

# Drug approval prediction based on the discrepancy in gene perturbation effects between cells and humans

Minhyuk Park,<sup>a</sup> Donghyo Kim,<sup>a</sup> Inhae Kim,<sup>b</sup> Sin-Hyeog Im,<sup>a,b</sup> and Sanguk Kim<sup>a,\*</sup>

<sup>a</sup>Department of Life Sciences, Pohang University of Science and Technology, Pohang, South Korea

<sup>b</sup>ImmunoBiome Inc., Pohang, South Korea



## Summary

**Background** Poor translation between *in vitro* and clinical studies due to the cells/humans discrepancy in drug target perturbation effects leads to safety failures in clinical trials, thus increasing drug development costs and reducing patients' life quality. Therefore, developing a predictive model for drug approval considering the cells/humans discrepancy is needed to reduce drug attrition rates in clinical trials.

**Methods** Our machine learning framework predicts drug approval in clinical trials based on the cells/humans discrepancy in drug target perturbation effects. To evaluate the discrepancy to predict drug approval (1404 approved and 1070 unapproved drugs), we analysed CRISPR-Cas9 knockout and loss-of-function mutation rate-based gene perturbation effects on cells and humans, respectively. To validate the risk of drug targets with the cells/humans discrepancy, we examined the targets of failed and withdrawn drugs due to safety problems.

**Findings** Drug approvals in clinical trials were correlated with the cells/humans discrepancy in gene perturbation effects. Genes tolerant to perturbation effects on cells but intolerant to those on humans were associated with failed drug targets. Furthermore, genes with the cells/humans discrepancy were related to drugs withdrawn due to severe side effects. Motivated by previous studies assessing drug safety through chemical properties, we improved drug approval prediction by integrating chemical information with the cells/humans discrepancy.

**Interpretation** The cells/humans discrepancy in gene perturbation effects facilitates drug approval prediction and explains drug safety failures in clinical trials.

**Funding** S.K. received grants from the Korean National Research Foundation (2021R1A2B5B01001903 and 2020R1A6A1A03047902) and IITP (2019-0-01906, Artificial Intelligence Graduate School Program, POSTECH).

**Copyright** © 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Machine learning; Drug approval; Drug safety; Clinical translation; Gene perturbation effect; Discrepancy

## Introduction

Drug attrition rates can be effectively reduced by developing a predictive model for drug approval before clinical trials. Many drug candidates fail in clinical trials because of safety issues, resulting in a considerable increase in drug development costs.<sup>1–3</sup> In addition, drug safety failures in clinical trials degrade the quality of life for patients, potentially causing additional health problems. One possible explanation for safety failures is related to drug-induced perturbations of target genes (either intended or off-targets) that can incur unacceptable risks to patients.<sup>4</sup> Thus, previous studies have explored the relationship between the side effects of drugs and the characteristics of drug targets, including the centrality of the protein interaction network<sup>5–7</sup> and gene expression profiles across

tissues.<sup>7,8</sup> In particular, genetic perturbation effects, which measure intolerance to genetic variation (e.g., CRISPR-Cas9 knockout), have been utilised to evaluate the safety of drug targets by assuming that targeting genes with intolerance to such effects is likely to cause severe side effects.<sup>5</sup> For example, genes that are intolerant to genetic perturbation effects on multiple cell types were difficult to be regarded as therapeutic targets due to potential toxicity when targeted by drugs.<sup>9,10</sup>

However, the discrepancy in drug target perturbation effects between *in vitro* (cells) and clinical studies (humans) is an important consideration to understand the safety failures of drugs in clinical trials. Even if the perturbation effect of a drug target is tolerant on cells, intolerance to the perturbation effects on humans may

eBioMedicine

2023;94: 104705

Published Online xxx

<https://doi.org/10.1016/j.ebiom.2023.104705>

1016/j.ebiom.2023.104705

104705

\*Corresponding author. Department of Life Sciences, Pohang University of Science and Technology, Pohang, 790-784, South Korea.

E-mail address: [sukim@postech.ac.kr](mailto:sukim@postech.ac.kr) (S. Kim).

### Research in context

#### Evidence before this study

The safety failure of drugs in clinical trials has negative effects on drug development, resulting in wasted resources, both in terms of time and money, as well as the reduced quality of life for patients. Effects of genetic perturbation, such as CRISPR-Cas9 screening in cells, have been utilised as a feature of target genes to assess drug safety, assuming that targeting genes with intolerance to such effects may lead to severe side effects. However, the discrepancy in drug target perturbation effects between cells and humans is an important consideration for understanding the safety failures of clinical trials due to poor translation between *in vitro* and clinical studies. Recently, the genetic perturbation effects on humans were estimated by modelling loss-of-function mutation rates in a large population, facilitating the comparison of the effects on cells and humans. These effects have the potential to be used to assess the safety of drugs and therapeutic targets in humans.

#### Added value of this study

We predicted drug approval using machine learning and evaluated the discrepancy in drug target perturbation effects

between cells and humans with over 2000 unapproved and approved drugs in clinical trials. We validated that genes with the cells/humans discrepancy in perturbation effects were likely to be risky drug targets in various drug datasets with safety issues, including failed and withdrawn drugs with severe side effects. Furthermore, the integration of our target gene-centric approach with the chemical-centric approach has improved the prediction of drug approval by complementing each other. Our analyses with large datasets can contribute to a significant advance in the drug discovery field. We provide our machine learning model for predicting drug approval to assess the safety of new drugs and target genes in clinical trials.

#### Implications of all the available evidence

Our discrepancy-based machine learning framework, which predicts drug approval and assesses the safety of drugs and target genes in clinical trials, has the potential to effectively reduce drug attrition rates during drug development. The results of our study provide biological insights for developing new candidate drugs with minimal toxicity and identifying safe therapeutic drug targets for patients.

cause severe safety problems, resulting in drug failures in clinical trials. For example, sibutramine, an appetite suppressant, does not affect cell viability.<sup>11</sup> However, it was withdrawn from several markets in 2010 due to its serious side effects, such as heart attack and stroke.<sup>12</sup> Indeed, *SLC6A3*, an off-target of sibutramine, has been implicated in serious human diseases, such as neuropsychiatric diseases.<sup>13</sup>

In this study, we evaluated the discrepancy in drug target perturbation effects between cells and humans in predicting drug approval in clinical trials. We hypothesised that failed drugs in clinical trials due to safety issues were associated with genes that are tolerant to perturbation effects on cells but intolerant to perturbation effects on humans. As those drugs frequently showed severe side effects on humans, despite passing toxicity tests in preclinical cell culture models, evaluating the cells/humans discrepancy for drug approval is necessary to understand poor translation from preclinical to clinical studies. CRISPR-Cas9 knockout can be used to analyse the perturbation effects of genes on cells,<sup>9,14</sup> but this method is not suitable for analysing the effects on humans. Instead, the effects on humans can be estimated by modelling loss-of-function mutation rates on genes with exome and genome sequencing data obtained from a large human population.<sup>15,16</sup> Because genes that are intolerant to perturbation effects are highly conserved in the human population, genes that have a depletion of loss-of-function mutations may be intolerant to perturbation effects by negative selection. Conversely, genes that have an enrichment of loss-of-

function mutations may be tolerant to perturbation effects. These estimated effects have been used to assess the safety of drugs and therapeutic targets in humans.<sup>8,17</sup>

We created a machine learning model to predict drug approval in clinical trials based on the discrepancy in drug target perturbation effects between cells and humans. Moreover, we found that genes that are tolerant to perturbation effects on cells but intolerant to perturbation effects on humans tend to be risky drug targets, associated with failed and withdrawn drugs with severe side effects. Our method provides a useful framework for predicting drug approval and evaluating the safety of drugs and target genes in clinical trials.

## Methods

### Perturbation effects of genes on cells and humans

Perturbation effects of genes on cells (cellular gene essentiality [CGE]) were derived from the fitness scores in the ProjectSCORE<sup>9</sup> database. Fitness scores were analysed using genome-wide CRISPR-Cas9 screening in 324 cell lines. A fitness score greater than 0 indicated a fatal effect on the cell line. The CGE was calculated by averaging fitness scores across 324 cell lines for each gene. Genes with CGE  $\geq 0$  and  $< 0$  were identified as essential and non-essential genes in cells, respectively.

Perturbation effects of genes on humans (organismal gene essentiality [OGE]) were derived from the loss-of-function observed/expected upper bound fraction (LOEUF) in the gnomAD<sup>16</sup> database. LOEUF was estimated by modelling the mutation rate from

sequencing data for a population of ~140,000 humans and calculated by comparing the observed vs. expected number of mutations in the gene. For example, a low LOEUF, which represents a significantly depleted number of observed vs. expected mutations in the gene, is considered to be an intolerant perturbation effect based on negative selection. The OGE was calculated by subtracting the LOEUF of each gene from 1.996 (the maximum value of LOEUF for all genes) to align the direction with that of the CGE. According to the threshold of intolerant gene perturbation effects defined by Karczewski et al.<sup>16</sup> (LOEUF < 0.35), genes with OGE > 1.646 were considered essential genes, whereas genes with OGE ≤ 1.646 were considered non-essential genes.

### Drug and drug–target interaction datasets

We obtained 2474 drugs (1404 approved and 1070 unapproved drugs) with the highest phase in clinical trials (max phase) that each drug had proceeded from the ChEMBL<sup>18</sup> (v30) database (Table S1). According to the definition from the ChEMBL, drugs with max phase 4 were annotated as approved drugs, and drugs with max phases 1, 2, and 3 were annotated as unapproved drugs. Drugs were categorised based on the Anatomical Therapeutic Chemical (ATC) code (Fig. S1). Analogous compounds with similar chemical formulas (such as the salts, hydrates, and radioisotopes of the compounds) were removed to ensure minimal overlap in structural properties among the drugs (Fig. S2a–c). Please see the “Supplementary Note” for a detailed description of drug dataset curation.

Information on drug–target interactions was obtained from the STITCH 5<sup>19</sup> database (combined interaction score ≥ 700, threshold of high confidence for drug–target interactions). We compiled 1070 unapproved drugs and 1404 approved drugs with 7697 targets. The average number of drug targets per drug was 18.0 (Fig. S3a;  $18.0 \pm 3.5$ ; mean value ± 95% confidence interval) and the average number of drugs per drug targets was 5.8 (Fig. S3b;  $5.8 \pm 0.3$ ). Please see “Supplementary Note” for a detailed description of drug–target interaction curation.

### Mapping target information on drugs using drug–target interaction

In addition to the CGE and OGE, we mapped the network (Network) and expression (Expression) information of drug targets on drugs using drug–target interactions. The CGE, OGE, Network, and Expression were mapped onto each drug by calculating the average of these values for drug targets. We validated that the average approach performed better than using the median or maximum values for predicting drug approval (Fig. S4). Therefore, the average approach was used in our study. Only drugs with CGE, OGE, Network, and Expression data available for more than 90% of drug targets were included in our analyses.

For network information on drug targets, we calculated the network degree and betweenness centrality of a protein–protein interaction network, which was obtained from the STRING<sup>20</sup> (v11.5) database (9606.protein.physical.links.v11.5.txt.gz). The network degree and betweenness centralities were calculated using NetworkX (v2.4) in Python.

For the expression information of drug targets, RNA tissue expression profiles were obtained from the Human Protein Atlas<sup>21</sup> (HPA v20.1) database (rna\_tissue\_consensus.tsv.zip). The expression profiles of 19,670 genes in 62 tissues were obtained. We calculated the average expression level across tissues and tissue broadness using Shannon entropy with Scipy (v1.6.0) in Python.

We validated our data curation for drug target information by conducting a comparative analysis with previously reported feature attributes (Fig. S5; see “Supplementary Note” for a detailed description of the validation of curated drug target information). Then, two network centrality and two expression profile values were integrated because of their correlation (Fig. S6). Integrating correlated features by averaging feature values was proposed to eliminate redundancy.<sup>22</sup> The integration of two network centrality values was achieved by averaging the ranks of each gene, which were determined based on its ascending order of degree and betweenness centrality. The averaged ranks were scaled from 0 to 1 using MinMaxScaler of Scikit-learn in Python. The same calculations were used to integrate the two expression profile values. The integrated network centrality and expression profile value were used as network information (Network) and expression information (Expression), respectively.

### Machine learning for drug approval prediction

We trained a random forest classifier using Scikit-learn (v0.24.2) in Python<sup>23</sup> to predict drug approval. The classifier consisted of 1000 trees with default settings because the number of trees did not affect the prediction performance significantly (Fig. S7a and b; see “Supplementary Note” for a detailed description of setting the number of trees in the random forest classifier). We chose the random forest [RF] classifier based on its superior performance to other models (stochastic gradient descent [SGD] classifier, soft vector machine [SVM] classifier, and logistic regression) in predicting drug approval (Fig. S8a–c; see “Supplementary Note” for a detailed description of comparison of prediction performance with those of other machine learning models). We used the discrepancy between CGE and OGE against the approval statuses of drugs in clinical trials to train the random forest classifier. To determine the discrepancy between CGE and OGE, we used both CGE and OGE as input in the classifier. The classifier also included Network, Expression, and chemical information as additional features.

We employed Monte Carlo cross-validation (1000 times) and randomly split the dataset into training (90%) and test (10%) sets to ensure reliable prediction performance. To ensure a balance of approved to unapproved drugs in each cross-validation, we randomly sampled approved drugs to match the size of unapproved drugs before splitting the dataset into training and test sets. Within each cross-validation, the balanced dataset was split so that the ratio of approved to unapproved drugs in each training and test set was maintained at 1:1. To compute the approval probability of drugs, we averaged the predicted probabilities of the random forest classifier across 1000 cross-validations. A drug was predicted as approved or unapproved based on whether its average predicted approval probability was  $\geq 0.5$  or  $< 0.5$ , respectively.

### Various metrics of prediction performance

We evaluated the prediction performance of the classifier by measuring the area under the precision–recall curve (AUPRC), area under the receiver operating characteristic curve (AUROC), accuracy, precision, recall, and sensitivity for the test set. To quantify the prediction results, we defined true positive (TP) as the number of correctly predicted unapproved drugs, true negative (TN) as the number of correctly predicted approved drugs, false positive (FP) as the number of incorrectly predicted approved drugs (approved drugs but predicted as unapproved drugs), and false negative (FN) as the number of incorrectly predicted unapproved drugs (unapproved drugs but predicted as approved drugs).

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{AUPRC} = \int_{-\infty}^{\infty} \text{Prec}(c) dP(Y \leq c)$$

$$\text{AUROC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx$$

where  $\text{Prec}(c)$  is the precision at the threshold value of  $c$  and  $\text{Recall}(c) = P(Y > c)$ . True positive rate (TPR) is by

definition equal to the recall defined above and false positive rate ( $\text{FPR}$ ) =  $\frac{FP}{FP + TN}$ .

### Sources of drugs having safety problems

To identify the association between the risk of drug targets and the cells/humans discrepancy in gene perturbation effects, we investigated target genes of failed drugs in clinical trials due to toxicity, drugs with severe side effects, and withdrawn drugs due to safety issues.

We obtained two datasets of failed drugs in clinical trials due to toxicity. The first dataset was the ClinTox dataset in MoleculeNet,<sup>24</sup> which contains 34 failed and 944 approved drugs with known drug–target interactions in the STITCH database. The second dataset was obtained from a study by Gayvert et al.,<sup>25</sup> which includes 67 failed and 697 approved drugs covered in the STITCH database.

We obtained the side effect profiles of drugs from two databases, ADReCS<sup>26</sup> (v3.1; downloaded in July 2022) and DrugCentral<sup>27</sup> (v2021; downloaded in July 2022). In the DrugCentral database, we filtered data for significant drug–side effect relationships by considering only side effects with a likelihood twice the likelihood threshold value. The sex specificity of side effects was not considered. We defined death drugs as those with the term “Death” as a side effect, based on the Medical Dictionary for Regulatory Activities (MedDRA; MedDRA preferred term: “Death”, MedDRA ID:10011906).<sup>28</sup> The ADReCS and DrugCentral databases included 341 and 533, and 321 and 669 drugs with and without the term “Death” as a side effect, respectively, all of which were covered in the STITCH database.

We obtained the 200 withdrawn drugs due to severe side effects from the ChEMBL<sup>29</sup> database and the dataset from Onakpoya et al.<sup>30</sup> covered in the STITCH database. The ChEMBL database classified the side effects of the withdrawn drugs based on the organ class, such as cardiotoxicity and neurotoxicity. The side effects of withdrawn drugs from Onakpoya et al. were manually classified according to the classifications provided by the ChEMBL database.

### Biological pathway enrichment analysis

To investigate the biological pathways associated with N2E, N2N, E2E, and E2N genes, we obtained gene sets of pathways from the BioPlanet<sup>31</sup> database. BioPlanet integrates public databases, including KEGG, Reactome, and WikiPathways, to eliminate redundancy across pathway databases. Functionally related pathways were grouped into several categories. For example, “axon guidance” and “signaling by NOTCH” were categorised under the nervous system. To identify pathway categories associated with gene sets (N2E, N2N, E2E, and E2N genes), we conducted enrichment analyses between pathways associated with each gene set and each category.

We conducted pathway enrichment analysis using g:Profiler<sup>32</sup> with a custom GMT file containing pathway

gene sets in the BioPlanet database. In g:Profiler, the statistical domain scope was set to “Custom over annotated genes”. The statistical domain scope describes the total genes used for random selection in the hypergeometric test. The total genes used in this analysis were limited to genes with information on BioPlanet pathways and gene perturbation effect (CGE and OGE) annotation. The enrichment was evaluated using a hypergeometric test. The statistical significance *P*-values from the hypergeometric tests were adjusted using the Benjamini-Hochberg false discovery rate (FDR).

### Additional chemical information

For chemical information, the components of drug-likeness rules were obtained using the simplified molecular-input line-entry system (SMILES) of drugs and codes from Gayvert et al.<sup>25</sup> (function “get-StructuralFeatures” of ProCTOR, <https://github.com/kgayvert/ProCTOR>). The SMILES strings of drugs were downloaded for all drugs from the ChEMBL (v30) database. The molecular weight, octanol-water partition coefficient log *P* (XlogP), hydrogen bond donor and acceptor count, polar surface area, formal charge, number of rings, rotatable bond counts, and refractivity data were extracted, and the rule outcomes (Rule of 5, Ghose’s rule, and Veber’s rule) were derived from these features. Weighted quantitative estimate of drug-likeness (wQED) values were computed using ProCTOR scripts.

### Principal component analysis (PCA)

For the PCA of unapproved and approved drugs, we used Scikit-learn in Python. The PCA was performed using four predicted probabilities (OGE + CGE, Network, Expression, and Chemical classifiers). Each score was scaled to show unit variance before the analysis using StandardScaler of Scikit-learn. We selected the first two principal components. The directions of each piece of information were derived from the principal axes in the feature space, representing the directions of the maximum variance.

### Statistical analysis

A *z*-score was calculated to describe the relationship with the mean of normally distributed data (random distribution), and significance was calculated by one- or two-tailed tests from the *z*-score. All data were shown as the mean ±95% confidence interval. Differences between the distributions of data were analysed by non-parametric tests that do not require a distribution to meet certain assumptions (especially if the data are normally distributed) to be analysed, such as the Mann-Whitney U, Wilcoxon signed-rank, and Kruskal-Wallis H test. The correlation was analysed by Spearman correlation analysis. All statistical analyses were performed using Scipy in Python.<sup>23</sup> A *P*-value of less than 0.05 was considered to indicate statistical significance.

### Role of the funding source

This study was funded by Korean National Research Foundation and Artificial Intelligence Graduate School Program in POSTECH. The supporting funders were not involved in the study design, data collection, data analyses, interpretation, or writing of report.

## Results

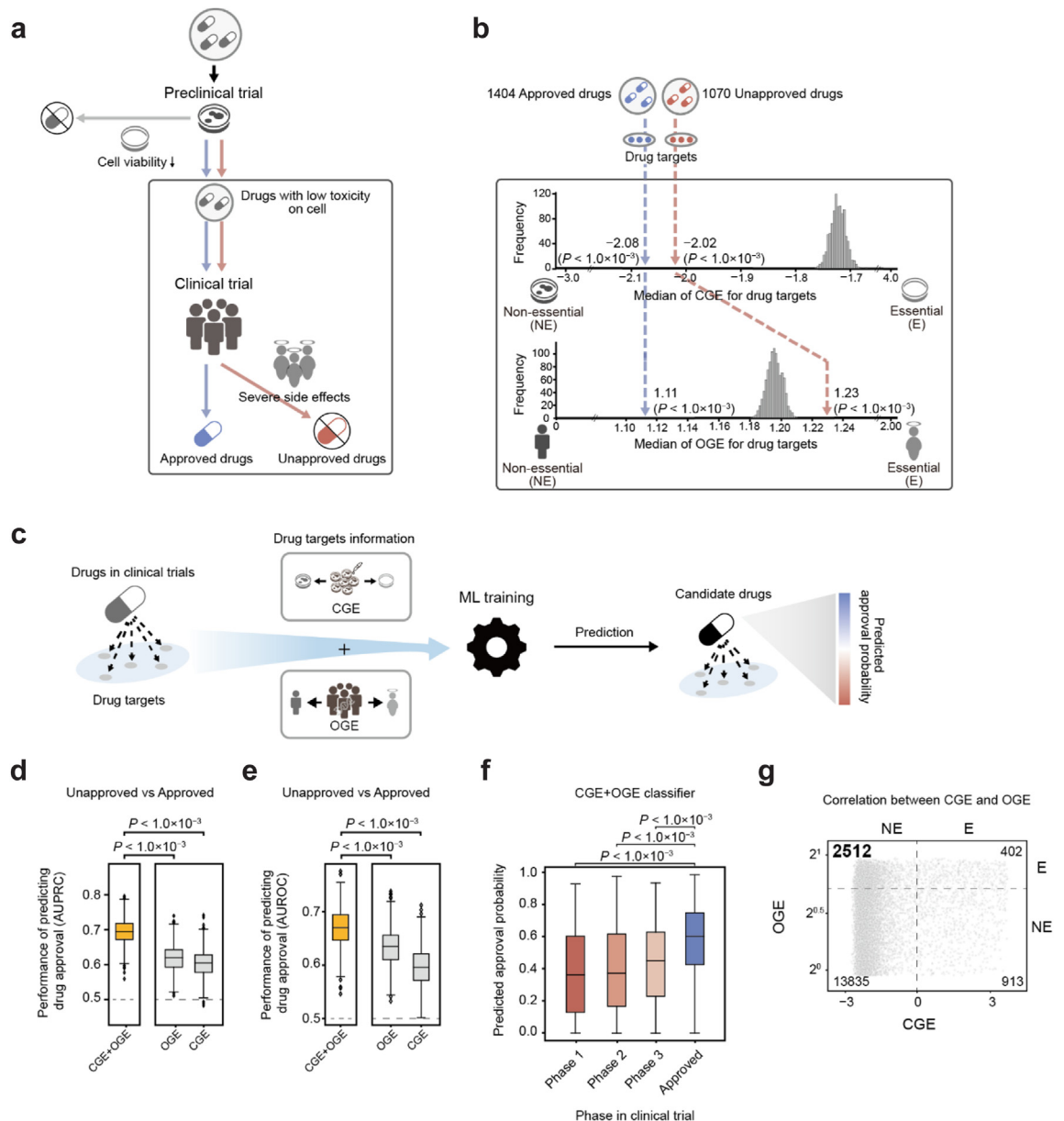
### Correlation between drug approval and the cells/humans discrepancy in gene perturbation effects

Drug approval in clinical trials can be evaluated by analysing the discrepancy in drug target perturbation effects between cells and humans. Drugs that have passed cell viability tests in preclinical trials are likely to be related to genes that are tolerant to perturbation effects on cells, regardless of their success or failure in clinical trials. Among these drugs, unapproved candidates that failed in clinical trials due to severe side effects are likely related to genes that are intolerant to perturbation effects on humans, despite their tolerance to perturbation effects on cells (Fig. 1a).

To identify the association between drug approval in clinical trials and the cells/humans discrepancy in drug target perturbation effects, we analysed the perturbation effects of genes in cells (cellular gene essentiality [CGE]), and humans (organismal gene essentiality [OGE]). The CGE and OGE were calculated for 17,662 genes (see “Methods”; Table S2). High CGE and OGE values are indicative of a strong intolerance to the perturbation effects of genes in cells and humans, respectively.

We found that drug approvals were correlated with the cells/humans discrepancy in gene perturbation effects. Specifically, unapproved drugs were associated with target genes that showed tolerance to perturbation effects on cells, but intolerance to perturbation effects on humans. However, approved drugs were associated with target genes that are tolerant to perturbation effects on both cells and humans. We compiled 7697 drug targets for 2474 drugs (1070 unapproved and 1404 approved drugs) from the STITCH 5<sup>19</sup> database, with drug approval statuses from the ChEMBL<sup>18</sup> database (see “Methods”). Drugs used in cancer therapy were excluded due to their inherent cytotoxicity and different profiles of acceptable side effects. The CGE and OGE values of drug targets were mapped onto drugs to characterise their approval statuses (see “Methods”; Table S3). As unapproved and approved drugs had passed cell viability tests in preclinical trials, the median CGEs of unapproved (*n* = 1070) and approved drugs (*n* = 1404) were significantly lower than the random distribution of the median CGEs of drugs in clinical trials (Fig. 1b; median CGE of unapproved drugs = −2.02, two-tailed test *P* < 1.0 × 10<sup>−3</sup>; median CGE of approved drugs = −2.08, *P* < 1.0 × 10<sup>−3</sup>). The random distribution refers to the median distribution of





**Fig. 1: Predicting drug approval in clinical trials with the CGE and OGE.** (a) Schematic for the drug development process. Drugs in clinical trials passed cell viability tests in preclinical trials. Drugs that failed clinical trials due to severe side effects are shown as unapproved drugs. Drugs that passed clinical trials without severe side effects are shown as approved drugs. (b) The CGE and OGE were mapped onto the unapproved and approved drugs. The numbers of unapproved and approved drugs are shown. The actual median CGE and OGE of 1070 unapproved and 1404 approved drugs are depicted with red and blue dashed lines, respectively. The histograms show the random distributions of median CGEs and OGEs for drugs in clinical trials (random shuffling of CGE and OGE 1000 times). Statistical significance ( $P$ ) was evaluated using a two-tailed test. (c) The overall framework for predicting drug approval using the CGE and OGE of drug targets with machine learning (ML). (d–e) The AUPRC and AUROC of the CGE + OGE, OGE, and CGE classifiers for predicting drug approvals in clinical trials. Unapproved drugs (drugs with max phase 1, 2, and 3) and approved drugs were predicted. Box plots represent the AUPRC and AUROC for Monte Carlo cross-validation (1000 times). Statistical significance ( $P$ ) of the differences between AUPRCs/AUROC was analysed using the Mann–Whitney U test. The grey dashed line denotes the baseline. (f) Distribution of predicted approval probabilities according to the clinical trial phase. Statistical significance ( $P$ ) of the difference in predicted approval probabilities between clinical phases was analysed using the Mann–Whitney U test. (g) The numbers of genes in each area are indicated. Dashed lines denote the thresholds of the CGE and OGE to discriminate between essential and non-essential genes. The y-axis was set to the  $\log_2$  scale. CGE, cellular gene essentiality; OGE, organismal gene essentiality.

CGEs mapped to drugs where the CGEs of drug targets are randomly shuffled 1000 times. According to the random distribution determined by us, this result indicates that the drugs in clinical trials, regardless of their approval statuses, are more related to drug targets that are tolerant to perturbation effects (“Non-essential”) on cells. However, the median OGE of unapproved drugs was significantly higher than the random distribution of median OGEs of drugs in clinical trials (median OGE of unapproved drugs = 1.23; two-tailed test  $P < 1.0 \times 10^{-3}$ ), whereas the median OGE of approved drugs was significantly lower (median OGE of approved drugs = 1.11;  $P < 1.0 \times 10^{-3}$ ). This result indicates that unapproved drugs are more associated with drug targets that are intolerant to perturbation effects (“Essential”) on humans, but approved drugs are more associated with drug targets that are tolerant to perturbation effects (“Non-essential”) on humans. In other words, among drugs that passed cell viability tests with low toxicity, unapproved drugs more frequently show intolerant effects on humans than approved drugs.

Given that drug approvals in clinical trials were correlated with the discrepancy in drug target perturbation effects between cells and humans, we hypothesised that the cells/humans discrepancy may serve as an important feature for predicting drug approval in clinical trials. To test our hypothesis, we developed a machine learning classifier to predict the probability of drug approval using the CGE and OGE of drug targets (Fig. 1c; see “Methods”). Indeed, considering the discrepancy significantly improved the prediction of drug approval in clinical trials. The AUPRC and AUROC of the classifier using the discrepancy between the CGE and OGE of drug targets (CGE + OGE) were  $0.70 \pm 0.002$  and  $0.67 \pm 0.002$  (mean value  $\pm$  95% confidence interval computed over 1000 Monte Carlo cross-validations), respectively, which were significantly higher than those of two stand-alone classifiers that relied on either OGE (AUPRC:  $0.62 \pm 0.002$ , Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ; AUROC:  $0.63 \pm 0.002$ ,  $P < 1.0 \times 10^{-3}$ ) or CGE (AUPRC:  $0.60 \pm 0.002$ ,  $P < 1.0 \times 10^{-3}$ ; AUROC:  $0.60 \pm 0.002$ ,  $P < 1.0 \times 10^{-3}$ ) alone (Fig. 1d and e; see “Methods”).

Additionally, we confirmed the reliability of our classifier using the discrepancy by assessing the correlation between the predicted probability of drug approval and the clinical phase of the drug (Table S4). In general, the probability of success in the early clinical phases is lower than that in the late phases of clinical trials.<sup>33</sup> Indeed, we observed that drugs with low predicted approval probabilities were associated with the early phases of clinical trials, whereas those with high predicted approval probabilities were associated with the late phases of clinical trials. Based on the statistically significant difference in approval probabilities between approved and unapproved drugs, our classifier could meaningfully classify drug approval. The approval

probability of the approved drugs ( $n = 1404$ ) predicted by the CGE + OGE classifier was  $0.58 \pm 0.01$  (mean value  $\pm$  95% confidence interval), which was significantly higher than those of the unapproved drugs in Phase 1 ( $n = 180$ ;  $0.38 \pm 0.04$ , Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ), Phase 2 ( $n = 493$ ;  $0.40 \pm 0.02$ ,  $P < 1.0 \times 10^{-3}$ ), and Phase 3 ( $n = 397$ ;  $0.44 \pm 0.02$ ,  $P < 1.0 \times 10^{-3}$ ) (Fig. 1f). These results indicate that the probability of drug approval predicted by our classifier can be a useful indicator of clinical approval.

We also validated the ability of the cells/humans discrepancy to predict drug approval using an independent test set. As the independent test set, we used the newly added drugs in a recently updated data version in the ChEMBL v30 database. Specifically, drugs in the previous ChEMBL v26 database (1012 unapproved and 1368 approved drugs) were used as a training set. Drugs newly added to the recent ChEMBL v30 database (58 unapproved and 36 approved drugs) did not overlap with those in the ChEMBL v26 database and were used as the test set. To ensure the representativeness of the test set for the actual population of newly developed drugs in the real-world, we refrained from re-sampling aimed at balancing unapproved and approved drugs in the test set. Again, the classifier using the discrepancy exhibited the improvement of prediction with the independent test set. The AUPRC of the CGE + OGE classifier was 0.80, which was higher than those of the two stand-alone classifiers that relied on either OGE (AUPRC = 0.66) or CGE (AUPRC = 0.67) alone (Fig. S9). Thus, we successfully predicted recently approved drugs (Table 1; seven among eight approved drugs after 2011).

Moreover, using the discrepancy in drug target perturbation effects to predict drug approval can reduce drug attrition rates in clinical trials. Drugs regarded as safe after cell viability tests frequently fail in clinical trials due to intolerance to drug-induced perturbation effects on humans. Our approach significantly reduced incorrectly predicted unapproved drugs, which were eventually unapproved but initially predicted as approved drugs based on the perturbation effects of genes on cells. Here, the recall was defined as the fraction of correctly predicted unapproved drugs among the total number of unapproved drugs. The recall of the CGE + OGE classifier was significantly improved compared to that of the CGE classifier (Fig. S10a; Wilcoxon signed-rank test  $P < 1.0 \times 10^{-3}$ ; 1000 Monte Carlo cross-validations). Moreover, approved drugs that were incorrectly predicted as unapproved drugs were also reduced by using the cells/humans discrepancy. The precision, defined as the fraction of correctly predicted unapproved drugs among the total number of drugs predicted as unapproved drugs, of the CGE + OGE classifier was improved compared to that of the CGE classifier (Fig. S10b; Wilcoxon signed-rank test  $P < 1.0 \times 10^{-3}$ ).

Predicting drug approval in clinical Phase 1 is necessary to confirm the ability of the CGE + OGE classifier to assess drug safety. Drug failure due to safety

Drug	Indication	First approval year	Predicted approval probability	CGE of drug	OGE of drug
Fosdenopterin	Metal Metabolism, Inborn Errors	2021	0.85	-1.9	<b>1.2</b>
Angiotensin	Cardiovascular diseases, septic shock	2017	0.98	-2.1	<b>1.0</b>
Vasopressin	Diabetes insipidus, Sepsis	2014	0.97	-1.9	<b>1.0</b>
Linaclotide	Irritable Bowel Syndrome, Constipation	2012	0.85	-2.0	<b>0.9</b>
Pasireotide	Acromegaly	2012	0.74	-2.3	<b>0.9</b>
Telaprevir	Virus diseases	2011	0.64	-2.0	<b>1.2</b>
Icatibant	Angioedemas, Hereditary	2011	0.62	-2.2	<b>1.1</b>
Belumosudil	Graft versus host disease	2021	0.34	-1.9	1.8

Predicted approval probabilities of recently approved drugs (after 2011) in an independent test set. Newly added drugs in the updated version of the ChEMBL database (v30) were employed as an independent test set. The successfully predicted approved drugs are presented in bold font.

**Table 1: Successfully predicted approved drugs using an independent test set.**

problems accounts for the largest proportion of failed drugs in Phase 1 trials. In addition, the first safety study on healthy volunteers is performed in Phase 1 trials. Consistent with the results of predicting unapproved drugs in Phases 1, 2, and 3, the prediction performance of the CGE + OGE classifier in predicting unapproved drugs in Phase 1 ( $n = 180$ ) was improved (Fig. S11a; OGE, Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ; CGE,  $P < 1.0 \times 10^{-3}$ ; 1000 Monte Carlo cross-validations). In addition, the misclassification rates of the classifier were significantly reduced for both unapproved and approved drugs (Fig. S11b and c; Recall, Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ; Precision,  $P < 1.0 \times 10^{-3}$ ).

The prediction performance of the CGE + OGE classifier was significantly improved by exploiting the cells/humans discrepancy. Our approach identified a considerable number of genes that were tolerant to perturbation effects on cells but intolerant to perturbation effects on humans. We identified 2512 such genes, accounting for approximately 14.2% of 17,662 genes (Fig. 1g). These genes can contribute to predicting drug approval based on the association between the cells/humans discrepancy and drug approval. Additionally, CGE and OGE were weakly correlated (Spearman's  $\rho = 0.08$  for 17,662 genes), highlighting their importance as key information for predicting drug approval in clinical trials.

### Classification of risky and safe targets

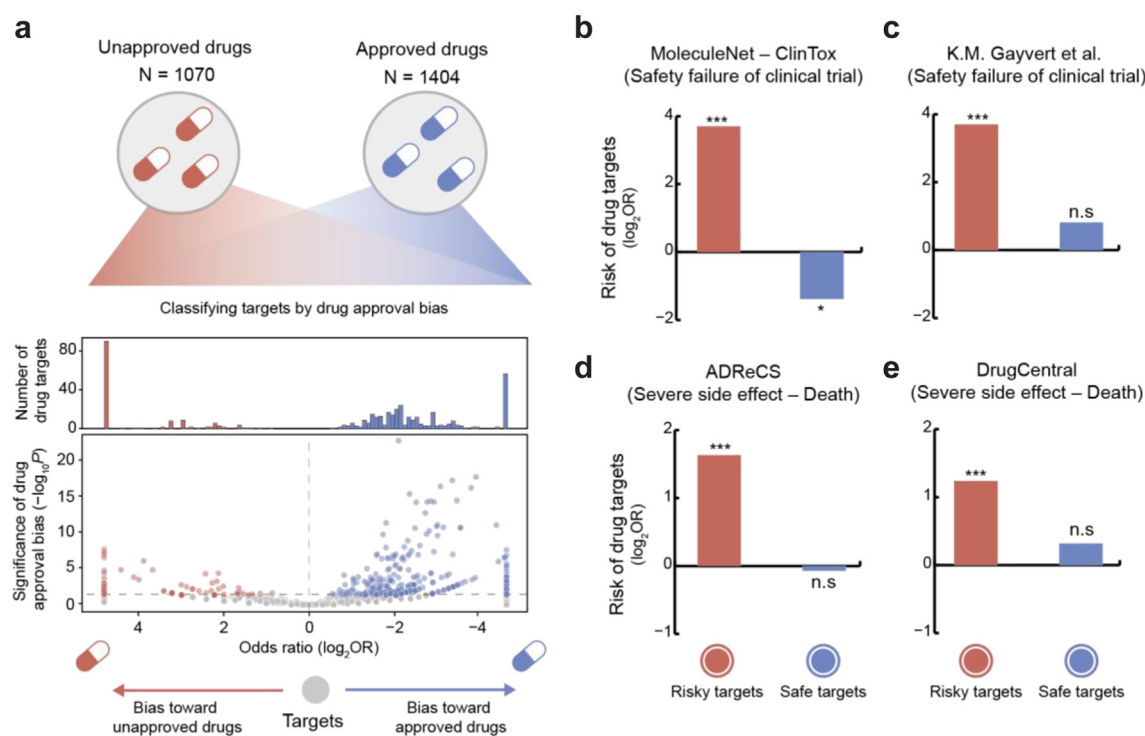
We classified risky and safe targets to investigate the relationship between the risk of drug targets and the cells/humans discrepancy. Risky and safe targets were classified based on the bias in approvals of target-associated drugs. Specifically, drug targets associated with unapproved drugs were regarded as risky targets, whereas those associated with approved drugs were regarded as safe targets. Drug approval bias was analysed using the odds ratio (OR) and  $P$ -value of Fisher's exact test for each drug target. With a  $P$ -value  $< 0.05$ , drug targets were classified as risky or safe for an  $OR > 1$ , or  $OR < 1$ , respectively. Using this approach, 141 risky and 313 safe targets were identified (Fig. 2a; Table S5).

The risky targets were identified by their association with drugs having safety problems. Drugs with risky and safe targets (risky and safe drugs, respectively) were compared to drugs that failed in clinical trials because of their toxicity (34 and 67 toxic drugs from the MoleculeNet and Gayvert et al., respectively) and drugs with the term “Death” as a severe side effect (341 and 321 death drugs from the ADRCS and Drug Central, respectively) (see “Methods”; Table S6). The risky drugs significantly overlapped with the toxic drugs (Fig. 2b and c;  $OR = 13.0$ , Fisher's exact test  $P < 1.0 \times 10^{-3}$ ;  $OR = 13.0$ ,  $P < 1.0 \times 10^{-3}$ ), whereas the safe drugs did not ( $OR = 0.4$ ,  $P = 4.4 \times 10^{-2}$ ;  $OR = 1.8$ ,  $P = 3.9 \times 10^{-1}$ ). “Death” term was the most frequently observed side effect in the boxed warning which describes severe side effects that can lead to death or serious injury.<sup>34</sup> The risky drugs also significantly overlapped with the death drugs (Fig. 2d and e;  $OR = 3.1$ , Fisher's exact test  $P < 1.0 \times 10^{-3}$ ;  $OR = 2.4$ ,  $P < 1.0 \times 10^{-3}$ ), whereas the safe drugs did not ( $OR = 1.0$ ,  $P = 0.8$ ;  $OR = 1.2$ ,  $P = 0.3$ ). These results suggest that the risk of drug target was related to the approvals of target-associated drugs. Moreover, relaxing the threshold for the statistical significance of drug approval bias showed consistent results with an increased number of risky and safe targets, emphasising the robustness of our method for classifying the risky and safe targets (Fig. S12a–e).

We confirmed that drugs in Phases 2 and 3 were also significantly related to drug safety issues. Notably, safety failures of drugs more frequently occurred in Phase 1 than in Phases 2 and 3.<sup>4</sup> To justify our classification that regarded drugs in Phases 2 and 3 as unapproved drugs with safety issues, we analysed the association between the risky targets from those drugs and drugs having safety problems. Applying the same methodology to classify the risky and safe targets, 123 risky and 225 safe targets were identified. Risky drugs associated with the 123 risky targets significantly overlapped with toxic and death drugs, whereas safe drugs associated with the 225 safe targets did not (Fig. S13a–d).

To verify the validity of our classification of the risky and safe targets based on drug approval bias, we





**Fig. 2: Classification of risky and safe targets.** (a) Classification of risky and safe targets based on bias in drug approvals. Red and blue dots indicate the risky and safe targets biased significantly toward unapproved and approved drugs, respectively. Drug approval bias was analysed using the odds ratio (OR) and P-value (P) of Fisher's exact test. Grey dots indicate drug targets without significant approval bias. The histogram shows the distribution of the risky and safe targets. Dashed lines denote the threshold for the classification of risky and safe targets. (b–e) The red and blue bars represent the enrichment (odds ratio [OR]) of drugs with safety problems among drugs with risky and safe targets, respectively. The statistical significance of enrichment was measured using Fisher's exact test. Asterisks denote statistical significance ( $***P < 1.0 \times 10^{-3}$ ;  $**P < 1.0 \times 10^{-2}$ ;  $*P < 5.0 \times 10^{-2}$ ; n.s.: not significant).

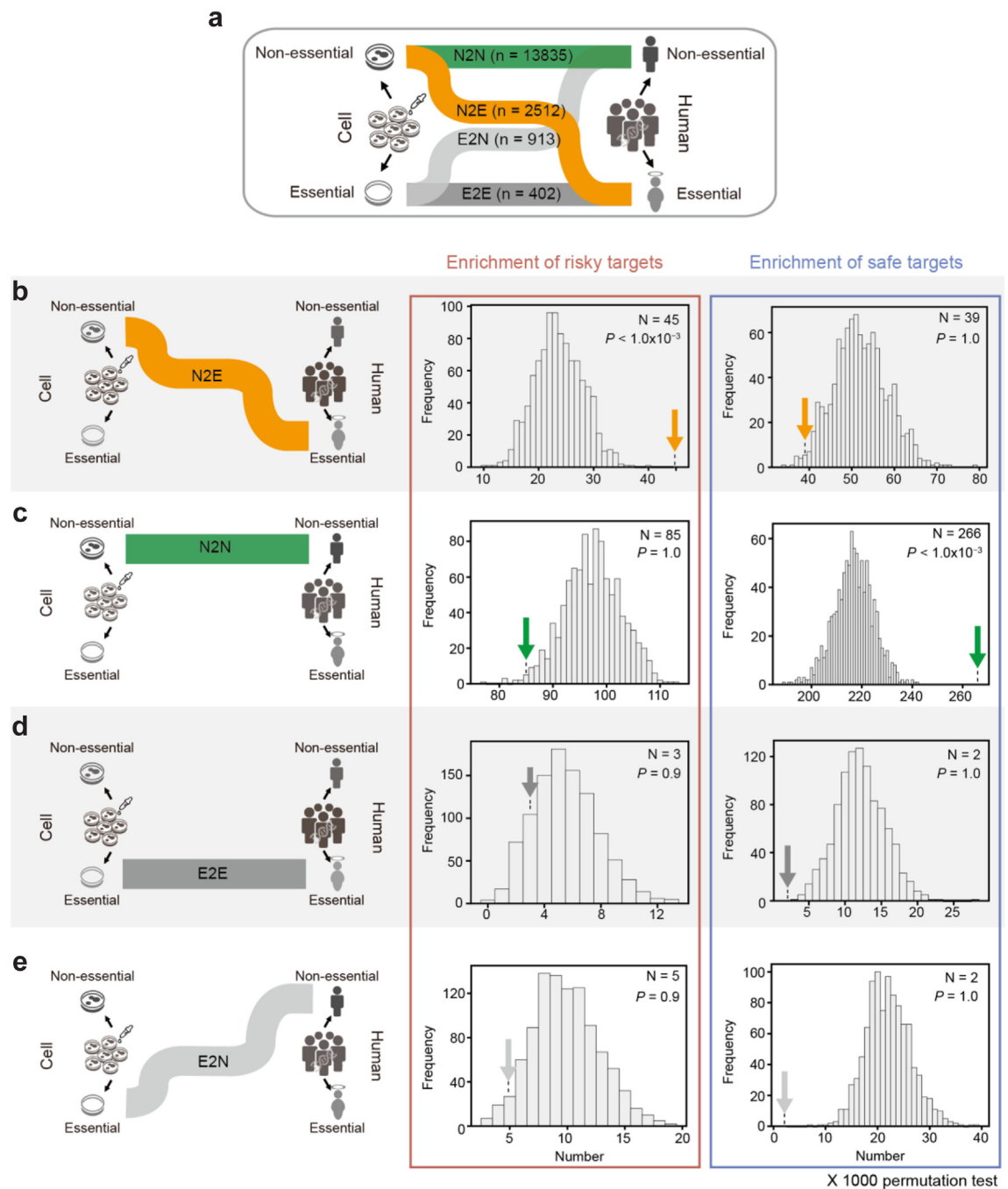
examined the number of associated drugs as well as the degree of drug approval bias per drug target. The number and degree per drug target should be large enough to establish a meaningful relationship between the risk of drug targets and drug approval bias. We found that the risky and safe targets classified using our method were statistically sound. Those drug targets with significant approval bias ( $n = 454$ ), compared with those without significant bias ( $n = 7243$ ), were associated with a greater number of associated drugs (Fig. S14a; Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ; Spearman's  $\rho = -0.38$  for 7697 drug targets) and a higher degree of drug approval bias (Fig. S14b;  $P < 1.0 \times 10^{-3}$ ; Spearman's  $\rho = -0.44$ ). These results indicate that the risky and safe targets classified using our method are explicitly shared by unapproved and approved drugs, respectively. Thus, the risk of drug targets can be assessed by the approval bias of the drug targets.

#### Enrichment of risky targets in non-essential genes in cells but essential genes in humans

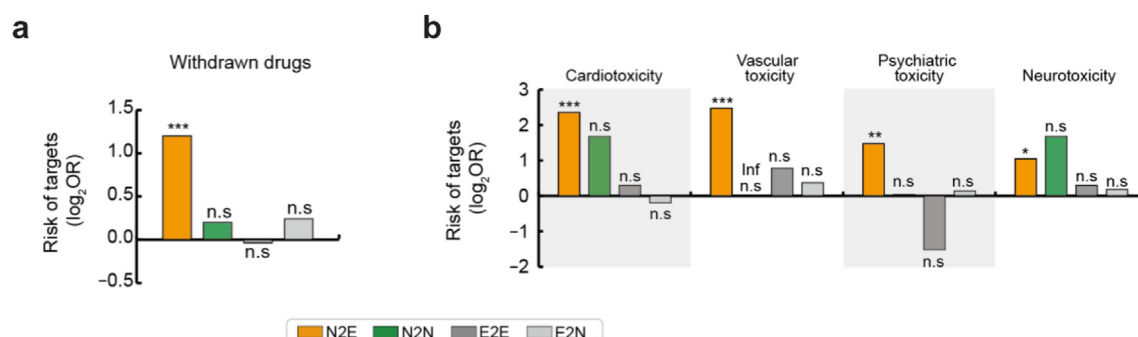
To characterise risky and safe targets with the discrepancy in gene perturbation effects between cells and

humans, we categorised 17,662 genes into four groups by comparing gene perturbation effects between cells and humans (Fig. 3a). We found that 2512 genes were non-essential in cells but essential in humans (N2E), 13,835 genes were non-essential in both cells and humans (N2N), 402 genes were essential in both cells and humans (E2E) and 913 genes were essential in cells but non-essential in humans (E2N).

We found that the risk of drug targets in clinical trials can be characterised by the cells/humans discrepancy. Specifically, the risky targets were associated with N2E genes, which were tolerant to perturbation effects on cells but intolerant to perturbation effects on humans (Fig. 3b; one-tailed test  $P < 1.0 \times 10^{-3}$  for 1000 permutation tests). In contrast, the safe targets were associated with N2N genes, which were tolerant to perturbation effects on both cells and humans (Fig. 3c; one-tailed test  $P < 1.0 \times 10^{-3}$  for 1000 permutation tests). These results indicate that drug targets that severely affect cell viability were significantly filtered out from preclinical trials. Accordingly, both risky and safe targets were associated with genes that were tolerant to perturbation effects on cells. However, the risky targets



**Fig. 3: Enrichment of risky/safe targets in gene sets based on a comparison of CGE and OGE.** (a) Sankey diagram showing the discrepancy of gene perturbation effects between cells and humans. The numbers of genes based on a comparison of CGE and OGE are shown. (b-e) Enrichment of risky/safe targets in each gene set (N2E, N2N, E2E, and E2N). Plots in red and blue boxes represent the enrichment of risky and safe targets, respectively. The histogram shows the expected numbers of risky/safe targets in each gene set for 1000 permutation tests. Arrows denote the observed number of risky and safe targets. The observed numbers of risky/safe targets (N) and statistical significance (P) are shown inside the parenthesis. Statistical significance was measured using a one-tailed test. The one-tailed test was performed with the observed and the expected numbers of the risky/safe targets for 1000 permutation tests in each gene set. CGE, cellular gene essentiality; OGE, organismal gene essentiality.



**Fig. 4: Association of the cells/humans discrepancy with drug withdrawal due to severe side effects.** (a) Enrichment of withdrawn drugs with severe side effects among drugs associated with each gene set (N2E, N2N, E2E, and E2N). (b) Enrichment analyses were performed with withdrawn drugs classified by side effect class. (a–b) Each bar represents the enrichment (odds ratio [OR]) of withdrawn drugs among drugs associated with each gene set (Inf, infinite). The statistical significance of enrichment was analysed using Fisher's exact test. Asterisks denote statistical significance ( $***P < 1.0 \times 10^{-3}$ ;  $**P < 1.0 \times 10^{-2}$ ;  $*P < 5.0 \times 10^{-2}$ ; n.s: not significant).

can be identified by the perturbation effects of genes on humans. In addition, the OGE of the risky targets ( $n = 141$ ) was significantly higher than that of the safe targets ( $n = 313$ ), and this difference was significantly larger than that of the CGE (Fig. S15a and b; Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ;  $P = 1.4 \times 10^{-2}$ ). Meanwhile, E2E and E2N genes, which were intolerant to perturbation effects on cells, were not associated with the risky or safe targets in clinical trials because they were likely to be filtered out by cell viability tests in preclinical trials (Fig. 3d and e).

Using various metrics of gene perturbation effects on cells and humans (Tables S7 and S8; see “Supplementary Note” for a detailed description of various metrics of gene perturbation effects), we confirmed the robustness of our results: the risky targets were associated with N2E genes, whereas the safe targets were associated with N2N genes. For various metrics of gene perturbation effects on cells, the class of fitness genes from Behan et al.,<sup>9</sup> probability of dependency from the DepMap database,<sup>14</sup> and fitness score of human pluripotent stem cells (hPSCs)<sup>35</sup> were investigated. Analysis for the other metric of gene perturbation effects on humans was expanded with the probability of being loss-of-function intolerant (pLI) from the ExAC database.<sup>15</sup> Essential genes in cells defined by CGE significantly overlapped with those defined by different metrics and two benchmark sets (Fig. S16a–c). Also, essential genes in humans defined by OGE significantly overlapped with those defined by pLI (Fig. S17). Through the same analysis with other metrics of gene perturbation effects, we found that the risky and safe targets were significantly enriched in N2E and N2N genes, respectively (Fig. S18a–g; Table S9). These results from unbiased gene perturbation effects on cells and humans emphasise our finding that the risk of drug targets in clinical trials can be assessed by the cells/humans discrepancy in gene perturbation effects.

Additionally, relaxing the threshold for the statistical significance of drug approval bias showed consistent results with an increased number of risky and safe targets (Fig. S19a–d), indicating the robustness for characterising the risk of drug targets using the cells/humans discrepancy in gene perturbation effects.

#### Association between N2E targets and withdrawn drugs with severe side effects

We further validated that N2E genes were likely to be the risky drug targets using withdrawn drugs having severe side effects (Table S10). Among 200 withdrawn drugs, 139 drugs targeted N2E genes (Fig. 4a; Odds ratio (OR) = 2.3, Fisher's exact test  $P < 1.0 \times 10^{-3}$ ). For example, thalidomide, used in hyperemesis, was withdrawn because of teratogenicity. The transcription factor *SP1*, an N2E gene, was inhibited by thalidomide, disrupting angiogenesis in the embryo.<sup>36</sup> Another example is sibutramine used to treat obesity, which was withdrawn because of side effects that included heart attack and stroke. Sibutramine inhibits the dopamine transporter *SLC6A3*, an N2E gene that has been implicated in several diseases, including neuropsychiatric and neurological disorders.<sup>13</sup> Withdrawn drugs with N2E targets associated with severe side effects, including the drugs mentioned above, are summarised in Table 2.

We found that N2E genes were likely to induce side effects associated with the central nervous system (CNS) and cardiovascular system, leading to drug withdrawal from markets. Side effects in the CNS and cardiovascular system are the most common reasons for safety failures in clinical trials.<sup>4</sup> Indeed, these side effects were frequently observed in withdrawn drugs with N2E targets. Withdrawn drugs that induced cardiovascular (38 drugs with cardiotoxicity and 34 drugs with vascular toxicity), psychiatric (39 drugs with psychiatric toxicity), and neurotoxic effects (38 drugs with neurotoxicity) were significantly associated with N2E targets (Fig. 4b;

Withdrawn drug	Indication	N2E targets	Reasons for withdrawal	References (PMID)
Novobiocin	Staphylococcal infections	AKT1	Jaundice, Hepatic failure, Blood dyscrasias	31445927
Isoxicam	Analgesia	PTGS2	Stevens-Johnson Syndrome, Toxic Epidermal Necrolysis	14528522
Pirprofen	Analgesia	PTGS2	Gastrointestinal toxicity, Hepatotoxicity	34234675
Miglustat	Gaucher's disease type 1	UGCG	Unexplained cognitive dysfunction	16109770
Mazindol	Obesity	SLC6A3	Drug abuse, Psychiatric toxicity	34650206
Phenmetrazine	Obesity	SLC6A3	Drug abuse, psychiatric	34650206
Pyrovalerone	Chronic fatigue syndrome	SLC6A3	Drug abuse, Hematological toxicity	34650206
Zopiclone	Hypnosedation	GABRA2	Carcinogenicity	31871774
Danthron	Constipation	RXRA	Carcinogenicity	12183441
Acitretin	Psoriasis	STAT3	Teratogenicity	17823373, 23788624
Dicyclomine	Allergy, hyperemesis	CHRM1,3,4	Psychiatric toxicity	34521861, 11392633
Thalidomide	Hyperemesis	SP1	Teratogenicity	10799645
Gatifloxacin	Bacterial infection	AKT1	Dysglycemia	26554652, 32717270
Sibutramine	Obesity	SLC6A3	Stroke and Heart attack, Cardiovascular toxicity,	34650206
Ketorolac	Analgesia	PTGS2	Renal toxicity, Hematological toxicity, Gastrointestinal toxicity, Dermatological toxicity	19832117, 34234675, 12518182, 27713354

Withdrawn drugs with reasons for withdrawal (side effects) and their N2E targets. References column contains literatures that support the relationship between phenotypic outcomes and perturbation of the N2E target, which relates the side effects to drug withdrawal.

**Table 2: Withdrawn drugs with their N2E targets and severe side effects.**

OR = 5.1, Fisher's exact test  $P < 1.0 \times 10^{-3}$ ; OR = 5.6,  $P < 1.0 \times 10^{-3}$ ; OR = 2.8,  $P = 5.4 \times 10^{-3}$ ; OR = 2.1,  $P = 4.8 \times 10^{-2}$ ). However, drugs with N2N, E2E, and E2N targets were not significantly associated with any withdrawn drugs having these side effects.

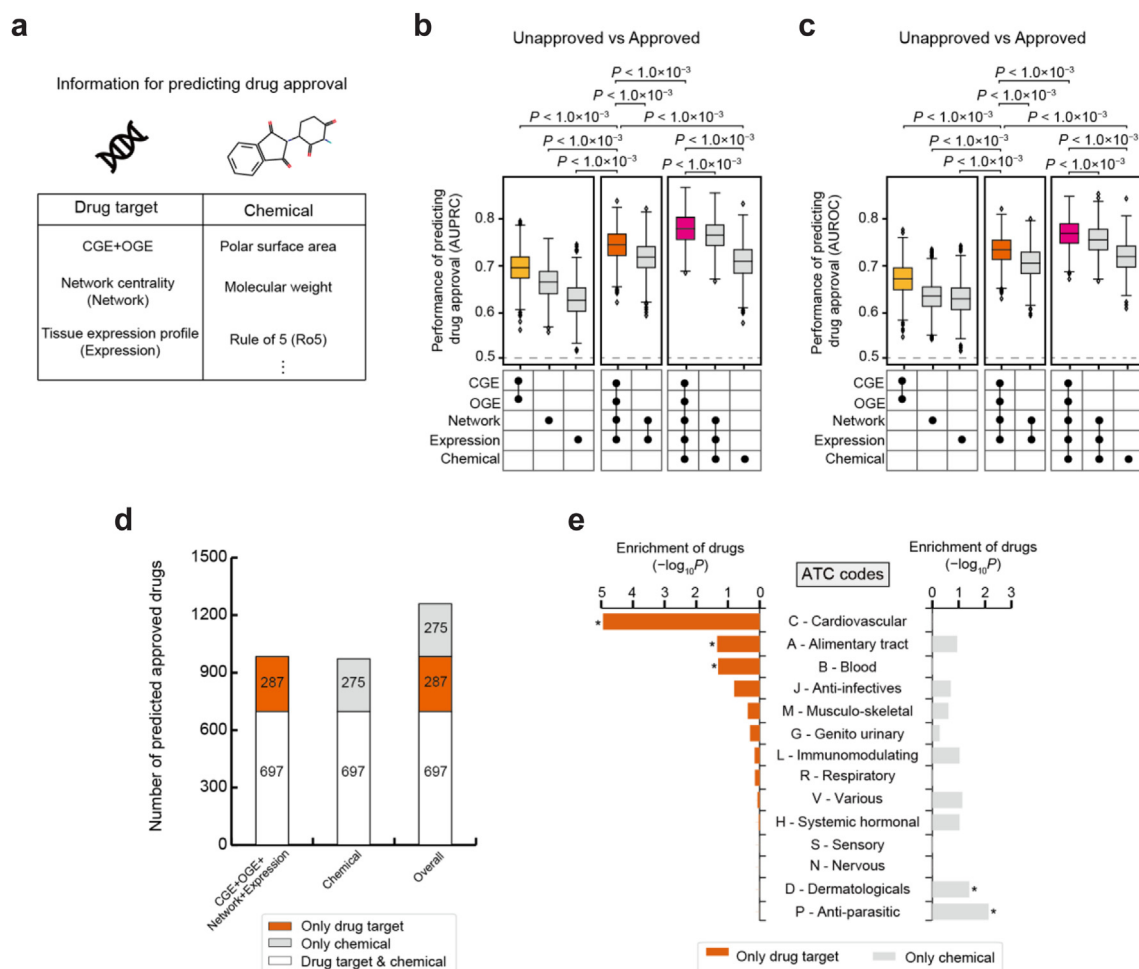
We found that the CNS and cardiovascular toxicity of withdrawn drugs with N2E targets were explained by the important roles of N2E genes in the nervous and circulatory system pathways (Fig. S20a; Tables S11 and S12; see "Methods"). For example, N2E genes play an important role in axon guidance. The perturbation of axon guidance induces neurological disorders.<sup>37</sup> Angiogenesis is another pathway related to N2E genes. Cardiotoxicity is induced by the perturbation of angiogenesis.<sup>38</sup> Meanwhile, G protein-coupled receptor (GPCR) signalling and metabolism-related pathways were associated with N2N genes (Fig. S20b). Importantly, GPCR signalling pathways are major pathways responsible for the therapeutic effect of 70% of approved small-molecule drugs.<sup>39</sup> Thus, N2N genes may be more desirable drug targets than N2E genes. E2E and E2N genes were associated with pathways for cellular processes essential for cell viability, such as RNA processing, DNA replication, and transcription (Fig. S20c and d).

### Improving drug approval prediction with additional information

Motivated by previous reports of assessing drug safety with diverse information, we examined whether the

performance of drug approval prediction was improved by integrating the cells/humans discrepancy with additional drug target and chemical information.

We found that additional information on drug targets and chemicals improved the prediction of drug approval. The classifier to predict drug approval was redesigned with additional information (Fig. 5a; Tables S13–S15). For drug target information, the centrality of the protein interaction network (Network) and tissue expression profiles (Expression) were used (see "Methods"). Reportedly, risky drug targets tend to have a high degree or betweenness centrality in the protein interaction network<sup>5,6,40</sup> and high or broad expression across tissues, such as the expression of housekeeping genes.<sup>7,8</sup> For chemical information, the components of the drug-likeness rules, such as molecular weight, number of hydrogen bond donors or acceptors, and polar surface area, were used (see "Methods"). The drug-likeness rules proposed by Lipinski (Rule of 5; Ro5), Veber, and Ghose are considered useful guidelines for the structural properties of desirable drugs with low toxicity.<sup>41–43</sup> The AUPRC and AUROC of the integrated classifier with additional drug target information (CGE + OGE + Network + Expression classifier) for predicting drug approval were significantly improved ( $0.74 \pm 0.002$  and  $0.73 \pm 0.002$ , respectively; mean value  $\pm$  95% confidence interval computed over 1000 Monte Carlo cross-validations) compared with those of the CGE + OGE classifier (Mann–Whitney U test  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ) and two stand-alone



**Fig. 5: Integration of drug target and chemical information to predict drug approval.** (a) Components of drug target and chemical information for predicting drug approval. (b–c) Prediction performance of drug approvals using a combination of drug target and chemical information. The components of each information combination are represented with filled dots in the matrix. Statistical significance ( $P$ ) associated with the difference in the AUPRCs and AUROCs was analysed using the Mann–Whitney U test. (d) Numbers of correctly predicted approved drugs. (e) Enrichment of correctly predicted approved drugs with various drug indications. Each drug was classified by the level 1 Anatomical Therapeutic Chemical (ATC) code. Statistical significance ( $P$ ) associated with enrichment was analysed using a hypergeometric test. Asterisks denote the statistical significance of enrichment ( $*P < 5.0 \times 10^{-2}$ ). CGE, cellular gene essentiality; OGE, organismal gene essentiality.

classifiers that relied on either network (Network classifier;  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ) or expression information (Expression classifier;  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ) alone (Fig. 5b and c). Furthermore, the AUPRC and AUROC of the integrated classifier with chemical information (CGE + OGE + Network + Expression + Chemical classifier) were significantly improved ( $0.78 \pm 0.002$  and  $0.77 \pm 0.002$ , respectively) compared with those of other classifiers using either drug target (CGE + OGE + Network + Expression classifier; Mann–Whitney U test  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ) or chemical information (Chemical classifier;  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ). Particularly, the AUPRC and AUROC of the integrated classifier were significantly improved compared with those of the

Network + Expression + Chemical classifier (Mann–Whitney U test  $P < 1.0 \times 10^{-3}$  and  $P < 1.0 \times 10^{-3}$ ). The improvement of drug approval prediction was also exhibited in various prediction performance metrics (Accuracy, Precision, and Specificity) (Fig. S21a–d; see “Methods”). These results suggest that the cells/humans discrepancy can complement Network, Expression, and Chemical information. The discrepancy needs to work together with those features for drug approval prediction. This is supported by a feature importance analysis of the classifier. The analysis indicated that both the drug target and chemical information contributed significantly to predicting drug approval (Fig. S22). Notably, the cells/humans discrepancy emerged as the most crucial feature.



The considerable difference in approval probabilities between approved and unapproved drugs resulted in a significant classification of drug approval (Fig. S23a and b). The approval probability of the approved drugs ( $n = 1404$ ) predicted by the integrated classifier was  $0.60 \pm 0.01$  (mean value  $\pm$  95% confidence interval), which was significantly higher than those of the unapproved drugs in Phase 1 ( $n = 180$ ;  $0.36 \pm 0.01$ , Mann–Whitney U test  $P < 1.0 \times 10^{-3}$ ), Phase 2 ( $n = 493$ ;  $0.38 \pm 0.01$ ,  $P < 1.0 \times 10^{-3}$ ), and Phase 3 ( $n = 397$ ;  $0.43 \pm 0.01$ ,  $P < 1.0 \times 10^{-3}$ ).

We confirmed that our classifier exhibited robust performance for predicting drug approval when the size of data used for the test increased (the test data sizes ranged from 0.1 to 0.9). When 30% of the data was used for the test, the AUPRC and AUROC of the classifier were  $0.77 \pm 0.001$  and  $0.76 \pm 0.001$  (purple random forest classifier consisted of 1000 trees; mean value  $\pm$  95% confidence interval over 1000 Monte Carlo cross-validation), respectively, which were comparable with those of the classifier ( $0.78 \pm 0.002$  and  $0.77 \pm 0.001$ ) when 10% of the data was used for the test (Fig. S7a and b). Although the performance of the classifier for predicting drug approval decreased as the size of the data used for the test increased (the size of the data used for the training decreased), the classifier exhibited that the AUPRC ( $0.71 \pm 0.001$ ) and AUROC ( $0.70 \pm 0.001$ ) were higher than 0.70 even in the most challenging test (10% and 90% of data used for training and test, respectively).

The improvement in drug approval prediction using the additional information is due to the orthogonality between the drug target and chemical information. We found that the chemical information had a different explanatory power than the drug target information. The orthogonality between the drug target and chemical information was analysed using principal component analysis (PCA; see “Methods”). Using PCA, we reduced the dimensionality of approval probabilities for four classifiers (CGE + OGE, Network, Expression, and Chemical classifiers) into two principal components. With 47.1% and 21.9% of variance attributed to principal components 1 and 2, respectively, the vector of the chemical information in PCA had a different direction relative to those of the drug target information (Fig. S24). Principal axes (eigenvector) of CGE + OGE, Network, and Expression information in feature space were (0.59, 0.09), (0.52, 0.26), and (0.50, 0.30), respectively, and those of direction angles were in Quadrant I. In contrast, that of Chemical information was (0.36, -0.92), and that of direction angle was in Quadrant IV.

Moreover, we found that the number of correctly predicted approved drugs was increased by integrating the drug targets and chemical information. Classifiers using the drug targets (CGE + OGE + Network + Expression) or chemical information alone (Chemical) correctly predicted 984 and 972 approved drugs,

respectively. However, the total number of correctly predicted approved drugs increased to 1259 when both drug target and chemical information were used (Fig. 5d), suggesting that the correct prediction of drug approval can be expanded by integrating drug target and chemical information in the classifier. Furthermore, the ability of drug target and chemical information to predict drug approval depends on the organ system in which the drug works. Drugs that were correctly predicted using only drug target or chemical information were investigated using the Anatomical Therapeutic Chemical (ATC) code. The cardiovascular system (ATC code C) and blood and blood-forming organs (ATC code B) associated with 287 drugs were correctly predicted by the CGE + OGE + Network + Expression classifier alone, whereas antiparasitic products (ATC code P) and dermatologicals (ATC code D) associated with 275 drugs were correctly predicted by the Chemical classifier alone (Fig. 5e). Thus, drug target information can be complemented with chemical information to predict drug approval for various indications in clinical trials.

Compared with the state-of-the-art model, ProCTOR,<sup>25</sup> which predicts the successes and failures of drugs in clinical trials, we confirmed that our classifier based on the cells/humans discrepancy was more robust in predicting drug approval using the small-sized training data. ProCTOR used drug target information, such as the centrality of the protein-interaction network, gene expression profiles across tissues, and chemical information. We identified that our classifier (CGE + OGE + Network + Expression + Chemical classifier) outperformed ProCTOR in the large and challenging sizes of test data (Fig. S25a; the test data sizes ranged from 0.6 to 0.9). For the small-sized test data, our classifier exhibited comparable prediction performance to ProCTOR. Moreover, the prediction performance of drug approval was significantly improved by adding the cells/humans discrepancy (CGE + OGE) to ProCTOR. Through feature importance analysis, we confirmed that not only features of ProCTOR but also the cells/humans discrepancy contributed significantly to predicting drug approvals. In the improved prediction achieved by integrating ProCTOR with the cells/humans discrepancy, the cells/humans discrepancy (CGE + OGE) complemented the features of ProCTOR by exhibiting a noteworthy feature importance score (Fig. S25b;  $0.15 \pm 0.0003$ ; mean value  $\pm$  95% confidence interval across the test data sizes ranged from 0.1 to 0.9). Meanwhile, more than half of the features used by ProCTOR play a substantial role in predicting drug approval. Importance scores of 11 out of 19 ProCTOR features were greater than 0.05 (under the assumption of equal importance among all 20 features, including those of ProCTOR and the cells/humans discrepancy, 0.05 of feature importance would be assigned to each feature). This result suggests that the features of ProCTOR are sufficiently valuable for drug approval

prediction, emphasizing the necessity of integrating these features with the cells/humans discrepancy.

The performance of a machine learning model depends on the selection of data representation, including feature engineering and dataset labelling, where they are applied.<sup>44</sup> We demonstrated the validity of our selection of data representation applied to the classifier for predicting drug approval. For example, CGE calculated by feature engineering, which averaged the fitness cores of 324 cell lines, was more appropriate than using all fitness scores (ACL). The CGE classifier (CGE + OGE + Network + Expression + Chemical classifier) exhibited a more robust predictive ability than the ACL classifier (ACL + OGE + Network + Expression + Chemical classifier) in predicting drug approval. When we assessed prediction performance by progressively expanding the size of the data used for the test, the CGE classifier outperformed the ACL classifier in the large and challenging size of test data (Fig. S26a; the test data sizes ranged from 0.6 to 0.9). While the ACL classifier exhibited higher prediction performance than the CGE classifier for the small-sized test data, the approval probabilities of drugs predicted by both classifiers exhibited a strong correlation (Fig. S26b; Spearman's  $\rho = 0.71$  for the approval probabilities of 2474 drugs; 30% of data used for the test). This indicates that the predictive ability of the two classifiers is comparable in the small-sized test data. Additionally, we confirmed that our dataset labelling method, which labels drugs withdrawn from certain countries as approved drugs due to their usage in other countries, did not considerably affect the training and performance of the classifier. The prediction performance remained comparable whether those drugs were included or excluded in the training set (Fig. S27). When the test data size was set at 0.1, the performance of the classifier exhibited no significant difference according to the inclusion or exclusion of the withdrawn drug in the training set (1000 Monte Carlo cross-validations; Mann–Whitney U test  $P = 1.2 \times 10^{-1}$ ). When the data size used for the test increased from 0.3 to 0.9, the performance improvement with the training set, including withdrawn drugs, was significant but very slight.

## Discussion

We found that the discrepancy between gene perturbation effects on cells (CGE) and humans (OGE) is a good indicator for predicting drug approval and assessing the risk of drug targets in clinical trials. Traditional 2D cell culture model-based screening has been used to evaluate drug safety and is a standard platform used in drug development because of its low cost and efficiency.<sup>45,46</sup> However, drugs that pass the cell viability test often fail in clinical trials because of their severe side effects in humans. In our study, the performance of drug approval prediction was improved using the discrepancy

between the CGE and OGE (Fig. 1d and e). In particular, the number of incorrectly predicted unapproved drugs, which were eventually unapproved but initially predicted as approved drugs by the CGE, was significantly reduced (Fig. S10a). This suggests that the CGE can be complemented with the OGE to reduce drug attrition rates and assess the risk of drug targets in clinical trials. Specifically, N2E genes, which were non-essential in cells but essential in humans, improved drug approval prediction, as these genes were associated with risky drug targets (Fig. 3b). Furthermore, N2E genes tend to be targeted by withdrawn drugs having severe side effects (Fig. 4a and b). Indeed, Caldu-Primo et al.<sup>47</sup> reported that genes that were tolerant to perturbation effects in *in vitro*, but intolerant to perturbation effects *in vivo* were associated with human diseases, especially brain/psychiatric disorders. Therefore, drug-induced perturbations of N2E genes have the potential to induce many abnormal phenotypes, including severe side effects.

Drug development can be more efficient by making decisions with information that is accessible before rather than during, or after clinical trials. We predicted drug approval with probability, a critical indicator of drug development, using the cells/humans discrepancy that was accessible before clinical trials. We found that drugs with higher approval probabilities were more likely to proceed to higher clinical trial phases (Fig. 1f). Lo et al.<sup>48</sup> and Siah et al.<sup>49</sup> previously predicted the approval probability of drugs using machine learning with clinical trial data such as the prior approval of drugs for another indication and the number of clinical trials with positive or negative results. They showed that the most important features for predicting drug approval included previous trial outcomes and trial accrual rates. However, these features are not available before clinical trials for new drugs. For instance, access to the accrual rate of planned participants associated with the statistical power of drug effects on sufficiently sizable human participants is limited for new drugs before clinical trials. The analysis of the discrepancy between the CGE and OGE of drug targets in this study, which was accessible information before clinical trials, allows the assessment of the safety and success potential of novel candidate drugs before clinical trials.

The OGE can represent the perturbation effects of drug targets on clinical trial participants with diverse genetic backgrounds. The genetic diversity of the participants is a critical factor in assessing drug safety in clinical trials<sup>50</sup> because of different outcomes for drug safety across ethnic groups.<sup>51,52</sup> As the OGE was estimated by modelling the mutation rates from a population of approximately ~140,000 including various races, the OGE can be generalisable perturbation effects of genes on humans.

We found that the CGE and OGE derived from genetic perturbation effects were good representatives of

drug-induced perturbation effects. Genetic perturbation effects, such as CRISPR-Cas9 knockout and genetic variant, are maintained throughout life, whereas the effect of drug-induced gene perturbation is maintained for the duration of use. However, this difference may not be a drawback when assessing the safety of the drug targets in our study. Drug approval in clinical trials was correlated with the discrepancy between the CGE and OGE (Fig. 1b). Unapproved drugs that passed the cell viability tests were related to genes that are tolerant to perturbation effects on cells. However, these drugs causing severe side effects on humans were related to genes that are intolerant to perturbation effects on humans. In recent reports, the genetic perturbation effect has been used to assess the safety of drug targets.<sup>5,8,53,54</sup> For example, Behan et al.<sup>9</sup> prioritised therapeutic targets with low toxicity using CRISPR-Cas9 knockout-based gene perturbation effects on cells. Whiffin et al.<sup>17</sup> considered the gene perturbation effect on humans estimated by modelling loss-of-function mutation rates to assess the safety of *LRRK2* as a candidate target for Parkinson's disease. Minikel et al.<sup>55</sup> reported that target genes of approved drugs, except cytotoxic chemotherapy agents used in cancer, tend to be tolerant to perturbation effects on humans. Together with these reports, our results demonstrated that analysis of genetic perturbation effects is a valuable approach for mimicking the drug perturbation effects.

For efficient drug development with minimal toxicity, the discrepancy in gene perturbation effects between pre-clinical animal models and humans also needs to be considered. Drugs evaluated in animal models, especially in mice which are the most widely employed and cost-effective model for drug development,<sup>56,57</sup> often fail in clinical trials due to safety problems.<sup>58–60</sup> These safety failures could be due to phenotypic differences in gene perturbation effects between mice and humans discussed by Han et al.<sup>61</sup> and Ha et al.<sup>62</sup> With these reports, we identified the discrepancy in gene perturbation effects between mice (mouse gene essentiality [MGE]) and humans (OGE). We found that a significant number of genes have a discrepancy between MGE and OGE (Table S16). Genes that were non-essential in mice but essential in humans ( $N_m2E$ ) were also enriched in the risky targets, whereas genes that were non-essential in both mice and humans ( $N_m2N$ ) were enriched in the safe targets (Fig. S28a and b, Table S17; Fig. S29a–d). Based on our results, considering the animal models/humans discrepancy will increase the confidence in evaluating drug safety for humans and finally reduce drug attrition rates in clinical trials.

Integrating drug target and chemical information improved the ability to predict drug approval in clinical trials. Chemical information has been widely used to filter out undesirable and toxic drugs during drug development.<sup>63</sup> Cai et al.<sup>64</sup> developed a machine-learning model based on chemical information to predict the

ability of the drug to be approved. However, chemical information may have limitations in drug approval prediction because of the deviation of drug-likeness rules for newly developed drugs.<sup>65</sup> Using drug target information can help overcome the limitations of using chemical information in drug assessment.<sup>64,66–73</sup> In this study, we found that the performance of the drug approval prediction model was improved by using both types of information (Fig. 5b and c). Furthermore, the indications of approved drugs that were correctly predicted using drug target information were different from those using chemical information (Fig. 5d and e). Consistent with our results, Gayvert et al.<sup>25</sup> reported an improved drug safety assessment by integrating chemical and drug target information. Notably, adding the cells/humans discrepancy for drug targets to the ProCTOR model developed by Gayvert et al. contributed to a significant improvement in the performance of drug approval prediction (Fig. S25). These results suggest that integrating drug target information, especially the cells/humans discrepancy, and chemical information is necessary to develop drugs for broad indications.

The process of drug approval is intricate and involves the comprehensive assessment of multiple factors, including safety, efficacy, pharmacokinetics, pharmacodynamics, and benefit-risk balance. Consequently, solely relying on safety assessment cannot reliably predict approval for some drugs (Fig. S23b). In addition to information for drug safety assessment, such as the cells/humans discrepancy, considering other important factors, such as efficacy, can make room for improvement of the accuracy and reliability of predicting drug approval.

#### Contributors

M.P., D.K., I.K. and S.K. conceived and designed the experiments. M.P. and S.K. devised the methodologies. M.P. curated the data and performed the experiments. M.P. and S.K. analysed the data. M.P., D.K. and S.K. wrote the paper. S.I. and S.K. supervised the study. M.P. and S.K. verified the underlying data. All authors read and approved the final version of the manuscript.

#### Data sharing statement

For the OGE, we used LOEUF<sup>16</sup> and pLI<sup>15</sup> from the gnomAD v2.1.1 (<https://gnomad.broadinstitute.org/>) database. For the CGE, we used the fitness scores from the ProjectSCORE (<https://score.depmap.sanger.ac.uk/>) database, group of fitness genes provided by Behan et al.,<sup>9</sup> probability of dependency from the DepMap (20Q4 v2; <https://depmap.org/portal/>) database, and fitness scores for hPSCs provided by Tables S1, S2 and S7 from Mair et al.<sup>35</sup> Drugs with clinical max phase and targets were downloaded from the ChEMBL v30 (<https://www.ebi.ac.uk/chembl/>) and the STITCH 5 (<http://stitch.embl.de/>) databases. Failed drugs due to toxicity were downloaded from the ClinTox dataset of the MoleculeNet (<https://moleculenet.org/>) database and scripts (<https://github.com/kgayvert/ProCTOR>) from Gayvert et al.<sup>25</sup> Side effects of drugs were obtained from the ARReCS (v3.1; <http://bioinf.xmu.edu.cn/ADReCS/>) and the DrugCentral (v2021; <https://drugcentral.org/>) databases. Withdrawn drugs were downloaded from the ChEMBL database and Additional file 3 from Onakpoya et al.<sup>30</sup> Pathways and categories were downloaded from the BioPlanet<sup>71</sup> (<https://tripod.nih.gov/bioplanet/>) database. The human protein interaction network was downloaded from the STRING (v11.5; <https://string-db.org/>) database. Normalized expression values across

tissues were downloaded from the Human Protein Atlas (HPA v20.1; <http://v20.proteinatlas.org/>). SMILES of drugs were downloaded from the ChEMBL database. All data used in this study are publicly available and can be found in “Supplementary Table”.

The source codes for reproduction of the results were developed in python 3.7.9 and R 4.0.5. The scripts used in this study are available at our GitHub repository (<https://github.com/SBILab/DrugApprovalPrediction>).

#### Declaration of interests

SI and IK are employees of ImmunoBiome Inc. A patent application has been jointly filed by ImmunoBiome (SI and IK) and POSTECH (SK and MP) related to the data described in the manuscript. This patent application is currently pending. Other authors do not have any conflicts of interests.

#### Acknowledgements

We thank all of the members of the Kim laboratory for helpful discussions. This study was supported by grants to S.K. from the Korean National Research Foundation (2021R1A2B5B01001903 and 2020R1A6A1A03047902 and IITP (2019-0-01906, Artificial Intelligence Graduate School Program, POSTECH).

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2023.104705>.

#### References

- Arrowsmith J, Miller P. Phase II and Phase III attrition rates 2011–2012. *Nat Rev Drug Discov*. 2013;12:569. <https://doi.org/10.1038/nrd4090>.
- Harrison RK. Phase II and phase III failures: 2013–2015. *Nat Rev Drug Discov*. 2016;15:817–818. <https://doi.org/10.1038/nrd.2016.184>.
- DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ*. 2016;47:20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
- Cook D, Brown D, Alexander R, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. 2014;13:419–431. <https://doi.org/10.1038/nrd4309>.
- Wang X, Thijssen B, Yu H. Target essentiality and centrality characterize drug side effects. *PLoS Comput Biol*. 2013;9:e1003119. <https://doi.org/10.1371/journal.pcbi.1003119>.
- Perez-Lopez AR, Szalay KZ, Türei D, et al. Targets of drugs are generally and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. *Sci Rep*. 2015;5:10182. <https://doi.org/10.1038/srep10182>.
- Piñero J, Gonzalez-Perez A, Guney E, et al. Network, transcriptomic and genomic features differentiate genes relevant for drug response. *Front Genet*. 2018;9:412. <https://doi.org/10.3389/fgene.2018.00412>.
- Duffy Á, Verbanck M, Dobbyn A, et al. Tissue-specific genetic features inform prediction of drug side effects in clinical trials. *Sci Adv*. 2020;6:eabb6242. <https://doi.org/10.1126/sciadv.abb6242>.
- Behan FM, Iorio F, Picco G, et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*. 2019;568:511–516. <https://doi.org/10.1038/s41586-019-1103-9>.
- Chang L, Ruiz P, Ito T, Sellers WR. Targeting pan-essential genes in cancer: challenges and opportunities. *Cancer Cell*. 2021;39:466–479. <https://doi.org/10.1016/j.ccell.2020.12.008>.
- Clements M, Millar V, Williams AS, Kalinka S. Bridging functional and structural cardiotoxicity assays using human embryonic stem cell-derived cardiomyocytes for a more comprehensive risk assessment. *Toxicol Sci*. 2015;148:241–260. <https://doi.org/10.1093/toxsci/kfv180>.
- Qureshi ZP, Seoane-Vazquez E, Rodriguez-Monguio R, Stevenson KB, Szeinbach SL. Market withdrawal of new molecular entities approved in the United States from 1980 to 2009. *Pharmacoepidemiol Drug Saf*. 2011;20:772–777. <https://doi.org/10.1002/pds.2155>.
- Reith MEA, Kortagere S, Wiers CE, et al. The dopamine transporter gene SLC6A3: multidisease risks. *Mol Psychiatry*. 2021;27:1031. <https://doi.org/10.1038/s41380-021-01341-5>.
- Meyers RM, Bryan JG, McFarland JM, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49:1779–1784. <https://doi.org/10.1038/ng.3984>.
- Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291. <https://doi.org/10.1038/nature19057>.
- Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–443. <https://doi.org/10.1038/s41586-020-2308-7>.
- Whiffin N, Armean IM, Kleinman A, et al. The effect of LRRK2 loss-of-function variants in humans. *Nat Med*. 2020;26:869–877. <https://doi.org/10.1038/s41591-020-0893-5>.
- Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res*. 2019;47:D930–D940. <https://doi.org/10.1093/nar/gky1075>.
- Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res*. 2016;44:D380–D384. <https://doi.org/10.1093/nar/gkv1277>.
- Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Uhlen M, Fagerberg L, Hallstrom BM, et al. Tissue-based map of the human proteome. *Science*. 2015;347:1260419. <https://doi.org/10.1126/science.1260419>.
- Rouillard AD, Hurler MR, Agarwal P. Systematic interrogation of diverse Omic data reveals interpretable, robust, and generalizable transcriptomic features of clinically successful therapeutic targets. *PLoS Comput Biol*. 2018;14:e1006142. <https://doi.org/10.1371/journal.pcbi.1006142>.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
- Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2018;9:513–530. <https://doi.org/10.1039/C7SC02664A>.
- Gayvert KM, Madhukar NS, Elemento O. A data-driven approach to predicting successes and failures of clinical trials. *Cell Chem Biol*. 2016;23:1294–1301. <https://doi.org/10.1016/j.chembiol.2016.07.023>.
- Cai M-C, Xu Q, Pan Y-J, et al. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic Acids Res*. 2015;43:D907–D913. <https://doi.org/10.1093/nar/gku1066>.
- Avram S, Bologna CG, Holmes J, et al. DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res*. 2021;49:D1160–D1169. <https://doi.org/10.1093/nar/gkaa997>.
- Mozzicato P. MedDRA. *Pharmaceut Med*. 2009;23:65–75. <https://doi.org/10.1007/BF03256752>.
- Hunter FMI, Bento AP, Bosc N, Gaulton A, Hersey A, Leach AR. Drug safety data curation and modeling in ChEMBL: boxed warnings and withdrawn drugs. *Chem Res Toxicol*. 2021;34:385–395. <https://doi.org/10.1021/acs.chemrestox.0c00296>.
- Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med*. 2016;14:10. <https://doi.org/10.1186/s12916-016-0553-2>.
- Huang R, Grishagin I, Wang Y, et al. The NCATS BioPlanet – an integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics. *Front Pharmacol*. 2019;10:1–13. <https://doi.org/10.3389/fphar.2019.00445>.
- Raudvere U, Kolberg L, Kuzmin I, et al. Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47:W191–W198. <https://doi.org/10.1093/nar/gkz369>.
- Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov*. 2019;18:495–496. <https://doi.org/10.1038/d41573-019-00074-z>.
- Wu L, Ingle T, Liu Z, et al. Study of serious adverse drug reactions using FDA-approved drug labeling and MedDRA. *BMC Bioinformatics*. 2019;20:97. <https://doi.org/10.1186/s12859-019-2628-5>.
- Mair B, Tomic J, Masud SN, et al. Essential gene profiles for human pluripotent stem cells identify uncharacterized genes and



- substrate dependencies. *Cell Rep.* 2019;27:599–615.e12. <https://doi.org/10.1016/j.celrep.2019.02.041>.
- 36 Stephens TD, Bunde CJW, Fillmore BJ. Mechanism of action in thalidomide teratogenesis. *Biochem Pharmacol.* 2000;59:1489–1499. [https://doi.org/10.1016/S0006-2952\(99\)00388-3](https://doi.org/10.1016/S0006-2952(99)00388-3).
  - 37 Van Battum EY, Brignani S, Pasterkamp RJ. Axon guidance proteins in neurological disorders. *Lancet Neurol.* 2015;14:532–546. [https://doi.org/10.1016/S1474-4422\(14\)70257-1](https://doi.org/10.1016/S1474-4422(14)70257-1).
  - 38 Touyz RM, Herrmann J. Cardiotoxicity with vascular endothelial growth factor inhibitor therapy. *NPJ Precis Oncol.* 2018;2:13. <https://doi.org/10.1038/s41698-018-0056-z>.
  - 39 Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16:19–34. <https://doi.org/10.1038/nrd.2016.230>.
  - 40 Kotlyar M, Fortney K, Jurisica I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods.* 2012;57:499–507. <https://doi.org/10.1016/j.ymeth.2012.06.003>.
  - 41 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997;23:3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1).
  - 42 Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem.* 2002;45:2615–2623. <https://doi.org/10.1021/jm020017n>.
  - 43 Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem.* 1999;1:55–68. <https://doi.org/10.1021/cc9800071>.
  - 44 Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell.* 2013;35:1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
  - 45 Jensen C, Teng Y. Is it time to start transitioning from 2D to 3D cell culture? *Front Mol Biosci.* 2020;7:1–15. <https://doi.org/10.3389/fmolb.2020.00033>.
  - 46 Kapaczynska M, Kolenda T, Przybyla W, et al. 2D and 3D cell cultures – a comparison of different types of cancer cell cultures. *Arch Med Sci.* 2016;14:910–919. <https://doi.org/10.5114/aoms.2016.63743>.
  - 47 Caldu-Primo JL, Verduzco-Martínez JA, Alvarez-Buylla ER, Davila-Velderrain J. In vivo and in vitro human gene essentiality estimations capture contrasting functional constraints. *NAR Genom Bioinform.* 2021;3:1–14. <https://doi.org/10.1093/nargab/lqab063>.
  - 48 Lo AW, Siah KW, Wong CH. Machine learning with statistical imputation for predicting drug approval. *Harvard Data Sci Rev.* 2019;1–38. <https://doi.org/10.1162/99608f92.5c5f0525>.
  - 49 Siah KW, Kelley NW, Ballerstedt S, et al. Predicting drug approvals: the Novartis data science and artificial intelligence challenge. *Patterns.* 2021;2:100312. <https://doi.org/10.1016/j.patter.2021.100312>.
  - 50 Clark LT, Watkins L, Piña IL, et al. Increasing diversity in clinical trials: overcoming critical barriers. *Curr Probl Cardiol.* 2019;44:148–172. <https://doi.org/10.1016/j.cpcardiol.2018.11.002>.
  - 51 Wright JT. Outcomes in hypertensive black and nonblack patients treated with chlorthalidone, amlodipine, and lisinopril. *JAMA.* 2005;293:1595. <https://doi.org/10.1001/jama.293.13.1595>.
  - 52 Carson P, Ziesche S, Johnson G, Cohn JN. Racial differences in response to therapy for heart failure: analysis of the vasodilator-heart failure trials. *J Card Fail.* 1999;5:178–187. [https://doi.org/10.1016/S1071-9164\(99\)90001-5](https://doi.org/10.1016/S1071-9164(99)90001-5).
  - 53 Nguyen PA, Born DA, Deaton AM, Nioi P, Ward LD. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun.* 2019;10:1579. <https://doi.org/10.1038/s41467-019-09407-3>.
  - 54 Carss KJ, Deaton AM, Del Rio-Espinola A, et al. Using human genetics to improve safety assessment of therapeutics. *Nat Rev Drug Discov.* 2022;22:0123456789. <https://doi.org/10.1038/s41573-022-00561-w>.
  - 55 Minikel EV, Karczewski KJ, Martin HC, et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature.* 2020;581:459–464. <https://doi.org/10.1038/s41586-020-2267-z>.
  - 56 Zuberi A, Lutz C. Mouse models for drug discovery. Can new tools and technology improve translational power? *ILAR J.* 2016;57:178–185. <https://doi.org/10.1093/ilar/ilw021>.
  - 57 Zambrowicz BP, Sands AT. Knockouts model the 100 best-selling drugs—will they model the next 100? *Nat Rev Drug Discov.* 2003;2:38–51. <https://doi.org/10.1038/nrd987>.
  - 58 Van Norman GA. Limitations of animal studies for predicting toxicity in clinical trials. *JACC Basic Transl S.* 2019;4:845–854. <https://doi.org/10.1016/j.jacbs.2019.10.008>.
  - 59 Wang B, Gray G. Concordance of noncarcinogenic endpoints in rodent chemical bioassays. *Risk Anal.* 2015;35:1154–1166. <https://doi.org/10.1111/risa.12314>.
  - 60 Bailey J, Thew M, Balls M. An analysis of the use of animal models in predicting human toxicology and drug safety. *Altern Lab Anim.* 2014;42:181–199. <https://doi.org/10.1177/026119291404200306>.
  - 61 Han SK, Kim D, Lee H, Kim I, Kim S. Divergence of noncoding regulatory elements explains gene–phenotype differences between human and mouse orthologous genes. *Mol Biol Evol.* 2018;35:1653–1667. <https://doi.org/10.1093/molbev/msy056>.
  - 62 Ha D, Kim D, Kim I, et al. Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. *Nucleic Acids Res.* 2022;50:1849–1863. <https://doi.org/10.1093/nar/gkac050>.
  - 63 Ursu O, Rayan A, Goldblum A, Oprea TI. Understanding drug-likeness. *WIREs Comput Mol Sci.* 2011;1:760–781. <https://doi.org/10.1002/wcms.52>.
  - 64 Cai C, Lin H, Wang H, et al. miDruglikeness: subdivisional drug-likeness prediction models using active ensemble learning strategies. *Biomolecules.* 2022;13:29. <https://doi.org/10.3390/biom13010029>.
  - 65 Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov.* 2007;6:881–890. <https://doi.org/10.1038/nrd2445>.
  - 66 Duran-Frigola M, Aloy P. Analysis of chemical and biological features yields mechanistic insights into drug side effects. *Chem Biol.* 2013;20:594–603. <https://doi.org/10.1016/j.chembiol.2013.03.017>.
  - 67 Juan-Blanco T, Duran-Frigola M, Aloy P. IntSide: a web server for the chemical and biological examination of drug side effects. *Bioinformatics.* 2015;31:612–613. <https://doi.org/10.1093/bioinformatics/btu688>.
  - 68 Duran-Frigola M, Mosca R, Aloy P. Structural systems pharmacology: the role of 3D structures in next-generation drug development. *Chem Biol.* 2013;20:674–684. <https://doi.org/10.1016/j.chembiol.2013.03.004>.
  - 69 Zhang C, Hong H, Mendrick DL, Tang Y, Cheng F. Biomarker-based drug safety assessment in the age of systems pharmacology: from foundational to regulatory science. *Biomark Med.* 2015;9:1241–1252. <https://doi.org/10.2217/bmm.15.81>.
  - 70 Keller TH, Pichota A, Yin Z. A practical view of ‘druggability’. *Curr Opin Chem Biol.* 2006;10:357–361. <https://doi.org/10.1016/j.cbpa.2006.06.014>.
  - 71 Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov.* 2002;1:727–730. <https://doi.org/10.1038/nrd892>.
  - 72 Liu A, Seal S, Yang H, Bender A. Using chemical and biological data to predict drug toxicity. *SLAS Discov.* 2023;28:53–64. <https://doi.org/10.1016/j.slasd.2022.12.003>.
  - 73 Hao Y, Romano JD, Moore JH. Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns.* 2022;3:100565. <https://doi.org/10.1016/j.patter.2022.100565>.