International Journal of Toxicology
2018, Vol. 37(6) 466-471

DOI: 10.1177/1091581818803880
journals.sagepub.com/home/ijt

Embracing Transparency Through Data Sharing

Stéphanie Boué[1], Michael Byrne[2], A. Wallace Hayes[3,4], Julia Hoeng[1], and Manuel C. Peitsch[1]

Abstract

Low rates of reproducibility and translatability of data from nonclinical research have been reported. Major causes of irreproducibility include oversights in study design, failure to characterize reagents and protocols, a lack of access to detailed methods and data, and an absence of universally accepted and applied standards and guidelines. Specific areas of concern include uncharacterized antibodies and cell lines, the use of inappropriate sampling and testing protocols, a lack of transparency and access to raw data, and deficiencies in the translatability of findings to the clinic from studies using animal models of disease. All stakeholders—academia, industry, funding agencies, regulators, nonprofit entities, and publishers—are encouraged to play active roles in addressing these challenges by formulating and promoting access to best practices and standard operating procedures and validating data collaboratively at each step of the biomedical research life cycle.

Keywords

reproducibility, antibody, data sharing, translatability, transparency

Introduction

Researchers and funding entities rely on the reproducibility of published discoveries to create new lines of research and to translate research findings into therapeutic or other applications.[1] The scientific community, however, has expressed ongoing concerns regarding the lack of reproducibility and translatability of published biomedical research data.[2,3] Data that are not reliable and robust can lead to assumptions that undermine and nullify the validity of subsequent research[2] and create distrust among funding agencies, the research community, and the general public.

A second area of concern, data reuse, relates to how data are shared and/or cited.[4] Encouragingly, the volume of open-access data and the rate of data dissemination have increased and can lead to new discoveries.[5] Potential issues in data sharing include country- and/or agency-specific policies on open data, how to ensure compliance with those policies, and the existence and adherence to standards necessary to produce reusable data. The key for data reusability is the annotation of the data, also called metadata, in a way that clarifies how the data have been produced and what the data exactly are. Producing reusable files requires effort from the researchers and should be encouraged/rewarded. This is especially relevant in the case of big data in system science. Hence, the balance between developing standards that can endure in evolving (and therefore complex) fields and minimizing the burden of compliance is delicate and is key for the adoption of any standard.[6]

Initiatives within organizations such as the US National Institutes of Health (NIH)[7] and the UK Wellcome Trust,[8] as well as efforts by some scientific journals, are aimed at improving the reproducibility of experimental data[9] and promoting access to data in the open-access literature[10] and code from computational analyses.[11] In particular, considerable attention is being paid to antibodies and animal models.[12] It has proven difficult to translate findings from animal models to achieve a better understanding of human disease. The challenge should be attributed not only to interspecies differences but also to a lack of rigor in study design, model validation, or the usage of (partially) inappropriate animal models, resulting in low reproducibility of studies.

This article is based on an exhibitor-hosted program session presented at the American College of Toxicology's 38th Annual Meeting in November 2017. Topics covered during the meeting included important considerations for enhancing

[1] PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
[2] Global Biological Standards Institute (GBSI), Washington, DC, USA
[3] Michigan State University, Institute for Integrative Toxicology, East Lansing, MI, USA
[4] University of South Florida, College of Public Health, Tampa, FL, USA

Corresponding Author:
Stéphanie Boué, PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland.
Email: stephanie.boue@pmi.com

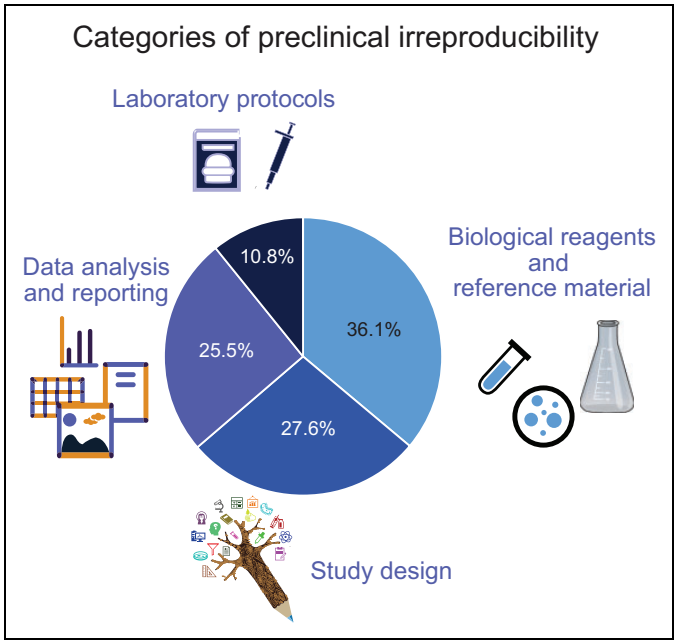

Figure 1. Categories of errors contributing to irreproducibility in preclinical research. Adapted from Freedman et al.[13]

the reproducibility and translatability of animal models, the importance of specimen preparation and antibody validation in improving the reproducibility of data collected from animal models, and potential avenues to sharing research transparently.

Reproducibility 2020: Progress and Priorities

Global Biological Standards Institute (GBSI) is a nonprofit entity founded in 2013 to improve the quality of preclinical research by advancing best practices and standards. Stakeholders include academic institutions, regulatory and funding agencies, industry, policy and professional organizations, and entities engaged in developing standards and promoting quality. To promote the reproducibility of research findings, GBSI supports standards initiatives in cell line authentication, antibody validation, and laboratory automation.

Multiple published analyses have reported low rates of reproducibility in preclinical studies, and the cost of US research that cannot be replicated has been estimated at $28 billion annually (Figure 1). The major causes of irreproducibility have been attributed to errors and inadequacies in study design, reagents and reference materials, laboratory protocols, and data analysis.[13] The core drivers contributing to irreproducibility are reporting and publication bias, underpowered studies, a lack of open access to methods and data, and a lack of clearly defined standards and guidelines in areas, such as reagent validation.[13] Growing concerns in the scientific community regarding funds spent on research that cannot be reproduced have prompted the NIH to suggest actions designed to promote reproducibility. The proposed interventions include training in good experimental design, checklists for reviewers of grant applications, and an online forum, PubMed Commons, which was developed as a pilot project to discuss published articles.[3,7] PubMed Commons, however, was recently discontinued after 5 years of existence due to a lack of usage, with comments submitted on only 6,000 of the 28 million articles indexed in PubMed. It will be important to follow alternatives to PubMed Commons that will allow scientists to share their comments, either on open peer-review platforms or, possibly, on other forums.

Antibodies are one class of reagents that contribute substantially to irreproducibility of research results. Typically, these widely used protein-binding reagents are validated and characterized rigorously in clinical studies but rarely in preclinical research, owing to a combination of a lack of standards, user apathy, and inconsistent quality assurance and control practices in manufacturing.[14,15] Even when antibodies have been validated and characterized, they may be misused by researchers with poor training in laboratory methods, resulting in unreliable and irreproducible findings. The GBSI is working to develop a qualitative scorecard for assessing antibody performance in techniques, such as Western blotting, immunohistochemistry, immunoprecipitation, and enzyme-linked immunosorbent assay. The intent of the scorecard is to provide a first-of-its-kind system to assess antibody performance using a point structure to reward the additional validation steps performed as well as the antibody's specificity and sensitivity to its intended target.

Another problem area affecting reproducibility is that variability in collection protocols and storage conditions of biospecimens can affect testing outcomes. Ideally, investigators would follow validated standard operating procedures (SOPs). However, there is currently a lack of SOPs, and specimens are often collected without using an SOP at all. When SOPs are used, they may not be adhered to consistently or made available to other research groups. Results may not be comparable where different protocols were used. Repositories such as BioSpecimen Commons (https://biospecimencommons.org/) offer researchers a database of SOPs for collection and analysis of a variety of biospecimen sources and types. Researchers can share and compare methods to ensure the use of appropriate samples and sampling techniques.

To improve reproducibility in nonclinical research, and by extension, the quality of scientific data, all stakeholders must play active roles. Funders should require prescreening of the proposed study design, reagent validation, cell line authentication, and open access to the protocols used, raw data, and publications. Research institutions should implement training programs for graduate students and postdoctoral researchers while holding their laboratory directors accountable for appropriate supervision. Journals should require detailed methods and more stringent reporting guidelines in research articles. The industry should provide validation data, including the results of internal replication procedures. Scientific societies should establish area-specific standards and competencies. All stakeholders should be involved in training and in post-publication

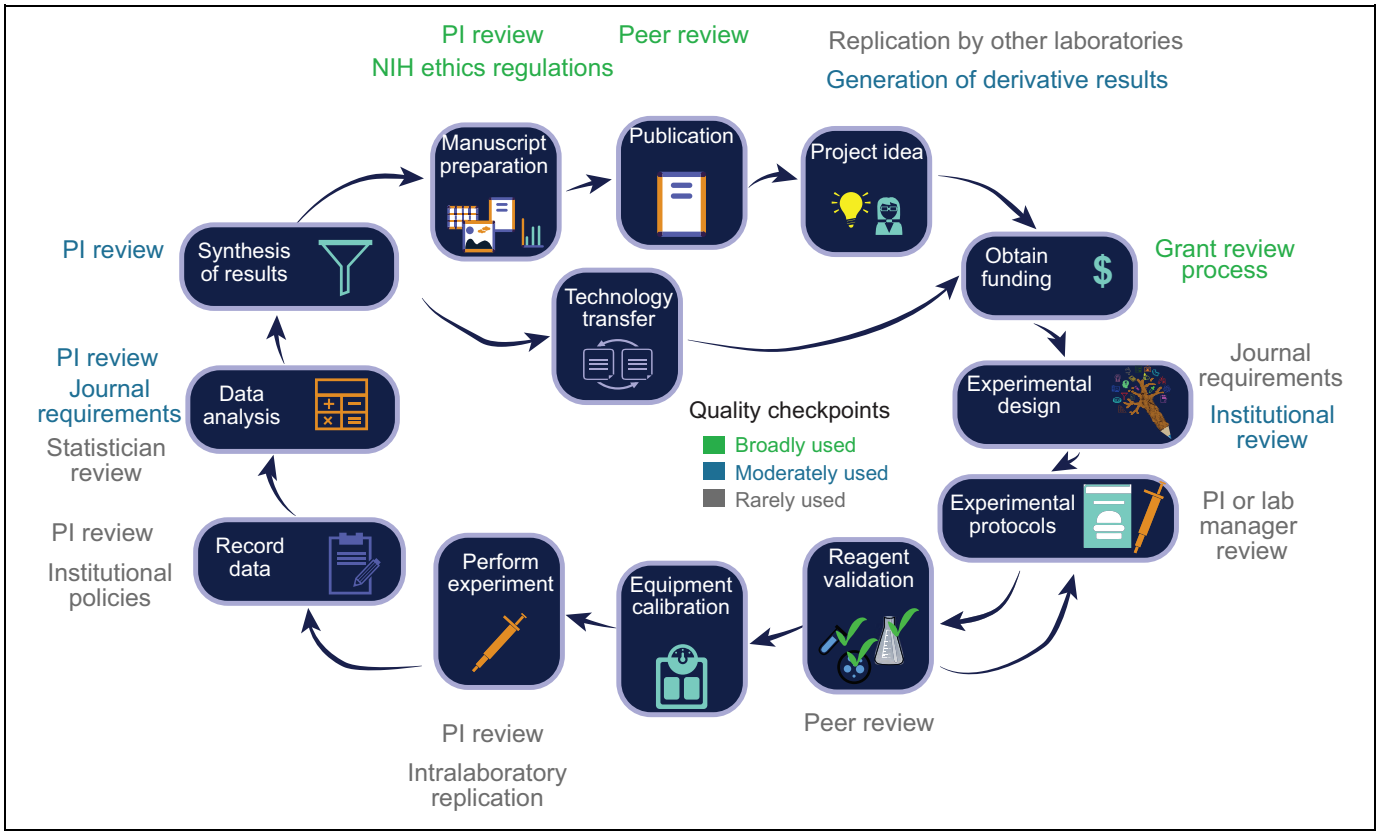

Figure 2. Common steps in the research life cycle. Adapted from Freedman et al.[12] PI indicates the principal investigator.

review. These collective efforts can improve research quality and enhance reproducibility.[12]

Reproducibility in Basic and Preclinical Research: Enhancing the Quality and Translatability of Animal Models

Two types of research utilize animal models: fundamental research, which includes hypothesis-driven mechanistic work that defines new concepts or identifies potential drug candidates, and nonclinical research, which includes proof-of-concept studies and safety/toxicity studies of new drugs or devices. Interspecies differences are a challenge in generalizing the findings of nonclinical research, and low reproducibility of studies with animal models of disease renders the findings less applicable. The reasons for low reproducibility include underpowered studies, poor experimental design, confirmation bias resulting from a lack of randomization and blinding, neglecting to account for sex-specific effects, inappropriate statistical analysis, and pressure to publish only positive results.[16] Reproducibility, however, is fundamental to maintaining funder and stakeholder trust and to justifying the funding of studies that may lead to drug development. Practices that result in poor reproducibility may lead to time and money spent on research that cannot be reproduced. Another unintended consequence is subjecting patients to clinical trials of agents that may not be effective. Figure 2 displays a schematic showing the 13 common steps of the research life cycle, each of which offers opportunities to improve the reproducibility of biomedical research.

Research in animals has long been a cornerstone of nonclinical investigation and is based on shared disease mechanisms and pathways. Nonetheless, modeling human disease in animals presents researchers with multiple challenges, including interindividual variability among humans owing to genetics, environmental contributions, age, diet, lifestyle, microbiome, and health status. Another challenge to translatability of nonclinical research findings is within-species differences in model animals.[17] Using inbred strains can minimize this variability, cut down on the number of animals required to detect differences between exposure groups, and maximize statistical power, as increased genetic diversity requires a larger sample to maintain power. However, maintaining genetic stability in colonies of inbred animals can become difficult over time. Translatability from animal models to humans can be done, but it has limitations. Minimizing bias to increase internal validity (ie, within an experiment, in a specific laboratory) can result in specific and reductionist experimental systems, which in turn lowers the external validity (ie, across laboratory and experiments) and generalizability of the findings.[17]

Importantly, before addressing the details of study design and conduct summarized below, it is of foremost importance to ensure that the animal models used are validated and predictive to address the clinical questions at hand (ie, that they are fit-for-

purpose).[18] One cannot expect that all clinical criteria are replicated perfectly in a single model, and gaining a detailed understanding of the strengths and limits of a model is crucial.

Reproducibility in nonclinical research involving animals can then be enhanced by accurate record-keeping, the cornerstone of Good Laboratory Practices; maintaining genetic stability in the experimental animal colony; using precise definitions and standard nomenclature in protocols and publications; critically reviewing the experimental design, power analysis, variables, metrics, and data analysis; applying appropriate statistical methods; and reporting complete and transparent findings. Improved reproducibility will extend the utility and predictive value of animal models of disease.[17]

INTERVALS: A Data- and Results-Sharing Platform Can Improve Transparency in Industry-Funded and Conducted Research

Sharing data in a way that enables reanalysis and reuse is not yet done in a systematic manner, despite the fact that it would clearly benefit the scientific community and society in general, especially when it relates to the scientific assessment of consumer products, including drugs. When the quality of science is questioned due to the affiliation of the scientists, the funding source, or the research topic, sharing data and results transparently is even more critical. For example, tobacco harm reduction is a controversial topic that would benefit greatly from the disclosure of methods, data, and results. Tobacco harm reduction at the population level depends in part on the availability of modified risk tobacco products (MRTPs) and their acceptance and use by smokers who do are unable or unwilling to quit smoking cigarettes.[19] Candidate MRTPs must undergo a scientific assessment process that includes comparison of their biological impact with that of cigarettes in nonclinical and clinical studies.

INTERVALS (https://www.intervals.science/) is an online platform developed by Philip Morris International Research & Development to demonstrate the scientific rigor, thoroughness, and precision required to assess the toxicology of candidate and potential MRTPs while enabling reuse of data sets, encouraging external verification of testing strategies, and informing the scientific community.[20]

In order to address reproducibility concerns, INTERVALS was built using the latest standards in data sharing and reproducible research, including FAIR (Findable, Accessible, Interoperable and Reusable) data principles, to gather detailed information on the design and conduct of studies. This should enable easy review of methods and results as well as reuse of data and the generation of new hypotheses.

The INTERVALS platform allows researchers to find relevant information on studies, (more) detailed protocols, and, most importantly, interoperable data files in a single platform to allow independent reanalysis of key findings, meta-analyses, and efficient data reuse. The INTERVALS toolbox also includes direct links to computational resources, such as AeroSolved (http://www.aerosolved.com/), a computational fluid dynamics code developed to simulate the generation, transport, evolution, and deposition of complex aerosol mixtures.

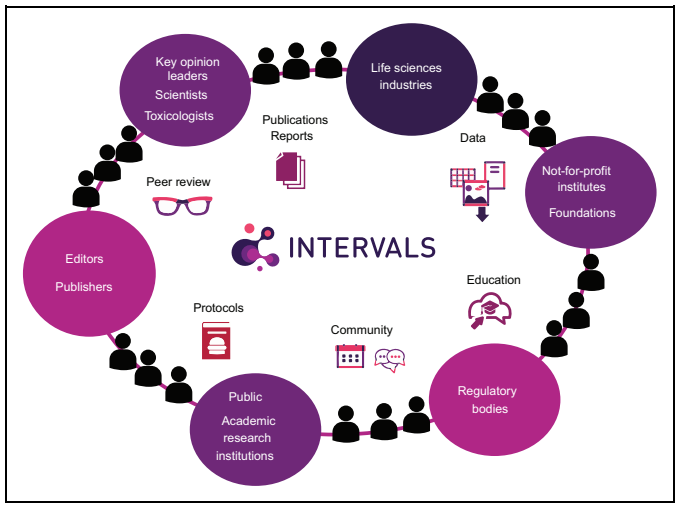

Figure 3. INTERVALS, a platform to enable transparency in tobacco harm reduction research.

INTERVALS also encourages communication by enabling constructive feedback on studies and protocols and will foster education by providing reference texts and media on diverse topics relevant to tobacco harm reduction and toxicology assessment of inhaled consumer products. With additional developments, INTERVALS is set to become a public repository for nonclinical and clinical data obtained during investigations into the toxicological effects of acute exposure and the course of disease progression following chronic exposure.[20] INTERVALS can serve as a hub for a community of scientists from industry, academia, nonprofit organizations, foundations, regulatory bodies, and publishers with a common interest in harm reduction (Figure 3).

Certainly as important as sharing data and results, the verification of methods used to generate them, as well as in-depth peer review of study design, results obtained, and interpretation thereof, are highly important. The sbv IMPROVER project (https://www.sbvimprover.com/)[21] was built to apply a crowdsourcing strategy to verify scientific processes and results in systems biology research in an industrial context. It has already proven useful for the collaborative discovery of signatures for specific diseases[22] or exposure status[23] and to investigate issues of general interest, including the translatability of rodent data to humans.[24] To date, the sbv IMPROVER crowdsourcing collaborations have included more than 600 scientists and have led to the successful completion of 4 challenges and 3 datathons as well as the publication of 15 manuscripts in peer-reviewed journals. To complement this challenge-based benchmarking of methods, in-depth peer review of studies, managed by a third party, are conducted. Panels of experts organized in consecutive peer-review panels are given access to study reports and/or publications as well as to the INTERVALS platform to get the best

possible overview of the research. They can anonymously and independently (via web) give their unbiased opinion on study design choices, study conduct appropriateness, quality of data, and objectivity of results interpretation.

Summary

Improving the reproducibility of research findings should be possible with the adoption of validated reagents, rigorous study design and analysis, adherence to SOPs, verification of methods with tools such as sbv IMPROVER, and transparency in sharing data and methods with platforms such as INTERVALS. Achieving high levels of reproducibility will require the cooperation and adoption of best practices by all stakeholders: researchers, academic institutions, funding entities, journals, industry, regulatory agencies, and nonprofit and professional groups. The establishment and adherence to validation and verification standards would optimize the use of research funds and facilitate the leveraging of research findings to meet therapeutic targets. This may be accomplished with a collaborative and transparent approach in industry-funded and conducted research to enable evidence-based decision making.

Authors' Note

Boue, S. and Hoeng, J. contributed to conception and design, contributed to interpretation, drafted manuscript, and critically revised manuscript; Byrne, M. and Hayes, A.W. contributed to conception and design, contributed to interpretation, and critically revised manuscript; Peitsch, M.C. contributed to conception, contributed to interpretation, and critically revised manuscript. All authors gave final approval and agreed to be accountable for all aspects of work ensuring integrity and accuracy. Stéphanie Boué, Julia Hoeng, and Manuel C. Peitsch are employees of Philip Morris International.

Acknowledgments

We thank Charles W. Frevert, DVM, ScD, for his contribution to the session. We also thank Dean Meyer, PhD, ELS, of Edanz Medical Writing, for providing medical writing support, which was funded by Philip Morris International.

Presentation

Some of the information in this communication was previously presented as an exhibitor-hosted program at the American College of Toxicology's 38th Annual Meeting, Palm Springs, CA, USA, on November 5, 2017.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Stéphanie Boué, Julia Hoeng, and Manuel C. Peitsch are employees of Philip Morris International.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: A. Wallace Hayes was funded by Philip Morris International for his part in organizing the American College of Toxicology session and for manuscript preparation.

ORCID iD

Stéphanie Boué http://orcid.org/0000-0003-0680-4168

References

1. Hewitt SM.Reproducibility: it is just good science. *J Histochem Cytochem*. 2016;64(4):223.
2. Begley CG, Ioannidis JPA. Reproducibility in science. *Improving Stand Basic Preclin Res*. 2015;116(1):116-126.
3. Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490(7419):187-191.
4. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. *PeerJ*. 2013;1:e175.
5. Xia J, Liu Y. Usage patterns of open genomic data. *Coll Res Libraries*. 2013;74(2):195-206.
6. Tenenbaum JD, Sansone SA, Haendel M. A sea of standards for omics data: sink or swim? *J Am Med Inform Assoc*. 2014;21(2): 200-203.
7. Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014;505(7485):612-613.
8. Butler D. Wellcome Trust launches open-access publishing venture. *Nat News*. 06 July 2016. DOI: doi:10.1038/nature.2016.20220
9. Announcement: Towards greater reproducibility for life-sciences research in nature. *Nature*. 2017;546(7656):8.
10. Bloom T, Ganley E, Winker M. Data access for the open access literature: PLOS's data policy. *PLoS Biol*. 2014;12(2): e1001797.
11. Stodden V, Guo P, Ma Z. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PLoS One*. 2013;8(6):e67111.
12. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: progress and priorities. *F1000Res*. 2017;6:604.
13. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. *PLoS Biol*. 2015;13(6): e1002165.
14. Bradbury A, Pluckthun A. Reproducibility: standardize antibodies used in research. *Nature*. 2015;518(7537):27-29.
15. Voskuil JL. The challenges with the validation of research antibodies. *F1000Resh*. 2017;6:161.
16. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10(9):712.
17. Zeiss CJ, Johnson LK. Bridging the gap between reproducibility and translation: data resources and approaches. *ILAR J*. 2017; 58(1):1-3.
18. Denayer T, Stöhr T, Van Roy M. Animal models in translational medicine: validation and prediction. *New Horiz Transl Med*. 2014;2(1):5-11.
19. Smith MR, Clark B, Ludicke F, et al. Evaluation of the tobacco heating system 2.2. Part 1: description of the system and the

scientific assessment program. *Regul Toxicol Pharmacol*. 2016; 81(suppl 2):S17-S26.

20. Boue S, Exner T, Ghosh S, et al. Supporting evidence-based analysis for modified risk tobacco products through a toxicology data-sharing infrastructure. *F1000Res*. 2017;6:12.
21. Meyer P, Alexopoulos LG, Bonk T, et al. Verification of systems biology research in the age of collaborative competition. *Nat Biotechnol*. 2011;29(9):811-815.
22. Tarca AL, Lauria M, Unger M, et al. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. *Bioinformatics*. 2013;29(22):2892-2899.
23. Poussin C, Belcastro V, Martin F, Boue S, Peitsch MC, Hoeng J. Crowd-sourced verification of computational methods and data in systems toxicology: a case study with a heat-not-burn candidate modified risk tobacco product. *Chem Res Toxicol*. 2017;30(4): 934-945.
24. Rhrissorrakrai K, Belcastro V, Bilal E, et al. Understanding the limits of animal models as predictors of human biology: lessons learned from the sbv IMPROVER species translation challenge. *Bioinformatics*. 2015;31(4):471-483.