feature

Securing reliability and validity in biomedical research: an essential task

Thomas Wilckens

The buzzword 'translational' dominates concepts to optimize value creation from science. This article discusses the impact of 'old' and contemporary data on hypothesis generation in relation to human physiology and in the effort to optimally implement translational sciences. I outline how dogmas and errors, sometimes perpetuated over decades, impact contemporary research and drug discovery projects. As a consequence and to improve value creation from science, a reevaluation of old data (i.e. of the validity and reliability of research with regard to human physiology) seems necessary. In line with this, the compliance of newly generated hypotheses, assays and tools with a conceptual focus on human physiology as the gold standard seems essential. To achieve improved research success, several measures need to be initiated and guided by industrial and academic leaders in concert to have an impact on the quality of research in the very near future. There is no 'holy grail', but in general terms, a constructive but critical approach – not just to contemporary biomedical research – seems mandatory to avoid the errors of the past and enable solutions to evolve dynamically.

A recent article in the *Financial Times* proposed 'drug research needs serendipity' [1]. What seems rather more needed is an incentive to identify and question pre-existing errors and dogmas, some of which have evolved over decades, and to re-evaluate essential data, which build the foundation of our contemporary research with respect to their relevance to human biology. This requirement could be defined as 'transcriptional science' (TS). I will use this paraphrase to describe a conceptual approach analogous to translational science, simply because in biology, transcription is the rate-limiting step for translation, and, if it goes awry, it can lead to false or irrelevant products.

To this author, translational science means that human physiology is positioned in the centre of all biomedical sciences to enable the right questions to translate biomedical research into new therapeutics [2–4]. Thus, the reliability and validity (Table 1) of a given study needs to be assessed *a priori*.

What, however, assures the reliability and validity of the data that are used to delineate translational hypotheses and generate the related experimental designs? An old Chinese saying implicates one aspect: old data should not be neglected but might require reconsideration in a new context (Fig. 1).

In general, the quality and value of a study are only as good as the design of the study and, equally importantly, the previous data and related interpretations on which the new hypothesis will be built. This is where the idea of TS steps in: TS aims to exclude data and related interpretations, which poorly relate to human conditions, whether generated *in vitro* or *in vivo*, for both hypothesis generation and the design of new studies. If possible, TS will re-evaluate data and extract the content that can be used for translational databases or hypothesis generation, a task that obviously requires manual data analysis and (re)interpretation. TS should serve as an essential prerequisite and quality control for translational research and related drug discovery projects; otherwise, errors will inevitably be perpetuated and invalidate all efforts made in translational sciences. TS requires that researchers approach their own work and the work of others (as well as the related interpretations) most critically, from ancient reports to

TABLE 1

Definitions.

Validity and reliability (partially adopted from Wikipedia.org and L.T.F. Gamut, *Logic, Language, and Meaning: Introduction to logic*, p. 115):

Reliability does not imply validity. Both terms are used in test theories and relate to the logic and accuracy of a test, i.e. an experimental design and the related results, if adapted to biomedical research.

A reliable measure is measuring something consistently, but one may not be measuring what is being intended to be measured. For example, while there are many reliable tests of specific biological reactions, not all of them would be valid for predicting, for example a glucose response to different stressors. In terms of accuracy and precision, reliability is analogous to precision, while validity is analogous to accuracy.

An example often used to illustrate the difference between reliability and validity in the experimental sciences involves a common bathroom scale.
If someone who is steps on a scale 10 times and gets readings of 25, 50, 100, 125, etc., the scale is not reliable. If the scale consistently reads "65", then it is reliable, but not valid. If it reads "80" each time, then the measurement is both reliable and valid. This is what is meant by the statement, "Reliability is necessary but not sufficient for validity."

Reliability requires better comparable experiments, while validity asks the question if the experiment is tailored to appropriately answer the questions being asked; i.e. if the experiment is valid in logic terms. In the dynamic context of increasing knowledge in biomedical research, both, reliability and validity of an experiment may require adjustment to the current status of science. I.e. in retrospective reliability and validity may need to be newly assessed for a given experiment, which may enable new hypothesis generation and even conclusions based on data generated earlier.

modern biomedical research and contemporary work.

Obviously, the quest for an approach that assures higher validity and reliability of data used for translational science implies that there are many dogmas and perpetuated errors in our scientific literature and community. Indeed, they do exist and might even have become stronger with time. Some of them seem to erode (i.e. drug-target selectivity as a predictor of desired therapeutic effects has been questioned), and the fact that most compounds are acting on multiple targets is implying new concepts [5]. Similarly, the animal models that are the cornerstones of a research field are being challenged [6].

The cortisol story

Here, just one example, which I consider of major impact for our contemporary therapeutic concepts and drug discovery related to a plethora of pathological conditions, will be briefly discussed to demonstrate some interrelationships between various mechanisms that can contribute to the constitution of a dogma.

Before getting into details, it should be mentioned that I am not questioning the undisputable negative effects of chronic excessive stress or hypercortisolism. However, that acute cortisol release enables coping with a variety of stressors to defend homeostasis and even enhances immunity [7] challenges the general view of the effects that 'glucocorticoids'

as a drug class might have on immunity and inflammation from a teleological viewpoint.

I argue that the almost standardized approach – that is, the interpolation of effects mainly generated by the use of synthetic glucocorticoids, which cannot be used in physiological concentrations *per definition*, or the supraphysiological use of cortisol, often in conjunction with the neglect of appropriate experimental design (there is no physiological state without cortisol present) – has perpetuated a false dogma: glucocorticoids as a drug class are, in general, considered to be immunosuppressive and anti-inflammatory, although the 'class' comprises compounds with highly different physiological and pharmacological profiles.

Cortisone therapy fell out of favor in the 1950s because of undesired effects observed in high non-physiological doses. The voices of Hench's contemporary colleagues, who emphasized that low, physiological replacement-like dosing regimens (i.e. lower than those currently considered low dose and using the endogenous cortisol, not prednisone or prednisolone) benefit patients and all the negative effects were due to the very high pharmacological regimens, were overheard [8,9]. When the patents for cortisone expired, new compounds were generated, which aimed to limit some of these undesired effects [10]. In this context, it is almost ironic that low-dose corticosteroid therapy with synthetic compounds, mainly prednisolone, has become a new standard [11].

Thus, in the very early days of corticosteroid research, dexamethasone (DEX) – a very potent, high-affinity synthetic steroid, which behaves completely differently to the endogenous hormone cortisol – became the gold standard. Fig. 2 summarizes how DEX differs from cortisol, with the sole exception that it also binds to the glucocorticoid receptors (GRs), but not the mineralocorticoid receptor (MR), which cortisol binds to with a higher affinity than the GRs. It was postulated that the sum of DEX's effects on various targets would eventually reflect the physiological functions of cortisol; although the logic behind this concept has already been questioned, more than ten years ago, the well-supported arguments had little to no impact [12,13].

The complex interactions of the GRs and a proposed regulation at the tissue level have been described recently [14]. In line with this, a variety of microarray studies have documented that the standard GR and MR agonists induce and repress an overlapping but not identical portfolio of genes: in human liver cells, for example, of a total of 300 genes that are variously regulated by cortisol or corticosterone (both binding to the MRs) and DEX (only GRs), only 25 are equivalently regulated by all three of the gold standard agonists (M. Cidlowski, personal communication). Thus, the interpolation from one compound to the other as a 'class effect' seems obsolete, not least because we know that targets like nuclear receptors might dynamically and highly specifically respond to different ligands [15].

A recent and elegant study, however, has demonstrated that minor changes in low physiological corticosteroid concentrations have a major impact on experimental arthritis (i.e. a decrease in local cortisol prevents bone destruction), which is in stark contrast with all clinical experiences with glucocorticoid treatment

温故知新

Drug Discovery Today

FIGURE 1

On Ko Chi Shin, 'The new ideas reside within the old.' Confucius (551–479 BC). As this saying indicates, to (re-)investigate and understand old ideas is essential for innovation.
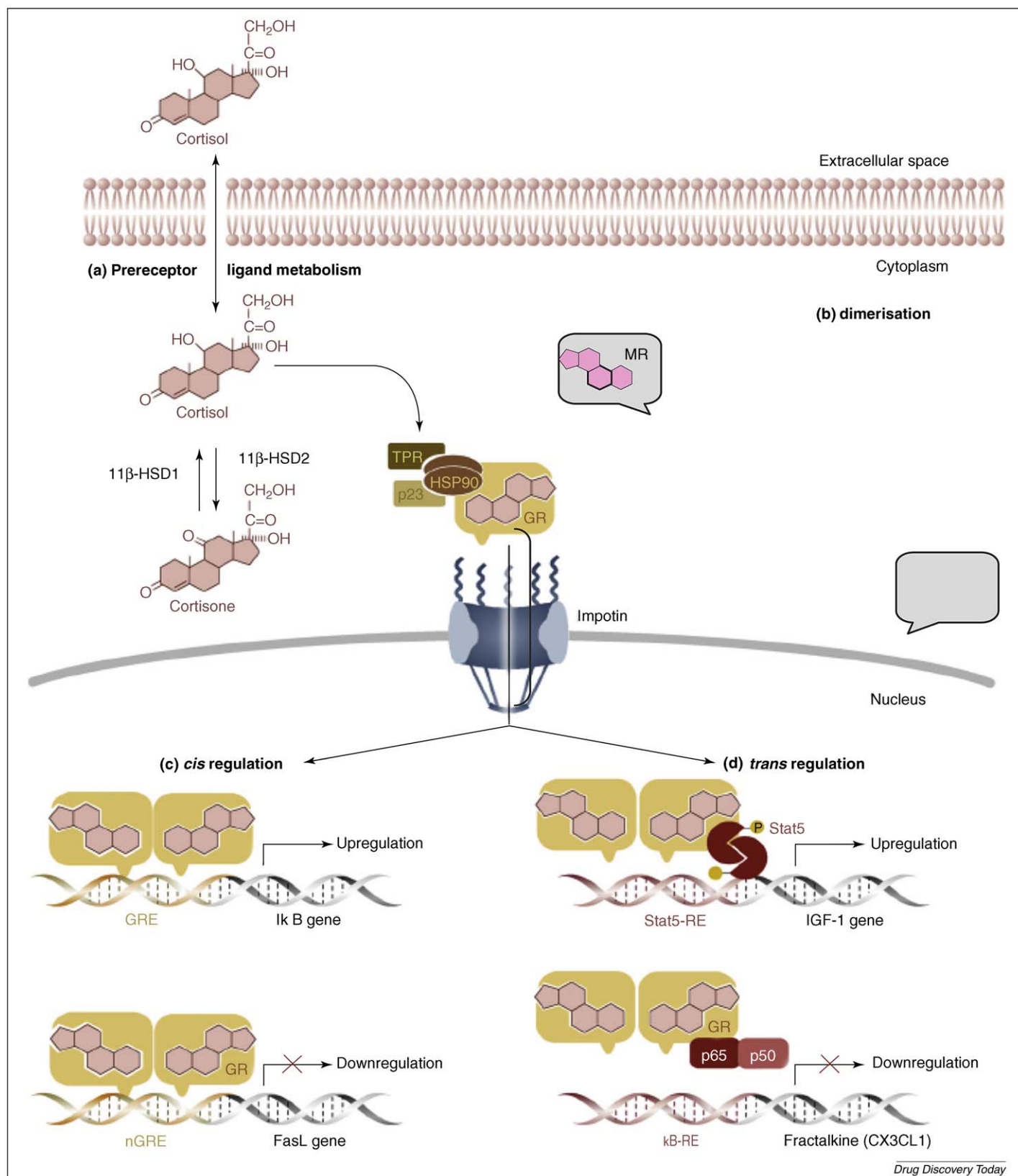
FIGURE 2

Selected differences affecting function: cortisol binding globulin; **(a)** metabolism by 11β-HSD; intracellular activation/recycling; binding to GR (five isoforms), MR (two isoforms); **(b)** GR and MR dimer formation; heterodimer formation occurs physiologically; receptor affinity; receptor–ligand interactions (conformational changes induced by ligand); receptor–ligand co-activator, repressor assembling; **(c,d)** transrepression/transactivation (differences between all GR ligands); ligand-dependent change of conformation and related effects. Endogenous cortisol differs from synthetic glucocorticoid receptor ligands (in particular, dexamethasone) at almost every step along the activation and metabolism pathways. Figure modified, with permission, from Ref. [14].

Perspective • FEATURE

in rheumatoid arthritis [16]. Interestingly, although this contention would have immediate consequences for cortisol research related to inflammation and immunity, the authors highlight other results as their major finding, rather than the conflict with current views. It seems there is a bias to stay within the accepted conceptual framework. Similar observations (i.e. a 'white hat bias') have been made in obesity research. In general, there seems to be a trend to interpret data within the mainstream framework of a research community [17]. The impact of this dramatic perpetuation of the bias in cortisol research on biomedical research and related drug discovery is best documented by the fact that the putative pro-inflammatory function and essential role of endogenous cortisol in cardiovascular disease, stroke and, possibly, many other diseases might have been overlooked for 50 years [18].

That DEX and similar synthetic glucocorticoids do not activate the MR, which is protected from occupation with cortisol in some, but not all, tissues by a cortisol-catabolizing enzyme [19], might explain why some of these proposed pro-inflammatory actions of cortisol via the MR [20] were overlooked for decades. The opportunities that will arise from these new discoveries could affect inflammatory conditions such as arthritis, asthma or even metabolic diseases [19,20]. It is also easy to imagine the immense consequences this observation could have had on public health if it had been discovered 40 years ago.

A dogma can evolve from various influences, including patenting interests and the standardization of mainstream thinking and experimental designs; protected by converging commercial and academic interests, it might predominate over decades of research and drug development.

In the following paragraphs, some ideas are presented to avoid a similar situation and create new value from existing data.

Data integration and interpretation: improving content

How could better 'content' in translational terms be achieved? Often, scientists will uncritically or in a biased manner extrapolate experimental conditions and related results from recent high-profile publications to design their studies because it is neither common nor appropriate to question the publications of scientific leaders. This procedure ensures that views and designs that, in retrospect, seem irrelevant from a translational, 'human physiology first' viewpoint are perpetuated. The new study, consequently, might result again in data with little physiological relevance, no matter how elegantly and elaborately it is performed. This consequence often seems to be overlooked by the editors and reviewers of leading journals, in which cutting-edge technologies sometimes dominate the evaluation of a manuscript over the generation and foundation of the hypothesis *per se*.

In conjunction with this constellation, systems biology [21], systems chemical biology [22], bioinformatics, semantics and other tools that aim to support translational science in general [23] also rely on the quality of data entered into the relevant database. As outlined above, however, these data might be confounded if they are not manually evaluated. Thus, all computational methods are certainly helpful to extract and organize data; however, they cannot replace content ranking by the human brain, which is still superior in validating complex data constellations and experimental designs.

Creating transcriptional (i.e. high-confidence content) databases might seem an insurmountable obstacle in light of the exponential growth of data. However, if each project were to start with a standardized approach to ranking and interpreting data in hypothesis generation, with respect to their relevance to the *in vivo* dynamics of human physiology, improved validity and reliability of contemporary research seems feasible. The standards for such a content ranking approach need to be generated first, which is challenging because even defining which questions should be asked and identifying common denominators in a given field of research might seem too complex. Nevertheless, it is anticipated that even simple standards and ranking tools could improve the quality of new data considerable. There must be both academic and commercial interest in further developing
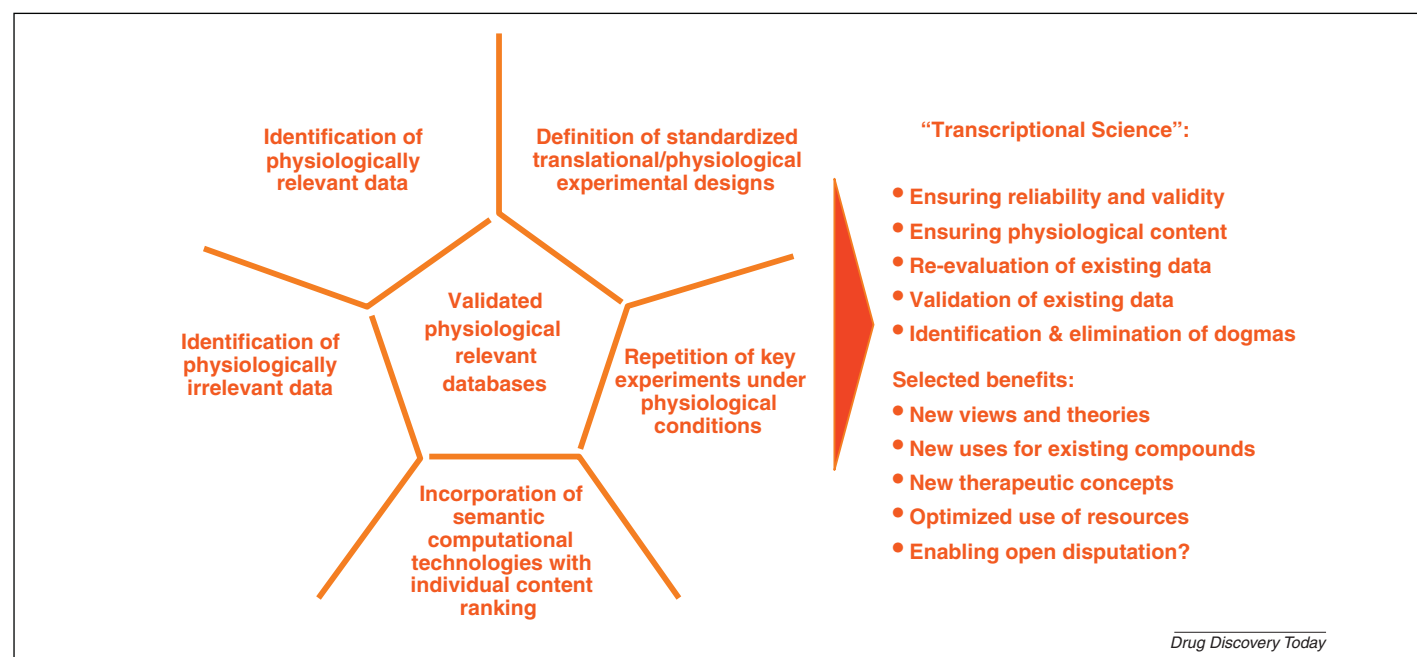

FIGURE 3

Transcriptional science: an essential prerequisite to enable translational sciences. If uncritically adopted, existing physiologically questionable or irrelevant data, related interpretations and dogmas will unequivocally be perpetuated and invalidate translational science *a priori*, including all financial investment.

this kind of knowledge base and content ranking, which is essential for informed decision making in any problem in biomedical sciences. Some aspects of the problem (i.e. too many neglected data) have been acknowledged [24], but the solution might require 'competitive collaboration', not only within the pharmaceutical industry [25] but also including academia.

Experimental design and standardization: biomarkers and more

When is an experiment physiologically relevant? It would be prudent to try to give a simple answer, but a direct consequence of this question is the fact that we need more standardized and truly physiologically relevant model systems to make data better comparable and, in particular, to rank them in a physiological context. First questions that come to mind are 'When is an experiment reflecting male or female conditions?' and 'How relevant is this after all?' In relation to the discussion above, small changes of cortisol dramatically impact on cellular functions (i.e. induce opposing functions in physiological concentrations) [26]. Findings like this force thought-provoking questions, such as 'How many in vitro assays, not just in this field, are of physiological relevance at all?'

Ultimately, TS implies that experiments that built the foundation of previous drug discovery programs and studies in progress must still be analyzed with respect to their physiological relevance, but they might also need to be repeated (e.g. to include new control conditions), perhaps often. At first, repeating 'old' studies might have little appeal for researchers driven by their hunt for impact factor scores. If appropriately addressed, however, the identification of new, more physiological functions for any given physiological pathway could open new and unexpected avenues for innovative treatments. This would ensure progress for both science and the scientist.

In addition to efforts to make data comparable, in particular for drug discovery, new standards for experiments might need to be agreed; generally defined and accepted biomarkers are one first approach. Regulatory authorities might require such initiatives sooner or later, but the consolidation of biomedical research forces a more efficacious use of resources now. If one only analyses the investments made in cortisol research since the patent expired in the 1950s (which was a major driver for new compounds at the time) and discovery projects based on data with limited translational relevance, it becomes obvious what resources could be released if our scientific approaches included a more aggressive evaluation of the existing fundaments on which we build our views and opinions, which often are prematurely interpreted as the truth (Fig. 3).

Transcriptional science, an integral part of translational sciences and its culture

In his review, titled 'Translational research: forging a new cultural identity', Barry Coller identified several challenges ahead, including the willingness to embrace change in general and to induce a new culture of scientific disputation [27]. Ensuring the reliability and validity of data generated in the paradigm of translational sciences will primarily require the implementation of a new culture in biomedical research, not another new terminology like 'transcriptional sciences', to assess data and scientific hypothesis more openly and with constructive criticism. Concepts such as endogenous angiogenic factors would not have been developed successfully if people like Judah Folkman had not resisted the opposition to their ideas and questioned dogmas throughout their careers.

Scientists in the industry and academia will have to appreciate that critical comments might help to improve the impact of their work, if the criticism is conveyed in an appropriate and constructive manner, ideally before the work is started. Open discussions and converging expertises are the only measures that will ultimately ensure a higher return of investment for societies.

'Open content ranking forums', established by publishers for specific areas of research, might be one approach worth exploring. Within companies, specific IT solutions for in-house content ranking by the companies' scientists, enabling them to participate proactively, could be developed. 'Constructive devil's advocates' (i.e. experts in challenging designs and ideas) could be trained and become part of project teams to promote a better success (translation) rate. In general, eliminating misleading concepts and experiments at the right time with the right questions should be rewarded – for example, by progression of a project to a milestone – both in academia and the industry. In addition, negative data or data that do not fit our current views should be more appreciated.

Challenging dogmas or falsifying accepted or new theories does not really exist as a research goal, although it is an equally logical and rewarding approach to generating knowledge [28]. After all, science gets exciting when things do not fit and new ground is touched. There is no holy grail to achieve a better return on investment in biomedical research related to personal effort and funding, but there are many ad hoc opportunities to improve the reliability and validity of experimental designs and extract additional value from old knowledge. To unfold this potential and secure progress in finding new cures for unmet needs, however, will require a concerted approach that can only be initiated and guided by leaders from academia and industry in concert.

References

1 Shawitz, D. and Taleb, N. (2008) Drug research needs serendipity. Financial Times
2 Littman, B.H. et al. (2007) What's next in translational medicine? Clin. Sci. (Lond.) 112, 217–227
3 Nathan, D.G. (2005) The several Cs of translational clinical research. J. Clin. Invest. 115, 795–797
4 Fitzgerald, G.A. (2005) Opinion: anticipating change in drug development: the emerging era of translational medicine and therapeutics. Nat. Rev. Drug Discov. 4, 815–818
5 Mestres, J. and Gregori-Puigjane, E. (2009) Conciliating binding efficiency and polypharmacology. Trends Pharmacol. Sci. 30, 470–474
6 Martin, B. et al. (2010) "Control" laboratory rodents are metabolically morbid: why it matters. Proc. Natl. Acad. Sci. U.S.A. 107, 6127–6133
7 Dhabhar, F.S. (2009) Enhancing versus suppressive effects of stress on immune function: implications for immunoprotection and immunopathology. Neuroimmunomodulation 16, 300–317
8 Jefferies, W.M. (1955) The present status of ACTH, cortisone and related steroids in clinical medicine. N. Engl. J. Med. 253, 441–446
9 Jefferies, W.M. (1967) Low-dosage glucocorticoid therapy: an appraisal of its safety and mode of action in clinical disorders, including rheumatoid arthritis. Arch. Intern. Med. 119, 265–278
10 Jefferies, W.M. (1996) Safe Use of Cortisol. Charles C Thomas Publisher Ltd
11 Hoes, J.N. et al. (2007) EULAR evidence-based recommendations on the management of systemic glucocorticoid therapy in rheumatic diseases. Ann. Rheum. Dis. 66, 1560–1567
12 Wilckens, T. (1995) Glucocorticoids and immune function: physiological relevance and pathogenic potential of hormonal dysfunction. Trends Pharmacol. Sci. 16, 193–197
13 Wilckens, T. and De Rijk, R. (1997) Glucocorticoids and immune function: unknown dimensions and new frontiers. Immunol. Today 18, 418–424
14 Gross, K.L. and Cidlowski, J. (2008) Tissue-specific glucocorticoid action: a family affair. Trends Endocrinol. Metab. 19, 331–339
15 Margolis, R.N. et al. (2009) Chemical approaches to nuclear receptors in metabolism. Sci. Signal. 2 mr5
16 Buttgereit, F. et al. (2009) Transgenic disruption of glucocorticoid signaling in mature osteoblasts and osteocytes attenuates K/BxN mouse serum-induced arthritis in vivo. Arthritis Rheum. 60, 1998–2007
17 Cope, M.B. and Allison, D. (2010) White hat bias: examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. Int. J. Obes. (Lond.) 34, 84–88
18 Funder, J.W. (2009) Reconsidering the roles of the mineralocorticoid receptor. Hypertension 53, 286–290
19 Cooper, M.S. and Stewart, P.M. (2009) 11{beta}-Hydroxysteroid Dehydrogenase Type 1 and Its Role in the Hypothalamus-Pituitary-Adrenal Axis, Metabolic

Syndrome, and Inflammation. *J. Clin. Endocrinol. Metab.* 94, 4645–4654

20 Rickard, A.J. and Young, M.J. (2009) Corticosteroid receptors, macrophages and cardiovascular disease. *J. Mol. Endocrinol.* 42, 449–459

21 Krishna, R. *et al.* (2007) Effective integration of systems biology, biomarkers, biosimulation, and modeling in streamlining drug development. *J. Clin. Pharmacol.* 47, 738–743

22 Oprea, T.I. *et al.* (2007) Systems chemical biology. *Nat. Chem. Biol.* 3, 447–450

23 Ruttenberg, A. *et al.* (2007) Advancing translational research with the Semantic Web. *BMC Bioinformatics* 8 (Suppl. 3), S2

24 Editorial, (2009) Data's shameful neglect. *Nature* 461, 145

25 Bingham, A. and Ekins, S. (2009) Competitive collaboration in the pharmaceutical and biotechnology industry. *Drug Discov. Today* 14, 1079–1081

26 Lim, H.Y. *et al.* (2007) Glucocorticoids exert opposing effects on macrophage function dependent on their concentration. *Immunology* 122, 47–53

27 Coller, B.S. (2008) Translational research: forging a new cultural identity. *Mt. Sinai J. Med.* 75, 478–487

28 Popper, K. (1934) *The Logic of Scientific Discovery*, Springer, Vienna

Thomas Wilckens
Jakob-Klar-Str. 7, 80796 München, Germany
thomas.wilckens@innventis.com
TWilckens@AOL.com