OPINION

Genomic evolution of cancer models: perils and opportunities

Uri Ben-David, Rameen Beroukhim and Todd R. Golub

Abstract | Cancer research relies on model systems, which reflect the biology of actual human tumours to only a certain extent. One important feature of human cancer is its intra-tumour genomic heterogeneity and instability. However, the extent of such genomic instability in cancer models has received limited attention in research. Here, we review the state of knowledge of genomic instability of cancer models and discuss its biological origins and implications for basic research and for cancer precision medicine. We discuss strategies to cope with such genomic evolution and evaluate both the perils and the emerging opportunities associated with it.

Cancer is a disease characterized by the genomic instability of somatic cells. Genomic evolution generates genetic and epigenetic diversity, and the resultant cellular heterogeneity constitutes a fertile molecular ground for further evolution. In recent years, largely thanks to the advance of single-cell omics and sequencing technologies, clonal heterogeneity and tumour evolution have been studied extensively, and their importance for cancer progression and for the clinical outcome of cancer treatments is now widely appreciated (reviewed previously[1,2]).

Any functional interrogation of human cancer cells must rely on patient-derived cancer models, such as patient-derived cell lines (PDCLs), patient-derived organoids (PDOs) and patient-derived xenografts (PDXs). The successful derivation of such models requires that the tumour cells adapt to new environmental conditions, in other words, distinct selection pressures, and their propagation continuously selects for the fittest and most rapidly proliferating cells[3–5]. Moreover, as cancer cells are often deficient in their ability to properly maintain genome integrity (reviewed previously[6]), their inherent genomic instability makes them susceptible to rapid acquisition of additional genetic insults throughout propagation. Non-patient-derived cancer models, such as genetically engineered mouse models (GEMMs), also experience genomic evolution at both the tumour level and the host level[7]. Cancer model evolution is thus emerging as an important aspect of cancer modelling.

In recent years, advances in the development of cancer models have greatly expanded their application in cancer precision medicine. First, large cohorts (also known as biobanks) of cancer models have been generated, and extensive genomic and phenotypic characterization of these models have been performed to uncover genotype–phenotype associations at the patient population level[8–31]. Second, patient-derived models are increasingly being used as avatars of their tumour of origin in an attempt to predict patient-specific drug response[31–35]. For both applications, cancer models ought to be faithful representations of the tumours from which they were derived and remain genomically and phenotypically stable throughout propagation. The proper use of cancer models thus requires critical evaluation of these underlying assumptions in light of the propensity of these models to evolve.

The evolution of cancer models bears potential consequences for another burning issue in cancer research — its reproducibility. The 'reproducibility crisis', that is, the inability to replicate results reported in the literature, has drawn much attention recently. Cancer research has been at the focus of this debate, following reports that only 11–25% of high-profile cancer studies could be replicated by an industrial laboratory[36,37]. For example, differences between large-scale drug screens of cancer cell lines have been observed and debated in the literature[38–40]. Although many explanations have been suggested to account for, and to some extent reconcile, such discrepancies[39–45], the potential contribution of model evolution to observed differences remains underexplored.

In this Opinion article, we summarize the emerging evidence for genomic evolution in cancer models, its biological origins and its functional consequences. We then highlight the implications for basic cancer research and for clinical translation, including cancer precision medicine. Finally, we suggest practical ways to mitigate the risks posed by genomic evolution and propose how to constructively build on this phenomenon in future research.

Model evolution: proof and prevalence

The factors shaping evolution (FIG. 1) can differ between GEMMs and patient-derived models and between PDCLs, PDXs and PDOs (TABLE 1). The rate of genomic evolution is determined by the genomic heterogeneity within the cell population and by the genomic stability of the individual cells. Quantitative assessment of these traits can therefore be used to follow genomic evolution and estimate its prevalence (BOX 1).

Genetically engineered mouse models

De novo tumour evolution. GEMMs are a powerful tool to study tumour heterogeneity and follow tumour evolution. They have been used extensively for these purposes, especially with the advance of technologies to edit the genome and to molecularly profile tumours (reviewed previously[7,46,47]). As GEMMs are primarily generated by manipulating a single gene, or a handful of genes, genomic evolution of the manipulated tissue must occur in order for tumours to form. Numerous studies have shown that diverse routes of genomic evolution could lead to the formation of molecularly distinct tumours within the same mouse model[48,49]. Nonetheless, genomic evolution is not stochastic; we and others have shown

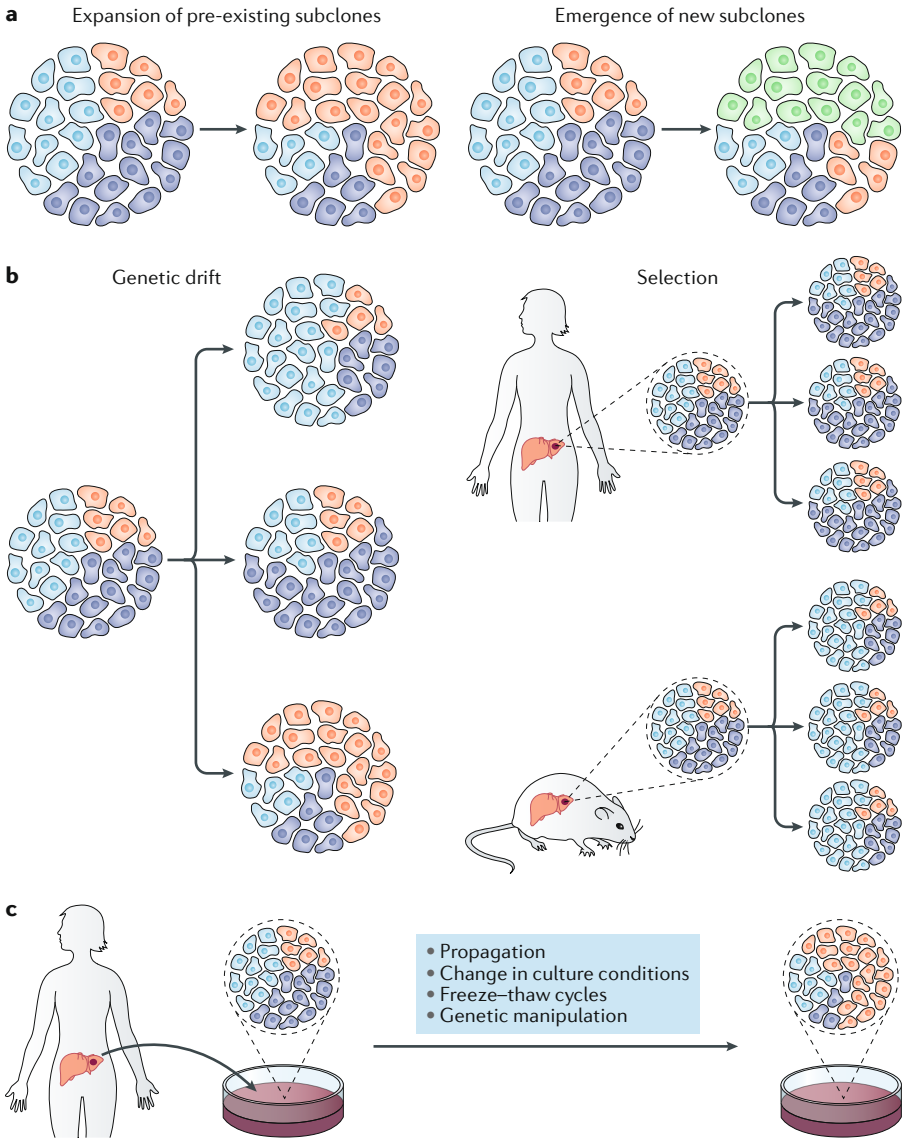


Fig. 1 | **The biological origins of cancer model evolution. a** | Genomic evolution could be the outcome of clonal dynamics that lead to the expansion of pre-existing subclones (left) or the outcome of the emergence of new subclones during the derivation or the propagation of the model (right). **b** | In both cases, such evolution could result from a genetic drift, which would lead to stochastic changes (left), or from clonal selection, which would lead to reproducible changes (right). Selection pressures are different between the natural tumour environment in the patient's body and the new environment of the model (for example, mouse in the case of patient-derived xenografts). **c** | Bottlenecks associated with model propagation can promote genomic evolution. In established cell lines, the main bottlenecks are extensive propagation, changes in culture conditions, multiple freeze–thaw cycles and genetic manipulations that involve viral infection and/or antibiotic selection.

that specific transgenes induce specific secondary genetic events, thus identifying driver-specific evolutionary trajectories of tumorigenesis[50–52]. Tumour formation in GEMMs is thus inherently associated with genomic evolution. Despite differences in the tumour genomic landscapes and in the relatively short time for tumour formation required in mice compared with humans[46], this type of evolution is clearly a desired trait of GEMMs, as it largely mimics tumour evolution in human patients[49,52,53].

Genetic instability in mouse colonies. Applying GEMMs for research involves their continuous breeding for the purpose of colony expansion and maintenance, which generates a risk of genomic diversification throughout the generations. Two populations of the same laboratory mouse strain will ultimately evolve in different directions if maintained and propagated separately. On the basis of spontaneous mutation rates, 0.96 deleterious germline mutations are expected to arise in wild-type laboratory mice each generation, and this number is much higher in genomically unstable mice[54]. As colony maintenance relies on inbred breeding, there is an ~25% likelihood for a new mutation to become homozygous and thus fixed in the population.

Indeed, different substrains of common mouse strains are acknowledged in the literature. For example, the commonly used C57BL/6 strain has evolved into multiple substrains, with the two major ones — C57BL/6J and C57BL/6N — differing in 34 single-nucleotide variants, 2 indels and 15 structural variants in coding genes[55]. This genetic variation is associated with marked phenotypic variation[55–58] and can severely mislead the interpretation of experimental results[59–61]. Therefore, the US Institute for Laboratory Animal Research (ILAR) assigns unique identifiers that designate the laboratory in which mice were maintained (see Related links), and strain diversification is thus reflected in proper mouse strain nomenclature.

Patient-derived models

Established cancer cell lines. The diversification of established cell lines (ECLs) in culture has been appreciated for decades, and some of the most commonly used cancer ECLs, such as the cervical cancer cell line HeLa and the breast cancer cell line MCF7, have become notorious for it[62–67]. In fact, similar to the mouse nomenclature, it has become common practice to designate the exact strain of HeLa cells used in one's study (for example, HeLa-CCL2, HeLa-S3 or HeLa-Kyoto). Nonetheless, ECLs are often considered to be clonal and stable for most applications, as evident by efforts to provide definitive characterization of their genomic landscapes and cellular dependencies[8–14], as well as by lack of routine documentation of culture history (for example, passage number is rarely tracked, reported or controlled for).

Recent molecular analyses comprehensively characterized the genomic variation that exists between strains of ECLs, revealing rather extensive heterogeneity and evolution. Comparing multiple strains of commonly used cancer and non-cancer ECLs, we found evidence for extensive genetic variation at all genetic levels — point mutations, rearrangements and copy number alterations — and affecting many cancer-related genes. A comparison between >100 ECLs cultured independently in 2 laboratories (in the USA and the UK) found an ~20% chance that a mutation would be detected in

Table 1 | **Determinants of genomic evolution in cancer models**

Model type	Major advantages for research	Factors impacting genomic evolution
Genetically engineered mouse models	• De novo tumorigenesis in vivo • Interactions with the microenvironment and with other cell types	• Somatic evolution of the tumour as an integral part of tumorigenesis • Germline evolution of the host throughout colony propagation
Patient-derived cell lines	Short time and few cell divisions from primary tumours to functional assays	• Physical constraints (2D) • Variations in culture conditions (media, passaging practices, etc.) • Continuous selection for rapidly proliferating cells
Established cell lines	• Widely accessible and easy to work with • Ample genomic data available	• Numerous cell divisions • Variations in culture conditions • Deficient mechanisms of genome maintenance inherited from tumour of origin
Patient-derived xenografts	• No growth on plastic • Functional investigation of human tumours in vivo	• Differences in physiology and metabolism between species • Immune-deficient environment • Site of transplantation • Multiple cell divisions within each passage
Patient-derived organoids	• Complex cellular interactions • Culture conditions mimic in vivo conditions better than 2D culture • 3D environment • Matched normal controls	• Variations in culture conditions • Deficient mechanisms of genome maintenance inherited from tumour of origin • Immune-deficient environment

only 1of the 2 compared strains[68]. Variation in gene expression mirrored genetic variation, and specific genetic alterations were often associated with a transcriptional signature of the genetically perturbed pathway[68]. Observations from 14 strains of HeLa cells revealed similar magnitudes of genetic and transcriptional variation, which translated into corresponding proteomic variation as well[69]. In both of these studies, routine maintenance of cells under standard culture conditions resulted in rapid genomic diversification within a handful of passages. In agreement with these findings, a recent analysis of RNA sequencing data from eight cell lines found genetic variability across laboratories, which was associated with gene expression variability[70]. Therefore, the genetic, transcriptional and proteomic landscape of cancer ECLs keeps evolving in culture, leading to genomic differences across separate cultures of the same ECL.

In addition to genetic variation, cell line diversification could be induced by epigenetic variation. Given their more transient nature, epigenetic marks would likely be more sensitive to culture conditions than DNA sequences, making cell lines susceptible to epigenetic instability. Indeed, instability of DNA methylation and erosion of X chromosome inactivation have been documented throughout the culture of human pluripotent stem cell lines[71,72]. Similarly, continuous propagation of human mammary epithelial cells is associated with silencing of the tumour suppressor $p16^{INK4A}$ by hypermethylation of its transcription start site[73], demonstrating that epigenetic instability can alter the tumorigenicity of cultured cell lines. Systematic analysis of epigenomic variation across strains of cell line cultures is yet to be performed.

New patient-derived cell lines. As described above, cancer ECLs, which were derived from patients many years ago, experience ongoing genomic evolution. To what extent would such evolution also affect recently derived early-passage PDCLs? Recently, systematic attempts to generate PDCLs have been initiated in order to increase the diversity of genetic alterations and tumour lineages available for cancer research[74–76]. It has been hoped that these freshly derived PDCLs would resemble their tumours of origin better than ECLs that have been cultured, and have evolved, over decades.

Although PDCLs were reported to largely retain the genomic features of primary tumours, differences between these models and their parental tumours were observed[77–80]. We recently analysed copy number alterations (CNAs) in 38 samples of PDCL models from 5 cancer types and determined how the genomic landscapes of these models evolved throughout their derivation (p0 or p1) and early propagation (through p20)[81]. We found evidence for continuous genomic evolution throughout passaging, with an average of ~20% of the genome differentially affected by CNAs between early and later passages[81]. Interestingly, CNA landscapes of PDCLs evolved more rapidly during the first few passages (lower than p5) than the later time points (greater than p10). This suggests that the derivation of the models is associated with genomic evolution and that the models eventually become more stable as they adapt to their new environment.

Patient-derived xenografts. PDXs are considered to be more physiologically relevant than cell lines and to mimic the human disease more accurately, as their generation and propagation do not involve their culture in artificial in vitro conditions (reviewed previously[4,33]). However, the in vivo xenograft environment is quite distinct from the original patient environment. First, metabolism and physiology differ between species. Second, PDXs are commonly transplanted subcutaneously, exposing the tumours to signalling cues, cellular interactions and mechanical constraints utterly different from their native microenvironment. And third, the lack of a functioning immune system in immune-compromised mice could alter tumour development and behaviour[82].

Indeed, PDXs undergo genomic evolution throughout their derivation and propagation. When engraftment and early-passage propagation of breast cancer PDX models were monitored using single-cell sequencing, extensive clonal dynamics were observed, which drastically changed the abundance of mutation clusters (and thus the allele fraction of mutations) throughout serial PDX passaging[83]. Similarly, the engraftment of human acute lymphoblastic leukaemia in mouse xenografts was shown to be associated with genomic evolution into a more aggressive malignancy[84], and engraftment propensity varied considerably between genetically distinct clones of acute myeloid leukaemia (AML)[85,86]. Our analysis of CNAs in 543 unique PDX models across 24 cancer types found that ~60% of the models acquired at least 1 large chromosomal aberration within a single passage and that ~90% acquired at least 1 such aberration within 4 passages[81]. Similar to the observation in PDCLs, genomic evolution was most rapid during PDX initiation and early passaging, and its rate

Box 1 | **Measuring genomic evolution in cancer models**

The ability to follow the genomic evolution of cancer models has considerably advanced in recent years. First, the drop in the costs of DNA and RNA sequencing has increased standard sequencing depths[121], enabling the detection of genetic alterations that are rare within the bulk tumour population and the characterization of the genomic composition of cancer models at multiple time points throughout their propagation[52,81,83]. Second, single-cell sequencing technologies now enable cellular heterogeneity to be studied at the resolution of individual cells[83,108,109], and improvements in the isolation and expansion of clones from single tumour cells enable their functional interrogation[115]. Third, analytical tools have been developed that use genomic data to infer the clonal structures of cell populations[107,122,123], and additional tools make use of genomic data to characterize and quantify signatures of genomic instability[124–126]. Together, these methods now enable the measurement of both the heterogeneity and the instability of cancer models in much greater detail than before. It is important to note that genomic heterogeneity and genomic instability, although conceptually related and quantitatively correlated[81,107], are not synonymous terms — heterogeneity refers to the genetic variation within the cell population, whereas instability refers to the stability of that genetic composition over time. Measuring both of these traits in cancer models is required to understand how these models evolve over time.

considerably decreased at later passages[81], consistent with strong selection pressures being associated with model initiation.

Importantly, the rate of genomic evolution, defined as the fraction of the genome altered per passage, was similar in PDXs and PDCLs[81]. Comparing matched PDXs, a median of ~12.5% of the genome was differentially affected by CNAs within four passages. One should note, however, that an in vitro passage normally involves many fewer cell divisions than an in vivo passage, so the rate of change per cell division may be lower in vivo than in vitro. Nevertheless, these findings challenge the notion that PDXs better preserve the genomic landscapes of primary tumours.

It is worth noting that PDXs are also susceptible to epigenomic evolution throughout their propagation. A recent study that compared the methylation patterns between primary non-small-cell lung cancer tumours and their derived PDXs reported somewhat low genome-wide Spearman correlations (ranging from 0.37 to 0.49)[87]. Although many of the observed differences were related to tumour purity, others potentially reflected epigenetic instability[87].

Patient-derived organoids. PDOs have emerged as physiologically relevant 3D model systems to study cancer in vitro (reviewed previously[3]). Long-term organoid cultures have been established from multiple epithelial cancer types by embedding cancer cells into a 3D matrix in medium containing tissue-specific growth factors that recapitulate the tumour niche. Cancer organoids have several advantages over 2D cell lines, including more complex cellular composition. The culture conditions also permit expansion of normal epithelial cells, providing the ability to obtain matched normal organoids from the same patients[3].

Several arguments suggest that tumour organoids may experience less genomic evolution than their 2D counterparts and thus better retain tumour genomic features. First, organoid derivation from primary tumours is more efficient than that of PDCLs[3], suggesting that their generation might be associated with less of a population bottleneck. Second, organoid culture conditions better recapitulate those of the original tissue[3], potentially alleviating some of the selection pressures entailed in the in vitro transition. Third, genomic profiles of PDOs were found to be highly similar to those of the tumours of origin[23,28,30,31], and PDO drug response could recapitulate patient response in the clinic[31,88].

Nonetheless, PDOs are not exempt from in vitro model evolution. Extensive genetic diversity has been recently described in colorectal PDOs[89]; the derivation and long-term culture of such highly heterogeneous organoids are likely to be associated with clonal dynamics that will alter the clonal composition of the model, similar to that observed in cell lines and PDXs[68,81,83,84]. Indeed, PDOs are not perfect genomic representations of their tumours of origin, and all of the major PDO cohorts reported to date have provided examples for differences in the status of both somatic point mutations and CNAs between matched primary tumours and PDOs[23,28,30,31,90]. Importantly, strong clonal dynamics have been recently described in PDOs, leading to rapid expansion of pre-existing minor subclones[90]. As in all other models, the inherent genomic instability of cancer cells is also likely to result in de novo genetic alterations on continuous propagation of organoid models[28]. Future studies will be required to characterize the extent of genomic evolution in PDOs as matched genomic data from multiple time points throughout PDO passaging become available.

Mechanisms of cancer model evolution

The genomic evolution observed in cancer models could be the result of pre-existing heterogeneity, as clonal dynamics of pre-existing tumour subclones would lead to changes in their relative abundance. A rare primary tumour subclone could expand and become the dominant subclone in the model, thereby altering the molecular landscape of the model. Alternatively, genomic evolution could be the result of ongoing genomic instability, which could lead to the accumulation and fixation of new genetic alterations following propagation of the model (FIG. 1a). Regardless of whether it reflects expansion of pre-existing subclones, the emergence of a new clone or both, genomic evolution could be either a stochastic or a deterministic process depending on whether it is a consequence of genetic drift or of genetic selection (FIG. 1b). In either case, bottlenecks associated with the model derivation and propagation would enable and expedite genomic evolution (FIG. 1c). However, if selection is involved, an important question is whether evolutionary trajectories in cancer models reflect evolutionary trajectories in the patients from which they are derived (FIG. 1b).

Clonal dynamics

Molecular heterogeneity has emerged as a fundamental characteristic of most tumour types[2]. Patient-derived cancer models would thus begin as heterogeneous cell populations. The extent to which models remain heterogeneous, and the strength of clonal dynamics during model establishment and passaging, could determine the stability of the models and whether they accurately mirror their parental tumours.

Clonal dynamics play an important role in the genomic evolution of cell lines, PDXs and PDOs. Stochastic cell transitions have shown clonal dynamics within cell line populations at the phenotypic level[91]. Genetic clonal dynamics have been recently described in cancer ECLs, in which the relative abundance of genetic subclones varies across strains of the same cell line[68]. Single-cell genomic analyses of breast cancer PDXs revealed that expansion of minor subclones could drastically change the clonal composition of the model compared with the original tumour[83]. Such clonal dynamics were also identified in PDXs of multiple additional cancer types[81], as well as in oesophageal cancer PDOs[90], and were

not limited to only the model derivation stage, as significant differences were also observed between early-passage and late-passage PDXs[81,83].

Genetic drift is likely to be involved in the clonal dynamics observed in models, especially during model initiation, given that models are typically generated from a biopsy taken from a small, randomly selected tumour region. However, in both cell lines and PDXs, selection has been shown to be a major driver of clonal dynamics. The expansion of minor subclones in PDXs was found to be reproducible; when the same tumour population was transplanted into different mice, expansion of the same minor subclones was observed, indicating non-stochastic, directional genomic evolution[81,83]. Similarly, DNA barcoding experiments in ECLs showed that minor changes in culture conditions, such as changing the medium from RPMI to DMEM, resulted in reproducible changes in barcode abundance[68]. Moreover, stronger selection pressures, such as drug exposure, resulted in stronger clonal dynamics. Therefore, selection-driven clonal dynamics clearly play an important role in genomic evolution of cancer models.

Ongoing genomic instability

Genomic instability is another fundamental trait of tumours, which is carried over from the primary tumours to the models derived from them. The nature and rate of the genetic alterations that arise in the model depend on the genome integrity mechanisms that are perturbed in the tumour of origin. Moreover, some types of genomic instability are exacerbated in model systems. For example, the fidelity of chromosome segregation was recently shown to depend on integrin function and thus diminish when cells were grown in 2D culture[92].

Evidence for ongoing genomic instability in cancer models is extensive. As expected, GEMMs generated by perturbation of genes directly related to genome integrity maintenance (most notably, *Trp53*) are more genomically perturbed than GEMMs generated with other classes of oncogenes or tumour suppressors[52,93,94]. Similarly, PDXs from p53-null tumours experience more copy number evolution than those from p53-wild-type tumours[81]. Furthermore, the mutation landscape variability across ECL strains is higher in ECLs with microsatellite instability (MSI) than in ECLs without it[68], which is in line with the hypermutation phenotype of MSI human tumours.

We recently demonstrated that single-cell clones derived from ECLs quickly became genetically, transcriptionally and phenotypically heterogeneous[68]. This means that ongoing genomic instability leads to the constant emergence of new subclones, which can eventually expand and alter the genomic composition of the model.

Bottlenecks of model propagation

The magnitude of genomic evolution depends on the heterogeneity of the population and on the stringency of the bottlenecks cells need to go through. Some of these bottlenecks are inherent to the nature of cancer models, whereas others depend on experimental practices that could be modified. It is therefore imperative to understand the bottlenecks associated with model derivation and propagation.

The first strong bottleneck that every model encounters is the founder effect associated with its establishment. The tumour biopsy from which a model is derived represents only a specific tumour region and therefore a local clonal composition[2]. At the same time, the need to survive in a new environment and adapt to markedly different conditions makes model initiation a highly selective process[5].

Routine propagation of cancer models continues to present bottlenecks for the cell population. Passaging of cell lines, PDOs or PDXs involves continuous competition that favours the fitter, more rapidly dividing cells. Propagation conditions inevitably vary, and the exact conditions (for example, media composition, batches of reagents or cellular densities) may affect model evolution. Bottlenecks are also introduced by freezing and thawing, another routine practice in the propagation of in vitro cultured models. Finally, genetic manipulations, including those considered to be neutral (such as the introduction of a reporter gene), introduce bottlenecks in the form of viral infection and antibiotic selection.

We recently found that in ECLs, most of the variability across strains was introduced through extensive passaging or genetic manipulations, whereas multiple freeze–thaw cycles did not seem to induce extensive genomic evolution[68]. However, the generalizability of this finding is yet to be confirmed.

Distinct trajectories of tumour evolution

As selection plays a major role in shaping the genomic landscapes of cancer models, it is important to assess whether model evolution mirrors the tumour evolution that naturally occurs in patients. Data from both cell lines and PDXs indicate divergent trajectories of tumour evolution in cancer models and in patients. An analysis of genomic and functional heterogeneity in AML revealed that the AML founding clone was not necessarily the AML-initiating clone in the mouse model[85]. In solid tumours, recurrent cancer-type-specific CNAs that are commonly observed in primary tumours tend to become even more recurrent during cancer progression (that is, metastases and recurrences); however, the same events tend to disappear in PDXs and in cell lines[68,81]. Therefore, genomic landscapes of tumours are shaped by distinct selection pressures during their evolution in the natural human environment and in the artificial model environment, leading to the gradual genomic divergence of cancer models from their tumours of origin.

Implications for research and medicine

The implications of genomic evolution of cancer models for research and clinical translation depend on whether it can alter their performance in functional assays. Different model applications are expected to be differentially affected by this phenomenon. Once potential risks are identified, mitigation strategies can be devised and implemented to properly assess, alleviate and control for genomic evolution.

Phenotypic consequences

It is conceivable that, although extensive genetic and epigenetic evolution occurs within cancer models, such evolution might not translate into biologically meaningful cellular properties. That is, this phenomenon could represent a molecular curiosity with little practical implication. Unfortunately, this does not seem to be the case. Genomic changes that arise throughout model propagation have been associated with marked phenotypic changes across all cancer model types (FIG. 2a).

In mice, a comparative phenotypic analysis of two substrains of the C57BL/6 strain demonstrated significant phenotypic differences between these strains in multiple physiological, biochemical and behavioural systems, ranging from blood pressure and eye morphology to bone structure and spatial memory[55]. Variation in morphological and behavioural traits was described between multiple C57BL/6 strains, and specific DNA-level differences were suggested to underlie these phenotypic differences[56,57]. Such genomically driven phenotypic differences can be carried over to GEMMs generated on the genetic backgrounds of these mice: for example, a *Dock2* copy number variant that spontaneously arose in C57BL/6 was

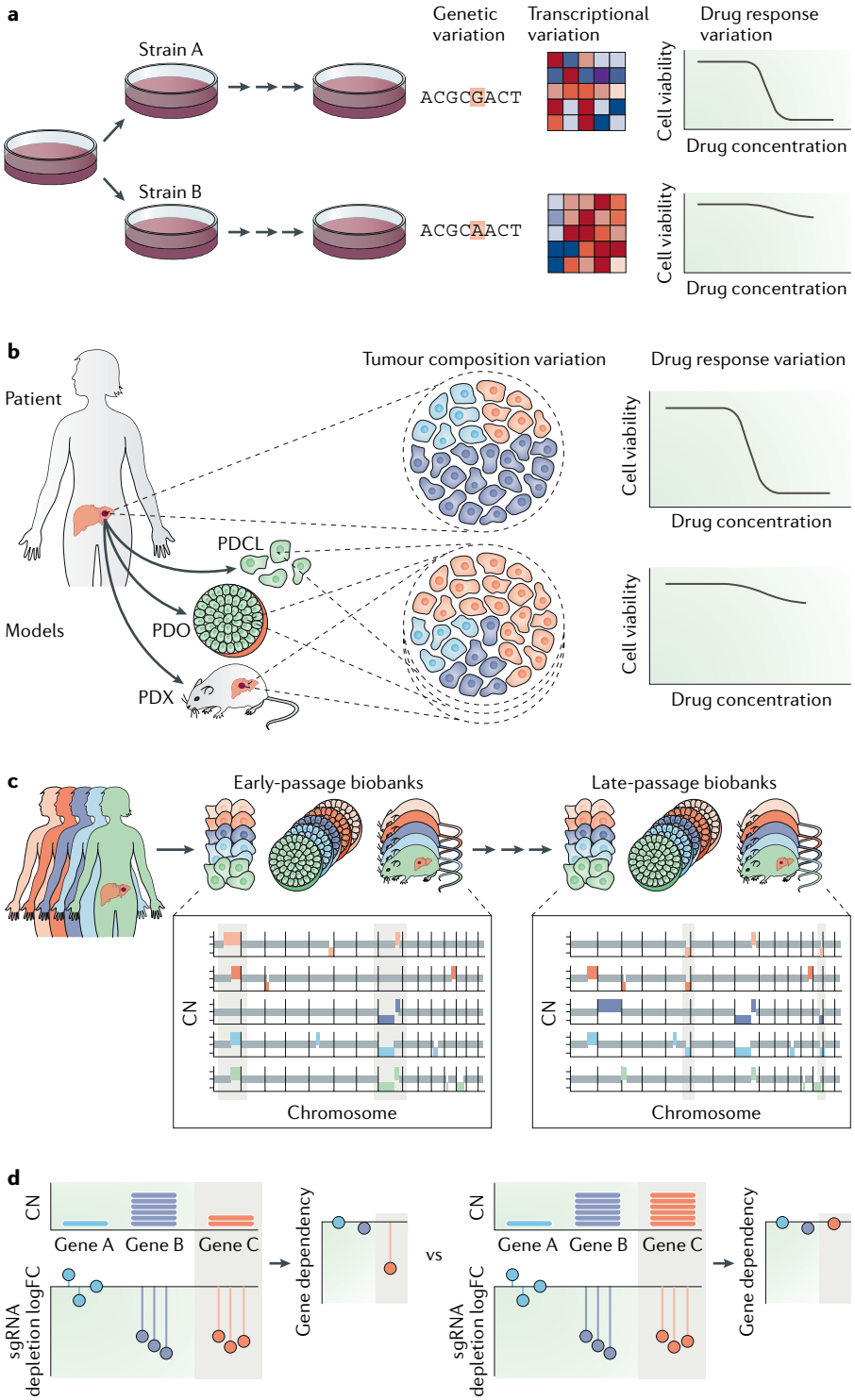


Fig. 2 | **Perils of cancer model evolution.** **a** | Genomic evolution leads to variability across strains of the same model. Genetic alterations acquired during model propagation can translate into differential gene expression patterns and result in disparate drug response[68]. **b** | Cancer models are used as tumour avatars to predict drug response in the patient from which the model was derived. However, if a rare subclone expands in the model and becomes dominant, its drug response may not reflect that of the primary tumour[81]. **c** | Cancer models are also used as cohorts, also known as biobanks, to characterize genotype–phenotype relationships (for example, cancer dependencies associated with specific mutations). Models are usually characterized upon their derivation, but functional experiments keep being conducted throughout model life. Continued passaging of the models may alter their genomic profiles, thus confounding analyses that make use of early-passage genomic profiles and late-passage functional experiments[81]. **d** | CRISPR–Cas9 screen results need to be corrected for copy number to account for the number of cuts (a phenomenon known as the CRISPR copy number effect). Changes in aneuploidy (and other copy number alterations) as a result of genomic evolution can jeopardize the accuracy of this computational correction. For example, if an amplification of gene C occurred between genomic characterization and CRISPR screening, and the original genomic characterization is used for correcting the copy number effect, this gene could be misidentified as a genetic dependency. Concepts shown in this figure are based on REFS[68,81]. CN, copy number; logFC, $\log_2$ fold change; PDCL, patient-derived cell line; PDO, patient-derived organoid; PDX, patient-derived xenograft; sgRNA, single guide RNA.

inadvertently backcrossed into multiple mutant mouse lines; as a result, these mice exhibited several immune phenotypes previously described in the context of *Dock2* deficiency[60].

In human cell lines, considerable phenotypic variation across strains of the same cell line has been described and linked to genomic variation. In our studies, 27 strains of the MCF7 cell line collected from multiple laboratories varied in their doubling time by as much as 3.5-fold under identical culture conditions. Morphological features, such as cell size and shape, largely varied as well, and morphological variation correlated well with genomic variation[68]. Similar differences in doubling time and morphological features were observed across 14 HeLa strains[69]. These HeLa strains also varied considerably in their susceptibility to *Salmonella* infection, which could be explained by differential expression of proteins associated with bacterial infection[69].

One of the most common and important applications of cancer models is their use for testing of drug sensitivity and resistance. Variability in drug sensitivity of the same cell line to the same drug is common in the literature: for example, reported half-maximal inhibitory concentration (IC_{50}) values of tamoxifen in wild-type MCF7 cells differ by >100-fold[10,95]. To evaluate whether such differences could result from genomic evolution, we screened 27 strains of the MCF7 cell line against 321 cancer-related compounds and compared the basal gene expression of sensitive and resistant strains (forward genetics approach). A striking variability was observed: for 48 of 55 (87%) drugs for which at least one strain was highly sensitive, another completely resistant strain was also identified. In the majority (82%) of cases, a differential gene expression signature of the drug mechanism of action was observed between the sensitive and resistant

strains[68], highlighting the idea that genomic variation indeed underlies the disparate drug response. Similarly, PDX response to some targeted therapies was associated with the existence of specific CNAs, and these associations were confirmed in ECLs[81]. Therefore, the functional effect of genomic evolution on drug response in PDXs may be as pronounced as the effect observed in cell lines.

Implications for tumour avatars

Avatar experiments, also known as co-clinical trials, match patients' drug responses to those of the models derived from them. The idea is that testing multiple drugs against a patient-derived cancer model could help direct the course of clinical treatment for that patient. Several such co-clinical trials reported a high degree of concordance between the drug response of patients and their PDXs and PDOs[24,31,88,96,97].

Although this strategy is intriguing and useful, several cases in which model drug response did not match patient's response have been described[24,31,88]. Such discrepancies could result from changes in clonal composition of the tumour model owing to genomic evolution (FIG. 2b). This could be especially important if genomic evolution affected driver events, such as *ESR1* status in ER$^+$ breast cancer (as observed throughout MCF7 cell line evolution[68]) or trisomy 7 and monosomy 10 in glioblastoma (as observed in glioblastoma PDXs[81]). In addition, differences in tumour microenvironment (for example, exposure to ligands and cytokines, matrix stiffness, extracellular matrix protein composition, and so on) can profoundly influence cell growth and therapeutic response (reviewed previously[98,99]). Further analyses are warranted to evaluate the extent to which evolution of cancer models, together with differences in the tumour microenvironment, limits their use as avatars and how these limitations can be overcome.

Implications for tumour model biobanks

Efforts to generate large panels of cancer models aim to represent the molecular diversity of patient populations and are commonly queried to identify phenotypes associated with specific molecular features (for example, drug sensitivity associated with a gene mutation). For this type of application, the similarity of a specific model to its tumour of origin is much less relevant. However, two important questions emerge in this setting: to what extent does the cohort as a whole represent the patient population, and how stable and reliable are the molecular features characterized in such cohorts?

Overall, large cohorts of cell lines, PDXs and PDOs mirror the genomic landscapes of their respective cancer types[23–27,30,100]. However, a couple of important concerns remain. First, continuous passaging can distance the model population from the patient population and/or skew the represented populations. Indeed, some hallmark genetic alterations were found to be less prevalent in high-passage than in low-passage PDXs[81]. Second, biobanks of cancer models are usually characterized only once, on their derivation, and the resultant data sets of molecular features are then widely used as 'lookup tables' by individual investigators. However, as models evolve, these lookup tables may be misleading — a given genetic alteration characterized in the model may be absent from the specific strain that happens to be available to an investigator and vice versa[68] (FIG. 2c).

As discussed above, genomic evolution is most rapid during the derivation and early propagation of cancer models. Genomic evolution then continues to shape the genomic landscape of the model on continuous passaging, albeit at a slower rate. It would therefore be optimal to characterize and use the models at a point when they are already sufficiently stable but are still good representations of their tumours of origin. In PDCLs and PDXs, such an optimal time point might be between p5 and p10 (REF.[81]), but this optimum would clearly be context and model specific.

Recently, a new practice emerged to first generate PDXs and then derive PDOs from them. This practice aims to have the best of both worlds, as the derivation success rate is generally higher in xenografts, whereas organoids are easier to manipulate and study at high throughput[23,101]. However, this strategy also adds strong bottlenecks to the process, as the model is being transferred twice, from a patient to a mouse and then from mouse into cell culture. It is likely that this multi-step derivation strategy would be associated with genomic evolution, especially given that the opposite transition — the transplantation of ECLs into mice — was shown to induce such evolution[81].

Implications for screening purposes

Genomic evolution of cancer models can greatly affect chemical and genetic screens. In a typical primary chemical screen, compounds are screened at a single high dose against one randomly selected strain of a cell line. The striking variability in drug response observed across strains of the same cell line suggests that very often whether a compound is identified as a hit in a screen may depend on the particular cell line strain that was screened. This problem may be exacerbated in genetic loss-of-function screens, which typically involve genetic manipulations and antibiotic selection that can act as strong bottlenecks that increase diversification.

CRISPR screens are probably most susceptible to this problem. First, the introduction of Cas9 and guide RNAs into cells often involves two selection steps, enabling more clonal selection. Second, genetic alterations that enable cells to tolerate DNA cutting may be selected in the process; for example, two recent studies showed that genome editing by CRISPR–Cas9 leads to a selection against cells with a functional p53 pathway[102,103]. Third, CRISPR screen results need to be corrected on the basis of the copy number landscapes of the screened cells, and this is commonly done using CNA profiles of the parental cell lines[13,104,105]; the more CNA evolution that took place from the time of the original profiling of the wild-type strain until the actual screen was performed with the CRISPR–Cas9 strain, the less accurate the computational correction of the copy number effect would be (FIG. 2d). Indeed, genome-wide CRISPR screens of two remote MCF7 cell line strains revealed distinct gene essentiality patterns, and some of the differential dependencies could be readily explained by genomic alterations of the underlying genes[68].

Implications for reproducibility

When results obtained in one laboratory cannot be replicated in another laboratory, one should consider the potential contribution of genomic evolution to the observed discrepancies. For example, mispairing C57BL/6 substrains of GEMMs and wild-type controls led to opposite results related to the role of *Jnk2* (also known as *Mapk9*) in liver injury[59]. Of note, many publications that make use of C57BL/6 mice do not report which specific substrain was used.

The recent genomic and phenotypic comparisons of multiple MCF7 and HeLa cell line strains provide another compelling demonstration of the potential of genomic evolution to jeopardize cancer research reproducibility. For example, studying sensitivity to proteasome inhibition using the MCF7 cell line can clearly lead to different conclusions on the basis of the strain used[68]. Importantly, however, although disagreements between large data sets of

chemical and genetic dependencies do exist[38–40], it is not clear yet to what extent this reflects genomic differences between strains.

Mitigation strategies

Many of the risks associated with cancer model evolution can be mitigated by adjusting model propagation strategies and experimental designs. Mitigation strategies can be divided into three classes: tracking and reporting model propagation; routine assessment of model genetic diversification; and alleviating propagation bottlenecks. In addition, robustness of findings should be confirmed across collections of model systems that represent the relevant aspects of cancer being modelled in different genomic and environmental contexts.

Track and report propagation. A requirement for a clear nomenclature of models — taking into account the model history and strain — could make investigators more aware of the phenomenon and prevent mispairing of controls. Keeping track of passage and generation information and reporting it in publications are desirable practices; the absolute passage number of a model does not necessarily carry useful information, but matching passage numbers can guarantee a uniform and appropriate within-study use of the model.

Assess diversification. Routine monitoring of genomic evolution could alleviate its detrimental outcomes. Assessment of diversification is needed after prolonged propagation or when models go through strong bottlenecks such as genetic manipulations or single-cell cloning. As the genetic distance between cell line strains correlates very well with the expression distance and the drug response distance[68], inexpensive commonly used genomic technologies, such as low-pass whole-genome sequencing, can be applied to accurately assess cell line diversification. Characterizing the genomic features of the model at the same time that it is subjected to functional experiments is therefore both desirable and feasible.

We developed Cell STRAINER (Cell STRAin Instability profilER; see Related links) to facilitate the routine assessment of cell line diversification. Users can upload cell line genomic data of their cell line strain (currently in the format of CNA profiles), and these data are compared with reference genomic data of the same cell line, as characterized by the Cancer Cell Line Encyclopedia (CCLE)[8]. Large genetic distance between the two strains is indicative of considerable cell line diversification, suggesting that other genomic features of the CCLE strain may also not apply to the tested strain.

When cell line misidentification and mycoplasma contamination emerged as major problems in cell line research, journals began to require authors to confirm cell line identity (using DNA fingerprinting) and mycoplasma contamination status. Similarly, we can envision that cell line genetic diversification assessment may become a prerequisite for publication in the future. Of note, the same genomic analysis can also confirm cell line identity and therefore replace DNA fingerprinting in most cases. As biobanks of PDXs and PDOs become more widely used, tools similar to Cell STRAINER should be developed for these models as well.

Minimize genomic evolution. Experimental practices can be adjusted to minimize genomic evolution and its phenotypic effects. First, it is advisable to avoid unnecessary passaging by using techniques that minimize cell culture after the initial characterization[106]. Second, bottlenecks should be reduced as much as possible. Practically, this means keeping culture conditions as constant as possible, as even minor changes can lead to clonal selection[68]. This also means that single-cell cloning should be avoided when population-based assays can be used instead. Better mimicking of the human tumour environment in the model — a desirable goal regardless — could also alleviate selection pressures and minimize genomic evolution.

Given that prolonged passaging leads to continuous model evolution, multiple frozen stocks should be prepared for large studies so that models can be used at comparable passage numbers throughout the entire course of the study. This could be especially important for large-scale screens, in which downstream follow-up experiments are often performed many months after the original screen.

Finally, some models are inherently unstable and are thus prone to much more rapid genomic evolution[81]. Sometimes the use of such models is unavoidable or even desirable (for example, when one wishes to study genomic instability); however, for some applications, these models can simply be replaced by alternative, more stable ones.

Emerging opportunities

Although the perils of genomic evolution of cancer models warrant caution, and applying mitigation strategies is recommended, this natural phenomenon could also be harnessed in creative ways. One might take advantage of the dynamic nature of cancer models to pursue novel avenues of research or to address longstanding open questions in new ways.

Glossary

Clonal dynamics
Changes in the relative abundance of tumour subclones throughout model propagation.

Copy number effect
In CRISPR screens, copy number changes result in a gene-independent anti-proliferative effect of Cas9-mediated DNA cleavage, confounding the measurement of gene essentiality. This effect can be corrected computationally using genome-wide copy number measurements.

Established cell lines
(ECLs). Models generated as patient-derived cell lines, followed by prolonged culture propagation. These models are not assumed to represent the specific tumours from which they were derived.

Founder effect
Genetic diversity that results when a cell population is descended from a small number of original cells.

Genetically engineered mouse models
(GEMMs). Models generated by genetically manipulating mice using genetic alterations that characterize human tumours.

Genetic drift
Stochastic changes in the clonal composition of the cancer cell population owing to chance disappearance and/or expansion of particular subclones.

Genetic selection
Directional changes in the clonal composition of the cancer cell population owing to growth advantage and/or disadvantage of particular subclones.

Ongoing genomic instability
Generation of de novo genetic alterations throughout model propagation, contributing to the genomic evolution of the model.

Patient-derived cell lines
(PDCLs). Models generated by the transferring of tumour cells into a 2D plastic dish using culture conditions that enable cells to proliferate.

Patient-derived organoids
(PDOs). Models generated by the embedment of tumour (or normal) cells into a 3D matrix using culture conditions that mimic the in vivo tumour niche.

Patient-derived xenografts
(PDXs). Models generated by the direct engraftment of resected human tumours into immune-deficient mice, followed by their serial transplantation between mice.

Pre-existing heterogeneity
Genetic diversity within the original tumour that contributes to the genomic evolution of the model.

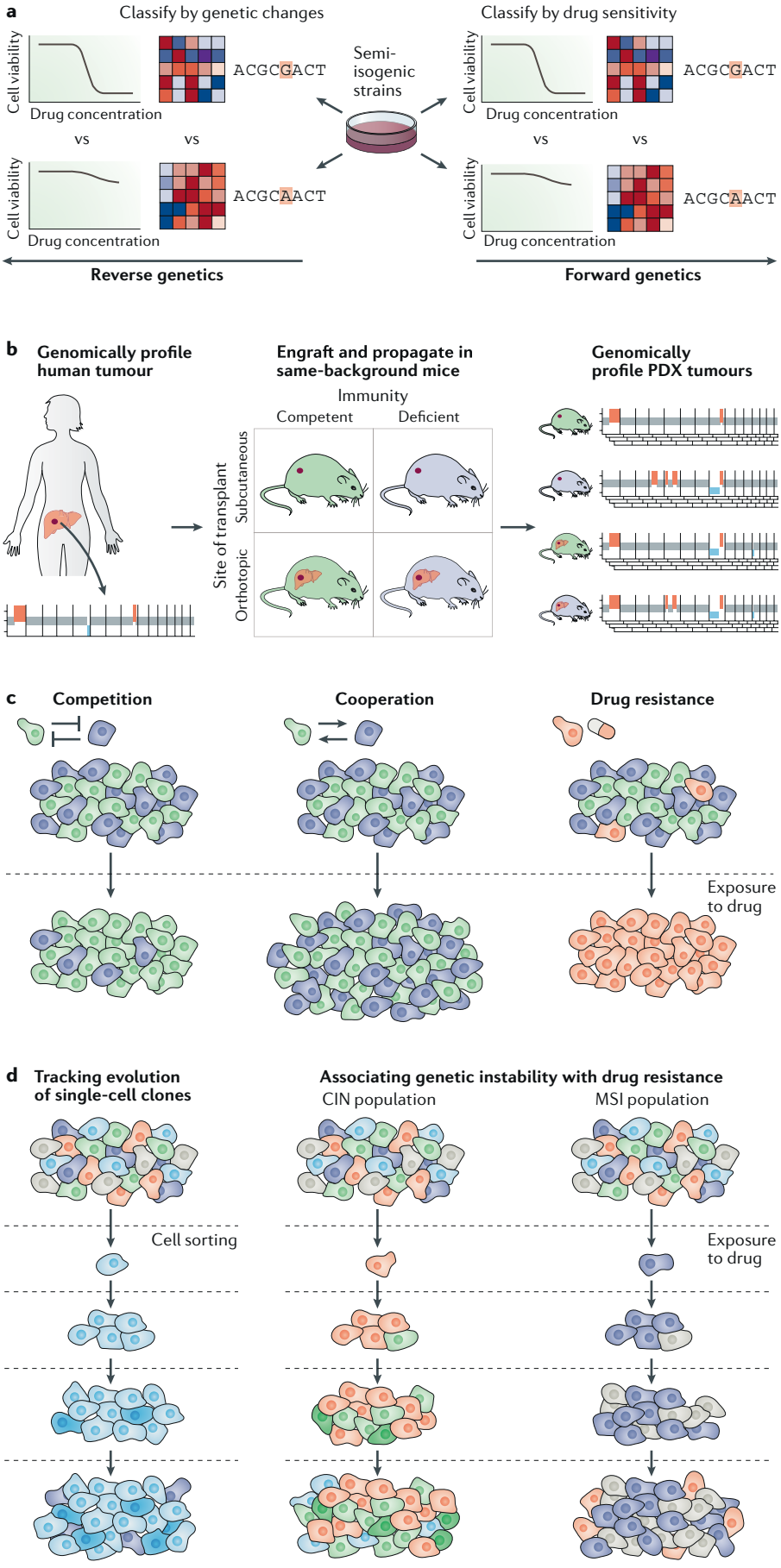


Fig. 3 | **New research opportunities presented by cancer model evolution. a** | The natural genomic evolution of cancer models generates semi-isogenic model strains that can be used for both reverse genetics and forward genetics. In reverse genetics experiments, the gene expression profiles and drug response patterns of model strains with a genetic alteration of interest can be compared with those of strains without that alteration. In forward genetics experiments, the genetic landscapes and gene expression profiles of model strains that are sensitive to a drug of interest can be compared with those of strains that are resistant to that drug. **b** | Genomic evolution can be used to study the selection pressures that shape the genetic landscapes of tumours. For example, patient-derived xenografts (PDXs) can be generated in mice hosts that share their genetic background but differ in their immune status and transplantation sites. The genomic profiles of these PDX models following propagation can be characterized and compared with that of the primary human tumour. The rate, extent and identity of genomic alterations in the various models can help to identify the components that determine the evolutionary trajectory of tumorigenesis. **c** | Cancer model evolution can be used to study cancer heterogeneity and cellular interactions. Competitive and cooperative interactions between tumour subclones can be dissected using single-cell genomics, genome editing and cell barcoding technologies. Drug sensitivity and resistance can also be studied by following tumour evolution and clonal dynamics following drug exposure. **d** | Mechanisms of genomic instability itself can be studied by following how single-cell clones become heterogeneous (left) or by following how drug resistance mechanisms differ between models that harbour distinct deficiencies in genome maintenance pathways (right). CIN, chromosomal instability; MSI, microsatellite instability.

Panels of near-isogenic cancer models

In mouse models, the realization that different substrains differ in their genomic and phenotypic features was successfully used to illuminate genotype–phenotype associations. For example, a genomic comparison of two C57BL/6 mouse substrains that differ significantly in their alcohol consumption identified candidate genes that explain this trait[56]. Application of a similar strategy uncovered the genetic alterations underlying differences in immune phenotypes[60].

The same approach can be applied to patient-derived models (FIG. 3a). Genomic evolution leads to the generation of genetically matched, near-isogenic models. These models are not fully isogenic, as more than a single event would normally separate each pair of strains. Nonetheless, panels of such near-isogenic models can be used to study specific molecular features that are variable across strains. As a proof

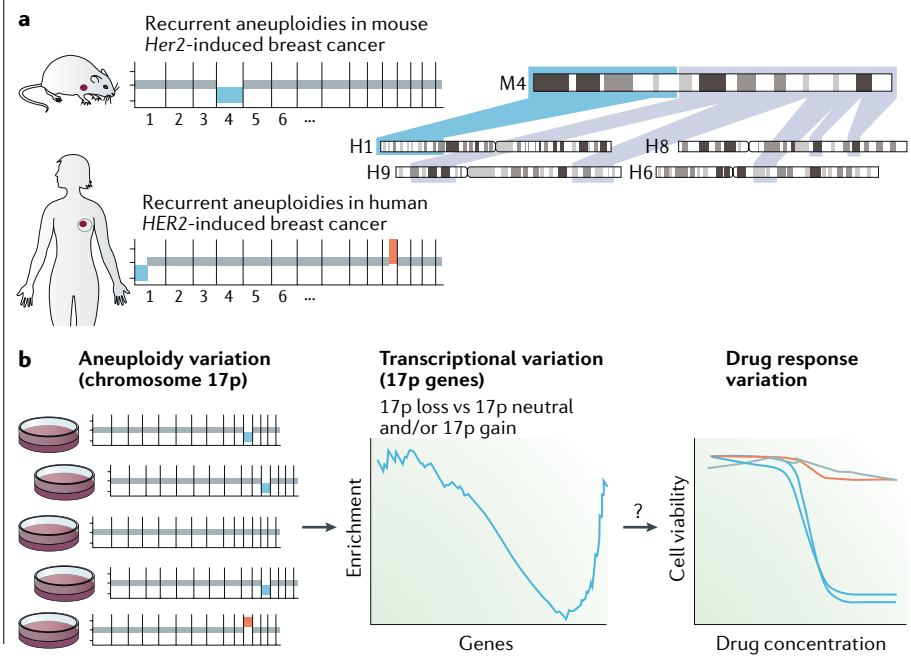
Box 2 | **Model evolution and cancer aneuploidy**

Aneuploidy demonstrates both the perils and the opportunities presented by cancer model evolution. Aneuploidy provides a unique lens through which one could follow tumour evolution, as this is a discrete event at the cellular level, is unique to cancer cells and can be detected by multiple methods[127,128]. In addition, owing to the high fitness cost associated with specific aneuploidies under most cellular circumstances[129], these events may be particularly sensitive to changes in selection pressures.

The study of aneuploidy in cancer models is exposed to all of the risks associated with model evolution. Variability in arm-level and whole-chromosome copy number status exists within established cell lines (ECLs), and aneuploidy landscapes rapidly evolve in patient-derived xenografts (PDXs) and patient-derived cell lines[81]. These genetic changes are associated with very consistent changes of gene expression along the affected chromosome(s) in the expected direction[68,81]. In both ECLs and PDXs, the existence of specific aneuploidies was associated with the response to specific anticancer drugs[81]. The gradual disappearance of recurrent cancer aneuploidies throughout the derivation and propagation of PDXs[81] emphasizes the potential risk that model evolution poses to accurately using these models as tumour avatars.

At the same time, model evolution can be used to advance the study of cancer aneuploidy. The artificial introduction of extra chromosomes into cells is tumour suppressive[130], whereas naturally occurring aneuploidy can be tumour promoting[131], as also suggested by the cancer-type-specific patterns of aneuploidy recurrence[132]. Therefore, naturally occurring aneuploid variants that arise during model evolution may be uniquely suitable for studying aneuploidy in a more cancer-relevant context.

Indeed, following aneuploidy landscapes in developing tumours in genetically engineered mouse models enabled us to identify aneuploidies that are associated with specific drivers, narrow down regions of interest within altered chromosomes and identify candidate genes that cooperate with the initial transgene to drive tumorigenesis[52] (see the figure, part **a**). Another promising direction is to use naturally occurring variation in aneuploidy landscapes within ECLs to study the cellular consequences of recurrent aneuploidies and to identify synthetic lethalities associated with these events (see the figure, part **b**).



of concept, comparison of gene expression and drug response between cell line strains with and without a given genetic alteration correctly identified cellular consequences of, and pathways perturbed by, that alteration[68]. This reverse genetics approach would be especially useful for studying the consequence of genetic events that are difficult to introduce experimentally, such as large chromosomal changes (BOX 2).

Similarly, taking a forward genetics approach as described above when comparing global gene expression levels between drug-sensitive and drug-resistant MCF7 cell line strains[68], such comparisons could uncover the genomics underlying variable phenotypes. Importantly, such comparisons provide more statistical power than similar comparisons of non-genetically matched cancer models.

However, this approach is limited to studying differential phenotypes and so is not useful for studying drugs that do not affect that particular cell line. Another limitation is that ongoing clonal dynamics may affect the stability of the panel, rendering the genomic profiling of the strains a moving target.

Understanding selection pressures

Because both positive and negative selection can alter the genomic landscapes of patient-derived tumours, cancer models could potentially be used to study selection itself. For example, the immune-deficient subcutaneous mouse environment of PDXs is different from the immune-competent organ-specific patient environment. Which of these components is most important for shaping the trajectory of tumour evolution? Following the rate, extent and identity of genomic alterations in subcutaneous versus orthotopic models, and in immune-deficient versus humanized mice, should help to clarify the relative importance of each of these factors (FIG. 3b). Similarly, following clonal dynamics in PDOs or cell lines cultured under various defined culture conditions may illuminate the role of specific media components (for example, growth factors) in shaping the genomic landscape of the tumour.

Studying heterogeneity

Tumour heterogeneity is a fundamental aspect of tumour biology with profound implications for drug response and clinical outcome[2,107]. Single-cell technologies have revolutionized the way cancer heterogeneity is studied and understood, and deeper sequencing allows more accurate reconstruction of tumour clonal composition[108,109]. The heterogeneous, dynamic nature of cancer models raises interesting opportunities for research (FIG. 3c).

Cooperative and competitive interactions.

There is growing evidence that cooperative and competitive interactions between tumour subclones can influence disease progression and clinical outcome (reviewed previously[110]). Using engineered subclones of a human breast cancer ECL, intra-tumour clonal heterogeneity was shown to drive tumour growth and dissemination[111]. Similarly, cooperation between subclones was found to promote invasion in a zebrafish melanoma xenograft model[112], and differentiated cell populations could increase invasiveness

and growth of cancer stem cell populations through factor secretion in pancreatic and glioblastoma cell lines, respectively[113,114]. Most recently, functional cooperativity was described between genetically distinct subclones derived from human paediatric brain tumours[115].

The realization that cancer models are naturally heterogeneous suggests that they can be used to dissect the mechanistic basis of clonal interactions. Single-cell profiling of cell lines might help to determine the abundance of existing subclones, and clonal dynamics throughout various types of perturbations and challenges can then be followed. Experiments such as the one described above can potentially be performed without cell engineering, taking advantage of the natural variation in the cell population to study non-cell-autonomous interactions.

Determinants of drug sensitivity.
The pre-existence of drug-resistant subclones was recently shown to be a major mechanism of drug resistance in cancer cell line populations[116,117]. Drug treatment of barcoded cell populations reproducibly induced the enrichment of the same specific barcodes, indicating selection of pre-existing resistant subclones[68,116]. Molecular, biochemical and functional analyses of sensitive and resistant subclones within a heterogeneous population could therefore be used to study drug sensitivity in cancer models.

Studying genomic instability
As ongoing genomic instability keeps shaping the genomic landscape of cancer models, these models can be used to study the cellular mechanisms underlying genomic instability. Indeed, cancer models have been used to study such mechanisms for many years (reviewed previously[118–120]). However, the recent advances in the field suggest new ways to study genomic instability in cancer models (FIG. 3d).

Single-cell clones of genomically unstable cell lines become genetically and transcriptionally heterogeneous. Studying the genomic evolution of single-cell clones from their initiation can therefore be a promising way to study how diversity is generated. Another idea would be to perform DNA barcoding experiments with cell lines that are deficient in distinct mechanisms of genome integrity maintenance in order to identify associations between mechanisms of genomic instability and drug resistance.

Concluding remarks

Genomic evolution is inevitable in living model systems. Like other sources of research irreproducibility, model diversification should be assessed routinely, reported appropriately and controlled for experimentally. It should not be viewed as a disaster, nor should it be ignored. The continuous improvement of cancer models over the past few years has yielded substantially better modelling of the human disease. Moving from a static to a more dynamic way of thinking about the genomics of cancer models could be another important step in this direction. It would help to minimize the detrimental effects of genomic evolution on cancer research and cancer precision medicine and at the same time open novel avenues of research that take advantage of these dynamics. Genomic evolution of cancer models can therefore contribute to the evolution of cancer research.

Uri Ben-David[1]*, *Rameen Beroukhim*[1,2,3,4]* *and Todd R. Golub*[1,2,3,5]*

[1]*Broad Institute of Harvard and MIT, Cambridge, MA, USA.*

[2]*Dana-Farber Cancer Institute, Boston, MA, USA.*

[3]*Harvard Medical School, Boston, MA, USA.*

[4]*Brigham and Women's Hospital, Boston, MA, USA.*

[5]*Howard Hughes Medical Institute, Chevy Chase, MD, USA.*

**e-mail: bendavid@broadinstitute.org; rameen_beroukhim@dfci.harvard.edu; golub@broadinstitute.org*

https://doi.org/10.1038/s41568-018-0095-3
Published online 21 December 2018

1. Greaves, M. Evolutionary determinants of cancer. *Cancer Discov.* **5**, 806–820 (2015).
2. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* **168**, 613–628 (2017).
3. Drost, J. & Clevers, H. Organoids in cancer research. *Nat. Rev. Cancer* **18**, 407–418 (2018).
4. Byrne, A. T. et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat. Rev. Cancer* **17**, 254–268 (2017).
5. Gillet, J. P., Varma, S. & Gottesman, M. M. The clinical relevance of cancer cell lines. *J. Natl Cancer Inst.* **105**, 452–458 (2013).
6. Jeggo, P. A., Pearl, L. H. & Carr, A. M. DNA repair, genome stability and cancer: a historical perspective. *Nat. Rev. Cancer* **16**, 35–42 (2016).
7. Kersten, K., de Visser, K. E., van Miltenburg, M. H. & Jonkers, J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol. Med.* **9**, 137–153 (2017).
8. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
9. Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
10. Basu, A. et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
11. Yang, W. et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
12. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 (2017).
13. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
14. Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).
15. Bertotti, A. et al. A molecularly annotated platform of patient-derived xenografts ("xenopatients") identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov.* **1**, 508–523 (2011).
16. Laurent, C. et al. Patient-derived xenografts recapitulate molecular features of human uveal melanomas. *Mol. Oncol.* **7**, 625–636 (2013).
17. Lin, D. et al. High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res.* **74**, 1272–1283 (2014).
18. Townsend, E. C. et al. The public repository of xenografts enables discovery and randomized phase II-like trials in mice. *Cancer Cell* **29**, 574–586 (2016).
19. Garman, B. et al. Genetic and genomic characterization of 462 melanoma patient-derived xenografts, tumor biopsies, and cell lines. *Cell Rep.* **21**, 1936–1952 (2017).
20. Krepler, C. et al. A comprehensive patient-derived xenograft collection representing the heterogeneity of melanoma. *Cell Rep.* **21**, 1953–1967 (2017).
21. Pauli, C. et al. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer Discov.* **7**, 462–477 (2017).
22. Stewart, E. et al. Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature* **549**, 96–100 (2017).
23. Beshiri, M. L. et al. A PDX/organoid biobank of advanced prostate cancers captures genomic and phenotypic heterogeneity for disease modeling and therapeutic screening. *Clin. Cancer Res.* **24**, 4332–4345 (2018).
24. Drapkin, B. J. et al. Genomic and functional fidelity of small cell lung cancer patient-derived xenografts. *Cancer Discov.* **8**, 600–615 (2018).
25. Gao, H. et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015).
26. Bruna, A. et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. *Cell* **167**, 260–274 (2016).
27. Gao, D. et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187 (2014).
28. van de Wetering, M. et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).
29. Lee, S. H. et al. Tumor evolution and drug response in patient-derived organoid models of bladder cancer. *Cell* **173**, 515–528 (2018).
30. Sachs, N. et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell* **172**, 373–386 (2018).
31. Vlachogiannis, G. et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018).
32. Izumchenko, E. et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann. Oncol.* **28**, 2595–2605 (2017).
33. Hidalgo, M. et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014).
34. Marangoni, E. et al. A new model of patient tumor-derived breast cancer xenografts for preclinical assays. *Clin. Cancer Res.* **13**, 3989–3998 (2007).
35. Zhang, X. et al. A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer Res.* **73**, 4885–4897 (2013).
36. Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **10**, 712 (2011).
37. Begley, C. G. & Ellis, L. M. Drug development: raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
38. Haibe-Kains, B. et al. Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
39. The Cancer Cell Line Encyclopedia Consortium & The Genomics of Drug Sensitivity in Cancer Consortium.

Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **528**, 84–87 (2015).
40. Haverty, P. M. et al. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
41. Franca, T. F. & Monserrat, J. M. Reproducibility crisis in science or unrealistic expectations? *EMBO Rep.* **19**, e46008 (2018).
42. Hunter, P. The reproducibility "crisis": reaction to replication crisis should not stifle innovation. *EMBO Rep.* **18**, 1493–1496 (2017).
43. Baker, M. Reproducibility crisis: blame it on the antibodies. *Nature* **521**, 274–276 (2015).
44. Loken, E. & Gelman, A. Measurement error and the replication crisis. *Science* **355**, 584–585 (2017).
45. Horbach, S. & Halffman, W. The ghosts of HeLa: how cell line misidentification contaminates the scientific literature. *PLOS ONE* **12**, e0186281 (2017).
46. Walrath, J. C., Hawes, J. J., Van Dyke, T. & Reilly, K. M. Genetically engineered mouse models in cancer research. *Adv. Cancer Res.* **106**, 113–164 (2010).
47. Tschaharganeh, D. F., Lowe, S. W., Garippa, R. J. & Livshits, G. Using CRISPR/Cas to study gene function and model disease in vivo. *FEBS J.* **283**, 3194–3203 (2016).
48. Andrechek, E. R. et al. Genetic heterogeneity of Myc-induced mammary tumors reflecting diverse phenotypes including metastatic potential. *Proc. Natl Acad. Sci. USA* **106**, 16387–16392 (2009).
49. Van Keymeulen, A. et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
50. Westcott, P. M. et al. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2015).
51. Nassar, D., Latil, M., Boeckx, B., Lambrechts, D. & Blanpain, C. Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma. *Nat. Med.* **21**, 946–954 (2015).
52. Ben-David, U. et al. The landscape of chromosomal aberrations in breast cancer mouse models reveals driver-specific routes to tumorigenesis. *Nat. Commun.* **7**, 12160 (2016).
53. Herschkowitz, J. I. et al. Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. *Proc. Natl Acad. Sci. USA* **109**, 2778–2783 (2012).
54. Uchimura, A. et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res.* **25**, 1125–1134 (2015).
55. Simon, M. M. et al. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol.* **14**, R82 (2013).
56. Mulligan, M. K. et al. Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. *Genes Brain Behav.* **7**, 677–689 (2008).
57. Liron, T., Raphael, B., Hiram-Bab, S., Bab, I. A. & Gabet, Y. Bone loss in C57BL/6J-OlaHsd mice, a substrain of C57BL/6J carrying mutated alpha-synuclein and multimerin-1 genes. *J. Cell. Physiol.* **233**, 371–377 (2018).
58. Kalish, S. et al. C57BL/6N mice are more resistant to ehrlich ascites tumors than C57BL/6J mice: the role of macrophage nitric oxide. *Med. Sci. Monit. Bas. Res.* **21**, 235–240 (2015).
59. Bourdi, M., Davies, J. S. & Pohl, L. R. Mispairing C57BL/6 substrains of genetically engineered mice and wild-type controls can lead to confounding results as it did in studies of JNK2 in acetaminophen and concanavalin A liver injury. *Chem. Res. Toxicol.* **24**, 794–796 (2011).
60. Mahajan, V. S. et al. Striking immune phenotypes in gene-targeted mice are driven by a copy-number variant originating from a commercially available C57BL/6 strain. *Cell Rep.* **15**, 1901–1909 (2016).
61. Mattapallil, M. J. et al. The Rd8 mutation of the Crb1 gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes. *Invest. Ophthalmol. Vis. Sci.* **53**, 2921–2927 (2012).
62. Jones, C. et al. Comparative genomic hybridization reveals extensive variation among different MCF-7 cell stocks. *Cancer Genet. Cytogenet.* **117**, 153–158 (2000).
63. Coser, K. R. et al. Antiestrogen-resistant subclones of MCF-7 human breast cancer cells are derived from a common monoclonal drug-resistant progenitor. *Proc. Natl Acad. Sci. USA* **106**, 14536–14541 (2009).
64. Nugoli, M. et al. Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications. *BMC Cancer* **3**, 13 (2003).
65. Kleensang, A. et al. Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. *Sci. Rep.* **6**, 28994 (2016).
66. Frattini, A. et al. High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci. Rep.* **5**, 15377 (2015).
67. Sato, S., Rancourt, A., Sato, Y. & Satoh, M. S. Single-cell lineage tracking analysis reveals that an established cell line comprises putative cancer stem cells and their heterogeneous progeny. *Sci. Rep.* **6**, 23328 (2016).
68. Ben-David, U. et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330 (2018).
69. Liu, Y. et al. Genomic, proteomic and phenotypic heterogeneity in HeLa cells across laboratories: implications for reproducibility of research results. Preprint at *bioRxiv* https://doi.org/10.1101/307421 (2018).
70. Fasterius, E. & Al-Khalili Szigyarto, C. Analysis of public RNA-sequencing data reveals biological consequences of genetic heterogeneity in cell line populations. *Sci. Rep.* **8**, 11226 (2018).
71. Weissbein, U., Plotnik, O., Vershkov, D. & Benvenisty, N. Culture-induced recurrent epigenetic aberrations in human pluripotent stem cells. *PLOS Genet.* **13**, e1006979 (2017).
72. Mekhoubad, S. et al. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012).
73. Locke, W. J. & Clark, S. J. Epigenome remodelling in breast cancer: insights from an early in vitro model of carcinogenesis. *Breast Cancer Res.* **14**, 215 (2012).
74. Boehm, J. S. & Golub, T. R. An ecosystem of cancer cell line factories to support a cancer dependency map. *Nat. Rev. Genet.* **16**, 373–374 (2015).
75. Liu, X. et al. ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* **180**, 599–607 (2012).
76. Hong, A. L. et al. Integrated genetic and pharmacologic interrogation of rare cancers. *Nat. Commun.* **7**, 11987 (2016).
77. Peng, S. et al. Tumor grafts derived from patients with head and neck squamous carcinoma authentically maintain the molecular and histologic characteristics of human cancers. *J. Transl Med.* **11**, 198 (2013).
78. Gunther, H. S. et al. Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene* **27**, 2897–2909 (2008).
79. Schulte, A. et al. A distinct subset of glioma cell lines with stem cell-like properties reflects the transcriptional phenotype of glioblastomas and overexpresses CXCR4 as therapeutic target. *Glia* **59**, 590–602 (2011).
80. Cifola, I. et al. Renal cell carcinoma primary cultures maintain genomic and phenotypic profile of parental tumor tissues. *BMC Cancer* **11**, 244 (2011).
81. Ben-David, U. et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat. Genet.* **49**, 1567–1575 (2017).
82. Villacorta-Martin, C., Craig, A. J. & Villanueva, A. Divergent evolutionary trajectories in transplanted tumor models. *Nat. Genet.* **49**, 1565–1566 (2017).
83. Eirew, P. et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
84. Clappier, E. et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J. Exp. Med.* **208**, 653–661 (2011).
85. Klco, J. M. et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* **25**, 379–392 (2014).
86. de Boer, B. et al. Prospective isolation and characterization of genetically and functionally distinct AML subclones. *Cancer Cell* **34**, 674–689 (2018).
87. Grasse, S. et al. Epigenomic profiling of non-small cell lung cancer xenografts uncover LRP12 DNA methylation as predictive biomarker for carboplatin resistance. *Genome Med.* **10**, 55 (2018).
88. Tiriac, H. et al. Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discov.* **8**, 1112–1129 (2018).
89. Roerink, S. F. et al. Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556**, 457–462 (2018).
90. Li, X. et al. Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. *Nat. Commun.* **9**, 2983 (2018).
91. Gupta, P. B. et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
92. Knouse, K. A., Lopez, K. E., Bachofner, M. & Amon, A. Chromosome segregation fidelity in epithelia requires tissue architecture. *Cell* **175**, 200–211 (2018).
93. Rowald, K. et al. Negative selection and chromosome instability induced by Mad2 overexpression delay breast cancer but facilitate oncogene-independent outgrowth. *Cell Rep.* **15**, 2679–2691 (2016).
94. Hingorani, S. R. et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
95. Sato, M., Glasebrook, A. L. & Bryant, H. U. Raloxifene: a selective estrogen receptor modulator. *J. Bone Miner. Metab.* **12**, S9–S20 (1994).
96. Kim, H. R. et al. Co-clinical trials demonstrate predictive biomarkers for dovitinib, an FGFR inhibitor, in lung squamous cell carcinoma. *Ann. Oncol.* **28**, 1250–1259 (2017).
97. Owonikoko, T. K. et al. Patient-derived xenografts faithfully replicated clinical outcome in a phase II co-clinical trial of arsenic trioxide in relapsed small cell lung cancer. *J. Transl Med.* **14**, 111 (2016).
98. Klemm, F. & Joyce, J. A. Microenvironmental regulation of therapeutic response in cancer. *Trends Cell Biol.* **25**, 198–213 (2015).
99. Morgan, M. M. et al. Personalized in vitro cancer models to predict therapeutic response: challenges and a framework for improvement. *Pharmacol. Ther.* **165**, 79–92 (2016).
100. Neve, R. M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527 (2006).
101. Gendoo, D. M. A. et al. Whole genomes define concordance of matched primary, xenograft, and organoid models of pancreas cancer. Preprint at *bioRxiv* https://doi.org/10.1101/209692 (2017).
102. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
103. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
104. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
105. Munoz, D. M. et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
106. Mitra, A., Mishra, L. & Li, S. Technologies for deriving primary tumor cells for use in personalized cancer therapy. *Trends Biotechnol.* **31**, 347–354 (2013).
107. Andor, N. et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
108. Janiszewska, M. et al. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat. Genet.* **47**, 1212–1219 (2015).
109. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
110. Tabassum, D. P. & Polyak, K. Tumorigenesis: it takes a village. *Nat. Rev. Cancer* **15**, 473–483 (2015).
111. Marusyk, A. et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58 (2014).
112. Chapman, A. et al. Heterogeneous tumor subpopulations cooperate to drive invasion. *Cell Rep.* **8**, 688–695 (2014).
113. Mateo, F. et al. SPARC mediates metastatic cooperation between CSC and non-CSC prostate cancer cell subpopulations. *Mol. Cancer* **13**, 237 (2014).
114. Ouchi, R., Okabe, S., Migita, T., Nakano, I. & Seimiya, H. Senescence from glioma stem cell differentiation promotes tumor growth. *Biochem. Biophys. Res. Commun.* **470**, 275–281 (2016).
115. Vinci, M. et al. Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat. Med.* **24**, 1204–1215 (2018).
116. Bhang, H. E. et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).
117. Hata, A. N. et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat. Med.* **22**, 262–269 (2016).

118. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
119. Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
120. Sansregret, L., Vanhaesebroeck, B. & Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat. Rev. Clin. Oncol.* **15**, 139–150 (2018).
121. Wetterstrand, K. A. DNA sequencing costs: data: data from the NHGRI Genome Sequencing Program (GSP). *NIH* https://www.genome.gov/27541954/dna-sequencing-costs-data/ (2018).
122. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
123. Silva, G. O. et al. SynthEx: a synthetic-normal-based DNA sequencing tool for copy number alteration detection and tumor heterogeneity profiling. *Genome Biol.* **18**, 66 (2017).
124. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048 (2006).
125. Zhang, S., Yuan, Y. & Hao, D. A genomic instability score in discriminating nonequivalent outcomes of BRCA1/2 mutations and in predicting outcomes of ovarian cancer treated with platinum-based chemotherapy. *PLOS ONE* **9**, e113169 (2014).
126. Burrell, R. A. et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492–496 (2013).
127. Ben-David, U. & Benvenisty, N. in *StemBook* (eds The Stem Cell Research Community) https://doi.org/10.3824/stembook.1.150.1 (Harvard Stem Cell Institute, Cambridge, MA, 2012).
128. Ben-David, U., Mayshar, Y. & Benvenisty, N. Virtual karyotyping of pluripotent stem cells on the basis of their global gene expression profiles. *Nat. Protoc.* **8**, 989–997 (2013).
129. Santaguida, S. & Amon, A. Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat. Rev. Mol. Cell Biol.* **16**, 473–485 (2015).
130. Sheltzer, J. M. et al. Single-chromosome gains commonly function as tumor suppressors. *Cancer Cell* **31**, 240–255 (2017).
131. Ben-David, U. et al. Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. *Nat. Commun.* **5**, 4825 (2014).
132. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

Acknowledgements
The authors thank I. Fung for her assistance with designing and preparing the figures. This work was supported by the Human Frontiers Science Program (U.B.-D.), the Howard Hughes Medical Institute (T.R.G.), the US National Institutes of Health (R01 CA188228; R.B.), the Gray Matters Brain Cancer Foundation (R.B.), the Bridge Project (R.B.), a Broad Institute SPARC award (R.B.) and a Broad Institute BroadNext10 grant (U.B.-D.).

Author contributions
All authors researched data for the article, contributed to discussion of the content and wrote, reviewed and edited the manuscript.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Cancer thanks J. Gray, R. Kimple and other anonymous reviewer(s) for their contribution to the peer review of this work.

RELATED LINKS
US Institute for Laboratory Animal Research (ILAR): http://dels.nas.edu/global/ilar/Lab-Codes
Cell STRAINER: https://cellstrainer.broadinstitute.org