How Qualification of 3D Disease Models Cuts the Gordian Knot in Preclinical Drug Development

Monika Schäfer-Korting and Christian Zoschke

Contents

M. Schäfer-Korting (✉) · C. Zoschke
Freie Universität Berlin, Institute of Pharmacy (Pharmacology and Toxicology), Berlin, Germany
e-mail: monika.schaefer-korting@fu-berlin.de; christian.zoschke@fu-berlin.de

Abstract

Preclinical research struggles with its predictive power for drug effects in patients. The clinical success of preclinically approved drug candidates ranges between 3% and 33%. Regardless of the approach, novel disease models and test methods need to prove their relevance and reliability for predicting drug effects in patients, which is usually achieved by method validation. Nevertheless, validating all models appears unrealistic due to the variety of diseases. Thus, novel concepts are needed to increase the quality of preclinical research.

Herein, we introduce qualification as a minimal standard to establish the relevance of preclinical models and test methods. Qualification starts with prioritizing and translating scientific requirements into technical parameters by quality function deployment. Qualified models use authenticated cells, which resemble the corresponding cells in humans in morphology and drug target expression. Moreover, disease models differ from normal models in the expression of relevant biomarkers. As a result, qualified test methods can discriminate effects of treatment standards and the effects of weakly effective or ineffective substances. Observer-blind readout, adequate data documentation, dropout inclusion, and a priori power studies are as crucial as realistic dosage regimens for qualified approaches. Here, we showcase the implementation of qualification. Adjusting the level of model complexity and qualification to three defined phases of preclinical research assures the optimal level of certainty at each step.

In conclusion, qualification strengthens the researchers' impact by defining basic requirements that novel approaches must fulfill while still allowing for scientific creativity. Qualification helps to improve the predictive power of preclinical research. Applied to human cell-based models, qualification reduces animal testing, since only effective drug candidates are subjected to final animal testing and subsequently to clinical trials.

Keywords

Aging · Analytical methods · Data quality · Neoplasms · Nonalcoholic fatty liver disease · Pharmacology · Preclinical drug evaluation · Validation study

Abbreviations

3R	Replacement, reduction, and refinement of animal experiments
ADME	Absorption, distribution, metabolism, and excretion
AUC	Area under the curve
CRISPR	Clustered regularly interspaced short palindromic repeats
DXM	Dexamethasone
EU	European Union
FDA	U.S. Food and Drug Administration
HOQ	House of quality

ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
iPS	Induced pluripotent stem cells
LC	Liquid chromatography
MS/MS	Tandem mass spectrometry
N/TERT	Human keratinocyte cell line
NASH	Nonalcoholic steatohepatitis
OECD	Organisation for Economic Co-operation and Development
QFD	Quality function deployment
RhS	Reconstructed human skin
STXM	Scanning transmission X-ray microscopy

1   Current Efficiency in Preclinical Research

Despite major efforts in 3R from academia, industry, and legislation, the number of animals being sacrificed in research every year still amounts to almost 9.4 and 1.8 million in the EU and Germany, respectively. Fundamental and applied research account for 44% and 15%, respectively, of the animals in Germany (European Commission 2019; Federal Ministry of Food and Agriculture 2019). Concomitantly, preclinical testing often fails to collect data, being relevant for human patients. Success rates in clinical trials are as low as 3% in oncology or 15% in neurology and immunology and question the current methodology of assessing investigational new drugs in preclinical science. The major shortcoming of preclinical research is related to the complex architecture of organs like the brain or the immune system as well as the heterogeneous nature of diseases. The translation from bench to bedside is only sightly more successful for vaccines and drugs against cardiovascular and infectious diseases (Table 1, Wong et al. 2019). Poor efficacy and safety stand out as major reasons for the termination of drug development projects (Arrowsmith and

Table 1 Disease-group-related success rates (%) in clinical Phase 1 and overall (success in Phases 2 and 3) in clinical drug development; from Wong et al. (2019)

Disease groups	Clinical Phase 1 passed transfer to Phase 2 (%)	Overall success rate (%)
Oncology	57.6	3.4
Central nervous system	73.2	15.0
Autoimmune/inflammation	68.8	15.1
Metabolic/endocrinology	76.2	19.6
Genitourinary	68.7	21.6
Infectious disease	70.1	25.2
Cardiovascular	73.3	25.5
Vaccines	76.8	33.4
Ophthalmology	87.1	32.6

Miller 2013), besides economic considerations of the pharmaceutical company (Waring et al. 2015).

Despite the marked reduction of drug candidates in early and advanced preclinical testing, still too many substances pass the preclinical phase but fail in clinical studies. High numbers of volunteers and patients exposed to non-efficacious or unsafe drugs demand a stricter preclinical selection. Recent developments in tissue engineering should enable addressing these questions with human cell-based disease models and rejecting unqualified drug candidates. Rethinking preclinical drug development will avoid expendable applications to human and animal test subjects and should reduce costs, from now $2{,}800 \times 10^6$ US dollars per year for preclinical research that is not reproducible (DiMasi et al. 2016).

1.1   Phases of Preclinical Research

Preclinical drug research comprises all tests from drug discovery to the first-in-human studies. The current approaches encompass in silico methods and high-throughput screenings, tests in disease models and pharmacokinetic investigations in vitro and in vivo, as well as regulatory toxicology and safety pharmacology studies. Preclinical research starts with simple models to save time and costs, while sophisticated approaches are used in later phases. This stepwise approach could be grouped into three phases. We suggest summarizing in silico methods and high-throughput screening in *Phase I*; simple pharmacological tests, regulatory toxicology, and safety pharmacology in *Phase II*; and sophisticated pharmacokinetic and pharmacodynamic tests in *Phase III*. Currently, animal experiments are predominantly used in *Phases II* and *III* (according to our definition) and still remain the backbone for preclinical drug research.

1.2   Models and Test Methods

Here, we define a preclinical model (normal, disease) as a system recapitulating the hallmarks of the human tissue in animal or cell culture. A test method is an approach to identify drug effects in the preclinical model, respectively. An efficient selection of drugs, suitable for human use, requires a tiered procedure with models and test methods that are as simple as possible but as complex as necessary. This means an increasing complexity from *Phase I* to *III*.

2   Reasons for Poor Translational Success

Five reasons stand out from the causes which limit the translational success of investigative new drugs from bench to bedside:

- *Animal models* are confounded by a major gap between animal and human biology (Seok et al. 2013; Warren et al. 2015). The animal-based disease frequently aligns poorly with the human indication of interest. New technologies like CRISPR/Cas offer new opportunities for more human-like disease models, but transgenic mice and rats remain genetically engineered rodents, except for the drug target.
- *Heterogeneity* is excluded, since young male animals from single inbred strains are preferred (Hartung 2013). In contrast, diseases affect male and female, young, adult, and senior patients with different genetic backgrounds.
- Currently, *cell culture practice* faces limitations with cells subcultivated in high passages or not authenticated (Hartung 2007) and generally the lack of quality control (OECD 2018).
- The use of *cell lines* can be unrepresentative of complex diseases. Moreover, monolayer cultures lack the tissue-specific extracellular matrix (Nallanthighal et al. 2019).
- *Study design* in preclinical research frequently does not apply to the same standard as in clinical trials. In particular, randomization and blinding (van Luijk et al. 2014) as well as statistical tests for differences are rarely considered.
- *Quality assurance and validation* for both the model itself and the pharmacological test method appear to be expendable with respect to disease models, although preclinical studies lack reproducibility (Begley et al. 2015; Simeon-Dubach et al. 2016).

Now the time has come to transform preclinical drug development into relevant and reproducible research, while avoiding suffering animals wherever possible (German Research Foundation 2019). Nevertheless, even less stressful testing in the animal will not overcome the genetic differences between animals and humans.

Toxicologists have addressed this issue by the development and validation of alternatives to animal testing (Leontaridou et al. 2017). Validated toxicological methods use reconstructed human epidermis for the evaluation of skin corrosion (OECD 2019a), skin irritation (OECD 2020), and phototoxicity (tier-2; OECD 2019b). Yet, it took about 25 years from the development of reconstructed human skin (Green et al. 1979) to regulatory acceptance of the respective test methods, and there is still a gap in fully accepting these in regulatory toxicology (Sauer et al. 2016).

Currently the so-called investigative toxicology shifts pharmaceutical toxicology from a descriptive to an evidence-based, mechanistic discipline. Outside the boundaries of *regulatory toxicology*, investigative toxicology embraces new technologies to predict human responses. European leaders in the pharmaceutical industry propose humanized in vitro test systems to improve preclinical decisions (Beilmann et al. 2019). However, toxicity studies in animals remain essential for *regulatory toxicology* because of the limited number of organs which can be reconstructed and the approach to investigation of the whole organism is at its infancy. In 2018 the European Medicines Agency started a consultation on the regulatory requirements for drug development, and the pharmaceutical industry requested harmonization with the US Food and Drug Administration.

3   From Validation to Qualification

The ICH M3(R2) guideline clearly states that in vitro alternative methods can be used to replace current standard methods, if validated and accepted by all regulatory authorities.

3.1   Validation

A validation study provides the documented approval that a model or a test method reproducibly shows the desired effect. The extensive requirements make validation highly time-consuming and costly (Basketter et al. 2012) and might prevent innovative methods from their application in preclinical drug development. Moreover, the broad spectrum of drug effects and the heterogeneity of diseases increase the complexity and stand against a timeline in using in vitro disease models.

The overarching goal of validation, proof-of-concepts, performance standards, and best practice guidelines is to demonstrate the *quality* of a model or test method. According to the Latin origin of *qualitas,* quality is defined by the nature of an object. Modern perceptions of quality assume that *quality* must be produced rather than assured in retrospective validations (Kamiske and Brauer 1993). *Quality* should originate from a company- or research group-wide spirit with a clear vision of *quality*; it is an inherent responsibility which cannot be delegated or outsourced, as it is the basis of scientific and industrial success. To apply the vision of *quality* into real work, management tools such as "quality management systems" or "failure mode and effects analysis" have been developed for industrial applications and recently translated to scientific and academic use (Dirnagl et al. 2018). Even evidence-based medicine strives to improve model and test method development (Lefevre and Balice-Gordon 2019). Most often the high aims of such guidance are perceived to stand against scientific creativity, publication output, and fundraising. Consequently, the compliance to such guidelines varies among research institutions which might contribute to the overall low success of preclinical research.

Since the best strategy is useless if not applied, the two major questions to be answered are:

- How to deploy quality planning and management in the development of novel in vitro methods for preclinical research?
- Which level of certainty is required for the model and test method, respectively?

We suggest starting with compiling the scientific requirements, which a model or test method must fulfill. According to the industrial definition "*Quality* is conformance to customer requirements" (Crosby 1996), these scientific requirements should be in accordance to the user of the model or test method, from industry or academia.

3.2   Quality Function Deployment: Learning from Industry

A relevant model or test method depends on the *quality* of its planning. Researchers aiming for a high impact and short time to application should focus on structured method planning, since in industry 80% of product flaws, which are occurring during product assembly and product use, originate from insufficient product design. The car industry took major profit from introducing quality function deployment (QFD) into the product planning. Thereby, QFD reduced the costs for product development to 40% of the initial value and diminished the changes necessary to optimize the original product design (Fig. 1).

The major parts of QFD include the formation of a quality planning team and the correlation matrix "house of quality" (HOQ, Zoschke 2009). The quality planning team should consist of the leading and first-stage researchers as well as technical assistants to involve all concerned group members into the planning. The HOQ fosters a systematic assessment, categorization, and prioritization of scientific requirements and technical parameters and clearly documents the results of the quality planning team's discussion. To the best of our knowledge, QFD has not yet been introduced to the development of in vitro models. Here, we present the first example with the development of an immunocompetent model of head and neck cancer for the evaluation of local drug effects (Fig. 2). The selection of scientific requirements and technical parameters and their weighting are the author's choice to serve as an example to further develop a recently published tumor oral mucosa model (Gronbach et al. 2020). However, the definition and weighting of requirements and parameters must be adapted to the disease model or test method of interest.

Fig. 1 Impact of quality function deployment in industry. The higher efforts and costs of structured product planning with quality function deployment (QFD) in the beginning of a car development are counterbalanced by shorter development times and less flaws of the product following start of production. Transfer to in vitro model and test method development should reduce the time-to-application similarly, from Zoschke (1993)
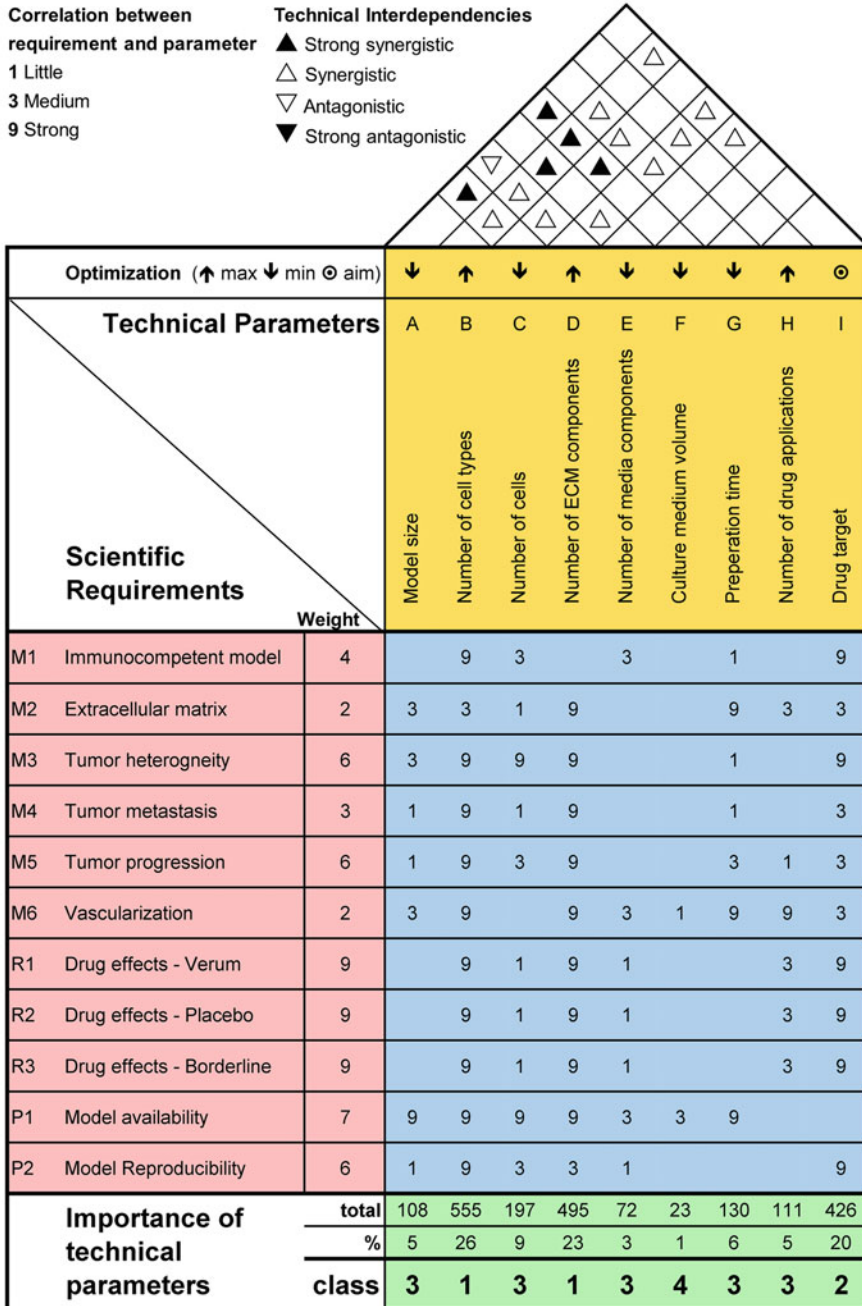
Correlation between requirement and parameter
1 Little
3 Medium
9 Strong

Technical Interdependencies
▲ Strong synergistic
△ Synergistic
▽ Antagonistic
▼ Strong antagonistic

		Optimization (↑ max ↓ min ⊙ aim)	↓	↑	↓	↑	↓	↓	↓	↑	⊙
	Technical Parameters		A	B	C	D	E	F	G	H	I
	Scientific Requirements	Weight	Model size	Number of cell types	Number of cells	Number of ECM components	Number of media components	Culture medium volume	Preperation time	Number of drug applications	Drug target
M1	Immunocompetent model	4		9	3		3		1		9
M2	Extracellular matrix	2	3	3	1	9			9	3	3
M3	Tumor heterogneity	6	3	9	9	9			1		9
M4	Tumor metastasis	3	1	9	1	9			1		3
M5	Tumor progression	6	1	9	3	9			3	1	3
M6	Vascularization	2	3	9		9	3	1	9	9	3
R1	Drug effects - Verum	9		9	1	9	1			3	9
R2	Drug effects - Placebo	9		9	1	9	1			3	9
R3	Drug effects - Borderline	9		9	1	9	1			3	9
P1	Model availability	7	9	9	9	9	3	3	9		
P2	Model Reproducibility	6	1	9	3	3	1				9
Importance of technical parameters		total	108	555	197	495	72	23	130	111	426
		%	5	26	9	23	3	1	6	5	20
		class	3	1	3	1	3	4	3	3	2

Fig. 2 House of quality (HOQ) for an immunocompetent 3D model of head and neck cancer. *red*: Scientific requirements with weighting. *yellow*: Technical parameters of the in vitro model. *blue*: Correlation between scientific requirements and technical parameters. *green*: Importance of technical parameters. *white*: Interdependencies between technical parameters, from Zoschke (1993)

First, the scientific requirements need to be listed in the left part (Fig. 2, red). Since not all requirements are equally important, the next step is to prioritize them. Therefore, a decision must be made for each requirement if it is less, equal, or more important than the other requirement. Having done this paired comparison for all requirements, a rank order of the requirements as well as a weighting factor will be obtained. Next the technical parameters need to be listed in the upper part of the HOQ (Fig. 2, yellow). Technical parameters can always be measured and quantified, which makes them very specific in contrast to scientific requirements. Next, the quality planning team determines the correlation between the technical requirements and the scientific requirements. Use an exponential scale with 1 for little correlation, 3 for medium correlation, and 9 for strong correlation; multiply the correlation by the weighting of the scientific requirements. The results are noted in the correlation matrix (Fig. 2, blue). Furthermore, the direction of optimization of each technical parameter is listed in the yellow part to better fulfill the scientific requirements. The roof of the HOQ serves to list the interdependencies between each technical parameter, since the optimization of one parameter can affect the optimization of another parameter synergistic or antagonistic (Fig. 2, white). If there is an antagonistic interference between two technical parameters, this will be a major target for innovation to overcome this antagonism. Moreover, in a synergistic interference, the deterioration of one parameter can also impair the other. The target values for each technical parameter are listed at the bottom of the correlation matrix (Fig. 2, green). The values are either minima, maxima, or a range which should be achieved for the in vitro model. The quality planning team can also assess the level of difficulty to achieve these target values. The final output of the HOQ is the importance of technical parameters. Therefore, the weighted correlations between the technical parameter and each scientific requirement are summed up. An overestimation of single parameters can be avoided by dividing the sum of one technical parameter by the sum of all technical parameters (relative relevance). Almost equally relevant technical parameters should be clustered together. The technical parameter with the highest value is most important, while the parameter with the lowest value needs to be addressed at last. After completing the HOQ, the quality planning team should check for plausibility by confirming that:

- Every scientific requirement correlates strongly to at least one technical parameter (no empty rows)
- Every technical parameter correlates strongly to at least one scientific requirement (no empty columns)
- The direction of optimization fits to the target value for each technical parameter
- Antagonistic technical interdependencies should be solved or prioritized
- The correlation matrix should be filled to at least one third to be able to prioritize technical parameters.

Additionally, the HOQ can be extended by comparisons of the approach to already existing models or test systems (Zoschke 2009). In conclusion, the HOQ cannot take the decision for the researcher, but the HOQ helps to systematically

translate vague scientific requirements into quantifiable and prioritized technical parameters. According to QFD, the outcome of the HOQ is the basis for the planning of the in vitro model or test method parts. Quality planning for the parts is the basis for planning the processes, which is finally leading to the final test method protocol. This deployment ensures the translation of the scientific requirements for the in vitro model or test method into feasible protocols for each step of the model or method development.

3.3 Qualification

We define *qualification* as minimal standard in model or test method development. *Qualification* comprises the sum of evidences that a model or test method is relevant for the disease at hand. *Qualification* is based on QFD and uses the state of the art in tissue engineering and testing in molecular medicine and pharmacology. *Qualification* does not replace validation but will provide a sufficient basis for decision-making in preclinical drug development.

3.4 Qualification of 3D In Vitro Models

Irrespective of the envisioned use in fundamental research or drug development, a qualified 3D in vitro model has to fulfill the following key features:

- Use of authenticated human cells
- Comparability between tissue morphology and function with the human disease
- Expression of drug targets in accordance to the human disease
- Concurring endpoints in models and patients, i.e., the change of biomarkers relevant for disease outcome

If relevant for the disease at hand, assessing graded drug effects should be preferred over simple yes or no assessments. The model features have to be reproducible over time and in different laboratories. *Qualification* is not limited to disease models but also applicable to models for normal tissue and the target structures used, e.g., in molecular modeling and high-throughput testing. Moreover, we suggest applying the same requirements for qualified normal models as described for qualified disease models. On the one hand, the normal models will serve as control to assess a potential *restitutio ad integrum* and on the other hand to provide insights into local adverse effects of drugs.

3.5 Qualification of Test Methods

Test procedures need to be qualified for preclinical drug research. Changes observed due to drug exposure are only signified if a suitable test protocol is used and the test

is run under quality assurance. Protocol adaptations during larger test series have to be avoided as they exclude comparisons over time. Hallmarks of a *qualified test method* in all phases include if not indicated otherwise:

- Relevant controls
  - Untreated
  - If available, an already approved standard treatment with maximum efficacy (*Phase III* only)
  - If available, an already approved treatment with minimal efficacy (*Phase III* only)
  - A treatment that showed no or insufficient efficacy in clinical trials (*Phase III* only)
- Observer-blind readout when using subjective endpoints
- Adequate data documentation, including dropouts
- Evaluation by explorative data analysis
- A priori definition of the relevant effect size and adapted sample size (power study, *Phase III*)
- Relevant dosage regimen and treatment period (*Phase III*)

Test methods can be qualified only by a range of performance standards, which are related to the respective drug targets. Thus, test methods designed for evaluating anticancer drugs are unlikely to be suitable for evaluating, e.g., antimicrobial endpoints. Yet, the transfer of a test method from one disease model to another disease model appears easier to achieve than a *qualification* from scratch. Testing of investigational new drugs with targets unrelated to the *qualification* process requires a requalification for the new performance standards.

3.6   Selection of Relevant Drug Doses

Currently, the extrapolation of drug doses from animal studies to first-in-human studies remains empirical. The most common approaches are dose-by-factor based on no observed adverse effect levels (NOAEL or benchmark dose), pharmacokinetically guided approaches, based on minimal anticipated biological effect levels, pharmacokinetic-pharmacodynamic models, similar drug approach, and data from human microdosing (Nair et al. 2018). Nevertheless, interspecies differences impede the calculation of human equivalent doses from animal data, despite of the introduction of correction factors.

A failed translation of preclinical dosage regimen into clinical treatments results in severe toxicity, prolonged dose escalation procedures or patients exposed to ineffective doses. Dose finding for anticancer drugs is in particular challenging, since they have steep dose-response curves and narrow therapeutic windows (Mathijssen et al. 2014). In vitro studies frequently use drug doses far higher than the maximum tolerated dose in cancer patients (Smith and Houghton 2013) and contribute to the highest attrition rate of anticancer drug candidates in clinical trials

(Wong et al. 2019). Moreover, ambiguous dosing in cell culture experiments due to different physical conditions, like volume of medium and number of cells used, hamper the reproducibility of in vitro experiments in different phases of preclinical research (Doskey et al. 2015).

Testing the efficacy of anticancer drugs in 3D in vitro models that recapitulate the tumor-specific extracellular matrix is crucial to emulate the drug uptake and metabolism in the tissue. The dense tumor stroma with extracellular matrix and cancer-associated fibroblasts (Mueller and Fusenig 2004; Minchinton and Tannock 2006) as well as the high interstitial fluid pressure may reduce drug uptake into the tumor despite their endothelial hyperpermeability (Saleem and Price 2008; Dewhirst and Secomb 2017). Thus, the failure of anticancer drugs in 3D models despite drug efficacy in monolayer cultures could be related to the absence of a tumor stroma (Cruz Rodriguez et al. 2019).

The determination of the drug concentration which is high enough to be active in 3D in vitro models, e.g., by automated UHPLC-MS/MS approaches (Joseph et al. 2020), might help to improve the translation of preclinical into clinical dosage regimen. Another approach uses time-dependent or maximum biomarker modulation as the matching metric, rather than a minimal threshold concentration (Spilker et al. 2017).

4 Current Strategies to Rethink Preclinical Drug Research

Our concept of qualification can be applied to various approaches in preclinical research. We highlight already existing strategies using human cell-based 3D in vitro models and novel test methods. These five strategies in preclinical research fulfill the criteria of qualification to different extents.

4.1 Strategy 1: Characterized Cell Lines

Studying the N/TERT keratinocyte cell line provides a deep insight into the MAPK/ERK pathway and revealed the impact of histone deacetylase modulation in skin diseases like psoriasis, atopic dermatitis, and cancer (Robertson et al. 2012). Inducing filaggrin knockdown in the N/TERT cell line and supplementing the Th-2 cytokine IL-31 result in a skin model with clinical signs of atopic dermatitis: fostered *Staphylococcus aureus* colonization, increased IL-8 levels, and reduced human β-defensin upregulation (van Drongelen et al. 2014a, b). Moreover, patient-derived material served to generate an iPS cell line for future use of, e.g., in vitro atopic dermatitis models in drug development (Devito et al. 2018).

In cancer research, human-based models revealed the impact of the dermal equivalent and the presence of a basement membrane on melanoma invasion (Commandeur et al. 2014). A 3D in vitro model of cutaneous squamous cell carcinoma was generated from primary human keratinocytes and fibroblasts as well as SCC-12 tumor cells, recapitulates the tumor histology, and predicts the activity of ingenol mebutate (Fig. 3). Ingenol mebutate induced abundant epidermal
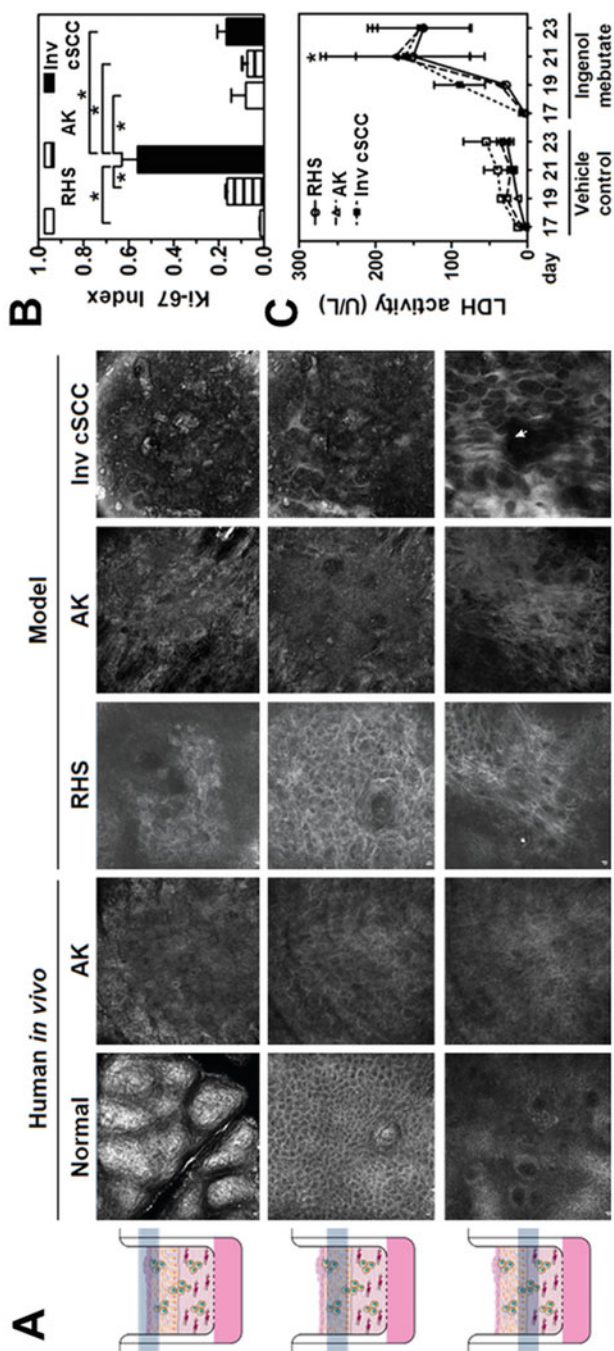
Fig. 3 3D skin cancer models and ingenol mebutate effects. (**a**) First row, stratum corneum: normal reconstructed human skin (RhS) with homogenous reflectance, actinic keratosis (AK) with beginning disruption, inhomogeneous appearance and detached cells. Second row, stratum granulosum: RhS with regular honeycomb pattern; AK with distinct epidermal foci of nuclear atypia; invasive cSCC model with dissipated honeycomb pattern. Third row, dermis: RhS with homogenous reflectance of collagen; AK model with few and invasive cSCC model with many irregular, bright collagen bundles. (**b**) RhS with slightly increased and skin cancer models with significantly decreased Ki-67 index ($p \leq 0.001$) upon ingenol mebutate treatment. (**c**) Lactate dehydrogenase (LDH) activity in the culture media of RhS and skin cancer models peaks after second ingenol mebutate treatment ($p \leq 0.05$). Graphs depict data of three batches and are presented as mean $\pm$ SEM, from Astner and Ulrich (2010) and Zoschke et al. (2016)

cell necrosis, acantholysis, and microvesicles in normal RhS (Zoschke et al. 2016). The epidermal growth factor receptor inhibitor erlotinib induced beneficial effects in another model of cutaneous squamous cell carcinoma and induced severe desquamation in the normal RhS (Commandeur et al. 2012).

This strategy is not limited to skin models but is also used for chronic kidney disease models. A human podocyte injury model of chronic kidney disease indicated that the renoprotection induced by sodium-glucose co-transporter 2 (SLGT-2) inhibitors is linked to normalized podocytes renewal and not to the lowering of blood glucose in type 2 diabetes. Correction of podocyte morphology and of associated cytoskeletal architecture renews the adhesion to the glomerulus membrane. Inhibitors of adenosine kinase reduce AMP formation and rescue cell adhesion and the actin cytoskeleton (Abraham et al. 2017).

4.2 Strategy 2: Primary Cells to Recapitulate Human Heterogeneity

An advanced 3D in vitro model of nonalcoholic steatohepatitis (NASH) was designed by co-culturing primary human hepatocytes in collagen sandwich with macrophages and stellate cells, separated by a porous transwell membrane (Fig. 4, Feaver et al. 2016). Tissue exposure to glucose, insulin, and free fatty acids corresponding to plasma levels in NASH patients repeatedly induced the lipotoxic milieu by activating key pathways spanning liver dysfunction in the hepatocytes. Triacylglycerides, diacylglycerides, cholesterol esters, and glucose levels increased significantly, and markers of inflammation (alanine amino transferase and caspase-generated cytokeratin 18, IL-6 and IL-8) as well as of fibrosis triggering TGF-β and osteopontin did so, too. Moreover, the secretion of smooth muscle α-actin increased.

Next, the model was challenged by the exposure to steady-state serum levels of obeticholic acid that targets the farnesoid X receptor in hepatocytes. The responses were compared to the vehicle control and the outcome in a clinical *Phase II* study (Hirschfield et al. 2015). Lipid accumulation declined by 25% with the most significant decrease in triacylglycerides. IL-6 and IL-8 declined significantly by 48% and 25% and other parameters of NASH including TGF-β and osteopontin were also reduced, indicating beneficial effects. Yet, intracellular cholesterol and several apolipoproteins including ApoB and ApoE increased (Feaver et al. 2016).

The interim analysis of 931 patients after 18 months treatment in clinical *Phase III* study "REGENERATE" indicated a significant improvement of key NASH factors and fibrosis by obeticholic acid 25 mg/d compared to placebo (Younossi et al. 2019). According to the positive outcome of the interim analysis, rapid drug EMA approval of obeticholic acid for NASH is applied for. The improvement of the intermediate endpoint (histology of liver biopsies) regarded as risk factors for the long-term outcome might be acceptable despite the lack of formal validation of intermediate clinical endpoints (Angulo et al. 2015). Yet already today, the *Phase III* study demonstrates the predictive power of this NASH model.
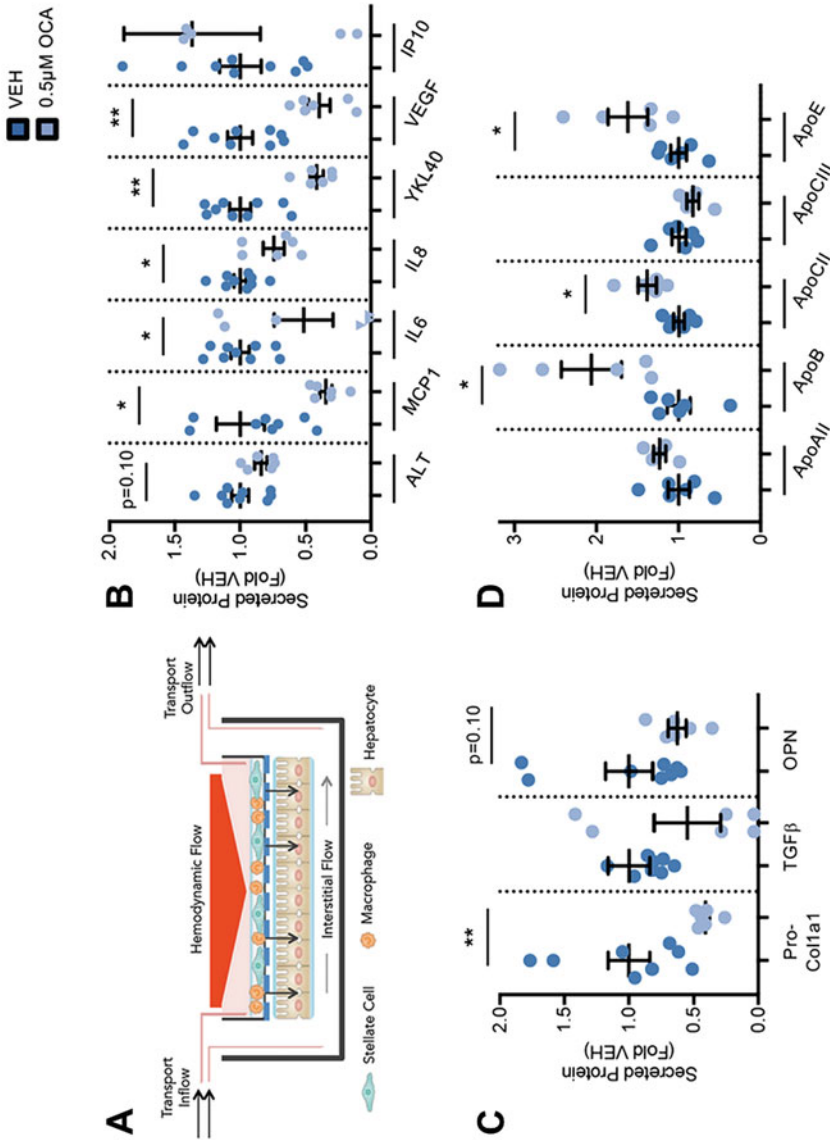
Fig. 4 NASH model used to evaluate obeticholic acid effects. (**a**) Liver sinusoidal hemodynamics were applied to the human liver system using a cone-and-plate viscometer incorporated into a transwell co-culture model of nonparenchymal cells (NPCs) (top of transwell) and hepatocytes (bottom of the transwell). Rotation of the cone (orange triangle) imparts shear stress onto the transwell. Medium is continually perfused to recapitulate interstitial flow, as indicated by the inflow and outflow ports. (**b, c**) Secreted analytes were measured in the media effluent from devices at day 10, $n \geq 5$ experiments, 3 donors. (**d**) Secreted apolipoproteins were measured in the media effluent from devices at day 10. $n = 4$ experiments, 2 donors. Triangles indicate samples that were below the lower limit of quantification. *$p < 0.05$, **$p < 0.01$, student's two-tailed t test, from Feaver et al. (2016)

Another NASH model used hepatic cells generated from human skin-derived precursors. Exposure of these cells to lipogenic (insulin, glucose, fatty acids) and pro-inflammatory factors (IL-1β, TNF-α, TGF-β) resulted in a characteristic NASH response. Elafibranor attenuated in vitro key features of NASH and significantly lowered the lipid load as well as the expression and secretion of inflammatory chemokines, being responsible for the recruitment of immune cells in vivo. This reduction in inflammatory response was mediated NFκB (Boeckmans et al. 2019). The clinical outcome, however, failed to meet the hepatic endpoint (Ratziu et al. 2016).

Treatments for end-stage liver disease, nonalcoholic liver disease in particular, are allograft liver or hepatocyte transplantation. Main obstacles are donor organ shortage and the need for efficient immunosuppression. While hepatocytes are more available than entire livers, the transplantation of hepatocytes tends to be less successful and requires more immunosuppression than the organ replacement. Allogeneic hepatocytes appear to be highly antigenic; alternatively, liver sinusoid endothelial cells or hepatic stellate cells may induce a loss of the antigenicity of hepatocytes in an allogeneic environment (Iansante et al. 2018). Immunosuppressive therapy following hepatocyte and liver transplantation include calcineurin inhibitors (cyclosporine, tacrolimus), everolimus, glucocorticoids, and basiliximab.

The suppression of immune responses was studied in a co-culture of primary human hepatocytes and allogeneic peripheral blood mononuclear cells (PBMC). Hepatocytes were isolated from six patients undergoing partial hepatectomy and grown as monolayers, while PBMC were isolated from blood of healthy donors and were added to the hepatocyte culture. Drug concentrations matched blood levels in patients receiving solid organ transplantation. Hepatocyte co-culture for 10 days strongly enhanced PBMC proliferation, and the secretion of Th-2 cell-associated cytokines strongly increased. Immunosuppressive drugs like everolimus efficiently suppressed the pro-inflammatory responses. A reduced metabolic activity of the hepatocytes, however, may indicate a potential toxicity of everolimus (Oldhafer et al. 2019). This interesting model demonstrates the immunosuppressive activity of the clinically used drugs. Given the correct identification of agents failing in the prophylaxis and therapy of allogenic rejections, the test may enable preclinical drug research on drug candidates most suitable for use in hepatocyte transplantation. The introduction of the missing innate immune system may improve predictive capacity. Moreover, these insights may allow for a pretest of the suitability of hepatocytes from donor livers for transplantation. Currently, hepatocytes are often isolated from livers unsuitable for transplantation, which appears to explain the lower success rate compared to liver transplantation (Iansante et al. 2018).

Primary cells are also essential to study the heterogeneity of aging processes and to evaluate differences in drug effects within the groups of aging. Monolayer cultures of fibroblasts from intrinsically aged human skin exhibited more signs of aging including DNA segments with chromatin alterations reinforcing senescence versus dermal fibroblasts from middle aged and young donors. Forty-three proteins confirmed the known hallmarks of aging and led to a consistent picture of eight biological categories involved in fibroblast aging, e.g., development and differentiation, cell death, and response to stress. Most of the age-associated alterations are
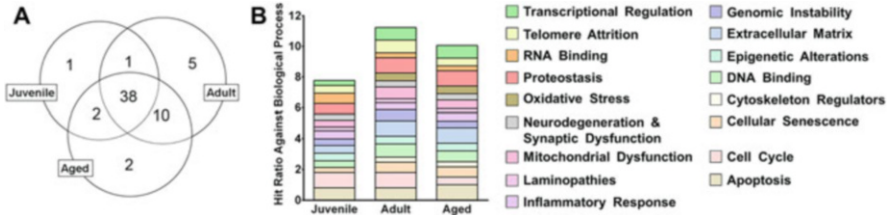
Fig. 5 Impact of normal human dermal fibroblast culture (fibroblast monolayers and reconstructed human skin) on gene expression. (**a**) Venn diagrams showing the number of genes altered due to culture conditions. (**b**) Hit ratios of the altered genes for different biological processes. The diagrams consider fold changes in gene expression > |1.3| and Ct values ≤ 35 for the 19 groups of biological processes; the maximum proportion of altered gene expression per biological process (hit ratio) = 1; from Hausmann et al. (2019)

likely caused posttranscriptionally (Waldera-Lupa et al. 2014, 2015). Next, fibroblasts from the donors aged 20–30 or 60–70 years were used to investigate the impact of age and body region on skin homeostasis, epidermal differentiation, and drug uptake on cell monolayers and reconstructed human skin. Fibroblasts from juvenile foreskin (<10 years old) served as control. 3D in vitro models containing aged fibroblasts differed from its juvenile and adult counterparts, especially in terms of the dermal extracellular matrix composition, IL-6 levels, and wound healing (Fig. 5). The region of the body from which fibroblasts are derived appears to affect the epidermal differentiation of the construct. Emulating patient heterogeneity in preclinical studies might improve the treatment of age-related skin (Hausmann et al. 2019).

4.3   Strategy 3: Patient-Derived Cells

The access to patient-derived cells is limited, and only a few subcultivations are feasible without cellular dedifferentiation. Plucking hair follicles offers a noninvasive approach for the generation of skin disease models. Only minor differences in morphology, ultrastructure, expression of important structural proteins, or barrier function are observed between normal reconstructed human skin and the in vivo counterpart generated from hair follicle-derived or interfollicular keratinocytes and fibroblasts (Löwa et al. 2018). Next, fibroblasts were isolated from plucked scalp hair follicles of six healthy volunteers and six atopic dermatitis patients. Some of the RhS with fibroblasts from atopic dermatitis patients show epidermal thickening and parakeratosis independent from filaggrin mutations. Moreover, the thymic stromal lymphopoietin and protease-activated receptor 2 are significantly upregulated in hyperproliferative RhS (Löwa et al. 2020).

For cancer research, tumor cells are used to generate patient-derived organoids in vitro and patient-derived xenografts in vivo. One of the largest collections of patient-derived material is the OncoTrack preclinical platform for colorectal cancer.

The biobank consists of 116 resected tissue samples with matched blood samples, comprising 89 primary tumors (stage I to IV) and 27 metastases from 106 patients. Organoids and xenografts are treated with drugs representing the therapeutic gold standard and experimental substances that address major pathways relevant in colorectal cancer. The OncoTrack study provides an unprecedented repository of data and models, which can be exploited further for improved drug discovery and understanding of cancer biology (Schütte et al. 2017).

4.4 Strategy 4: New Technologies in Tissue Engineering

The ongoing change in drug development will significantly increase the need for standardized tissues in high numbers. *Bioprinting* of the tissues, in particular, has the potential to enhance the delivery of the essential test platforms. For example, functional cardiac constructs can be printed. The inclusion of conductive gold nanorods improved the electrical propagation between adjacent cardiomyocytes (Zhu et al. 2017). Inter alia, bioprinted cardiac tissue reflects the activity of β-adrenoceptor and m-receptor antagonist as well as the reversibility of the effects after removal as reviewed recently (Lind et al. 2017). Bioprinting allows for the generation of models closer to the human morphology and the control of culture environment. Injecting the cell suspension into a micromold can ensure cell cluster growth sufficient nutrient supply to avoid cell death and the formation of blood vessel (Huh et al. 2013; Prabhakarpandian et al. 2013; Hagiwara and Koh 2020).

Transforming the human-on-the-chip technology to the *patient-on-the-chip* by the use of miniaturized disease models is ahead of us. For example, a cancer chip has been developed for drug testing in a vascularized tumor model (Nashimoto et al. 2020). Tissue banks, providing vital tissues and replicable cells of defined quality over years (Palechor-Ceron et al. 2019), should allow the inclusion of human heterogeneity into *Phase III* of preclinical drug development.

The implementation of human-based testing may even open up the path to improve the therapeutic outcome of the most severe, non-acute diseases by *personalized therapy*.

4.5 Strategy 5: Comparing New Test Methods to Current Standards

The ongoing introduction of high-end analytics will allow for a much more detailed insight into pharmacokinetics and pharmacodynamics. Recently the label-free quantification of drugs at the highest local resolution of 70 ± 5 nm became possible by scanning transmission X-ray microscopy (Fig. 6a, Yamamoto et al. 2015, 2017). STXM and LC-MS/MS quantified dexamethasone equally in reconstructed human skin (Fig. 6b). Moreover, this study compared the drug penetration into reconstructed human skin with human and SKH1-mouse skin ex vivo. SKH-1 is reported to be the most human-like (Radbruch et al. 2017). The inter-model
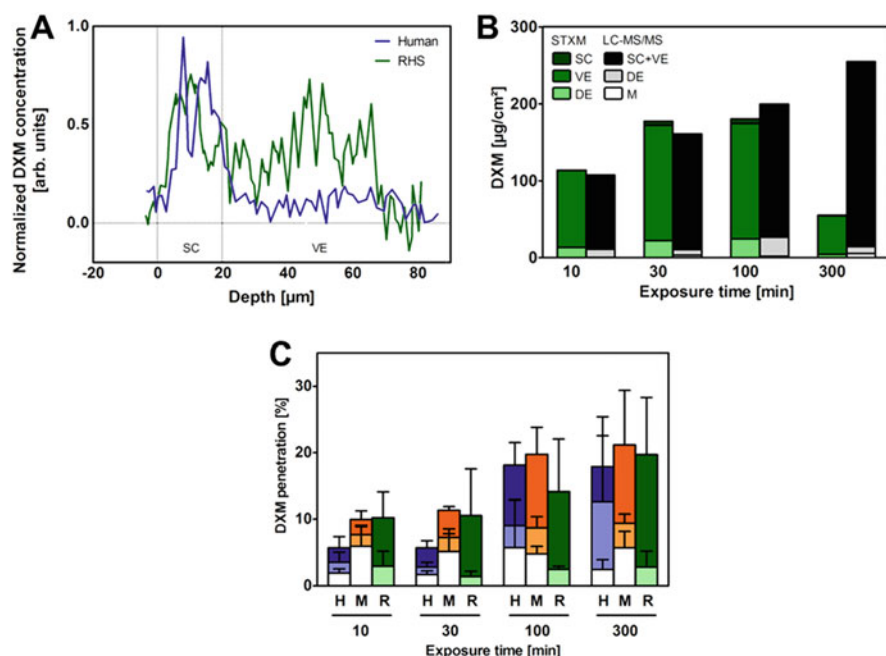
Fig. 6 Dexamethasone penetration of RhS, human, and murine skin determined by LC-MS/MS and STXM. Dexamethasone (DXM) in hydroxyethyl cellulose gel (600 µg/cm^2 DXM, 70% ethanol) was applied topically for up to 300 min. (**a**) Spatial analysis of DXM concentrations in human skin and RhS (STXM) following 10 min of exposure. Tissue surface: 0 µm. (**b**) STXM quantification shows the same skin penetration results as observed by LC-MS/MS for $t \leq 100$ min. ($n = 1$). (**c**) DXM slowly penetrates into human skin compared to murine skin and RhS. Grouped bars in order from left to right: human skin (H), murine skin (M), reconstructed human skin (R). Stacking order from top to bottom: epidermis (dark), dermis (light), heat separation water (white; human and mouse). LC-MS/MS measurements, mean ± SD, $n = 3$, from Wanjiku et al. (2019)

comparison revealed an overall similar dexamethasone uptake with minor differences in the penetration rate (Fig. 6c, Wanjiku et al. 2019).

5   Phases of Innovative Preclinical Drug Research

To keep the efforts for the qualification of novel models and test methods as low as possible, we suggest categorizing preclinical research in three phases and defined the requirements for qualified approaches accordingly (Fig. 7). Human cell-based models are preferred in all phases of preclinical testing. The combination of different in vitro models will provide higher levels of predictive power than relying on only one sophisticated model. Although human cell-based models are already revolutionizing fundamental and applied research, today, the entire organism cannot be recapitulated in vitro. Although animal tests have the clear advantage of getting an insight into systemic drug effects, risk will remain as seen in the almost fatal
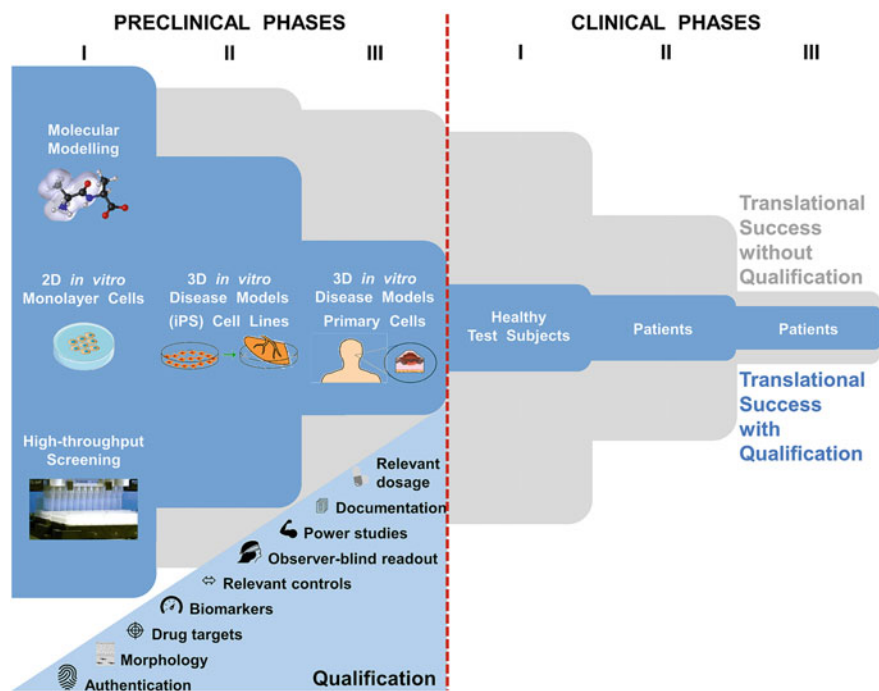
Fig. 7 Assumed impact of qualification on the predictive power of preclinical research. *iPS* induced pluripotent stem cells

cytokine release in the first-in-human study of TGN 1112 (Costello et al. 2018). To date, we consider final toxicology testing in animals prior to first-in-human studies indispensable for those drug candidates which passed preclinical efficacy tests.

5.1   Preclinical *Phase I*

In *Phase I*, physicochemical parameters of drug candidates like molecular size and hydro-/lipophilicity are the basis for in silico methods like molecular modeling, read-across, and quantitative structure-activity relationship screening. This is exemplified by the in silico identification of hit and lead structures for G-protein-coupled receptors (Wacker et al. 2017) as well as by recent breakthroughs in the treatment of HBV, HCV, and HIV as well as of cancer and severe eosinophilic asthma.

The vast knowledge of essential physicochemical features of drugs (Egner and Hillig 2008) helps to predict their bioavailability and the drugability of the pharmacological target (Zuang et al. 2018). Progress in machine learning allows calculating drug absorption, distribution, metabolism, and excretion very quickly (Tao et al. 2015).

Subsequently, high-throughput screening provides a first insight into the profile of drug effects. Moreover, biotransformation has to be assessed. The metabolism of substances into carcinogenic intermediates would complicate the therapeutic use of this drug candidate. Furthermore, the FDA has recently published a guidance document to plan and evaluate studies on drug-drug interactions (FDA 2020).

Focusing on anticancer drugs, screening can be done in authenticated well-characterized human tumor cell lines, but the genetic aberrations and epigenetic modifications will increase with the number of subcultivations. The effects of the tumor-specific extracellular matrix on tumor progression and drug efficacy cannot be captured by monolayer cultures at all. Tumor cell lines are known to be more sensitive to drug treatment than patients, as observed in the false-positive prediction of effects for more than 90% of the test agents in large surveys (Palechor-Ceron et al. 2019).

If substances appear active without severe adverse effects in *Phase I*, the drug candidate will pass to *Phase II* preclinical development.

5.2   Preclinical *Phase II*

The desired drug effects predicted by in silico approaches and high-throughput screening need to be verified in qualified models that reflect the human disease. Using the right targets, adequate biomarkers and endpoints will allow narrowing the panel of drug candidates generated in *Phase I*. Models in *Phase II* include different cell types, extracellular matrix, and tissue architecture to obtain a more precise effect profile of the drug candidate. However, the complexity of models should increase stepwise with models in *Phase II* still based on co-cultures of cell lines and/or iPS cells (Zoschke et al. 2016; Wolff et al. 2019). Another example is the investigation in the barrier function of skin models generated from the N-TERT keratinocyte cell line, which corresponds to skin models generated from primary human cells (van Drongelen et al. 2014a). Yet, the filaggrin knockdown did not alter *stratum corneum* lipids in the cell-line-based skin models (van Drongelen et al. 2013), but in primary cell-based, skin models (Vávrová et al. 2014). Slight deviations to the human patient as well as the loss of patient heterogeneity will be tolerated in *Phase II* studies to limit the numbers of repeats, necessary to observe effects with statistical significance. Nevertheless, disease models in *Phase II* need to be qualified as outlined in the qualification section.

5.3   Preclinical *Phase III*

Models in *Phase III* include different cell types, extracellular matrix, and tissue architecture to obtain a more precise effect profile of the drug candidate. Testing in *Phase III* needs to consider patient heterogeneity by using primary human cells as well as pharmacodynamics and pharmacokinetics. Models and test protocols must be qualified for predicting human responses.

The lack of sufficient amounts of patient-derived cells is increasingly addressed by the establishment of various biobanks (Simeon-Dubach et al. 2016; Palechor-Ceron et al. 2019), but cannot satisfy the needs yet. Together with potential ethical concerns in biopsy taking, e.g., in children, this limitation further supports testing only the most promising drug candidates in on primary cell-based models in *Phase III*. Current approaches to retransform iPS cells of various donors might facilitate the use of patient-derived material, but the potential dedifferentiation as well as the fact that all these cultures are juvenile (embryonic) tissues already showed some limitations in aging research (Christensen et al. 2018).

The use of flow-through chambers in organ-on-a-chip cultures continuously supplies fresh medium, removes waste, and induces sheer stress related to the blood flow (Prantil-Baun et al. 2018). These dynamic culture conditions increase the cultivation times to 28 days, potentially useful for evaluating the efficacy and toxicity of several treatment cycles. Moreover, microfluidic platforms are suitable to investigate cancer metastasis (Lin et al. 2020) as well as hematopoietic stem cells (Sieber et al. 2018). The human-on-the-chip technology can connect several tissue chambers to an in vivo organ system (Maschmeyer et al. 2015).

The major challenge in preclinical drug development will be the transition to the *patient-on-the-chip*. Beyond efficacy testing, the consideration of the higher vulnerability of patients can provide a more relevant toxicological risk analysis, currently lacking in preclinical toxicity testing (Menshykau 2017). Finally, the drug candidates need to pass the standard tests of regulatory toxicology and safety pharmacology.

In particular, *Phase III* studies must be conducted in accordance to clinical trial protocols: blinding, randomization, and proper controls. Dosage should consider both the effective concentrations as derived from *Phase I and II* studies as well as pharmacokinetic calculations of *Phase I*.

6   The Price of Quality

The implementation of qualification and quality function deployment into model or test method development involves higher costs in the beginning and might slow down the time to the first publication. Moreover, these concepts require the availability of clinical data. Taking the development of models for skin aging as an example, reliable clinical data are scarce (Hausmann et al. 2020). Preclinical models for evaluating drug effects can hardly be better than the clinical knowledge of the disease. Once human cell-based in vitro models have been qualified, their modular design offers the opportunity to manipulate single parameters to better understand the underlying mechanisms of the disease. Despite several strategies in nonclinical research that already use parts of qualification, the final proof-of-concept is still to be awaited. Therefore, we suggest developing and qualifying a disease model first and perform a full validation subsequently. The best proof of our concept will be an increased success rate of investigative new drugs in clinical trials that have been evaluated before by qualified models in preclinical drug development.

The time of wasting costs and time for poorly predictive models and test methods in preclinical research should come to an end. The increased efforts in model development will pay off, since publishing "just another disease model" will cost more time and money than developing a model and test method that fulfill and have a real impact on preclinical research.

References

Abraham VC, Miller LN, Pratt SD, Putman B, Kim L, Gopalakrishnan SM, King A (2017) Implementation of a human podocyte injury model of chronic kidney disease for profiling of renoprotective compounds. Eur J Pharmacol 815:219–232

Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, Mills PR, Keach JC, Lafferty HD, Stahler A, Haflidadottir S, Bendtsen F (2015) Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 149:389–397.e310

Arrowsmith J, Miller P (2013) Trial watch: phase ii and phase iii attrition rates 2011-2012. Nat Rev Drug Discov 12:569–569

Astner S, Ulrich M (2010) Confocal laser scanning microscopy. Hautarzt 61:421–428

Basketter DA, Clewell H, Kimber I, Rossi A, Blaauboer B, Burrier R, Daneshian M, Eskes C, Goldberg A, Hasiwa N, Hoffmann S, Jaworska J, Knudsen TB, Landsiedel R, Leist M, Locke P, Maxwell G, McKim J, McVey EA, Ouedraogo G, Patlewicz G, Pelkonen O, Roggen E, Rovida C, Ruhdel I, Schwarz M, Schepky A, Schoeters G, Skinner N, Trentz K, Turner M, Vanparys P, Yager J, Zurlo J, Hartung T (2012) A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. ALTEX 29:3–91

Begley CG, Buchan AM, Dirnagl U (2015) Robust research: institutions must do their part for reproducibility. Nature 525:25–27

Beilmann M, Boonen H, Czich A, Dear G, Hewitt P, Mow T, Newham P, Oinonen T, Pognan F, Roth A, Valentin JP, Van Goethem F, Weaver RJ, Birk B, Boyer S, Caloni F, Chen AE, Corvi R, MTD C, Daneshian M, Ewart LC, Fitzgerald RE, Hamilton GA, Hartung T, Kangas JD, Kramer NI, Leist M, Marx U, Polak S, Rovida C, Testai E, Van der Water B, Vulto P, Steger-Hartmann T (2019) Optimizing drug discovery by investigative toxicology: current and future trends. ALTEX 36:289–313

Boeckmans J, Buyl K, Natale A, Vandenbempt V, Branson S, De Boe V, Rogiers V, De Kock J, Rodrigues RM, Vanhaecke T (2019) Elafibranor restricts lipogenic and inflammatory responses in a human skin stem cell-derived model of NASH. Pharmacol Res 144:377–390

Christensen K, Roudnicky F, Patsch C, Burcin M (2018) Requirements for using iPSC-based cell models for assay development in drug discovery. Adv Biochem Eng Biotechnol 163:207–220

Commandeur S, Sparks SJ, Chan HL, Gao L, Out JJ, Gruis NA, van Doorn R, El Ghalbzouri A (2014) In-vitro melanoma models: invasive growth is determined by dermal matrix and basement membrane. Melanoma Res 24:305–314

Commandeur S, van Drongelen V, de Gruijl FR, El Ghalbzouri A (2012) Epidermal growth factor receptor activation and inhibition in 3d in vitro models of normal skin and human cutaneous squamous cell carcinoma. Cancer Sci 103:2120–2126

Costello R, Kissenpfennig A, Martins PN, McDaid J (2018) Development of transplant immunosuppressive agents – considerations in the use of animal models. Expert Opin Drug Discov 13:1041–1053

Crosby P (1996) Quality is still free: making quality certain in uncertain times. McGraw-Hill, New York

Cruz Rodriguez N, Lineros J, Rodriguez CS, Martinez LM, Rodriguez JA (2019) Establishment of two dimensional (2d) and three-dimensional (3d) melanoma primary cultures as a tool for in vitro drug resistance studies. Methods Mol Biol 1913:119–131

Devito L, Donne M, Kolundzic N, Khurana P, Hobbs C, Kaddour G, Dubrac S, Gruber R, Schmuth M, Mauro T, Ilic D (2018) Induced pluripotent stem cell line from an atopic dermatitis patient heterozygous for c.2282del4 mutation in filaggrin: Kcli001-a. Stem Cell Res 31:122–126

Dewhirst MW, Secomb TW (2017) Transport of drugs from blood vessels to tumour tissue. Nat Rev Cancer 17:738–750

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of r&d costs. J Health Econ 47:20–33

Dirnagl U, Kurreck C, Castanos-Velez E, Bernard R (2018) Quality management for academic laboratories: burden or boon? Professional quality management could be very beneficial for academic research but needs to overcome specific caveats. EMBO Rep 19:e47143

Doskey CM, van't Erve TJ, Wagner BA, Buettner GR (2015) Moles of a substance per cell is a highly informative dosing metric in cell culture. PLoS One 10:e0132572

Egner U, Hillig RC (2008) A structural biology view of target drugability. Expert Opin Drug Discov 3:391–401

European Commission (2019) Report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015–2017 [cited 2020/01/10]. Available from https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581689520921&uri=CELEX:52020DC0016

FDA (2020) In vitro drug interaction studies – cytochrome p450 enzyme- and transporter-mediated drug interactions: guidance for industry. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), Washington

Feaver RE, Cole BK, Lawson MJ, Hoang SA, Marukian S, Blackman BR, Figler RA, Sanyal AJ, Wamhoff BR, Dash A (2016) Development of an in vitro human liver system for interrogating nonalcoholic steatohepatitis. JCI Insight 1:e90954

Federal Ministry of Food and Agriculture (2019) Report on the use of animals used for scientific purposes in germany [cited 2020/01/10]. Available from: https://www.bmel.de/DE/Tier/Tierschutz/_texte/Versuchstierzahlen2018.html

German Research Foundation (2019) Animal experimentation in research: the 3rs principle and the validity of scientific research. https://www.dfgde/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/handreichung_sk_tierversuche_enpdf

Green H, Kehinde O, Thomas J (1979) Growth of cultured human epidermal cells into multiple epithelia suitable for grafting. Proc Natl Acad Sci U S A 76:5665–5668

Gronbach L, Wolff C, Klinghammer K, Stellmacher J, Jurmeister P, Alexiev U, Schäfer-Korting M, Tinhofer I, Keilholz U, Zoschke C (2020) A multilayered epithelial mucosa model of head neck squamous cell carcinoma for analysis of tumor-microenvironment interactions and drug development. Biomaterials (in revision)

Hagiwara M, Koh I (2020) Engineering approaches to control and design the in vitro environment towards the reconstruction of organs. Develop Growth Differ 00:1–9

Hartung T (2007) Food for thought... on cell culture. ALTEX 24:143–152

Hartung T (2013) Look back in anger – what clinical studies tell us about preclinical work. ALTEX 30:275–291

Hausmann C, Vogt A, Kerscher M, Ghoreschi K, Schäfer-Korting M, Zoschke C (2020) Optimizing skin pharmacotherapy for older patients: the future is at hand but are we ready for it? Drug Discov Today 25:851–861

Hausmann C, Zoschke C, Wolff C, Darvin ME, Sochorová M, Kovacik A, Wanjiku B, Schumacher F, Tigges J, Kleuser B, Lademann J, Fritsche E, Vávrová K, Ma N, Schäfer-Korting M (2019) Fibroblast origin shapes tissue homeostasis, epidermal differentiation, and drug uptake. Sci Rep 9:2913

Hirschfield GM, Mason A, Luketic V, Lindor K, Gordon SC, Mayo M, Kowdley KV, Vincent C, Bodhenheimer HC Jr, Pares A, Trauner M, Marschall HU, Adorini L, Sciacca C, Beecher-Jones T, Castelloe E, Bohm O, Shapiro D (2015) Efficacy of obeticholic acid in patients with primary biliary cirrhosis and inadequate response to ursodeoxycholic acid. Gastroenterology 148:751–761.e758

Huh D, Kim HJ, Fraser JP, Shea DE, Khan M, Bahinski A, Hamilton GA, Ingber DE (2013) Microfabrication of human organs-on-chips. Nat Protoc 8:2135–2157

Iansante V, Mitry RR, Filippi C, Fitzpatrick E, Dhawan A (2018) Human hepatocyte transplantation for liver disease: current status and future perspectives. Pediatr Res 83:232–240

Joseph J, Gronbach L, García-Miller J, Cruz L, Wuest B, Keilholz U, Zoschke C, Parr M (2020) Automated real-time tumor pharmacokinetic profiling in 3d models: a novel approach for personalized medicine. Pharmaceutics 12:E413

Kamiske G, Brauer J (1993) Modern quality management. In: Hansen W, Jansen H, Kamiske G (eds) Quality mangement in companies. Springer, Heidelberg, pp 137–161

Lefevre IA, Balice-Gordon RJ (2019) Learning from principles of evidence-based medicine to optimize nonclinical research practices. In: Bespalov A, Michel MC, Steckler T (eds) Handbook of experimental pharmacology: "good research practice in non-clinical pharmacology and biomedicine". Springer, Cham, pp 35–54

Leontaridou M, Urbisch D, Kolle SN, Ott K, Mulliner DS, Gabbert S, Landsiedel R (2017) The borderline range of toxicological methods: quantification and implications for evaluating precision. ALTEX 34:525–538

Lin Z, Luo G, Du W, Kong T, Liu C, Liu Z (2020) Recent advances in microfluidic platforms applied in cancer metastasis: circulating tumor cells' (CTCS) isolation and tumor-on-a-chip. Small 16:e1903899

Lind JU, Busbee TA, Valentine AD, Pasqualini FS, Yuan H, Yadid M, Park SJ, Kotikian A, Nesmith AP, Campbell PH, Vlassak JJ, Lewis JA, Parker KK (2017) Instrumented cardiac microphysiological devices via multimaterial three-dimensional printing. Nat Mater 16:303–308

Löwa A, Graff P, Kaessmeyer S, Hedtrich S (2020) Fibroblasts from atopic dermatitis patients trigger inflammatory processes and hyperproliferation in human skin equivalents. J Eur Acad Dermatol Venereol 34:e262–e265

Löwa A, Vogt A, Kaessmeyer S, Hedtrich S (2018) Generation of full-thickness skin equivalents using hair follicle-derived primary human keratinocytes and fibroblasts. J Tissue Eng Regen Med 12:e2134–e2146

Maschmeyer I, Lorenz AK, Schimek K, Hasenberg T, Ramme AP, Hubner J, Lindner M, Drewell C, Bauer S, Thomas A, Sambo NS, Sonntag F, Lauster R, Marx U (2015) A four-organ-chip for interconnected long-term co-culture of human intestine, liver, skin and kidney equivalents. Lab Chip 15:2688–2699

Mathijssen RH, Sparreboom A, Verweij J (2014) Determining the optimal dose in the development of anticancer agents. Nat Rev Clin Oncol 11:272–281

Menshykau D (2017) Emerging technologies for prediction of drug candidate efficacy in the preclinical pipeline. Drug Discov Today 22:1598–1603

Minchinton AI, Tannock IF (2006) Drug penetration in solid tumours. Nat Rev Cancer 6:583–592

Mueller MM, Fusenig NE (2004) Friends or foes – bipolar effects of the tumour stroma in cancer. Nat Rev Cancer 4:839–849

Nair A, Morsy MA, Jacob S (2018) Dose translation between laboratory animals and human in preclinical and clinical phases of drug development. Drug Dev Res 79:373–382

Nallanthighal S, Heiserman JP, Cheon DJ (2019) The role of the extracellular matrix in cancer stemness. Front Cell Dev Biol 7:86

Nashimoto Y, Okada R, Hanada S, Arima Y, Nishiyama K, Miura T, Yokokawa R (2020) Vascularized cancer on a chip: the effect of perfusion on growth and drug delivery of tumor spheroid. Biomaterials 229:119547

OECD (2018) Guidance document on good in vitro method practices (GIVIMP). OECD series on testing and assessment, no. 286. OECD Publishing, Paris. https://doi.org/10.1787/9789264304796-en

OECD (2019a) Test no. 431: In vitro skin corrosion: reconstructed human epidermis (RHE) test method. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264264618-en

OECD (2019b) Test no. 432: In vitro 3T3 NRU phototoxicity test. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264071162-en

OECD (2020) Test no. 439: In vitro skin irritation: reconstructed human epidermis test method. OECD guidelines for the testing of chemicals, Section 4. OECD Publishing, Paris. https://doi.org/10.1787/9789264242845-en

Oldhafer F, Wittauer EM, Falk CS, DeTemple DE, Beetz O, Timrott K, Kleine M, Vondran FWR (2019) Alloresponses of mixed lymphocyte hepatocyte culture to immunosuppressive drugs as an in-vitro model of hepatocyte transplantation. Ann Transplant 24:472–480

Palechor-Ceron N, Krawczyk E, Dakic A, Simic V, Yuan H, Blancato J, Wang W, Hubbard F, Zheng YL, Dan H, Strome S, Cullen K, Davidson B, Deeken JF, Choudhury S, Ahn PH, Agarwal S, Zhou X, Schlegel R, Furth PA, Pan CX, Liu X (2019) Conditional reprogramming for patient-derived cancer models and next-generation living biobanks. Cell 8:1327

Prabhakarpandian B, Shen MC, Nichols JB, Mills IR, Sidoryk-Wegrzynowicz M, Aschner M, Pant K (2013) Sym-bbb: a microfluidic blood brain barrier model. Lab Chip 13:1093–1101

Prantil-Baun R, Novak R, Das D, Somayaji MR, Przekwas A, Ingber DE (2018) Physiologically based pharmacokinetic and pharmacodynamic analysis enabled by microfluidically linked organs-on-chips. Annu Rev Pharmacol Toxicol 58:37–64

Radbruch M, Pischon H, Ostrowski A, Volz P, Brodwolf R, Neumann F, Unbehauen M, Kleuser B, Haag R, Ma N, Alexiev U, Mundhenk L, Gruber AD (2017) Dendritic core-multishell nanocarriers in murine models of healthy and atopic skin. Nanoscale Res Lett 12:64

Ratziu V, Harrison SA, Francque S, Bedossa P, Lehert P, Serfaty L, Romero-Gomez M, Boursier J, Abdelmalek M, Caldwell S, Drenth J, Anstee QM, Hum D, Hanf R, Roudot A, Megnien S, Staels B, Sanyal A (2016) Elafibranor, an agonist of the peroxisome proliferator-activated receptor-α and -δ, induces resolution of nonalcoholic steatohepatitis without fibrosis worsening. Gastroenterology 150:1147–1159.e1145

Robertson ED, Weir L, Romanowska M, Leigh IM, Panteleyev AA (2012) Arnt controls the expression of epidermal differentiation genes through hdac- and egfr-dependent pathways. J Cell Sci 125:3320–3332

Saleem A, Price PM (2008) Early tumor drug pharmacokinetics is influenced by tumor perfusion but not plasma drug exposure. Clin Cancer Res 14:8184–8190

Sauer UG, Hill EH, Curren RD, Raabe HA, Kolle SN, Teubner W, Mehling A, Landsiedel R (2016) Local tolerance testing under reach: accepted non-animal methods are not on equal footing with animal tests. Altern Lab Anim 44:281–299

Schütte M, Risch T, Abdavi-Azar N, Boehnke K, Schumacher D, Keil M, Yildiriman R, Jandrasits C, Borodina T, Amstislavskiy V, Worth CL, Schweiger C, Liebs S, Lange M, Warnatz HJ, Butcher LM, Barrett JE, Sultan M, Wierling C, Golob-Schwarzl N, Lax S, Uranitsch S, Becker M, Welte Y, Regan JL, Silvestrov M, Kehler I, Fusi A, Kessler T, Herwig R, Landegren U, Wienke D, Nilsson M, Velasco JA, Garin-Chesa P, Reinhard C, Beck S, Schafer R, Regenbrecht CR, Henderson D, Lange B, Haybaeck J, Keilholz U, Hoffmann J, Lehrach H, Yaspo ML (2017) Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to egfr inhibitors. Nat Commun 8:14262

Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H, Hennessy L, Finnerty CC, Lopez CM, Honari S, Moore EE, Minei JP, Cuschieri J, Bankey PE, Johnson JL, Sperry J, Nathens AB, Billiar TR, West MA, Jeschke MG, Klein MB, Gamelli RL, Gibran NS, Brownstein BH, Miller-Graziano C, Calvano SE, Mason PH, Cobb JP, Rahme LG, Lowry SF, Maier RV, Moldawer LL, Herndon DN, Davis RW, Xiao W, Tompkins RG, Inflammation, Host Response to Injury LSCRP (2013) Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A 110:3507–3512

Sieber S, Wirth L, Cavak N, Koenigsmark M, Marx U, Lauster R, Rosowski M (2018) Bone marrow-on-a-chip: long-term culture of human haematopoietic stem cells in a three-dimensional microfluidic environment. J Tissue Eng Regen Med 12:479–489

Simeon-Dubach D, Zeisberger SM, Hoerstrup SP (2016) Quality assurance in biobanking for pre-clinical research. Transfus Med Hemother 43:353–357

Smith MA, Houghton P (2013) A proposal regarding reporting of in vitro testing results. Clin Cancer Res 19:2828–2833

Spilker ME, Chen X, Visswanathan R, Vage C, Yamazaki S, Li G, Lucas J, Bradshaw-Pierce EL, Vicini P (2017) Found in translation: maximizing the clinical relevance of nonclinical oncology studies. Clin Cancer Res 23:1080–1090

Tao L, Zhang P, Qin C, Chen SY, Zhang C, Chen Z, Zhu F, Yang SY, Wei YQ, Chen YZ (2015) Recent progresses in the exploration of machine learning methods as in-silico adme prediction tools. Adv Drug Deliv Rev 86:83–100

van Drongelen V, Alloul-Ramdhani M, Danso MO, Mieremet A, Mulder A, van Smeden J, Bouwstra JA, El Ghalbzouri A (2013) Knock-down of filaggrin does not affect lipid organization and composition in stratum corneum of reconstructed human skin equivalents. Exp Dermatol 22:807–812

van Drongelen V, Danso MO, Mulder A, Mieremet A, van Smeden J, Bouwstra JA, El Ghalbzouri A (2014a) Barrier properties of an n/tert-based human skin equivalent. Tissue Eng Part A 20:3041–3049

van Drongelen V, Haisma EM, Out-Luiting JJ, Nibbering PH, El Ghalbzouri A (2014b) Reduced filaggrin expression is accompanied by increased staphylococcus aureus colonization of epidermal skin models. Clin Exp Allergy 44:1515–1524

van Luijk J, Bakker B, Rovers MM, Ritskes-Hoitinga M, de Vries RBM, Leenaars M (2014) Systematic reviews of animal studies; missing link in translational research? PLoS One 9: e89981

Vávrová K, Henkes D, Strüver K, Sochorová M, Skolová B, Witting MY, Friess W, Schreml S, Meier RJ, Schäfer-Korting M, Fluhr JW, Küchler S (2014) Filaggrin deficiency leads to impaired lipid profile and altered acidification pathways in a 3d skin construct. J Invest Dermatol 134:746–753

Wacker D, Stevens RC, Roth BL (2017) How ligands illuminate gpcr molecular pharmacology. Cell 170:414–427

Waldera-Lupa DM, Kalfalah F, Florea AM, Sass S, Kruse F, Rieder V, Tigges J, Fritsche E, Krutmann J, Busch H, Boerries M, Meyer HE, Boege F, Theis F, Reifenberger G, Stühler K (2014) Proteome-wide analysis reveals an age-associated cellular phenotype of in situ aged human fibroblasts. Aging (Albany NY) 6:856–878

Waldera Lupa DM, Kalfalah F, Safferling K, Boukamp P, Poschmann G, Volpi E, Götz-Rösch C, Bernerd F, Haag L, Huebenthal U, Fritsche E, Boege F, Grabe N, Tigges J, Stühler K, Krutmann J (2015) Characterization of skin aging-associated secreted proteins (saasp) produced by dermal fibroblasts isolated from intrinsically aged human skin. J Invest Dermatol 135:1954–1968

Wanjiku B, Yamamoto K, Klossek A, Schumacher F, Pischon H, Mundhenk L, Rancan F, Judd MM, Ahmed M, Zoschke C, Kleuser B, Rühl E, Schäfer-Korting M (2019) Qualifying x-ray and stimulated raman spectromicroscopy for mapping cutaneous drug penetration. Anal Chem 91:7208–7214

Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, Pairaudeau G, Pennie WD, Pickett SD, Wang J, Wallace O, Weir A (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov 14:475–486

Warren HS, Tompkins RG, Moldawer LL, Seok J, Xu W, Mindrinos MN, Maier RV, Xiao W, Davis RW (2015) Mice are not men. Proc Natl Acad Sci U S A 112:E345

Wolff C, Zoschke C, Kalangi SK, Reddanna P, Schäfer-Korting M (2019) Tumor microenvironment determines drug efficacy in vitro – apoptotic and anti-inflammatory effects of 15-lipoxygenase metabolite, 13-hpotre. Eur J Pharm Biopharm 142:1–7

Wong CH, Siah KW, Lo AW (2019) Estimation of clinical trial success rates and related parameters. Biostatistics 20:273–286

Yamamoto K, Flesch R, Ohigashi T, Hedtrich S, Klossek A, Patoka P, Ulrich G, Ahlberg S, Rancan F, Vogt A, Blume-Peytavi U, Schrade P, Bachmann S, Schäfer-Korting M, Kosugi N,

Rühl E (2015) Selective probing of the penetration of dexamethasone into human skin by soft x-ray spectromicroscopy. Anal Chem 87:6173–6179

Yamamoto K, Klossek A, Flesch R, Rancan F, Weigand M, Bykova I, Bechtel M, Ahlberg S, Vogt A, Blume-Peytavi U, Schrade P, Bachmann S, Hedtrich S, Schäfer-Korting M, Rühl E (2017) Influence of the skin barrier on the penetration of topically-applied dexamethasone probed by soft x-ray spectromicroscopy. Eur J Pharm Biopharm 118:30–37

Younossi ZM, Ratziu V, Loomba R, Rinella M, Anstee QM, Goodman Z, Bedossa P, Geier A, Beckebaum S, Newsome PN, Sheridan D, Sheikh MY, Trotter J, Knapple W, Lawitz E, Abdelmalek MF, Kowdley KV, Montano-Loza AJ, Boursier J, Mathurin P, Bugianesi E, Mazzella G, Olveira A, Cortez-Pinto H, Graupera I, Orr D, Gluud LL, Dufour JF, Shapiro D, Campagna J, Zaru L, MacConell L, Shringarpure R, Harrison S, Sanyal AJ (2019) Obeticholic acid for the treatment of non-alcoholic steatohepatitis: interim analysis from a multicentre, randomised, placebo-controlled phase 3 trial. Lancet 394:2184–2196

Zhu K, Shin SR, van Kempen T, Li YC, Ponraj V, Nasajpour A, Mandla S, Hu N, Liu X, Leijten J, Lin YD, Hussain MA, Zhang YS, Tamayol A, Khademhosseini A (2017) Gold nanocomposite bioink for printing 3d cardiac constructs. Adv Funct Mater 27:1605352

Zoschke C, Ulrich M, Sochorová M, Wolff C, Vávrová K, Ma N, Ulrich C, Brandner JM, Schäfer-Korting M (2016) The barrier function of organotypic non-melanoma skin cancer models. J Control Release 233:10–18

Zoschke M (1993) Quality function deployment (qfd). In: Hansen W, Jansen HH, Kamiske GF (eds) Quality management in the company. Springer, Berlin, pp 1–22

Zoschke M (2009) Quality function deployment (qfd). In: Kamiske GF (ed) Quality techniques for engineers. Symposion Publishing, Düsseldorf, pp 137–161

Zuang V, Dura A, Asturiol Bofill D, Viegas Barroso J, Batista Leite S, Belz S, Berggren E, Bernasconi C, Bopp S, Bouhifd M, Bowe G, Campia I, Casati S, Coecke S, Corvi R, Gribaldo L, Grignard E, Halder M, Holloway M, Kienzler A, Landesmann B, Madia F, Milcamps A, Morath S, Munn S, Paini A, Pistollato F, Price A, Prieto Peraita M, Richarz A, Sala Benito J, Wilk-Zasadna I, Wittwehr C, Worth A, Whelan M (2018) Eurl ecvam status report on the development, validation and regulatory acceptance of alternative methods and approaches. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/818599