Preclinical mouse solid tumour models: status quo, challenges and perspectives

Nicolas Gengenbacher[1], Mahak Singhal[1,2]* and Hellmut G. Augustin[1–3]*

Abstract | Oncology research in humans is limited to analytical and observational studies for obvious ethical reasons, with therapy-focused clinical trials being the one exception to this rule. Preclinical mouse tumour models therefore serve as an indispensable intermediate experimental model system bridging more reductionist *in vitro* research with human studies. Based on a systematic survey of preclinical mouse tumour studies published in eight scientific journals in 2016, this Analysis provides an overview of how contemporary preclinical mouse tumour biology research is pursued. It thereby identifies some of the most important challenges in this field and discusses potential ways in which preclinical mouse tumour models could be improved for better relevance, reproducibility and translatability.

Immune checkpoint blockade
A therapeutic approach aimed at restoring or enhancing antitumour immune response by blocking signalling pathways that naturally limit immune reactions to prevent autoimmunity.

Autochthonous
Arising in its natural site; refers to *de novo* tumours that evolve out of normal cells within a living organism, in contrast to transplanted tumours, which are referred to as non-autochthonous.

[1]Division of Vascular Oncology and Metastasis, German Cancer Research Center (DKFZ–ZMBH Alliance), 69120 Heidelberg, Germany.
[2]Department of Vascular Biology and Tumor Angiogenesis (CBTM), Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany.
[3]German Cancer Consortium, 69120 Heidelberg, Germany.
**These authors contributed equally to this work.*

Correspondence to H.G.A.
augustin@angiogenese.de

doi:10.1038/nrc.2017.92
Published online 27 Oct 2017

Despite increasing ethical concerns with experiments in laboratory animals, preclinical mouse tumour models are without alternative in the armamentarium of oncology-related research tools. As with all models, they are a surrogate of the 'real' world. Different models serve different needs, ranging from high-throughput, non-cellular screens to complex *in vivo* investigations. Currently, only manipulatable *in vivo* models are suitable to mimic the complexity of human cancer as an evolutionary process of neoplastically transformed cells that are in intimate crosstalk with their local and systemic environment as well as with the plethora of different immune cells. The comparative discussion of *in vitro* versus *in vivo* models should therefore not be reduced to 'better' or 'worse' but should centre on the issue of 'which model for which question' (REF. 1).

There are numerous examples of breakthrough discoveries in preclinical mouse tumour models that have paved the way for clinical application in humans, such as the first demonstration of the efficacy of immune checkpoint blockade[2]. Generally speaking, it may be fair to conclude that fundamental discovery research employs the experimental model that is most suitable to answer the question. The issue of human translatability is secondary. This is different for more translation-related applied research, be it the disease-related experimental validation of a potential therapeutic target or the preclinical validation of a novel therapeutic compound. Of note, the enormous pace of human tumour genomic analysis has dramatically accelerated oncology research in recent years and thereby strongly increased the need for preclinical tumour models with high translatability.

There are a number of recent authoritative reports summarizing the repertoire of available mouse models for different tumour types, and the reader is referred to the published literature for more information[3–7]. We have retrieved preclinical oncology studies published in eight journals in 2016 (BOX 1) in order to quantitatively analyse contemporary mouse cancer biology research with a focus on solid tumour models. Following an overview of available cancer mouse model categories, this Analysis article critically assesses how these models are actually used, with the aim of identifying existing bottlenecks in bench-to-bedside translational oncology research. Our goal is to stimulate discussion on the ways in which the translational value of existing models can be advanced and on how best to develop new models.

Overview of mouse model categories

Cell line-derived models. Most of our current understanding of cancer and its hallmarks is based on the establishment of long-term *in vitro* cultured tumour cell lines and their *in vivo* inoculation in mice. To this day, these cell line models, which are not autochthonous, remain by far the most commonly used mouse models in basic and translational oncology (82% (506/618) of articles published in 2016; FIG. 1a). Low-cost, synchronous tumour growth and easy technical manipulability enable the application of these models for rapid identification and validation of cancer-relevant genes as well as preclinical evaluation of drug candidates. Allotransplantation of mouse cells into syngeneic immunocompetent inbred mice (allografts) and xenotransplantation of human cells into immunocompromised mice (xenografts) represent the two broadest categories of cell line models. In both types of model, tumour cells can be injected ectopically (mostly subcutaneously), orthotopically to mimic tumour growth

Box 1 | **Methodology of the systematic literature analysis**

Study selection. To survey contemporary preclinical mouse tumour biology research, all oncology research articles published between January and August 2016 in the interdisciplinary journals *Cell*, *Nature*, *Science* and the *Proceedings of the National Academy of Sciences USA* as well as in the cancer-specific journals *Cancer Cell*, *Cancer Discovery*, *Cancer Research* and *Oncogene* were retrieved and systematically reviewed for mouse tumour experiments. For interdisciplinary journals, the title and abstract of all original articles published during the respective time frame were manually screened, and all oncology-related articles were retrieved for the analysis. For cancer-specific journals, all original articles from this time frame were included. A detailed list of the 949 studies included in this analysis is provided as Supplementary information S1 (table). The goal was to analyse a sizeable recent cohort of studies that representatively reflects contemporary preclinical mouse tumour biology research. The inclusion of almost 1,000 publications has given the analysis sufficient depth for a detailed subgroup analysis of different experimental approaches and different tumour types. The focus on eight selected journals potentially implied a bias that could have challenged the representativeness of the analysis. The selection of journals was primarily based on the journal's impact factor, which is a widely employed, albeit disputed, measure of quality. Yet, we believe that even if biased, this list should represent the higher end of more advanced model application.

Methodology. All included studies were evaluated at the study level, but only those containing data derived from mouse tumour models were used for experiment-level data extraction. For this subset, the following 11 parameters were collected: aim of the study, mouse model category, cancer type, genetic background, immune status, inclusion of primary tumour data, engraftment site of non-autochthonous models, inclusion of metastasis data, metastasis assay type, experimental readout and therapeutic approach. All parameters were recorded in an Excel file using multi-selection drop-down lists. Detailed information of all list values is provided in Supplementary information S2 (table). Subsets of this data set (for example, all breast cancer studies) were analysed with Excel using the list filter function 'contains'.

in its organ of origin or systemically (mostly intraperitoneally, intravenously or intracardially) to study metastatic spread.

Starting in the mid-1950s, cell line allografts with an emphasis on leukaemia were employed primarily in preclinical drug development to screen for potential cancer therapeutics[8]. Although a small number of cytotoxic drugs were successfully identified, this approach suffered from an overall low predictive value for efficacy in human trials[9,10]. Hence, since the 1990s, xenografts largely established from human cell lines that are part of well-characterized panels, that is, the NCI60 panel (REF. 11) and more recently the Genomics of Drug Sensitivity in Cancer (GDSC) and the Cancer Cell Line Encyclopedia (CCLE) panels[12–14], have replaced allografts as the predominant tool for preclinical drug evaluation. Although cell line xenografts can often predict efficacy of cytotoxic agents, these models unfortunately fail to faithfully predict clinical activity for most targeted therapies[7,15].

Criticized for being overly reductionist, cell line-derived models suffer from two major shortcomings. First, owing to their rapid non-autochthonous growth, cell line-derived tumours, compared with their human counterparts, display a dramatically perturbed tissue architecture with a severely altered microenvironment, including changes in the vascular, lymphatic and immune compartments[16,17]. Second, loss of genetic heterogeneity and irreversible changes in gene expression are imposed by long-term *in vitro* propagation[18–20].

Patient-derived xenograft models. Circumventing *in vitro* culture and its potential artefacts, patient-derived xenografts (PDXs) are typically generated by subcutaneous implantation of fresh, surgically derived human tumour material into immunodeficient mice. PDXs have been shown to stably retain molecular, genetic and histopathological features of their originating tumours, at least over limited passages of *in vivo* expansion[21], and are therefore currently the only model system that can directly incorporate the vast inter-patient and intra-tumour heterogeneity that is inherent to human cancer. Detailed methodological aspects of PDX generation would go beyond the scope of this Analysis and are outlined in REFS 22–25.

Several retrospective comparisons of drug responses in patients and their corresponding xenografts, an approach that was pioneered in the mid-1980s (REF. 26), have provided strong evidence that PDXs can faithfully predict therapy response (REFS 21,25,27). Consequently, as a promising tool for personalized medicine, so-called co-clinical trials aim to establish PDXs from patients who are enrolled in clinical trials in avatar mice that can be used in parallel to explore multiple therapeutic options in an individualized manner[28]. The fact that PDXs are derived from human tumours allows direct evaluation of clinically approved drugs[7,29–31].

As a population-based instead of a personalized approach, several large-scale PDX repositories have been implemented, including the EurOPDX consortium, the US National Cancer Institute (NCI) repository of patient-derived models, the Public Repository of Xenografts (PRoXe), the Children's Oncology Group (COG) cell culture and xenograft repository, the Pediatric Preclinical Testing Consortium (PPTC) and the Novartis Institutes for Biomedical Research PDX Encyclopedia (NIBR PDXE)[3]. Containing high numbers of PDXs, which are comprehensively characterized on a molecular and genetic level, these resources promise to serve as powerful screening platforms to profile candidate therapeutics, identify biomarkers of drug sensitivity and elucidate mechanisms of therapy resistance[3,32]. For example, recent analyses of large colorectal cancer PDX cohorts revealed several biomarkers of reduced sensitivity to epidermal growth factor receptor (EGFR) inhibitors, some of which might be druggable by rational combination therapies[33–35]. Likewise, the validity of PDX studies for compound screening purposes was shown by employing a 'one animal per model per treatment' trial design in approximately 1,000 highly heterogeneous PDXs across six cancer types, assessing a total of 62 different drug candidates[32]. Despite such promises, PDX studies remain, at least on an academic level, severely limited by high cost and logistical difficulties, which is also reflected by the fact that only 7% (43/618) of the reviewed articles published in 2016 contain PDX data (FIG. 1b).

As with all preclinical models, PDXs are handicapped by several built-in limitations. First, engraftment rates not only depend on the recipient mouse strain and original patient sample quality but may also strongly vary between different tumour types and grades[36,37], and differences in engraftment can limit the genetic complexity represented by PDXs[38,39]. The engraftment rate itself may even serve as

Allografts
Tumours transplanted from one individual to another of the same species.

Xenografts
Cells or tissues that are transplanted between two different species, such as human and mouse.

Orthotopically
The engraftment of tumours into the 'natural' anatomical site or organ in which they usually arise.

Co-clinical trials
Trials in which an ongoing human clinical trial is mirrored by simultaneous studies in mice.

Avatar mice
A mouse into which a patient's tumour tissue is grafted to generate a 'personalized' model that is then used to identify an optimal therapeutic strategy.

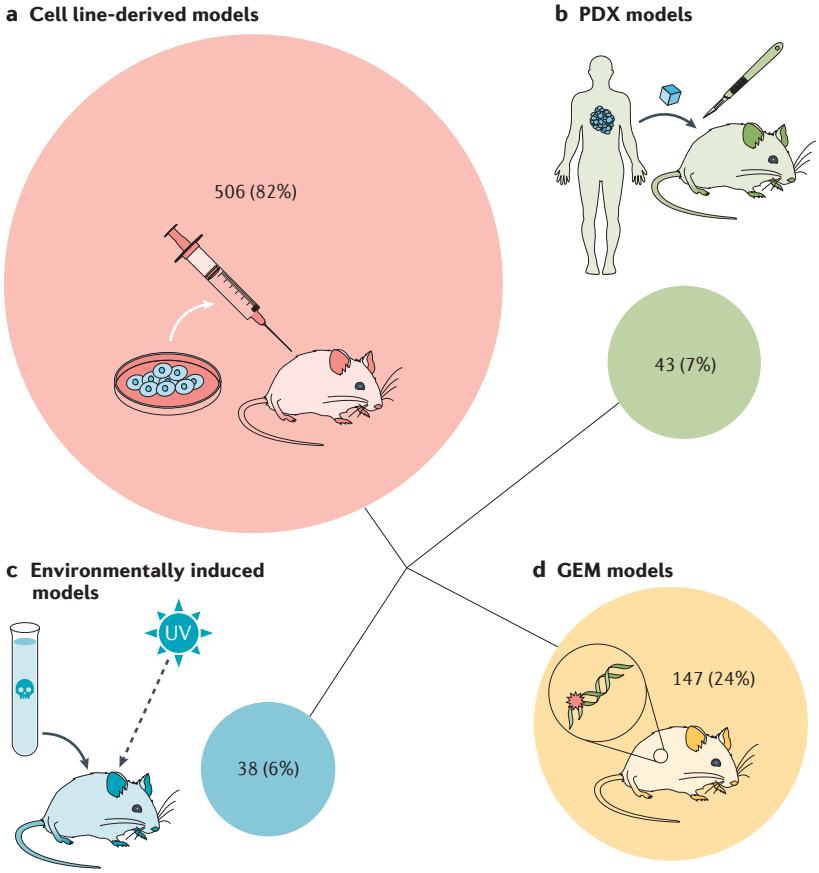

Figure 1 | **Schematic representation of the frequency distribution of different categories of tumour models in preclinical studies published in 2016. a** | The majority of all analysed mouse tumour articles (82%; $n_{\text{mouse model}} = 618$) employed cell line-derived models. **b** | Patient-derived xenograft (PDX) models were used in 7% of studies. **c** | Environmentally induced tumour models were used in 6% of studies. **d** | Genetically engineered mouse (GEM) models were employed in 24% of studies. The percentages add up to more than 100% because 109 studies employed more than one category of tumour model (a Venn diagram indicating intersections of all categories is included as Supplementary information S3 (figure)). The frequency distribution of different model use in each journal analysed is included as Supplementary information S4 (table). UV, ultraviolet.

'One animal per model per treatment' trial design
Also known as '1 × 1 × 1' trial design; a type of trial design in which individual mice (instead of groups of mice) from large patient-derived xenograft collections are used to evaluate drug response in a heterogeneous study population.

Carcinogen bioassay
A standardized measurement of an animal response to an environmental exposure in order to estimate its cancer-causing potential.

a predictive biomarker, as patients from whom PDXs can be established show the worst prognosis, as demonstrated for various indications including breast cancer[40], pancreatic ductal adenocarcinoma (PDAC)[41] and renal cell carcinoma[42]. Another limitation is that although human stroma is initially present following engraftment, it is ultimately replaced by mouse stroma components following *in vivo* passaging. This not only modifies paracrine tumour microenvironmental interactions owing to cross-species signalling incompatibilities but also restricts the utility of PDXs for examining human-specific microenvironment-targeted therapy[21,43]. Finally, the immunocompromised background required for successful engraftment of PDXs precludes their use to study immune cell function and analyse immunotherapeutic strategies[4].

Environmentally induced models. Historically, the origin of modelling cancer *in vivo* dates back to the mid-1910s with the demonstration that topical application of coal tar induces skin tumour formation in rabbits[44] and mice[45]. Subsequently, a wide range of suspected environmental cancer-causing agents, including chemicals, radiation and pathogens, were evaluated in animals, in an approach that was later standardized as the carcinogen bioassay (reviewed in REFS 46,47). Indeed, roughly 25% of all carcinogens were first identified in rodents before being epidemiologically linked to human cancer[47]. A plethora of environmentally induced mouse models for multiple tumour types have since been established, including skin[48,49], bladder[50], lung[51], liver[52] and colon cancer[53]. Robust tumour induction protocols for these cancers are based on environmental stimuli that are highly relevant to human tumorigenesis. Many of these models closely recapitulate the phenotypic and genetic heterogeneity of their human counterparts. For example, two recent reports confirmed substantial molecular similarities by comparing the mutational landscapes of carcinogen-induced skin squamous cell carcinomas[54] and chemically induced *Kras*-driven lung cancers[55] in mice with corresponding patient material. Furthermore, such environmental models exhibit *de novo* tumour growth and represent all stages of multistep carcinogenesis. They may therefore be particularly beneficial in defining genetic risk factors and assessing prevention strategies[5,56]. For example, via the exploitation of differences in cancer susceptibility, multiple modifier alleles of lung carcinogenesis were revealed by employing a urethane-based tumour induction protocol in 21 different inbred mouse strains[57].

Despite their biological relevance, environmental models are rarely used in current cancer research (6% (38/618); FIG. 1c). In particular, the inherently long latency and high variability of penetrance pose major difficulties for proper study design (for example, choosing adequate animal numbers and identifying relevant time points). A recent study[58] highlights the technical challenges involved in modelling especially subtle environmental exposures, such as dietary changes. By long-term use of a combined choline-deficient and high-fat diet, two intrahepatic immune subsets (natural killer T cells and cytotoxic T cells) were identified as major drivers of non-alcoholic fatty liver disease and hepatocellular carcinoma. Yet, this elegant study employed an experimental approach that was based on a tumour incidence of only 25% after one year of challenge.

Genetically engineered mouse models. Over the past three decades, the increasing understanding of the genetic aberrations underlying tumorigenesis has fuelled the generation of diverse genetically engineered cancer models that reproduce the genetic events in an autochthonous *in vivo* setting, allowing *de novo* tumour formation in a native immune-proficient microenvironment. These advances have made genetically engineered mouse (GEM) models the second most common type of mouse model in oncology research in our sample of studies published in 2016 (24% (147/618); FIG. 1d).

The first GEM tumour models in the mid-1980s were 'simple' transgenic mice, harbouring randomly integrated oncogenes under the control of a tissue-specific minimal promoter (so-called oncomice[59]). Thereafter,

locus-specific gene-targeting methodology[60] enabled the generation of global tumour suppressor gene (TSG) knockout animals[61]. At that time, the observation of *de novo* tumorigenesis *in vivo* provided proof of the causal role of mutated oncogenes and TSGs in cancer. A wide range of conditional models based on site-specific recombinase (SSR) systems, such as Cre–*loxP* and Flp–*FRT*, has been developed since, allowing a spatially and temporally controlled introduction of human-cancer-relevant mutations into mice (extensively reviewed in REFS 6,17,62). Inducibility is achieved via exogenous SSR delivery (for example, in the form of adenoviral Cre[63]) to an accessible somatic tissue, tissue-specific expression of an SSR fused with a hormone-responsive nuclear receptor domain (for example, the 4-hydroxytamoxifen-binding domain of the oestrogen receptor[64]) or use of the doxycycline-inducible *Tet* system[65].

GEM models not only faithfully recapitulate molecular and histopathological features of human disease but also have a strong predictive power for drug response and resistance[6,66]. For example, a study assessing standard-of-care chemotherapy as well as multiple targeted therapeutic approaches in *Kras*-driven GEM models of non-small-cell lung carcinoma and PDAC could retrospectively reproduce human clinical therapeutic responses[67]. This also prompted the use of GEM models in co-clinical trial studies, in which parallel treatments of multiple genetically defined mouse cohorts can be used for patient stratification by identifying genetic biomarkers of drug resistance and responsiveness as well as for evaluation of combination therapies[28]. Displaying *de novo* tumorigenesis in a fully immunocompetent host environment, GEM models can be additionally employed in prevention studies[5] and for testing of immunotherapies[4].

Despite numerous beneficial attributes, GEM models also share three major disadvantages. First, the often synchronous overexpression or inactivation of potent oncogenes and TSGs, respectively, bypasses major bottlenecks to malignant transformation, resulting in a reduced clonal heterogeneity compared with human tumours[68,69]. Second, preclinical evaluation of metastasis is challenging, as most GEM models must be sacrificed before developing metastatic disease owing to a heavy multifocal primary tumour load[6]. Third, the generation of GEM models is very expensive and time consuming.

Status quo and challenges

Mouse tumour models constitute the most widely used preclinical research tool in oncology. In fact, of the reviewed cohort of 949 oncology studies published in 2016, 65% (618/949) employed mouse tumour models. Of these, 89% (548/618) focused on a single cancer type, whereas only a smaller sub-cohort examined multiple tumour types (FIG. 2a). Around 18% (109/618) of studies employed more than one model category to validate their findings (FIG. 2a). In order to identify major challenges of current mouse tumour model technology, we analysed a cohort of publications from 2016 more systematically to record the key modelling parameters for each study (BOX 1). In this section, we present a summary of the key findings of this analysis, on the basis of which we discuss some of the challenges for future advancement of mouse tumour models. We focus on the modelling of clinically relevant cancer types, better mimicking the natural course of tumour progression and metastasis, improving the assessment of therapeutic efficacy and improving data representation and interpretation.

Modelling different cancer types. There is considerable variation in the frequency of mouse studies related to different cancer types. Melanoma, leukaemia, brain, breast, colorectal, hepatocellular, lung, pancreatic and prostate tumours are most frequently represented in preclinical tumour studies published in 2016. By contrast, bladder, cervical, head and neck, kidney and thyroid cancer are strongly under-represented (FIG. 2b). This finding might be indicative of an imbalance in funding and cancer-type-specific research interest, mirroring in part the relative incidence, mortality and number of ongoing clinical trials for a given indication[70]. Consequently, these factors translate into a limited availability of mouse models for understudied tumour types[71–74].

Cell line transplantation models are over-represented in studies of all tumour types published in 2016. Yet, a subclass analysis revealed some remarkable differences. Melanoma, breast and brain tumour research is primarily based on grafted cell line models, whereas GEM models account for 60% (28/47) of pancreatic cancer studies (FIG. 2c). This can likely be attributed to the early discovery of a pancreatic tissue-specific promoter in 1996 (REF. 75) and the establishment of the first PDAC GEM model by tissue-specific expression of both $Kras^{G12D}$ and $Trp53^{R172H}$ (KPC model) in 2005 (REF. 76). The continuously low use of GEM models in some tumour types is likely due to the limited availability of specific promoters for the cell of origin as well as an insufficient understanding of the expression kinetics of available tissue-specific promoters, which together can make the development of GEM models for specific cancer types challenging[17].

Environmentally induced tumour models are frequently used for hepatocellular and colorectal cancer research, contributing to around 28% (9/32) and 17% (12/69) of studies, respectively (FIG. 2c). These models closely mimic clinical features of precancerous tissues, such as chronic hepatitis[58] or colitis[77], that are initiated by exposure to environmental factors. Given the power of such models, it is somewhat surprising that they are not used more frequently for cancer types for which an environmental aetiology is well defined, for example, hormones for breast cancer[78,79], inflammatory agents for pancreatic cancer[80], ultraviolet radiation for melanoma[81] and chemical carcinogens for lung cancer[82]. Additionally, certain cancer types, for example, brain tumours[83], lack a substantial environmental aetiology, and this fact is reflected by the absence of corresponding mouse models (FIG. 2c).

PDX models are most commonly used in lung and pancreatic cancer studies (9% (8/88) and 13% (6/47), respectively; FIG. 2c). Despite substantial technical advances, low engraftment rates for certain tumour types remain the major limitation for the advancement of PDX

Site-specific recombinase (SSR) systems
Enzymatic systems, such as Cre–*loxP* or Flp–*FRT*, that rearrange genomic target segments that have been marked by specific DNA recognition sites.

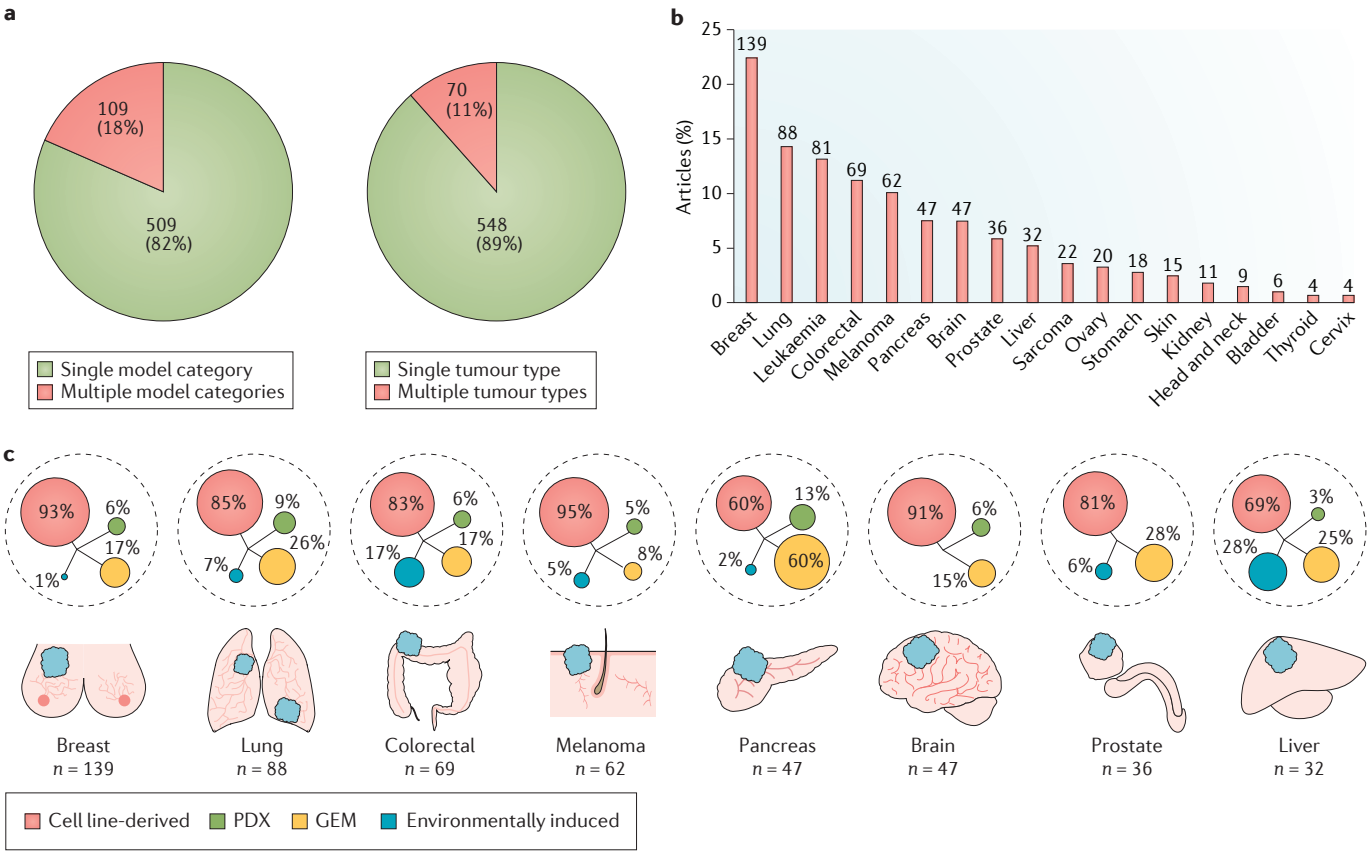

Figure 2 | **Modelling of different tumour types in mouse model studies. a** | Absolute numbers and percentages of preclinical mouse tumour articles published in 2016 utilizing one versus multiple model categories as well as studies focusing on a single versus multiple cancer types ($n_{\text{mouse model}} = 618$). **b** | Percentages of different tumour types in all analysed articles. Absolute numbers are indicated above each bar. **c** | Distribution of model categories within the eight most frequent solid cancer types. The percentages add up to more than 100% because some studies employed more than one category of tumour model. Absolute numbers are indicated in TABLE 1. GEM, genetically engineered mouse; PDX, patient-derived xenograft.

models; this limitation is reflected by the absence of prostate PDXs in the reviewed studies published in 2016 (REF. 37). Therefore, it would be desirable for academia and industry to collaborate to establish additional PDXs for under-represented cancer types.

Mimicking tumour progression and metastasis. Precise modelling of tumour progression includes replicating the natural course of events ranging from tumour initiation to metastatic outgrowth[17]. The *de novo* growth of tumours in GEM and environmentally induced models closely recapitulates the early stages of tumour progression, but the multifocal nature of the resulting primary tumours limits the lifespan of a mouse, thereby restricting metastatic progression[84–86]. In turn, non-autochthonous models fail to mimic the genetic events leading to tumour initiation. Instead, such models offer a focal primary tumour, which can be resected to allow metastatic growth to become the experimental end point. In such models, the primary tumour is initiated by transplantation of cancer cells or patient-derived material either at an orthotopic or ectopic site in 42% (209/493) and 67% (331/493) of studies (the percentages add up to >100% because 47/493 studies employed orthotopic and ectopic models), respectively (FIG. 3a). The site of engraftment strongly affects the metastatic potential of tumour cells. Orthotopic transplantation of cell line-based models has been extensively described to be superior with respect to modelling tumour microenvironmental interactions, therapeutic response and metastatic patterns[10,87,88]. Similarly, orthotopic PDXs have a higher metastatic incidence than subcutaneous implants[89–91]. Still, in the reviewed studies published in 2016, a majority of non-autochthonous models, especially those modelling visceral organ-derived cancers, relied on ectopic engraftment, whereas orthotopic models were preferably used in brain and breast cancer studies (TABLE 1). This imbalanced distribution is likely due to ease of access to the target anatomical site. Nevertheless, orthotopic tumour models for various other cancer types have been developed through efforts to improve surgical procedures[92–95].

Although metastasis accounts for approximately 90% of cancer-related mortality[96], only 25% (152/618) of studies focused on metastasis (FIG. 3b). Experimental metastasis models, wherein cancer cells are injected directly into the systemic circulation, recapitulate only

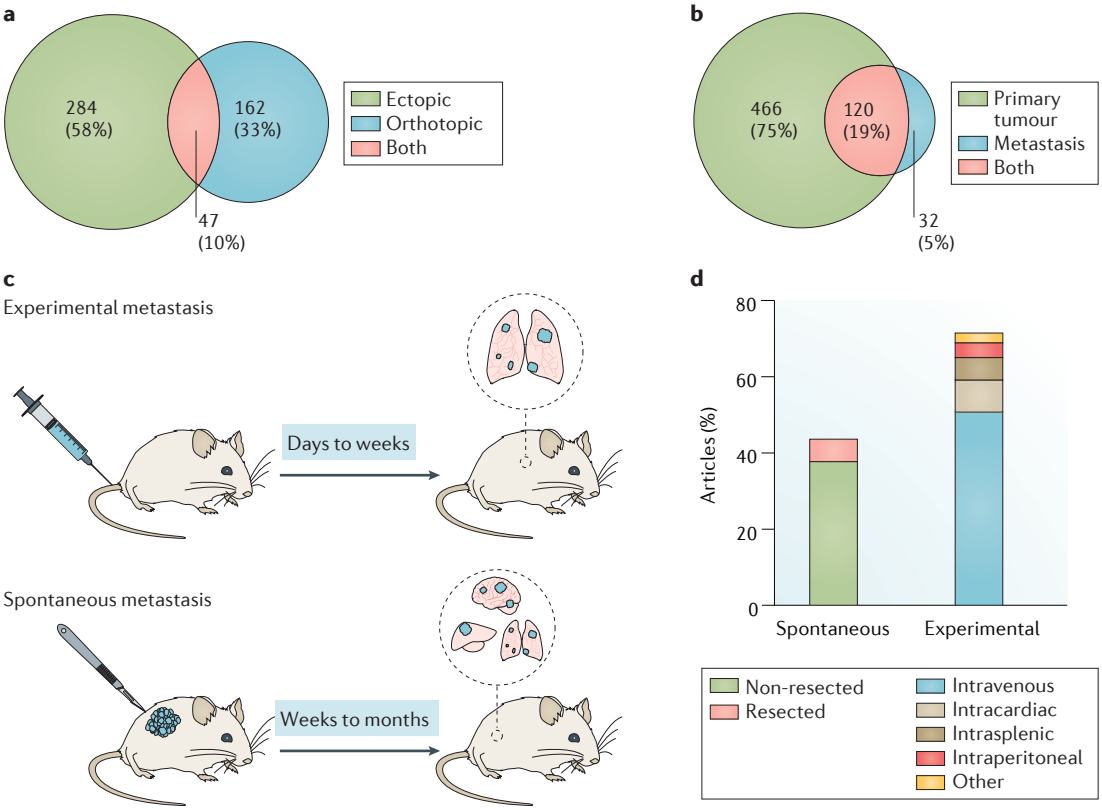

Figure 3 | **Modelling tumour progression and metastatic disease in mouse tumour model studies.** **a** | Non-autochthonous cell line-derived and patient-derived tumour models require either orthotopic engraftment into the tissue of origin or ectopic engraftment at a foreign site. Their distribution is shown as absolute numbers and percentages out of all primary tumour studies utilizing non-autochthonous models ($n_{\text{non-autochthonous model}} = 493$). **b** | Distribution of articles studying primary and/or metastatic disease ($n_{\text{mouse model}} = 618$). **c** | In experimental metastasis models, tumour cells are systemically injected via diverse routes (intravenous (often through the tail vein), intracardiac, intrasplenic or intraperitoneal). These models are characterized by rapid and uniform formation of metastases but only capture late steps of the metastatic cascade (survival in the circulation and metastatic colonization). Spontaneous metastasis models truthfully recapitulate the natural course of metastatic disease, including initial steps such as tumour cell dissemination from the primary tumour. However, long latency may require the surgical resection of the primary tumour to allow sufficient time for metastatic progression. **d** | Frequency distribution of different types of experimental and spontaneous metastasis models within the sub-cohort of all metastasis studies ($n_{\text{metastasis}} = 152$). The percentages add up to more than 100% because 23 studies employed multiple metastasis models. Sixty-seven studies utilized spontaneous metastasis models (58 non-resected, 9 resected), and 100 studies used experimental metastasis models (77 intravenous, 13 intracardiac, 9 intrasplenic, 6 intraperitoneal and 2 others).

metastatic colonization and circumvent the primary disease (FIG. 3c). Such models have been employed in 66% (100/152) of all metastasis-related studies (FIG. 3d). These models primarily metastasize to a single organ, strongly influenced by the site of injection. They thereby fail to recapitulate typical clinical metastatic spread[97]. Injection of cancer cells into the tail vein, heart or spleen (77% (77/100), 13% (13/100) and 9% (9/100) of experimental metastasis studies) provokes subsequent metastasis to the lung, brain or liver, respectively. Importantly, these models lack a latency phase, which is a clinically important feature[98]. Cancer recurrence is frequently observed in patients with breast[99] or prostate[100] cancer even after very long periods of remission. A better mechanistic understanding of these clinical observations could be obtained from mouse models mimicking tumour dormancy. Spontaneous models (employed by 44% (67/152) of all metastasis studies; FIG. 3d) recapitulate the entire metastatic cascade, wherein tumour cells disseminate from the primary tumour to multiple distant sites and grow into overt metastases, which can require a latency phase of weeks to months (FIG. 3c). Therefore, to prevent mice from succumbing to the primary disease, tumour resection is often required to make metastases rate-limiting[86,101]. Intriguingly, only 13% (9/67) of spontaneous metastasis studies included surgical removal of the primary tumour (FIG. 3d). In conclusion, the clinical relevance of metastatic models could be improved by combining orthotopic tumour transplantation and surgical removal of the primary tumour.

Tumour dormancy
Undetectable and asymptomatic tumour cells that remain in patients who have been clinically disease-free for a long period of time. Tumours can clinically recur from this population of cells.

Assessing clinical therapeutic efficacy. From the US Food and Drug Administration (FDA) approval of nitrogen mustard in 1949 to rituximab in 1997, there has been

a major paradigm shift in cancer therapies from DNA-damaging cytotoxic chemotherapy towards targeted approaches based on molecular mechanisms. Reflecting this change, 27 novel targeted agents, including 19 kinase inhibitors and 8 monoclonal antibodies, have received FDA approval in the 2004–2014 period, whereas only 10 new cytotoxic drugs were approved in the same time period[102]. Cell line-based models, still the predominant tools for preclinical drug evaluation, serve as a useful platform for the screening of cytotoxic compounds but often fail to predict the clinical outcome of targeted therapies. This is also reflected by the low FDA approval rate of only 5% of oncology therapies that moved through clinical trials between 2006 and 2015 (REFS 103,104). GEM models are thought to be better predictors of therapeutic response rates. For example, treatment with the multitargeted tyrosine kinase inhibitor sunitinib in a GEM model of pancreatic neuroendocrine tumours (RIP-Tag2 mice) demonstrated increased survival[105], which was subsequently confirmed in a clinical trial that led to FDA approval of sunitinib in 2011 for patients with these tumours[106]. However, a quantitative correlation of GEM-model-based preclinical studies and the success rate of corresponding clinical trials has yet to be evaluated. PDXs have also been successfully deployed to predict the clinical efficacy of therapeutic interventions and have led to the establishment of avatar mouse studies to analyse personalized therapeutic efficacy in real time[107]. Regardless of the great promise that both GEM and PDX models offer for evaluating therapeutic efficacy, only 16% (42/261) and 12% (31/261) of the analysed therapeutic studies published in 2016 employed these models, respectively. Instead, cell line-based models continue to remain, by far, the most frequently used experimental approach in this context (89% (231/261); FIG. 4a).

The angiogenesis inhibitor bevacizumab was, in 2004, the first clinically approved cancer therapy targeting an element of the microenvironment[108]. This has strongly stimulated research into the mechanisms of tumour microenvironmental interactions and the validation of microenvironment-targeted therapies. Infiltrating immune cells are a key component of the tumour microenvironment. Proof-of-principle studies indicating that the immune milieu can be modified from a pro-tumour to antitumour response have led to the FDA approval of immunomodulatory therapies such as the dendritic cell-based vaccine sipuleucel-T for castration-resistant prostate cancer[109] and the immune checkpoint inhibitor ipilimumab (which targets cytotoxic T lymphocyte-associated antigen 4 (CTLA4)) for melanoma[110]. Autochthonous mouse tumour models are not only fully immunocompetent but also represent

Table 1 | **Subgroup analysis of preclinical mouse tumour model studies published in 2016**

Type of study	All	Breast	Lung	Colorectal	Melanoma	Pancreas	Brain	Prostate	Liver
In vivo mouse tumour studies	618	139	88	69	62	47	47	36	32
• Cell line-derived	506	129	75	57	59	28	43	29	22
• PDX	43	9	8	4	3	6	3	0	1
• GEM model	147	24	23	12	5	28	7	10	8
• Environmentally induced	38	1	6	12	3	1	0	2	9
Immunocompetent	265	58	38	33	38	31	13	13	18
Immunocompromised	426	105	61	46	35	23	40	27	17
Primary tumour	586	129	80	63	57	47	47	33	32
• Ectopic	331	50	58	51	50	22	23	21	19
• Orthotopic	209	82	16	5	9	18	29	5	8
Metastasis	152	57	28	16	24	7	2	12	6
• Spontaneous	67	29	7	5	6	6	1	4	4
• Spontaneous with tumour resection	9	5	1	0	1	0	1	1	0
• Spontaneous without tumour resection	58	24	6	5	5	6	0	3	4
• Experimental	100	38	22	14	21	1	1	9	2
• Experimental intravenous	77	29	20	7	20	1	1	7	2
• Experimental intracardiac	13	10	2	0	1	0	0	3	0
• Experimental intrasplenic	9	2	3	6	0	0	0	0	0
• Experimental intraperitoneal	6	0	0	1	1	0	0	0	0
Therapeutic	261	63	38	28	36	21	21	18	8
• Prevention	36	7	1	4	7	5	2	2	0
• Intervention	108	24	20	13	15	10	10	8	4
• Insufficient information	117	32	17	11	14	6	9	8	4

Shown are absolute numbers of all indicated subsets of 949 oncology studies published in 2016. GEM, genetically engineered mouse; PDX, patient-derived xenograft.

the corresponding cellular dynamics of multiple microenvironmental components, which are largely compromised in cell line-based models. Therefore, these types of models should be better for preclinically testing therapies targeting the tumour microenvironment. For example, the combination of the chemotherapy gemcitabine and an immunostimulatory CD40 agonist failed to regress autochthonous KPC tumours, although this therapy significantly reduced tumour burden in cell line-based models[111]. Mechanistically, $CD8^+$ T cells failed to infiltrate into KPC tumours due owing the presence of an intact tissue architecture.

PDX models preserve the histological features of their originating tumour and have the potential to represent a wide range of cancer subtypes[3,112]. The predictive value of PDX models was recently demonstrated for gastric cancer, wherein the antiangiogenic multikinase inhibitor regorafenib effectively inhibited tumour growth in PDXs representing diverse histological subtypes such as tubular, papillary, mucinous and poorly cohesive subtypes, as well as uncommon variants[113,114]. Likewise, in a subsequent randomized phase II trial, regorafenib treatment prolonged progression-free survival (PFS) in a highly heterogeneous cohort of patients with refractory gastric adenocarcinoma[115]. Nevertheless, the use of immunocompromised animals and the replacement of human stromal components with mouse counterparts during *in vivo* passaging limit the use of PDXs for preclinical testing of microenvironment-targeted therapies.

Preclinical experiments in drug development utilize a wide range of immunodeficient hosts (75% (195/261) of therapeutic studies; FIG. 4b). Interestingly, 57% (112/195) of studies with immunocompromised animals rely on athymic nude mice (FIG. 4c), despite the fact that these animals have an intact B cell and innate immune compartment, which partially limits engraftment of many solid human tumours and prevents engraftment of haematological cancers[116]. Efforts to humanize the immune system, by replacing mouse genes or tissues with human counterparts, will foster translational research in the

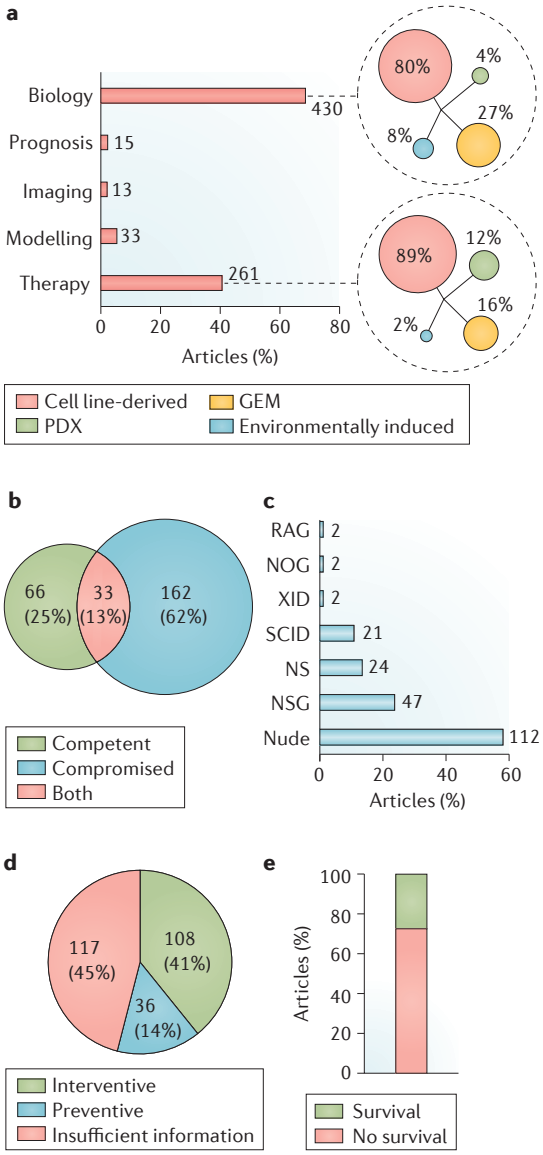

Figure 4 | **Experimental approaches in mouse tumour model studies. a** | The left panel shows the frequency distribution of mouse tumour articles published in 2016 depending on the type of study ($n_{\text{mouse model}} = 618$). Absolute numbers are indicated next to each bar. The distribution of model categories represented within the basic biology and therapy articles ($n_{\text{biology}} = 430$; $n_{\text{therapy}} = 261$) is shown on the right. For biology-focused and therapy-focused studies, respectively, cell line-derived models were used in 344 and 231 studies, patient-derived xenograft (PDX) models in 16 and 31 studies, genetically engineered mouse (GEM) models in 117 and 42 studies and environmentally induced models in 33 and 6 studies. **b** | The frequency of preclinical therapeutic studies that employed immunocompetent and/or immunocompromised models is shown. **c** | Immunodeficient mice can be subdivided into multiple host strains: nude mice harbour the *Foxn1*^{nu} mutation, resulting in deficiency of mature T cells; X-linked immunodeficient (XID) mice carrying the *Btk*^{xid} mutation show impaired B cell function; severe combined immunodeficient (SCID) (S) mice carrying the *Prkdc*^{scid} mutation and RAG mice lacking functional *Rag1* or *Rag2* are deficient for T and B cells; and multigenic models crossed with the NOD/Shi (N) background and carrying specific mutations in the interleukin 2 receptor γ-subunit gene (G) additionally lack innate immune function characterized by natural killer cell deficiency and impaired macrophage and dendritic cell functionality. Such multigenic models include NS, NOG and NSG mice. Shown is their distribution among all therapeutic studies that utilized immunodeficient mice ($n_{\text{immunodeficient}} = 195$). The percentages add up to more than 100% because 15 studies employed multiple immunodeficient models. Absolute numbers are indicated next to each bar. **d** | Distribution of preclinical therapeutic articles ($n_{\text{therapy}} = 261$) that focused on preventive versus interventive therapy. Nearly half of the articles could not be classified because sufficient information such as criteria of therapy initiation was not indicated. **e** | Frequency of preclinical therapeutic studies ($n_{\text{therapy}} = 261$) that employed survival as an end point. Seventy-five studies indicated overall survival, and one study indicated progression-free survival.

years to come[117–119]. PDXs in humanized mice should faithfully reflect the human population dynamics and clonal heterogeneity of cancer and could serve as a screening platform for next-generation targeted therapies including immunomodulatory agents.

Data representation and interpretation. Therapeutic approaches can be classified as interventive and preventive treatments. In preclinical settings, an interventive therapy would be administered to a mouse only following the diagnosis of a tumour by standard readouts such as visible tumour mass, palpation or non-invasive imaging. In turn, a preventive treatment would be administered before the detection of overt disease. Surprisingly, only 55% (144/261) of therapeutic studies reported defined criteria for therapy initiation (41% (108/261) interventive and 14% (36/261) preventive), whereas 45% (117/261) did not include sufficient information (FIG. 4d), calling into question the predictive utility as well as the reproducibility of the studies in this latter group. Of note, only 7 out of the 36 analysed prevention studies employed autochthonous *de novo* tumour models in which tumour incidence, the classical readout for human prevention trials, could be assessed, while the majority utilized non-autochthonous transplantation models.

Gold standard clinical end points include PFS and overall survival (OS). In preclinical research, investigators frequently use tumour growth curves and tumour mass as study readouts to estimate therapeutic response, whereas a survival end point was employed in only 29% (76/261) of studies (FIG. 4e). Using OS as an end point improves the predictive power of preclinical studies[67]. Further, OS end points allow the evaluation of long-term survivors (durable complete remission). For example, it was predicted based on preclinical studies that dual inhibition of CTLA4 and programmed cell death protein 1 (PD1) could provide a long-term survival benefit to patients with melanoma compared with monotherapy[120]. This observation was recapitulated in a clinical trial, in which the median PFS was significantly increased from 2.9 months with ipilimumab monotherapy to 11.5 months when coadministered with nivolumab (which targets PD1)[121]. Similarly, initial clinical results suggest an improved OS (REF. 122). Such examples confirm the need for robust preclinical readouts to anticipate the success of a clinical trial.

Stringent analysis and proper representation of preclinical tumour data could yield valuable biological insights. Waterfall plots represent the response rates of individual patients based on a particular clinical parameter such as percentage change in tumour burden from baseline[123] (FIG. 5a). They provide an ease of visualization and interpretation, which allows clinicians to stratify patients' response rates as progressive disease, partial response or stable disease in accordance with Response Evaluation Criteria in Solid Tumours (RECIST) recommendations. Clinical-like interpretation of waterfall plots could also be achieved in preclinical research by depicting growth curves of individual tumour-bearing mice. Recently, the combination of radiotherapy and anti-CTLA4 treatment was assessed in mice with bilateral flank tumours (with irradiation of one tumour). In this study, anti-CTLA4 treatment led to 17% complete remission of the unirradiated tumours when comparing tumour growth curves of individual mice[124] (FIG. 5b). Despite the clear benefits of this kind of data representation, individual growth curves were depicted in only 7 of the 261 analysed therapeutic studies.

Fine-tuning adjustments in data analysis and representation could also support the mechanistic interpretation of experimental tumour data. Tumour growth is often plotted as simple end point measurements of tumour volume or weight, which allows no judgement of kinetics. Temporal analyses are usually depicted as linear growth curves. Yet, tumours grow exponentially; that is, subtle differences in the early stages of tumour growth may translate into substantial size differences at later stages[125]. In this case, the relative difference between experimental groups becomes larger the longer the experiment lasts. Log-transformation of exponential growth curves instead shows tumour growth rates[125]. An impressive dissociation of linear tumour growth curves, insinuating a major biological effect, could translate into parallel growth curves in a log-transformed graphical representation and indicate that there was in fact no difference biologically. An example of such temporally varying tumour growth behaviour was identified in the preclinical validation of the angiogenesis-regulating TIE2 ligand angiopoietin 2 (ANGPT2)[126]. At the end of the study, tumour volume in *Angpt2*-deficient mice was significantly reduced compared with wild-type mice. Yet, log-transformation of experimental tumour growth data identified a transient window during early tumour growth in which the absence or presence of host-derived ANGPT2 made a difference. Later tumour growth rates were identical (FIG. 5c,d), suggesting that ANGPT2 was dispensable for later stages of tumour progression. Intriguingly, this finding appeared to be recapitulated in later ovarian cancer clinical trials in which the angiopoietin inhibitor trebananib showed an effect on PFS (that is, short-term effect) but failed to prolong OS (that is, long-term effect)[127], an outcome shared by many clinical studies[128]. Thus, a successful preclinical study design not only aims to mimic the human disease but also incorporates clinical study criteria, readouts, end points and post-experimental data analysis.

Perspectives

The shift in preclinical mouse tumour biology research from fundamental discovery research towards more translational target validation and therapy research increases the need for advanced preclinical mouse tumour models that better mimic the pathogenesis and the progression of human tumours as well as their response to therapy. Generally speaking, GEM, PDX and environmentally induced models are superior to cell line models, and an increasing number of studies employ such models. Yet, as outlined in this Analysis, the vast majority of studies continue to employ cell line-based models. While it may be desirable to gradually replace those in the future, the more advanced models are more time consuming and cost intensive. Nevertheless, we

Response Evaluation Criteria in Solid Tumours (RECIST). A rule set aimed at defining whether a patient with a tumour improves (objective response), stays the same (stable disease) or worsens (progressive disease) under treatment.

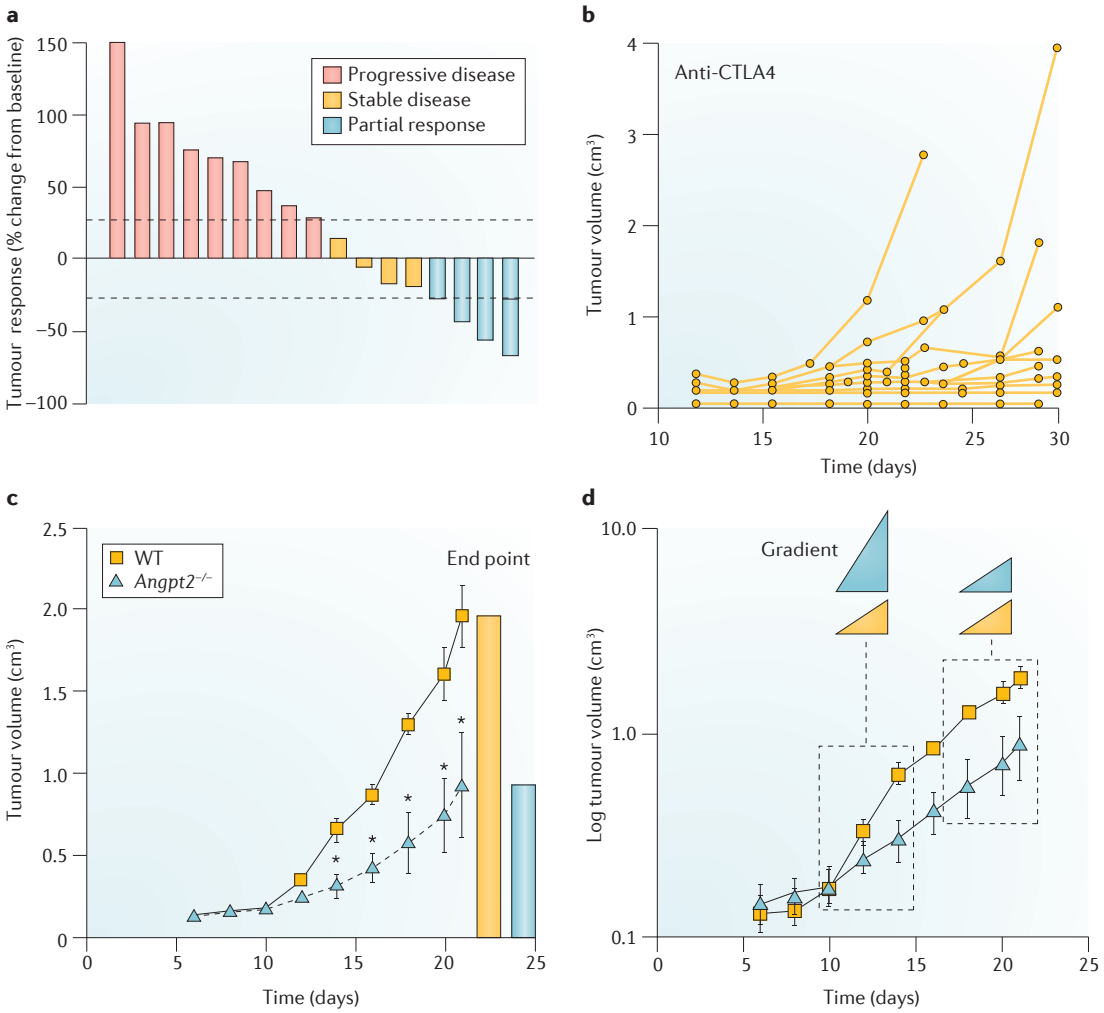

Figure 5 | **Examples of data representation in preclinical mouse tumour studies. a** | Individual patient response rates in clinical trials are often represented as waterfall plots allowing stratification into progressive disease, stable disease or partial response. Dashed lines mark thresholds for progressive disease (in red) or partial response (in blue), according to Response Evaluation Criteria in Solid Tumours (RECIST) guidelines. **b** | In a preclinical mouse tumour experiment, plotting of individual growth curves similarly allows the distinction between responders and non-responders. Shown are response rates of unirradiated tumours upon anti-cytotoxic T lymphocyte-associated antigen 4 (CTLA4) treatment in mice carrying bilateral flank tumours wherein the contralateral tumours were irradiated. **c** | Most preclinical studies focus on end points and growth of tumours. Differences at early stages of tumour growth will increase over time simply as an epiphenomenon of the exponential growth. **d** | By contrast, log-transformation of the same data shows tumour growth rates (indicated as curve gradients), which may be more suitable to detect a transient biological or therapeutic effect. Data in parts **a** and **b** are expressed as mean ± standard error of the mean; *$P < 0.001$ compared with wild type (WT) mice. Parts **a** and **b** are adapted from REF. 124, Macmillan Publishers Limited. Parts **c** and **d** are adapted by permission from the American Association for Cancer Research: Nasarre, P. *et al.* Host-derived angiopoietin-2 affects early stages of tumor development and vessel maturation but is dispensable for later stages of tumor growth. *Cancer Res.* **69**, 1324–1333 (2009) http://dx.doi.org/10.1158/0008-5472.CAN-08-3030 (REF. 126).

believe that initial findings from cell line models should at least be validated in more sophisticated models before clinical consideration. Moreover, new models are being developed at a great pace (BOX 2). While this diversification provides unique opportunities, it could further complicate preclinical mouse tumour biology research and reduce standardization efforts, which would be desirable for better comparison of studies. In this section, we discuss some considerations, beyond the development of new models, that could improve the quality of preclinical mouse tumour research.

Mimicking human therapy. Mouse tumours are studied as a surrogate of human tumours. As such, more emphasis should be placed on designing mouse experiments so that they come as close as possible to the human situation. In humans, tumours are treated after they have been diagnosed (intervention) or measures are taken to reduce the incidence of a cancer (prevention). In mouse experiments, many published therapeutic studies (29/261) directly initiate therapy upon inoculation of tumour cells (cell line studies or PDXs). As a rule of thumb, one may ask what the relevance of such a trial

would be if this approach is primarily affecting tumour engraftment, which in essence is meaningless for human patients. Even more bothersome, a substantial fraction of studies do not supply sufficient details of the treatment protocol to reliably conclude if it is a prevention or an intervention study.

Preclinical mouse studies often present slower tumour growth as an indicator of drug efficacy. By contrast, in a human clinical trial, according to RECIST guidelines, a successful therapy must induce a complete remission or >30% reduction in tumour mass[129]. In fact, slowed tumour growth would be classified as progressive disease, indicating a lack of drug efficacy. It is foreseeable that the rate of successful translation of preclinical drugs into the clinic would be much higher if only those candidates that qualified in mouse studies according to clinical criteria advanced to clinical trials. In that respect, preclinical therapeutic research would additionally benefit from the inclusion of typical clinical end points (PFS, OS) as well as data representation standards such as Kaplan–Meier curves and waterfall plots.

Modelling metastasis and therapeutic resistance. Since the implementation of surgery for primary tumours more than 100 years ago[130], metastasis has become the primary cause of cancer-related mortality. However, in preclinical oncology, metastasis experiments are still commonly conducted without surgical intervention. When primary tumour growth precedes metastasis, it will always be the primary disease that will be decisive for the experiment. Metastasis will only be rate-limiting if the tumour is resected[101]. Grafted cell line-based tumours generally grow focally and can therefore often be surgically removed[131] more easily than the multifocal tumours in GEM models. For these, a potential solution might be GEM model-derived allografts, which metastasize from single orthotopically transplanted tumour fragments or organoids that can subsequently be resected to allow sufficient time for metastasis formation to occur[132]. Thus, metastasis research would likely advance faster if surgical models were employed more widely.

Overcoming drug resistance is another unmet clinical need in oncology. The complex clonal heterogeneity of human tumours generally offers multiple avenues for resistance to therapy. Accordingly, cell line models that are very homogenous often fail to identify translatable mechanisms of resistance[133]. By contrast, PDX cohorts as well as genetically diverse collections of GEM models have already proven to be valuable tools for identifying biomarkers and druggable mechanisms of resistance[3,28]. It can be assumed that a wider availability of these models would further boost drug resistance research. Studying relapse after a successful therapeutic intervention requires models in which the relapsing disease can be reliably distinguished from independently emerging *de novo* tumours. Accordingly, the identification of relapse mechanisms would likely benefit from a more concentrated development of suitable focal tumour models that go beyond a simple inoculation of tumour cell lines, that is, focally inducible GEM models[134].

Monitoring tumour growth. Many preclinical studies are limited to simple end point readouts (for example, tumour weight, tumour volume or survival). While these are clinically the most important parameters, end point readouts neglect the temporal kinetics of tumour growth. FIGURE 5 shows an example in which a biological difference affected growth only transiently. The figure shows an important biological finding but makes the point that growth rates may oftentimes be more insightful than linear growth curves. Moreover, care must be taken to ensure that such differences in early tumour growth are not introduced artefactually as an experimental epiphenomenon. Cell line experiments in particular are very sensitive in the early stages of tumour growth when subtle variations (for example, parental cells versus cells with a gene overexpressed or deleted) may substantially affect engraftment. In conclusion, the validity of many experimental findings, particularly in cell line studies, could likely be improved if greater care was taken to monitor tumour growth over time as closely as possible.

Standardization. Our systematic analysis of a large cohort of tumour studies published in 2016 revealed substantial interstudy variability. For example, when

Box 2 | **Perspectives in mouse tumour modelling technologies**

Cell line-derived models. Novel serum-free culture systems for bladder[144], pancreatic[145] and ovarian[146] cancer cell lines promise to retain histopathological and genomic features of their originating tumours. It remains to be established if these systems will translate into a higher predictive preclinical value of xenografts made from these cell lines.

Patient-derived xenografts. Current modelling efforts include orthotopic patient-derived xenografts (PDXs) that enable spontaneous metastasis formation[89] and models that capture different stages of disease progression, that is, PDXs derived from circulating tumour cells, metastases, minimal residual disease or relapsed tumours (reviewed in REF. 3). Additionally, several approaches to humanize mice (for example, transplantation of human haematopoietic stem cells and genetic integration of major histocompatibility complex components, inhibitory natural killer cell receptors and cytokines) are under development and aim to introduce functional components of the human immune system into PDXs[119].

Environmentally induced models. Initiatives in modelling genetic diversity, such as the collaborative cross[147] or the diversity outbred[148] initiatives, aim to generate genetically complex mouse populations by inter-crossing common inbred strains. Such approaches are likely to boost the identification of genetic cancer risk factors when combined with environmental tumour induction protocols.

Genetically engineered mouse models. Novel strategies (reviewed in REF. 6) aimed at accelerating *in vivo* cancer gene validation via genetic engineering include the genetically engineered mouse (GEM) model–embryonic stem cell strategy[149] as well as piggyBac and Sleeping Beauty transposon-based insertional mutagenesis[150]. CRISPR–Cas9 technology in particular is considered a breakthrough for facilitating rapid germline[151] and somatic cancer modelling[152,153] (extensively reviewed in REF. 154). The discovery of novel site-specific recombinase (SSR) systems such as Dre–*rox*[155], ΦC31 (REF. 156), Vika–*vox*[157], Nigri–*nox* and Panto–*pox*[158] promises to enable complex multi-SSR modelling[159].

Organoid-based models. Recently developed 3D *in vitro* cultures established out of normal, pre-malignant or malignant tissue have been shown to retain hierarchical organization and genomic features of their originating organs. Such organoid cultures derived from patients or GEM models can be genetically and pharmacologically manipulated, allowing high-throughput *in vitro* drug screens[160]. Novel orthotopic transplantation models of organoids for pancreatic and colorectal cancer not only recapitulate characteristic histopathology but also capture all disease stages from pre-malignancy to metastatic disease[161–163]. Accordingly, organoid cultures are likely to further amplify the development of highly predictive genetically engineered as well as patient-derived models.

3R principles
Guidelines aimed at improving animal welfare and the quality of *in vivo* experiments by developing alternative models (replacement), limiting the number of animals used (reduction) and minimizing the suffering of animals (refinement).

Animal Research: Reporting *In Vivo* Experiments (ARRIVE) guidelines
A framework for proper design, analysis and reporting of mouse studies.

randomly comparing 20 studies that have employed the same model, one cannot ignore the substantial variation in different laboratories. There is substantial concern about the reproducibility of preclinical oncology research studies[135]. Model variability may be an important contributor to the difficulties in reproducing a mouse tumour experiment. Cell line studies in particular employ cancer cells that have been passaged for several decades. It is common knowledge among scientists who work with cell lines that the same cell line from two different sources can behave very differently in the same experiment[136]. As such, a community-wide effort to standardize the use of mouse tumour models could strongly advance the quality of oncology research. As discussed above, several repositories have been organized for PDX models. It is conceivable that similar repositories could be introduced for cell line and GEM models. Complementary white papers could define standards for specific models. Electronic databases on preclinical tumour models, such as the recently established Models in Translational Oncology (MiTO) database[137], could similarly contribute towards this end.

There is also a substantial need for a concerted effort towards better standardization due to ethical reasons: the 3R principles define replacement, reduction and refinement as key elements of responsible animal experimentation[138]. Replacement may not be possible in oncology research in most cases owing to the systemic nature of tumour growth. Similarly, reduction is only possible by performing better experiments, but the multitude of important questions to be addressed suggests that an increase in the number of animal experiments is more likely than a reduction. That leaves refinement as the key to ethical animal experimentation in the field of oncology. We believe that standardization of mouse tumour models could be one of the most important measures to better refine models and improve the quality of mouse tumour research.

Cell line experiments versus GEM models. Cell line experiments continue to constitute the majority of preclinical mouse tumour studies. It would be better if cell line studies were gradually replaced by more advanced models. Yet, considering the time and cost of GEM models, it does not seem realistic that they could replace cell line studies in the foreseeable future. It would likely be possible to establish GEM repositories similar to the ongoing efforts for PDXs. These could enable the distribution of standardized (and extensively genomically characterized) GEM model-derived tumour fragments or organoids to investigators throughout the world. The inoculation of tumour fragments would bypass the problems associated with cell lines and conveniently make GEM models available to laboratories throughout the world without the need to actually breed them. Such concerted effort would greatly contribute to the standardized use of GEM models.

Reporting. Our analysis also identified major limitations in the reporting standards of preclinical mouse tumour model data, especially regarding comprehensive outlines of experimental protocols and timelines, including criteria for therapeutic initiation. Detailed reporting standards for *in vivo* experimentation, such as the Animal Research: Reporting *In Vivo* Experiments (ARRIVE) guidelines[139] and the Guidance for the Description of Animal Research in Scientific Publications[140], have been widely endorsed by scientific journals and funding agencies, but their rigorous implementation has not been fully achieved[141]. Journals that publish oncology research should be encouraged to post in their instructions to authors a detailed checklist outlining the specific and standardized reporting requirements for mouse tumour experiments. Recent initiatives from top-tier journals to enforce compliance with such guidelines are heading in the right direction[142,143]. Together with a concerted effort aimed at standardizing widely employed tumour models, such reporting standards will contribute to improving the transparency, reproducibility, clinical significance and quality of preclinical mouse tumour model research.

Conclusions

Mouse models are powerful and indispensable tools in basic as well as translational tumour biology research. New models are developed at an unprecedented rate and continuously expand the toolbox of cancer researchers. In parallel, a concerted effort aimed at addressing some of the issues raised in this Analysis article could substantially help to make more effective use of the broad repertoire of available tools. This includes efforts to better mimic the dynamics of human tumours and their therapy, to standardize the available tumour models, to make GEM and other advanced models more readily available, to monitor tumour growth kinetically and not just as an end point and to improve the reporting standards of mouse tumour experiments.

1. Marx, V. Models: stretching the skills of cell lines and mice. *Nat. Methods* **11**, 617–620 (2014).
2. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
3. Byrne, A. T. *et al.* Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat. Rev. Cancer* **17**, 254–268 (2017).
A comprehensive, state-of-the-art review summarizing bottlenecks and perspectives on current PDX technology.
4. Zitvogel, L., Pitt, J. M., Daillère, R., Smyth, M. J. & Kroemer, G. Mouse models in oncoimmunology. *Nat. Rev. Cancer* **16**, 759–773 (2016).
5. Le Magnen, C., Dutta, A. & Abate-Shen, C. Optimizing mouse models for precision cancer prevention. *Nat. Rev. Cancer* **16**, 187–196 (2016).
6. 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).
An insightful review about the development of novel GEM tumour models and their application in basic and translational oncology.
7. Day, C.-P., Merlino, G. & Van Dyke, T. Preclinical mouse cancer models: a maze of opportunities and challenges. *Cell* **163**, 39–53 (2015).
8. Suggitt, M. & Bibby, M. C. 50 years of preclinical anticancer drug screening: empirical to target-driven approaches. *Clin. Cancer Res.* **11**, 971–981 (2005).
9. Voskoglou-Nomikos, T., Pater, J. L. & Seymour, L. Clinical predictive value of the *in vitro* cell line, human xenograft, and mouse allograft preclinical cancer models. *Clin. Cancer Res.* **9**, 4227–4239 (2003).
10. Talmadge, J. E., Singh, R. K., Fidler, I. J. & Raz, A. Murine models to evaluate novel and conventional therapeutic strategies for cancer. *Am. J. Pathol.* **170**, 793–804 (2007).
11. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).

12. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
13. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
14. Haverty, P. M. *et al.* Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
15. Johnson, J. I. *et al.* Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br. J. Cancer* **84**, 1424–1431 (2001).
16. Sikder, H. *et al.* Disruption of Id1 reveals major differences in angiogenesis between transplanted and autochthonous tumors. *Cancer Cell* **4**, 291–299 (2003).
17. Frese, K. K. & Tuveson, D. A. Maximizing mouse cancer models. *Nat. Rev. Cancer* **7**, 654–658 (2007).
18. Hausser, H.-J. & Brenner, R. E. Phenotypic instability of Saos-2 cells in long-term culture. *Biochem. Biophys. Res. Commun.* **333**, 216–222 (2005).
19. Daniel, V. C. *et al.* A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture *in vitro*. *Cancer Res.* **69**, 3364–3373 (2009).
20. Gillet, J.-P. *et al.* Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *Proc. Natl Acad. Sci. USA* **108**, 18708–18713 (2011).
21. Hidalgo, M. *et al.* Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014).
22. Alkema, N. G. *et al.* Biobanking of patient and patient-derived xenograft ovarian tumour tissue: efficient preservation with low and high fetal calf serum based methods. *Sci. Rep.* **5**, 14495 (2015).
23. Choi, Y. Y. *et al.* Establishment and characterisation of patient-derived xenografts as paraclinical models for gastric cancer. *Sci. Rep.* **6**, 22172 (2016).
24. Pavía-Jiménez, A., Tcheuyap, V. T. & Brugarolas, J. Establishing a human renal cell carcinoma tumorgraft platform for preclinical drug testing. *Nat. Protoc.* **9**, 1848–1859 (2014).
25. Tentler, J. J. *et al.* Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* **9**, 338–350 (2012).
26. Fiebig, H. H., Schuchhardt, C., Henss, H., Fiedler, L. & Löhr, G. W. Comparison of tumor response in nude mice and in the patients. *Behring Inst. Mitt.* **74**, 343–352 (1984).
27. Siolas, D. & Hannon, G. J. Patient-derived tumor xenografts: transforming clinical samples into mouse models. *Cancer Res.* **73**, 5315–5319 (2013).
28. Clohessy, J. G. & Pandolfi, P. P. Mouse hospital and co-clinical trial project — from bench to bedside. *Nat. Rev. Clin. Oncol.* **12**, 491–498 (2015).
29. Morelli, M. P. *et al.* Prioritizing phase I treatment options through preclinical testing on personalized tumorgraft. *J. Clin. Oncol.* **30**, e45–e48 (2012).
A case study exemplifying the application of a personalized PDX model as a patient's 'avatar' to evaluate a battery of potential therapeutic options.
30. Azaro, A. *et al.* A first-in-human phase I trial of LY2780301, a dual p70 S6 kinase and Akt Inhibitor, in patients with advanced or metastatic cancer. *Invest. New Drugs* **33**, 710–719 (2015).
31. Girotti, M. R. *et al.* Application of sequencing, liquid biopsies, and patient-derived xenografts for personalized medicine in melanoma. *Cancer Discov.* **6**, 286–299 (2016).
32. Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015).
A landmark study introducing the 'one animal per model per treatment' trial design to perform efficacy, biomarker and resistance studies in PDX models on a population level.
33. Bardelli, A. *et al.* Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov.* **3**, 658–673 (2013).
34. Kavuri, S. M. *et al.* HER2 activating mutations are targets for colorectal cancer treatment. *Cancer Discov.* **5**, 832–841 (2015).
35. Zanella, E. R. *et al.* IGF2 is an actionable target that identifies a distinct subpopulation of colorectal cancer patients with marginal response to anti-EGFR therapies. *Sci. Transl. Med.* **7**, 272ra12 (2015).
36. Landis, M. D., Lehmann, B. D., Pietenpol, J. A. & Chang, J. C. Patient-derived breast tumor xenografts facilitating personalized cancer therapy. *Breast Cancer Res.* **15**, 201 (2013).
37. Lawrence, M. G. *et al.* Establishment of primary patient-derived xenografts of palliative TURP specimens to study castrate-resistant prostate cancer. *Prostate* **75**, 1475–1483 (2015).
38. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
39. Klco, J. M. *et al.* Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* **25**, 379–392 (2014).
40. Eyre, R. *et al.* Patient-derived mammosphere and xenograft tumour initiation correlates with progression to metastasis. *J. Mammary Gland Biol. Neoplasia* **21**, 99–109 (2016).
41. Garrido-Laguna, I. *et al.* Tumor engraftment in nude mice and enrichment in stroma-related gene pathways predict poor survival and resistance to gemcitabine in patients with pancreatic cancer. *Clin. Cancer Res.* **17**, 5793–5800 (2011).
42. Sivanand, S. *et al.* A validated tumorgraft model reveals activity of dovitinib against renal cell carcinoma. *Sci. Transl. Med.* **4**, 137ra75 (2012).
43. Delitto, D. *et al.* Patient-derived xenograft models for pancreatic adenocarcinoma demonstrate retention of tumor morphology through incorporation of murine stromal elements. *Am. J. Pathol.* **185**, 1297–1303 (2015).
44. Yamagiwa, K. & Ichikawa, K. Experimentelle Studie über die Pathogenese der Epithelialgeschwülste [German]. *Mitt. Med. Fak. Tokio* **15**, 295–344 (1915).
45. Tsutsui, H. Über das künstlich erzeugte Carcinoid in der Maus. *Gann* **12**, 17–21 (1918).
46. Gaylor, D. W. Are tumor incidence rates from chronic bioassays telling us what we need to know about carcinogens? *Regul. Toxicol. Pharmacol.* **41**, 128–133 (2005).
47. Kemp, C. J. Animal models of chemical carcinogenesis: driving breakthroughs in cancer research for 100 years. *Cold Spring Harb. Protoc.* **2015**, 865–874 (2015).
An exhaustive review on the history, recent developments and future perspectives of chemically induced animal models of cancer.
48. Abel, E. L., Angel, J. M., Kiguchi, K. & DiGiovanni, J. Multi-stage chemical carcinogenesis in mouse skin: fundamentals and applications. *Nat. Protoc.* **4**, 1350–1362 (2009).
49. Balmain, A. & Yuspa, S. H. Milestones in skin carcinogenesis: the biology of multistage carcinogenesis. *J. Invest. Dermatol.* **134**, E2–E7 (2014).
50. Kobayashi, T., Owczarek, T. B., McKiernan, J. M. & Abate-Shen, C. Modelling bladder cancer in mice: opportunities and challenges. *Nat. Rev. Cancer* **15**, 42–54 (2015).
51. Gurley, K. E., Moser, R. D. & Kemp, C. J. Induction of lung tumors in mice with urethane. *Cold Spring Harb. Protoc.* http://dx.doi.org/10.1101/pdb.prot077446 (2015).
52. Gurley, K. E., Moser, R. D. & Kemp, C. J. Induction of liver tumors in mice with N-ethyl-N-nitrosourea or N-nitrosodiethylamine. *Cold Spring Harb. Protoc.* **2015**, 941–942 (2015).
53. Gurley, K. E., Moser, R. D. & Kemp, C. J. Induction of colon cancer in mice with 1,2-dimethylhydrazine. *Cold Spring Harb. Protoc.* http://dx.doi.org/10.1101/pdb.prot077453 (2015).
54. 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).
55. Westcott, P. M. K. *et al.* The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2015).
56. Steele, V. E. & Lubet, R. A. The use of animal models for cancer chemoprevention drug development. *Semin. Oncol.* **37**, 327–338 (2010).
57. Liu, P. *et al.* Candidate lung tumor susceptibility genes identified through whole-genome association analyses in inbred mice. *Nat. Genet.* **38**, 888–895 (2006).
58. Wolf, M. J. *et al.* Metabolic activation of intrahepatic $CD8^+$ T cells and NKT cells causes nonalcoholic steatohepatitis and liver cancer via cross-talk with hepatocytes. *Cancer Cell* **26**, 549–564 (2014).
59. Hanahan, D., Wagner, E. F. & Palmiter, R. D. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev.* **21**, 2258–2270 (2007).
60. Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244**, 1288–1292 (1989).
61. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
62. Jonkers, J. & Berns, A. Conditional mouse models of sporadic cancer. *Nat. Rev. Cancer* **2**, 251–265 (2002).
63. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
64. Vooijs, M., Jonkers, J. & Berns, A. A highly efficient ligand-regulated Cre recombinase mouse line shows that *LoxP* recombination is position dependent. *EMBO Rep.* **2**, 292–297 (2001).
65. Schonig, K., Schwenk, F., Rajewsky, K. & Bujard, H. Stringent doxycycline dependent control of CRE recombinase *in vivo*. *Nucleic Acids Res.* **30**, e134 (2002).
66. Singh, M., Murriel, C. L. & Johnson, L. Genetically engineered mouse models: closing the gap between preclinical data and trial outcomes. *Cancer Res.* **72**, 2695–2700 (2012).
67. Singh, M. *et al.* Assessing therapeutic responses in Kras mutant cancers using genetically engineered mouse models. *Nat. Biotechnol.* **28**, 585–593 (2010).
A seminal paper showing the predictive power of GEM models when therapeutic studies are conducted according to clinical criteria.
68. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21**, 846–853 (2015).
69. Tabassum, D. P. & Polyak, K. Tumorigenesis: it takes a village. *Nat. Rev. Cancer* **15**, 473–483 (2015).
70. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA Cancer J. Clin.* **67**, 7–30 (2017).
71. Ding, J. I. E. *et al.* Current animal models of bladder cancer: awareness of translatability (review). *Exp. Ther. Med.* **8**, 691–699 (2014).
72. Hillman, G. G. in *Tumor Models in Cancer Research* 2nd edn (ed. Teicher, B. A.) 287–305 (Humana Press, 2010).
73. Kim, S. Animal models of cancer in the head and neck region. *Clin. Exp. Otorhinolaryngol.* **2**, 55–60 (2009).
74. Larmour, L. I., Jobling, T. W. & Gargett, C. E. A review of current animal models for the study of cervical dysplasia and cervical carcinoma. *Int. J. Gynecol. Cancer* **25**, 1345–1352 (2015).
75. Offield, M. F. *et al.* PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–995 (1996).
76. 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).
77. Tanaka, T. *et al.* A novel inflammation-related mouse colon carcinogenesis model induced by azoxymethane and dextran sodium sulfate. *Cancer Sci.* **94**, 965–973 (2003).
78. Rudel, R. A., Fenton, S. E., Ackerman, J. M., Euling, S. Y. & Makris, S. L. Environmental exposures and mammary gland development: state of the science, public health implications, and research recommendations. *Environ. Health Perspect.* **119**, 1053–1061 (2011).
79. Travis, R. C. & Key, T. J. Oestrogen exposure and breast cancer risk. *Breast Cancer Res.* **5**, 239–247 (2003).
80. Barone, E., Corrado, A., Gemignani, F. & Landi, S. Environmental risk factors for pancreatic cancer: an update. *Arch. Toxicol.* **90**, 2617–2642 (2016).
81. Volkovova, K., Bilanicova, D., Bartonova, A., Letasiova, S. & Dusinska, M. Associations between environmental factors and incidence of cutaneous melanoma. *Environ. Health* **11**, S12 (2012).
82. Field, R. W. & Withers, B. L. Occupational and environmental causes of lung cancer. *Clin. Chest Med.* **33**, 681–703 (2012).
83. Connelly, J. M. & Malkin, M. G. Environmental risk factors for brain tumors. *Curr. Neurol. Neurosci. Rep.* **7**, 208–214 (2007).

84. Guerin, E., Man, S., Xu, P. & Kerbel, R. S. A model of postsurgical advanced metastatic breast cancer more accurately replicates the clinical efficacy of antiangiogenic drugs. *Cancer Res.* **73**, 2743–2748 (2013).
85. Runge, A. *et al.* An inducible hepatocellular carcinoma model for preclinical evaluation of antiangiogenic therapy in adult mice. *Cancer Res.* **74**, 4157–4169 (2014).
86. Francia, G., Cruz-Munoz, W., Man, S., Xu, P. & Kerbel, R. S. Mouse models of advanced spontaneous metastasis for experimental therapeutics. *Nat. Rev. Cancer* **11**, 135–141 (2011).
87. Killion, J. J., Radinsky, R. & Fidler, I. J. Orthotopic models are necessary to predict therapy of transplantable tumors in mice. *Cancer Metastasis Rev.* **17**, 279–284 (1998).
88. Hackl, C. *et al.* Metronomic oral topotecan prolongs survival and reduces liver metastasis in improved preclinical orthotopic and adjuvant therapy colon cancer models. *Gut* **62**, 259–271 (2013).
89. Hiroshima, Y. *et al.* Patient-derived mouse models of cancer need to be orthotopic in order to evaluate targeted anti-metastatic therapy. *Oncotarget* **7**, 71696–71702 (2016).
90. Hiroshima, Y. *et al.* Establishment of a patient-derived orthotopic xenograft (PDOX) model of HER-2-positive cervical cancer expressing the clinical metastatic pattern. *PLoS ONE* **10**, e0117417 (2015).
91. Hoffman, R. M. Patient-derived orthotopic xenografts: better mimic of metastasis than subcutaneous xenografts. *Nat. Rev. Cancer* **15**, 451–452 (2015).
92. Justilien, V. & Fields, A. P. Utility and applications of orthotopic models of human non-small cell lung cancer (NSCLC) for the evaluation of novel and emerging cancer therapeutics. *Curr. Protoc. Pharmacol.* **62**, 14.27 (2013).
93. Mittal, V. K., Bhullar, J. S. & Jayant, K. Animal models of human colorectal cancer: current status, uses and limitations. *World J. Gastroenterol.* **21**, 11854–11861 (2015).
94. Qiu, W. & Su, G. H. Development of orthotopic pancreatic tumor mouse models. *Methods Mol. Biol.* **980**, 215–223 (2013).
95. Zhao, G.-J. *et al.* Establishment of an orthotopic transplantation tumor model of hepatocellular carcinoma in mice. *World J. Gastroenterol.* **18**, 7087–7092 (2012).
96. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168**, 670–691 (2017).
97. Saxena, M. & Christofori, G. Rebuilding cancer metastasis in the mouse. *Mol. Oncol.* **7**, 283–296 (2013).
98. Aguirre-Ghiso, J. A. Models, mechanisms and clinical evidence for cancer dormancy. *Nat. Rev. Cancer* **7**, 834–846 (2007).
99. Karrison, T. G., Ferguson, D. J. & Meier, P. Dormancy of mammary carcinoma after mastectomy. *J. Natl Cancer Inst.* **91**, 80–85 (1999).
100. Pfitzenmaier, J. *et al.* Telomerase activity in disseminated prostate cancer cells. *BJU Int.* **97**, 1309–1313 (2006).
101. Kerbel, R. S. A decade of experience in developing preclinical models of advanced- or early-stage spontaneous metastasis to study antiangiogenic drugs, metronomic chemotherapy, and the tumor microenvironment. *Cancer J.* **21**, 274–283 (2015).
102. Savage, P. & Mahmoud, S. Development and economic trends in cancer therapeutic drugs: a 5-year update 2010–2014. *Br. J. Cancer* **112**, 1037–1041 (2015).
103. Mullard, A. Parsing clinical success rates. *Nat. Rev. Drug Discov.* **15**, 447 (2016).
104. Thomas, D. W. *et al. Clinical development success rates 2006–2015* (BIO, BioMedTracker and Amplion, 2016).
105. Pietras, K. & Hanahan, D. A multitargeted, metronomic, and maximum-tolerated dose "chemo-switch" regimen is antiangiogenic, producing objective responses and survival benefit in a mouse model of cancer. *J. Clin. Oncol.* **23**, 939–952 (2005).
106. Raymond, E. *et al.* Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N. Engl. J. Med.* **364**, 501–513 (2011).
107. Malaney, P., Nicosia, S. V. & Dave, V. One mouse, one patient paradigm: new avatars of personalized cancer therapy. *Cancer Lett.* **344**, 1–12 (2014).
108. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
109. Kantoff, P. W. *et al.* Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N. Engl. J. Med.* **363**, 411–422 (2010).
110. Robert, C. *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).
111. Beatty, G. L. *et al.* Exclusion of T cells from pancreatic carcinomas in mice is regulated by $Ly6C^{low}$ $F4/80^{+}$ extratumoral macrophages. *Gastroenterology* **149**, 201–210 (2015).
112. Cassidy, J. W., Caldas, C. & Bruna, A. Maintaining tumor heterogeneity in patient-derived tumor xenografts. *Cancer Res.* **75**, 2963–2968 (2015).
113. Hu, B. *et al.* Gastric cancer: classification, histology and application of molecular pathology. *J. Gastrointest. Oncol.* **3**, 251–261 (2012).
114. Huynh, H., Ong, R. & Zopf, D. Antitumor activity of the multikinase inhibitor regorafenib in patient-derived xenograft models of gastric cancer. *J. Exp. Clin. Cancer Res.* **34**, 132 (2015).
115. Pavlakis, N. *et al.* Regorafenib for the treatment of advanced gastric cancer (INTEGRATE): a multinational placebo-controlled phase II trial. *J. Clin. Oncol.* **34**, 2728–2735 (2016).
116. Shultz, L. D. *et al.* Human cancer growth and therapy in immunodeficient mouse models. *Cold Spring Harb. Protoc.* **2014**, 694–708 (2014).
117. Holzapfel, B. M., Wagner, F., Thibaudeau, L., Levesque, J. P. & Hutmacher, D. W. Concise review: humanized models of tumor immunology in the 21 st century: convergence of cancer research and tissue engineering. *Stem Cells* **33**, 1696–1704 (2015).
118. Morton, J. J. *et al.* XactMice: humanizing mouse bone marrow enables microenvironment reconstitution in a patient-derived xenograft model of head and neck cancer. *Oncogene* **35**, 290–300 (2016).
A potential breakthrough study on humanized mouse models reporting the engraftment of *ex vivo*-expanded human haematopoietic stem cells into immunodeficient mice.
119. Walsh, N. C. *et al.* Humanized mouse models of clinical disease. *Annu. Rev. Pathol.* **12**, 187–215 (2017).
120. Curran, M. A., Montalvo, W., Yagita, H. & Allison, J. P. PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proc. Natl Acad. Sci. USA* **107**, 4275–4280 (2010).
121. Larkin, J. *et al.* Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* **373**, 23–34 (2015).
122. Hodi, F. S. *et al.* Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol.* **17**, 1558–1568 (2016).
123. Gillespie, T. W. Understanding waterfall plots. *J. Adv. Pract. Oncol.* **3**, 106–111 (2012).
124. Twyman-Saint Victor, C. *et al.* Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. *Nature* **520**, 373–377 (2015).
125. Demidenko, E. Three endpoints of *in vivo* tumour radiobiology and their statistical estimation. *Int. J. Radiat. Biol.* **86**, 164–173 (2010).
126. Nasarre, P. *et al.* Host-derived angiopoietin-2 affects early stages of tumor development and vessel maturation but Is dispensable for later stages of tumor growth. *Cancer Res.* **69**, 1324–1333 (2009).
127. Sheridan, C. Amgen's angiopoietin blocker fails in ovarian cancer. *Nat. Biotechnol.* **33**, 5–6 (2015).
128. Miller, K. *et al.* Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N. Engl. J. Med.* **357**, 2666–2676 (2007).
129. Nishino, M., Jagannathan, J. P., Ramaiya, N. H. & Van den Abbeele, A. D. Revised RECIST guideline version 1.1: what oncologists want to know and what radiologists need to know. *AJR Am. J. Roentgenol.* **195**, 281–289 (2010).
130. DeVita, V. T. Jr & Rosenberg, S. A. Two hundred years of cancer research. *N. Engl. J. Med.* **366**, 2207–2214 (2012).
131. Srivastava, K. *et al.* Postsurgical adjuvant tumor therapy by combining anti-angiopoietin-2 and metronomic chemotherapy limits metastatic growth. *Cancer Cell* **26**, 880–895 (2014).
An illustrative example of the application of mouse models of postsurgical metastasis for evaluating efficacy of adjuvant therapy.
132. Day, C.-P. *et al.* "Glowing Head" mice: a genetic tool enabling reliable preclinical image-based evaluation of cancers in immunocompetent allografts. *PLoS ONE* **9**, e109956 (2014).
133. Rottenberg, S. & Borst, P. Drug resistance in the mouse cancer clinic. *Drug Resist. Updat.* **15**, 81–89 (2012).
134. Gurlevik, E. *et al.* Adjuvant gemcitabine therapy improves survival in a locally induced, R0-resectable model of metastatic intrahepatic cholangiocarcinoma. *Hepatology* **58**, 1031–1041 (2013).
An excellent example of using a focally inducible GEM model to study tumour relapse and metastasis.
135. Morrison, S. J. Time to do something about reproducibility. *eLife* **3**, e03981 (2014).
136. Baker, M. Reproducibility: respect your cells! *Nature* **537**, 433–435 (2016).
137. Galuschka, C., Proynova, R., Roth, B., Augustin, H. G. & Muller-Decker, K. Models in translational oncology: a public resource database for preclinical cancer research. *Cancer Res.* **77**, 2557–2563 (2017).
138. Flecknell, P. Replacement, reduction and refinement. *ALTEX* **19**, 73–78 (2002).
139. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
A perspective article providing a set of rules that preclinical researchers should follow in order to improve the reporting of animal experimentation.
140. National Research Council (US) Institute for Laboratory Animal Research. Guidance for the description of animal research in scientific publications. (*Natl Acad. Press,* 2011).
141. Baker, D., Lidster, K., Sottomayor, A. & Amor, S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol.* **12**, e1001756 (2014).
142. [No authors listed.] Announcement: towards greater reproducibility for life-sciences research in Nature. *Nature* **546**, 8 (2017).
143. Marcus, E. & whole Cell team. A STAR Is Born. *Cell* **166**, 1059–1060 (2016).
144. Hofner, T. *et al.* Development and characteristics of preclinical experimental models for the research of rare neuroendocrine bladder cancer. *J. Urol.* **190**, 2263–2270 (2013).
145. Noll, E. M. *et al.* CYP3A5 mediates basal and acquired therapy resistance in different subtypes of pancreatic ductal adenocarcinoma. *Nat. Med.* **22**, 278–287 (2016).
146. Ince, T. A. *et al.* Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. *Nat. Commun.* **6**, 7419 (2015).
147. Reilly, K. M. Using the collaborative cross to study the role of genetic diversity in cancer-related phenotypes. *Cold Spring Harb. Protoc.* http://dx.doi.org/10.1101/pdb.prot079178 (2016).
148. Churchill, G. A., Gatti, D. M., Munger, S. C. & Svenson, K. L. The diversity outbred mouse population. *Mamm. Genome* **23**, 713–718 (2012).
149. Huijbers, I. J. *et al.* Using the GEMM-ESC strategy to study gene function in mouse models. *Nat. Protoc.* **10**, 1755–1785 (2015).
150. Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat. Protoc.* **12**, 289–309 (2017).
151. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
152. Maresch, R. *et al.* Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nat. Commun.* **7**, 10770 (2016).
153. Weber, J. *et al.* CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc. Natl Acad. Sci. USA* **112**, 13982–13987 (2015).
154. Sánchez-Rivera, F. J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat. Rev. Cancer* **15**, 387–395 (2015).
A thorough review detailing applications of CRISPR–Cas9 technology in oncology research.

155. Anastassiadis, K. *et al.* Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Dis. Models Mech.* **2**, 508–515 (2009).
156. Imayoshi, I., Hirano, K., Kitano, S., Miyachi, H. & Kageyama, R. *In vivo* evaluation of ΦC31 recombinase activity in transgenic mice. *Neurosci. Res.* **73**, 106–114 (2012).
157. Karimova, M. *et al.* Vika/*vox*, a novel efficient and specific Cre/*lox*P-like site-specific recombination system. *Nucleic Acids Res.* **41**, e37 (2013).
158. Karimova, M., Splith, V., Karpinski, J., Pisabarro, M. T. & Buchholz, F. Discovery of Nigri/*nox* and Panto/*pox* site-specific recombinase systems facilitates advanced genome engineering. *Sci. Rep.* **6**, 30130 (2016).
159. Schönhuber, N. *et al.* A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. *Nat. Med.* **20**, 1340–1347 (2014).
160. van de Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).
161. Boj, S. F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
162. O'Rourke, K. P. *et al.* Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nat. Biotechnol.* **35**, 577–582 (2017).
163. Roper, J. *et al. In vivo* genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nat. Biotechnol.* **35**, 569–576 (2017).
Refs 162 and 163 are two compelling examples of the transformative potential of tumour organoid technology for mouse modelling.

Acknowledgements
The authors regret that, because of space limitations, they could not cite all original research articles and related references on this topic. Work in the authors' laboratory is supported by funds from the Deutsche Forschungsgemeinschaft, the Helmholtz Association and the European Union.

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

Competing interests statement
The authors declare no competing interests.

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

DATABASES
Models in Translational Oncology (MiTO): http://www.mito.dkfz.de

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (table) | S3 (figure) | S4 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF