New Horizons in Translational Medicine 2 (2014) 5–11

Contents lists available at ScienceDirect

New Horizons in Translational Medicine

journal homepage: www.elsevier.com/locate/nhtm

Research Articles

Animal models in translational medicine: Validation and prediction

Tinneke Denayer, Thomas Stöhr*, Maarten Van Roy

Ablynx NV, Dept. Pharmacology, Zwijnaarde, Belgium

ARTICLE INFO

Available online 27 August 2014

Keywords:
Animal model
Drug development
Translational value
Fit-for-purpose validation

ABSTRACT

Despite large investments in drug development, the overall success rate of drugs during clinical development remains low. One prominent explanation is flawed preclinical research, in which the use and outcome of animal models is pivotal to bridge the translational gap to the clinic. Therefore, the selection of a validated and predictive animal model is essential to address the clinical question. In this review, the current challenges and limitations of animal models are discussed, with a focus on the fit-for-purpose validation. Moreover, guidance is provided on the selection, design and conduct of an animal model, including the recommendation of assessing both efficacy and safety endpoints. In order to improve the clinical translation, the use of humanized mouse models and preclinical applications of clinical features are discussed. On top, the translational value of animal models could be further enhanced when combined with emerging alternative translational approaches.

Focal points:

- Bedside
  Animal models are essential for translation of drug findings from bench to bedside. Hence, critical evaluation of the face and predictive validity of these models is important. Reversely, clinical bedside findings that were not predicted by animal testing should be back translated and used to refine the animal models.
- Benchside
  Proper design, execution and reporting of animal model results help to make preclinical data more reproducible and translatable to the clinic.
- Industry
  Design of an animal model strategy is part of the translational plan rather than (a) single experiment (s). Data from animal models are essential in predicting the clinical outcome for a specific drug in development.
- Community
  Review, standardization and refinement of animal models by disease expert groups helps to improve rigor of animal model testing. It is important that the applied animal models are validated fit-for-purpose according to stringent criteria and reproducible.
- Governments
  As during drug development fit-for-purpose animal models are key for success in clinical translation, financial investments and support from the government to develop, optimize, validate and run such translation tools are important. Over time, this will be of benefit for patients and healthcare institutions.
- Regulatory agencies
  Preclinical testing of a drug in an animal model is not a prerequisite for regulatory agencies before entering clinical trials, but does unquestionably provide valuable data on the expected clinical performance of the drug. Hence, testing in animal models is largely recommended from both a business and patient perspective. In addition, inclusion of safety parameters in animal models will help to build the required safety data package of drugs in development.

* Correspondence to: Ablynx NV, Technologiepark 21, 9052 Zwijnaarde, Belgium. Tel.: +32 9 262 0049; fax: +32 9 262 0001.
E-mail address: thomas.stoehr@a2m-pharma.com (T. Stöhr).

http://dx.doi.org/10.1016/j.nhtm.2014.08.001
2307-5023/

1. Introduction

Success rates for drugs during clinical development remain low despite the human genome project and other molecular biology approaches having identified a large number of potential new drug targets. Moreover, progress in the design of small molecule drugs (e.g. computer aided drug design) and the development of new biological drug formats such as Nanobodies® [1], has yielded a plethora of promising drug candidates addressing these potential new targets. As a consequence, a higher number of drugs interacting with less validated targets have entered clinical development during the last decade. However, this had led to a decline in the success rates for drugs during clinical development. The main reason for these failures is the lack of efficacy during phases II and III of clinical development [2,3]. The possible failure rates could be decreased by having more stringent success criteria during the non-clinical stages of the drug development process. This is especially true during target validation and the preclinical proof of concept stage for a given development candidate. Success during these stages is heavily dependent on the selected animal models that allow for assessment of the target validity, and that can predict clinical efficacy of a specific compound. A model is a simple representation of a complex system. Consequently, an animal model for a human disease is by no means attempting to reproduce the human disease with all its complexities in an animal but rather to model specific aspects of a disease. Whenever using an animal model, it is thus of utmost importance to define a specific question and to ensure that the chosen model is fit-for-purpose.

When adequately designed and conducted, animal models can contribute invaluable information to our knowledge of biology and medicine, including the discovery and development of new drugs. However, better design and conduct as well as further development of animal models is warranted. The current paper highlights some aspects for improving the translational value of animal models.

2. Proper design, conduct and reporting of experiments

The best validated animal model (for validation criteria see further below) is not able to yield conclusive data when the experimental design is flawed or the execution of the study is not well controlled. A number of factors which should be considered in conducting animal studies and that are well-known though not always followed are highlighted below:

(1) Time course of treatment: In animal models, treatment is frequently initiated either before or shortly after the disease pathology is initiated, i.e. before or early during the course of the disease (prophylactic treatment). This is in contrast to the clinical situation in which treatment is normally started after onset of symptoms and clear diagnosis (therapeutic treatment). Thus, a potential pharmacological effect could be overestimated in an animal model, simply because therapeutic intervention occurs earlier in the disease process as compared to the clinical situation.
(2) Animal characteristics and background: The species and strain of animals selected for a particular model need to be carefully selected. First of all, the drug to be tested should be fully cross reactive to the animal target. This is especially an issue for many biological drugs that are often cross reactive to non-human primates only. In addition, the age, gender and health status of the animals should be matched as closely as possible to the clinical condition. Malignancies, Alzheimer's disease or osteoarthritis are diseases of the elderly population and thus screening of new drug candidates in young animals can give misleading results [4]. Finally, depending on the pathology and the target of interest, very special conditions in the design need to be considered. The anti T-cell co-stimulatory receptor CD28 antibody TGN1412 caused a severe cytokine storm in its first human volunteer trial [5,6], an effect that was not observed in animal models using non-human primates. One potential reason that might have contributed to this discrepancy is the fact that the antibody caused the activation of memory T-cells in the human volunteers, but less so in non-human primates despite a near 100% sequence identity of the target [7].
(3) Subjective endpoints: Many outcomes used in animal models are dependent upon subjective interpretation. While subjective evaluations are generally a very efficient way to score behavioral endpoints, it can create bias if the scorer is aware of the animal's treatment. Thus, ideally the experimenter should be unaware of the treatments or manipulations of the animals he is dealing with. In addition, the intra- and inter-operator variation of most subjective measures is high, and standardization is difficult.
(4) Reproducibility of experimental animal results: While most experimental set-ups are very much standardized in a particular lab, slightly different parameters in another one may yield different results [8–10]. Repetition of experimental findings in slightly different models (e.g. a tumor xenograft model with a different cell line of the same cancer type) might thus help to ensure the observed effect is generalizable to a broader context.
(5) Group size: For ethical and other reasons, the number of animals used for biomedical experiments is minimized. However, this needs to be balanced with the statistical power required to generate solid data in order to either verify, or to reject the experimental hypothesis.
(6) Reporting: recent review articles have addressed the issue of insufficient reporting of experimental animal data which makes it difficult for others to reproduce those experiments [11–14]. Thus, for this question, the reader is referred to the published reviews. An additional problem is that experiments with a positive outcome are more likely to be published than negative results.

3. Implementing safety

During the evaluation of a drug candidate, the assessment of efficacy and safety is normally performed in different experiments, with a possible overlap in some doses. While efficacy is usually assessed in disease models without monitoring of side effects, safety is examined in healthy animals at high dose. The doses effective in the disease model are then compared to that in the safety evaluation to yield a safety margin. This margin might be systematically overestimated for two reasons:

(1) Healthy animals normally used for the safety assessment might be less sensitive for potential side effects as compared to diseased subjects.
(2) The efficacy in a disease model could be overestimated when disregarding potential side effects making it impossible to administer corresponding doses in a clinical setting. Moreover, they could directly interfere with correct interpretation of endpoints. For instance, many drugs at high doses could have unspecific effects on motor activity while many behavioral efficacy parameters are directly or indirectly depending on a motor response. Any unspecific effect of a drug on motor activity can therefore be falsely interpreted as a false positive effect.

The risk for this possible efficacy overestimation could be mitigated by using animal disease models for assessing the safety

of drugs in parallel with the standard safety testing. For instance, animals suffering from chronic epilepsy have been shown to be more susceptible than healthy animals to cognitive impairing side effects of the anticonvulsant drug valproate [15]. Consequently, including safety and quality of life endpoints (e.g. overall activity, body weight, food consumption) in animal efficacy studies should become standard.

4. Validation of an animal model and a preclinical strategy

Animal models can be validated according to a number of different criteria:

(1) Face validity: The similarity in biology and symptoms between the animal model and the human disease. Although important for the validation of a disease model, assessing face validity is often hampered by lack of understanding the biology underlying the disease symptoms.
(2) Predictive validity: Demonstration that clinically effective interventions demonstrate a similar effect in the model. This is often difficult to achieve given incomplete correlation between animal and human disease mechanisms, and the inability of approved human drugs to be active in the appropriate animal model species.
(3) Target validity: The target under investigation should have a similar role in the disease model as in the clinical situation. One classical example is the beta-3 adrenergic receptor which has an important role in the energy metabolism of rodents but not in humans [16].

These criteria are typically used to provide a general validation of a model. However, since animal models can be used for quite a variety of different purposes (e.g. examining pathomechanisms, benchmarking compounds against standard of care, providing proof of concept for a new target), it is important that the validation provided for a given model is fit-for-purpose. For instance, face validity may be more important when researching potential pathomechanisms whereas the predictive validity has a higher priority for models to be used for benchmarking.

A model, by definition, is not a perfect replication of the clinical condition. Thus, not all criteria can be met by a single model. However, a combination of different models can eventually come closer to the clinical situation than a single, even highly sophisticated model. In an attempt to define an optimal combination of models, Sams-Dodd [17] has proposed a model validity scoring system which, in an extended version, is shown in Table 1.

This scoring system uses five different criteria:

(1) Species: The closer a species comes to human, the more likely it is that the pathophysiology of the disease is similar to humans.
(2) Complexity: The more complex the test system is, the more probable that the relevant mechanisms are included. For instance, an *in vitro* ion-channel test may detect the effect of a test compound on conductance of a cardiac ion-channel whereas an *ex vivo* or an *in vivo* test system can evaluate its effect on the overall cardiac effect e.g. on contractibility.
(3) Disease simulation: Current models use different principles to induce the disease of interest. The simplest models do not even attempt to induce a disease but simply look at a measure in healthy individuals, e.g. the use of memory to predict cognitive enhancing effects of drug candidates to treat Alzheimer's disease. Somewhat more complex is the use of drugs to induce disease symptoms such as phencyclidine or amphetamine to elicit psychotic like symptoms. For many disorders the etiology has not been fully elucidated making it nearly impossible to truly simulate the disease in a model. Infectious diseases may be one of the exceptions on the ease of replicating a disease simulation as shown for example by the development of a sophisticated neonatal lamb model for respiratory syncytial virus (RSV) infection [18].
(4) Predictivity: A drug effect in an experimental model can, in principle, be observed in two different manners: a quantal or a graded response. A quantal response does just indicate as to whether a drug is active or not. A graded response, however, allows one to distinguish between drugs or doses with higher and lower activity. Such models allow the comparison of different drugs and may help to decide as to whether an experimental compound has a similar or even superior efficacy to the existing standard of care.
(5) Face validity: This criterion can be further differentiated depending on whether just one symptom of a disease is modeled or a set of symptoms and whether this includes core symptoms as for instance defined in the ICD-10 (www.who.int).

This scoring system can help to assemble a screening cascade/combination of models which altogether has maximal validity. For example, the neonatal lamb model for RSV infection scores high for all criteria other than species and could thus be ideally complemented by an *ex vivo* human cell culture model [19].

5. Humanization of models

Humanized mouse models in which immunodeficient mice are engrafted with human cells or tissues, are considered extremely useful as they permit functional research studies *in vivo* and hence support clinical translation. Dependent on the human disease and question addressed, different humanized models and mouse strains are utilized [20]. Most commonly used are the human tumor xenograft models for study of cancer, and the humanized mouse models that mimic the human immune system.

5.1. Human tumor models (e.g. PDX)

Many preclinical animal models in oncology drug development fail to accurately predict the clinical efficacy of novel anticancer agents,

Table 1
Proposed validity scoring system. Adapted from [17].

Criterion	Value	Score
Species	Human	4
	Non-human primate	3
	Non-human mammal	2
	Non-mammal	1
Disease simulation	True	4
	Complex	3
	Pharmacological	2
	No	1
Face validity	> 1 core symptom	4
	1 core symptom	3
	1 symptom	2
	No	1
Complexity	*In vivo*	4
	Tissue	3
	Cellular	2
	Sub-cellular/molecular	1
Predictivity	Graded for all pharmacology principles	4
	Graded for certain pharmacology principles	3
	All or none for certain pharmacology principles	2
	No or not shown	1

largely due to their inability to reflect the complexity and heterogeneity of human tumors. Patient-derived xenograft (PDX) models, where surgically resected primary tumor samples are engrafted directly from patients onto immunodeficient mice, maintain more of the molecular, genetic and histological heterogeneity of their parental tumors [21]. The possible advantages of PDX models compared to other preclinical alternatives in oncology are listed in Table 2. Well characterized PDX models represent an information-rich preclinical resource for analysis of drug activity, including novel drug combinations, as well as predictive biomarker discovery [22,23]. PDX models can be predictive for efficacy of cytotoxic drugs in patients, when these chemotherapeutics are tested in mice using pharmacokinetically clinically equivalent drug doses [24,25], and they offer a route towards personalized medicine for cancer patients. As a proof of principle for the latter, tumor graft models developed for sarcoma, melanoma, adenocarcinoma [26] and ovarian cancer [27] demonstrated strong correlation with clinical efficacy when tumor graft response was used to guide treatment for patients. Although all these examples highlight the potential for these PDX models to become the standard for modeling human cancers, there are still a number of disadvantages related to the use of PDX models (Table 2).

5.2. Mice with humanized immune system

Over the last decade, much progress has been achieved in the development of appropriate humanized mouse models to serve preclinical translational research of the human immune system *in vivo*. Following discovery of the first immunodeficient strains (nude and SCID mice), various series of new strains have been generated [20]. Fig. 1 presents a simplified overview of the development of various immonodeficient mice for humanized mouse models. A remarkable improvement was obtained by introducing the mutated IL-2 receptor gamma chain (*IL-2Ry*) into the parent NOD/SCID and RAG1/2$^{-/-}$ immunodeficient mouse strains [28,29]. These strains feature multiple immunodeficiencies, including defects in T-, B- and natural killer cells, reduced macrophage and dendritic cell function. After transplantation of human hematopoietic stem cells (HSCs), these mice develop well-differentiated multi-lineage hematopoietic cells. For instance, T-cell subpopulations including CD4^{+} and CD8^{+} cells develop in these mice, while they are not differentiated in the earlier generation NOD/SCID immunodeficient mice. Further improvement in recapitulating the *bona fide* human immune system was achieved by introducing human genes for various cytokines (e.g. IL-2, IL-4, IL-6, GM-CSF) or HLA class I and II [30,31]. The engrafted functional human immune systems are capable of T- and B-cell dependent immune responses, antibody production, anti-viral responses, and allograft rejection. Applications of these mouse models with a reconstituted human immune system are various, and include the investigation of cancer immunotherapy, regenerative medicine, human stem cell transplantation and vaccines. For instance, HIV-infected humanized mice are used to evaluate new therapies regulating chronic immune activation and replication of HIV [32]. Another example of the use of humanized mouse models is the preclinical cancer research with novel drug candidates such as chimeric antigen receptors (CARs) [33].

6. Introducing clinical trial features

The translational gap between animal models and clinical trials can be bridged by making animal model testing more clinical trial like.

6.1. Important clinical endpoints are not established or not assessable in animal models

Quality of life is an important endpoint in clinical trials for a variety of different chronic diseases, mostly to support reimbursement. This clinical evaluation is most often assessed by questionnaires, a methodology that obviously cannot be used in experimental animals. However, there are recent attempts in animals to model pain questionnaires which serve as primary outcome measures in clinical trials.

Table 2
Advantages and disadvantages of PDX compared to other preclinical oncology models. Adapted from [21,41,42].

Model	Advantages	Disadvantages
Syngeneic mouse tumor models, including genetically engineered mice (GEMs)	– Possibility to model tumor development and premalignant neoplastic stages – Tumor heterogeneity exists with multiple lesions in one mouse – Tumor microenvironment (stroma and immune system) is representative of the studied tumor – Studies on defined mutations possible, including the analysis of the effects of these mutations in many genetic backgrounds	– Target tissue expression pattern can be different between mouse and human – Tumor and microenvironment are both murine and human/mouse cross-reactive compounds are required – Tumor development in animals slow and variable – Development costly and time consuming – Limited number of genes can be engineered
Human tumor cell derived xenograft on immunodeficient mice	– Allows a rapid analysis of response to a therapeutic regimen – Source of material virtually unlimited for immortal cell lines	– Immunodeficient mice cannot adequately capture the intact human immune component – Human tumor microenvironment is not represented – Orthotopic implant is often technically complicated – *In vitro* passages induce artificial drift leading to outgrowth of cells with different characteristics than primary tumors – Poor predictive value
Patient-derived xenograft (PDX) on immunodeficient mice	– No evolutionary selection pressure from *in vitro* culture assuring a realistic reflection of original tumor heterogeneity – Stromal component is representative of the parental tumor in the initial passages – One tumor model = one patient – Offer the opportunity to evaluate tumors from metastatic sites or tumors that have developed resistance to multiple treatments – Studies have shown very good correlation between response in PDX models and clinical response in patients	– Immunodeficient mice cannot adequately capture the intact human immune component – Tumor development in animals is slow (tumor graft latency from 2 to 12 months) – Low engraftment rates for some tumor types (varying from 23–75%) and only a limited source of original material – Orthotopic implant is often technically complicated – Expensive and labor intensive to establish and maintain PDX bank

For instance, the human brief pain inventory was successfully modeled in dogs suffering from cancer pain [34].

6.2. Animal models use other endpoints than the preferred endpoints in clinical trials

Disease models are often designed to give a readout that is easily measureable, objective and gives a large window between healthy and diseased subjects. While this makes the experimental design of the animal study more robust and reproducible, this may not always be the endpoints requested from regulatory authorities for pivotal clinical trials. In preclinical tumor models for instance the tumor size is often the primary endpoint, while overall survival is commonly used in oncology clinical trials. The latter does not necessarily need to be directly dependent upon tumor mass. For ethical reasons, survival is rarely an acceptable endpoint in experimental animals. Progression free survival which often is a secondary endpoint in clinical studies could, however, be included in preclinical models.

6.3. Predictive biomarker discovery in animal models

An integrated preclinical approach using PDX models together with systems biology can for instance enable the discovery and development of predictive biomarkers for classifying clinical tumor responsiveness to novel agents [22,23]. More specifically, a novel targeted therapy can be screened in a cohort of tumor-specific PDX models to determine efficacy, using a sensitive (regression) *versus* resistant (progression) classification system. Multiple layers of 'omics' technologies (e.g. genome sequencing, transcriptome profiling, proteomics, metabolomics) can then be employed to characterize these PDX models to derive an "integrative classifier" to predict sensitive and resistant models in response to the drug. If the assay development seems feasible, this classifier or predictive biomarker could be submitted to preclinical validation by correct prediction of efficacy in an independent cohort of PDX models. If the classifier achieved a high level of accuracy in this experiment, biomarker-driven clinical trials could be designed based on the prevalence of the identified biomarkers and their link with efficacy.

6.4. Back translation from clinical trials to animal models

Clinical success also represents lessons learnt for preclinical research. Hence, it is important to feed back the clinical data to further improve and fine tune the predictive value of the next generation of drugs as guided by biomarkers. Proteomics are emerging biomarkers in preclinical and clinical research, a process in which the translation and back translation is crucial in order to improve the patient outcomes [35]. Alternatively, in case of clinical failure, back translational research in animal models might support the understanding of the clinical data and mode of action of drug.

7. Future perspectives

During the last decades, investment in development of new technologies and refinement of existing technologies has been considerably higher in areas such as molecular biology or clinical trial biomarkers than it has been in development of more predictive animal models. Exceptions are the areas outlined above e.g. the development of humanized experimental animals. As a consequence, animal modeling is more and more debated and few improvements have been implemented during the last two decades. Improving the quality of animal models involves three

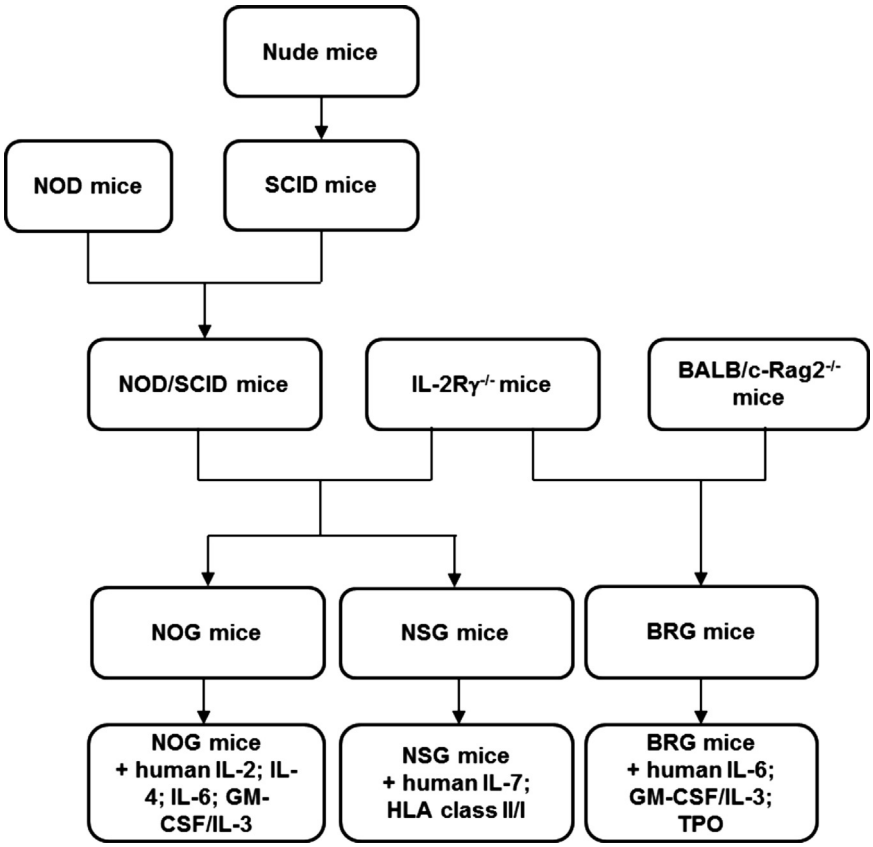

Fig. 1. A simplified schematic overview of the historical development of various immonodeficient mice for humanized mouse models. Adapted from [20].

things: improving the performance quality of existing models (e.g. proper design and execution of experiments), improving the way in which animal models are used in the decision making process and investing in the development of more sophisticated and clinical relevant and predictive models.

Several alternatives for animal model testing have been proposed. One of those potential alternatives is the use of experimental clinical trials, e.g. in which disease is modeled in healthy subjects, which allow the early clinical testing of drug candidates without prior extensive animal model testing. Following a number of failures in clinical development of analgesic drugs, experimental clinical pain protocols have been developed for providing early clinical proof of concept within the boundaries of a phase I clinical trial design [36,37]. While these human models have helped to bridge the translational gap, they constitute rather an addition in the translational research armamentarium than a substitute for well-designed animal models. Their major advantage (i.e. involving human subjects) should be balanced with the often simplistic design of experimental clinical trials. A more recent approach is in silico modeling or also called quantitative systems pharmacology [38–40]. By modeling disease mechanisms, the efficacy of drugs addressing known or novel targets on clinical endpoints and biomarkers is predicted. Ideally, quantitative systems pharmacology is combined with disease models in a way that certain specific hypothesis are generated that can subsequently be assessed in experimental animals and finally be fed back into the in silico model to potentially refine the hypothesis. Thus these approaches are rather an addition to than a substitution of animal models.

In summary, animal models when carefully selected, designed and conducted are an important part of any translational drug development strategy. Their translational value can be further enhanced when combined with other translational tools such as quantitative systems pharmacology, biomarkers or experimental clinical trials.

Executive summary

- Animal models are essential to bridge the translational gap between preclinical and clinical research.
- More appropriate preclinical testing in fit-for-purpose animal models might result in increased clinical success rates for drugs in development.
- Fit-for-purpose validation of an animal model takes into account the specific objectives to be addressed by the model.
- Humanization of animal models has helped to improve the clinical translation, especially in oncology and inflammatory diseases.
- The translational value of animal experiments can be enhanced by the inclusion of safety parameters and more clinically relevant endpoints.
- Design of a translational strategy including systems pharmacology and exploratory clinical trials is preferred over the focus on isolated animal experiments.

References

[1] S.C. Williams, Small nanobody drugs win big backing from pharma, Nat. Med. 19 (11) (2013) 1355–1356.
[2] J. Arrowsmith, Trial watch: phase II failures: 2008–2010, Nat. Rev. Drug Discov. 10 (5) (2011) 328–329.
[3] J. Arrowsmith, Trial watch: phase III and submission failures: 2007–2010, Nat. Rev. Drug. Discov. 10 (2) (2011) 87.
[4] M.N. Bouchlaka, G.D. Sckisel, M. Chen, A. Mirsoian, A.E. Zamora, E. Maverakis, D.E. Wilkins, K.L. Alderson, H.H. Hsiao, J.M. Weiss, et al., Aging predisposes to acute inflammatory induced pathology after tumor immunotherapy, J. Exp. Med. 210 (11) (2013) 2223–2237.
[5] S. Hansen, R.G. Leslie, TGN1412: scrutinizing preclinical trials of antibody-based medicines, Nature 441 (7091) (2006) 282.
[6] R. Stebbings, L. Findlay, C. Edwards, D. Eastwood, C. Bird, D. North, Y. Mistry, P. Dilger, E. Liefooghe, I. Cludts, et al., "Cytokine storm" in the phase I trial of monoclonal antibody TGN1412: better understanding the causes to improve preclinical testing of immunotherapeutics, J. Immunol. 179 (5) (2007) 3325–3331.
[7] H. Attarwala, TGN1412: from discovery to disaster, J. Young Pharm. 2 (3) (2010) 332–336.
[8] J.C. Crabbe, D. Wahlsten, B.C. Dudek, Genetics of mouse behavior: interactions with laboratory environment, Science 284 (5420) (1999) 1670–1672.
[9] S.H. Richter, J.P. Garner, B. Zipser, L. Lewejohann, N. Sachser, C. Touma, B. Schindler, S. Chourbaji, C. Brandwein, P. Gass, et al., Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study, PLoS One 6 (1) (2011) e16461.
[10] D. Wahlsten, P. Metten, T.J. Phillips, S.L. Boehm 2nd, S. Burkhart-Kasch, J. Dorow, S. Doerksen, C. Downing, J. Fogarty, K. Rodd-Henricks, et al., Different data from different labs: lessons from studies of gene–environment interaction, J. Neurobiol. 54 (1) (2003) 283–311.
[11] C.R. Hooijmans, R. de Vries, M. Leenaars, J. Curfs, M. Ritskes-Hoitinga, Improving planning, design, reporting and scientific quality of animal experiments by using the Gold Standard Publication Checklist, in addition to the ARRIVE guidelines, Br. J. Pharmacol. 162 (6) (2011) 1259–1260.
[12] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, PLoS Biol. 8 (6) (2010) e1000412.
[13] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, Osteoarthr. Cartil. 20 (4) (2012) 256–260.
[14] A.S. Rice, D. Cimino-Brown, J.C. Eisenach, V.K. Kontinen, M.L. Lacroix-Fralish, I. Machin, J.S. Mogil, T. Stohr, Animal models and the prediction of efficacy in clinical trials of analgesic drugs: a critical appraisal and call for uniform reporting standards, Pain 139 (2) (2008) 243–247.
[15] Y. Lamberty, D.G. Margineanu, H. Klitgaard, Absence of negative impact of levetiracetam on cognitive function and memory in normal and amygdala-kindled rats, Epilepsy Behav. 1 (5) (2000) 333–342.
[16] C. Weyer, J.F. Gautier, E. Danforth Jr, Development of beta 3-adrenoceptor agonists for the treatment of obesity and diabetes – an update, Diabetes Metab. 25 (1) (1999) 11–21.
[17] F. Sams-Dodd, Strategies to optimize the validity of disease models in the drug discovery process, Drug Discov. Today 11 (7–8) (2006) 355–363.
[18] R.J. Derscheid, M.R. Ackermann, Perinatal lamb model of respiratory syncytial virus (RSV) infection, Viruses 4 (10) (2012) 2359–2378.
[19] R. Villenave, S. Thavagnanam, S. Sarlang, J. Parker, I. Douglas, G. Skibinski, L. G. Heaney, J.P. McKaigue, P.V. Coyle, M.D. Shields, et al., In vitro modeling of respiratory syncytial virus infection of pediatric bronchial epithelium, the primary target of infection in vivo, Proc. Natl. Acad. Sci. USA 109 (13) (2012) 5040–5045.
[20] R. Ito, T. Takahashi, I. Katano, M. Ito, Current advances in humanized mouse models, Cell. Mol. Immunol. 9 (3) (2012) 208–214.
[21] D. Siolas, G.J. Hannon, Patient-derived tumor xenografts: transforming clinical samples into mouse models, Cancer Res. 73 (17) (2013) 5315–5319.
[22] T.M. Pitts, A.C. Tan, G.N. Kulikowski, J.J. Tentler, A.M. Brown, S.A. Flanigan, S. Leong, C.D. Coldren, F.R. Hirsch, M. Varella-Garcia, et al., Development of an integrated genomic classifier for a novel agent in colorectal cancer: approach to individualized therapy in early development, Clin. Cancer Res. 16 (12) (2010) 3193–3204.
[23] J.J. Tentler, A.C. Tan, C.D. Weekes, A. Jimeno, S. Leong, T.M. Pitts, J.J. Arcaroli, W.A. Messersmith, S.G. Eckhardt, Patient-derived tumour xenografts as models for oncology drug development, Nat. Rev. Clin. Oncol. 9 (6) (2012) 338–350.
[24] H.H. Fiebig, A. Maier, A.M. Burger, Clonogenic assay with established human tumour xenografts: correlation of in vitro to in vivo activity as a basis for anticancer drug discovery, Eur. J. Cancer 40 (6) (2004) 802–820.
[25] R.S. Kerbel, Human tumor xenografts as predictive preclinical models for anticancer drug activity in humans: better than commonly perceived-but they can be improved, Cancer Biol. Ther. 2 (4 Suppl. 1) (2003) S134–S139.
[26] M. Hidalgo, E. Bruckheimer, N.V. Rajeshkumar, I. Garrido-Laguna, E. De Oliveira, B. Rubio-Viqueira, S. Strawn, M.J. Wick, J. Martell, D. Sidransky, A pilot clinical study of treatment guided by personalized tumorgrafts in patients with advanced cancer, Mol. Cancer Ther. 10 (8) (2011) 1311–1316.
[27] S.J. Weroha, M.A. Becker, S. Enderica-Gonzalez, S.C. Harrington, A.L. Oberg, M. J. Maurer, S.E. Perkins, M. AlHilli, K.A. Butler, S. McKinstry, et al., Tumorgrafts as in vivo surrogates for women with ovarian cancer, Clin. Cancer Res. 20 (5) (2014) 1288–1297.
[28] L.D. Shultz, M.A. Brehm, J.V. Garcia-Martinez, D.L. Greiner, Humanized mice for immune system investigation: progress, promise and challenges, Nat. Rev. Immunol. 12 (11) (2012) 786–798.
[29] Q. Zhou, J. Facciponte, M. Jin, Q. Shen, Q. Lin, Humanized NOD-SCID IL2rg(-/-) mice as a preclinical model for cancer research and its potential use for individualized cancer therapies, Cancer Lett. 344 (1) (2014) 13–19.
[30] R. Danner, S.N. Chaudhari, J. Rosenberger, J. Surls, T.L. Richie, T.D. Brumeanu, S. Casares, Expression of HLA class II molecules in humanized NOD.Rag1KO. IL2RgcKO mice is critical for development and function of human T and B cells, PLoS One 6 (5) (2011) e19826.

[31] L.D. Shultz, Y. Saito, Y. Najima, S. Tanaka, T. Ochi, M. Tomizawa, T. Doi, A. Sone, N. Suzuki, H. Fujiwara, et al., Generation of functional human T-cell subsets with HLA-restricted immune responses in HLA class I expressing NOD/SCID/IL2r gamma(null) humanized mice, Proc. Natl. Acad. Sci. USA 107 (29) (2010) 13022–13027.
[32] M. Singh, P. Singh, D. Vaira, M. Amand, S. Rahmouni, M. Moutschen, Minocycline attenuates HIV-1 infection and suppresses chronic immune activation in humanized NOD/LtsZ-scidIL-2Rgamma mice, Immunology 142 (4) (2014) 562–572.
[33] D. Abate-Daga, K.H. Lagisetty, E. Tran, Z. Zheng, L. Gattinoni, Z. Yu, W.R. Burns, A.M. Miermont, Y. Teper, U. Rudloff, et al., A novel chimeric antigen receptor against prostate stem cell antigen mediates tumor destruction in a humanized mouse model of pancreatic cancer, Hum. Gene Ther. (2014), 10.1089/hum.2013.209, in press.
[34] D.C. Brown, R. Boston, J.C. Coyne, J.T. Farrar, A novel approach to the use of animals in studies of pain: validation of the canine brief pain inventory in canine bone cancer, Pain Med. 10 (1) (2009) 133–142.
[35] P.C. Guest, M.G. Gottschalk, S. Bahn, Proteomics: improving biomarker translation to modern medicine? Genome Med. 5 (2) (2013) 17.
[36] L. Arendt-Nielsen, M. Curatolo, A. Drewes, Human experimental pain models in drug development: translational pain research, Curr. Opin. Investig. Drugs 8 (1) (2007) 41–53.
[37] A.E. Olesen, T. Andresen, C. Staahl, A.M. Drewes, Human experimental pain models for assessing the therapeutic efficacy of analgesic drugs, Pharmacol. Rev. 64 (3) (2012) 722–779.
[38] G. Maxwell, C. Mackay, Application of a systems biology approach to skin allergy risk assessment, Altern. Lab. Anim. 36 (5) (2008) 521–556.
[39] B.J. Schmidt, J.A. Papin, C.J. Musante, Mechanistic systems modeling to guide drug discovery and development, Drug Discov. Today 18 (3–4) (2013) 116–127.
[40] H. Geerts, A. Spiros, P. Roberts, R. Carr, Quantitative systems pharmacology as an extension of PK/PD modeling in CNS research and development, J. Pharmacokinet. Pharmacodyn. 40 (3) (2013) 257–265.
[41] P.A. Konstantinopoulos, U.A. Matulonis, Current status and evolution of preclinical drug development models of epithelial ovarian cancer, Front. Oncol. 3 (2013) 296.
[42] M. Moro, G. Bertolini, M. Tortoreto, U. Pastorino, G. Sozzi, L. Roz, Patient-derived xenografts of non small cell lung cancer: resurgence of an old model for investigation of modern concepts of tailored therapy and cancer stem cells, J. Biomed. Biotechnol. 2012 (2012) 568567.