A systems approach defining constraints of the genome architecture on lineage selection and evolvability during somatic cancer evolution

Albert Rübben[1,2,*] and Ole Nordhoff[1]

[1]Independent Institute of Systems Sciences Aachen, 52064 Aachen, Germany
[2]Department of Dermatology, RWTH Aachen University, 52074 Aachen, Germany
*Author for correspondence (albert.ruebben@post.rwth-aachen.de)

Biology Open 2, 49–62
doi: 10.1242/bio.20122543
Received 17th July 2012
Accepted 15th October 2012

Summary

Most clinically distinguishable malignant tumors are characterized by specific mutations, specific patterns of chromosomal rearrangements and a predominant mechanism of genetic instability but it remains unsolved whether modifications of cancer genomes can be explained solely by mutations and selection through the cancer microenvironment.

It has been suggested that internal dynamics of genomic modifications as opposed to the external evolutionary forces have a significant and complex impact on Darwinian species evolution. A similar situation can be expected for somatic cancer evolution as molecular key mechanisms encountered in species evolution also constitute prevalent mutation mechanisms in human cancers. This assumption is developed into a systems approach of carcinogenesis which focuses on possible inner constraints of the genome architecture on lineage selection during somatic cancer evolution. The proposed systems approach can be considered an analogy to the concept of evolvability in species evolution.

The principal hypothesis is that permissive or restrictive effects of the genome architecture on lineage selection during somatic cancer evolution exist and have a measurable impact. The systems approach postulates three classes of lineage selection effects of the genome architecture on somatic cancer evolution: i) effects mediated by changes of fitness of cells of cancer lineage, ii) effects mediated by changes of mutation probabilities and iii) effects mediated by changes of gene designation and physical and functional genome redundancy. Physical genome redundancy is the copy number of identical genetic sequences. Functional genome redundancy of a gene or a regulatory element is defined as the number of different genetic elements, regardless of copy number, coding for the same specific biological function within a cancer cell. Complex interactions of the genome architecture on lineage selection may be expected when modifications of the genome architecture have multiple and possibly opposed effects which manifest themselves at disparate times and progression stages.

Dissection of putative mechanisms mediating constraints exerted by the genome architecture on somatic cancer evolution may provide an algorithm for understanding and predicting as well as modifying somatic cancer evolution in individual patients.

Key words: Carcinogenesis, Evolvability, Genome architecture, Somatic cancer evolution

Introduction

Cancer as a devastating but variable disease has always been subject of various theories, many of them sharing an explicit or implicit holistic or systems approach. Today, the scientific community has widely accepted the concept that cancer is caused by changes within the cellular genome (Knudson, 1971; Loeb et al., 1974; Fearon and Vogelstein, 1990) and that carcinogenesis and cancer progression resembles an evolutionary process (Foulds, 1954; Nowell, 1976; Gatenby and Vincent, 2003). This view may be regarded as a specification of Boveri's idea that somatic mutations are the basis of cancer (Boveri, 1914). Although the genetic hypothesis of cancer has been discussed for nearly a century, there is still an ongoing debate on the impact of the various molecular mechanisms in causing and defining cancer (Duesberg et al., 2005).

The two most prominent positions may be described as the sequential gene mutation and selective clonal expansion theory and the mutator phenotype and selective clonal expansion theory.

The first theory states that activation or inactivation of a gene or a set of genes by specific mutations conveys a selective advantage to a cell resulting in clonal expansion of cells harbouring these mutations. The degree of malignancy increases with successive rounds of mutation acquisition and clonal expansion forming a two-stage or a multistage process of carcinogenesis and tumor progression. Historically, analysis of statistical data on cancer incidences first led to the Armitage–Doll multistage model which did not incorporate clonal expansion (Nordling, 1953; Armitage and Doll, 1954). In order to provide a better fit with the available biological data, the model was changed into a model with two stages and clonal expansion (Armitage and Doll, 1957). Through

the study of hereditary and sporadic retinoblastomas, Knudson postulated a two hit mechanism which included the process of growth and differentiation in normal tissues (Knudson, 1971). Subsequently, Moolgavkar, Venzon and Knudson introduced a two-stage model with clonal expansion of intermediate cells (Moolgavkar and Venzon, 1979; Moolgavkar and Knudson, 1981) which was later extended to a multistage model (Moolgavkar and Luebeck, 1992). Molecular analysis of clinical material then confirmed that both, clonal expansions and stepwise accumulation of specific mutations were necessary for carcinogenesis in most human cancers (Fearon and Vogelstein, 1990).

The second theory may be regarded as an extension of the first hypothesis. It concludes that most cancers exhibit a mutator phenotype as normal mutation rates are insufficient to account for the multiple mutations observed in human tumors. This concept was introduced by Loeb in 1974 (Loeb et al., 1974; Loeb, 1991; Loeb and Christians, 1996) and later confirmed by the molecular analysis of hereditary non-polyposis colorectal cancer (Lynch et al., 1966; Aaltonen et al., 1994). The importance of chromosomal instability for somatic cancer evolution was postulated by Nowell (Nowell, 1976). The mutator phenotype may be induced by gene mutations enhancing the mutation rate per base pair or by mutations inducing chromosomal instability (Beckman and Loeb, 2005a). Preneoplastic or neoplastic cells with a mutator phenotype will undergo the same successive rounds of mutation acquisition and clonal expansions as predicted by the first theory. The requirement of a mutator phenotype to initiate carcinogenesis has been questioned by supporters of the first theory (Tomlinson and Bodmer, 1999) although one may argue that the first theory does not rule out that genetic instability may be acquired during later stages of malignant tumor progression.

Significant differences between the two theories can be found with regards to the phylogenetic tree of cell clones generated during the process of somatic tumorigenesis and tumor progression. Under the assumptions of the first theory, the progression of cancer through its first stages is sequential and forms a single lineage due to the low normal mutation rate which reduces the chances that a new positively selected phenotype arises more than once during early stages of clonal expansion. The characteristic of single lineage progression is not the total lack of subclone formation at a specific cancer stage but the feature that all tumor cells of one tumor stage find their most recent common ancestor in the cell population of the previous stage (Tsao et al., 1999). The second theory does not explicitly comment on this aspect but it could be demonstrated that cancers arising on the grounds of inherited genetic instability display subclone divergence already within the preneoplastic tumor cell compartment (Tsao et al., 1999; Shibata et al., 1993). In this case, the phylogenetic tree of somatic cancer evolution splits up early into multiple branches which has been named multilineage progression (Tsao et al., 1999) or subclone divergence (Welch et al., 1984). Multilineage progression has also been shown in genetically unstable cutaneous lymphoma (Rübben et al., 2004). Little and Wright have formulated a stochastic multi pathway model of carcinogenesis which incorporates genomic instability and which is able to explain tumor incidence data (Little and Wright, 2003). The difference between single lineage and multilineage tumor evolution has an impact on cancer diagnosis and treatment. The degree of genetic and phenotypical heterogeneity should be higher in cancers arising through a multi pathway evolutionary process. Phenotypical heterogeneity negatively affects the prognostic value of a molecular diagnostic test and it should similarly reduce the efficacy of therapies which target specific mutations (Shibata, 2012). It is interesting to note that the hypothesis of cancer stem cells (Reya et al., 2001) favors multilineage progression of cancer cell clones. The postulated tumor cells with stem cell-like properties of cell renewal represent stable most common ancestors within a phylogenetic tree of somatic tumorigenesis harbouring cancer subclones. Data generated with cell lines *in vitro* seem to support this view (Marzi et al., 2007).

Despite the discussed differences, the endpoint of carcinogenesis in both theories is a neoplastic cell clone or a group of genetically related cell clones which have adapted to all external selective constraints imposed by its microenvironment by accumulation of advantageous mutations.

The scope of the present contribution is to propose a complementary systems approach of carcinogenesis which focuses on possible inner constraints of the genome architecture on lineage selection during somatic cancer evolution. These inner constraints may have an independent impact on the process of somatic tumorigenesis and tumor progression. The proposed conceptual framework may be considered an analogy to the concept of evolvability in Darwinian species evolution (Conrad, 1990; Shapiro, 1999; Poole et al., 2003; Brookfield, 2009). Dissection of putative mechanisms mediating constraints exerted by the genome architecture on somatic cancer evolution may provide an algorithm for understanding, predicting and modifying somatic cancer evolution in individual patients.

In a first step, a conceptual framework will be developed which defines and systemizes putative constraints of the genome architecture on somatic cancer evolution. In a second step, hypotheses derived from this conceptual framework will be proposed which can be submitted to experimental validation. Finally, an algorithm will be put forward which may be used to modify somatic cancer evolution for therapeutic purposes. In addition, results from a computer simulation of somatic cancer evolution based on unequal sister chromatid exchange will be presented in the figures in order to illustrate putative restrictive or permissive effects of the genome architecture on somatic cancer evolution.

Systems concept

Basic assumptions

1. The presented mechanistic approach is embedded in the classical framework of a multistage and evolutionary cancer model where individual somatic cancer evolution starts with one cell and proceeds with successive rounds of mutation acquisitions, clonal expansions and partial elimination of tumor subclones by various mechanisms (Beckman and Loeb, 2005a).
2. The approach does not differentiate between carcinogenesis and cancer progression as somatic cancer evolution contains both aspects and as the diagnosis of cancer is based on histopathology and clinics but not on molecular grounds. With regards to the proposed systems approach, the terms preneoplastic cell(s) or cancer cell(s) will, therefore, be replaced by “cell(s) of cancer lineage (CCL)”.
3. It is assumed that cells of cancer lineage interact with their microenvironment. This interaction leads to positive or negative selection of subclones. The phenotype of a CCL is

determined by its transcriptome and proteome. The microenvironment is dependent on the clinical tumor stage.

4. The term "genome architecture" stands for the cell genome as a whole and has been chosen to underline that its influence on somatic cancer evolution is not only determined by coding and regulatory sequences but also by the physical arrangement of coding and regulatory sequences within chromosomes or chromosome fragments as well as by chromatin organization and nuclear topology of chromosomes. It relates to the term "genome system architecture", used by Shapiro in order to describe species evolution as a systems engineering process (Shapiro, 1999).

   The genome architecture comprises five structural aspects:

   (i) base composition of genes, of regulatory elements and of foreign (viral) DNA/RNA sequences
   (ii) presence and copy number of wild type or mutated genes, of regulatory elements, and of foreign (viral) DNA/RNA sequences
   (iii) physical arrangement of genes of regulatory elements, and of foreign (viral) DNA/RNA sequences within chromosomes or mitochondrial DNA, including rearrangements of chromosomes or chromosome fragments
   (iv) chromatin organization including epigenetic modifications, interaction of DNA with RNAs and nuclear proteins and nuclear topology of chromosomes
   (v) status of gene transcription.

   Chromatin organization, epigenetic modifications and transcriptional status are interrelated but shall be considered independently as not all changes of chromatin organization or epigenetic modifications are reflected by changes in gene transcription.

5. Genomic instability is considered a hallmark of neoplastic disease which alters the genome architecture although mutations leading to genomic instability may not represent the initiating events in all human cancers.
6. It is admitted that genomic instability is mediated by various but specific mechanisms and that genomic instability is the result of mutations of specific genes in most cases although infection with oncogenic viruses and chronic inflammation may substitute for specific mutations of host cell DNA (zur Hausen, 2000; Yan et al., 2006).
7. A basic assumption of the proposed mechanistic approach is that various types of genomic instability promote different and specific mutations of individual genes, of regulatory elements, of non-coding DNA, of chromosomes and chromosome fragments depending on the base composition, sequence context, physical arrangement of the target DNA sequences in the genome and depending on the transcriptional status of the cell (Fig. 1). As mutation acquisition itself is a stochastic process, specificity in this context means that the type of genomic instability affects mutation probabilities in daughter cells of CCL and that these mutations probabilities may be estimated *a priori*.

   The type of genomic instability is dependent on the transcriptome and proteome and is thereby not only a function of the genome architecture but also modulated by the interaction of the CCL with the microenvironment (Fig. 1). Even in the absence of genomic instability, changes of chromatin organization induced by changes of transcriptional activity may modify gene mutation probability as it has been

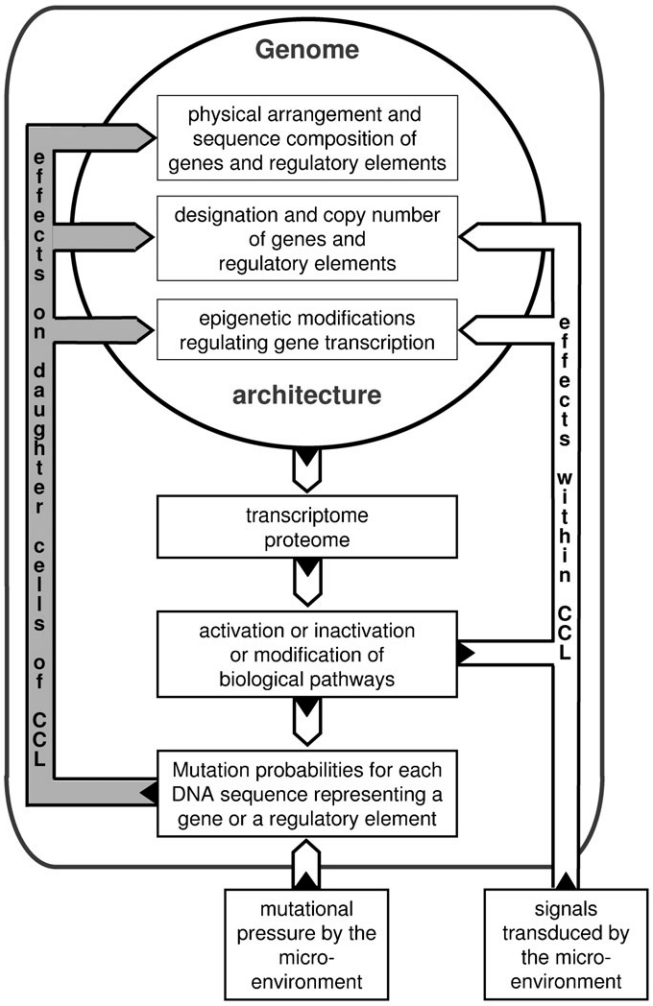

Fig. 1. Interdependency of cancer genome architecture and somatic cancer evolution (CCL = cell of cancer lineage).

shown that transcriptional active DNA regions may display more efficient repair of base damage and less efficient repair of DNA breaks (Bailey and Bedford, 2006).

The mechanisms leading to genomic instability are multiple as multiple different systems exist which maintain genome integrity and which target different forms of DNA damage. The best known examples of genomic instability are post replicative mismatch repair (MMR) deficiency leading to mutations of short repetitive sequences such as microsatellite DNA (microsatellite instability = MIN) and chromosomal instability (= CIN) which induces rearrangement, loss, amplification or duplication of chromosomes or chromosome fragments resulting in aneuploidy. Besides, various other mechanisms exist such as defective nucleotide excision repair (NER), defective base excision repair (BER) and defective cell cycle checkpoint controls.

DNA-repair systems such as MMR and NER are constituted of multiple proteins and defects of different factors within a DNA repair system may result in differing mutation frequencies, mutation spectra and mutation hotspots

within coding and non-coding DNA. This has been demonstrated in bacterial, yeast and animal models (Schaaper and Dunn, 1987; Habraken et al., 1996; Marti et al., 2003; Harfe and Jinks-Robertson, 2000; Denver et al., 2005; Denver et al., 2006; Yao et al., 1999; Kuraguchi et al., 2001; Andrew et al., 2000).

Different biological mechanisms and pathways preserve chromosomal stability. Mutations which reduce efficiency of repair of double strand DNA breaks, mutation which induce telomere loss or centrosome amplification, defects in homologous recombination as well as suppression of hyperrecombination have been implicated in chromosomal instability (Lobachev et al., 2002; Traverso et al., 2003; Sabatier et al., 2005; Michor, 2005; Deans et al., 2003; Saunders, 2005; Bailey and Murnane, 2006).

It has been shown that human tumors with deficiency of MMR display different mutation patterns at target genes involved in carcinogenesis than tumors with chromosomal instability (CIN) (Zhou et al., 2002). Distinct mutation spectra at target genes could also be found when comparing hereditary non-polyposis colorectal cancers (HNPCC) with germline hMSH2 mutations and HNPCC with germline hMLH1 mutations (Fujiwara et al., 1998).

Finally, the observation that hereditary diseases with specific defects of genomic stability control such as Lynch syndrome or Li–Fraumeni syndrome are associated with distinct patterns of cancer susceptibility clearly indicates that the non-random nature of mutations arising on the background of genomic instability associated with defined molecular defects has a very strong impact on somatic cancer evolution (Marra and Boland, 1995; Birch et al., 2001). Likewise it has been postulated “that tumor genomes are shaped by selection for alterations that promote survival and growth advantage, as well as by the particular dysfunctions in genes responsible for maintenance of genetic integrity” (Snijders et al., 2003).

8. It is further assumed that a specific mechanism of genomic instability may represent the only form of genomic instability within a cell of cancer lineage or it may coexist with other forms of genomic instability. The latter possibility seems less frequent in human cancers but, nonetheless, has been demonstrated repeatedly (Rübben et al., 2000; Camps et al., 2004; Vauhkonen et al., 2006; Trautmann et al., 2006).

Constraints of the genome architecture on somatic cancer evolution

Lineage selection effects exerted by the genome architecture

A change of the genome architecture of a cell of cancer lineage may have immediate effects within one cancer evolution stage or may induce long-term effects on lineage selection. The immediate effects of the modified genome architecture on cancer evolution manifest themselves prior to the acquisition of additional mutations. The destiny of a new tumor cell subclone will be determined by changes in tumor cell fitness within the original or a new anatomical compartment. Provided that no extinction of the subclone occurs prior to acquisition of additional mutations, the initial mutation will also affect further evolution of daughter cell lineages within the original microenvironment or at later stages of clinical tumor progression. These later effects of the initial mutation are the effects on lineage selection.

Within the scheme of classical Darwinian species evolution, lineage selection describes an evolutionary process which favors long-term fitness whenever there is a conflict between the short- and the long-term effects of selection (Nunney, 1989; Nunney, 1999). In analogy to Darwinian species evolution, subclones which arise during the process of somatic cancer evolution may be regarded as variants or new species which evolve and colonize different ecological niches within the human body as metastases.

Although governed by the same process of evolution, differentiation between immediate effects and effects on lineage selection is reasonable as a mutation conferring selection advantage to a cell of cancer lineage within its original microenvironment may be responsible for elimination of the resulting cell clone or cell line at subsequent clinical progression stages. The opposite case may also be true, when the genome architecture of a CCL reduces the proliferation rate at an early progression stage but becomes a positive selective factor at later stages of somatic cancer evolution. Three classes of effects on lineage selection may be distinguished within the proposed conceptual framework.

Lineage selection effects mediated by CLL fitness

As depicted above, changes of the genome architecture may enhance or decrease fitness of a CCL during later stages of tumor cell evolution. A clinical example of a fitness enhancing effect during somatic cancer evolution is the simple constellation when a mutation, such as the acquisition of drug resistance, occurs in cells of cancer lineage without apparent selective pressure and becomes a critical selective advantage for further evolution of CCLs at later clinical stages. In terms of Darwinian evolution of species, this represents preadaptation (Budd, 2006). Preadaptation is an important feature of classical Darwinian evolution and evolution of drug resistance has been described and quantified in evolutionary models (Iwasa et al., 2006).

Lineage selection effects mediated by changes of mutation probability

The second class of lineage selection effects is mediated by changes of mutation probabilities of individual genes, regulatory elements, chromosomes and chromosome fragments.

Mutations resulting in genomic instability induce lineage selection effects as stochastic mutations which arise in genetically unstable CCLs are non-random with regards to mutation probability and mutation patterns. Mutations which induce genomic instability have a strong influence on further evolution of a tumor cell lineage as they will increase or decrease the probability that a specific fitness enhancing or fitness decreasing mutation will occur at later stages of tumor cell evolution. The fact, that genetic instability is a hallmark of most human cancers (Beckman and Loeb, 2005a) and is often already detected in preneoplastic lesions (Tsao et al., 1999; Cohn et al., 2001), by itself demonstrates that cells of cancer lineage harbouring mutations leading to genetic instability tend to be positively selected during the process of tumor cell evolution. Beckman and Loeb have also evaluated the impact of genomic instability on negative clonal selection in somatic cancer evolution by a deterministic mathematical approach (Beckman and Loeb, 2005b). They conclude that effects of negative clonal selection can be expected but that both forms of genomic instability (MIN, CIN) are not detrimental to somatic cancer evolution within realistic estimates of mutation frequencies and of the net number of reduced-fitness loci. Nevertheless, their data

suggest the existence of an upper limit of the mutation rate, beyond which negative clonal selection begins to limit survival of mutator clones. Additional information derives from the application of the quasispecies model to somatic cancer evolution (Brumer and Shakhnovich, 2004; Brumer et al., 2006). It could be deduced by mathematical modeling that cancers with CIN differ from cancers with MIN with respect to their tolerance of high mutation rates.

The importance of chromosomal instability as an initiating event of cancer development has also been evaluated by different mathematical approaches which model evolutionary dynamics quantitatively (Michor, 2005; Nowak et al., 2004).

Lineage selection effects mediated by changes of gene designation and physical and functional genome redundancy

The third class of lineage selection effects results of changes of functional gene designation and physical and functional genome redundancy.

A physical DNA stretch in a CCL representing a gene or a regulatory element may have various biological functions. With focus on somatic cancer cell evolution, one may divide the functions in two basic categories with few specific attributes: first, functions with regards to cell survival within a specific clinical stage and microenvironment of a CCL and second, functions with regards to the transition of a CLL to a new clinical stage or microenvironment during the process of somatic cancer evolution. For each biological function of a gene or regulatory element, attributes of both categories may be combined.

Within a clinical stage or microenvironment, functions of wild type or mutated genes or regulatory elements may be essential to sustain cell survival. Genes or regulatory elements with one or more essential functions are designated essential genes within the proposed mechanistic systems approach. Genes or regulatory elements displaying functions which modify the fitness of a CCL in a given clinical stage or microenvironment may be considered non-essential fitness modifying genes while genes with functions without any apparent consequence on cellular fitness will be considered fitness neutral genes.

With regards to the role in somatic cancer evolution, functions of wild type or mutated genes or regulatory elements may inhibit, promote or accelerate the evolutionary process. Genes which inhibit tumor cell evolution, are considered tumor suppressor genes or gatekeeper genes (Sidransky, 1996; Kinzler and Vogelstein, 1997) while genes which promote carcinogenesis or tumor progression when they are amplified, activated or have acquired a gain of function mutation are oncogenes. Genes or regulatory elements which are implicated in the maintenance of the genomic integrity are called caretaker genes. Mutations of caretaker genes but also changes of expression levels of caretaker genes may lead to genomic instability and thus accelerate tumor cell evolution (Kinzler and Vogelstein, 1997).

The designation of a gene or regulatory element within this scheme, i.e. an essential gene, a non-essential fitness modifying gene, a fitness neutral gene, a tumor suppressor/gatekeeper gene, an oncogene or a caretaker gene is fundamental for lineage selection effects through changes of physical and functional genome redundancy.

Physical genome redundancy is the copy number of identical genetic sequences. Functional genome redundancy of a gene or a regulatory element is defined as the number of different genetic elements, regardless of copy number, coding for the same specific biological function within a cancer cell. Functional genome redundancy is lost when only one specific DNA sequence, regardless of its copy number, assumes a specific biological function within a cancer cell at a specific progression state during the process of somatic cancer cell evolution. Functional genome redundancy of a gene or a regulatory element acts together with copy number changes in order to affect lineage selection effects.

A mutation within a cell of cancer lineage may reduce or amplify the copy number of a specific gene or regulatory element without a direct effect on tumor cell fitness in the original anatomical compartment or progression stage. The resulting increase or decrease of physical genome redundancy may, nevertheless, determine the functional effect of a secondary mutation which arises later during tumor evolution and which targets an additional copy of the gene or regulatory element.

The two-hit process postulated by Knudson (Knudson, 1971) falls within this class of lineage selection effects. The first inactivating mutation results in a reduction of physical redundancy of the tumor suppressor gene RB. In case of hereditary retinoblastoma, the first-hit mutation is present in germline and found in all patient's cells. The mutation has no direct phenotypic effect as the second copy of the RB gene is sufficient for tumor suppressor function. If a second mutation targets the intact RB gene in a cell of the retina, tumor suppressor function is deleted and a tumor will arise. The observation that retinoblastoma arises early childhood in hereditary retinoblastoma suggests that the RB gene probably has no functional redundancy in cells of the retina.

A more complex situation arises, when a large DNA deletion reduces physical or functional genome redundancy of multiple genes. For example, one copy of a tumor suppressor gene could be co-deleted with one copy of a neighboring essential gene. Reduction of physical genome redundancy of the essential gene, which by definition has no functional redundancy, would display a restricting effect on subsequent tumor cell evolution as it restricts transition to a progression state with complete loss of tumor suppressor function to mutations which do not delete the remaining copy of the essential gene.

Inactivating mutations targeting all copies of one gene may result in loss of functional redundancy of a second gene implicated in the same essential biological pathway. This change in gene designation will restrict the spectrum of viable mutation in the derived cell lineage. Fig. 2 demonstrates an example of combined loss of physical and functional genome redundancy following deletion of both copies of a tumor suppressor gene.

Duplications or further amplifications of chromosomes or chromosome fragments may have an immediate effect on CCL fitness by enhancing the copy number of oncogenes. On the other hand, co-amplified sequences may contain genes which may act as tumor suppressor genes at later stages of tumor progression. Therefore, amplifications of genetic material may also display lineage selection effects.

Functional genome redundancy may not only be altered by changes of copy number of intact or mutated genes but also by changes in the designation of genes and regulatory elements. For example, a mutation which constitutively activates a protein within a signal transduction pathway consisting of multiple proteins and which is essential for cell survival would not only shorten the pathway but would also create an essential gene

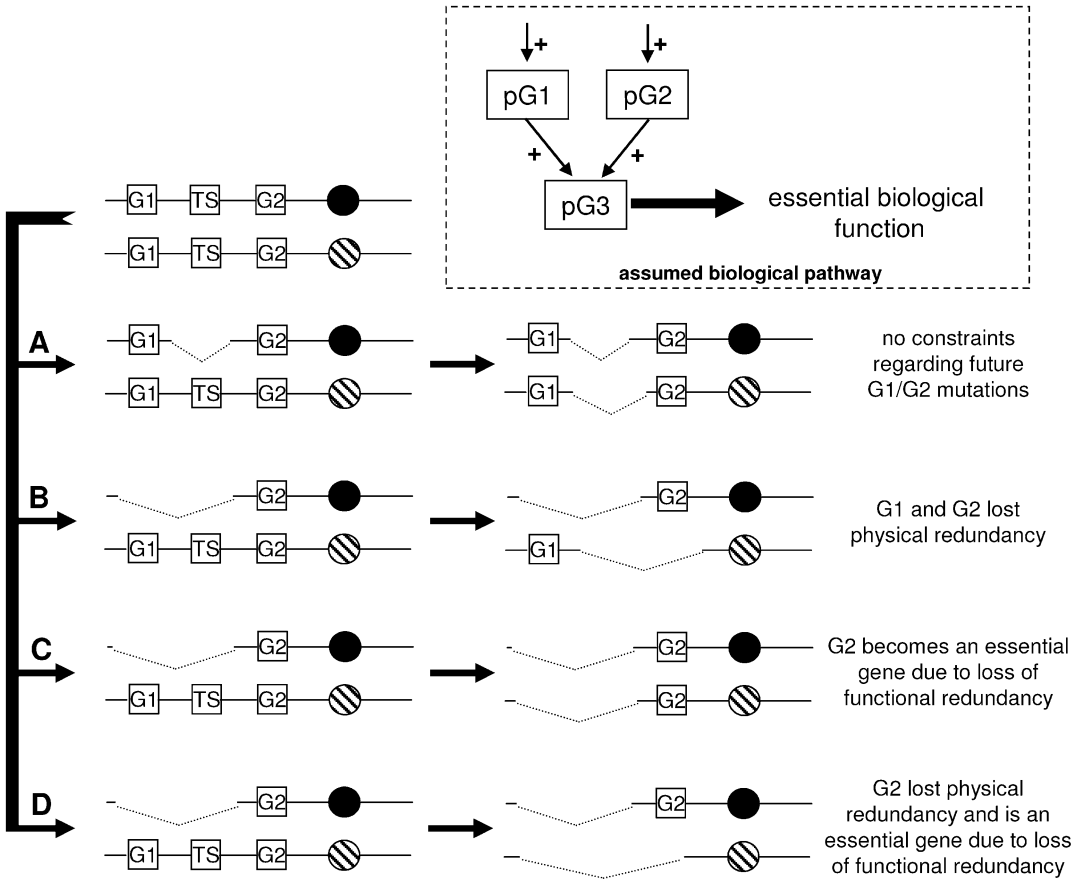

Fig. 2. Schematic drawing of two-hit deletions of tumor suppressor gene alleles (TS) which have different impacts on physical and functional redundancy of neighboring genes (G1 and G2). G1 and G2 shall substitute for induction of an essential gene (G3). Deletion sequence A has no effects on further mutations of G1 or G2. Deletion sequences B, C and D result in losses of physical and/or functional genome redundancy of G1 or G2 which affect the vulnerability of the cancer cell lineages towards further loss of function mutations of these genes.

which is only present as a single sequence. It would further change the designation of genes with functions upstream in the signal transduction pathway from essential to non-essential fitness modifying or fitness neutral genes. These genes may then be more readily deleted hemizygously or homozygously in genetically unstable CCL resulting in further reduction of functional genome redundancy. There are data suggesting that a reduction of redundancy in cancer cells might be found at cdk genes where aberrant expressions and amplifications are frequently detected (Berthet and Kaldis, 2007; Schmidt et al., 1994).

Besides shortening of signal transduction pathways, mutations may also induce rewiring of signal transduction pathways (Pawson and Warner, 2007) and remodelling of key biologic processes including alterations of the protein–protein, protein–DNA and RNA–DNA interactome with similar putative effects on multiple pathways and processes.

These simple examples of lineage selection effects certainly do not warrant the elaboration of a new systems approach of the carcinogenic process but more complex interactions of the genome architecture on lineage selection may be expected when modifications of the genome architecture have multiple and possibly opposed effects which manifest themselves at disparate times and progression stages. Complexity will also rise with the number and entanglement of involved genes and pathways.

It has been suggested that the internal dynamics of genomic modifications as opposed to the external evolutionary forces have a significant and complex impact on Darwinian species evolution (Dover, 2000). A similar situation can be expected for somatic cancer evolution as the key mechanisms encountered in species evolution such as duplications, rearrangements or deletions of genes (Demuth et al., 2006) also constitute prevalent mutation mechanisms in cancers with chromosomal instability.

Testable hypotheses resulting from the proposed systems approach

The human genome and the biological pathways regulated by the genome are highly complex due to the large number of genes and the various interactions of the multiple gene products. Complexity of the proposed systems approach further increases, as it does not only describe one single cell of cancer lineage with one genome architecture but tries to explain somatic cancer evolution with multiple cells of cancer lineage harbouring different genome architectures and evolving in different clinical progression states. It is evident, that description of somatic tumor cell evolution which comprises all known genes and regulatory

elements in cells of cancer lineage from a tumor cell population of realistic size is not feasible with present knowledge and technology. On the other hand, in order to validate the proposed systems approach one may deduce testable hypotheses which focus on isolated aspects and analyze few genetic elements.

Principal hypothesis

The principal hypothesis of the proposed systems approach is that permissive or restrictive effects of the genome architecture on lineage selection during somatic cancer evolution exist and have a measurable impact.

While it is becoming ever more evident that clinically distinguishable malignant tumors are characterized by specific patterns of chromosomal rearrangements (Curtin et al., 2005; Gunawan et al., 2007), it remains unsolved whether these chromosomal rearrangements can be explained solely by positive or negative selection through the cancer microenvironment or whether they are also stabilized by permissive or restrictive effects of the genome architecture on somatic cancer evolution as hypothesized within the proposed framework.

If the clinical course of cancer is considered a dynamic process regulated by biological needs and stochastic events which can be described mathematically, than knowledge of constraints of the tumor cell architecture on lineage selection during somatic tumor cell evolution might be used to modify tumor cell evolution as well as the clinical course of the malignant disease in individual cancer patients.

The principal hypothesis may be tested by a deductive approach in which detectable changes of the cancer genome architecture are explained by biological effects of the cancer genome architecture on somatic cancer evolution. This is based on the systems concept that the genome of CCL shapes the genome of CCL-daughter-cells but it necessitates sufficient knowledge of evolutionary pathways as well as of biological functions of involved genes or regulatory elements.

The hypothesis could also be tested by *in vitro* experiments using cancer cell lines. Cancer cell lines show relative karyotype stability after several passages (Fan and Li, 1987; Fischer et al., 1985; Boukamp et al., 1997) but may respond with karyotype changes to drug exposure (Tew et al., 1983). Experiments using well defined cancer cell lines and culture conditions might be conceived to dissect external selection effects of the culture media as experimental surrogate of the *in vivo* microenvironment from permissive or restrictive effects of the genome architecture.

In a recursive approach, additional hypotheses following from the proposed systems approach might be analyzed focusing on defined genes and chromosomal regions.

Hypotheses considering the sequence of mutation acquisition

An additional hypothesis of the proposed mechanistic systems approach is that the earliest changes within the genome architecture of a CCL induce the strongest constraints on somatic cancer evolution and that these constraints have a significant impact on the clinical and molecular behavior of individual malignant tumors arising in humans.

The designation of a gene as an oncogene, a tumor suppressor gene or an essential gene is highly dependent on cell differentiation status which may be deduced by its transcriptome and proteome. Therefore one may further postulate, that initial chromosomal gains and losses of a cell of cancer lineage should reflect the tissue origin of the CCL.

Within the proposed systems approach it is further hypothesized that constraints of the genome architecture are responsible in part for multilineage progression with an early split of cancer subclone lineages. This hypothesis derives from the hypothesis that mutations with the strongest constraints on somatic cancer evolution are likely to occur in early progression stages and may thereby define a stable and early most common ancestor.

Hypothesis of pseudo-cancer-stem-cells

During the process of somatic tumor evolution, a specific pattern of chromosomal losses and gains present in the majority of CCLs may restrict acquisition of chromosomal rearrangements essential for further tumor progression while specific mutations only present in a subset of all CCL may allow tumor progression through all stages. These CCLs with the full progression potential may imitate the behavior of cancer stem cells. One may hypothesize that a minor subset of a tumor cell population may behave like pseudo-cancer-stem-cells as their pattern of chromosomal gains and losses is permissive for further somatic tumor cell evolution while evolution of the predominating tumor cell clones is blocked by constraints of its genome architecture. Constraints of the genome architecture on somatic tumor cell evolution may, therefore, constitute an alternative explanation of stem cell-like behavior of cancer cells.

Hypotheses based on the physical arrangement of genes and regulatory elements as major restrictive factor in cells of cancer lineages with chromosomal instability

Assuming that loss of tumor suppressor gene functions by homozygous as well as hemizygous deletions are strongly selected during somatic cancer cell evolution in chromosomally unstable CCLs as demonstrated for CDKN2A in malignant melanoma and pancreatic cancer (Rübben et al., 2000; Fountain et al., 1992; Caldas et al., 1994), one may expect that size and localization of deletions in viable CCLs will be restricted by the presence of essential genes (as defined above) and oncogenes in the vicinity of the targeted tumor suppressor genes. In case of homozygous deletions of tumor suppressor genes, asymmetrical DNA losses extending to opposite directions on sister chromatids with small overlapping regions targeting only the tumor suppressor gene should prevail as this avoids complete and lethal loss of nearby essential genes and of active or potential oncogenes. In the absence of additional tumor suppressor genes in the vicinity of a strongly selected tumor suppressor gene, smaller deletions around this locus should equally be favored against larger deletions during somatic cancer evolution as minimizing hemizygous loss of essential genes might reduce the vulnerability of the CCL to further deletion mutations. One may also expect that deletions of chromosomal material occur more frequently within chromosomal regions with a lower density of essential genes while amplifications of chromosomal material should be found predominantly in regions not only harbouring oncogenes but also containing essential genes and fitness enhancing genes as amplification of essential genes in one progression stage reduces the probability of deleterious homozygous loss at subsequent cell divisions.

Hypotheses based on putative effects resulting from combinations of different forms of genetic instability

Assuming that loss of tumor suppressor gene function is strongly selected during somatic cancer cell evolution of CCLs with

chromosomal instability, one may expect that a pre-existing point mutation which inactivates one tumor suppressor gene copy reduces the above described restrictions on size and localization of a second hit deletion. Acquisition of microsatellite instability within a tumor cell subclone already displaying chromosomal instability might act in the same direction as the presence of multiple small mutations induced by microsatellite instability might favor further larger loss of chromosomal material. Abundant loss of heterozygosity could be demonstrated in a melanoma subclone with microsatellite as well as chromosomal instability (Rübben et al., 2000). One may hypothesize that CCLs with both forms of genetic instability acquire larger deletion mutations targeting tumor suppressor genes more rapidly due to initially reduced constraints of the genome architecture. On the other hand, only few malignant tumors seem to display both forms of genetic instability, which suggests that rapid loss of genetic material or genetic heterozygosity finally results in reduced cellular fitness which eliminates subclones displaying both mechanisms of genetic instability.

Hypotheses based on effects resulting from modifications of signal transduction pathways

As already described, shortening or rewiring of a signal transduction pathway essential for cell survival or malignant progression by an activating mutation or by a mutation leading to a functionally active fusion protein could favor deletion or inactivation of upstream genes and regulatory elements during somatic tumor cell evolution, which, in turn, may create restrictive effects at later progression stages. This hypothesis may easily be tested for activating mutations of oncogenes whose predominant roles in the malignant transformation of the malignancy has already been certified. On the other hand, the reverse assumption may also be true: the importance as well as suitability as therapeutic target of a mutation which results in shortening or rewiring of a signal transduction pathway should be reflected by the degree of deletions of upstream genes and of inactivation of redundant signal transduction pathways.

Hypotheses predicting tertiary therapeutic target genes resulting from constraints of the tumor cell architecture on somatic tumor cell evolution

Activated oncogenes which initiate malignant transformation of a specific malignancy may be considered primary therapeutic target genes in oncology. Highly efficient medicaments have already been developed which target cancer driving oncogene mutations in a number of defined human malignancies (Verweij et al., 2008). Fully inactivated tumor suppressor genes may be regarded as secondary therapeutic targets. Although abrogated tumor suppressor pathways represent stable cancer signatures in case of mutations of both tumor suppressor alleles, they are more difficult to exploit for pharmacologic or genetic therapies.

The proposed systems approach postulates a third group of target genes which do not belong to oncogenes or tumor suppressor genes but have become putative therapeutic targets due to constraints of the tumor cell architecture. This class of genes may comprise homozygously deleted genes which are frequently lost in specific cancers due to their physical vicinity to tumor suppressor genes as well as genes which have become essential genes in malignant tumors but not in normal tissue due to loss of redundancy within central cellular biochemical pathways.

The approach of using co-deleted genes as therapeutic targets in human cancers has been developed early and outside the proposed systems approach and already entered clinical trials (Chen et al., 1996; Capella and Caldas, 2005; Varshavsky, 2007; Kindler et al., 2009).

Nevertheless, there is a difference between choosing candidate genes only through analysis of copy number changes in cancer genomes and a search for therapeutic targets following the proposed systems approach. Within the proposed systems approach, the genome is considered dynamic and a search for putative genetic targets comprises the analysis of external and internal constraints which may stabilize the targeted genetic changes. Moreover, in order to exploit a homozygous gene deletion therapeutically, the mutation must not necessarily be present in the cancer cell population. Theoretically, it should be possible to target a cancer in which co-deletion of a gene or of a gene function associated with a more aggressive subclone is expected to arise with high probability due to constraints from external selection and tumor cell architecture on lineage selection. Such a cancer could be treated prophylactically with a compound that acts in cells with the expected mutation in order to prevent or greatly delay occurrence of the aggressive cancer subclone. One alternative therapeutic approach could be the introduction of a metabolic restriction by a drug which drives the tumor cell population into a newly rearranged genotype which then may not be able to evolve further into the more aggressive subclone due to restriction by the genome architecture of the selected genotype.

Algorithm for modifying cancer progression in human cancers

Large amounts of data on the human genome as well as on various human cancer genomes have been obtained in recent years by genome-wide sequencing, genome-wide expression analysis (DeRisi et al., 1996; Schena, 1996), proteome analysis, genome-wide copy number analysis as well as by combinations of these methods (Green et al., 2011). Computational approaches such as cluster analysis (Eisen et al., 1998) or self-organizing maps (Tamayo et al., 1999) have been applied in order to extract biological information from extensive data-points generated by these techniques. Patterns detected in gene expression or in copy number changes of specific cancers suggest that cancer genomes emerge and evolve through a multitude of different but still finite evolutionary pathways. One may expect that the restrictive effects of the genome architecture on lineage selection during somatic cancer evolution will be contained in these data-sets as well as in patterns and clusters generated by supervised or unsupervised data-mining algorithms using the available data.

Nevertheless, comparing genome-scale mutation data as well as transcriptome and proteome data of multiple cancers of the same entity might not be sufficient to identify target genes for anticancer therapy which do not belong to oncogenes or tumor suppressor genes but have become putative therapeutic targets due to constraints of the cancer cell genome architecture on somatic tumor evolution in individual patients. Several reasons may account for the limitations of an approach which extracts patterns from genome wide analysis of multiple cancers of one entity: pattern analysis will be more likely to detect genes which are overexpressed or homozygously deleted, i.e. the classical oncogenes or tumor suppressor genes, than genes, which show only low or temporarily restricted expression in a cancer or which are hemizygously deleted but nevertheless represent essential

genes for survival of the cancer. The magnitude of expression of a gene as well as the differential expression of a gene in cancer tissue does not *per se* indicate whether the gene and its expression are essential for survival of cancer cells. In addition, even if progression of cancer genomes might follow only a finite number of evolutionary pathways, it can be expected that many genes might become putative therapeutic targets only in few patients due to the random nature of the mutation process and the multitude of involved genes. Furthermore, it has to be taken into account, that genome-scale gene expression and gene copy number data are not obtained from single genomes but from cancer cell populations which are highly heterogeneous due to underlying genetic instability. Therefore, predominating gene expression patterns or copy number changes detected in cancer tissue of a specific progression stage might not be representative of the cancer cells which have the ability to progress to more advanced tumor stages and which should be targeted by therapy.

In order to overcome these limitations but still use the wealth of information generated by genome-scale molecular analysis techniques, an integrative approach is proposed to search for putative tertiary therapeutic targets in human cancers. It primarily focuses on the comparison of genome-wide molecular data obtained from cancer material of the same patient but of different progression stages as the constraints of the cancer genome architecture on somatic cancer evolution can best be detected when considering the phylogenetic tree of cancer subclones. In addition, construction of a phylogenetic tree of cancer subclones and their evolution allows eliminating confounding data-points generated by genetic instability and by admixture of cancer subclones which represent evolutionary dead ends.

The proposed algorithm for modifying somatic cancer progression (Fig. 3) consists of the following steps:

1. Acquisition of tumor material of a human cancer with multiple samples of an individual patient representing unaffected tissue, primary tumor as well as regional and distant metastases and construction of a map of clinical cancer progression in the patient based on the anatomical and temporal occurrence of the primary tumor and its metastases.
2. Extraction of DNA, RNA and proteins of unaffected tissue and of samples of the primary tumor as well as of regional and distant metastases from the patient representing clinical cancer progression.
3. Determination of the cancer genome architecture of the analyzed tissue samples by determination of RNA and protein expression, of karyotypes, of gene losses and gains as well as of gene allele losses in tumor tissues. Determination of the mechanisms of genetic instability based on these data.
4. Construction of a preliminary phylogenetic tree of somatic cancer evolution based on steps 2 and 3 by supervised or unsupervised clustering of the genetic data obtained by step 3.
5. Construction of a phylogenetic tree of somatic cancer evolution of identifiable cancer subclone populations by integration of data generated by steps 3 and 4. This recursive analysis is necessary as individual tissue samples may contain multiple cancer subclone populations with different genome architectures.
6. Establishment of stage and cancer subclone specific cancer genome maps. Cancer genome maps contain multilayered data on:

6.1. mutations of genes and regulatory sequences, copy number changes of genes and regulatory sequences, changes of chromosome structure and chromosome number
6.2. gene expression status
6.3. putative or ascertained protein function networks (interactomes) and their assumed or demonstrated modifications based on mutation status within the cancer genomes and based on published data
6.4. assumed mutation probabilities of genes and regulatory elements based on genetic data obtained by step 3
6.5. putative assignment of mutated or wild type genes or regulatory elements of cancer genomes at specific progression stages in individual patients as: essential genes, non-essential fitness modifying genes, fitness neutral genes, tumor suppressor genes/gatekeeper genes, oncogenes, caretaker genes. The assignment process consists of:
6.5.1. assignment of functions of mutated or wild type genes or regulatory elements by their putative functions derived from data obtained by step 6.3
6.5.2. assignment of functions of mutated or wild type genes or regulatory elements by identification of positive or negative selection of their physical presence in cancer genomes through the process of tumor progression in individual patients based on phylogenetic maps established by step 5 and by considering mutation probabilities determined by step 6.4
6.5.3. identification of mutations signatures which are indicative of the designation of genes or regulatory elements of cancer genomes at different progression stages; mutation signatures may, *inter alia*, comprise:
6.5.3.1. asymmetrical deletions on homologous chromosomes which limit homozygous loss of nearby genes. These genes are likely to represent essential genes or oncogenes
6.5.3.2. DNA deletions or loss of function mutations suggesting the presence of tumor suppressor genes
6.5.3.3. gain of function mutations or translocations, DNA duplications or lack of DNA deletions suggesting the presence of essential genes or oncogenes
6.5.4. identification of essential genes or oncogenes by considering loss or reduction of physical or functional genome redundancy based on data generated by steps 6.1 to 6.3
6.6. recursive modification of 6.3, if necessary, by embedment of data generated through steps 6.1 to 6.5.

7. Integration of knowledge generated through steps 5 and 6 in order to explain somatic cancer evolution in individual patients on the basis of genetic instability, external selection forces and the internal effects of the genome architecture on genes and regulatory elements.
8. Identification of putative genes or regulatory elements which may be targeted in order to kill cancer cells or to slow or to stop clinical cancer progression in the patient. Target genes or regulatory elements may be identified by:

8.1. their designation as essential genes or essential regulatory elements in cancer cells of the patient but not in normal tissue cells based on knowledge gained by steps 5 to 7
8.2. their demonstrable or deducible impact on the process of somatic cancer evolution in the analyzed patient based on the phylogenetic tree of somatic cancer evolution and on the constructed multilayered cancer genome maps. The

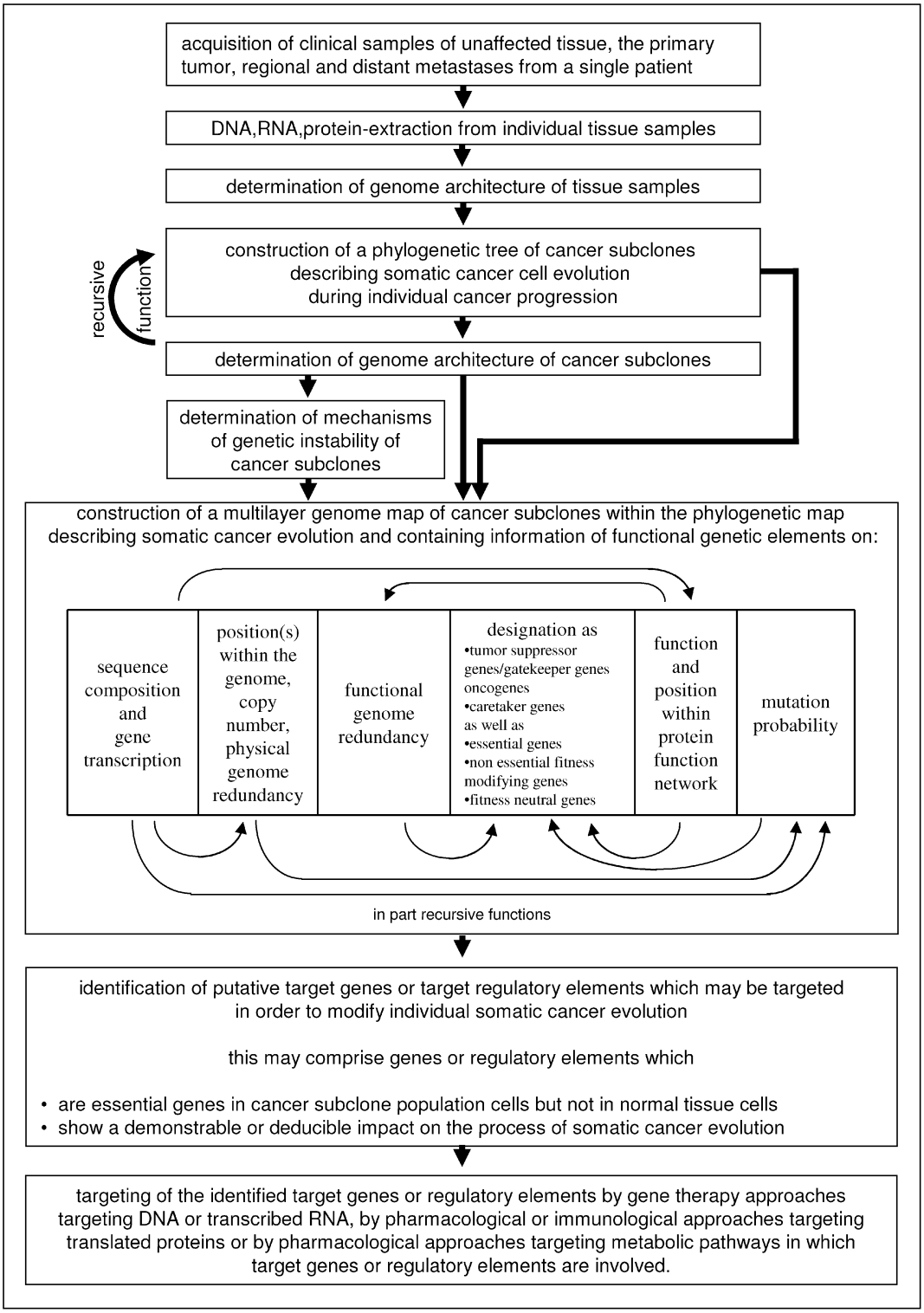

Fig. 3. Algorithm for analyzing and modifying somatic cancer progression in individual cancer patients.

impact on the process of somatic cancer evolution may be demonstrated by stabilization of the genome architecture despite a high mutation probability or by a permissive effect on further modification of the genome architecture during somatic cancer evolution.

9. *In silico* and *in vitro* evaluation of genes or regulatory elements identified by step 8 as potential targets of gene-therapeutic, immunological or pharmaceutical interventions designed to kill cancer cells or to slow or stop cancer progression during individual somatic cancer evolution.

10. Treatment of the patient with the most suitable gene therapy or immunological or pharmaceutical intervention in order to eliminate cancer cells or to slow or stop cancer progression during somatic cancer evolution.

Conclusion

A systems approach of carcinogenesis has been proposed which focuses on putative inner constraints of the genome architecture on lineage selection during somatic cancer evolution.

It postulates three classes of lineage selection effects of the genome architecture on somatic cancer evolution: i) effects mediated by changes of fitness of cells of cancer lineage, ii) effects mediated by changes of mutation probabilities and iii) effects mediated by changes of gene designation and by changes of physical and functional genome redundancy.

Clinical and molecular data as well as the analogy with the concept of evolvability in Darwinian evolution suggest that restrictive effects of the genome architecture on lineage selection during somatic cancer evolution exist and are measurable.

Testable hypotheses which focus on isolated aspects and analyze few genetic elements have been derived from the systems approach and an algorithm has been proposed which may be used to modify somatic cancer evolution for therapeutic purposes in individual patients.

In the future, substantial molecular and clinical data must be compiled to provide the necessary proof of concept of the systems approach and the hypotheses developed within the described systems concept.

Materials and Methods

As addition to the developed concept, the authors have designed a computer simulation of somatic cancer evolution of genetically unstable cells based on a very simplified mechanism of unequal sister chromatid exchange (USCE) in order to provide a crude impression of possible restrictive or permissive effects of the cancer genome architecture on cancer cell evolution. The simulation (CSim-2000 6.14) has been programmed under C++ and can be run under Linux. The program package of the simulation can be downloaded under the GNU Free Documentation License at http://ftp.bioinformatics.org/pub/canevolvability.

Unequal sister chromatid exchange proceeds through unequal pairing of sister chromatids, chromatid breakage and chromatid rejoining. USCE has been proposed as one possible mechanism of carcinogenesis (Killen et al., 2009) and USCE seems to be enhanced in cancer prone Bloom's syndrome (Pall, 1981; Sengupta et al., 2003). The aim of this simulation was to demonstrate the effects of the physical arrangement of genes as a major restrictive factor influencing tumor cell evolution in cells with chromosomal instability. USCE was chosen as mutation mechanism as it can be easily coded and because it leads to either duplications or to deletions of chromosomal DNA. According to the proposed systems concept, deletions and duplications exert strong lineage selection constraints on somatic tumor cell evolution. It is obvious that USCE does not reflect the multitude of mutation mechanisms observed in cancer cells with chromosomal instability.

The computer simulation starts with one cell with two chromosomes containing 16 genes each, four essential genes (E1 to E4), four neutral genes (N1 to N4), four oncogenes (O1 to O4) and four tumor suppressor genes (T1 to T4). The order of these genes on the chromosomes can be varied as well as the frequency of mutations, the maximum length of genes involved in USCE and the number of cell divisions.

The simulation only considers the end result of reciprocal USCE where one chromatid loses and the sister chromatid gains a contiguous stretch of genetic material (Fig. 4).

The simplified mutation process consists of four random steps. First a random generator selects one cell to mutate (Fig. 4, R1). After doubling of the chromosomes of this cell, a random generator chooses one chromatid to change (Fig. 4, R2). The random generator then picks one gene on the chromatid as first breakpoint (Fig. 4, R3) and finally the random generator selects a second breakpoint within a predefined length of genes (Fig. 4, R4). The chromatid segment defined by these two breakpoints is transferred to the sister chromatid resulting in a duplication on the sister chromatid and a deletion on the donor chromatid.

The computer simulation allows simulating a linear chromosome where the mutation probability of each gene is also dependent of its relative position on the chromatid or a ring chromosome with uniform mutation probability. In the linear chromosome model, the maximum number of transferred genes from the donor chromatid to the acceptor chromatid is dependent on the number of genes in telomere direction of the first breakpoint whereas in the ring chromosome model the maximum number of transferred genes is the total number of genes on the donor chromatid.

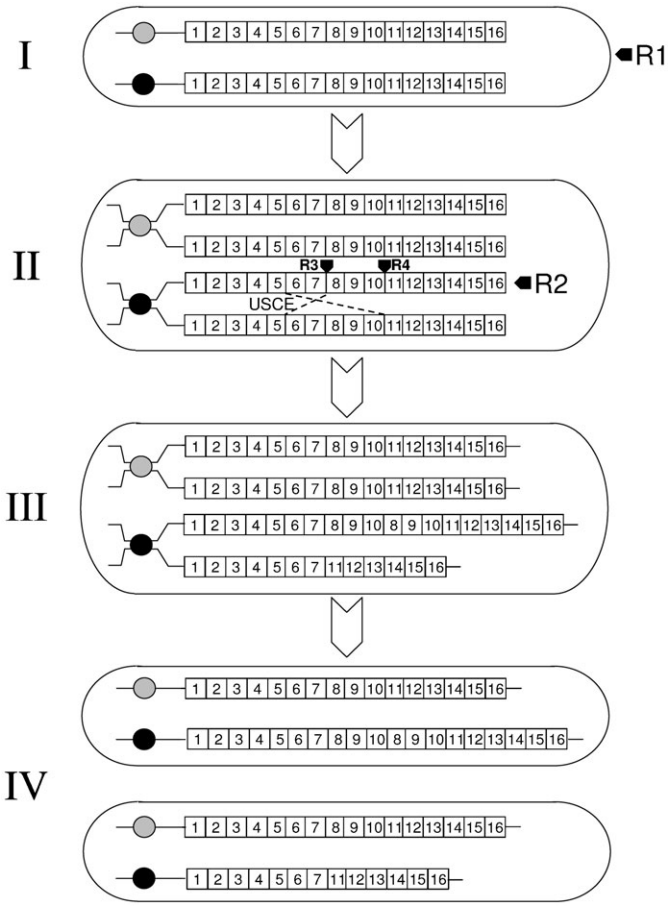

Fig. 4. Simplified simulation of unequal sister chromatid exchange (USCE) using four random events R1–R4 determined by random generator in a cell with one chromosome pair harbouring 16 genes on each chromatid. I = starting cell, II = DNA duplication, III = result of USCE, IV = daughter cells after cytokinesis.

The simulation differentiates between five stages. The simulation starts with cells in stage one. If one cell loses both copies of tumor suppressor gene one (T1) or both copies of tumor suppressor gene two (T2) it proceeds to stage two. If one cell acquires, in addition, an extra copy of oncogenes one or two (O1 or O2) it advances to stage three. With an additional loss of both copies of tumor suppressor genes three or four, the cell enters stage four and an additional extra copy of oncogenes three or four (O3 or O4) allows progression to the final stage five. Duplications or losses of neutral genes have no effects. If a cell loses both copies of one of the four essential genes E1 to E4, the cell is eliminated. The cell is also eliminated when the total number of genes exceeds 64.

Within the simulation, proliferation rates and mutation rates do not differ between stages. The only negative selection effect occurs through elimination of non-viable cells with either homozygous loss of essential genes or a total gene number exceeding 64.

The program registers the number of cells in each stage, the number of transitions between stages, the distribution of total gene number and the individual evolution history of the cell which first reaches the final stage.

A typical example of results which can be obtained with the simulation program is shown in Fig. 5. The simulations were performed with 40 cell divisions, a mutation frequency of 1:100 and ring chromosomes. Three different genomes, G1, G2, G3, as shown in Fig. 5, were used for the starting cells. For each genome, the simulation was repeated 20 times. The plot in Fig. 5 shows the number of cells in stages one to five for each simulation run and each starting genome. As each simulation represents a stochastic process, the number of cells in different stages differs for each individual simulation (data not shown).

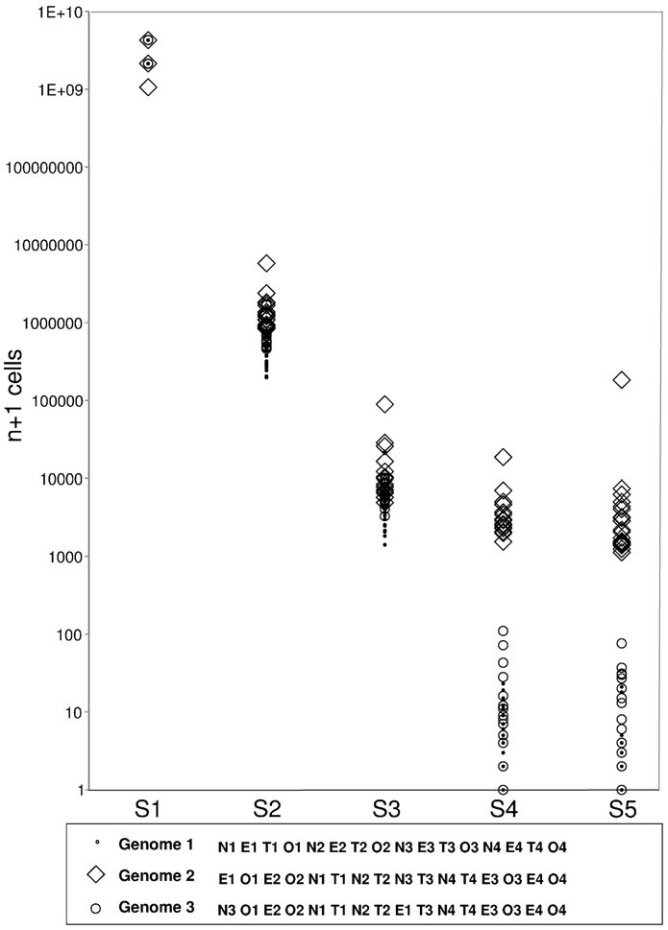

Fig. 5. Results of 20 simulation runs with three different starting genomes. Axis of ordinate shows number of cells in stages S1 to S5 after end of simulation.

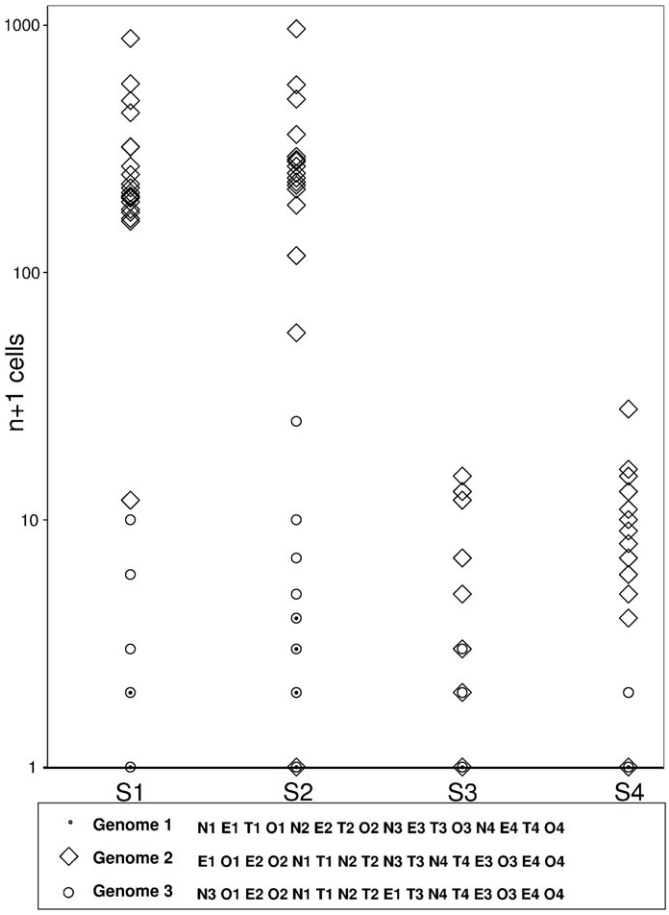

Fig. 6. Results of 20 simulation runs with three different starting genomes. Axis of ordinate shows number of cells which progressed directly to stage 5 from stages S1 to S4 during simulation.

Despite some overlap within individual results which represents the stochastic nature of the simulation process, the different starting genomes demonstrated strikingly different ability to progress to stage five. Genes in genome G1 were arranged as: N1E1T1O1N2E2T2O2N3E3T3O3N4E4T4O4. Only few cells evolved to stage five with the chosen parameters. In contrast, simulations starting with genome G2 progressed rapidly to the last stage. The numbers of cells in stage five were more than three orders of magnitude higher compared to simulations starting with genome G1. This is not surprising as oncogenes were grouped together with the essential genes while tumor suppressor genes were grouped with the neutral genes in genome G2: E1O1E2O2N1T1N2T2N3T3N4T4E3O3E4O4. This arrangement facilitated duplications of oncogenes and posed fewer restrictions to losses of oncogenes.

Genome G3 differed from genome G2 only by the switch of genes E1 and N3: N3O1E2O2N1T1N2T2E1T3N4T4E3O3E4O4. Thereby, the essential gene E1 was positioned between two tumor suppressor genes. Such a change of gene designation could reflect a therapeutic intervention. As a result, the ability to progress to stage five is greatly reduced in genome G3.

Fig. 6 displays the stages from where cells progress to stage five. In all simulations most cells reached stage five from stage one or two. Simulations with genome G2 displayed highest cell numbers and more cells reached stage five from stages three and four while no cells of these stages reached stage five in simulations with genome G1. Thus, all simulations demonstrated multilineage progression as hypothesized by the proposed systems concept. Nevertheless, it must be considered that selection is a major factor shaping the phylogenetic tree of somatic cancer evolution *in vivo* and that positive selection has been deliberately omitted in the simulation.

The simulation program behaves as postulated, but it certainly does not provide a proof that somatic cancer cell evolution in man is indeed strongly affected by restrictive effects of the genome architecture on lineage selection. Molecular and clinical data must deliver the necessary confirmations of the hypotheses developed within the described systems concept.

Acknowledgements

A.R. developed the concept, performed data analysis, and wrote the manuscript. O.N. designed the simulation program CSim-2000 6.14 together with Daniel Robertz. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing Interests

A.R. has applied for a patent based on key components of the described systems approach. O.N. has no competing interests to declare.

References

Aaltonen, L. A., Peltomäki, P., Mecklin, J. P., Järvinen, H., Jass, J. R., Green, J. S., Lynch, H. T., Watson, P., Tallqvist, G., Juhola, M. et al. (1994). Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res.* **54**, 1645-1648.

Andrew, S. E., Xu, X. S., Baross-Francis, A., Narayanan, L., Milhausen, K., Liskay, R. M., Jirik, F. R. and Glazer, P. M. (2000). Mutagenesis in PMS2- and MSH2-deficient mice indicates differential protection from transversions and frameshifts. *Carcinogenesis* **21**, 1291-1295.

Armitage, P. and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1-12.

Armitage, P. and Doll, R. (1957). A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br. J. Cancer* **11**, 161-169.

Bailey, S. M. and Bedford, J. S. (2006). Studies on chromosome aberration induction: what can they tell us about DNA repair? *DNA Repair (Amst.)* **5**, 1171-1181.

Bailey, S. M. and Murnane, J. P. (2006). Telomeres, chromosome instability and cancer. *Nucleic Acids Res.* **34**, 2408-2417.

Beckman, R. A. and Loeb, L. A. (2005a). Genetic instability in cancer: theory and experiment. *Semin. Cancer Biol.* **15**, 423-435.

Beckman, R. A. and Loeb, L. A. (2005b). Negative clonal selection in tumor evolution. *Genetics* **171**, 2123-2131.

Berthet, C. and Kaldis, P. (2007). Cell-specific responses to loss of cyclin-dependent kinases. *Oncogene* **26**, 4469-4477.

Birch, J. M., Alston, R. D., McNally, R. J., Evans, D. G., Kelsey, A. M., Harris, M., Eden, O. B. and Varley, J. M. (2001). Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* **20**, 4621-4628.

Boukamp, P., Popp, S., Altmeyer, S., Hülsen, A., Fasching, C., Cremer, T. and Fusenig, N. E. (1997). Sustained nontumorigenic phenotype correlates with a largely stable chromosome content during long-term culture of the human keratinocyte line HaCaT. *Genes Chromosomes Cancer* **19**, 201-214.

Boveri, T. (1914). *Zur Frage Der Entstehung Maligner Tumoren*. Jena: Gustav Fischer.

Brookfield, J. F. (2009). Evolution and evolvability: celebrating Darwin 200. *Biol. Lett.* **5**, 44-46.

Brumer, Y. and Shakhnovich, E. I. (2004). Importance of DNA repair in tumor suppression. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**, 061912.

Brumer, Y., Michor, F. and Shakhnovich, E. I. (2006). Genetic instability and the quasispecies model. *J. Theor. Biol.* **241**, 216-222.

Budd, G. E. (2006). On the origin and evolution of major morphological characters. *Biol. Rev. Camb. Philos. Soc.* **81**, 609-628.

Caldas, C., Hahn, S. A., da Costa, L. T., Redston, M. S., Schutte, M., Seymour, A. B., Weinstein, C. L., Hruban, R. H., Yeo, C. J. and Kern, S. E. (1994). Frequent somatic mutations and homozygous deletions of the p16 (*MTS1*) gene in pancreatic adenocarcinoma. *Nat. Genet.* **8**, 27-32.

Camps, J., Morales, C., Prat, E., Ribas, M., Capellà, G., Egozcue, J., Peinado, M. A. and Miró, R. (2004). Genetic evolution in colon cancer KM12 cells and metastatic derivates. *Int. J. Cancer* **110**, 869-874.

Capella, G. and Caldas, C. (2005). MTAP homozygous deletion: an Achilles heel of human cancers ready for clinical use? *Cancer Biol. Ther.* **4**, 347.

Chen, Z. H., Zhang, H. and Savarese, T. M. (1996). Gene deletion chemoselectivity: codeletion of the genes for $p16^{INK4}$, methylthioadenosine phosphorylase, and the α- and β-interferons in human pancreatic cell carcinoma lines and its implications for chemotherapy. *Cancer Res.* **56**, 1083-1090.

Cohn, D. E., Mutch, D. G., Herzog, T. J., Rader, J. S., Dintzis, S. M., Gersell, D. J., Todd, C. R. and Goodfellow, P. J. (2001). Genotypic and phenotypic progression in endometrial tumorigenesis: determining when defects in DNA mismatch repair and *KRAS2* occur. *Genes Chromosomes Cancer* **32**, 295-301.

Conrad, M. (1990). The geometry of evolution. *Biosystems* **24**, 61-81.

Curtin, J. A., Fridlyand, J., Kageshita, T., Patel, H. N., Busam, K. J., Kutzner, H., Cho, K. H., Aiba, S., Bröcker, E. B., LeBoit, P. E. et al. (2005). Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* **353**, 2135-2147.

Deans, B., Griffin, C. S., O'Regan, P., Jasin, M. and Thacker, J. (2003). Homologous recombination deficiency leads to profound genetic instability in cells derived from *Xrcc2*-knockout mice. *Cancer Res.* **63**, 8181-8187.

Demuth, J. P., De Bie, T., Stajich, J. E., Cristianini, N. and Hahn, M. W. (2006). The evolution of mammalian gene families. *PLoS ONE* **1**, e85.

Denver, D. R., Feinberg, S., Estes, S., Thomas, W. K. and Lynch, M. (2005). Mutation rates, spectra and hotspots in mismatch repair-deficient *Caenorhabditis elegans*. *Genetics* **170**, 107-113.

Denver, D. R., Feinberg, S., Steding, C., Durbin, M. and Lynch, M. (2006). The relative roles of three DNA repair pathways in preventing *Caenorhabditis elegans* mutation accumulation. *Genetics* **174**, 57-65.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. and Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457-460.

Dover, G. (2000). How genomic and developmental dynamics affect evolutionary processes. *Bioessays* **22**, 1153-1159.

Duesberg, P., Li, R., Fabarius, A. and Hehlmann, R. (2005). The chromosomal basis of cancer. *Cell. Oncol.* **27**, 293-318.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.

Fan, Y. S. and Li, P. (1987). Cytogenetic studies of four human lung adenocarcinoma cell lines. *Cancer Genet. Cytogenet.* **26**, 317-325.

Fearon, E. R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767.

Fischer, H., Schwechheimer, K., Heider, M., Bernhardt, S. and Zang, K. D. (1985). Establishment and characterization of a human glioblastoma cell line with a stable karyotype and nullisomy 13. *Cancer Genet. Cytogenet.* **17**, 257-268.

Foulds, L. (1954). The experimental study of tumor progression: a review. *Cancer Res.* **14**, 327-339.

Fountain, J. W., Karayiorgou, M., Ernstoff, M. S., Kirkwood, J. M., Vlock, D. R., Titus-Ernstoff, L., Bouchard, B., Vijayasaradhi, S., Houghton, A. N., Lahti, J. et al. (1992). Homozygous deletions within human chromosome band 9p21 in melanoma. *Proc. Natl. Acad. Sci. USA* **89**, 10557-10561.

Fujiwara, T., Stolker, J. M., Watanabe, T., Rashid, A., Longo, P., Eshleman, J. R., Booker, S., Lynch, H. T., Jass, J. R., Green, J. S. et al. (1998). Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. *Am. J. Pathol.* **153**, 1063-1078.

Gatenby, R. A. and Vincent, T. L. (2003). An evolutionary model of carcinogenesis. *Cancer Res.* **63**, 6212-6220.

Green, M. R., Aya-Bonilla, C., Gandhi, M. K., Lea, R. A., Wellwood, J., Wood, P., Marlton, P. and Griffiths, L. R. (2011). Integrative genomic profiling reveals conserved genetic mechanisms for tumorigenesis in common entities of non-Hodgkin's lymphoma. *Genes Chromosomes Cancer* **50**, 313-326.

Gunawan, B., von Heydebreck, A., Sander, B., Schulten, H. J., Haller, F., Langer, C., Armbrust, T., Bollmann, M., Gasparov, S., Kovac, D. et al. (2007). An oncogenetic tree model in gastrointestinal stromal tumours (GISTs) identifies different pathways of cytogenetic evolution with prognostic implications. *J. Pathol.* **211**, 463-470.

Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996). Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr. Biol.* **6**, 1185-1187.

Harfe, B. D. and Jinks-Robertson, S. (2000). Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae*. *Genetics* **156**, 571-578.

Iwasa, Y., Nowak, M. A. and Michor, F. (2006). Evolution of resistance during clonal expansion. *Genetics* **172**, 2557-2566.

Killen, M. W., Stults, D. M., Adachi, N., Hanakahi, L. and Pierce, A. J. (2009). Loss of Bloom syndrome protein destabilizes human gene cluster architecture. *Hum. Mol. Genet.* **18**, 3417-3428.

Kindler, H. L., Burris, H. A., 3rd, Sandler, A. B. and Oliff, I. A. (2009). A phase II multicenter study of L-alanosine, a potent inhibitor of adenine biosynthesis, in patients with MTAP-deficient cancer. *Invest. New Drugs* **27**, 75-81.

Kinzler, K. W. and Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761-763.

Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68**, 820-823.

Kuraguchi, M., Yang, K., Wong, E., Avdievich, E., Fan, K., Kolodner, R. D., Lipkin, M., Brown, A. M., Kucherlapati, R. and Edelmann, W. (2001). The distinct spectra of tumor-associated Apc mutations in mismatch repair-deficient Apc^{1638N} mice define the roles of MSH3 and MSH6 in DNA repair and intestinal tumorigenesis. *Cancer Res.* **61**, 7934-7942.

Little, M. P. and Wright, E. G. (2003). A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Math. Biosci.* **183**, 111-134.

Lobachev, K. S., Gordenin, D. A. and Resnick, M. A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**, 183-193.

Loeb, L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075-3079.

Loeb, L. A. and Christians, F. C. (1996). Multiple mutations in human cancers. *Mutat. Res.* **350**, 279-286.

Loeb, L. A., Springgate, C. F. and Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. *Cancer Res.* **34**, 2311-2321.

Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L. and Krush, A. J. (1966). Hereditary factors in cancer. Study of two large midwestern kindreds. *Arch. Intern. Med.* **117**, 206-212.

Marra, G. and Boland, C. R. (1995). Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J. Natl. Cancer Inst.* **87**, 1114-1125.

Marti, T. M., Mansour, A. A., Lehmann, E. and Fleck, O. (2003). Different frameshift mutation spectra in non-repetitive DNA of MutSα- and MutLα-deficient fission yeast cells. *DNA Repair (Amst.)* **2**, 571-580.

Marzi, I., D'Amico, M., Biagiotti, T., Giunti, S., Carbone, M. V., Fredducci, D., Wanke, E. and Olivotto, M. (2007). Purging of the neuroblastoma stem cell compartment and tumor regression on exposure to hypoxia or cytotoxic treatment. *Cancer Res.* **67**, 2402-2407.

Michor, F. (2005). Chromosomal instability and human cancer. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 631-635.

Moolgavkar, S. H. and Knudson, A. G., Jr. (1981). Mutation and cancer: a model for human carcinogenesis. *J. Natl. Cancer Inst.* **66**, 1037-1052.

Moolgavkar, S. H. and Luebeck, E. G. (1992). Multistage carcinogenesis: population-based model for colon cancer. *J. Natl. Cancer Inst.* **84**, 610-618.

Moolgavkar, S. H. and Venzon, D. J. (1979). Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Math. Biosci.* **47**, 55-77.

Nordling, C. O. (1953). A new theory on cancer-inducing mechanism. *Br. J. Cancer* **7**, 68-72.

Nowak, M. A., Michor, F., Komarova, N. L. and Iwasa, Y. (2004). Evolutionary dynamics of tumor suppressor gene inactivation. *Proc. Natl. Acad. Sci. USA* **101**, 10635-10638.

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science* **194**, 23-28.

Nunney, L. (1989). The maintenance of sex by group selection. *Evolution* **43**, 245-257.

Nunney, L. (1999). Lineage selection and the evolution of multistage carcinogenesis. *Proc. Biol. Sci.* **266**, 493-498.

Pall, M. L. (1981). Gene-amplification model of carcinogenesis. *Proc. Natl. Acad. Sci. USA* **78**, 2465-2468.

Pawson, T. and Warner, N. (2007). Oncogenic re-wiring of cellular signaling pathways. *Oncogene* **26**, 1268-1275.

Poole, A. M., Phillips, M. J. and Penny, D. (2003). Prokaryote and eukaryote evolvability. *Biosystems* **69**, 163-185.

Reya, T., Morrison, S. J., Clarke, M. F. and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111.

Rübben, A., Babilas, P., Baron, J. M., Hofheinz, A., Neis, M., Sels, F. and Sporkert, M. (2000). Analysis of tumor cell evolution in a melanoma: evidence of mutational and selective pressure for loss of $p16^{ink4}$ and for microsatellite instability. *J. Invest. Dermatol.* **114**, 14-20.

Rübben, A., Kempf, W., Kadin, M. E., Zimmermann, D. R. and Burg, G. (2004). Multilineage progression of genetically unstable tumor subclones in cutaneous T-cell lymphoma. *Exp. Dermatol.* **13**, 472-483.

Sabatier, L., Ricoul, M., Pottier, G. and Murnane, J. P. (2005). The loss of a single telomere can result in instability of multiple chromosomes in a human tumor cell line. *Mol. Cancer Res.* **3**, 139-150.

Saunders, W. (2005). Centrosomal amplification and spindle multipolarity in cancer cells. *Semin. Cancer Biol.* **15**, 25-32.

Schaaper, R. M. and Dunn, R. L. (1987). Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc. Natl. Acad. Sci. USA* **84**, 6220-6224.

Schena, M. (1996). Genome analysis with gene expression microarrays. *Bioessays* **18**, 427-431.

Schmidt, E. E., Ichimura, K., Reifenberger, G. and Collins, V. P. (1994). CDKN2 (p16/MTS1) gene deletion or CDK4 amplification occurs in the majority of glioblastomas. *Cancer Res.* **54**, 6321-6324.

Sengupta, S., Linke, S. P., Pedeux, R., Yang, Q., Farnsworth, J., Garfield, S. H., Valerie, K., Shay, J. W., Ellis, N. A., Wasylyk, B. et al. (2003). BLM helicase-dependent transport of p53 to sites of stalled DNA replication forks modulates homologous recombination. *EMBO J.* **22**, 1210-1222.

Shapiro, J. A. (1999). Genome system architecture and natural genetic engineering in evolution. *Ann. N. Y. Acad. Sci.* **870**, 23-35.

Shibata, D. (2012). Cancer. Heterogeneity and tumor history. *Science* **336**, 304-305.

Shibata, D., Schaeffer, J., Li, Z. H., Capella, G. and Perucho, M. (1993). Genetic heterogeneity of the c-K-ras locus in colorectal adenomas but not in adenocarcinomas. *J. Natl. Cancer Inst.* **85**, 1058-1063.

Sidransky, D. (1996). Is human *patched* the gatekeeper of common skin cancers? *Nat. Genet.* **14**, 7-8.

Snijders, A. M., Fridlyand, J., Mans, D. A., Segraves, R., Jain, A. N., Pinkel, D. and Albertson, D. G. (2003). Shaping of tumor and drug-resistant genomes by instability and selection. *Oncogene* **22**, 4370-4379.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. and Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 2907-2912.

Tew, K. D., Moy, B. C. and Hartley-Asp, B. (1983). Acquired drug resistance is accompanied by modification in the karyotype and nuclear matrix of a rat carcinoma cell line. *Exp. Cell Res.* **149**, 443-450.

Tomlinson, I. and Bodmer, W. (1999). Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nat. Med.* **5**, 11-12.

Trautmann, K., Terdiman, J. P., French, A. J., Roydasgupta, R., Sein, N., Kakar, S., Fridlyand, J., Snijders, A. M., Albertson, D. G., Thibodeau, S. N. et al. (2006). Chromosomal instability in microsatellite-unstable and stable colon cancer. *Clin. Cancer Res.* **12**, 6379-6385.

Traverso, G., Bettegowda, C., Kraus, J., Speicher, M. R., Kinzler, K. W., Vogelstein, B. and Lengauer, C. (2003). Hyper-recombination and genetic instability in BLM-deficient epithelial cells. *Cancer Res.* **63**, 8578-8581.

Tsao, J. L., Tavaré, S., Salovaara, R., Jass, J. R., Aaltonen, L. A. and Shibata, D. (1999). Colorectal adenoma and cancer divergence. Evidence of multilineage progression. *Am. J. Pathol.* **154**, 1815-1824.

Varshavsky, A. (2007). Targeting the absence: homozygous DNA deletions as immutable signposts for cancer therapy. *Proc. Natl. Acad. Sci. USA* **104**, 14935-14940.

Vauhkonen, H., Vauhkonen, M., Sajantila, A., Sipponen, P. and Knuutila, S. (2006). Characterizing genetically stable and unstable gastric cancers by microsatellites and array comparative genomic hybridization. *Cancer Genet. Cytogenet.* **170**, 133-139.

Verweij, J., Seynaeve, C. and Sleijfer, S. (2008). GIST as the model of paradigm shift towards targeted therapy of solid tumors: update and perspective on trial design. *Adv. Exp. Med. Biol.* **610**, 144-154.

Welch, D. R., Krizman, D. B. and Nicolson, G. L. (1984). Multiple phenotypic divergence of mammary adenocarcinoma cell clones. I. *In vitro* and *in vivo* properties. *Clin. Exp. Metastasis* **2**, 333-355.

Yan, B., Wang, H., Rabbani, Z. N., Zhao, Y., Li, W., Yuan, Y., Li, F., Dewhirst, M. W. and Li, C. Y. (2006). Tumor necrosis factor-α is a potent endogenous mutagen that promotes cellular transformation. *Cancer Res.* **66**, 11565-11570.

Yao, X., Buermeyer, A. B., Narayanan, L., Tran, D., Baker, S. M., Prolla, T. A., Glazer, P. M., Liskay, R. M. and Arnheim, N. (1999). Different mutator phenotypes in *Mlh1*- versus *Pms2*-deficient mice. *Proc. Natl. Acad. Sci. USA* **96**, 6850-6855.

Zhou, X. P., Loukola, A., Salovaara, R., Nystrom-Lahti, M., Peltomäki, P., de la Chapelle, A., Aaltonen, L. A. and Eng, C. (2002). *PTEN* mutational spectra, expression levels, and subcellular localization in microsatellite stable and unstable colorectal cancers. *Am. J. Pathol.* **161**, 439-447.

zur Hausen, H. (2000). Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J. Natl. Cancer Inst.* **92**, 690-698.