Trends in Cancer

CellPress REVIEWS

Review

In Vivo Cancer-Based Functional Genomics

Chennan Li[1,2] and Andrea L. Kasinski[1,2,3,*]

Pinpointing the underlying mechanisms that drive tumorigenesis in human patients is a prerequisite for identifying suitable therapeutic targets for precision medicine. In contrast to cell culture systems, mouse models are highly favored for evaluating tumor progression and therapeutic response in a more realistic *in vivo* context. The past decade has witnessed a dramatic increase in the number of functional genomic studies using diverse mouse models, including *in vivo* clustered regularly interspaced short palindromic repeats (CRISPR) and RNA interference (RNAi) screens, and these have provided a wealth of knowledge addressing multiple essential questions in translational cancer research. We compare the multiple mouse systems and genomic tools that are commonly used for *in vivo* screens to illustrate their strengths and limitations. Crucial components of screen design and data analysis are also discussed.

Highlights

In vivo screens are in high demand for identifying therapeutic targets for precision medicine.

Selecting a suitable mouse model and an appropriate genomic screening tool is essential for addressing a specific research question.

Recent advances in clustered regularly interspaced short palindromic repeats (CRISPR) toolkits and small guide (sg)RNA design largely enhance the accuracy and diversity of screening outcome.

New computational pipelines are highly favored for the next generation of sgRNA design and screening data analysis.

In Vivo Functional Genomic Screens To Identify Drivers of Tumorigenesis

Human cancers are classified as complex genetic diseases. Indeed, even a single type of cancer is genetically heterogeneous, thereby making it difficult to effectively target these tumors with a single therapeutic agent. To overcome this challenge, multiagent and personalized targeted therapeutic agents against genetic drivers are being investigated and are making their way into clinical trials [1,2]. Unfortunately, the identification of genetic drivers for various cancers is incomplete, leaving many drivers undiscovered. A key aspect to the future success of targeted therapeutics relies on elucidating these unknown drivers and identifying the cohort of patients whose tumors are addicted to the driver. Because mouse models such as the mutant *Kras* model [3,4] represent the most faithful *in vivo* systems for studying human cancer and therapeutic efficacy, *in vivo* functional genomic studies that can fill this knowledge gap are in high demand.

Genetic tools are essential for functional genomic studies. Early studies used retroviruses or transposon systems to identify genetic elements that contribute to carcinogenesis directly *in vivo* [5,6]. Nevertheless, with the advent of novel technologies such as **RNA interference** (**RNAi**; see Glossary) and **clustered regularly interspaced short palindromic repeats (CRISPR)**, precision gene targeting is now achievable and has revolutionized the way that cancer drivers are discovered. Although multiple research questions have been cleverly addressed through CRISPR screening approaches in cell culture [7–9], an increasing number of studies using CRISPR *in vivo* to identify genes that are crucial for the various processes of carcinogenesis have been reported [10–12]. *In vivo* screens using CRISPR have revealed several benefits that are not possible using the previous strategies, as well as limitations (Box 1 and Table 1). Recent advances in the field have largely enhanced the capacity and versatility of these genomic toolkits that are uncovering cancer-causing genes and are beginning to be used for the development of personalized medicine. This review provides comparisons between the various mouse models and genomic tools that are commonly used in human cancer research, and discusses recent progress in several essential aspects for conducting a successful *in vivo* screen using CRISPR. The ultimate goal is to provide a theoretical and technical basis for researchers in the field so that they can select an appropriate experimental system to address their specific question.

[1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
[2]Bindley Biosciences Center, Purdue University, West Lafayette, IN 47907, USA
[3]Purdue Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA

*Correspondence: akasinski@purdue.edu (A.L. Kasinski).

Recent Advances in the Use of RNAi and CRISPR for Functional Genomic Studies

Among the genetic tools available, RNAi and the CRISPR system are widely used technologies for genome-wide functional interrogation that both benefit from high-precision gene targeting (Box 1). The high accuracy of genetic manipulation is dictated by the ability of the short hairpin RNA (shRNA) or the small guide RNA (sgRNA) to direct a nuclease in a sequence-dependent manner to either a transcript or genomic region, leading to target cleavage. Compared with traditional mutagenic tools such as transposons, the precision in gene targeting afforded by shRNAs and sgRNAs permits systematic genotype-to-phenotype studies. Although shRNA and sgRNA systems can easily be harnessed for conducting high-throughput screens, there are also multiple challenges with these systems that can impact on the fidelity of the research findings.

First, on-target gene editing needs to be achieved at high efficiency. In the case of CRISPR-mediated knockout, the goal is to completely abolish the expression of a target gene. Traditionally, sgRNAs used in screens are designed to target early exons of protein-coding genes to facilitate gene knockout, while taking into consideration the requirement for a **protospacer adjacent motif (PAM)** [13]. However, the efficiency of obtaining a gene knockout is complicated by features unique to the sgRNA sequence, to the Cas9 enzyme, and to the genome – such as the bias towards the use of a specific nucleotide for maximal target cleavage [14,15], the occurrence of **in-frame mutations**, and lack of accessibility of the sgRNA to the target gene owing to the chromatin state [16]. Similarly to CRISPR, efficient targeting by RNAi also requires specific features in the target gene and in the small interfering RNA (siRNA) sequence [17].

Box 1. RNAi, CRISPR, and Transposons – Working Mechanisms

RNA interference (RNAi) relies upon the ability of small interfering RNAs (siRNAs) or short hairpin RNA (shRNAs) to post-transcriptionally downregulate target genes through perfect basepairing between the siRNA/shRNA and the target mRNA (Figure IA) ([82] for review). This feature supports high specificity in gene targeting, making RNAi a suitable tool for genotype-to-phenotype evaluation. It is important to recognize that RNAi reduces gene expression at the transcript level, without causing genetic perturbations. In addition, because the processing and activity of RNAi requires machinery that is essential to endogenous miRNA activity ([83,84] for review), it is possible that RNAi could alter miRNA activity and therefore may lead to erroneous discovery [26]. Some of these challenges have been addressed with the development of the CRISPR/Cas9 system.

CRISPR uses a DNA endonuclease, Cas9 ,that is directed by a small guide RNA (sgRNA) to catalyze a DNA double-strand break. Similarly to siRNA and shRNA, the specificity of gene targeting by the CRISPR system is dictated by sequence complementarity with the sgRNA (Figure IB); however, in the case of CRISPR the sgRNA associates with the DNA and not the mRNA ([85] for review). The key feature of the CRISPR system is its ability to cause complete and permanent disruption of the gene at the genomic level, often resulting in loss of functional products as a result of error-prone repair mechanisms. Therefore, CRISPR allows the identification of genes that are essential for a particular disease-related phenotype. Another benefit is the versatile applications in modulating genes, such as inducing transcriptional activation or transcriptional repression (Figure IB), and single-nucleotide editing [86–88], which largely expand its use in a variety of screening applications [89–91].

Although not currently as widespread, the transposon system was among the first systems employed *in vivo* to identify functional genetic elements in cancer. This system consists of a transposon and a catalytic enzyme, transposase. Transposase binds to the two ends of a transposon and catalyzes excision of the transposon and translocation to a different genomic region. This results in a novel mutation that potentially disrupts the expression of the gene into which the transposon becomes inserted. Depending on the regulatory elements (such as a polyA or promoter sequences) that are incorporated within a transposon, activation or repression of target gene expression may occur. The most common tools used in mammalian systems are Sleeping Beauty (SB) and PiggyBac (PB). Both SB and PB favorably integrate into the transcriptional regions of any gene, but differ in their preference of insertion sites (PB, transcriptional start-site and TTAA motif; SB, gene body and TA motif) and the degree to which transposition occurs across the genome (see Table 1 in main text) ([92] for review). Most transposon-mediated screens are conducted directly in GEMMs. In general, following ubiquitous or tissue-specific activation of transposon mobilization, the resulting tumors can be harvested for identification of the genetic elements responsible. To identify genetic contributions, experimental pipelines are applied that usually involve genomic fragmentation and purification of transposon–genome junctions (such as the Splinkerette PCR method) for high-throughput sequencing, followed by extensive read mapping and statistical analysis [93].

Glossary

Clustered regularly interspaced short palindromic repeats (CRISPR): a genetic tool that uses a Cas9 nuclease guided by a small guide RNA (sgRNA) to catalyze a double-strand DNA break through sequence complementarity.
Ectopic implantation: an *in vivo* model in which cells are often subcutaneously implanted into mice to evaluate tumor progression in a nonlocal environment.
Genetically engineered mouse model (GEMM): an *in vivo* model in which one or more genetic mutations are endogenously induced in murine tissue. Genetic manipulation can be achieved globally at the embryonic stage or induced either temporally or spatially.
Humanized mouse: an immunodeficient mouse model that is reconstituted with human hematopoietic stem cells (HSCs) or peripheral blood mononuclear cells (PBMCs), with the aim of evaluating a phenotype or therapeutic response in the presence of a human-like immune system.
In-frame mutation: a genetic mutation that does not cause a shift in the open reading frame, thus leading to a change in one or a few amino acid(s), and therefore may or may not be deleterious for target protein function.
Long noncoding RNA (lncRNA): a group of noncoding RNA molecules greater than 200 nt in length.
miRNA: a group of small noncoding RNA molecules that are 18–22 nt in length that have been increasingly recognized for their functional roles in human development and various diseases.
Orthotopic cell implantation: an *in vivo* model in which cells are implanted into a murine organ that is analogous to its human counterpart from which the cells were originally derived.
Patient-derived xenograft (PDX): tissue harvested from a human patient is implanted into mice. PDXs are favored for their ability to faithfully represent the heterogeneity that was present in the original patient tumors.
Protospacer adjacent motif (PAM): a short sequence of 3–5 nt, located adjacent to the DNA sequences targeted by the sgRNA, that is required for Cas9 activity.
RNA interference (RNAi): a genetic tool that uses a short hairpin RNA (shRNA, ~21 nt in length) or small interfering RNA (siRNA) for target mRNA

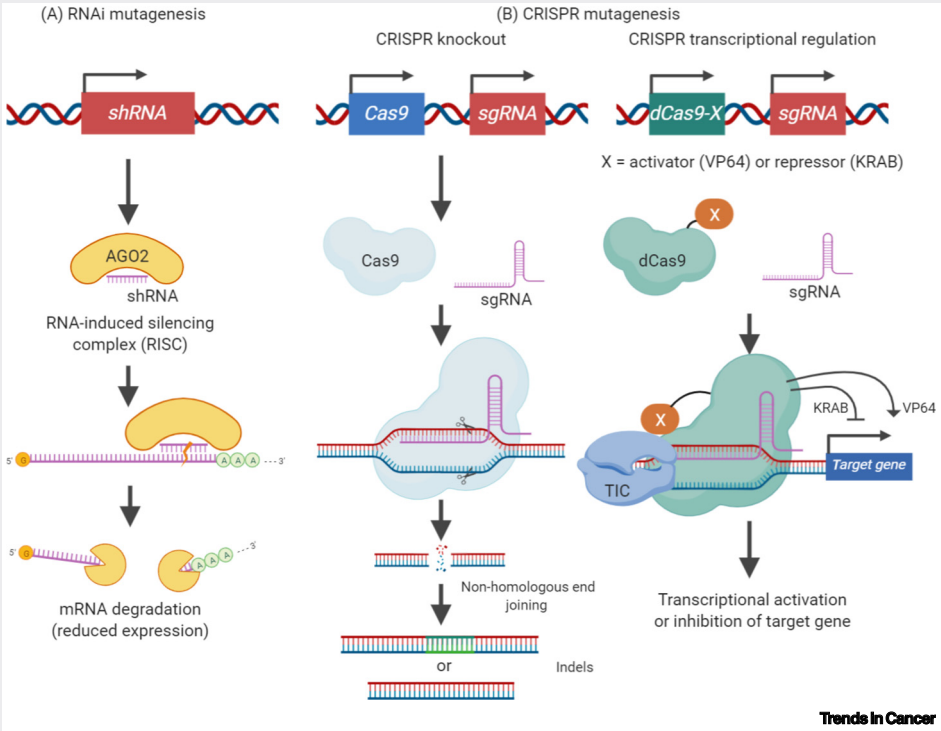

Figure I. RNA Interference (RNAi-) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Mediated Genetic Perturbation. (A) The most common method employed to induce RNAi is through the use of a constitutively expressed short hairpin RNA (shRNA) that targets the gene of interest. Upon expression of the shRNA from its transgenic locus, the guide strand of the shRNA is loaded into AGO2, which forms the RNA-induced silencing complex (RISC). RISC identifies target genes through perfect sequence complementarity between the shRNA and target mRNA, and then catalyzes mRNA cleavage, which is followed by target mRNA decay resulting in reduced target gene expression. (B) The CRISPR system is generally used for genetic knockout and for transcriptional activation or inhibition studies. To execute a genetic knockout, the Cas9 nuclease and a small guide RNA (sgRNA) are expressed. A specific genomic locus is identified by the sgRNA and cleaved by Cas9, and is then repaired by the host non-homologous end joining (NHEJ) machinery, unless a homologous template is provided. Error-prone NHEJ results in small insertions or deletions (indels) that often lead to functional gene knockout. To activate or repress a gene using CRISPR, a promoter-targeting sgRNA and a catalytically dead Cas9 (dCas9) fused to a transcriptional activator (such as VP64) or inhibitor (such as KRAB) are expressed. Once dCas9/sgRNA identifies a target gene, the regulatory module added to dCas9 interacts with the transcription initiation complex (TIC), which up- or down-regulates the target gene depending on the regulatory enzyme. As a side note, other molecular modules such as epigenetic regulators or cytidine deaminases can also be attached to dCas9 for diverse applications.

degradation catalyzed by Argonaute proteins.

Single-cell RNA sequencing: a recently developed technology that allows transcriptomic analysis of each individual cell from the entire sample.

Single-cell lineage tracing: a technology that often uses genome-integrated barcodes to trace each individual cell during tumorigenic progression.

Target-capture sequencing: a technology that aims to amplify and sequence the expected knockout sites. In brief, a molecular inversion probe (MIP) designed for each sgRNA target site binds to the genomic region where mutations are expected. Sequences that contain the target site are then used as a template for DNA synthesis and subsequent sequencing to identify the presence and abundance of the mutation.

Driven by accumulated experimental evidence and models, several computational prediction algorithms [sequence scan for CRISPR gRNA efficiency (SSC) [15,18], CRISPRscan [19], and siDirect [20]] can help to increase on-target efficacy, which has seen marked improvement in recent years.

A second challenge includes off-target effects, where genes that are not intended to be targeted are affected [21,22]. The main off-targets of RNAi include genes that share partial complementarity with the siRNA and are therefore subject to miRNA-like targeting [23]. Various bioinformatic algorithms have been developed to identify these potential off-targets, and these are helpful in siRNA design and in analyzing the resulting screening data [24,25]. In addition, as indicated in Box 1, RNAi can globally impact on **miRNA**-mediated gene regulation to various degrees [26], and this should not be neglected when performing the experiment. Regarding CRISPR,

Table 1. Three Major Tools Used for High-Throughput Loss-of-Function Assays in Mammalian Systems[a]

	Advantages	Disadvantages
Transposon	Direct induction of endogenous transposon mobilization *in vivo* without the need for virus infection Suitable for genome-wide mutagenesis [94] Allows precise identification of novel mutations driving the disease, facilitating downstream characterization of their true genetic influence	Random mutagenesis Local effects (local hopping) may limit the number of genes that can be evaluated (PB has a higher efficiency in genome coverage than SB [93]) Negative selection cannot be conducted A relatively difficult method is employed to identify transposon insertion sites [93]
RNAi	Sequence-guided targeting Gene silencing without genomic editing Loss-of-function libraries are available [40,95] Capable of targeting genes (such as lncRNAs and circular RNAs) that overlap with other genes [96] Genomic integration of LV-delivered shRNA constructs for use as sequencing barcodes can serve as a surrogate for identifying causal genes	Off-target effects related to miRNA-like silencing of unintended genes [23] Potential sequestration of the endogenous miRNA biogenesis machinery [26] The degree of gene silencing varies [97], which may limit the efficiency of identifying essential genes Current RNAi technologies mainly target cytosolic mRNAs, although RNAi-mediated silencing occurs in both cytoplasm and nucleus through different mechanisms [98]
CRISPR	Sequence-guided targeting Capable of knockout at the genomic level Suitable for evaluating essential genes Genome-wide knockout/CRISPRi libraries are available [13,52,89] CRISPR base editors can also allow precise premature termination of target gene expression [32] Genomic integration of LV-delivered sgRNA constructs for use as sequencing barcodes can serve as a surrogate for identifying causal genes	Off-target effects related to Cas9 toxicity [29] and promiscuous Cas9/sgRNA binding [27,28] Gene targeting efficiency varies owing to sgRNA on-target efficacy and chromatin state [14–16] May affect multiple genes located at the same genetic locus Potential influence of host DNA repair systems [99]

[a]Abbreviations: CRISPRi, CRISPR interference; PB, PiggyBac; SB, Sleeping Beauty.

undesired off-target effects occur because of mismatches between the sgRNA and an unintended target that still permit nuclease activity [27,28]. Nuclease-associated toxicity can also occur when targeting a gene with a high copy number, leading to multiple genomic aberrations that may be inefficiently repaired, resulting in G2 cell-cycle arrest [29]. To address these issues, novel strategies have been developed, including the discovery and use of CRISPR-associated protein (Cas) variants (Sniper-Cas9 for enhanced specificity [30], Cas13 for RNA targeting [31], and CRISPR-STOP for silencing high copy-number genes without causing massive genomic damage [32]). Progress in sgRNA design [33] is also helping to reduce off-targeting. Guides with high off-target potential can be predicted through computational algorithms [34] or models that have been experimentally validated [15], and these can provide guidance for more rational sgRNA design. As will be discussed in a later section, these bioinformatic programs to predict on- and off-target efficiencies can also be applied to interpretation of the screening data ([35] for extensive analysis of the computational algorithms used in posterior analysis of CRISPR genomic editing).

An additional mechanism that can lead to off-target effects, specifically related to the CRISPR system, includes genomic overlap between the target gene of interest and an unintended target. This off-targeting often occurs between protein-coding genes and **long noncoding RNAs** (lncRNAs) or between two lncRNAs. For instance, the lncRNAs *HOTAIR* and *HOXC*, or *MALAT1* and *TALAM1*, are encoded from opposite strands. One is therefore unable to deduce whether the molecular reason for the phenotype is caused by alteration of the gene encoded from the sense or antisense strand. Because many lncRNAs share this genomic context, the number of lncRNAs that can be evaluated through CRISPR-mediated knockout is potentially restricted. In this case, targeting individual promoters for transcriptional silencing or knocking out any unique region can overcome these targeting challenges [36,37]. Careful evaluation of the genetic context will aid in designing more accurate targeting agents, leading to reduced off-target

effects. These are all areas of intense investigation and will undoubtedly advance the future of CRISPR-Cas-mediated screens.

General Methodology for Conducting an *In Vivo* CRISPR-Mediated Screen

Given the versatility of CRISPR toolkits and their prevailing use in functional genomic studies, the CRISPR system will be used as an example to illustrate key elements that are necessary for conducting a large-scale, pooled, *in vivo* genetic screen ([38] for review). For researchers interested in conducting an *in vivo* RNAi screen, similar methodology can be employed because CRISPR and RNAi screens share a very similar research design and comparable toolkits (Table 1 for comparison).

Three major steps are involved in an *in vivo* CRISPR screen, as follows. (i) A prerequisite for conducting an *in vivo* screen includes selecting the appropriate mouse model and an experimental design based on the research question. **Ectopic implantation** of cancer cells is technologically the easiest strategy to evaluate the intrinsic tumorigenic potential of the selected cell line. Cells are often implanted in the subcutis of an immunodeficient mouse for assessment of tumor growth or therapeutic response. However, to investigate more realistic stepwise tumor formation, an **orthotopic cell implantation** model or a **genetically engineered mouse model (GEMM)** is often used because they generate tumors in the correct anatomical location, which supports the influence of the correct microenvironmental factors. For example, orthotopic engraftment of cultured mammary cancer epithelial cells into the cleared fat pad of a recipient mouse results in the formation of mammary cancer in the correct anatomical location [39]. A similar orthotopic model was used to conduct an RNAi screen to identify pathways essential for breast tumorigenesis [40]. However, owing to the lack of an intact immune system in the orthotopic model, a GEMM may prove more robust, depending on the research question. With an intact immune system, GEMMs allow one to genetically manipulate cells endogenously that are supported by the correct microenvironmental factors, and permit monitoring tumor initiation and progression [3,4]. Such GEMMs, including various *Kras*-mutated models, are excellent systems for evaluating the entire process of tumorigenesis and response to therapeutics [3,4,41,42]. Owing to these outstanding features, both systems have been highly recognized for their potential to evaluate metastasis and therapeutic efficacy [3,43]. To directly identify therapeutic targets for precision medicine, **patient-derived xenograft (PDX)** models are by far the best *in vivo* system. PDX models can support large-scale genetic or drug screens [44,45]. The use of these mouse models in functional genomic screens is illustrated in Figure 1 (a flow chart for decision making is given in Figure 2, comparisons between different *in vivo* systems in Table 2, and a list of representative *in vivo* screens in Table 3).

(ii) The second step for conducting an *in vivo* CRISPR screen entails performing the mutagenesis, which is achieved through the coordinated delivery of Cas9 and a pool of sgRNAs, typically lentivirus-encoded owing to their ability to be integrated into the genome. This results in a population of mutagenized cells, each carrying a distinct sgRNA barcode and a unique genetic mutation. Integration is essential for 'barcoding' the cells, which is later used to identify the genes that when targeted produced the phenotype. Mutagenesis can be performed in cells in culture that will ultimately be implanted in a mouse model, or directly in a Cas9-expressing mouse model [13,46,47]. In either case, achieving a high degree of library representation (often at 500-fold coverage of the library) is crucial for identifying statistically significant genetic contributions.

(iii) The third and final step involves selecting cells that are associated with a particular phenotype. For *in vivo* screens, the mouse is used as a phenotypic sorter for selecting mutant cells that gain or lose tumorigenic potential. The time of selection needs to be determined for each cell line or

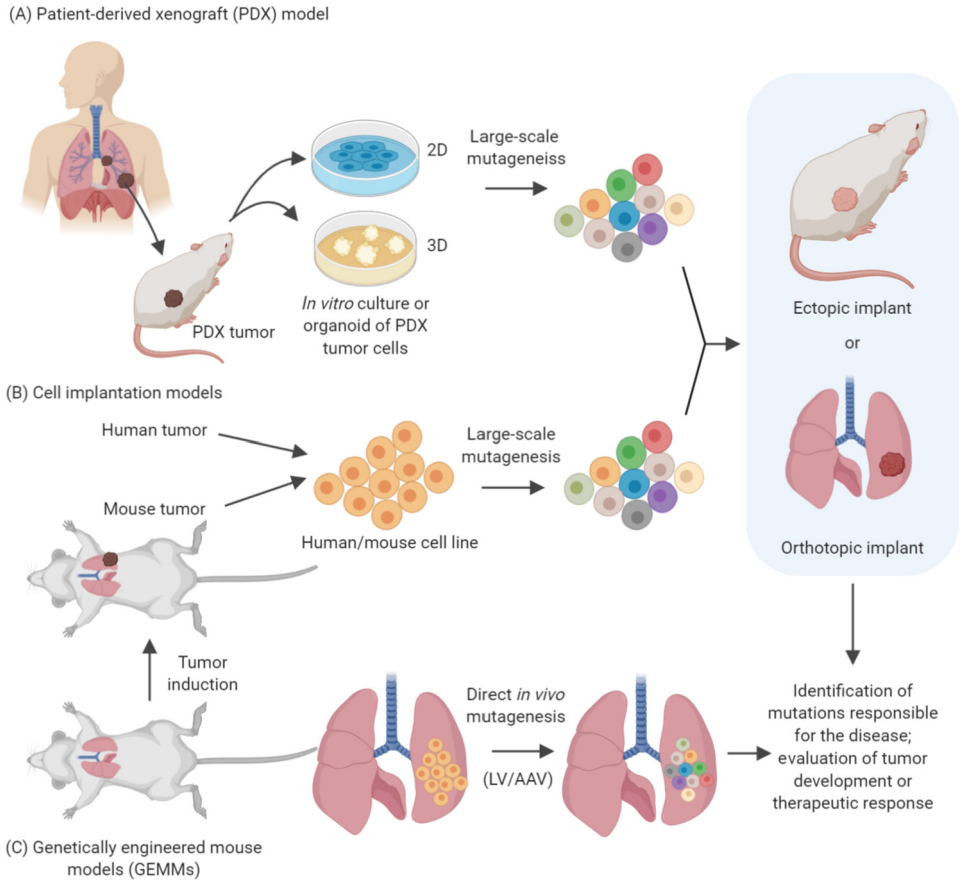

Figure 1. Mouse Models in Genetic Screening. (A) Patient-derived xenografts (PDX) are developed by transplanting a tumor from a patient into an immunodeficient murine host. To identify potential therapeutic targets for the patient, PDX tumors are harvested, cultured *in vitro* (in 2D) or as organoids (in 3D) for genetic manipulation and evaluation of tumor growth or therapeutic response *in vivo*. (B) The cell implantation model employs a cancer cell line derived from a human tumor or from a genetically engineered mouse (GEM) tumor. The cell line is transduced with a shRNA/CRISPR library and mutant cells are implanted ectopically or orthotopically to evaluate their tumorigenic or metastatic potential. (C) In addition, the tumor can be induced through transgene activation in a genetically engineered mouse model (GEMM). To identify potential drivers of cancer in an immunocompetent system, a direct *in vivo* mutagenesis screen can be performed. This is generally achieved by delivering a screening library (in the form of lentivirus, LV; or adeno-associated virus, AAV) to the murine tissue of interest, followed by the identification of shRNA/sgRNA barcodes or genetic mutations that are strongly associated with the phenotype that develops. Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; sgRNA, single guide RNA; shRNA, short hairpin RNA.

tumor type depending on the growth kinetics and the nature of the screen. Harvesting tumor samples too early or too late will result in selective pressure that is either too strong or too weak, potentially leading to the identification of excessive false negatives or false positives, respectively. Following selection, surviving cells are isolated and a pool of mutations, or barcodes (the sgRNAs in this case) associated with the various mutations, are identified through PCR amplification of the integrated viral loci, or the targeted gene loci and are identified by high-throughput sequencing. The sequence reads obtained after selection are then compared with the representation of reads before selection using various bioinformatic pipelines. In a positive-selection screen, mutations that drive tumor or metastatic growth are enriched in the resulting tumors and their over-representation can easily be determined. In a negative-selection

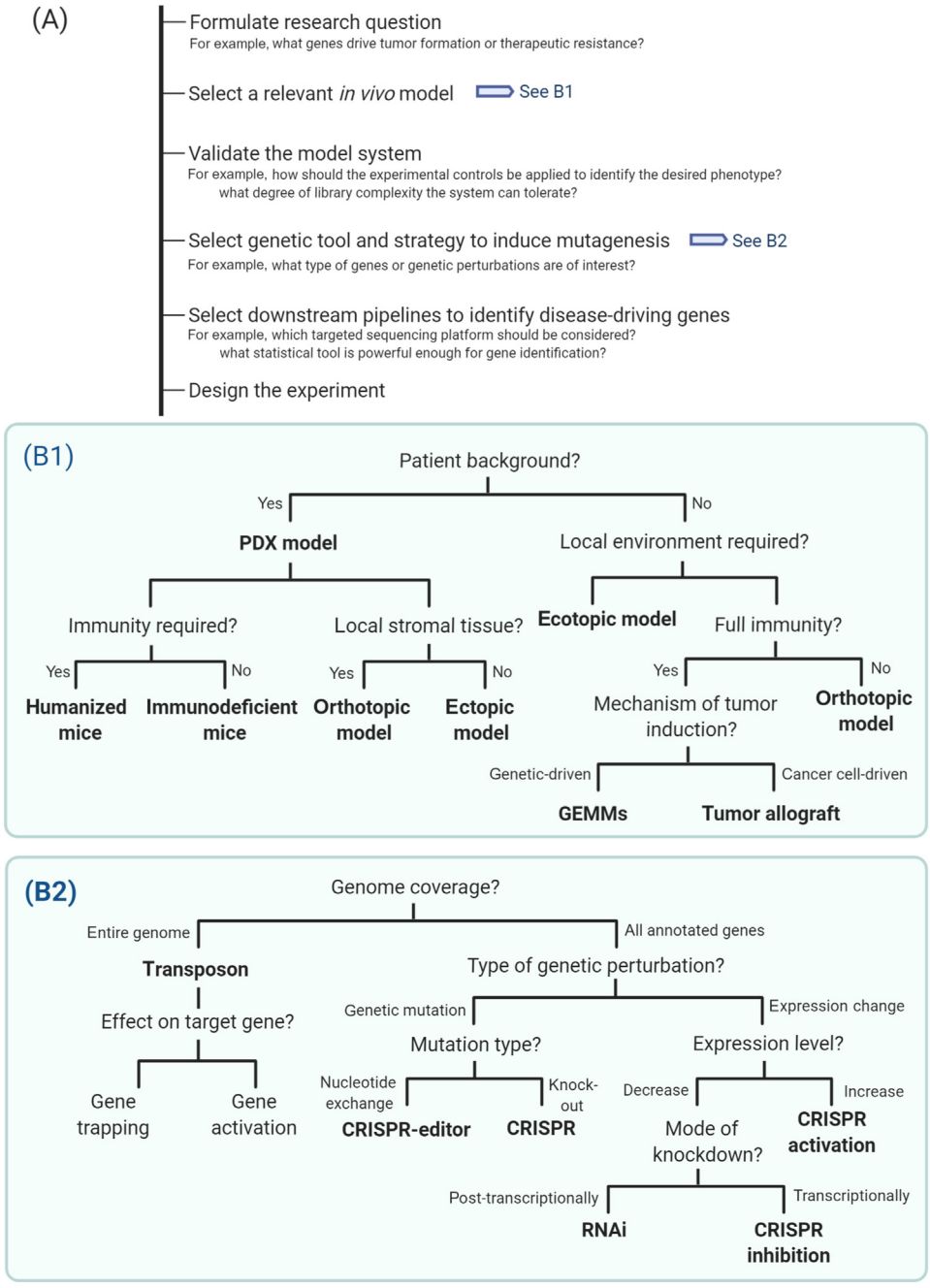

Figure 2. Guidance for Decision Making in an *In Vivo* CRISPR/RNAi Screening Design. (A) Flow chart depicting the various steps one needs to consider before conducting an *in vivo* screen. (B) Steps to decide on selecting appropriate mouse model (B1) and genetic tools (B2) for functional *in vivo* interrogation. Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; GEMM, genetically engineered mouse model; PDX, patient-derived xenograft; RNAi, RNA interference.

screen, sgRNAs that target genes that are essential for tumor growth are under-represented. Recently, **single-cell RNA sequencing** has been adapted for CRISPR screens, which allows in-depth mechanistic evaluation of the sgRNA-induced transcriptomic change and the resulting

Table 2. Comparison of Multiple Mouse Models for Transplantation of Human Cancer Cells and Genetically Engineered Mouse Models

	Advantages	Disadvantages
Ectopic implant	Palpable tumor Easy to establish a tumor and track tumor growth Supports high cell seeding density Suitable for genome-scale screening	Lack of local microenvironment Lack of an intact immune system May fail to support metastasis of human metastatic tumors in general [100] Cell selection over time in 2D may potentially alter the gene expression and tumorigenicity status of the primary tumor [102] Limited fidelity in estimating therapeutic efficacy in clinic [103]
Orthotopic implant	Correct local anatomical features Supports a relatively authentic local environment for a more accurate evaluation of multiple events in tumorigenesis, including the development of metastasis [100] Faithfully represents chemotherapeutic response observed in patients [43,101] Suitable for large-scale screening	Lack of intact immune system Requires more technical expertise, such as surgical placement of cells in difficult-to-access murine organs such as the pancreas [104] Cells are selected over time in 2D May be difficult to monitor tumor progression Limited models (hematopoietic, liver, and mammary systems) have been used in large-scale screens [40,80]
Intravenous implant	Suitable for studying the metastatic potential of tumor cells (extravasation and local tumor expansion) Suitable for large-scale screens	Lack of an intact immune system Cells are selected over time in 2D Difficult to monitor tumor progression, and often requires the use of bioluminescence for tracking metastatic growth
PDX	High accuracy in representing genetic and cell heterogeneity in human tumor Experimental systems have been established to evaluate genetic influences specific to a patient's disease [44,81] Suitable for identifying precision therapeutics	Difficulty in validating targets from a screen owing to tumor heterogeneity Usually require very high library representation owing to the heterogeneity within the patient-derived tumor Graft efficiency differs between surgical specimens [105]
GEMMs	Tumors initiate at the correct anatomical location Intact immune system Influence of tumor microenvironment Pure genetic background, suitable for studying genetic interactions *in vivo* Can recapitulate the therapeutic response in human patients [3]	Requires technical expertise such as delivery of transgenic material, noninvasive tumor monitoring, and maintenance of transgenic strains Difficult to monitor tumor progression Lack of genetic heterogeneity and tumor evolution that take place in human patients

phenotype [9]. **Single-cell lineage tracing** has also been used to identify cancer-causing genes in a single-cell context [48], and this could be beneficial for *in vivo* screens using the PDX model that often harbors a high degree of cell heterogeneity.

In the following sections, several key aspects that are essential for performing a CRISPR screen are discussed, including the selection of an appropriate (i) library, (ii) *in vivo* sgRNA delivery system, and (iii) data analysis pipeline.

Selecting or Designing a CRISPR Library for an *In Vivo* Screen

To gain broad insight into the molecular contribution to carcinogenesis it is necessary to use a large or even a genome-scale library that targets the majority of human or murine genes, including noncoding genes and gene regulatory elements. Several robust genome-scale sgRNA libraries are publicly available. For example, significant findings have been reported using the genome-scale CRISPR knockout (GeCKO) libraries [7,13,49–51] that were designed to target both protein-coding genes and miRNA genes with 4–6 sgRNAs per gene. Many genes identified using these libraries were validated both in cell culture and in human clinical data. Nevertheless, owing to advances in sgRNA design that have enhanced on-target efficacy and/or decreased off-target effects, newer and more optimized libraries have been generated, including the Brunello library (knockout), the Dolcetto library (CRISPR inhibition), and the Calabrese library (CRISPR activation) [52,53]. Generally, the newer libraries use fewer sgRNAs, but the sgRNAs contained are highly efficient targeting constructs [52,53]. To illustrate the advanced design features, Table 4 summarizes and compares three major CRISPR knockout libraries that are most commonly

used. It is also important to note that several nontargeting sgRNAs are routinely used as controls, which are crucial for conducting statistical analysis of the data. To further control for nonspecific nuclease toxicity related to genomic damage, a set of sgRNAs that target nongenic regions in the genome (referred to as 'safe-targeting guide') have been developed and can potentially serve as better controls for growth-related screens [54]. In addition to targeting protein-coding genes and miRNAs, additional sgRNAs are emerging that target other DNA sequences, including regions harboring lncRNAs. In one example, paired sgRNAs were designed to target the promoter and exonic region within lncRNAs [36]. Importantly, lncRNA genes that overlapped with surrounding coding sequences were not included, and this avoided complications in interpreting the finding. Recent studies have also described new libraries with higher flexibility in genomic interrogation, such as dual knockout [55,56], gene knockout together with activation of another gene [57], a regulated CRISPR system using an anti-CRISPR tool [58], and ligand-deactivated sgRNAs [59]. For instance, the dual-knockout system is beneficial for identifying gene pairs that contribute to carcinogenesis. Indeed, nearly all cancers are driven by more than one genetic alteration (such as mutated *Kras* and *Trp53* that together drive lung adenocarcinoma in mice [60]). Nonetheless, in an *in vivo* setting, there is often a limit to the number of cells that can be assayed within a particular anatomical location (Table 2). This makes it difficult and sometimes impossible to perform a full-scale genomic screen. An alternative strategy is to use a smaller library (sometimes referred to as a sublibrary) that targets a subset of genes that belong to a particular biological pathway. Examples include libraries targeting cellular kinases or cell cycle regulators. Additional sublibraries have been constructed that target genes belonging to a specific type of molecule, such as lncRNAs or circular RNAs [10,36,61,62]. Selecting or designing a sublibrary reduces the number of targeting constructs, and thus enhances the sensitivity of the screen through increasing the representation of each sgRNA. For instance, Chow *et al.* constructed a sgRNA library that targets the mouse orthologs of 50 tumor-suppressor genes that are significantly mutated in human cancer [61]. It is also important to note that careful examination of the sgRNA design algorithm is necessary for generating sgRNAs that are efficient in targeting and capable of causing functional mutations such as knockouts of the target genes. Several online resources provide knowledge-driven algorithms for sgRNA design (i.e., GuideScan [33]), and these make it easy to develop a user-specific screening library *in situ.*

Selection of an Appropriate Library Delivery Method for *In Vivo* Use

Once the library is selected or designed, it is essential to deliver it to the target cells or tissues with high efficiency. Current systems for transducing CRISPR libraries *in vivo* mostly use either adeno-associated viruses (AAVs) or lentiviruses (LVs). When selecting the delivery strategy, the major concern lies in the trade-off between the potential unwanted phenotypic effects due to viral integration (in the case of LV) and the need to identify mutations through extensive procedures such as **target-capture sequencing** (in the case of AAV). For both systems, the researcher needs to conduct appropriate statistical analysis of the screening data, and all positive results need to be carefully validated before a conclusion can be made. These steps are particularly crucial for LV-mediated screens because potentially irrelevant mutations introduced by LV insertions may skew the fidelity of the research finding. One way in which nearly all libraries overcome this deficiency is through the inclusion of four to six sgRNAs that target the same gene. Following selection, if more than one sgRNA is altered in representation, this is unlikely to be a result of the integration site. No matter which virus is being used, it is recommended that the cells are transduced at a low multiplicity of infection (MOI), often less than the MOI of 0.5 that is employed in cell culture [49]. By this design every cell is expected to be transduced with at most one viral particle, resulting in a single genetic perturbation per cell. This will help to avoid a perplexing validation process. Although not as high-throughput, other strategies for delivering targeting constructs *in vivo* include direct lipid-based transfection [63], microinjection/electroporation [64,65], the

Table 3. Representative Studies Using Pooled CRISPR/RNAi *In Vivo* Screens in Mouse Models

Genes expected to be identified	Screening library	Positive/negative screen	Strategy to identify mutations	Mouse model	Method of cancer induction[a]	Genes validated[b]	Refs
Suppressors of metastasis	Whole-genome LV-sgRNA (GeCKO, 67,405 sgRNAs)	Positive	Barcodes, fold change in abundance of multiple sgRNAs	Ectopic implant (*Nu/Nu*)	Sc injection of murine NSCLC Cas9 line	*Nf2* *Pten* *Trim72* *Cdkn2a* *Fga* *Cryba4* *miR-152* *miR-345*	[49]
Mediators of therapeutic resistance	LV-sgRNA (9872 sgRNAs), targeting 2368 genes selected from relevant GO terms	Positive (enrichment)	Barcodes, STARS algorithm	Ectopic implant (wild-type and *Tcra*$^{-/-}$ syngeneic)	Sc injection of murine melanoma B16 Cas9 cell line	***Ptpn2*** *H2-T23* *Ripk1* *Stub1*	[10]
Maintainers of breast cancer	LV-shRNA (700 shRNAs), targeting 133 user-selected genes (metabolic and transporters)	Negative	Barcodes, genes targeted by multiple shRNAs that are significantly depleted	Orthotopic implant (NOD; *scid*)	Ot injection of human breast cancer MCF10DCIS.com cell line	***PHGDH*** *GMPS* *SLC16A3* *PYCR1* *VDAC1*	[40]
Suppressors of leukemia therapeutic resistance	LV-sgRNA (175 sgRNAs), targeting promoters of 25 DNA damage genes	Positive	Barcodes, fold change in the abundance of multiple sgRNAs	Orthotopic implant (*Eμ–Myc p19*$^{Arf-/-}$ syngeneic)	Iv injection of murine B cell lymphoma (BCL) dCas9-VP64 cell line	No validation studies for the screening result	[95]
Suppressors of *Myc*-driven lymphoma	Retroviral (LMS)-shRNA (2300 shRNAs), targeting cancer 1000 gene sets	Positive	Barcodes, shRNAs with the highest enrichment	Orthotopic implant (*Eμ–Myc* syngeneic)	Iv injection of normal murine HSCs taken from Eμ–Myc mice	***Mek1*** ***Rad17*** *Ang2* *Sfrp1* *Numb*	[80]
Suppressors of *Kras*^{G12D} lung adenocarcinoma	LV-Cre-shRNA (27 500 shRNAs), targeting 4625 signaling genes	Positive	Barcodes, user-set algorithm based on the occurrence of multiple shRNAs in multiple tumors	GEMM (*LSL-Kras*^{G12D})	It injection of viruses encoding the library	***EphA2***	[106]

Suppressors of *p53*$^{-/-}$ glioblastoma	AAV-Cre-sg-p53-sgRNA (288 sgRNAs), targeting 49 mouse tumor-suppressor genes	Positive	Target-capture sequencing and VarScan, user-set algorithm for the identification of significant mutations and candidate genes	GEMM (*LSL-Cas9*)	Ic injection of viruses encoding the library	Co-occurring gene pairs	[61]
Maintainers (epigenetic regulators) of PDAC PDX tumor	LV-shRNA (2410 shRNAs), targeting 236 epigenetic regulators	Negative	Barcodes (fold change in abundance), RSA gene scoring (top-scoring three shRNAs)	PDX-2 implant (NSG)	Sc injection of human PDAC xenografts, and murine GEMM PDAC tumors	*PHF5A* *SMC2* *WDR5*	[81]
Maintainers (epigenetic regulators) of melanoma PDX tumor	LV-shRNA (2410 shRNAs), targeting 236 epigenetic regulators	Negative	Barcodes (fold change in abundance), user set algorithm (based on the median of log fold change of all six shRNAs)	PDX-2 implant (NSG)	Sc injection of human melanoma xenografts	*BAZIB* *SMARCA4* *CHD4* *KMT2D*	[44]
Negative regulators of T cell infiltration into a tumor	Whole-genome LV-sgRNA (cloned from the mouse GeCKO libraries, 129 209 sgRNAs)	Positive	Barcodes; RIGER and MAGECK analysis	GEMM (Cas9 mouse) for harvesting T cells, and Iv injection for blood reconstitution of mutagenized T cells	Ot/Sc implantation of triple-negative breast cancer cells made to be recognizable by the T cells used in the study	***Dhx37*** *Odc1*	[11]
Essential maintainers of leukemia progression	Whole-genome LV-sgRNA (Brie, 78 637 sgRNAs)	Negative	Barcodes; fold change in the abundance of multiple sgRNAs	GEMM (BCR-ABL/NUP98-HOXA9 bcCML)	Iv injection of murine Cas9 leukemic cells in a B6 transplant recipient	***Stau2***	[12]

[a]Abbreviations: GO, gene ontology; Ic, intracranial; Iv, intravenous; It, intratracheal; LMS, PDAC, pancreatic ductal adenocarcinoma; NSCLC, non-small-cell lung carcinoma; Ot, orthotopic; RSA, redundant siRNA activity; Sc, subcutaneous.
[b]Genes highlighted in bold font have been further characterized in the manuscript.

Table 4. Comparison between Three Major Human-Targeting Genome-Wide CRISPR Knockout Libraries

	GeCKO human library [13]	Brunello human Library [52]	Bassik human library [54]
Target	19 050 human genes (including protein-coding genes and miRNAs)	19 114 human genes (protein-coding only)	20 500 human genes (protein-coding only)
sgRNAs/gene	4 for each miRNA, 6 for each protein-coding gene	4	~10
Total number of sgRNAs	123 411	76 441	Varies between 17 438 and 31 324 per sublibrary
Control sgRNAs	2000 nontargeting	1000 nontargeting	5600 nontargeting, and 6750 safe-targeting controls that target nonfunctional nongenic regions
Knockout efficiency	Varies	sgRNAs selected for high on-target efficiency and low off-target effects	sgRNAs with various lengths designed for lower off-target activity
Side note	Two half-libraries	Single library	9 sublibraries in which different categories of genes are targeted

use of nanomaterials [66], and hydrodynamic delivery of transposons for genomic integration of CRISPR constructs [67]. Many of these methods have thus far only been used for therapeutic application and not for screens. Other targeted delivery approaches, such as those used to deliver miRNAs [68–70], may also have a future in targeted sgRNA delivery. Overall, selection of a library delivery strategy depends on the general experimental design and technical considerations. In particular, when direct *in vivo* mutagenesis is proposed, the transgene delivery efficiency and the tropism of the selected method need to be understood because these can vary among different tissues in mice [71]. It is therefore recommended to conduct preliminary experiments to assay *in vivo* targeting efficiency of the intended cells/tissue to ensure that the library is sufficiently represented.

High-Throughput Sequencing and Bioinformatic Analysis of Screening Results

Efficient identification of candidate genes driving the investigated cellular phenotype requires robust methods for sequencing the mutations. Equally importantly, it is essential to use a set of bioinformatic and statistical tools that are capable of teasing out crucial genes from a large pool of irrelevant genes. If a focused screen is conducted using a small library, or a strong phenotypic selection is applied such that only a few mutations are present in the selected cells, direct cloning and sequencing of the sgRNAs can be performed at low cost and in a relatively short amount of time. Nevertheless, in most cases, large-scale screens result in an extensive number of genetic mutations (or sgRNAs) that are enriched or depleted at varying levels. Therefore, high-throughput next-generation sequencing (NGS) is necessary to resolve the differences in representation of the sgRNAs. Current strategies involve targeted amplification and sequencing of either the sgRNA barcodes that are inserted into the genome or sequencing of the endogenous locus where the knockouts are expected to take place, depending on which library delivery system is used. Following mapping of sequencing reads to the library or to the human genome using software such as the variant detecting software VarScan [72,73], the relative abundance of reads indicative of each mutation can be determined. To identify statistically significant mutations, it is then crucial to select a bioinformatic pipeline suitable for the chosen library. For sgRNA libraries with multiple sgRNAs targeting each gene, altered representation of more than one sgRNA following selective pressure suggests that the candidate gene likely plays a relevant role. Using a similar principle, the RNAi gene enrichment ranking (RIGER) and the model-based analysis of genome-wide CRISPR-Cas9 knockout (MAGeCK) algorithms have been widely used for data analysis of RNAi and CRISPR screens [7,74,75], and these have proved to be robust in identifying functional genes. It is important to note that the latter pipeline, MAGeCK, generates a sgRNA enrichment score (or a beta value) for each gene based on the fold increase/decrease in read counts of all 4–6

sgRNAs [76], which is a more informative ranking system. Other bioinformatic pipelines such as caRpools [77] and newer algorithms that use different statistical models for data analysis such as CRISPRBetaBinomial [78] are also available.

To tackle the challenge related to the on- or off-target issues reminiscent of the CRISPR system, many learning- or knowledge-driven tools have been developed, such as sequence scan for CRISPR gRNA efficiency (SSC) [15]. These algorithms can predict the targeting efficacy and the probability of off-target events based on existing screening datasets [15]. This knowledge can then be used not only for designing highly effective sgRNAs but also for correcting the screening data based on different sgRNA efficiencies. Because genomic targeting efficacy is influenced by multiple biological factors, including the targeting sequence [14,15] and chromatin state [16], these algorithms are therefore beneficial for analyzing the screening data and have proved to be essential for understanding many biological questions related to CRISPR targeting. Similarly, multiple algorithms including the genome-wide enrichment for seed sequence matches (GESS) [24] and gespeR that incorporate potential off-targets have been used in interpreting data from RNAi screening [25]. These bioinformatic programs together with accumulated experimental data are useful for advancing data interpretation [79], and therefore largely support the power of high-throughput screening.

With enhanced understanding of the biology of the CRISPR system and improvements in screen design, significant advances have been made in the corresponding bioinformatic tools. Regardless, experimentally validating whether the highly efficient sgRNAs are indeed capable of creating the desired knockout in various cell types is necessary to fully support the contribution of the gene/genetic perturbation to the observed phenotype.

Concluding Remarks and Future Perspectives

Albeit in its infancy, *in vivo* screens using state-of-the-art mutagens have shown great promise in the discovery of important cancer genes and therapeutic targets. Newer mouse models such as the **humanized mouse** and GEMMs have made it possible to address diverse research questions related to the complexity of particular diseases, which was previously difficult. Furthermore, the advent of genomic editing tools has fostered a wide range of genetic studies that are able to obtain a high precision of genetic perturbation. Targeted mutagenesis, particularly with the development of CRISPR editing, allows a better interpretation of the causal relationship between a gene (or a mutation) and a disease phenotype. In recent years multiple new discoveries have been made in cancer research using RNAi/CRISPR screens, including those conducted *in vivo* [10,80,81]. Further progress in the experimental strategies will pave the way towards a more complete understanding of cancer biology and future medicine.

Although conducting functional genomic studies *in vivo* has multiple benefits, such as directly pinpointing clinically relevant mechanisms, there are several questions that the current technology cannot adequately address, but will hopefully be solved with future advances (see Outstanding Questions). Most of the unanswered questions relate to the complexity of the tumor microenvironment and general concerns with gene targeting. First, it is well established that each cell type in the microenvironment contributes to tumorigenesis and therapeutic response in different ways. It is also highly likely that more than one gene is dysregulated in these tumor-associated tissues, resulting in altered intercellular signaling and thereby creating a niche that favors disease progression. Indeed, *in vivo* screens have been conducted to identify genes in immune cells, including CD8 T lymphocytes, that modulate tumor progression [11]. Nevertheless, other genetic drivers or modulators remain to be discovered in the wide range of cells that reside in the tumor microenvironment. This knowledge will be a prerequisite for identifying important

Outstanding Questions

Is the current sgRNA design strategy sufficient for efficiently targeting a wide range of genomic elements including various noncoding RNAs and regulatory elements? Further advances in sgRNA design require (i) knowledge of diverse uncharacterized noncoding sequences and the biological mechanisms that are necessary to target these regions using CRISPR, and (ii) further refinement of computational tools to obtain more precise targeting efficiency and reduced off-target effects, followed by vigorous experimental validation.

Do the identified targets translate into useful targets for therapeutic intervention? Human tumors are highly heterogeneous, and the immune system and the tumor microenvironment are intimately involved in human tumorigenesis. Most systems used for conducting *in vivo* CRISPR screens do not faithfully recapitulate either tumor heterogeneity or the tumor microenvironment, and those that do are usually conducted in immunocompromised animals.

How do we identify targeted therapeutics for individual cancer patients? Traditional strategies involve the identification of common genetic markers and applying a single targeted therapy to patients with the aberration, and this often results in an ineffective therapeutic outcome owing to tumor heterogeneity between patients. Personalized *in vivo* genetic or drug screens using mouse models will be crucial for identifying patient-specific drivers. It is therefore necessary to expand functional genomic research by using patient-derived xenograft models to establish a scientific basis for future precision medicine and to streamline the process to provide timely therapy for patients.

pathways of intercellular communication and for understanding how heterogeneity contributes to disease.

Second, heterogeneity among patients generates another layer of complexity. The genetic causes of the disease and therapeutic responses vary among patients, as indicated by the frequent occurrence of targeted therapeutic resistance. Identifying the major cause(s) is the first step for successfully treating a patient with the most effective therapy. PDX mouse models can serve as a good proxy for identifying patient-specific genetic or epigenetic drivers and evaluating therapeutic outcome in a relatively accurate manner [42,43,81]. However, owing to the multiple constraints described in Table 2, only a few studies have used PDX models for *in vivo* genetic screens (Table 3). Because the PDX model, together with the humanized system, are the best models for recapitulating the histopathological features in patient tumors, there is a definite need to expand research using PDX models with tissues from more patients and with cancerous tissues isolated at various stages. Doing so will generate insight into the heterogeneity among patients and the underlying mechanisms that will be necessary for designing precision therapeutics.

Lastly, accurately identifying functional genes requires further development of experimental systems of RNAi/CRISPR screens. The accuracy of gene targeting and editing relies on the nature of the targeting reagents such as Cas9/sgRNA. Insights into the molecular mechanism of CRISPR gene targeting will lead to enhanced design of high-efficiency sgRNAs and nucleases, which will enable the screening methodology to more accurately reflect the true genetic impact. Furthermore, as we learn more about genomic information and cancer biology, additional screening libraries with new design principles, or libraries targeting a different set of genes or the noncoding space will be necessary to identify additional genomic alterations that function in cancer. Continued efforts in designing bioinformatic algorithms should then follow to assist scientists in identifying authentic and crucial genes that contribute to human cancer.

Author Contributions

C.L. and A.L.K. drafted and edited the manuscript.

Acknowledgments

Figures were generated using BioRender.com.

References

1. Simon, R. and Roychowdhury, S. (2013) Implementing personalized cancer genomics in clinical trials. *Nat. Rev. Drug Discov.* 12, 358–369
2. Tannock, I.F. and Hickman, J.A. (2016) Limits to personalized cancer medicine. *N. Engl. J. Med.* 375, 1289–1294
3. Singh, M. *et al.* (2010) Assessing therapeutic responses in Kras mutant cancers using genetically engineered mouse models. *Nat. Biotechnol.* 28, 585–593
4. Kasinski, A.L. and Slack, F.J. (2012) MiRNA-34 prevents cancer initiation and progression in a therapeutically resistant K-ras and p53-induced mouse model of lung adenocarcinoma. *Cancer Res.* 72, 5576–5587
5. Nusse, R. and Varmus, H.E. (1982) Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* 31, 99–109
6. Dorr, C. *et al.* (2015) Transposon mutagenesis screen identifies potential lung cancer drivers and CUL3 as a tumor suppressor. *Mol. Cancer Res.* 13, 1238–1248
7. Golden, R.J. *et al.* (2017) An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* 542, 197–202
8. Feldman, D. *et al.* (2019) Optical pooled screens in human cells. *Cell* 179, 787–799
9. Datlinger, P. *et al.* (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14, 297–301
10. Manguso, R.T. *et al.* (2017) In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* 547, 413–418
11. Dong, M.B. *et al.* (2019) Systematic immunotherapy target discovery using genome-scale in vivo CRISPR screens in CD8 T cells. *Cell* 178, 1189–1204
12. Bajaj, J. *et al.* (2020) An in vivo genome-wide CRISPR screen identifies the RNA-binding protein Staufen2 as a key regulator of myeloid leukemia. *Nat. Cancer* 1, 410–422
13. Shalem, O. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening. *Science* 343, 84–88
14. Doench, J.G. *et al.* (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267
15. Xu, H. *et al.* (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* 25, 1147–1157
16. Schwartz, U. *et al.* (2019) Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. *Nucleic Acids Res.* 47, 1239–1254
17. Reynolds, A. *et al.* (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.* 22, 326–330
18. Li, W. *et al.* (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* 16, 281
19. Moreno-Mateos, M.A. *et al.* (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* 12, 982–988

20. Naito, Y. *et al.* (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.* 32, 124–129
21. Tsai, S.Q. *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–198
22. Frock, R.L. *et al.* (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33, 179–186
23. Jackson, A.L. *et al.* (2006) Widespread siRNA 'off-target' transcript silencing mediated by seed region sequence complementarity. *RNA* 12, 1179–1187
24. Sigoillot, F.D. *et al.* (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat. Methods* 9, 363–366
25. Schmich, F. *et al.* (2015) GespeR: a statistical model for deconvoluting off-target-confounded RNA interference screens. *Genome Biol.* 16, 220
26. Grimm, D. *et al.* (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441, 537–541
27. Hsu, P.D. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832
28. O'Geen, H. *et al.* (2015) A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* 43, 3389–3404
29. Aguirre, A.J. *et al.* (2016) Genomic copy number dictates a gene-independent cell response to CRISPR-Cas9 targeting. *Cancer Discov.* 6, 914–929
30. Lee, J.K. *et al.* (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* 9, 3048
31. Konermann, S. *et al.* (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* 173, 665–676
32. Kuscu, C. *et al.* (2017) CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* 14, 710–712
33. Perez, A.R. *et al.* (2017) GuideScan software for improved single and paired CRISPR guide RNA design. *Nat. Biotechnol.* 35, 347–349
34. Listgarten, J. *et al.* (2018) Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* 2, 38–47
35. Chuai, G.H. *et al.* (2017) In silico meets in vivo: towards computational CRISPR-based sgRNA design. *Trends Biotechnol.* 35, 12–21
36. Zhu, S. *et al.* (2016) Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* 34, 1279–1286
37. Liu, Y. *et al.* (2018) Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat. Biotechnol.* 36, 1203–1210
38. Humphrey, S.E. and Kasinski, A.L. (2015) RNA-guided CRISPR-Cas technologies for genome-scale investigation of disease processes. *J. Hematol. Oncol.* 8, 31
39. Miller, F.R. *et al.* (1981) Preferential growth of mammary tumors in intact mammary fatpads. *Cancer Res.* 41, 3863–3867
40. Possemato, R. *et al.* (2011) Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* 476, 346–350
41. Kasinski, A.L. *et al.* (2015) A combinatorial microRNA therapeutics approach to suppressing non-small cell lung cancer. *Oncogene* 34, 3547–3555
42. Edmonds, M.D. *et al.* (2016) MicroRNA-31 initiates lung tumorigenesis and promotes mutant KRAS-driven lung cancer. *J. Clin. Invest.* 126, 349–364
43. Ambrogio, C. *et al.* (2014) Modeling lung cancer evolution and preclinical response by orthotopic mouse allografts. *Cancer Res.* 74, 5978–5988
44. Bossi, D. *et al.* (2016) In vivo genetic screens of patient-derived tumors revealed unexpected frailty of the transformed phenotype. *Cancer Discov.* 6, 650–663
45. Gao, H. *et al.* (2015) High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* 21, 1318–1325
46. Wang, T. *et al.* (2015) Identification and characterization of essential genes in the human genome. *Science* 350, 1096–1101
47. Platt, R. *et al.* (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455
48. Michlits, G. *et al.* (2017) CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat. Methods* 14, 1191–1197
49. Chen, S. *et al.* (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis resource genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260
50. Yau, E.H. *et al.* (2017) Genome-wide CRISPR screen for essential cell growth mediators in mutant KRAS colorectal cancers. *Cancer Res.* 77, 6330–6340
51. Pal, A.S. *et al.* (2020) Loss of SUV420H2 promotes EGFR inhibitor resistance in NSCLC through upregulation of MET via LINC01510. *BioRxiv* Published online March 18, 2020. https://doi.org/10.1101/2020.03.17.995951
52. Doench, J.G. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191
53. Sanson, K.R. *et al.* (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* 9, 5416
54. Morgens, D.W. *et al.* (2017) Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* 8, 15178
55. Han, K. *et al.* (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* 35, 463–474
56. Najm, F.J. *et al.* (2018) Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* 36, 179–189
57. Boettcher, M. *et al.* (2017) Decoding directional genetic dependencies through orthogonal CRISPR/Cas screens. *BioRxiv* Published online March 25, 2017 https://doi.org/10.1101/120170
58. Nakamura, M. *et al.* (2019) Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells. *Nat. Commun.* 10, 194
59. Kundert, K. *et al.* (2019) Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat. Commun.* 10, 2127
60. DuPage, M. *et al.* (2009) Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protoc.* 4, 1064–1072
61. Chow, R.D. *et al.* (2017) AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. *Nat. Neurosci.* 20, 4–6
62. Li, S. *et al.* (2020) Screening for functional circular RNAs using the CRISPR-Cas13 system. *BioRxiv* Published online March 25, 2020. https://doi.org/10.1101/2020.03.23.002865
63. Zhang, L. *et al.* (2017) Lipid nanoparticle-mediated efficient delivery of CRISPR/Cas9 for tumor therapy. *NPG Asia Mater.* 9, e441
64. Wu, Y. *et al.* (2019) Generating viable mice with heritable embryonically lethal mutations using the CRISPR-Cas9 system in two-cell embryos. *Nat. Commun.* 10, 2883
65. Gurumurthy, C.B. *et al.* (2019) Creation of CRISPR-based germline-genome-engineered mice without ex vivo handling of zygotes by i-GONAD. *Nat. Protoc.* 14, 2452–2482
66. Lee, K. *et al.* (2017) Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* 1, 889–901
67. Xu, C. *et al.* (2017) piggyBac mediates efficient in vivo CRISPR library screening for tumorigenesis in mice. *Proc. Natl. Acad. Sci. U. S. A.* 114, 722–727
68. Orellana, E.A. *et al.* (2017) FolamiRs: ligand-targeted, vehicle-free delivery of microRNAs for the treatment of cancer. *Sci. Transl. Med.* 9, eaam9327
69. Orellana, E.A. *et al.* (2019) Enhancing microRNA activity through increased endosomal release mediated by nigericin. *Molecular Ther. Nucleic Acids* 16, 505–518
70. Myoung, S. and Kasinski, A.L. (2019) Strategies for safe and targeted delivery of microRNA therapeutics. In *MicroRNAs in Diseases and Disorders: Emerging Therapeutic Targets*

(Peplow, P.V. *et al.*, eds), pp. 386–415, Royal Society of Chemistry
71. Wang, C. *et al.* (2003) Recombinant AAV serotype 1 transduction efficiency and tropism in the murine brain. *Gene Ther.* 10, 1528–1534
72. Koboldt, D.C. *et al.* (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25, 2283–2285
73. Koboldt, D.C. *et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576
74. Jain, I.H. *et al.* (2016) Hypoxia as a therapy for mitochondrial disease. *Science* 352, 54–61
75. Marceau, C.D. *et al.* (2016) Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* 535, 159–163
76. Li, W. *et al.* (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554
77. Winter, J. *et al.* (2016) CaRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. *Bioinformatics* 32, 632–634
78. Jeong, H.H. *et al.* (2019) Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives. *Genome Res.* 29, 999–1008
79. Riba, A. *et al.* (2017) Explicit modeling of siRNA-dependent on- and off-target repression improves the interpretation of screening results. *Cell Syst.* 4, 182–193
80. Bric, A. *et al.* (2009) Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. *Cancer Cell* 16, 324–335
81. Carugo, A. *et al.* (2016) In vivo functional platform targeting patient-derived xenografts identifies WDR5-Myc association as a critical determinant of pancreatic cancer. *Cell Rep.* 16, 133–147
82. Wilson, R.C. and Doudna, J.A. (2013) Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* 42, 217–239
83. Kasinski, A.L. and Slack, F.J. (2011) MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer* 11, 849–864
84. Orellana, E.A. and Kasinski, A.L. (2015) MicroRNAs in cancer: a historical perspective on the path from discovery to therapy. *Cancers* 7, 1388–1405
85. Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911
86. Qi, L.S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183
87. Perez-Pinera, P. *et al.* (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* 10, 973–976
88. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424
89. Gilbert, L.A. *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661
90. Konermann, S. *et al.* (2014) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588
91. Hess, G.T. *et al.* (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13, 1036–1042
92. Denicola, G.M. *et al.* (2015) The utility of transposon mutagenesis for cancer studies in the era of genome editing. *Genome Biol.* 16, 229
93. Friedrich, M.J. *et al.* (2017) Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nat. Protoc.* 12, 289–309
94. Chang, H. *et al.* (2019) Efficient genome-wide first-generation phenotypic screening system in mice using the piggyBac transposon. *Proc. Natl. Acad. Sci. U. S. A.* 116, 18507–18516
95. Braun, C.J. *et al.* (2016) Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3892–E3900
96. Lennox, K.A. and Behlke, M.A. (2016) Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* 44, 863–877
97. Huesken, D. *et al.* (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.* 23, 995–1001
98. Avivi, S. *et al.* (2017) Visualizing nuclear RNAi activity in single living human cells. *Proc. Natl. Acad. Sci. U. S. A.* 114, E8837–E8846
99. Haapaniemi, E. *et al.* (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930
100. Fidler, I.J. (1990) Critical factors in the biology of human cancer metastasis: twenty-eighth G.H.A. Clowes Memorial Award Lecture. *Cancer Res.* 50, 6130–6138
101. Veeranki, O.L. *et al.* (2019) A novel patient-derived orthotopic xenograft model of esophageal adenocarcinoma provides a platform for translational discoveries. *Dis. Model. Mech.* 12, dmm041004
102. Mouriaux, F. *et al.* (2016) Effects of long-term serial passaging on the characteristics and properties of cell lines derived from uveal melanoma primary tumors. *Invest. OphthanImol. Vis. Sci.* 57, 5288–5301
103. Garber, K. (2006) Realistic rodents? Debate grows over new mouse models of cancer. *J. Natl. Cancer Inst.* 98, 1176–1178
104. Reyes, G. *et al.* (1996) Orthotopic xenografts of human pancreatic carcinomas acquire genetic aberrations during dissemination in nude mice. *Cancer Res.* 56, 5713–5719
105. Baccelli, I. *et al.* (2013) Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* 31, 539–544
106. Yeddula, N. *et al.* (2015) Screening for tumor suppressors: loss of ephrin receptor A2 cooperates with oncogenic KRas in promoting lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6476–E6485