducing damage at the DNA level and the risk R_2 of propagating a primary damage to successive higher levels of biological organization against protective barriers. R_1 is a stochastic quantity and addresses both biophysical and primary biochemical mechanisms at intracellular target sites, such as the DNA–histone complexes. The incidence of a relevant radiogenic damage rises proportionally to dose over a certain dose range. With no or constant-level protective mechanisms, the extent of initial damage could possibly linearly determine the degree or extent of a final clinical outcome. This is, indeed, a prevailing assumption among many cancer risk assessors and, in particular, among epidemiologists. R_2 is less direct and more complicated to assess, because it comprises the sum of effects of physiological barriers as they are up-regulated mainly by low doses in terms of adaptive responses involving cells and signaling systems at all organizational levels, as outlined above. R_2 , thus, includes predominantly non-linear reactions, which are largely controlled by complex schemes of gene expression. Whereas animal and cell experimental data conform to this model, there is a lack of unequivocal epidemiological data for model validation.

R_1 is expressed here by the risk coefficient P_{ind} , the probability of radiation-induced DNA damage per unit dose D of radiation, which would develop later to clinical cancer assuming no or some constant rate of protection irrespective of the value of D . P_{ind} is assumed to be constant over a certain range of D and thus conforms to the conventional proportionality constant α in the well known expression of the linear non-threshold dose-risk function: $R = \alpha D$. Hence, $R_1 = P_{\text{ind}} D$.

R_2 is expressed here by the product $P_{\text{prot}} \times f(D, t_p)$, which represents the fractional cumulative probability of adaptive protection; it describes the inhibition of damage propagation, as a function of the variable D and of the parameter “duration of protection effectiveness t_p ”. R_2 is an empirically obtained probability that does not unravel

individual quantitative contributions of mechanistic components responsible for the protective phenomena in the whole system, as they are observed experimentally. The dose response function for $P_{\text{prot}} \times f(D, t_p)$ is illustrated in Fig. 4. A value of 0 means no adaptive protection, and a value of 1 means full protection of damage propagation by sufficient adaptive responses in the system. Protection in this model operates against both the probability P_{spo} of the appearance of spontaneous cancer, derived, for example, from non-radiogenic DNA damage at the times of observation, as well as against the radiation-induced cancer expressed by the term $P_{\text{ind}} \times D$ to indicate that $P_{\text{prot}} \times f(D, t_p)$ also affects $P_{\text{ind}}D$. Thus, the value of $P_{\text{prot}} \times f(D, t_p) \times (P_{\text{spo}} + P_{\text{ind}}D)$ summarily encompasses the probabilities of total low-dose induced adaptive protection against propagation of any DNA damage to cause cancer (Fig. 5). The net risk of radiation induced cancer R_n , in this model, thus, amounts to $R_n = P_{\text{ind}}D - P_{\text{prot}} \times f(D, t_p) \times (P_{\text{spo}} + P_{\text{ind}}D)$. This simplifies an approach published elsewhere [36, 38, 39, 64]. Note that, with increasing D , the term $P_{\text{prot}} \times f(D, t_p) \times (P_{\text{spo}} + P_{\text{ind}}D)$ tends towards zero, notwithstanding the protective contribution by apoptosis (Fig. 4). Moreover, the term also reaches zero with t_p becoming too short for P_{prot} to operate, for instance, when adaptive protection is not allowed to develop. The difference between the two dose–risk functions, $R_1 - R_2$, yields the net dose–risk function R (Fig. 5, solid line in between), which conforms, at least qualitatively, to a large set of experimental and epidemiological data ([26, 27, 65–68]; for a review see [68]).

The discussion of risks would be incomplete without considering repeated or chronic low dose rate exposures. Dose rates may be described in terms of mean time

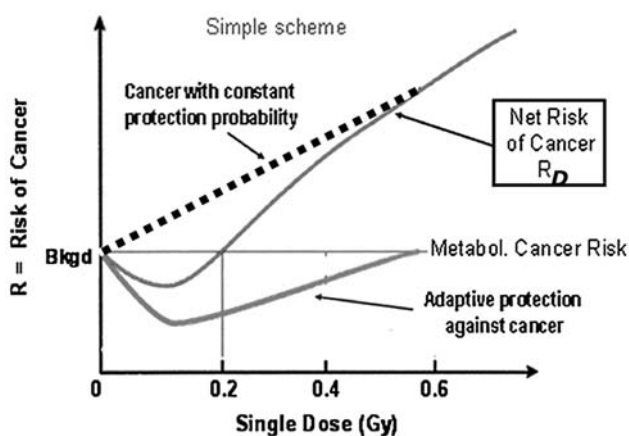


Fig. 5 Schematic representation of effects of low-dose radiation. Straight line linear dependency of radiation-induced DNA damage R_1 on dose D : $R_1 = P_{\text{ind}}D$. Lower curved line dose–risk function $R_2 = P_{\text{prot}}f(D, t_p) (P_{\text{spo}} + P_{\text{ind}}D)$, describing inhibition of damage propagation by adaptive protection. Solid line in between difference, $R_1 - R_2$, corresponds to the net dose–risk function R

intervals t_x between consecutive energy deposition events in the exposed micro-mass, for a given radiation quality [69–71]. An example is the chronic exposure of mice to tritiated water throughout life [72]. Thymic lymphoma induction and life shortening only appeared at dose rates above 1 mGy per day. This corresponds to about one event of 1 mGy occurring per micro-mass within less than 1 day, i.e., at a t_x shorter than 1 day [28]. There was no thymic lymphoma induced and no life shortening observed when t_x was longer than 1 day. This assessment is in line with other observations after chronic low dose rate exposures [56, 61, 68]. Occupational exposures in humans usually deliver much lower dose rates than discussed above and thus provide for relatively long t_x . For long t_x , immediate and adaptive protections are expected to operate within the cell’s capacities [68] and are likely the reason for repeated epidemiological observations of reduced cancer incidence below the background incidence at chronic low dose rate exposures [25–27]. The analysis of the mortality of 45,468 Canadian nuclear power industry workers after chronic low-dose exposure to ionizing radiation [73] quotes: “For all solid cancers combined, the categorical analysis shows a significant reduction in risk in the 1–49 mSv category compared to the lowest category (<1 mSv) with a relative risk of 0.699 (95% CI: 0.548, 0.892). Above 100 mSv the risk appeared to increase.”

Network approaches to radiation biology

The previous section has made it clear that many factors govern the processes that begin with radiation exposure and ultimately lead or do not lead to the development of cancer. They include “external” aspects, such as the quality, dose, and timing of radiation, as well as “internal” aspects like the efficacy of various protective barriers and adaptive responses. While it is possible to rationalize intuitively the behaviors of linear chains of causes and effects, intuition often fails us when we try to predict the response of more complex or even adaptive systems. For instance, if counteracting processes are present, with one tending to cause disease and the other one triggering protective responses, as is the case with the process of radiation exposure, our unaided mind is no longer able to integrate all individual responses quantitatively and to predict the ultimate outcome at the systems level with a sufficient degree of reliability. The same difficulties emerge in the ubiquitous situation where defined diverging and converging branches dominate a multi-faceted system and where individual different system components simultaneously contribute to the dynamics of the system with different magnitudes and possibly different signs. It is therefore necessary to work toward mathematical

descriptions that permit rigorous quantification as well as the scaling to large networks and systems. This section alludes to some current issues related to network analysis in biology. It is exemplified with a discussion of how we have moved in recent years from simple gene–disease associations to the recognition that complex traits, such as diabetes or cancer, can only be explained through considerations of genomic networks and beyond.

Genomic systems analysis of complex traits

With several maturing technologies that enable low-cost, high-throughput characterization of DNA variations that are either naturally occurring or induced by exposure to radiation or other exogenous agents, we are entering a new era in which DNA variations that lead to phenotypic variation will be identified on unprecedented scales. In fact, a number of genome-wide association studies have already leveraged the availability of high-throughput genotyping technologies to identify polymorphisms in genes that associate with diseases like age-related macular degeneration [74–76], diabetes [77, 78], obesity [79], and cancer [80], to name just a few. However, while variations in DNA that associate with complex phenotypes like disease provide a peek into pathways that underlie these phenotypes, such associations are usually devoid of biological context, so that elucidating the functional role such genes play in disease can linger for years, or even decades, as has been the case for ApoE, an Alzheimer's susceptibility gene identified nearly 15 years ago [81].

The information that defines how variations in DNA induced, say, by radiation exposure influence complex physiological processes, flows through transcriptional and other molecular, cellular, tissue, and organismic networks. In the past it was not possible to comprehensively monitor these types of intermediate phenotypes that comprise the hierarchy of networks that drive complex phenotypes like disease. However, today's DNA microarrays have radically changed the way we study genes, enabling a more comprehensive look at the role they play in everything from the regulation of normal cellular processes to complex diseases like obesity, diabetes, and cancer. In their typical use, microarrays allow researchers to screen thousands of genes for differences in expression or differences in how genes are connected in molecular networks [82] under altered experimental conditions of interest. The data produced in this fashion are often used to discover genes and, more generally, characteristics of networks that differ between normal and disease-associated tissues, to model and predict continuous or binary measures, to predict patient survival, and to classify disease or tumor sub-types. Because gene expression levels in any given sample are measured

simultaneously, researchers are able to identify genes whose expression levels at the time of measurement are correlated, implying possibly coordinated regulation under specific conditions.

Integrating DNA variation and functional genomic data can provide a path to inferring causal associations between genes, their properties, and disease. In the past, causal associations between genes and traits have been investigated using time series experiments, gene knockouts or transgenics, and RNAi-based knockdown, viral-mediated over-expression, or chemical activation/inhibition of genes of interest. A more systematic and arguably relevant source of perturbation to make such inferences regarding genes and disease are naturally occurring or environmentally-induced DNA polymorphisms, where gene expression and other molecular phenotypes in a number of species have been shown to be significantly heritable and at least partially under the control of specific genetic loci [83–92]. By examining the effects that naturally occurring or artificially induced variations in DNA have on variations in gene expression traits in human or experimental populations, higher-order phenotypes (including diseases) can be examined with respect to these same DNA variations and ultimately ordered with respect to genes to infer causal control [93–96]. The power of this integrative genomics strategy rests in the molecular processes that transcribe DNA into RNA and then RNA into protein, so that information on how variations in DNA impact complex physiological processes often flow directly through transcriptional networks. As a result, integrating DNA variation, transcription, and phenotypic data has the potential to enhance identification of the associations between DNA variation and complex phenotypes like disease, as well as characterize those parts of the molecular networks that drive disease.

Gene transcripts have been identified that are associated with complex disease phenotypes [91, 97], are alternatively spliced [98], elucidate novel gene structures [99–101], can serve as biomarkers of disease or drug response [102], lead to the identification of disease subtypes [91, 103, 104], and elucidate mechanisms of drug toxicity [105]. Changes in gene expression often reflect changes in a gene's activity and the impact a gene has on different phenotypes. Because gene expression is a quantitative trait, association methods can be directly applied to such traits to identify genetic loci that control them, or in the case of radiation-induced DNA changes, the transcriptional responses associated with such changes can be assessed. In turn, variations in DNA that control for expression traits may also associate with higher-order phenotypes affected by expression changes in corresponding genes of interest, providing a path to directly identify genes controlling for phenotypes of interest. In the context of naturally occurring DNA variations in a

population of interest, identifying the heritable traits and the extent of their genetic variability might provide insight about the evolutionary forces contributing to the changes in expression that associate with biological processes that underlie diseases like cancer and diabetes, beyond what can be gained by looking at the transcript abundance data alone.

It is now well established that gene expression is a significantly heritable trait [83, 88, 89, 91, 96, 106–111]. If a gene expression trait is highly correlated with a disease trait of interest, and if the corresponding gene physically resides in a region of the genome that is associated with a complex phenotype, then knowing that the expression trait is also genetically linked to a region coincident with its physical location provides an objective and direct path toward identifying candidate causal genes for the complex phenotype [83, 88, 89, 91, 96, 106–111]. The DNA variation information therefore enables the dissection of the covariance structure for two traits of interest into genetic and non-genetic components, where the genetic component can then be leveraged to support whether an expression and disease trait are related in a causal, reactive, or independent manner (with respect to the expression trait). Elucidating causal relationships in this way is possible given the flow of information from changes in DNA to changes in RNA and protein function. That is, given that two traits are controlled by the same DNA locus, there are just three basic ways in which these two traits can be related with respect to a given locus: (1) the two traits are independently impacted by the common DNA locus, (2) the first trait is more immediately impacted by the DNA locus and in turn affects the second trait, or (3) the second trait is more immediately impacted by the DNA locus and in turn affects the first trait [96, 112, 113].

Expression traits detected as significantly correlated with a higher-order complex phenotype (e.g., disease) may reflect a causal relationship between the expression trait and phenotype, either because the expression trait contributes to, or is causal for, the complex phenotype, or because the expression trait is reactive to, or a marker of, the clinical phenotype. However, correlation may also exist in cases when the two traits are not causally associated. Two traits may appear correlated due to confounding factors such as tight linkage of causal mutations in DNA [96] or may arise independently from a common genetic source. The A^y mouse provides an example of correlations between eumelanin RNA levels and obesity phenotypes induced by an allele that acts independently on these different traits, causing both decreased levels of eumelanin RNA and an obesity phenotype. More generally, a clinical and expression trait for a particular gene may depend on the activity of a second gene, in such a manner that conditional on the second gene, the clinical and expression traits are independent.

Correlation data alone cannot indicate which of the possible relationships between gene expression traits and a complex phenotype are true. For example, given two expression traits and a complex phenotype detected as correlated in a population of interest, there are 112 ways to order the traits with respect to one another. To see this, consider traits as nodes in a network, in which case there are five possible ways that the two nodes can be connected: (1) connected by an undirected edge, (2) connected by a directed edge moving left-to-right, (3) connected by a directed edge moving right-to-left, (4) connected by a directed edge moving right-to-left and a directed edge moving left-to-right, (5) not connected by an edge. Since there are three pairs of nodes, there are $5 \times 5 \times 5 = 125$ possible graphs. However, since we start with the assumption that the traits are all correlated with one another, we exclude 12 of the 125 possible graphs in which one node is not connected to either of the other two nodes, in addition to excluding the graph in which none of the nodes are connected, leaving us with 112 possible graphs. This stands in contrast to the case indicated above where variations in DNA affecting two phenotypes of interest are leveraged as a causal anchor, reducing the number of graphs to consider to three.

Large-scale networks

The reductionist view of traditional biology has motivated the identification of single genes associated with disease as a means of initial exploration into disease pathways. However, even in cases where genes are involved in pathways that are well known, it is unclear whether the gene causes disease via the known pathway or whether the gene is involved in other pathways or more complex networks that lead to disease. This is the case with TGFBR2, a recently identified and validated obesity susceptibility gene [96]. While TGFBR2 plays a central role in the well-studied TGF- β signaling pathway, TGFBR2 and other genes in this signaling pathway are correlated with hundreds of other genes [96, 114], so that it is possible that perturbations in one or more of these other genes, or in TGFBR2 itself, may drive diseases like obesity by influencing other parts of the network beyond the TGF- β signaling pathway. Because this type of complexity appears to be the rule rather than the exception, it is beneficial—if not mandatory—to consider single genes in the context of larger gene regulatory networks. Indeed, for complex phenomena like radiation damage, such a network view may be a necessary prerequisite for establishing the context within which to interpret the role of a candidate gene.

Mathematical network models provide a convenient framework for characterizing the multiple roles of genes

and proteins, and an enormous effort has been devoted in recent times to generate biological data elucidating networks and providing parameters for network modeling. Examples include molecular interaction networks, gene transcription and regulation networks, as well as stoichiometric and fully regulated metabolic pathway systems. In the simplest case, networks are represented as graphs comprised of nodes and edges. For gene regulatory networks, the nodes typically represent genes and the edges (links) represent some relationship between the two associated genes. For example, an edge may indicate that the corresponding expression traits are correlated in a given population of interest [115], that the corresponding proteins interact [116], or that changes in the activity of one gene lead to changes in the activity of the other gene [96].

Three issues have to be kept in mind when setting up biological network experiments and choosing mathematical network representations: the potential uses of network data toward understanding complex biological systems and the insights they may convey; the statistical and bioinformatics techniques that are available—or still need to be developed—for dealing with such data; and the limitations of data and representations, along with the scope for extending and refining them in the future. As an example, it has been suggested that it is the importance of interactions among genes that is crucial for understanding most complex phenotypes, given that these are emergent properties of complex networks. Ignoring the details of connectivity of interactions may make it difficult, if not impossible, to quantify the functional relevance of genes identified for disease in case–control association studies. However, characterizing the significance of such interactions is associated with an enormous multiple-testing problem, and it will be necessary to develop efficient statistical techniques for this purpose. These techniques should assess the importance of potential interactions within a gene network and will help with the development of more realistic models for the link between genotype and phenotype.

One should also recognize that as of yet, most functional analyses of phenotypes in the context of biological networks (mostly metabolic, transcriptional and protein–interaction networks) have focused primarily on simple model organisms (such as *E. coli* or *Saccharomyces cerevisiae*) rather than mammalian systems that are better models for radiation damage and its consequences. Moreover, interventions and perturbations have predominantly relied on relatively coarse actions, such as the knock-out or knock-down (using RNAi) of individual genes or pairs of genes. Nevertheless, these rudimentary methods have taught us a lot and will without doubt be expanded and refined to a degree that they can be applied to higher organisms and more complex traits.

Another significant issue concerning biological networks is the implicit assumption that network data are essentially correct. However, it must be acknowledged that, at present, network datasets are often incomplete and noisy. In particular, the use of protein–protein interaction data, with false-positive and false-negative rates estimated to be in the range of 30–70% and 50–90%, respectively, may be a limiting factor of what we can reliably infer for the underlying system, if we focus only on this single dimension. In other words, most of the inferred interactions we see are probably not there in actuality, while most interactions that are known to exist are not picked up by high-throughput assays, such as yeast 2-hybrid or TAP-tagging approaches. While this is a severe problem, methodologies are quickly improving and may soon lead to crisper and more reliable information. As important as noise and uncertainties is the incompleteness of network datasets. Missing interactions can lead to wrong interpretations, because a subnet drawn from some larger network can have very different properties than the true network (see Fig. 6). In order to make meaningful inferences from network data about the biological system, it is therefore important to include the effects of noise and incompleteness into statistical analyses from the outset.

Finally, one must acknowledge that different networks may be functionally coupled. Specifically, proteins as well as genes form their own networks, but transcriptional networks, influence the expression of proteins, which as enzymes, in turn, affect metabolite concentrations or act as transcription factors that regulate gene expression. Thus, bioinformatics tools will be needed to assess several levels simultaneously and to account for the fact that proteins or genes are embedded in interacting networks and that we may strictly not be able to study the effects of genes on their own, because the network induces dependencies among interacting biomolecules that are disrupted in isolation experiments. Extending the complications caused by these interdependencies, we often implicitly assume that we can understand biological interactions in terms of networks that are describable with static, mathematical graphs. In reality, about all biological networks change over time and their topology and regulatory structure depend dynamically on their internal and external conditions. Including time-dependence and conditionality in the description of biological networks will presumably be one of the most challenging problems of statistical bioinformatics and computational systems biology.

In spite of challenges with respect to computational techniques as well as to the generation and curation of data, many disciplines, especially engineering, have approached systems with mathematical means, and some of these techniques are now being adapted for the analysis of biological phenomena. One of the earliest approaches goes

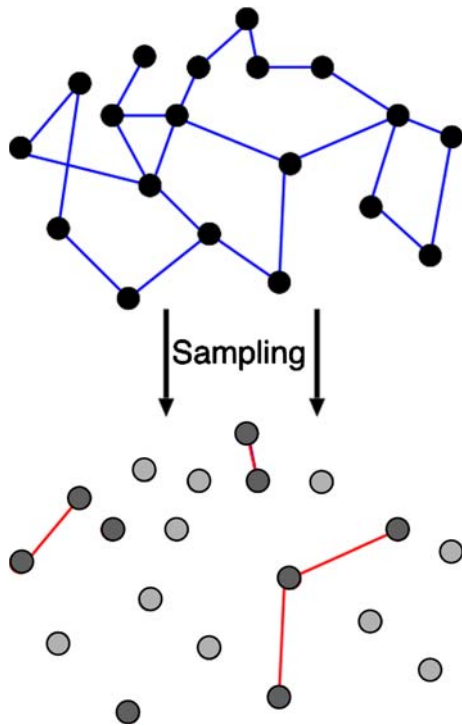


Fig. 6 A subnet (dark gray nodes in lower part) will generally have very different properties than the true network (top): interactions involving those nodes which were not included in the subnet (light gray) cannot be observed, and the incomplete nature of network data may introduce severe bias into the network analysis. In some cases, it is possible to make statistical predictions about properties of the true network from incomplete network data

back to Boltzmann, who addressed large assemblies of molecules with methods of statistical mechanics. However, this approach is better suited for unorganized systems, such as ideal gases, than highly organized biological systems, in which each component has a specific role.

Linear network models

For organizationally complex systems, efficacious analyses are still in their infancy. The great divide comes with the question of whether the system is linear or nonlinear. Linear models have very nice mathematical properties that allow sophisticated analyses even for very large systems. For instance, it has become a straightforward task (e.g., in the electrical power industry) to optimize some objective in a system of thousands of variables. In biology, linear systems models come in different varieties that call for different methodologies. Within the realm of deterministic systems models, stoichiometric network models are without doubt the most important. In these models, each component (typically a metabolite) is represented as a variable S_i , which is an element in a vector S , the connectivity between

these variables is coded as a “stoichiometric matrix” N , and each flux rate entering or leaving a metabolite pool is coded as a vector v . The formulation of a dynamic model with these components is easy, namely

$$\frac{dS}{dt} = Nv. \quad (9)$$

At steady state, the left-hand side becomes 0, because by definition no metabolite concentration changes, and the resulting equation becomes algebraic and can be analyzed with a variety of methods of linear algebra and operations research [117].

A different linear approach is based on stochastic ideas and, for instance, uses Markov chains, in which each transition between states (or nodes) is governed by a random process [118]. Stochastic Petri net models generalize on these concepts [119]. Yet another statistical approach addresses the reconstruction (or identification) of networks with Bayesian methods [120, 121]. Bayesian networks are directed acyclic graphs that while limited with respect to representing temporal information or feedback loops, allow for the explicit representation of causal associations among nodes in the network. With Bayesian network reconstruction methods taking gene expression data as the only source of input, many relationships between genes in such a setting will be Markov equivalent (symmetric), so that inferring networks that are actually predictive have not met with much success. To break this symmetry, Zhu et al. [121] incorporated expression quantitative trait (eQTL) data as prior information to more reliably establish the correct direction among expression traits.

Bayesian network methods have been applied previously to reconstruct networks comprised only of expression traits, as well as to networks comprised of both expression and disease traits, where the aim has been to identify those portions of the network that are driving a given disease trait. Forming candidate relationships among genes was carried out using an extension of standard Bayesian network reconstruction methods [122]. In the first approach to extend this method using genetic data, DNA variations found to associate with transcript abundances of each gene considered in the network were incorporated into the reconstruction process. It is well known that searching for the best possible network linking a moderately sized set of genes is an NP-hard problem (i.e., there is no known algorithm able to identify the best possible network in polynomial time). Exhaustively searching for the optimal network with hundreds of genes is presently a computationally intractable problem. Therefore, various simplifications are typically applied to reduce the size of the search space and to reduce the number of parameters that need to be estimated from the data. Two simplifying

assumptions to achieve such reductions are commonly employed. First, while any gene in a biological system could potentially control many other genes, a given gene can be restricted so that it is allowed to be controlled by a reduced set of other genes. Second, the set of genes that can be considered as possible causal drivers (parent nodes) for a given gene can be restricted using the type of causality arguments discussed in previous sections, as opposed to allowing for the possibility of any gene in the complete gene set to serve as a parent node. The expression data linked to specific genetic loci in this case can be leveraged as prior information to restrict the types of relationships that can be established among genes. As already indicated, correlation measures are symmetric and so can indicate association but not causality. However, incorporating DNA variation information and its association with gene expression traits can be used to help sort out causal relationships. The different tests described above on making causal inferences between pairs of traits provide a way to sort out such relationships explicitly. Zhu et al. [115] leveraged the DNA variation data in a related way by incorporating it as a structure prior in their network reconstruction algorithm (described below), enhancing the ability to infer causal relationships among the nodes in the network [121].

With the various constraints and measures defined above, the goal in reconstructing whole gene networks is to find a graphical model M (a gene network) that best represents the relationship between genes, given a gene expression data set, D , of interest. That is, given data D , we seek to find the model M with the highest posterior probability $P(M|D)$. The prior probability $p(M)$ of model M is

$$p(M) = \prod_{X \rightarrow Y} p(X \rightarrow Y), \quad (10)$$

where the product is taken over all paths in the network (M) under consideration.

The algorithm employed by Zhu et al. [115] to search through all possible models to find the network that best fits the data is similar to the local maximum search algorithm implemented by Friedman et al. [123]. Zhu et al. [121] recently demonstrated, via simulation of biologically realistic networks, that the integration of genetic and expression data in this fashion to reconstruct gene networks leads to networks that are more predictive than networks reconstructed from expression data alone. This approach was more recently applied to genetic and expression data generated from a segregating population of yeast. Again networks constructed by integrating genetic and expression data were shown to have superior predictive power compared to networks constructed from expression data alone, where, in this case, the predictions were prospectively validated experimentally.

Nonlinear network models

In contrast to the one and only linear structure, infinitely many nonlinear model structures are available. This creates challenging issues not only with their analysis, but also immediately at the very beginning of the modeling process, when the most suitable model structure is to be chosen. The choices fall into two categories. One may select an ad hoc model, in which each process is described with some function that seems to be most appropriate. The closer the biological phenomenon is to a process whose physics is well understood, the better the chances of success are with this method. The alternative is a canonical nonlinear model. In this case, the model structure is prescribed, and the model design phase consists of adapting parameter settings within this structure to the phenomenon under investigation. While this approach may seem limiting, it has proven to be very successful for a variety of analyses of biological systems. Two prominent canonical forms are Lotka–Volterra systems [124], which are particularly well suited for ecological systems, and models with the framework of biochemical systems theory [125, 126], which was developed for the analysis of metabolic and genetic systems. Interestingly, while both approaches have fixed mathematical structures, they are very flexible in their repertoires of responses and capable of modeling essentially any smooth nonlinearities, including stable limit cycle oscillations and chaos [124, 127].

It is beyond the scope of this article to review the many interesting features of nonlinear canonical forms, but there is rich literature on very many aspects (for reviews, see [128–132]). Suffice it to say that a biological network diagram, showing which components affect each other, can be translated almost automatically into a dynamic mathematical model [133]. Once formulated in mathematical terms, many computational methods that were tailored specifically for these types of systems can be applied to diagnose and refine the model, to interrogate the model with respect to experiments not yet executed, and to manipulate the model in a rational fashion.

The main bottleneck of all nonlinear analyses is the determination of suitable parameter values. These have traditionally been obtained from the bottom up. In other words, each step within a model is studied before in isolation and parameters are computed from focused biological experiments. In the second phase of model design, the representations of all steps are merged and the resulting dynamics of the integrated model is compared to observations and leads to suggestions for refinements. Because the model in most cases does not produce responses as observed in the biological system, a string of refinements and re-estimations has to be executed before a reasonable model is obtained. More recently, this approach

has been complemented with parameter estimation attempts from global in vivo data, especially in the form of time series [134]. These estimations are computationally more challenging, but have the potential of capturing biological reality in a shorter turn-around time. Essentially, all methods of parameter estimation, however, do not scale well and run into problems for biological systems of even moderate size. In other words, one may expect that parameter estimation will continue to be the primary hold-up of systems biology.

Conclusion

It is clear that any predictive model describing the details of radiogenic carcinogenesis at the different levels of organization will have to be nonlinear. It is also evident that we are far from being able to construct such models with current biological and computational means. On the biological side, we are still missing many detail measurements that will be needed to compose comprehensive models. On the computational side, models of the necessary size and complexity would challenge our current diagnostic and analytical tools. Nonetheless, while a true understanding of radiation exposure and carcinogenesis is not yet within reach, this article has attempted to point out some of the issues that need to be addressed. It is evident that true understanding will be based on quantitative systems models, because the human mind is simply not able to process systems of such complexity unaided. This recognition of the complexity of radiation biology is the first step toward such models. The next steps will consist of the development of partial models and accompanying techniques, as we demonstrated them here, and these must eventually be merged into larger structures that will deepen our understanding of radiogenic carcinogenesis.

Acknowledgments The authors are indebted to Dr. Herwig Paretzke and Dr. Ronald D. Neumann for valuable contributions to the section of low-dose radiation effects and to anonymous reviewers and editors for constructive suggestions. This work was supported by NASA NSCOR Grant NNJ04HJ12G and NIH Grant CA-78496-07 (P.H., PI) and an NSF Molecular and Cellular Biosciences Grant (E.O.V., PI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring institutions.

References

- http://www.accessexcellence.org/AE/AEC/CC/historical_back_ground.html
- <http://www.accessexcellence.org/AE/AEC/CC/radioactivity.html>
- Müller HJ (1927) Artificial transmutation of the gene. *Science* 66:84–87
- Timoféeff-Ressovsky NW, Zimmer KG, Delbrück M (1935) Über die Natur der Genmutation und der Genstruktur. *Nachr Ges Wiss Göttingen FG VI Biol NF* 1:189–245
- Friedberg EC (2003) DNA damage and repair. *Nature* 421:436–440
- Friedberg EC (2002) The intersection between the birth of molecular biology and the DNA repair and mutagenesis field. *DNA Repair* 1:855–867
- Zimmer KG (1992) The target theory. In: Cairns J, Stent GS, Watson JD (eds) *Phage and the origins of molecular biology*, Expanded edn. Cold Spring Harbor Laboratory, NY
- Friedland W, Dingfelder M, Jacob P, Paretzke HG (2005) Calculated DNA double-strand break and fragmentation. *Radiat Phys Chem* 72:279–286
- Ballarini F, Ottolenghi A (2004) A model of chromosome aberration induction and CML incidence at low doses. *Radiat Environ Biophys* 43:165–171
- Moolgavkar SH, Dewanji A, Venzon DJ (1988) A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor. *Risk Anal* 8:383–392
- Luebeck EG, Moolgavkar SH (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci USA* 99:15095–15100
- Heidenreich WF (2005) Heterogeneity of cancer risk due to stochastic effects. *Risk Anal* 25:1589–1594
- Black WC, Welch HG (1993) Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* 328:1237–1243
- Feuer EJ, Wun LM (1999) DEVCAN: probability of developing or dying of cancer. Version 4.0. National Cancer Institute, Bethesda
- Sachs RK, Chan M, Hlatky L, Hahnfeldt P (2005) Modeling intercellular interactions during carcinogenesis. *Radiat Res* 164:324–331
- Moolgavkar SH, Luebeck EG (2003) Multistage carcinogenesis and the incidence of human cancer. *Genes Chromosomes Cancer* 38:302–306
- Brugmans MJ, Rispens SM, Bijwaard H, Laurier D, Rogel A, Tomásek L, Tirmarçhe M (2004) Radon-induced lung cancer in French and Czech miner cohorts described with a two-mutation cancer model. *Radiat Environ Biophys* 43:153–165
- Heidenreich WF, Brugmans MJ, Little MP, Leenhouts HP, Paretzke HG, Morin M, Lafuma J (2000) Analysis of lung tumour risk in radon-exposed rats: an intercomparison of multi-step modeling. *Radiat Environ Biophys* 39:253–265
- Heidenreich WF, Paretzke HG (2004) Interpretation by modeling of observations in radon radiation carcinogenesis. *Radiat Prot Dosimetry* 112:501–507
- Bhowmick NA, Chytil A, Plieth D, Gorska AE, Dumont N, Shappell S, Washington MK, Neilson EG, Moses HL (2004) TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* 303:848–851
- Radisky DC, Bissell MJ (2004) Cancer. Respect thy neighbor! *Science* 303:775–777
- Reiss M, Barcellos-Hoff MH (1997) Transforming growth factor-beta in breast cancer: a working hypothesis. *Breast Cancer Res Treat* 45:81–95
- Gompertz B (1825) On the nature of the function expressive of the law of human mortality and on the new mode of determining the value of life contingencies. *Philos Trans R Soc A* 115:513–580
- Hahnfeldt P, Panigrahy D, Folkman J, Hlatky L (1999) Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer Res* 59(19):4770–4775
- Luckey TD (1980) *Hormesis with ionizing radiation*. CRC, Boca Raton

26. Tubiana M, Aurengo A, Averbek D, Masse R (2006) Recent reports on the effect of low doses of ionizing radiation and its dose–effect relationship. *Radiat Environ Biophys* 44:245–251
27. Pollycove M, Feinendegen LE (2001) Biologic response to low doses of ionizing radiation: detriment versus hormesis. Part 2: Dose responses of organisms. *J Nucl Med* 42:26N–32N
28. Feinendegen LE, Pollycove M, Neumann RD (2007) Whole body responses to low-level radiation exposure. New concepts in mammalian radiobiology. *Exp Hematol* 35:37–46
29. Mothersill C, Seymour CB (2006) Radiation-induced bystander effects and the DNA paradigm: an “out of field” perspective. *Mutat Res* 59:5–10
30. Kadhim MA, Hill MA, Moore SR (2006) Genomic instability and the role of radiation quality. *Radiat Prot Dosimetry* 122:221–227
31. Morgan WF, Day JP, Kaplan MI, McGhee REM, Limoli CL (1996) Genomic instability induced by ionizing radiation. *Radiat Res* 146:247–258
32. ICRU, I. C. o. R. U. a. M. (1983) Microdosimetry. ICRU-report 36, Bethesda
33. Hall EJ (2000) Radiobiology for the radiologist. Lippincott/Williams & Wilkins, Philadelphia/Baltimore
34. Fliedner TM, Dörr H, Meineke V (2005) Multi-organ involvement as a pathogenetic principle of the radiation syndromes: a study involving 110 case histories documented in SEARCH and classified as the bases of haematopoietic indicators of effect. *Br J Radiol Suppl* 27:1–8
35. Arthur C, Guyton MD, Hall JE (2000) Textbook of medical physiology WB Saunders, USA
36. Feinendegen LE, Loken MK, Booz J, Muehlensiepen H, Sondhaus CA, Bond VP (1995) Cellular mechanisms of protection and repair induced by radiation exposure and their consequences for cell system responses. *Stem Cells* 13(Suppl 1):7–20
37. Feinendegen LE, Bond VP, Sondhaus CA, Altman KI (1999) Cellular signal adaptation with damage control at low doses versus the predominance of DNA damage at high doses. *C R Acad Sci III, Sci Vie* 322:245–251
38. Feinendegen LE, Bond VP, Sondhaus CA (2000) The dual response to low-dose irradiation: induction vs. prevention of DNA damage. In: Yamada T, Mothersill C, Michael BD, Potten CS (eds) Biological effects of low dose radiation. *Excerpta Medica. International Congress Serie 1211*, Elsevier, Amsterdam, pp 3–17
39. Feinendegen LE, Pollycove M, Sondhaus CA (2004) Responses to low doses of ionizing radiation in biological systems. *Non-linearity Biol Toxicol Med* 2:143–171
40. Rothkamm K, Löbrich M (2003) Evidence for a lack of DNA double-strand break repair in human cells exposed to very low X-ray doses. *Proc Natl Acad Sci USA* 100:5057–5062
41. Bond VP, Varma M, Feinendegen LE, Wu CS, Zaider M (1995) Application of the HSEF to assessing radiation risks in the practice of radiation protection. *Health Phys* 68:627–631
42. Cleaver J (1968) Defective repair replication of DNA in Xeroderma Pigmentosum. *Nature* 218:652–656
43. Amundson SA, Lee RA, Koch-Paiz CA, Bittner ML, Meltzer P, Trent JM, Fornace AJJ (2003) Differential responses of stress genes to low dose-rate gamma irradiation. *Mol Cancer Res* 1:445–452
44. Chandra J, Samali A, Orrenius S (2000) Triggering and modulation of apoptosis by oxidative stress. *Free Radic Biol Med* 29:323–333
45. Finkel T, Holbrook NJ (2000) Oxidants, oxidative stress and the biology of aging. *Nature* 408:239–247
46. Zamboglou N, Porschen W, Muehlensiepen H, Booz J, Feinendegen LE (1981) Low dose effect of ionizing radiation on incorporation of iodo-deoxyuridine into bone marrow cells. *Int J Radiat Biol* 39:83–93
47. Feinendegen LE, Muehlensiepen H, Lindberg C, Marx J, Porschen W, Booz J (1984) Acute and temporary inhibition of thymidine kinase in mouse bone marrow cells after low-dose exposure. *Int J Radiat Biol* 45:205–215
48. Olivieri G, Bodycote J, Wolff S (1984) Adaptive response of human lymphocytes to low concentration of radioactive thymidine. *Science* 223:594–597
49. Wolff S, Afzal V, Wienke JK, Olivieri G, Michaeli A (1988) Human lymphocytes exposed to low doses of ionizing radiations become refractory to high doses of radiation as well as to chemical mutagens that induce double-strand breaks in DNA. *Int J Radiat Biol* 53:39–49
50. Kondo S (1988) Altruistic cell suicide in relation to radiation hormesis. *Int J Radiat Biol* 53:95–102
51. James SJ, Makinodan T (1990) T-cell potentiation by low dose ionizing radiation: possible mechanisms. *Health Phys* 59:29–34
52. Feinendegen LE, Neumann RD (2005) Physics must join with biology in better assessing risk from low-dose irradiation. *Radiat Prot Dosimetry* 117:346–356
53. Feinendegen LE, Bond VP, Sondhaus CA, Muehlensiepen H (1996) Radiation effects induced by low doses in complex tissue and their relation to cellular adaptive responses. *Mutat Res* 358:199–205
54. Feinendegen LE (2005) Evidence for beneficial low level radiation effects and radiation hormesis. *Br J Radiol Suppl* 78:3–7
55. Franco N, Lamartine J, Frouin V, Le Minter P, Petat C, Leplat JJ, Libert F, Gidrol X, Martin MT (2005) Low-dose exposure to gamma rays induces specific gene regulations in normal human keratinocytes. *Radiat Res* 163:623–635
56. Pollycove M, Feinendegen LE (2003) Radiation-induced versus endogenous DNA damage: possible effect of inducible protective responses in mitigating endogenous damage. *Hum Exp Toxicol* 22:290–306
57. Sen K, Sies H, Baeurle P (eds) (2000) Redox regulation of gene expression. Academic, San Diego
58. Feinendegen LE, Neumann RD (eds) (2000) Cellular responses to low doses of ionizing radiation. In: Workshop of the US Department of Energy (DOE), Washington, DC, and the National Institutes of Health (NIH), Bethesda, MD, 27–30 April 1999, Mary Woodward Lasker Center, Cloister, NIH, DOE report publication
59. Moolgavkar SH, Knudson AGJ (1981) Mutation and cancer: a model for human carcinogenesis. *J Natl Cancer Inst* 66:1037–1052
60. Heidenreich WF, Hoogenweem R (2001) Limits of applicability for the deterministic approximation of the two-step clonal expansion model. *Risk Anal* 21:103–105
61. Scott BR (2004) A biological-based model that links genomic instability, bystander effects, and adaptive response. *Mutat Res* 568:129–143
62. Schöllnberger H, Stewart RD, Mitchel REJ (2005) Low-LET-induced radioprotective mechanisms within a stochastic two-stage cancer model. *Dose Response* 3:508–518
63. Leonard BE (2007) Adaptive response and human benefit: part I. A microdosimetry dose-dependent model. *Int J Radiat Biol* 83:115–131
64. Feinendegen LE (2003) Relative implications of protective responses versus damage induction at low-dose and low-dose rate exposures, using the microdose approach. *Radiat Prot Dosimetry* 104:337–346
65. Azzam EI, de Toledo SM, Raaphorst GP, Mitchel REJ (1996) Low-dose ionizing radiation decreases the frequency of neoplastic transformation to a level below the spontaneous rate in C3H 10T1/2 cells. *Radiat Res* 146:369–373

66. Redpath JL, Antoniono RJ (1998) Introduction of an adaptive response against spontaneous neoplastic transformation in vitro by low-dose gamma radiation. *Radiat Res* 149:517–520
67. Mitchel RJE, Jackson JS, Morrison DP, Carlisle SM (2003) Low doses of radiation increase the latency of spontaneous lymphomas and spinal osteosarcomas in cancer-prone radiation-sensitive Trp53 heterozygous mice. *Radiat Res* 159:320–327
68. Tapio S, Jacob V (2007) Radioadaptive response revisited. *Radiat Environ Biophys* 46:1–12
69. Feinendegen LE, Bond VP, Booz J (1994) The quantification of physical events within tissue at low levels of exposure to ionizing radiation. *ICRU News* 2:9–13
70. Feinendegen LE, Booz J, Bond VP, Sondhaus CA (1985) Microdosimetric approach to the analysis of cell responses at low dose and low dose rate. *Radiat Prot Dosimetry* 13:299–306
71. Feinendegen LE, Graessle D. (2002) Energy deposition in tissue during chronic irradiation and the biological consequences. In: *Chronic irradiation: tolerance and failure in complex biological systems*. Br J Radiol Suppl, vol 26, British Institute of Radiology, London, pp 6–14
72. Yamamoto O, Seyama T, Ito A, Fujimoto N (1998) Oral administration of tritiated water (THO) in mouse. III: Low dose-rate irradiation and threshold dose-rate for radiation risk. *Int J Radiat Biol* 73:535–541
73. Zablotska LB, Ashmore JP, Jowe GR (2004) Analysis of mortality among Canadian nuclear power industry workers after chronic low-dose exposure to ionizing radiation. *Radiat Res* 161:633–641
74. Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA (2005) Complement factor H polymorphism and age-related macular degeneration. *Science* 308:421–424
75. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, Schetz-Boutaud N, Agarwal A, Postel EA, Pericak-Vance MA (2005) Complement factor H variant increases the risk of age-related macular degeneration. *Science* 308:419–421
76. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. *Science* 308:385–389
77. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A, Styrkarsdóttir U, Magnusson KP, Walters GB, Pálsdóttir E, Jónsdóttir T, Gudmundsdóttir T, Gylfason A, Saemundsdóttir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdóttir U, Gulcher JR, Kong A, Stefansson K (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323
78. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshzhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885
79. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu X, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF (2006) A common genetic variant is associated with adult and childhood obesity. *Science* 312:279–283
80. Easton DF, Pooley NP, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Luccarini C, Conroy D, Shah M, Munday H, Jordan C, Perkins B, West J, Redman K, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, Macpherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Scutella C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, Aghmesheh M, Amor D, Andrews L, Antill Y, Armes J, Armitage S, Arnold L, Balleine R, Begley G, Beilby J, Bennett I, Bennett B, Berry G, Blackburn A, Brennan M, Brown M, Buckley M, Burke J, Butow P, Byron K, Callen D, Campbell I, Clarke C, Colley A, Cotton D, Cui J, Culling B, Cummings M, Dawson SJ, Dixon J, Dobrovic A, Dudding T, Edkins T, Eisenbruch M, Farshid G, Fawcett S, Field M, Fergair F, Fleming J, Forbes J, Friedlander M, Gaff C, Gardner M, Gattas M, George P, Giles G, Gill G, Goldblatt J, Greening S, Grist S, Haan E, Harris M, Hart S, Hayward N, Hopper J, Humphrey E, Jenkins M, Jones A, Kefford R, Kirk J, Kollias J, Kovalenko S, Lakhani S, Leary J, Lim J, Lindeman G, Lipton L, Lobb L, Maclurcan M, Mann G, Marsh D, McCredie M, McKay M, McLachlan SA, Meiser B, Milne R, Mitchell G, Newman B, O'Loughlin I, Osborne R, Peters L, Phillips K, Price M, Reeve J, Reeve T, Richards R, Rinehart G, Robinson B, Rudzki B, Salisbury E, Sambrook J, Saunders C, Scott C, Scott E, Scott R, Seshadri R, Shelling A, Southey M, Spurdle A, Suthers G, Taylor D, Tennant C, Thorne H, Townshend S, Tucker K, Tyler J, Venter D, Visvader J, Walpole I, Ward R, Waring P, Warner B, Warren G, Watson E, Williams R, Wilson J, Winship I, Young MA, Bowtell D, Green A, Defazio A, Gertig D, Webb P, Manermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–1093
81. Peacock ML, Warren JT Jr, Roses AD, Fink JK (1993) Novel polymorphism in the A4 region of the amyloid precursor protein gene in a patient without Alzheimer's disease. *Neurology* 43:1254–1256
82. Schadt EE, Lum PY (2006) Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J Lipid Res* 47(12):2601–2613
83. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296:752–755
84. DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* 172:1155–1164
85. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, Mueller M, Hummel O, Monti J, Zidek V, Musilova A, Kren V, Causton H, Game L, Born G, Schmidt S, Muller A, Cook SA, Kurtz TW, Whittaker J, Pravenec M, Aitman TJ (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet* 37:243–253
86. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* 29:389–395

87. Klose J, Nock C, Herrmann M, Stuhler K, Marcus K, Bluggel M, Krause E, Schalkwyk LC, Rastan S, Brown SD, Bussow K, Himmelbauer H, Lehrach H (2002) Genetic analysis of the mouse brain proteome. *Nat Genet* 30:385–393
88. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, Schadt EE (2004) Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 75:1094–1105
89. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743–747
90. Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nat Genet* 32:261–266
91. Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, Colinao V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302
92. Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET (2005) Genome-wide associations of gene expression variation in humans. *PLoS Genet* 1:e78
93. Kulp DC, Jagalur M. (2006) Causal inference of regulator-target pairs by gene mapping of expression phenotypes. *BMC Genomics* 7:125
94. Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, Wang S, Drake TA, Lusk AJ, Schadt EE (2006) Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J Neurochem* March 15 (Epub ahead of print)
95. Mehrabani M, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, Suh M, Armour C, Edwards S, Lamb J, Lusk AJ, Schadt EE (2005) Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat Genet* 37:1224–1233
96. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, Lusk AJ (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 37:710–717
97. Karp CL, Grupe A, Schadt E, Ewart SL, Keane-Moore M, Cuomo PJ, Kohl J, Wahl L, Kuperman D, Germer S, Aud D, Peltz G, Wills-Karp M (2000) Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol* 1:221–226
98. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302:2141–2144
99. Mural RJ, Adams MD, Myers EW, Smith HO, Miklos GL, Wides R, Halpern A, Li PW, Sutton GG, Nadeau J, Salzberg SL, Holt RA, Kodira CD, Lu F, Chen L, Deng Z, Evangelista CC, Gan W, Heiman TJ, Li J, Li Z, Merkulov GV, Milshina NV, Naik AK, Qi R, Shue BC, Wang A, Wang J, Wang X, Yan X, Ye J, Yooseph S, Zhao Q, Zheng L, Zhu SC, Biddick K, Bolanos R, Delcher AL, Dew IM, Fasulo D, Flanigan MJ, Huson DH, Kravitz SA, Miller JR, Mobarry CM, Reinert K, Remington KA, Zhang Q, Zheng XH, Nusskern DR, Lai Z, Lei Y, Zhong W, Yao A, Guan P, Ji RR, Gu Z, Wang ZY, Zhong F, Xiao C, Chiang CC, Yandell M, Wortman JR, Amanatides PG, Hladun SL, Pratts EC, Johnson JE, Dodson KL, Woodford KJ, Evans CA, Gropman B, Rusch DB, Venter E, Wang M, Smith TJ, Houck JT, Tompkins DE, Haynes C, Jacob D, Chin SH, Allen DR, Dahlke CE, Sanders R, Li K, Liu X, Levitsky AA, Majoros WH, Chen Q, Xia AC, Lopez JR, Donnelly MT, Newman MH, Glodok A, Kraft CL, Nodell M, Ali F, An HJ, Baldwin-Pitts D, Beeson KY, Cai S, Carnes M, Carver A, Caulk PM, Center A, Chen YH, Cheng ML, Coyne MD, Crowder M, Danaher S, Davenport LB, Desilets R, Dietz SM, Doup L, Dullaghan P, Ferreira S, Fosler CR, Gire HC, Gluecksmann A, Gocayne JD, Gray J, Hart B, Haynes J, Hoover J, Howland T, Ibegwam C, Jalali M, Johns D, Kline L, Ma DS, MacCawley S, Magoon A, Mann F, May D, McIntosh TC, Mehta S, Moy L, Moy MC, Murphy BJ, Murphy SD, Nelson KA, Nuri Z, Parker KA, Prudhomme AC, Puri VN, Qureshi H, Raley JC, Reardon MS, Regier MA, Rogers YH, Romblad DL, Schutz J, Scott JL, Scott R, Sitter CD, Smallwood M, Sprague AC, Stewart E, Strong RV, Suh E, Sylvester K, Thomas R, Tint NN, Tsonis C, Wang G, Wang G, Williams MS, Williams SM, Windsor SM, Wolfe K, Wu MM, Zaveri J, Chaturvedi K, Gabrielian AE, Ke Z, Sun J, Subramanian G, Venter JC, Pfannkoch CM, Barnstead M, Stephenson LD (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* 296:1661–1671
100. Schadt EE, Edwards SW, GuhaThakurta D, Holder D, Ying L, Svetnik V, Leonardson A, Hart KW, Russell A, Li G, Cavet G, Castle J, McDonagh P, Kan Z, Chen R, Kasarskis A, Margarint M, Caceres RM, Johnson JM, Armour CD, Garrett-Engele PW, Tsinoremas NF, Shoemaker DD (2004) A comprehensive transcript index of the human genome generated using microarrays and computational approaches. *Genome Biol* 5:R73
101. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scherer S, Boguski MS (2001) Experimental annotation of the human genome using microarray technology. *Nature* 409:922–927
102. DePrimo SE, Wong LM, Khatri DB, Nicholas SL, Manning WC, Smolich BD, O'Farrell AM, Cherrington JM (2003) Expression profiling of blood samples from an SU5416 phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC Cancer* 3:3
103. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003) PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267–273
104. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
105. Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, Buratto B, Roberts C, Schadt E, Ulrich RG (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* 175:28–42
106. Alberts R, Terpstra P, Bystrykh LV, de Haan G, Jansen RC (2005) A statistical multiprobe model for analyzing *cis* and *trans* genes in genetical genomics experiments with short-oligonucleotide arrays. *Genetics* 171:1437–1439
107. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, Hsu HC, Mountz JD, Baldwin NE, Langston MA, Threadgill DW, Manly KF, Williams RW (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 37:233–242

108. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 437:1365–1369
109. Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. *Trends Genet* 17:388–391
110. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, Fischer J, Maatz H, Kren V, Pravenec M, Hubner N, Aitman TJ (2006) Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet* 2:e172
111. Petretto E, Mangion J, Pravanec M, Hubner N, Aitman TJ (2006) Integrated gene expression profiling and linkage analysis in the rat. *Mamm Genome* 17:480–489
112. Schadt EE (2005) Exploiting naturally occurring DNA variation and molecular profiling data to dissect disease and drug response traits. *Curr Opin Biotechnol* 16:647–654
113. GuhaThakurta D, Xie T, Anand M, Edwards SW, Li G, Wang SS, Schadt EE (2006) *Cis*-regulatory variations: a study of SNPs around genes showing *cis*-linkage in segregating mouse populations. *BMC Genomics* 7:235
114. Schadt EE, Sachs A, Friend S (2005) Embracing complexity, inching closer to reality. In: *Science STKE* 2005, p 40
115. Zhu J, Lum PY, Lamb J, GuhaThakurta D, Edwards SW, Thieringer R, Berger JP, Wu MS, Thompson J, Sachs AB, Schadt EE (2004) An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res* 105:363–374
116. Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, Rual JF, Kennedy S, Dybbs M, Bertin N, Kaplan JM, Vidal M, Ruvkun G (2005) Functional genomic analysis of RNA interference in *C. elegans*. *Science* 308:1164–1167
117. Palsson BO (2006) *Systems biology*. Cambridge University Press, Cambridge
118. Ross SM (1985) *Introduction to probability models*. Academic, Orlando
119. Matsuno H, Tanaka Y, Aoshima H, Doi A, Matsui M, Miyano S (2003) Biopath-ways representation and simulation on hybrid functional Petri nets. In *Silico Biol* 3:32
120. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan G. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308:523–529
121. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE (2007) Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol* 3:e69
122. Chiellini C, Bertacca A, Novelli SE, Gorgun CZ, Ciccarone A, Giordano A, Xu H, Soukas A, Costa M, Gandini D, Dimitri R, Bottone P, Cecchetti P, Pardini E, Perego L, Navalesi R, Folli F, Benzi L, Cinti S, Friedman JM, Hotamisligil GS, Maffei M. (2002) Obesity modulates the expression of haptoglobin in the white adipose tissue via TNFalpha. *J Cell Physiol* 190:251–258
123. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *J Comput Biol* 7:601–620
124. Peschel M, Mende W. (1986) *Do we live in a Volterra world?*. Akademie-Verlag, Berlin
125. Savageau MA (1969) *Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation*. *J Theor Biol* 25:370–379
126. Savageau MA (1969) *Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions*. *J Theor Biol* 25:365–369
127. Savageau MA, Voit EO (1987) Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math Biosci* 87:83–115
128. Savageau MA (1976) *Biochemical systems analysis: a study of function and design in molecular biology*. Addison-Wesley Pub. Co. Advanced Book Program, Reading
129. Voit EO (1991) *Canonical nonlinear modeling: S-system approach to understanding complexity*. Van Nostrand Reinhold, New York
130. Voit EO (2000) *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press, New York
131. Voit EO (2000) *Canonical modeling: review of concepts with emphasis on environmental health*. *Environ Health Perspect* 108(Suppl 5):895–909
132. Torres NV, Voit EO (2002) *Pathway analysis and optimization in metabolic engineering*. Cambridge University Press, New York
133. Goel G, Chou I-C, Voit EO (2006) *Biological systems modeling and analysis: a biomolecular technique of the 21st century*. *J Biomol Tech* 17:252–269
134. Voit EO (2004) The dawn of a new era of metabolic systems analysis. *Drug Discov Today BioSilico* 2:182–189