Perspectives

Mutation-selection balance and compensatory mechanisms in tumour evolution

Erez Persi [1,†], Yuri I. Wolf [1], David Horn [2], Eytan Ruppin [3], Francesca Demichelis [4,5], Robert A. Gatenby [6], Robert J. Gillies [7,†] and Eugene V. Koonin [1,†]

1 National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
2 School of Physics and Astronomy, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
3 Cancer Data Science Lab, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20894, USA
4 Department for Cellular, Computational and Integrative Biology, University of Trento, Trento 38123, Italy
5 Caryl and Israel Englander Institute for Precision Medicine, New York Presbyterian Hospital, Weill Cornell Medicine, New York , NY 10065, USA
6 Integrated Mathematical Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, 33612, USA
7 Department of Cancer Physiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA

† **To whom correspondence should be addressed.**
Correspondence to: **E.P. (**erezpersi@gmail.com), **R.J.G.** (Robert.Gillies@moffitt.org) or **E.V.K.** (koonin@ncbi.nlm.nih.gov).

Abstract

Intratumour heterogeneity and phenotypic plasticity, sustained by a range of somatic aberrations, as well as epigenetic and metabolic adaptations, are the principal mechanisms that enable cancers to resist treatment and survive under environmental stress. A comprehensive picture of the interplay between different somatic aberrations, from point mutations to whole-genome duplications, in tumour initiation and progression is lacking. We posit that different genomic aberrations generally exhibit a temporal order, shaped by a balance between the levels of mutations and selective pressures. Repeat instability emerges first, followed by larger aberrations, with compensatory roles leading to robust tumour fitness maintained throughout the tumour progression. A better understanding of the interplay between genetic aberrations, the microenvironment, and epigenetic and metabolic cellular states, is essential for early detection, prevention, and development of efficient therapeutic strategies.

[H1] Introduction

Despite the impressive recent progress and the development of new, promising therapeutic avenues, cancer remains a puzzling phenomenon, with limited ability to treat many cancer types [1]. The primary reason for this limited success is the enormous genetic heterogeneity and phenotypic plasticity of tumours, which is a major cause of resistance to treatment [2]. According to the genetic paradigm of tumour evolution, cancers evolve by the acquisition of driver mutations that confer selective advantage [3,4], as has been demonstrated by numerous studies [5-10]. Tumour evolution involves complex genomic aberrations that occur on every length scale including point mutations, complex insertions and deletions [11], microsatellite and repeat instability [12-15], gene copy number alterations, chromosomal instability and aneuploidy [16-22], massive genomic rearrangements [23,24], and whole-genome duplications [25]. Similar somatic clonal evolution also occurs in unicellular organisms [26,27] and human normal tissues [28-30], emphasizing that the impact of mutations on cellular fates is highly context specific.

Genomic aberrations arise randomly [31,32] but can be further driven and selected by local environmental conditions [33-35]. Different environmental exposures (e.g., UV radiation or smoking) can affect the generation and fate of mutations and leave distinctive signatures [36]. These aberrations generate intratumour heterogeneity, which allows cancer cells to cope with microenvironmental assaults, immune surveillance, and therapy, eventually resulting in tumour invasion and metastasis [37-41]. Although much is known about how different aberrations affect tumour progression and clinical outcome, it is less clear how they act jointly to maintain the fitness of tumours across time.

Beyond (and partially owing to) genome evolution, cancers evolve through phenotypic plasticity, mediated by metabolic adaptations and epigenetic modifications, which tightly interact with each other, sense and rapidly respond to changing environmental factors [42-44]. The role of phenotypic plasticity in cancer progression has been extensively described and further contributes to the ability of cancers to develop diverse strategies to proliferate under harsh and fluctuating conditions [35,45-47], promoting aggressive and invasive phenotypes [37,41,48]. However, how this plasticity is linked to the tumor mutational landscape is incompletely understood.

In this Perspectives article we aim to provide insights into the temporal dynamics of different aberrations, highlighting their apparent compensatory roles in maintaining tumour fitness during different phases of cancer evolution. Starting with the smallest scale of point mutations and progressively moving to larger aberrations, we present evidence of a specific temporal order in the acquisition of different types of genetic alterations, whereby repeat instability emerges early in tumour evolution, followed by larger aberrations. We examine the interplay between different types of aberrations in the context of population genetics, pointing to similarities between their dynamics and roles in species and cancer evolution. Additionally, we highlight how metabolism, epigenetics and environmental stress could be involved in tumour initiation and the induction of oncogenic events, providing further support to the emerging temporal pattern of different aberrations. Lastly, we discuss the impact of interactions between mutations (epistasis) as well as between mutations and the metabolic and epigenetic states of cancer cells. Such interactions lead to rich fitness landscapes, but also to vulnerabilities, as tumours develop dependencies on specific mutations and conditions. We discuss how interactions shape the evolutionary trajectories of tumours and can be exploited for developing more efficient therapeutic strategies, by generalizing the

concept of synthetic lethality. We conclude that further elucidation of the context- and time-dependent acquisition of mutations in different phases of tumour evolution and their interaction with the environment can be expected to improve our ability to identify and distinguish between features involved in cancer initiation and progression, and ultimately, to improve methods for early detection and the efficacy of combinatorial treatments in advanced cases.

[H1] Point mutations, selection and fitness

In population genetics, the two key variables that determine the evolutionary fates of evolving populations, including populations of cancer cells, are the effective population size (*Ne*) and the point mutation rate (μ) [39,49] (**BOX 1)**. μ determines the supply of point mutations (beneficial and deleterious), whereas *Ne* determines how likely is a mutation to propagate or be eliminated from the population. Large *Ne* entails dominance of purifying selection, whereby deleterious mutations are efficiently removed from the large population, which is the prevalent evolutionary regime in unicellular organisms, particularly prokaryotes. By contrast, small *Ne* translates into dominance of neutral evolution by genetic drift, whereby mutations are fixed in a population due to stochastic sampling fluctuations, which is the common regime of evolution in multicellular eukaryotes [50-52]. Across the diversity of life forms, the product *Ne* x μ is a marker of macroevolution, which delineates major clades: it exhibits high values in prokaryotes and gradually decreases with organismal size and complexity in unicellular eukaryotes and plants, reaching the lowest values in animals, particularly vertebrates [49]. Furthermore, when μ is high and *Ne* is small, many deleterious mutations can accumulate, leading to successive loss of the fittest genotypes, in a process known as Muller's ratchet [53]. In some cases, Muller's ratchet can lead to extinction of the population, a phenomenon called mutational meltdown [54]. Conversely, when μ is low and *Ne* is large, different members of the population can acquire similar beneficial mutations, leading to parallel evolution [55,56]. Between these extremes, for intermediate *Ne* and μ, a population can exhibit complex dynamics, through the interplay between clonal spatial organization, competition (clonal interference) and genetic interactions (epistasis) [39]. Therefore, the evolutionary regimes of species are often assessed by exploring the status of these two variables, referred to as the **mutation-selection diagram**. This diagram can be also constructed using the strength of selection, *dN/dS*, the ratio between non-synonymous and synonymous substitution rates, instead of *Ne*, because *Ne* and *dN/dS* are directly related [57,58] (**BOX 1)**.

Cancer is a unique testing ground for exploring the mutation-selection diagram because: different cancers exhibit a wide range of mutation rates [59] largely reflecting characteristics of different tissues [31,32]; the strength of selection can be assessed from patients' mutational profiles; and because tumour fitness (**BOX 1**) can be assessed using clinical data, assuming that poor outcomes reflect the capacity of cancer cells to proliferate, migrate, invade tissues and colonize. Although the mutation rate can vary greatly over space and time, the mutation load (ML) — that is the number of non-silent mutations in tumour genomes — can be used to assess the average mutation rate, provided the time interval from the birth of the neoplastic cell to the tumour state is known. Using this approach, data-driven population-genetics models of tumour evolution estimate that selection is typically weak [60]. Nonetheless, there exists a trade-off between driver mutations that enhance tumour fitness [61] and deleterious passenger mutations (**BOX 1**), reflecting a mutation-selection balance [62], such that the dependency of tumour fitness on ML is non-monotonic and

exhibits a critical state beyond which the number of deleterious mutations becomes substantial and decreases tumour fitness [63,64].

These properties are captured by pan-cancer data analysis (**Fig. 1**). First, analysis of the association between ML and tumour fitness (clinical outcome), using Cox regression analysis [65] (**BOX1**), reveals a dual regime [66], whereby (**i**) in low ML cancers, there is a positive correlation between ML and tumour fitness (i.e. the higher the ML the worse the clinical outcome), which is likely to reflect the impact of driver mutations, whereas (**ii**) in high ML cancers, the correlation between ML and tumour fitness is negative (a further increase in ML is associated with better outcome), apparently due to the impact of deleterious mutations. The damaging effect of deleterious passenger mutations is independent of immune response, as was established in experiments of human cell line and mouse models [67]. Second, most of the individual genes [68,69] and most tumour genomes on average [66,70] evolve close to neutrality ($dN/dS \approx 1$) although some subclones appear to be subject to substantial positive selection [71]. Nonetheless, deviations from neutrality exist, indicating (weak) positive selection in low ML tumours, probably reflecting the accumulation of driver mutations in initial phases, and substantial negative selection at extreme ML, probably reflecting purifying selection that prevents mutational meltdown of the tumour cell population [66,72,73]. These properties show how positive and negative selection maintain the fitness of tumours in initial and advanced phases, respectively. Consistent with this view, deviations from neutrality correspond to decreased tumour fitness as demonstrated by the association of neutrality with the worst clinical outcome [66].

Thus, the mutation-selection diagram of tumour evolution reveals a dual regime, with ML and *dN/dS* being the key variables that determine the fitness of tumours in different phases, with maximum fitness at intermediate ML and under neutral evolution. Next, we discuss the behaviour of other somatic aberrations in the context of this mutation-selection balance.

[H1] Microsatellite and repeat instability

The next length-scale of aberrations that can affect more than a single locus involves changes in short repetitive motifs (typically <100 bp) in coding and non-coding regions, hereafter referred to as 'repeats' (**Fig. 2**). Repeats display complex patterns, including tandem duplications, interspersed repeats, repetitive domains, and overrepresentation of motifs in low complexity regions. Repeats are hotspots of evolution [74-76] that emerge, primarily, through replication slippage and recombination [77-79]. Protein repeats serve as building blocks of various macromolecular complexes [80,81], and are involved in a variety of biological processes by mediating both protein–protein and protein–nucleic acid interactions, notably, transcription regulation, intracellular trafficking, immunity and stress-response [82-86].

Repeats play important roles both in microevolution and in speciation. In microevolution, in diverse life forms, variations in the copy-number of repeat units facilitate the acquisition of new functions and new phenotypic traits, for example, cell-adhesion in yeast allowing for immune evasion [87], skull and limb morphology in dogs [88], and regulation of circadian clock to changing temperature in flies [89]. Repeats are key regulators of gene expression [90] and RNA structure [91,92], as well as brain development [93-95] and social behaviour in mammals [96]. Analyses of species proteomes further unravel the dynamics and evolution of point mutations in repeat units. Such mutations have functional consequences, as exemplified by the rapid

evolution of the zinc-finger array of the *PRDM9* gene, which promotes meiotic recombination, and is directly implicated in mammalian speciation [97,98]. Nearly universally, following duplication, new repeats in protein-coding regions are relaxed of selective constraints, diverge by accumulating mutations at elevated rates under positive selection [99], and subsequently, become conserved as they acquire new functions [99,100] (**Fig. 2a**). This mutational susceptibility eventually results in repeat expansion and sequence divergence among the repeat units, such that the number of repetitive 'words' that compose repeat arrays (repeat vocabulary) in proteomes, which represents the compound effect of unit duplication and propagation of mutations, increases over long spans of evolution, with an inverse relation to the ordering of species by $Ne \times \mu$ [49,101,102] (**Fig. 2b)**. Thus, repeat propagation is affected by the strength of purifying selection, with low repeat abundance in unicellular organisms (high $Ne \times \mu$, i.e. strong purifying selection) and high abundance in multicellular eukaryotes (low $Ne \times \mu$, i.e. weak purifying selection and substantial genetic drift). Therefore, repeat copy number variation and divergence represent a distinct evolutionary mechanism that can rapidly generate genetic variability, and is likely to compensate for the relatively low mutation rates in multicellular eukaryotes.

The rapid evolution of repeats comes at the cost of promoting genetic diseases. Variation in the number of repeats (mostly expansion, but also contraction), hereafter referred to as repeat instability (RI), is implicated in a variety of neurological disorders, including Huntington disease, fragile X disorders, Friedrich ataxia, spinocerebellar ataxia, frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) [103-111]. Repeat instability in different regions of genes (i.e., introns, 5′ and 3′ untranslated regions (UTRs) and coding sequences) is associated with different diseases [105,109,110]. For example, the expansion of intronic TTTCA and GGGGCC repeats is associated with ALS and FTD, respectively[112,113], The most prevalent variation involves short tandem repeats of alanine and glutamine tracts, but other types and complex repeat variations are known, with unit sizes ranging from a few base pairs (microsatellites) to 10–1000 bp [105,108] and larger structural variations [109]. The extent of variation often correlates with disease severity [105,108], as the expansion of repeats leads to RNA toxicity, autophagy stress and epigenetic dysregulation (e.g., nucleosome assembly exclusion, or promotor-bound gene silencing) which promote disease [105,107,110].

RI is also implicated in cancer somatic evolution, which often exhibits aberrant DNA replication and repair that promote RI [114,115]. The best-studied type of RI implicated in cancer is microsatellite instability (MSI), partially due to technical difficulties inherent in the identification of larger repeats with current next-generation sequencing (NGS) technologies [116,117]. MSI is a widespread phenomenon affecting most cancer types [12-14], and shows distinct dynamics and roles in tumour evolution (**Fig. 2c**). At the pan-cancer level, across patients, MSI is inversely proportional to ML [14], and is a prognostic marker of favourable clinical outcome [13,118]. This indicates tolerance for RI at low mutational burden, but as cancer progresses and the burden increases, MSI is not tolerated. Indeed, generalizing from MSI to RI, by accounting for a wider range of repeats in the analysis of NGS data, we have recently shown that the inverse relationship between ML and RI is captured in individual patients [15]. Relative to healthy tissue, RI increased significantly in primary tumour samples, but then gradually reduced in metastases, whereas ML continued to increase during tumour progression to metastases, accurately capturing tumour phylogeny in individual patients (**Fig. 2c**). Thus, RI was high when ML was low (in primary samples) but was low when ML was high (in metastatic samples). Normal tissues adjacent to tumours exhibited similar RI to that of tumours, suggesting that RI might be induced by microenvironmental cues, even prior to the fixation of a driver,

oncogenic mutation and pathological evidence. We address the link between somatic aberrations, epigenetics and environmental stress in the dedicated section below.

These observations suggest a distinct, compensatory role of RI in cancer evolution, which can be explained in the context of the dual mutation-selection evolutionary regime. In the initial phase, RI is most pronounced, possibly compensating for the low number of cancer drivers and allowing the cells to cope with environmental stress. In this regime, ML is low, and RI is positively selected for. Later in evolution, as tumours acquire more mutations, ML increases and RI becomes a vulnerability to cancer cells, consistent with the association of high-MSI with better prognosis (i.e., decreased tumour fitness) [13]. In this regime, RI is reduced both by cell autonomous mechanisms associated with the high burden of deleterious mutations (*cf*. **Fig. 1**), and by the immune response to tumour neoantigens, which become prominent at high MLs [119-122]. The immune system then exercises purifying selection as demonstrated for microsatellite-unstable colon cancer, where strong purifying selection eliminated antigen-presenting tumours, whereas immune-adapted MSI tumours metastasized [123,124]. These findings also indicate that RI plays an adaptive role in immune evasion (reminiscent of the repeat variation in yeast mentioned above [87]) and metastasis formation.

In summary, the effects of selection on the propagation of mutations in repeats in species evolution (**Fig. 2a** and **2b**) and on the repeat copy-number in cancer evolution (**Fig. 2c**) are similar, demonstrating that repeats can propagate by drift under weak purifying selection but are eliminated by strong purifying selection. Positive selection, both in species (**Fig. 2a**) and in cancer evolution (**Fig. 2c**), drives repeat propagation in the initial phases and rapidly generates genetic and functional variability. This variability confers fast adaptability, possibly in response to environmental factors, and might be a universal rapid compensatory mechanism under evolutionary regimes of low mutation rates and environmental stress. The importance of point mutations in repeats in species evolution suggests that, beyond copy-number variations, mutations in new repeat copies may have functional implications in cancer, as has been demonstrated in colon cancer [125] but remains largely unexplored.

[H1] Larger genomic aberrations

Next, we address the impact of large genomic aberrations, such as aneuploidy and chromosomal instability (CIN), manifested by whole-arm gains or losses, as well as by focal somatic copy number alterations (SCNA) and whole-genome duplications (WGD). What is the role of these large genomic aberrations in maintaining tumour fitness, and how do they act through different phases and contexts of tumour evolution? In addressing these questions, we emphasize the temporal behaviours of different aberrations and their effects on clinical outcome (**Fig. 3**).

Aneuploidy is the most common genomic aberration in cancer that affects nearly 90% of cancers, in a highly tissue-specific manner, and covers about 25% of a typical tumour genome [18-22]. In species evolution, across diverse life forms including humans, aneuploidy is strongly deleterious, presumably as a consequence of a highly imbalanced gene expression [22]. Nonetheless, its effects on the fitness of cells are strongly context-specific. Under physiological conditions, aneuploidy is mostly detrimental, but under selective pressures, it can be beneficial, in particular, conferring drug resistance by increasing karyotypic and phenotypic diversity [19,20,22]. This dual effect of aneuploidy has been demonstrated in yeast under

normal [126] and stress [127] conditions. Consistently, in cancer, aneuploidy can act either as a tumour driver or as a suppressor [128,129]. Nonetheless, recent studies indicate that CIN-induced aneuploidy is mostly a driver of tumour evolution, which correlates with tumour progression [130] and directly drives metastasis formation via activation of the cGAS–STING DNA-sensing pathway, in response to DNA spilled by aneuploid cells into the cytosol, promoting cellular invasion [131]. Aneuploidy is a hallmark of late stages in tumour evolution, appearing at increasingly high rates as a tumour progresses, possibly marking the transformation to an invasive phenotype [22,132]. Indeed, the association of aneuploidy with clinical outcome further supports its dominant role as a tumour driver: in the vast majority of cancers, aneuploidy is associated with adverse clinical outcome, except for a few cancer types, notably haematopoietic malignancies [22,133]. Most cancer types show a positive association between driver mutations and aneuploidy, which appear to synergistically contribute to poor clinical outcome [134]. Nonetheless, extreme levels of aneuploidy appear to decrease tumour fitness, thereby leading to improved clinical outcome [135-137], presumably due to the increased genomic burden in the tumour cells. Indeed, excess levels of rates of chromosomal mis-segregation lead to tumour suppression and cell death [138]. Nonetheless, in tumours, CIN levels can be regulated through mutations in anaphase-promoting complex/cyclosome APC/C, dampening the effect of aneuploidy [139]. Thus, similarly to the case of ML, there seems to be an optimal aneuploidy level for tumour progression, but the critical level is not easily crossed during tumour evolution, and is likely to depend on the distribution of cancer genes affected by aneuploidy [140].

At the pan-cancer level, aneuploidy is significantly positively correlated with ML and inversely correlated with MSI [21]. Consistently, and opposite to the adverse effects of high aneuploidy on patient survival, high MSI is strongly associated with better clinical outcome [13]. These findings further strengthen the conclusion that MSI, and more generally, RI, contributes to tumour fitness at early stages, whereas later in tumour evolution aneuploidy becomes the dominant mechanism. Indeed, RI and MSI levels are low in high ML tumour genomes [14], and are significantly reduced in the transition to metastatic states [15], presumably due to immune surveillance [123]. By contrast, aneuploidy negatively correlates with immune signatures, apparently, being tolerated as cancers evolve mechanisms to evade immune control, but often positively correlates with proliferation and with the transition to metastatic states [21,22,131,132,141]. Recent observations further identify multinucleated neoplastic cells as a group of cells with unique capabilities for conferring renewal potential and coping with stress and therapy, leading to aggressive phenotypes [142,143].

Taken together, these findings propose possible complementary roles of ML, RI/MSI, SCNA and aneuploidy in tumour evolution (**Fig. 3)**. Specifically, RI (and MSI) appear to play an important role in tumour initiation and achieves optimal levels at early stages, whereas aneuploidy is likely to become important later, with optimal effects observed at extreme levels. In late stages of tumour evolution, high levels of RI (MSI) are not tolerated, and aneuploidy assumes the dominant role as a driver. Furthermore, the association of SCNA (focal and broad) with the clinical outcome shows that medium levels of SCNA are associated with the worst outcome, whereas low and high levels are associated with better prognosis [144]. Thus, focal SCNA appear to be associated with high tumour fitness, whereas the addition of CIN, once it leads to a high genomic burden, could reduce tumour fitness. A recent study [145] has shown that focal SCNA indeed portents worse prognosis than broad SCNA, in support of the observed trend, even if focal SCNA affect only about 10% of a typical cancer genome [18]. Thus, focal SCNA seem to reach the

optimal level for tumour fitness at intermediate stages, but are likely to remain high and contribute to tumour fitness throughout the course of progression (**Fig. 3**).

These findings suggest ordering of the types of genomics changes by time and the increasing length scale of the aberrations along the course of tumour progression: first, RI (MSI), at the short length scale, is a major contributor at early stages; second, focal SCNA, at the intermediate length scale, become important at intermediate stages, as RI declines; and third, aneuploidy, at the longest length scale, dominates late in tumour evolution (**Fig. 3**). Along the same lines, differences between the evolution of gene and repeat copies following duplication are pertinent also in species evolution. Whereas repeat copies evolve rapidly, under combined effects of positive selection and strongly relaxed purifying selection, presumably driving the evolution of new functions via neofunctionalization [99], gene duplicates typically evolve much slower, under weak, relaxed purifying selection, and are mostly implicated in subfunctionalization of ancestral genes [146-148]. This difference could, in part, stem from the different length scales of the affected sequences, whereby new repeats can be strongly affected by a single mutation, whereas new genes are less susceptible to mutations.

Finally, the largest aberration is WGD, a common phenomenon in cancers, which affects about 30% of patients with advanced cancer. WGD occurs relatively early in tumour evolution and accumulates in cancer cells in a manner that is positively correlated with tumour proliferation, and with poor clinical outcome [25]. Bielski et al.[25] have found that WGD occurs only following the acquisition of specific driver mutations, which lead to either p53 dysfunction or E2F-mediated cell-cycle arrest malfunction, but precedes SCNA, hence WGD possibly serves as a precursor to facilitate SCNA. WGD was also similarly frequent in primary tumours and metastases, with some cancers (e.g., pancreatic, prostate, and non-small cell lung cancers) exhibiting higher WGD frequency in the metastatic state. WGD could promote tolerance to aneuploidy [149], as indicated by the elevated levels of aneuploidy and chromosomal instability in tumours exhibiting WGD, consistent with its role as a precursor to SCNA. Conceivably, WGD dampens the unbalancing effect of aneuploidy, thereby creating aneuploidy-permissive conditions [22]. Furthermore, in the context of the mutation-selection balance, the observation that the deleterious effects of mutations can be mitigated by ploidy in asexual reproduction [150] further indicates that large somatic aberrations play a role in evolution when ML becomes substantial, reducing the risk of mutational meltdown.

The temporal order depicted in **Fig. 3** represents only an average or a common evolutionary scenario of tumour progression, from which many deviations exist. Arguably, however, some of these exceptional cases can be explained within the framework developed here. The context-dependent recurrences and fates of mutations reviewed here imply that, beyond this consensus order, a variety of genomic aberrations can be induced and become a dominant driver under specific molecular and environmental conditions, in a certain tumour type. Clearly, some large-scale structural genomic events are crucial for the initiation of certain cancers, such as TEL-AML1 fusions in acute lymphoblastic leukaemia (ALL)[151], gains of specific chromosomes in glioblastomas [152], and chromothripsis in childhood medulloblastomas, kidney and other cancers [153-155]. Such early events are distinct from the phenomenon of CIN discussed above which occurs, mostly, in late phases of tumour evolution. In Barrett's esophagus, although aneuploidy is clearly a late event, it preceded transformation to invasive phenotypes [156], and CNA rates increased prior to WGD and were maintained, indicating that WGD was not a precursor for CIN in this

case [157]. Furthermore, some tumours can harbour enormous numbers of mutations, as in the case of children born with biallelic mismatch repair (MMR) deficiency that acquire a POLE/POLD1 hypermutator phenotype in some central nervous system (CNS) and gastrointestinal (GI) cancers [14]. These tumours are associated with poor clinical outcome, suggesting that high ML does not reduce tumour fitness in this case (*cf*. **Fig 1**). In GI tumours, unlike most cancers, aneuploidy, but not the number of driver mutations in cancer genes, has been shown to uniquely correlate with poor clinical outcome [134]. Thus, in these cases aneuploidy appears to be the dominant driver, which may explain the tolerance for hypermutators, in a manner reminiscent of the effect of WGD on mitigating the risk of a mutational meltdown [73].

[H1] Phenotypic plasticity

Beyond genome evolution, tumours evolve through phenotypic plasticity, which involves extensive metabolic adaptations and epigenetic modifications. The metabolic and epigenetic circuits interact with each other, enabling cells to sense and respond to environmental cues, adapt to harsh conditions, determine cell fates and embrace diverse strategies to survive [35,43,44,46,158]. How might these metabolic and epigenetic changes be linked to the induction of genomic aberrations in cancer cells, beyond the baseline stochastic variability of normal cells? How do these complex interactions affect the fitness landscape of tumours, and can such effects be exploited for therapeutics?

Metabolic adaptations are universal hallmarks of cancer [37] that emerge as a series of adaptive strategies, ordered in time, that help tumours to overcome physiological and microenvironmental barriers [45]. Adaptation to loss of basement membrane induced apoptosis marks the transition to hyperplasia, followed by adaptations to nutrient deprivation, such as insufficient growth factors and oxygen availability (hypoxia) as cells proliferate into the lumen [45]. At these early stages, adaptation to hypoxia is accompanied by a switch to aerobic glycolysis, known as the 'Warburg effect' [159], which supports tumour cell proliferation [160] and promotes the selection of acid-producing cells [161]. Consequently, the tumour microenvironment becomes not only hypoxic by also acidic, with concomitant transition of the intracellular environment to a redox-stressed, alkaline state [45,162,163]. Adaptation to these harsh conditions confers invasive capabilities through diverse mechanisms [35,42,45], including suppression of immune surveillance [164,165] and proliferation of aggressive tumour phenotypes [166]. Eventually, to invade and metastasize, cells must adapt to fluctuations in the blood flow and develop supporting vasculature [41].

The metabolic changes in cancer impose stress on cells, which in turn can induce mutations that increase genetic variability and hence facilitate adaptation [35], a fundamental factor of evolution under stress in all organisms [167,168]. This causality could be mediated by epigenetic signals. In unicellular organisms, stress-induced mutagenesis mechanisms employ SOS-response [169] and subsequent error-prone DNA repair [170,171] which promote mutations. These pathways have parallels in cancer [172]. More generally, stress can induce mutations through diverse mechanisms that affect the DNA repair machinery, such as excess levels of reactive oxygen species (ROS)[35]. Stress can be transduced by epigenetic signals, via the production of a variety of metabolites (e.g., acetyl-CoA and α-ketoglutarate) that regulate enzymatic activity in the nucleus and are essential for epigenetic imprinting [44]. Consequently, some genes in the epigenetic machinery, known as mediators, can confer renewal potential to tumour cells (i.e., cancer stem cells) in early stages of tumorigenesis, even prior to the emergence of driver mutations [43] (e.g., loss of imprinting

of *IGF2* or aberrant WNT/β-catenin signalling [173]). Other genes comprising the epigenetic machinery, such as those involved in chromatin remodelling, DNA methylation, and histone modification (e.g., *PRDM9*, *ARID*, *DNMT* and *HDA* genes), and/or those acting further upstream, sensing and responding to cellular environmental factors (e.g., *APC*, *KRAS*, *STAT3* and *CTCF*), are tightly involved in cancer progression and are often mutated [43].

What is the correspondence between the temporal order of different types of metabolic stress in cancer and the temporal order of somatic aberrations? Stress can induce diverse types of aberrations. For example, in colorectal cancer cell lines, glucose deprivation can select for cells with an oncogenic KRAS mutation [174]. In murine tumour cells, hypoxia can induce SCNA and chromosomal rearrangements [175]. Critically, however, among all types of genomic aberrations (*cf.* **Fig 3**), RI appears to be the most susceptible to hypoxia and oxidative stress. This effect has been demonstrated in *Escherichia coli*, following DNA oxidative damage [176], and in human cell-line experiments, under microenvironmental stress conditions that mimic the conditions in cancer, such as hypoxia and oxidative stress [33,177]. Mechanistically, under hypoxia, the transcription factor HIF1α induces MSI by downregulating MutSα [34]. Because hypoxia and oxidative stress occur early in tumorigenesis, this link implies a primary role of RI as a mechanism of escape from stress in early stages.

Less is known about potential effects of stress on larger aberrations, such as aneuploidy. Aneuploidy is believed to be mainly mediated by mutations in genes that control replication fidelity and chromosome segregation [178]. Once aneuploidy emerges, it can cause epigenetic changes [179,180], but a causal relationship between epigenetic changes and aneuploidy in tumour evolution remains uncertain. Some studies have suggested that aneuploidy can be induced by stress factors, such as exposure to carcinogens (e.g., cadmium, nickel and arsenite), which cause DNA methylation in DNMT genes and subsequent chromosomal instability [181]. In yeast, under pH and heat stress, aneuploidy is transiently induced, and then, vanishes once the stress is removed [182,183]. Indeed, acidosis, a relatively late condition in cancer, appears to be involved in the induction of large aberrations, such as chromosomal breaks and translocations, but the exact mechanisms remain obscure [35]. Generally, these findings are consistent with aneuploidy being a mechanism that confers fitness advantage under stress in late stages of tumorigenesis.

The entirety of these observations indicates an important role of stress-induced mutagenesis in cancer that is consistent with the temporal order of mutations (**Fig. 3**), whereby RI emerges first in response to stress, along with other oncogenic events, whereas larger aberrations follow, presumably as mutations accumulate and the environmental conditions change. Moreover, these observations suggest that RI could be involved in the onset of cancer, even prior to fixation of oncogenic events, via stress-induced signals, which affect the epigenetic machinery and subsequently promote RI. This causality appears plausible, given that metabolic stress typically occurs early in carcinoma *in situ*, and even in precancer ducts [45,161,184], and in light of the intimate dependence of RI on DNA metabolic processes [105,107] and the induction of RI through epigenetic mechanisms, as exemplified in the cases of mutated CTCF binding regulatory factor and the DNA methyltransferase 1 (DNMT1) deficiency that are implicated in genomic imprinting and chromatin remodelling [185-187]. This concept accounts for the observation that RI is also manifested in tissues adjacent to tumours (*cf.* **Fig. 2c**), which are exposed to and sense similar environmental stresses as those that affect cancer cells in the tumour microenvironment. Furthermore, environmental stress that develops during the lifetime of an individual, prior to cancer onset, is likely to

contribute to cancer initiation, which could explain the notable propensity for cancer in individuals with obesity and type 2 diabetes [188].

Finally, it is important to consider the impact of phenotypic adaptations on the fitness landscape of tumours. Phenotypic adaptation to temporal changes in environmental conditions is a fundamental driving force of organismal evolution [189]. Temporal environmental changes affect the probability of fixation of mutations and the efficacy of selection that acts on them [190]. In the case of cancer cells, this dependency can blur the difference between driver and passenger mutations, highlighting the context-dependent nature of mutations [191]. Fluctuations also promote evolution of generalist traits, for example, when tumours adapt to blood flow changes, eventually evolving towards invasiveness [41]. Generalist traits evolving through the tumour history might underlie its capacity to metastasize and explain the difficulties in treating metastases [192].

Adding to this complexity is the high prevalence of genetic interactions between mutations (epistasis) in cancer [193]. Epistasis, the different fitness effect of co-occurring mutations compared to the sum of the effects of the corresponding individual mutations, leads to complex, rugged fitness landscapes [194,195]. Thus, tumours likely exhibit a more complex fitness landscape than the simple dual regime of low and high mutational burdens that is captured by the current analyses and population-genetic models of tumour evolution (*cf.* **Fig. 1**). This rich landscape likely reflects vulnerabilities that can be exploited for therapeutics, in the setting of synthetic lethality, where two events together are deleterious, whereas each of them alone is not [196,197]. Notable examples include synthetic lethality between mutations in the BRCA oncogenes and poly(ADP-ribose) polymerase (PARP) inhibitors [198], whereby PARP inhibitors and BRCA deficiency (leading to impaired DNA repair), combined, push tumours into the regime of high levels of instability, associated with good prognosis; and, between mutations in the ARID1A oncogene, a chromatin remodeller, which leads to aggressive ovarian cancer, and EZH2 methyltransferase [199]. Synthetic lethal pairs exist also between different types of somatic aberrations, as in the case of WRN helicase and MSI in several cancers [200]. In this case, DNA mismatch repair leading to MSI is essential for cancer progression, whereas depletion of WRN, a RecQ-like DNA helicase, selectively hampers those MSI cancers.

The concept of synthetic lethality can be further expanded to interactions not only between mutations, but also between the metabolic state of cells and their mutational profile. For example, cancer cells show elevated production of reactive oxygen species (ROS) which promotes tumour progression, but also represents a therapeutic opportunity because excess of ROS can trigger apoptosis and/or increase the mutation rate, thereby pushing cancers towards the Muller's ratchet regime and increasing the likelihood of a mutational meltdown [201]. Thus, gene targets whose knockdown increases ROS production might be synthetic lethal pairs for high ROS cancers. Another example is the identification of anti-cancer metabolic targets with amplified effects at low intracellular pH, forming an effective synthetic pair between gene targets (e.g., GAPDH) and the intracellular state of cells to combat aggressive phenotypes [166]. Thus, a better understanding of the interactions in different phases of tumour evolution, not only between mutations, but also between mutations and phenotypic perturbations, could lead to new, efficient therapeutic approaches.

[H1] Conclusions and future directions

The key parameters that determine the evolutionary fate of evolving populations are mutation and selection. Both population genetic theory of tumour evolution and empirical observations support the existence of a dual regime, whereby the accumulation of driver mutations promotes tumour evolution in the initial phases, with signatures of positive selection, but as the number of deleterious passenger mutations increases, cancer fitness drops, exacerbated by immune response, with a signature of purifying selection on the tumour genome. This dual mutation-selection regime appears to be a major determinant of tumour evolution that shapes the evolutionary trajectories of tumours, resulting in a distinct temporal order in the accumulation of different somatic aberrations, with compensatory interactions leading to robust tumour fitness (**Fig. 3**). Specifically, RI appears to be an early event, presumably, induced by environmental stress, and even preceding the appearance of driver mutations, whereas focal CNA and aneuploidy mainly gain prominence in intermediate and late phases of tumour progression, respectively. The contrasting associations of these aberrations with clinical outcomes further support their distinct roles in tumour evolution, and may suggest how cancers remain fit during different phases of evolution, by embarrassing diverse mechanisms. The validity of this picture of tumour evolution and the extent of variability across patients and cancer types should be further investigated by tracking the evolution of tumours in individual patients. Future studies making use of liquid biopsies to longitudinally follow individual tumour trajectories in patients might provide the necessary comprehensive data to address those questions on a variety of tumour types. Understanding the dynamics of the relationship between different aberrations and elucidating how environmental factors might induce these aberrations should provide mechanistic insights into the factors that contribute to cancer initiation and progression.

Beyond genome evolution, cancer cells evolve by epigenetic and metabolic adaptations to microenvironmental conditions which promote phenotypic plasticity. Both genetic variability and phenotypic plasticity contribute to cancer proliferation, survival and the ability to invade, metastasize and resist therapy. This plasticity is mediated by a tight, complex, context-dependent interactions between the environment, the epigenome and metabolism. Such interactions likely lead to more complex fitness landscapes than the simple dual regime captured by the current models and mutational data analysis. Understanding the nature of the interactions between different types of genetic and non-genetic changes occurring in cancer cells can unravel vulnerabilities that might be exploited for therapeutics.

Acknowledgements

We thank members of the Koonin group at the NCBI and members of the Moffitt Cancer Center for useful discussions and feedback. E.P. and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). R.J.G. and R.A.G. acknowledge support of the Physical Sciences in Oncology Program U54CA193489. F.D. is supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 648670).

Author contributions

E.P. researched data for the article. E.P. and E.V.K. wrote the article and reviewed/edited the manuscript before submission. All authors substantially contributed to discussion of content.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Genetics thanks P. Campbell, C. Maley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

References

1. Atlanta: American Cancer Society. Cancer Facts & Figures 2019. www.cancer.org (2019).
2. Vasan, N., Baselga, J. & Hyman, D.M. **A view on drug resistance in cancer**. *Nature* **575**, 299-309 (2019).
3. Cairns, J. **Mutation selection and the natural history of cancer**. *Nature* **255**, 197-200 (1975).
4. Nowell, P.C. **The clonal evolution of tumor cell populations**. *Science* **194**, 23-8 (1976).
5. Gerlinger, M. et al. **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing**. *N. Engl. J. Med*. **366**, 883-892 (2012).
6. Ding, L. et al. **Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing**. *Nature* **481**, 506-10 (2012).
7. Beltran, H. et al. **Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer**. *Nat. Med*. **22**, 298-305 (2016).
8. Faltas, B.M. et al. **Clonal evolution of chemotherapy-resistant urothelial carcinoma**. *Nat. Genet*. **48**, 1490-1499 (2016).
9. Hu, Z. et al. **Quantitative evidence for early metastatic seeding in colorectal cancer**. *Nat. Genet*. **51**, 1113-1122 (2019).
10. Nayar, U. et al. **Acquired HER2 mutations in ER+ metastatic breast cancer confer resistance to estrogen receptor-directed therapies**. *Nat. Genet*. **51**, 207-216 (2019).
11. Ye, K. et al. **Systematic discovery of complex insertions and deletions in human cancers**. *Nat. Med*. **22**, 97-104 (2016).
12. Wooster, R. et al. **Instability of short tandem repeats (microsatellites) in human cancers**. *Nat. Genet*. **6**, 152-6 (1994).
13. Hause, R.J., Pritchard, C.C., Shendure, J., Salipante, S.J. **Classification and characterization of microsatellite instability across 18 cancer types**. *Nat. Med.* **22**, 1342-1350 (2016).
14. Campbell, B.B. et al. **Comprehensive Analysis of Hypermutation in Human Cancer**. *Cell* **171**, 1042-1056.e10 (2017).
15. Persi, E. et al. **Proteomic and genomic signatures of repeat instability in cancer and adjacent normal tissues**. *Proc. Natl. Acad. Sci. U S A* **116**, 16987-16996 (2019).
16. Winge, Ö. **Zytologische Untersuchungen über die Natur maligner Tumoren**. *Z. Zellforsch* **10**, 683–735 (1930).
17. Abdel-Rahman, W.M. et al. **Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement**. *Proc. Natl. Acad. Sci. U S A* **98**, 2538-43 (2001).
18. Beroukhim, R. et al. **The landscape of somatic copy-number alteration across human cancers**. *Nature* **463**, 899-905 (2010).
19. Gordon, D.J., Resio, B. & Pellman, D. **Causes and consequences of aneuploidy in cancer**. *Nat. Rev. Genet.* **13**, 189-203 (2012).
20. Sansregret, L., Vanhaesebroeck, B. & Swanton, C. **Determinants and clinical implications of chromosomal instability in cancer**. *Nat. Rev. Clin. Oncol*. **15**, 139-150 (2018).
21. Taylor, A.M. et al. **Genomic and Functional Approaches to Understanding Cancer Aneuploidy**. *Cancer Cell* **33**, 676-689.e3 (2018).
22. Ben-David, U. & Amon, A. **Context is everything: aneuploidy in cancer**. *Nat. Rev. Genet*. **21**, 44-62 (2020).
23. Stephens, P.J. et al. **Massive genomic rearrangement acquired in a single catastrophic event during cancer development**. *Cell* **144**, 27-40 (2011).
24. Baca, S.C. et al. **Punctuated evolution of prostate cancer genomes**. *Cell* **153**, 666-77 (2013).
25. Bielski, C.M. et al. **Genome doubling shapes the evolution and prognosis of advanced cancers**. *Nat. Genet*. **50**, 1189-1195 (2018).
26. Sniegowski, P.D., Gerrish, P.J. & Lenski, R.E. **Evolution of high mutation rates in experimental populations of E. coli**. *Nature* **387**, 703–705 (1997).

27. Lang, G.I. et al. **Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations**. *Nature* **500**, 571-4 (2013).
28. Martincorena, I. et al. **High burden and pervasive positive selection of somatic mutations in normal human skin**. *Science* **348**, 880-6 (2015).
29. Martincorena, I. et al. **Somatic mutant clones colonize the human esophagus with age**. *Science* **362**, 911-917 (2018).
30. Yizhak, K. et al. **RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues**. *Science* **364**, pii: eaaw0726 (2019).
31. Tomasetti, C. & Vogelstein, B. **Variation in cancer risk among tissues can be explained by the number of stem cell divisions**. *Science* **347**, 78-81 (2015).
32. Tomasetti, C., Li, L. & Vogelstein, B. **Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention**. *Science* **355**, 1330-1334 (2017).
33. Reynolds, T.Y., Rockwell, S. & Glazer, P.M. **Genetic instability induced by the tumor microenvironment**. *Cancer Res*. **56**, 5754-7 (1996).
34. Koshiji, M. et al. **HIF-1alpha induces genetic instability by transcriptionally downregulating MutSalpha expression**. *Mol. Cell* **17**, 793-803 (2005).
35. Gillies, R.J., Verduzco, D. & Gatenby, R.A. **Evolutionary dynamics of carcinogenesis and why targeted therapy does not work**. *Nat. Rev. Cancer* **12**, 487-93 (2012).
36. Alexandrov, L. et al. **Signatures of mutational processes in human cancer**. *Nature* **500**, 415–421 (2013).
37. Hanahan, D. & Weinberg, R.A. **Hallmarks of cancer: the next generation**. *Cell* **144**, 646-74 (2011).
38. Burrell, R.A., McGranahan, N., Bartek, J. & Swanton, C. **The causes and consequences of genetic heterogeneity in cancer evolution**. *Nature* **501**, 338-45 (2013).
39. Lipinski, K.A. et al. **Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine**. *Trends Cancer* **2**, 49-63 (2016).
40. McGranahan, N. & Swanton, C. **Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future**. *Cell* **168**, 613-628 (2017).
41. Gillies, R.J., Brown, J.S., Anderson, A.R.A & Gatenby, R.A. **Eco-evolutionary causes and consequences of temporal changes in intratumoural blood flow**. *Nat. Rev. Cancer* **18**, 576-585 (2018).
42. DeBerardinis, R.J. & Chandel, N.S. **Fundamentals of cancer metabolism**. *Sci. Adv*. **2**, e1600200 (2016).
43. Feinberg, A.P., Koldobskiy, M.A. & Göndör, A. **Epigenetic modulators, modifiers and mediators in cancer aetiology and progression**. *Nat. Rev. Genet*. **17**, 284-99 (2016).
44. Kinnaird, A., Zhao, S., Wellen, K.E. & Michelakis, E.D. **Metabolic control of epigenetics in cancer**. *Nat. Rev. Cancer* **16**, 694-707 (2016).
45. Gatenby, R.A. & Gillies, R.J. **A microenvironmental model of carcinogenesis**. *Nat. Rev. Cancer* **8**, 56-61 (2008).
46. Gatenby, R.A., Gillies, R.J. & Brown, J.S. **Of cancer and cave fish**. *Nat. Rev. Cancer* **11**, 237–238 (2011).
47. Lloyd, M.C. et al. **Darwinian Dynamics of Intratumoral Heterogeneity: Not Solely Random Mutations but Also Variable Environmental Selection Forces**. *Cancer Res*. **76**, 3136-44 (2016).
48. Maley, C.C. et al. **Classifying the evolutionary and ecological features of neoplasms**. *Nat. Rev. Cancer* **17**, 605-619 (2017).
49. Lynch, M. & Conery, J.S. **The origins of genome complexity**. *Science* **302**, 1401-4 (2003).
50. Kimura, M. **The neutral theory of molecular evolution** (Cambridge University Press, Cambridge, 1983).
51. Lynch, M. **The Origins of Genome Architecture** (*Sinauer Associates Inc* (1723), **2007**).
52. Koonin, E.V. **The Logic of Chance: The Nature and Origin of Biological Evolution** (*FT Press Science*, 1st Edition, 2012).
53. Muller, H.J. **The relation of recombination to mutational advance**. *Mutat. Res*. **106**, 2-9 (1964).
54. Haigh, J. **The accumulation of deleterious genes in a population--Muller's Ratchet**. *Theor. Popul. Biol*. **14**, 251-67 (1978).
55. Dolja, V.V. & Koonin, E.V. **Common origins and host-dependent diversity of plant and animal viromes**. *Curr. Opin. Virol*. **1**, 322-31 (2011).

56. Reuter, S. et al. **Parallel independent evolution of pathogenicity within the genus Yersinia**. *Proc. Natl. Acad. Sci. U S A* **111**, 6768-73 (2014).
57. Kimura, M. **On the probability of fixation of mutant genes in a population**. *Genetics* **47**, 713-719 (1962).
58. Kryazhimskiy, S. & Plotkin, J.B. **The population genetics of dN/dS**. *PLoS Genet.* **4**, e1000304 (2008).
59. Lawrence, M.S. et al. **Mutational heterogeneity in cancer and the search for new cancer-associated genes**. *Nature* **499**, 214-218 (2013).
60. Bozic, I. et al. **Accumulation of driver and passenger mutations during tumor progression**. *Proc. Natl. Acad. Sci. U S A*, **107**, 18545-50 (2010).
61. Futreal, P.A. et al. **A census of human cancer genes**. *Nat. Rev. Cancer* **4**, 177-83 (2004).
62. Goyal, S. et al. **Dynamic mutation-selection balance as an evolutionary attractor**. *Genetics* **191**, 1309-19 (2012).
63. McFarland, C.D., Korolev, K.S., Kryukov, G.V., Sunyaev, S.R. & Mirny, L.A. **Impact of deleterious passenger mutations on cancer progression**. *Proc. Natl. Acad. Sci. U S A*, **110**, 2910-5 (2013).
64. McFarland, C.D., Mirny, L.A. & Korolev, K.S. **Tug-of-war between driver and passenger mutations in cancer and other adaptive processes**. *Proc. Natl. Acad. Sci. U S A*, **111**, 15138-15143 (2014).
65. Cox, D.R. **Regression models and life-tables**. *J. R. Stat. Soc. B.* **34**, 187–220 (1972).
66. Persi, E., Wolf, Y.I., Leiserson, M.D.M, Koonin, E.V. & Ruppin, E. **Criticality in tumor evolution and clinical outcome**. *Proc. Natl. Acad. Sci. U S A* **115**, E11101-E11110 (2018).
67. McFarland, C.D. et al. **The Damaging Effect of Passenger Mutations on Cancer Progression**. *Cancer Res.* **77**, 4763-4772 (2017).
68. Martincorena, I. et al. **Universal Patterns of Selection in Cancer and Somatic Tissues**. *Cell* **171**, 1029-1041.e21 (2017).
69. Weghorn, D. & Sunyaev, S. **Bayesian inference of negative and positive selection in human cancers**. *Nat. Genet.* **49**, 1785-1788 (2017).
70. Williams, M.J., Werner, B., Barnes, C.P., Graham, T.A. & Sottoriva, A. **Identification of neutral tumor evolution across cancer types**. *Nat. Genet.* **48**, 238-244 (2016).
71. Williams, M.J. et al. **Quantification of subclonal selection in cancer from bulk sequencing data.** *Nat. Genet.* **50**, 895–903 (2018).
72. Kumar, S. et al. **Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences.** *Cell* **180**, 915-927.e16 (2020).
73. López, S. et al. **Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution**. *Nat. Genet.* **52**, 283–293 (2020).
74. Kashi, Y. & King, D.G. **Simple sequence repeats as advantageous mutators in evolution**. *Trends Genet.* **22**, 253–259 (2006).
75. Gemayel, R., Vinces, M.D., Legendre, M. & Verstrepen, K.J. **Variable tandem repeats accelerate evolution of coding and regulatory sequences**. *Ann. Rev. Gen.* **44**, 445-77 (2010).
76. Chavali, S. et al. **Constraints and consequences of the emergence of amino acid repeats in eukaryotic proteins**. *Nat. Struct. Mol. Biol.* **24**, 765-777 (2017).
77. Levinson, G. & Gutman, G.A. **Slipped-strand mispairing: a major mechanism for DNA sequence evolution**. *Mol. Biol. Evol.* **4**, 203-221 (1987).
78. Charlesworth, B., Sniegowski. P. & Stephan, W. **The evolutionary dynamics of repetitive DNA in eukaryotes**. *Nature* **371**, 215-20 (1994).
79. Pâques, F., Leung, W.Y. & Haber, J.E. **Expansions and contractions in a tandem repeat induced by double-strand break repair.** *Mol. Cell Biol.* **18**, 2045-2054 (1998).
80. Hoelz, A., Debler, E.W. & Blobel, G. **The structure of the nuclear pore complex**. *Annu. Rev. Biochem.* **80**, 613-43 (2011).
81. Pick, E., Hofmann, K. & Glickman, M.H. **PCI complexes: Beyond the proteasome, CSN, and eIF3 Troika**. *Mol. Cell* **35**, 260-4 (2009).
82. Klug, A. & Rhodes, D. '**Zinc fingers': a novel protein motif for nucleic acid recognition**. *Trends in Biochemical Sciences* **12**, 464-469 (1987).

83. Neer, E.J., Schmidt, C.J., Nambudripad, R. & Smith, T.F. **The ancient regulatory-protein family of WD-repeat proteins**. *Nature* **371**, 297-300 (1994).
84. Bell, J.K. et al. **Leucine-rich repeats and pathogen recognition in Toll-like receptors**. *Trends Immunol* **24**, 528-33 (2003).
85. Mosavi, L.K., Cammett, T.J., Desrosiers, D.C. & Peng, Z.Y. **The ankyrin repeat as molecular architecture for protein recognition**. *Protein Sci*. **13**, 1435-48 (2004).
86. Chavali, S., Singh, A.K., Santhanam, B. & Babu M. **Amino acid homorepeats in proteins.** *Nat. Rev. Chem.* **4**, 420–434 (2020).
87. Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. **Intragenic tandem repeats generate functional variability**. *Nat. Genet*. **37**, 986-90 (2005).
88. Fondon, J.W.3rd & Garner, H.R. **Molecular origins of rapid and continuous morphological evolution**. *Proc. Natl. Acad. Sci. USA*, **101**, 18058-63 (2004).
89. Sawyer, L.A. et al. **Natural variation in a Drosophila clock gene and temperature compensation**. *Science* **278**, 2117-2120 (1997).
90. Gymrek, M. et al. **Abundant contribution of short tandem repeats to gene expression variation in humans**. *Nat. Genet*. **48**, 22-9 (2016).
91. Santos-Pereira, J.M. & Aguilera, A. **R loops: new modulators of genome dynamics and function**. *Nat. Rev. Genet*. **16**, 583-97 (2015).
92. Jain, A. & Vale, R.D. **RNA phase transitions in repeat expansion disorders**. *Nature* **546**, 243-247 (2017).
93. Bibel, M. & Barde, Y.A. **Neurotrophins: key regulators of cell fate and cell shape in the vertebrate nervous system**. *Genes Dev*. **14**, 2919-37 (2000).
94. Nithianantharajah, J. & Hannan, A.J. **Dynamic mutations as digital genetic modulators of brain development, function and dysfunction**. *Bioessays* **29**, 525-35 (2007).
95. Kannan, M. et al. **WD40-repeat 47, a microtubule-associated protein, is essential for brain development and autophagy**. *Proc. Natl. Acad. Sci. U S A* **114**, E9308-E9317 (2017).
96. Hammock, E.A. & Young, L.J. **Microsatellite instability generates diversity in brain and sociobehavioral traits**. *Science* **308**, 1630-4 (2005).
97. Segurel, L., Leffler, E.M. & Przeworski, M. **The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans**. *PLoS Biol*. **9**, e1001211 (2011).
98. Davies, B. et al. **Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice**. *Nature* **530**, 171-176 (**2016**).
99. Persi, E., Wolf, Y.I. & Koonin, E.V. **Positive and Strongly Relaxed Purifying Selection Drive the Evolution of Repeats in Proteins.** *Nat. Commun.* **7,** 13570 (2016).
100. Schaper, E., Gascuel, O. & Anisimova, M. **Deep conservation of human protein tandem repeats within the eukaryotes.** *Mol. Biol. Evol.* **31**, 1132–1148 (2014).
101. Marcotte, E.M., Pellegrini, M., Yeates, T.O. & Eisenberg, D. **A census of protein repeats**. *J. Mol. Biol*. **293**, 151–160 (1999).
102. Persi, E. & Horn, D. **Systematic analysis of compositional order of proteins reveals new characteristics of biological functions and a universal correlate of macroevolution.** *PLoS Comput. Biol.* **9**, e1003346 2013 (2013).
103. Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J. & Gentles, A.J. **Amino acid runs in eukaryotic proteomes and disease associations**. *Proc. Nat. Acad. Sci. USA* **99**, 333–338 (2002).
104. Gatchel, J.R. & Zoghbi, H.Y. **Diseases of unstable repeat expansion: mechanisms and common principles**. *Nat. Rev. Genet.* **6**, 743-755 (2005).
105. Pearson, C.E., Nichol Edamura, K. & Cleary, J.D. **Repeat instability: mechanisms of dynamic mutations.** *Nat. Rev. Genet*. **6**, 729-42 (2005).
106. Mirkin, S.M. **Expandable DNA repeats and human disease**. *Nature* **447**, 932-40 (2007).
107. La Spada, A.R. & Taylor, J.P. **Repeat expansion disease: progress and puzzles in disease pathogenesis**. *Nat. Rev. Genet*. **11**, 247-58 (2010).

108. López Castel, A., Cleary, J.D. & Pearson, C.E. **Repeat instability as the basis for human diseases and as a potential target for therapy**. *Nat. Rev. Mol. Cell Biol*. **11**, 165-70 (2010).
109. Carvalho, C.M. & Lupski, J.R. **Mechanisms underlying structural variant formation in genomic disorders**. *Nat. Rev. Genet*. **17**, 224-38 (2016).
110. Hannan, A.J. **Tandem repeats mediating genetic plasticity in health and disease**. *Nat. Rev. Genet*. **19**, 286-298 (2018).
111. Khristich, A.N. & Mirkin, S.M. **On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability**. *J. Biol. Chem.* **295**, 4134-4170 (2020).
112. DeJesus-Hernandez, M. et al. **Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS**. *Neuron* **72**, 245-56 (2011).
113. Ishiura, H. et al. **Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy**. *Nat. Genet*. **50**, 581-590 (2018).
114. Loeb, L.A., Springgate, C.F. & Battula, N. **Errors in DNA replication as a basis of malignant changes**. *Cancer Res*. **34**, 2311–2321 (1974).
115. Duval, A. & Hamelin, R. **Mutations at coding repeat sequences in mismatch repair deficient human cancers: Toward a new concept of target genes for instability**. *Cancer Res*. **62**, 2447–2454 (2002).
116. Treangen, T.J. & Salzberg, S.L. **Repetitive DNA and next-generation sequencing: Computational challenges and solutions**. *Nat. Rev. Genet.* **13**, 36–46 (2011).
117. El-Metwally, S., Hamza, T., Zakaria, M. & Helmy, M. **Next-generation sequence assembly:Four stages of data processing and computational challenges**. *PLoS Comput. Biol*. **9**, e1003345 (2013).
118. Popat, S., Hubner, R. & Houlston, R.S. **Systematic review of microsatellite instability and colorectal cancer prognosis.** *J. Clin. Oncol*. **23**, 609-18 (2005).
119. Li, B. et al. **Landscape of tumor-infiltrating T cell repertoire of human cancers**. *Nat. Genet*. **48**, 725-32 (2016).
120. Germano, G. et al. **Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth**. *Nature* **552**, 116-120 (2017).
121. Yarchoan, M., Hopkins, A. & Jaffee, E.M. **Tumor Mutational Burden and Response Rate to PD-1 Inhibition**. *N. Engl. J. Med*. **377**, 2500-2501 (2017).
122. Yarchoan, M., Johnson, B.A.3rd., Lutz, E.R., Laheru, D.A. & Jaffee, E.M. **Targeting neoantigens to augment antitumour immunity**. *Nat. Rev. Cancer* **17**, 209-222 (2017).
123. Mlecnik, B. et al. **Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability**. *Immunity* **44**, 698-711 (2016).
124. Mlecnik, B. et al. **The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis**. *Sci. Transl. Med*. **8**, 327ra26 (2016).
125. Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. & Perucho, M. **Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis**. *Nature* **363**, 558–561 (1993).
126. Torres, E.M. et al. **Effects of aneuploidy on cellular physiology and cell division in haploid yeast**. *Science* **317**, 916-24 (2007).
127. Pavelka, N. et al. **Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast**. *Nature* **468**, 321-5 (2010).
128. Weaver, B.A., Silk, A.D., Montagna, C., Verdier-Pinard, P. & Cleveland, D.W. **Aneuploidy acts both oncogenically and as a tumor suppressor.** *Cancer Cell* **11**, 25-36 (2007).
129. Rowald, K. et al. **Negative Selection and Chromosome Instability Induced by Mad2 Overexpression Delay Breast Cancer but Facilitate Oncogene-Independent Outgrowth**. *Cell Rep*. **15**, 2679-91 (2016).
130. Sotillo, R., Schvartzman, J.M., Socci, N.D. & Benezra, R. **Mad2-induced chromosome instability leads to lung tumour relapse after oncogene withdrawal**. *Nature* **464**, 436-40 (2010).
131. Bakhoum, S.F., Ngo, B., Laughney, A.M., Cavallo, J.A. & Murphy, C.J. **Chromosomal instability drives metastasis through a cytosolic DNA response**. *Nature* **553**, 467-472 (2018).
132. Gao, C. et al. **Chromosome instability drives phenotypic switching to metastasis**. *Proc. Natl. Acad. Sci. U S A* **113**, 14793-14798 (2016).

133. Danielsen, H.E., Pradhan, M. & Novelli, M. **Revisiting tumour aneuploidy - the place of ploidy assessment in the molecular era.** *Nat. Rev. Clin. Oncol*. **13**, 291-304 (2016).
134. Auslander, N., Wolf, Y.I. & Koonin, E.V. **Interplay between DNA damage repair and apoptosis shapes cancer evolution through aneuploidy and microsatellite instability**. *Nat. Commun.* **11**, 1234 (2020).
135. Birkbak, N.J. et al. **Paradoxical relationship between chromosomal instability and survival outcome in cancer.** *Cancer Res*. **71**, 3447-52 (2011).
136. Roylance, R. et al. **Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer**. *Cancer Epidemiol. Biomarkers Prev*. **20**, 2183-94 (2011).
137. Zasadil, L.M. et al. **High rates of chromosome missegregation suppress tumor progression but do not inhibit tumor initiation**. *Mol. Biol. Cell* **27**, 1981-9 (2016).
138. Silk, A.D. et al. **Chromosome missegregation rate predicts whether aneuploidy will promote or suppress tumors.** *Proc. Natl. Acad. Sci. U S A* **110**, E4134-41 (2013).
139. Sansregret, L. et al. **APC/C Dysfunction Limits Excessive Cancer Chromosomal Instability**. *Cancer Discov*. **7**, 218-233 (2017).
140. Davoli, T. et al. **Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome**. *Cell* **155**, 948-962 (2013).
141. Davoli, T., Uno, H., Wooten, E.C. & Elledge, S.J. **Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy.** *Science* **355**, eaaf8399 (2017).
142. Díaz-Carballo, D. et al. **A Distinct Oncogenerative Multinucleated Cancer Cell Serves as a Source of Stemness and Tumor Heterogeneity.** *Cancer Res*. **78**, 2318-2331 (2018).
143. Pienta, K.J., Hammarlund, E.U., Axelrod, R., Brown, J.S. & Amend, S.R. **Poly-aneuploid cancer cells promote evolvability, generating lethal cancer.** *Evolutionary Applications* **00**, 1–9 (2020).
144. Andor, N. et al. **Pan-cancer analysis of the extent and consequences of intratumor heterogeneity.** *Nat. Med*. **22**, 105-13 (2016).
145. Smith, J.C. & Sheltzer, J.M. **Systematic identification of mutations and copy number alterations associated with cancer patient prognosis.** *Elife*, 7:e39217 (2018).
146. Lynch, M. & Conery, J.S. **The evolutionary fate and consequences of duplicate genes**. *Science* **290**, 1151-1155 (2000).
147. Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. **Selection in the evolution of gene duplications**. *Genome Biol*. **3**, RESEARCH0008 (2002).
148. Innan, H. & Kondrashov, F. **The evolution of gene duplications: classifying and distinguishing between models**. *Nat. Rev. Genet.* **11**, 97-108 (2010).
149. Dewhurst, S.M. et al. **Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution**. *Cancer Discov*. **4**, 175-185 (2014).
150. Maciver, S.K. **Asexual Amoebae Escape Muller's Ratchet through Polyploidy**. *Trends Parasitol*. **32**, 855-862 (2016).
151. Ford A.M. et al. **Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia**. *Proc. Natl. Acad. Sci. U S A* **95**, 4584–4588 (1998).
152. Gerstung, M. et al. **The evolutionary history of 2,658 cancers.** *Nature* **578**, 122–128 (2020).
153. Rausch, T. et al. **Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations**. *Cell* **148**, 59-71 (2012).
154. Mitchell, T.J. et al. **Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal**. *Cell* **173**, 611-623.e17 (2018).
155. Cortés-Ciriano I. et al. **Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing.** *Nat Genet.* **52**, 331–341 (2020).
156. Paulson, T.G. et al. **Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma.** *Clin. Cancer Res*. **15**, 3305-3314 (2009).
157. Martinez, P. et al. **Evolution of Barrett's esophagus through space and time at single-crypt and whole-biopsy levels**. *Nat. Commun*. **9**, 794 (2018).

158. Ordway, B., Swietach, P., Gillies, R.J. & Damaghi, M. **Causes and Consequences of Variable Tumor Cell Metabolism on Heritable Modifications and Tumor Evolution**. *Front. Oncol.* **10**, 373 (2020).
159. Warburg, O. **On the origin of cancer cells**. *Science* **123**, 309-14 (1956).
160. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. **Understanding the Warburg effect: the metabolic requirements of cell proliferation**. *Science* **324**, 1029-33 (2009).
161. Gillies, R.J., Robey, I. & Gatenby, R.A. **Causes and consequences of increased glucose metabolism of cancers**. *J. Nucl. Med*. **49**, Suppl 2:24S-42S (2008).
162. Webb, B.A., Chimenti, M., Jacobson, M.P. & Barber, D.L. **Dysregulated pH: a perfect storm for cancer progression**. *Nat. Rev. Cancer* **11**, 671-7 (2011).
163. Corbet, C. & Feron, O. **Tumour acidosis: from the passenger to the driver's seat**. *Nat. Rev. Cancer* **17**, 577-593 (2017).
164. Brand, A. et al. **LDHA-Associated Lactic Acid Production Blunts Tumor Immunosurveillance by T and NK Cells**. *Cell Metab.* **24**, 657-671 (2016).
165. Pilon-Thomas, S. et al. **Neutralization of Tumor Acidity Improves Antitumor Responses to Immunotherapy**. *Cancer Res*. **76**, 1381-90 (2016).
166. Persi, E. et al. **Systems analysis of intracellular pH vulnerabilities for cancer therapy**. *Nat Commun*. **9**, 2997 (2018).
167. Hoffmann, A.A. & Hercus, M.J. **Environmental Stress as an Evolutionary Force**. *BioScience* **50**, 217–226 (2000).
168. Ram, Y. & Hadany, L. **The evolution of stress-induced hypermutation in asexual populations**. *Evolution* **66**, 2315-28 (2012).
169. McKenzie, G.J., Harris, R.S., Lee, P.L. & Rosenberg, S.M. **The SOS response regulates adaptive mutation**. *Proc. Natl. Acad. Sci. U S A* **97**, 6646-51 (2000).
170. Rattray, A.J. & Strathern, J.N. **Error-prone DNA polymerases: when making a mistake is the only way to get ahead**. *Annu. Rev. Genet*. **37**, 31-66 (2003).
171. Krutyakov, V.M. **Eukaryotic error-prone DNA polymerases: The presumed roles in replication, repair, and mutagenesis**. *Mol. Biol.* **40**, 1–8 (2006).
172. Fitzgerald, D.M., Hastings, P.J. & Rosenberg, S.M. **Stress-Induced Mutagenesis: Implications in Cancer and Drug Resistance**. *Annu. Rev. Cancer Biol*. **1**, 119-140 (2017).
173. Clara, J.A., Monge, C., Yang, Y. & Takebe, N. **Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update**. *Nat. Rev. Clin. Oncol.* **17**, 204-232 (2020).
174. Yun, J. et al. **Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells**. *Science* **325**, 1555-9 (2009).
175. Young, S.D., Marshall, R.S. & Hill, R.P. **Hypoxia induces DNA overreplication and enhances metastatic potential of murine tumor cells**. *Proc. Natl. Acad. Sci. U S A* **85**, 9533-7 (1988).
176. Jackson, A.L., Chen, R. & Loeb, A.L. **Induction of microsatellite instability by oxidative DNA damage**. *Proc. Natl. Acad. Sci. U S A* **95**, 12468–12473 (1998).
177. Chatterjee, N., Lin, Y., Santillan, B.A., Yotnda, P. & Wilson, J.H. **Environmental stress induces trinucleotide repeat mutagenesis in human cells**. *Proc. Natl. Acad. Sci. U S A* **112**, 3764-9 (2015).
178. Burrell, R.A. et al. **Replication stress links structural and numerical cancer chromosomal instability**. *Nature* **494**, 492-496 (2013).
179. Matzke, M.A., Mittelsten Scheid, O. & Matzke, A.J. **Rapid structural and epigenetic changes in polyploid and aneuploid genomes**. *Bioessays* **21**, 761-7 (1999).
180. Tang, Y.C. & Amon, A. **Gene copy-number alterations: a cost-benefit analysis**. *Cell* **152**, 394-405 (2013).
181. Herrera, L.A., Prada, D., Andonegui, M.A. & Dueñas-González, A. **The epigenetic origin of aneuploidy**. *Curr. Genomics* **9**, 43-50 (2008).
182. Chen, G., Bradford, W.D., Seidel, C.W. & Li, R. **Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy**. *Nature* **482**, 246-50 (2012).
183. Yona, A.H. et al. **Chromosomal duplication is a transient evolutionary solution to stress**. *Proc. Natl. Acad. Sci. U S A*, **109**, 21010-5 (2012).

184. Espina, V. & Liotta, L.A. **What is the malignant nature of human ductal carcinoma in situ?** *Nat. Rev. Cancer*, **11**, 68-75 (2011).
185. Filippova, G.N. et al. **CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus**. *Nat. Genet.* **28**, 335-43 (2001).
186. Libby, R.T. et al. **CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination**. *PLoS Genet.* **4**, e1000257 (2008).
187. Dion, V., Lin, Y., Hubert, L. Jr, Waterland, R.A. & Wilson, J.H. **Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline**. *Hum. Mol. Genet*. **17**, 1306-17 (2008).
188. Gallagher, E.J. & LeRoith, D. **Obesity and Diabetes: The Increased Risk of Cancer and Cancer-Related Mortality**. *Physiol. Rev*. **95**, 727-48 (2015).
189. Kussell, E. & Leibler, S. **Phenotypic diversity, population growth, and information in fluctuating environments**. *Science* **309**, 2075-2078 (2005).
190. Cvijović, I., Good, B.H., Jerison, E.R. & Desai, M.M. **Fate of a mutation in a fluctuating environment**. *Proc. Natl. Acad. Sci. U S A* **112**, E5021-8 (2015).
191. Gatenby, R.A., Cunningham, J.J. & Brown, J.S. **Evolutionary triage governs fitness in driver and passenger mutations and suggests targeting never mutations**. *Nat. Commun.* **5**, 5499 (2014).
192. Schiffman, J.D., White, R.M., Graham, T.A., Huang, Q. & Aktipis, A. **The Darwinian Dynamics of Motility and Metastasis**. In C. C. Maley & M. Greaves (Eds.), Frontiers in Cancer Research: Evolutionary Foundations, Revolutionary Directions (pp. 135–176). Springer New York (2016).
193. Ashworth, A., Lord, C.J. & Reis-Filho, J.S. **Genetic interactions in cancer progression and treatment**. *Cell* **145**, 30-8 (2011).
194. Phillips, P.C. **Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems**. *Nat. Rev. Genet*. **9**, 855-67 (2008).
195. Poelwijk, F.J., Tănase-Nicola, S., Kiviet, D.J. & Tans, S.J. **Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes**. *J. Theor. Biol*. **272**, 141-4 (2011).
196. O'Neil, N.J., Bailey, M.L. & Hieter, P. **Synthetic lethality and cancer**. *Nat. Rev. Genet*. **18**, 613-623 (2017).
197. Huang, A., Garraway, L.A., Ashworth, A. & Weber, B. **Synthetic lethality as an engine for cancer drug target discovery**. *Nat. Rev. Drug Discov*. **19**, 23-38 (2020).
198. Lord, C.J. & Ashworth, A. **PARP inhibitors: Synthetic lethality in the clinic**. *Science* **355**, 1152-1158 (2017).
199. Bitler, B.G. et al. **Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers**. *Nat. Med*. **21**, 231-8 (2015).
200. Chan, E.M. et al. **WRN helicase is a synthetic lethal target in microsatellite unstable cancers**. *Nature* **568**, 551-556 (2019).
201. Trachootham, D., Alexandre, J. & Huang, P. **Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach?** *Nat. Rev. Drug Dis.* **8**, 579-91 (2009).

Figures

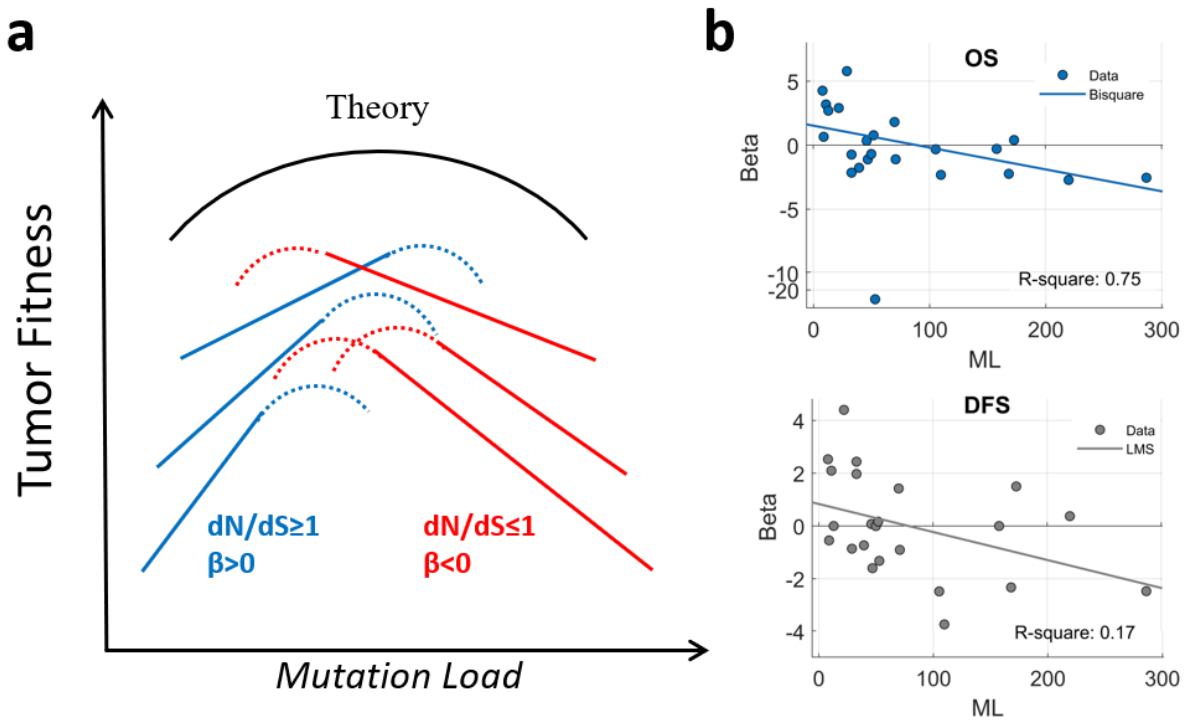

Figure 1. The mutation-selection diagram of tumour evolution. a | The dependency of tumour fitness on the point mutation load (ML) The graph illustrates the dual evolutionary regime of cancers (coloured curves). In initial phases (low mutation loads, blue lines), tumour progression is characterized by an increasing number of positively selected drivers (*dN/dS>1*) and increased tumour fitness (poor clinical outcome, *β>0*). By contrast, later in tumour progression, accumulation of deleterious passenger mutations becomes critical, leading to better clinical outcome (*β<0*, red lines) and purifying selection (*dN/dS<1*), thereby avoiding mutation meltdown at extreme mutation loads. Solid coloured lines denote observed behaviour, whereas dashed coloured curves denote theoretically inferred behaviour. *β* is the prognostic factor (Hazard ratio, $HR=e^{\beta}$) derived from Cox regression analysis of patient survival data and *dN/dS* is the ratio between the rates of non-silent (*dN*) and silent (*dS*) mutations. (**Box 1**). Collectively, these empirical results corroborate theoretical predictions (black curve) [63,64]. **b |** The observed relationship between the prognostic factor β of Cox regression analysis and the median ML, estimated across patients, in 23 cancer types from The Cancer Genome Atlas (TCGA), from which **Fig 1a**. is derived. Each point represents a cancer type, showing the change in evolutionary regimes, from β>0 (i.e., mutations lead to adverse outcome) for low ML cancers, to β<0 (i.e., mutations lead to favourable outcome) for high ML cancers. Results are shown for both the overall survival (OS) and the disease-free survival (DFS) data. Parts **a** and **b** are adapted with permission from ref[66].

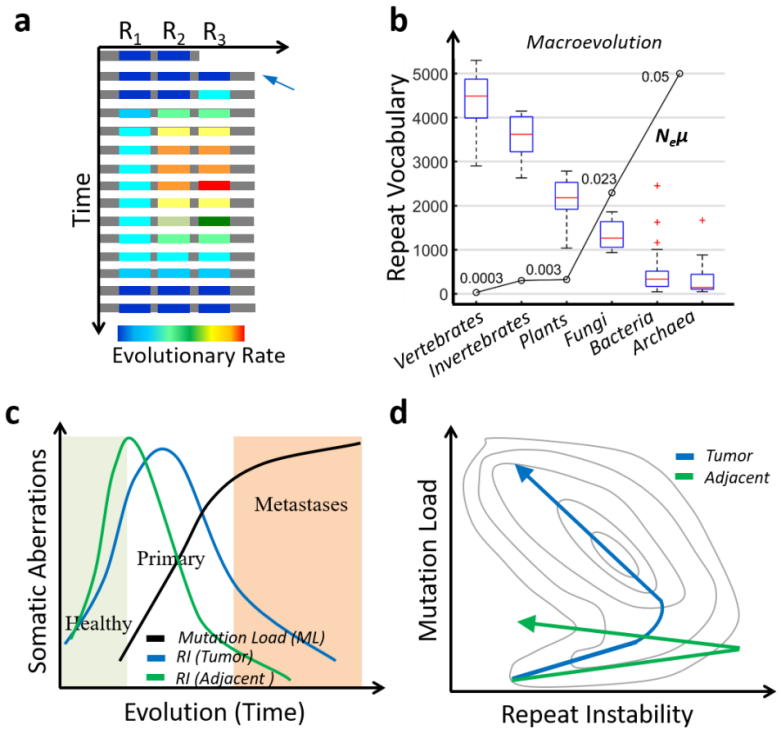

Figure 2. The dynamics and role of repeat instability in species and cancer evolution. a | Accelerated evolution of repeats following duplication of a new copy (R3, arrow) in a protein, illustrated by the evolutionary rates (colour code) of all repeat units (R1–R3)[99]. The new copy is most relaxed of selective constraints, and rapidly diverges at an elevated rate (red), under positive selection and relaxed purifying selection. This evolutionary divergence can yield a new function, followed by fixation, conservation and slowing down of evolutionary rates (blue). **b |** Over longer spans of organismal evolution, following several cycles of repeat unit duplications and fixation of mutations in new copies, this process translates into a rich vocabulary of repetitive elements, which can be measured by the number of 'words' (here, amino-acid triplets) that compose the repeat arrays in proteomes of different species[102] (boxes). This repeat diversity provides a marker of macroevolution that anticorrelates with $Ne \times \mu$ (black curve, where Ne is the effective population size and μ is the mutation rate per nucleotide site per generation; median values derived from ref[49] are shown). **c |** A proposed model of evolutionary repeat dynamics in cancer and normal tissues. In cancer evolution, repeat instability (RI) is most pronounced at early stages and, presumably, is positively selected for, compensating for the low number of driver mutations. By contrast, at later stages, RI is substantially reduced by negative selection because, as cancer progresses, the genomic burden becomes detrimental to the tumour and an immune response to neoantigens is evoked. Adjacent tissues show mutational signatures similar to those in tumours, presumably because they respond to similar environmental stress factors, but exhibit a faster dynamic[15]. **d |** The dynamics in **Fig. 2c** can be translated into a tumour fitness landscape, where both adjacent and tumour cells climb the landscape through increased RI, as an escape mechanism from stress. Cells that acquire driver point mutations become neoplastic (blue) and further climb the landscape, whereas adjacent cells (green) do not harbour driver mutations and retain normal function. At late stages, RI is reduced in both tumours and adjacent tissue by purifying selection. Part **a** is adapted from ref[99]. Part **b** is adapted from ref[102]. Part **c** is adapted from ref[15].

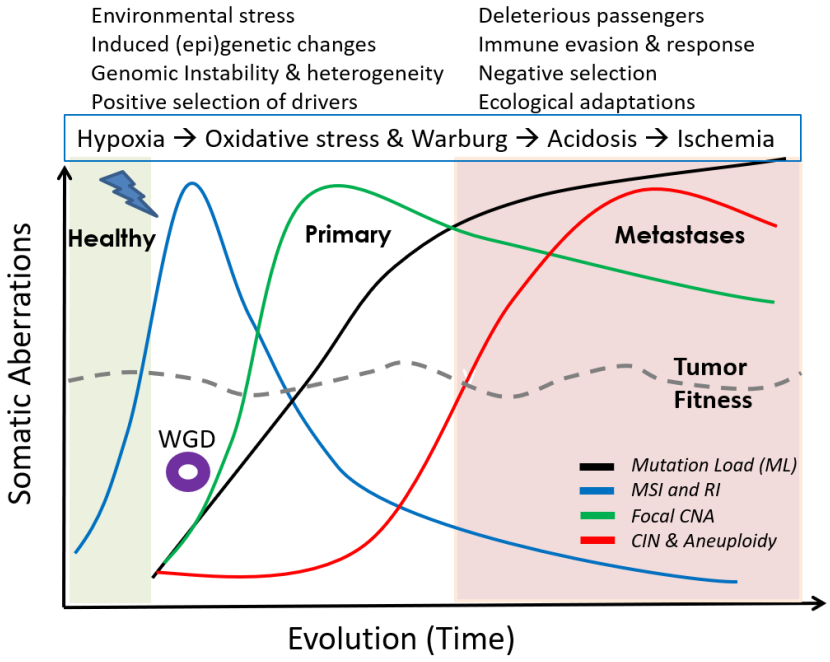

Figure 3. **Proposed compensatory relationships between different types of genomic aberrations during tumour evolution.** The extent of somatic aberrations, normalized to maximum, is illustrated as a function of time. A healthy cell transforms into a neoplastic cell following epigenetic changes and/or acquisition of a driver mutation, which could occur randomly or induced by environmental stress [35]. The point mutation load (ML) increases nearly irreversibly with time and cancer progression (black curve). Microsatellite instability (MSI), and repeat instability (RI) are induced in early stages of the primary tumour, presumably compensating for the low number of drivers [14,15]. Later in the tumour evolution, the genomic burden increases, and MSI and RI decline, via both cell autonomous mechanisms and through the evoked immune response to neoantigens which elicits purifying selection. In individual patients, this reduction is most evident during the transition to metastatic states. As high MSI becomes a vulnerability [13], aneuploidy acts as a compensatory mechanism at late stages [21,22]. Conceivably, the immune system suppresses the accumulation of aneuploidy at early stages, but as cancers evolve mechanisms to avoid immune surveillance, aneuploidy is more tolerated, especially when preceded by whole-genome duplication (WGD) (shown by the circle). Aneuploidy can facilitate metastasis formation and the emergence of resistant phenotypes and mitigate Muller's ratchet. However, extreme levels of aneuploidy become a vulnerability. Focal somatic copy number alterations (SCNA) appear to serve as a compensatory mechanism in intermediate stages, after MSI and RI reduce and before aneuploidy accumulation confers a fitness advantage [144,145]. The cumulative action of these compensatory mechanisms may provide for the maintenance of robust tumour fitness over the course of evolution (dashed curve). The temporal order of metabolic stress (upper panel) parallels the induction of somatic aberrations, whereby hypoxia and oxidative stress can induce RI and point mutations, whereas larger aberrations are likely to depend on the accumulation of mutations in DNA repair and replication machineries and/or can be induced by acidosis (see the main text).

BOX 1: Key variables and concepts of populations genetics and cancer genomics

[b1] Mutation rate (μ)

The number of mutations per nucleotide site per generation. Typically, μ is high in prokaryotes and unicellular eukaryotes, but substantially lower in multicellular eukaryotes.

[b1] Effective population size (N_e)

The number of individuals in a population that participate in reproduction. Typically, N_e is high in prokaryotes and unicellular eukaryotes and substantially lower in multicellular eukaryotes. Thus, the product N_e x μ defines the characteristic evolutionary regimes of major clades [49]. For a population of N cancer cells, each with a driver mutation conferring tumorigenic renewal potential, N_e can be approximated by N.

[b1] Selection coefficient (s)

The effect of a mutation on the fitness of a genotype. $s>0$ denotes selective advantage, whereas $s<0$ denotes selective disadvantage (i.e., a deleterious mutation).

[b1] Fixation probability

The probability that a mutation spreads across the population and will become fixed, reaching a stable allele frequency in the population. Kimura [57] has shown that the fixation probability (for a diploid) is related to N_e x s: $P_{fix} = (1-e^{-4NesQ})/(1-e^{-4Nes})$, where q is the initial allele frequency, such that for a new mutation ($q=1/2N$) and assuming $N_e=N$, $P_{fix} = 2s/(1-e^{-4NS})$.

[b1] Selection (dN/dS)

Acting on a protein-coding sequence: the ratio between the substitution rate of non-silent mutations per non-synonymous site (dN), that is mutations that change the amino acid sequence, and the rate of synonymous mutations (dS), that is, mutations that do not result in amino acid changes. In a haploid, assuming that synonymous mutations are neutral ($s=0$, with substitution rate μ) and non-synonymous mutations have selection coefficient s ($s>0$), $dN/dS = \mu N$ x $P_{fix} / \mu = 2Ns/(1-e^{-2NS})$; hence the direct correspondence between dN/dS and N_e [58].

[b1] Positive selection

Excess level of non-synonymous mutations over synonymous mutations, $dN/dS>1$, $s>0$.

[b1] Purifying (negative) selection

Excess of synonymous mutations over non-synonymous mutations, $dN/dS<1$, $s<0$.

[b1] Genetic drift

Changes in allele frequency due to random sampling of progeny in the process of reproduction. In populations with small N_e, random fluctuations due to sampling can lead to fixation of not only beneficial but also of moderately deleterious mutations.

[b1] Neutral evolution

Evolution by fixation of selectively neutral mutant alleles ($s=0$). According to the modern theory of molecular evolution [50], most of the variation within and between species is (nearly) neutral.

[b1] Muller's ratchet

A situation in which the fittest clone in an asexual population carries an excess of deleterious mutations, which can lead to mutational meltdown and extinction of the entire population [53,54]. In sexual populations, Muller's ratchet is unlikely because recombination can mitigate the accumulation of deleterious mutations.

[b1] Mutation load (ML)

The total number of non-silent, that is, protein changing, point mutations in a tumour genome of a patient, relative to blood or matched-normal tissue of the same patient, which serve as proxies for the germline genome. ML represents the integral of the mutation rate (μ)_for non-silent mutations across the genome and over time T, from the emergence of a neoplastic cell to the current time. ML correlates with time and the age of a tissue [29], and assuming T is known and sufficiently large, the average μ can be estimated from ML [60]. ML usually includes a small fraction (< 5%) of in-frame and out-of-frame insertions and deletions (indels) and splice variants.

[b1] Genomic burden

For reversible mutations, such as repeat instability (RI), somatic copy-number alterations (SCNA) and chromosomal instability (CIN), the extent of aberration is often measured by the genomic burden, defined as the total amount (or fraction) of DNA affected by the respective type of mutations, relative to wild-type. Genomic burden is proportional to the integral over time of the net rate of amplification and deletions, $\mu^{amp}{}_i - \mu^{del}{}_i$, summed (in absolute value) across regions i.

[b1] Driver and passenger mutations

A cancer driver mutation is any type of mutation that confers a cell with a selective advantage ($s>0$), whereas a passenger mutation is either neutral or deleterious ($s\leq 0$). Driver mutations often occur in a small subset of genes, such as oncogenes and tumour suppressors, known as cancer genes [61].

[b1] Tumour fitness

The fitness of a tumour is defined here as the inverse of the clinical outcome (survival rate) of patients. Favourable outcomes (high survival rates) correspond to low tumour fitness, and adverse outcomes (low survival rate) correspond to high tumour fitness. Accordingly, the effect of a mutation on the fitness of tumours is defined as the inverse of its correlation with the clinical outcome (survival rate). This is a biologically intuitive and practical definition that weighs in the compound effect of the capacity of cancer cells to proliferate, migrate, and eventually, invade and colonize tissues.

[b1] Cox regression

A semi-parameterized approach that fits the survival data to a hazard function [$h(t) = -d[\log S(t)]/dt$, where $S(t)$ is the survival probability at time t] and tests the effect of variables (X) under the "proportional hazard" assumption [$h(X,t) = ho(t)e^{X\beta}$; ho the baseline hazard], namely, that the tested hazard functions are log-linearly scaled by a constant factor beta (β), which determines the hazard ratio, HR ($HR = e^{\beta}$) [65]. $\beta > 0$ ($HR > 1$) indicates poor survival, whereas $\beta < 0$ ($HR < 1$) indicates better survival, for sufficiently large X.

Table of contents blurb

Although cancer genetics analyses have often focused on individual mutations of classic cancer genes, a wealth of cancer sequencing data is allowing a more comprehensive understanding of the cumulative effects of mutations genome-wide. In this Perspectives article, the authors propose how the burden of different types of mutations — from point mutations to large-scale chromosomal aberrations — has distinct and compensatory roles on tumour fitness and selection during different stages of cancer evolution.