Policy Forum

DOI: 10.5582/ddt.2026.01031

Are we ready to replace animal models? A perspective on regulatory challenges of human-relevant drug testing systems

Tanbin Liu, Hongzhou Lu*

The Third People's Hospital of Shenzhen and Second Affiliated Hospital Southern University of Science and Technology, Shenzhen, China.

SUMMARY: This perspective examines whether animal models can be gradually replaced by human-relevant New Approach Methodologies (NAMs) in drug development and regulatory evaluation, synthesizing recent advances across hepatology, oncology, neurotoxicology, and cardiac safety assessment. Although the relevant technologies have become increasingly mature, their acceptance at the regulatory level is still lagging due to factors such as limited standardization, insufficient reproducibility of results, and unclear approval pathways. Building on a systematic comparison of existing evidence, this article proposes a three-stage transition pathway: the first stage (auxiliary stage)—using NAMs in parallel with animal experiments; the second stage (partial replacement stage)—using NAMs as the preferred efficacy evaluation model for human-specific diseases; and the third stage (regulatory integration stage). Only through gradual validation can animal models be replaced responsibly, thereby increasing the success rate of translational research and reducing the use of animal experiments. Overcoming the persistent challenges of limited standardization, poor reproducibility, and unclear regulatory pathways is the central prerequisite for this transition.

Keywords: New Approach Methodologies (NAMs), regulatory acceptance, three stage roadmap, translational research

1. Introduction

For decades, animal models have served as the gold standard for preclinical drug evaluation. However, they have consistently demonstrated limited capacity to predict human therapeutic responses. This limitation is particularly pronounced in areas such as infectious diseases, immunology, and hepatology. A salient illustration is oncology, where overall clinical trial failure rates exceed 90%—a figure that, in part, reflects the inadequate translational validity of conventional animal-based preclinical systems (*1-3*). In parallel, mounting scientific efforts and increasingly rigorous ethical standards have catalyzed rapid advances in human-relevant *in vitro* research platforms, encompassing organ-on-a-chip technologies, induced pluripotent stem cell (iPSC)-derived models, and computational toxicology approaches (*4,5*). The formal recognition of New Approach Methodologies (NAMs) by the U.S. Food and Drug Administration (FDA) marks a significant inflection point in this trajectory, signaling a broader paradigm shift in regulatory thinking (*6,7*). Despite this momentum, a critical bottleneck persists: the transition from technological feasibility to regulatory acceptance remains incomplete. In the absence of clear, internationally harmonized validation and qualification frameworks, NAMs risk being relegated to exploratory research contexts rather than being adopted as recognized standards for regulatory submissions. Bridging this regulatory gap is therefore essential to realizing the full translational potential of these advanced methodologies.

Unlike prior reviews that focused primarily on technological advances in organoids or organ-on-a-chip platforms, the present article specifically focuses on the transition pathway from technological maturity to regulatory acceptance. We propose a three-stage model stratified by disease type and the maturity of NAM platforms and emphasize a fit for purpose validation framework rather than a binary argument for or against complete animal model replacement. This perspective is intended to bridge the gap between scientific innovation and regulatory practice.

2. Why animal models fall short: A clinical perspective

From the perspective of clinical researchers, the gap between animal trial outcomes and human clinical

effectiveness is striking. In the realm of liver disorders, numerous potential NASH medications that demonstrated effectiveness in mouse models were found to be useless in human clinical trials (*8,9*); similarly, hepatitis B virus (HBV)—owing to its highly restricted species tropism, rendering it able to efficiently infect only humans and higher primates—lacks acceptable animal models, markedly hampering preclinical antiviral testing (*10*). In the field of neurodegenerative disease research, dozens of potential drugs for Alzheimer's disease that alleviated amyloid pathology in transgenic mice ultimately failed to result in cognitive improvement in human patients (*11-13*). In oncology, mouse xenograft models typically fail to correctly represent the human tumor microenvironment and immune evasion mechanisms, explaining why many immunotherapies that are effective in mice have little impact on people (*14-16*). Even in the study of infectious diseases and sepsis, rodent models fail to replicate the unique cytokine responses and Toll-like receptor signaling pathways in humans, leading to repeated setbacks in clinical translation (*17-19*). This situation is not uncommon in biomedical research, and a well known adage succinctly captures this persistent disconnect: "The drug candidate demonstrated marked preclinical efficacy but failed to be clinically useful."

These failures share a common mechanistic basis. First, there are significant differences in the immune system: the evolutionary pathways of the innate and adaptive immune systems in mice and humans are vastly different; mice's responses to viral infections or checkpoint inhibitors often differ greatly from actual responses in patients (*20,21*). Secondly, the viral life cycle is species-specific: HBV, HCV, and many emerging viruses cannot replicate in conventional laboratory animals without extensive genetic modification. Third, there are significant differences in drug metabolism kinetics and toxicity characteristics: a drug that is metabolized safely in the rat liver may produce toxic metabolites in human liver cells, and *vice versa*. These differences are not merely theoretical; they have directly led to several catastrophic drug recall events, such as the cytokine storm caused by TGN1412: this compound appeared safe in monkey experiments but nearly resulted in the death of human volunteers (*22*).

3. The emerging approach: Human-relevant efficacy modeling

In recent years, remarkable developments in different therapeutic domains have revealed that human-related biological systems are no longer simple conceptual assumptions but have become mature platforms confirmed by functional testing. In the area of hepatology, a recent study effectively replicated the whole life cycle of the hepatitis B virus (HBV) from invasion to antiviral response using liver organoids grown from iPSCs (*23*). In the field of immuno-oncology, microphysiological systems constructed based on patient-derived organoids, perfusable vascular networks, and tumor-associated macrophages have confirmed the macrophage-mediated immunosuppressive effect of immunotherapies in various solid tumor models, including prostate cancer and hepatocellular carcinoma (*24-26*). In cardiac toxicity research, a beating heart-on-a-chip platform based on human iPSC-derived cardiomyocytes achieved an accuracy of 91.6% in detecting drug-induced QT interval prolongation (*27*). In the realm of neurotoxicology, brain organoids containing iPSC-derived microglia efficiently reproduced the microglia-mediated inflammatory response to developing neurotoxins, a route commonly missed in normal animal research (*28*).

These examples highlight a shift in validation logic: from "Does it behave like a sick animal?" to "Can it predict clinical outcomes in humans?" However, not all NAM platforms are equally mature. We propose a simple two-tier stratification:

Tier 1 (relatively mature): iPSC-cardiomyocyte models for cardiotoxicity (CiPA paradigm, > 90% retrospective accuracy) and liver organoids/chips for DILI efficacy. These have received positive regulatory signals in defined contexts of use.

Tier 2 (exploratory): Multi-organ systems, neuroimmune models, and vascularized tumor chips. These suffer from low throughput, high batch-to-batch variability, and lack of cross-laboratory validation. They remain research tools and are not regulatory-ready.

4. The real problem: Technological readiness does not equal regulatory readiness

However, human-relevant platforms cannot fully replace animal models in the short term. An objective assessment must acknowledge several crucial shortcomings. Even for Tier 1 platforms, organoid and organ on a chip technologies vary significantly across laboratories (*e.g.*, extracellular matrices, culture media, differentiation protocols, and endpoints), making direct comparison of results difficult. Second, the level of batch-to-batch variability is high: even within the same laboratory, iPSC derived organoids often lack consistency in size, cellular composition, and functional maturity. Third, there is a lack of long-term toxicity and systemic response data. Most current platforms operate for only a few days to weeks and cannot integrate the neuroendocrine-immune axis, limiting their ability to assess adverse reactions such as delayed cardiotoxicity, cytokine release syndrome, and tissue remodeling. Standard liver chips typically support culture for only 21–28 days, insufficient to capture drug induced liver injury (DILI) that may manifest after months of treatment. Similarly, kidney on-a-chip models are largely restricted to acute injury endpoints, with few capable of recreating progressive fibrosis or chronic tubular atrophy. These gaps represent major hurdles for regulatory acceptance, as many drug withdrawals are

due to late onset toxicities not detectable in short term assays. Finally, regulatory certification pathways remain unclear. Although the U.S. FDA's NAM program is a positive signal, no global consensus exists on what validation data suffice to replace animal testing for specific indications. Thus, technological maturity does not equal regulatory maturity. Without coordinated efforts to standardize production processes, ensure the reproducibility of benchmark results, and establish fit for purpose certification frameworks, human based systems risk becoming scientific curiosities rather than regulatory tools. Acknowledging these shortcomings does not negate the value of NAMs but rather points the way toward the next critical improvements. For Tier 2 platforms, these challenges are even more pronounced.

5. Current regulatory landscape of NAMs

A systematic comparison of major regulatory bodies reveals significant heterogeneity in NAM acceptance. The U.S. FDA has taken the most proactive stance through its NAM program and ISTAND initiative, issuing draft guidance on liver and cardiac safety models and accepting organoid data for certain IND submissions (*e.g.*, rare disease gene therapies). The European Medicines Agency (EMA) has incorporated NAMs into its 3Rs strategy and published a reflection paper on organ-on-chip technology, but it requires parallel animal data for most indications. Japan's Pharmaceuticals and Medical Devices Agency (PMDA) has supported the development of iPSC-based cardiotoxicity assays (CiPA) and accepts them as supplemental evidence, and yet it retains animal testing as the default standard. China's National Medical Products Administration (NMPA) has recently issued guidance on organoid research but has not formally recognized any NAM platform for regulatory submission; ongoing pilot projects focus on liver toxicity screening. The International Council for Harmonisation (ICH) has not issued unified NAM guidelines, though ICH S5 (reproductive toxicology) and ICH M3 (nonclinical safety) are undergoing review for potential updates. This fragmented landscape underscores the urgent need for global coordination.

6. A path forward: The three phase transition model

We are not yet ready for a complete replacement—but the transition is both necessary and achievable. This paper therefore proposes a three-phase roadmap for the responsible integration of NAMs into regulatory practice as shown in Figure 1. Given that technological maturity and regulatory readiness vary across disease contexts and platform types, the phases need not advance simultaneously across all therapeutic areas; the timelines below (~2028, 2028-2033, 2033 and beyond) are illustrative projections intended to guide planning and

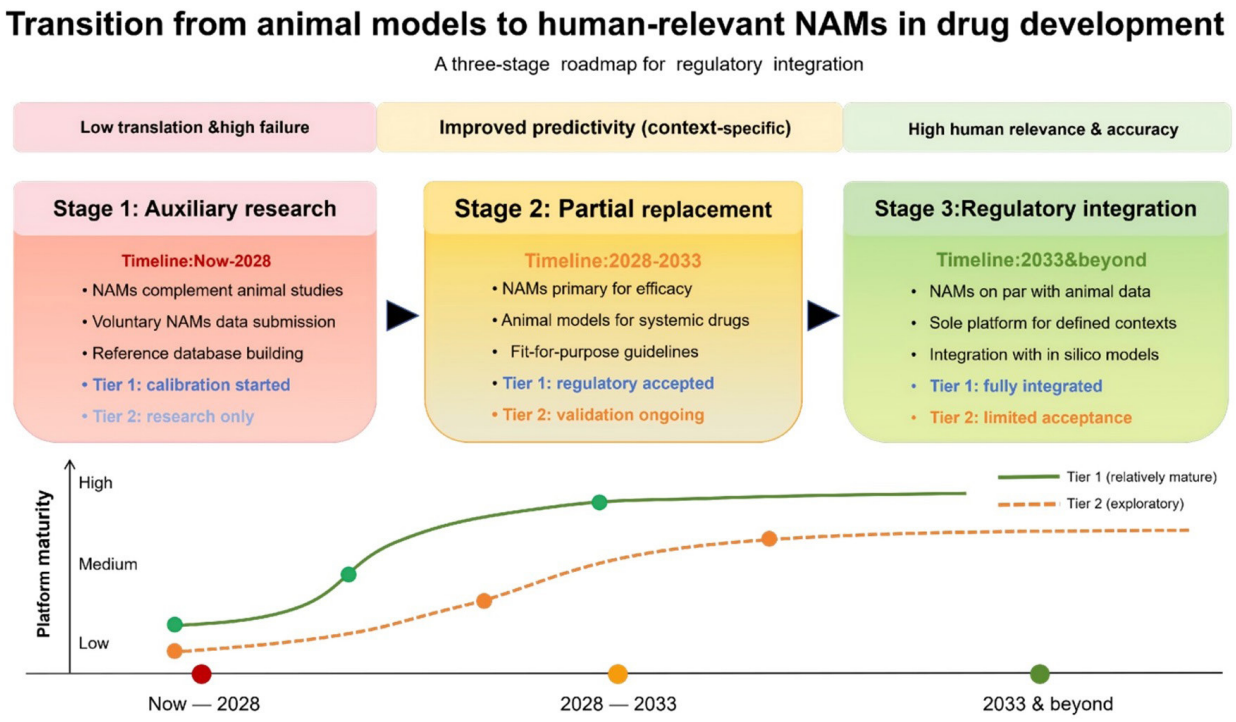

Figure 1. Proposed three phase roadmap for integrating NAMs into regulatory drug development. Stage 1 (auxiliary, ~2028): NAMs used in parallel with animal studies; voluntary data submission; reference databases established; Tier 1 platforms begin calibration. Stage 2 (partial replacement, 2028–2033): NAMs become the primary efficacy model; fit-for-purpose guidelines issued; Tier 1 platforms accepted, Tier 2 platforms undergo validation. Stage 3 (full integration, 2033+): NAM data on par with animal data for IND submissions; sole platform for defined contexts; integrated with in silico models; ICH guidelines harmonized. Timelines are illustrative; Tier 1 (relatively mature) platforms progress faster than Tier 2 (exploratory) platforms.

are not fixed mandates.

Stage 1: Auxiliary research (near term, illustratively by ~2028)

In this phase, NAMs function as complementary tools to conventional animal research. For each candidate drug, sponsors need to concurrently conduct conventional animal efficacy and toxicology studies as well as a succession of human-relevant NAM assessments—such as using iPSC-derived liver organoids or liver chips to demonstrate effectiveness and early toxicity analysis for liver-targeted therapies. Regulatory bodies do not yet accept NAM data alone, but they will begin to develop reference databases to study the association between organoid responses and human efficacy. This phase involves extremely minimal regulatory revisions, including a proposal for sponsors to voluntarily provide NAM data. The fundamental purpose is to acquire real-world evidence: When organoid models predict effectiveness, do they have a greater ability to predict the outcomes of Phase II clinical trials compared to animal experiments? During this stage, efforts should also begin to establish reference standards and interlaboratory calibration protocols for highly mature NAMs (*e.g.*, liver chips and iPSC cardiomyocytes), whereas less mature platforms (*e.g.*, multi organ systems and neuroimmune models) would remain primarily research tools.

Stage 2: Partial replacement (mid term, illustratively 2028–2033)

By the late 2020s, specific therapeutic areas can, after thorough validation, gradually transition to a partial replacement model. For human-specific pathogens such as HBV and human immunodeficiency virus (HIV) as well as newly emerging viruses that lack natural animal hosts, NAMs should become the primary efficacy evaluation model. In contrast, for immunotherapies or drugs requiring complex systemic pharmacokinetic assessment, animal models remain essential in Stage 2, with NAMs serving in a supportive role. Similarly, in drug screening, for certain categories of compounds (such as direct-acting antiviral drugs and low-molecular-weight compounds with simple metabolic pathways), patient-derived organoid panels with diverse genetic backgrounds could potentially reduce or replace the use of dogs and rats, pending systematic validation and regulatory qualification. Regulatory agencies will issue clear guidelines: for example, "For HBV antiviral drugs, new drug clinical trial (IND) applications can rely on efficacy data obtained from organoids; efficacy evaluation does not require the use of animal models and is only needed to assess safety endpoints that NAMs cannot yet cover". During this stage, we encourage regulatory agencies to define fit-for-purpose validation criteria for specific NAM platforms, recognizing that not all NAMs are equally mature.

Stage 3: Integration into the regulatory framework (long term, illustratively 2033 and beyond)

In the final stage, NAMs are expected to become a standard and widely recognized core component of new drug clinical trial (IND) submission materials. Organoid and organ-on-a-chip data will be on par with animal experiment data, with both serving as sources of evidence, each with its unique advantages and limitations. For certain clearly defined situations (such as drugs metabolized by the liver or viruses that directly cause cell damage), NAMs can even serve as the sole platform for evaluating efficacy and safety, thereby completely replacing animal experiments. In parallel, in silico models—such as quantitative systems pharmacology (QSP), physiologically based pharmacokinetic (PBPK) modeling, and machine learning-based toxicity predictors—are expected to be integrated with organoid and organ on-a-chip data, collectively comprising a comprehensive NAM toolbox for regulatory decision making. Achieving this goal not only requires continued maturation of the technology but also necessitates the revision of relevant ICH guidelines and the promotion of global coordination among the NMPA, FDA, EMA, and PMDA. Although the transformation process may be relatively slow, as long as we start now to systematically build an evidence base, this goal is entirely achievable.

Figure 1 shows the three stage transition model. The timelines (~2028, 2028–2033, 2033 and beyond) are illustrative and assume differential progression across Tier 1 and 2 platforms. Tier 1 platforms (*e.g.*, liver chips and iPSC cardiomyocytes) are expected to enter Stage 2 earlier than Tier 2 platforms (*e.g.*, multi organ systems).

7. Limitations of this perspective

Our proposed roadmap is intentionally optimistic. It may underestimate the lasting need for animal models of complex systemic diseases (*e.g.*, autoimmune disorders and chronic neuroinflammation) where human-relevant platforms are still immature. Global regulatory harmonization remains a formidable political and logistical challenge. This perspective is meant to stimulate discussion, not to serve as a regulatory template.

8. Conclusion

"Are we ready to replace animal models?" This question seems to elicit a binary answer, but reality indicates that the answer lies along a gradient. From a technical perspective, human-relevant systems such as liver organoids have, in defined contexts of use, demonstrated performance that can surpass animal models. Research on the HBV liver organoid model is just one of many

signs that in the future we will no longer rely on mice or monkeys to replace human patients. However, the regulatory system that was meticulously built over half a century based on animal data has yet to keep pace with this change. As mentioned at the beginning, the real bottleneck is increasingly shifting from the capacity for innovation to acceptance by the regulatory system.

We cannot simply ask regulatory agencies to abandon their cautious approach—their mission is to protect public health and to ensure incidents like TGN1412 do not happen again. But by adopting a three-stage model, we can accelerate this transition process: first as an auxiliary means, then achieving partial replacement, and finally achieving full regulatory integration. The path is clear, and the tools are in place. What we need now is a shared commitment from researchers, sponsors, and regulators to move forward with caution and avoid unnecessary delays. Achieving this vision requires not only continued technological refinement but also harmonized regulatory frameworks across the FDA, EMA, PMDA, NMPA, and ICH, as outlined in Section 5.

Funding: This work was supported by grants from the National Key R&D Program of China (2025ZD01904500), the Special Funds for Strategic Emerging Industry in Shenzhen (Grant No. F-2022-Z99-502266), the Shenzhen Clinical Research Center for Emerging Infectious Diseases (No. LCYSSQ20220823091203007), the Guangdong Key Lab for Diagnosis & Treatment of Emerging Infectious Diseases, and the Shenzhen Key Laboratory for Infectious Diseases.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Maeda H, Khatami M. Analyses of repeated failures in cancer therapy for solid tumors: Poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs. Clin Transl Med. 2018; 7:11.
2. Van Norman GA. Limitations of animal studies for predicting toxicity in clinical trials: Is it time to rethink our current approach? JACC Basic Transl Sci. 2019; 4:845-854.
3. Mak IW, Evaniew N, Ghert M. Lost in translation: Animal models and clinical trials in cancer treatment. Am J Transl Res. 2014; 6:114-118.
4. Ingber DE. Human organs-on-chips for disease modelling, drug development and personalized medicine. Nat Rev Genet. 2022; 23:467-491.
5. Youhanna S, Kemas AM, Preiss L, Zhou Y, Shen JX, Cakal SD, Paqualini FS, Goparaju SK, Shafagh RZ, Lind JU, Sellgren CM, Lauschke VM. Organotypic and microphysiological human tissue models for drug discovery and development-Current state-of-the-art and future perspectives. Pharmacol Rev. 2022; 74:141-206.
6. Avila AM, Bebenek I, Bonzo JA, *et al*. An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs). Regul Toxicol Pharmacol. 2020; 114:104662.
7. Yao J, Peretz J, Bebenek I, Avila A, Alapatt T, Lee B, Patel D, Brown P, Davis-Bruno K. FDA/CDER/OND experience with New Approach Methodologies (NAMs). Int J Toxicol. 2026; 45:136-156.
8. Hunter H, de Gracia Hahn D, Duret A, *et al*. Weight loss, insulin resistance, and study design confound results in a meta-analysis of animal models of fatty liver. Elife. 2020; 9:e56573.
9. Gallage S, Avila JEB, Ramadori P, Focaccia E, Rahbari M, Ali A, Malek NP, Anstee QM, Heikenwalder M. A researcher's guide to preclinical mouse NASH models. Nat Metab. 2022; 4:1632-1649.
10. Liu Y, Maya S, Ploss A. Animal models of hepatitis B virus infection-Success, challenges, and future directions. Viruses. 2021; 13:777.
11. Granzotto A, Vissel B, Sensi SL. Lost in translation: Inconvenient truths on the utility of mouse models in Alzheimer's disease research. Elife. 2024; 13:e90633.
12. Weekman EM, Sudduth TL, Caverly CN, Kopper TJ, Phillips OW, Powell DK, Wilcock DM. Reduced efficacy of anti-Abeta immunotherapy in a mouse model of amyloid deposition and vascular cognitive impairment comorbidity. J Neurosci. 2016; 36:9896-9907.
13. Kodamullil AT, Iyappan A, Karki R, Madan S, Younesi E, Hofmann-Apitius M. Of Mice and men: Comparative analysis of neuro-inflammatory mechanisms in human and mouse using cause-and-effect models. J Alzheimers Dis. 2017; 59:1045-1055.
14. Lonberg N. The Problem with syngeneic mouse tumor models. Cancer Immunol Res. 2025; 13:456-462.
15. Floe LVB, Pedersen MG, Moller BK. Animal models in preclinical evaluation of CAR-T cell therapy: Advantages and limitations. Biochim Biophys Acta Rev Cancer. 2025; 1880:189455.
16. Hendriks H. Advancing oncology drug development: Innovative approaches to enhance success rates while reducing animal testing. Biochim Biophys Acta Rev Cancer. 2025; 1880:189467.
17. Seok J, Warren HS, Cuenca AG, *et al*. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A. 2013; 110:3507-3512.
18. Stortz JA, Raymond SL, Mira JC, Moldawer LL, Mohr AM, Efron PA. Murine models of sepsis and trauma: Can we bridge the gap? ILAR J. 2017; 58:90-105.
19. Jameson SC, Masopust D. What is the predictive value of animal models for vaccine efficacy in humans? Reevaluating the potential of mouse models for the human immune system. Cold Spring Harb Perspect Biol. 2018; 10:a029132.
20. Masubuchi T, Chen L, Marcel N, *et al*. Functional differences between rodent and human PD-1 linked to evolutionary divergence. Sci Immunol. 2025; 10:eads6295.
21. Zschaler J, Schlorke D, Arnhold J. Differences in innate immune response between man and mouse. Crit Rev Immunol. 2014; 34:433-454.
22. Attarwala H. TGN1412: From discovery to disaster. J Young Pharm. 2010; 2:332-336.
23. Liu T, Xu J, Chen X, Li J, Ke J, Xu J, Lu H, Wu F. Modeling of the hepatitis B virus life cycle and the efficacy of antivirals in human iPSC-derived hepatic

organoids. Biosci Trends. 2026; 20:235-244.
24. Bains RS, Raju TG, Semaan LC, Block A, Yamaguchi Y, Priceman SJ, George SC, Shirure VS. Vascularized tumor-on-a-chip to investigate immunosuppression of CAR-T cells. Lab Chip. 2025; 25:2390-2400.
25. Zou Z, Lin Z, Wu C, Tan J, Zhang J, Peng Y, Zhang K, Li J, Wu M, Zhang Y. Micro-engineered organoid-on-a-chip based on mesenchymal stromal cells to predict immunotherapy responses of HCC patients. Adv Sci (Weinh). 2023; 10:e2302640.
26. Sanchez-de-Diego C, Yada RC, Sethakorn N, *et al*. Engineering the bone metastatic prostate cancer niche through a microphysiological system to report patient-specific treatment response. Commun Biol. 2025; 8:961.
27. Visone R, Lozano-Juan F, Marzorati S, Rivolta MW, Pesenti E, Redaelli A, Sassi R, Rasponi M, Occhetta P. Predicting human cardiac QT alterations and pro-arrhythmic effects of compounds with a 3D beating heart-on-chip platform. Toxicol Sci. 2023; 191:47-60.
28. Yuan NY, Richards WD, Parham KT, Clark SG, Greuel K, Polzin B, Smith SW, Lebakken CS. Neural organoids incorporating microglia to assess neuroinflammation and toxicities induced by known developmental neurotoxins. Curr Res Toxicol. 2025; 9:100252.

Received May 10, 2026; Revised June 9, 2026; Accepted June 16, 2026.

**Address correspondence to:*
Hongzhou Lu, The Third People's Hospital of Shenzhen and Second Affiliated Hospital Southern University of Science and Technology, Shenzhen, China.
E-mail: luhongzhou@szsy.sustech.edu.cn

Released online in J-STAGE as advance publication June 23, 2026.