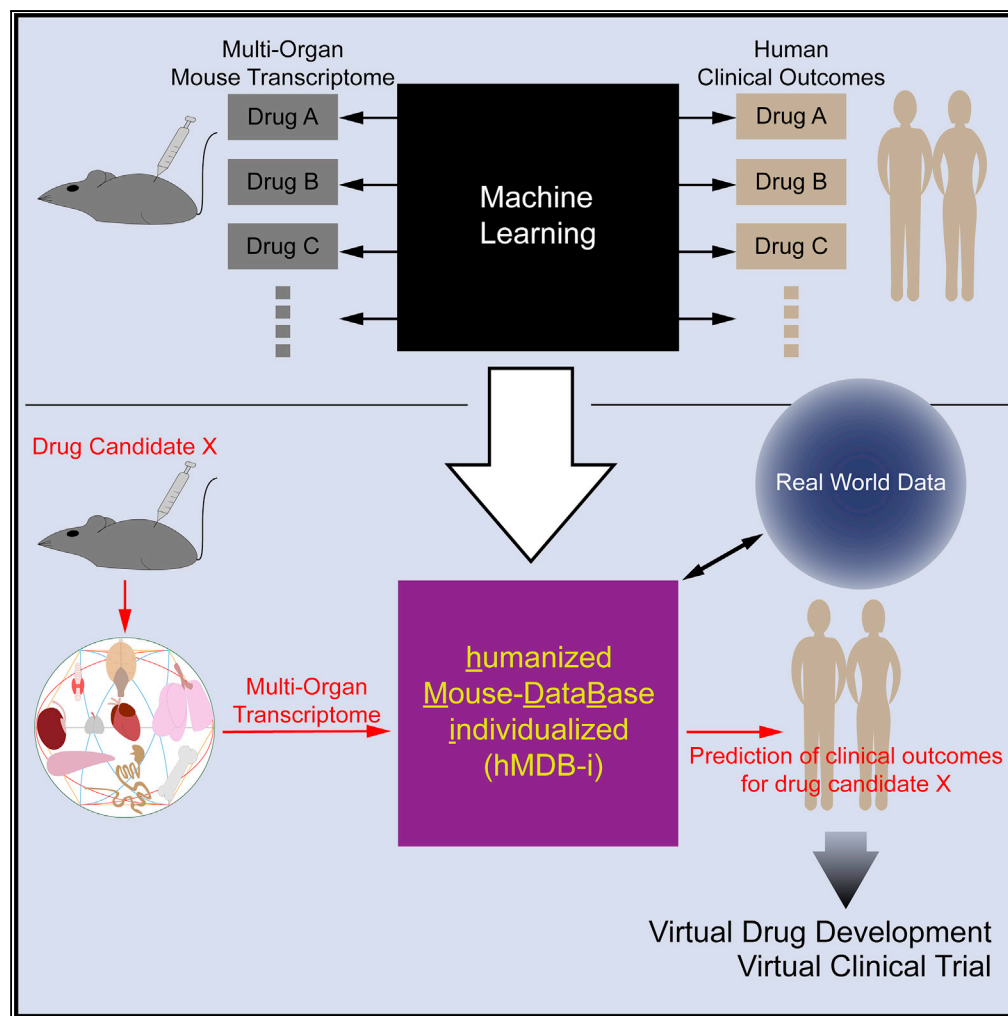


## Article

# Predicting Human Clinical Outcomes Using Mouse Multi-Organ Transcriptome



Satoshi Kozawa,  
Fumihiko Sagawa,  
Satsuki Endo,  
Glicia Maria De  
Almeida, Yuto  
Mitsuishi, Thomas  
N. Sato

island1005@gmail.com

## HIGHLIGHTS

Mouse-transcriptome humanized by machine learning predicts human clinical outcomes

This system is referred to as humanized Mouse DataBase individualized, hMDB-i

hMDB-i predicts drug adverse events and therapeutic indications

hMDB-i is a bias-free and modality-independent system for virtual drug development

## DATA AND CODE

### AVAILABILITY

GSE142068

Kozawa et al., iScience 23,  
100791  
February 21, 2020 © 2019 The  
Author(s).  
[https://doi.org/10.1016/  
j.isci.2019.100791](https://doi.org/10.1016/j.isci.2019.100791)



## Article

# Predicting Human Clinical Outcomes Using Mouse Multi-Organ Transcriptome

Satoshi Kozawa,<sup>1,2,3</sup> Fumihiko Sagawa,<sup>1,2,3</sup> Satsuki Endo,<sup>1,2,3</sup> Glicia Maria De Almeida,<sup>1</sup> Yuto Mitsuishi,<sup>1</sup> and Thomas N. Sato<sup>1,2,3,4,5,6,7,\*</sup>

## SUMMARY

Approximately 90% of pre-clinically validated drugs fail in clinical trials owing to unanticipated clinical outcomes, costing over several hundred million US dollars per drug. Despite such critical importance, translating pre-clinical data to clinical outcomes remain a major challenge. Herein, we designed a modality-independent and unbiased approach to predict clinical outcomes of drugs. The approach exploits their multi-organ transcriptome patterns induced in mice and a unique mouse-transcriptome database “humanized” by machine learning algorithms and human clinical outcome datasets. The cross-validation with small-molecule, antibody, and peptide drugs shows effective and efficient identification of the previously known outcomes of 5,519 adverse events and 11,312 therapeutic indications. In addition, the approach is adaptable to deducing potential molecular mechanisms underlying these outcomes. Furthermore, the approach identifies previously unsuspected repositioning targets. These results, together with the fact that it requires no prior structural or mechanistic information of drugs, illustrate its versatile applications to drug development process.

## INTRODUCTION

Unexpected adverse events (AEs) and/or the lack of the expected efficacy in human subjects fails drug development. The average success rate of drug candidates through clinical trials is 13.8% (Wong et al., 2019). This low success rate costs US\$161M–1.8B per drug candidate on drug developers (e.g., pharmaceutical and biotech companies) (Morgan et al., 2011), leading to drug price hike and rising medical cost. Hence, effective prediction of clinical outcomes from the pre-clinical studies improves the success rate of drug development and reduces the drug price and medical cost.

A major impediment in drug development is the unpredictability of human outcomes on the basis of the pre-clinical results using cells, organs, and/or animal models. To overcome this problem, numerous technological solutions, both experimental and computational approaches, are currently undertaken. The uses of human cells and organs *in vitro* and *in vivo* models are among such approaches. Human cells such as induced human pluripotent cells (human iPSCs) (Elitt et al., 2018; Ko and Gelb, 2014; Meseguer-Ripolles et al., 2018) and an organ(s)-on-a-chip consisting of human cells (Oleaga et al., 2016; Rezaei Kollahchi et al., 2016) are currently utilized as drug screening and *in vitro* toxicity assays. As *in vivo* models, partially humanized mouse models, such as those where the liver is nearly 100% composed of human cells, is exploited (Tateno et al., 2004). In addition to such experimental approaches, computational tools are also invented and used. In particular, applications of machine learning algorithms for predicting clinical outcomes are in fashion (Shah et al., 2019; Vamathevan et al., 2019). They exploit the big-data sets representing structural and functional features of drugs and their target information.

Although both of such experimental and computational approaches have shown some success and promise, there are certain limitations with these existing approaches. The existing machine learning approaches require prior knowledge about the characteristics (such as structure) of drugs and their mechanisms of actions (such as target molecules). In addition, many of the computational approaches are specialized for the drugs of specific modality such as small-molecule compounds (i.e., mixed structures and mechanisms). Hence, they have difficulty dealing with the mixture of compounds. The existing experimental approaches are often “biased”, as they can design assay system according to what the testers want to examine. For example, the testers need to know which organ(s) or phenotype(s) to examine for the drug effect(s) prior to designing the experiment(s). Furthermore, these presently available experimental and computational approaches fail to recapitulate the organismal level body-wide drug effects in human. Hence, an approach that recapitulates organismal biology but does not require any prior knowledge about the drug structure or

<sup>1</sup>Karydo TherapeutiX, Inc., Kyoto, Japan

<sup>2</sup>ERATO Sato Live Bio-Forecasting Project, Kyoto, Japan

<sup>3</sup>The Thomas N. Sato BioMEC-X Laboratories, Advanced Telecommunications Research Institute International, 2-2-2 Hikoridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

<sup>4</sup>Biomedical Engineering, Cornell University, Ithaca, NY, USA

<sup>5</sup>Centenary Institute, Sydney, Australia

<sup>6</sup>V-iClinix Laboratory, Nara Medical University, Nara, Japan

<sup>7</sup>Lead Contact

\*Correspondence:

island1005@gmail.com

<https://doi.org/10.1016/j.isci.2019.100791>



mechanisms (i.e., modality-independent and unbiased) is expected to advance and facilitate the drug development process.

At the whole-body level, drugs could act on not only one or a few specific organ(s), but also multiple organs showing complex effects (Berger and Iyengar, 2011). Drugs could also act on one organ that influences the functions of other organs via inter-organ cross talks (Droujinine and Perrimon, 2013, 2016). Hence, measuring the drug effects on large number of organs may provide useful features that could be exploited in predicting the drug effects at the whole-body level. However, this assay is only possible with animal models but not with human subjects.

We previously conducted multi-organ multi-model transcriptome studies and identified model-specific and organ-specific transcriptome patterns in mice (Kozawa et al., 2018). Furthermore, weighted correlation network analysis (WGCNA) of the gene expression across multiple organs identified a number of putative organ-to-organ cross talks, some of which we genetically validated (Kozawa et al., 2018). On the basis of this previous study, we considered that such multi-organ transcriptome patterns could be exploited as the unique features of drugs at the whole-body level to predict its clinical outcomes. Furthermore, if successful, the approach provides a modality-independent and unbiased system requiring no prior knowledge of the structures/mechanisms of drugs, as the only requirement is that the drugs can be administered to mouse.

In the current study, we develop and show the performance of such system. We design a hybrid approach of experimental and machine-learning methods to utilize the multi-organ transcriptome patterns induced by drugs in mice to predict their clinical outcomes in human. We evaluate its prediction performance for both AEs and therapeutic indications by cross-validation. In addition, we also test whether the approach could be adapted to deducing molecular and cellular mechanisms underlying such predictions. Furthermore, we illustrate a possibility of applying this approach to identifying drug repositioning targets.

## RESULTS

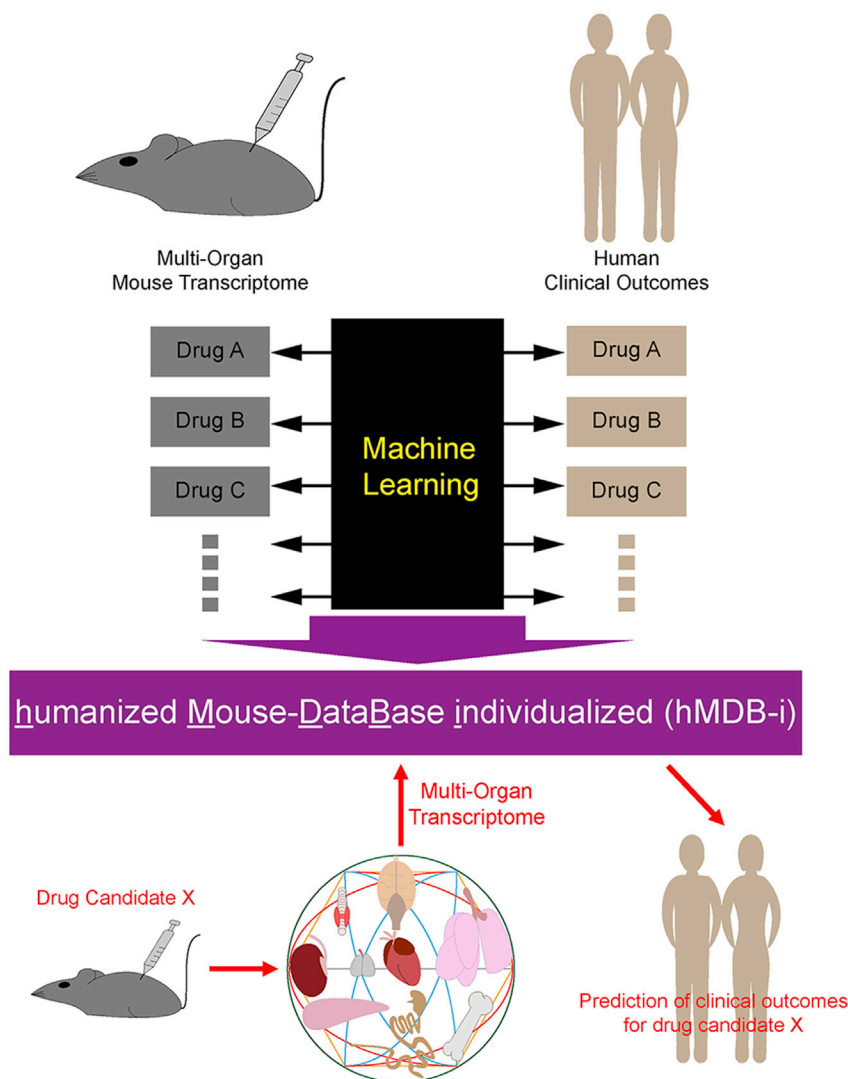
### “Humanizing” Mouse Datasets

The concept of our approach is described in Figure 1. First, we generate transcriptome data across 24 organs from the mice to which we administer each drug with known clinical outcomes. Next, for each human clinical criterion (e.g., AE, therapeutic indication), we train a machine learning model with the 24-organ transcriptome data induced by the drug in mice and its known outcomes in human. The outcome data are classified according to the demographic profiles (e.g., sex, age) of the individuals. Consequently, the drug-induced 24-organ mouse transcriptome patterns are associated with the individualized human clinical outcomes in the machine learning model/database, hence referred to as “humanized Mouse-DataBase, individualized (hMDB-i).” To predict clinical outcomes of a new drug candidate X, we generate the 24-organ transcriptome data from the mice to which X is administered. Such data are used as the input data into hMDB-i to predict X’s putative individualized clinical outcomes (i.e., the output).

### AEs Prediction by Support Vector Machine Algorithm

We generated the 24-organ transcriptome datasets with 15 drugs of various modalities (i.e., small molecules, antibody, and peptide) with diverse known therapeutic indications (see Transparent Methods, Figure 1 and Table S1). The incidence of AEs reported for each drug in each sex and each age group were compiled from US Food and Drug (FDA) Adverse Event Reporting System (FAERS). For each AE (a total of 5,519 AEs is reported for one or more of the 15 drugs), we train a support vector machine (SVM) algorithm with the drug-induced 24-organ transcriptome patterns to predict the individualized outcomes.

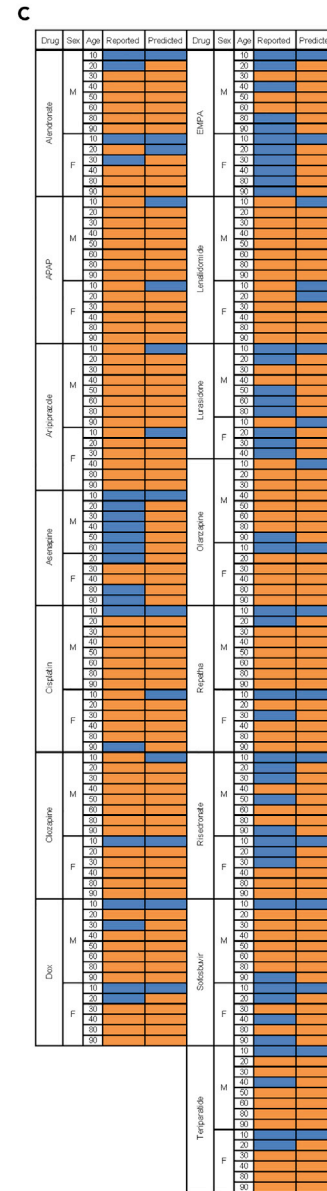
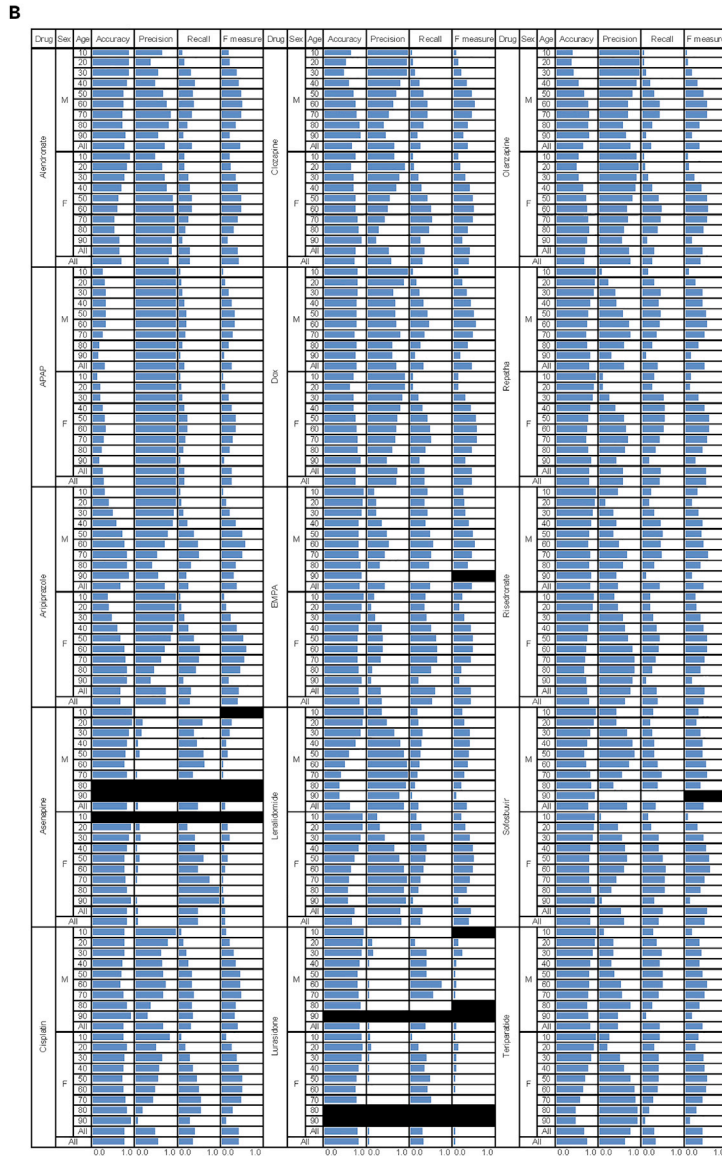
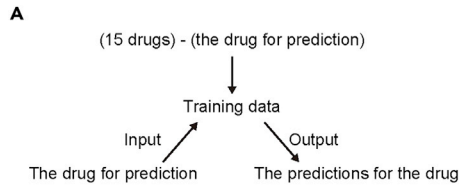
We performed cross-validation to evaluate the effectiveness of this approach to predict the outcomes of the AEs (see Transparent Methods, Figure 2A). We first train an SVM algorithm with the 14 drug-induced 24-organ mouse transcriptome data. We then input the 24-organ mouse transcriptome data of the omitted drug as “an unseen drug data” (see Transparent Methods, Figure 2A). We performed such evaluations for all 15 drugs by omitting one drug at a time from the training data. We used an SVM classifier to predict whether each AE would be observed by each drug for each sex and age-group (see Transparent Methods). The scores of accuracy, precision, recall, and F-measure of the prediction result of the AEs reported for each sex and age group for each drug are summarized in Figure 2B and Table S2.



**Figure 1. Conceptual Framework of hMDB-i**

See text for the description.

The result shows that the accuracy scores are over 0.7 for the 200 of 264 sex/age-groups across the 15 drugs and over 0.9 for the 52 sex/age-groups. This indicates that more than 70% of the AEs predicted to either appear or not appear for the drug are indeed reported or not reported, respectively, in these sex/age-groups. The precision scores are over 0.5 for the 156 of the 264 sex/age-groups, indicating that more than 50% of the AEs predicted to appear for the drug are indeed reported in these sex/age-groups. If the question is whether an AE predicted to appear for a drug is indeed reported in at least one or more of the sex/age group, 13 drugs (alendronate, acetaminophen, aripiprazole, cisplatin, clozapine, doxycycline, empagliflozin, lenalidomide, olanzapine, evolocumab [Repatha], risenedronate, sofosbuvir, teriparatide) show the precision scores of over 0.5 (see "All" row in each drug in Figure 2B), indicating that more than 50% of the AEs predicted to appear for these drugs are indeed reported at some sex/age groups. Furthermore, for nine drugs (alendronate, acetaminophen, aripiprazole, cisplatin, clozapine, doxycycline, lenalidomide, olanzapine, teriparatide), the precision scores of 0.9–1.0 are found at least in one or more the sex/age groups. The recall score indicates how many of the reported AEs are indeed predicted for the drug in each sex/age group. The recall scores are relatively lower across all drugs and sex/age groups, indicating that the approach misses many of the reported AEs. However, for such drugs (asenapine and lurasidone), the recall scores of 0.7–1.0 are found for certain sex/age groups. Sample size differences among different



**D**

Drug	TP	TN	FP	FN	Accuracy	Precision	Recall	F-measure
Alendronate	51	2810	676	18				
APAP	95	2330	600	36				
Aprepitant	435	3011	538	220				
Aspirin	79	1740	758	9				
Ciprofloxacin	69	2163	795	34				
Clozapine	129	2174	677	50				
Dox	88	2247	693	33				
EMIPA	30	2195	841	15				
Lisinopril	270	2291	533	136				
Lisinopril	92	1540	767	13				
Rosuvastatin	220	2195	545	102				
Sildenafil	103	2224	656	28				
Sildenafil	57	2260	741	9				
Temazepam	70	2232	750	9				

**Figure 2. Summary of the AE Prediction Results by SVM/hMDB-i**

(A) Hold-out scheme.

(B) Confusion matrix summarizing the AE prediction results for each drug. Scores of accuracy (Accuracy), precision (Precision), recall (Recall), and F-measure (F measure) for each drug (Drug) are indicated by bars (light blue) for each age group (Age) and sex (Sex). The scale of the scores (0.0–1.0) are indicated at the bottom. The black filled boxes across all scores indicate that no AEs are reported for this drug in this sex/age group. When the denominator is 0 in the F-measure calculation, the score is incalculable, hence shown as black filled box.

(C) The prediction results of death event for each drug (Drug) are summarized for each age group (Age) and sex (Sex) using FAERS. The reported outcomes (Reported) are shown as orange and light-blue filled boxes indicating death event is reported and not-reported, respectively. The prediction results (Predicted) are shown as orange and light-blue filled boxes indicating death event is predicted to occur and not-to-occur, respectively. The age groups are 10: 10–19 years old, 20: 20–29 years old, 30: 30–39 years old, 40: 40–49 years old, 50: 50–59 years old, 60: 60–69 years old, 70: 70–79 years old, 80: 80–89 years old, 90: 90–99 years old, All: 10–99 years old. F, females; M, males; APAP, Acetaminophen; Dox, doxycycline; EMPA, empagliflozin; Repatha, evolocumab. The raw data are available as [Table S2](#).

(D) The confusion matrix showing the prediction result using SIDER4.1 reference database. The raw data are available as [Table S3](#).

sex/age groups and the number of the reports for each AE may significantly impact on the accuracy, precision, and recall scores. In fact, the number of the reported AEs for asenapine and lurasidone is significantly smaller than those of the other drugs and their precision is lower.

In FAERS, many of the AEs (e.g., diarrhea, abdominal pain) are common among many drugs. The outcomes of rare but serious AEs (SAEs) are often more important to know in advance to clinical trials. Hence, we assessed the prediction performance on SAEs. The top of such SAE hierarchy is “death event” (Sonawane et al., 2018). The result shows that the correct outcomes of death event are predicted for 148 of 204 sex/age groups across the 15 drugs (Figure 2C). The death events are reported in a total of 141 sex/age groups, and our prediction approach missed its incidence only in 12 groups (Figure 2C).

To examine potential influences of reference-specific errors, biases, and/or cofounders that are often found in any real-world data such as FAERS (Harpaz et al., 2013; Tatonetti et al., 2012), we validated our prediction with another reference, Side Effect Resource (SIDER) (<http://sideeffects.embl.de>). Several differences between FAERS and SIDER must be noted. SIDER4.1 version does not include evolocumab (Repatha). The outcomes reported in SIDER4.1 are not classified according to sex/age groups. The SEs and AEs are not necessarily represented by the same terms. Considering these differences, we examined the outcome predictions for 14 drugs (i.e., 15 drugs minus evolocumab) without the sex/age group classifications. In addition, the SEs/AEs common to both reference databases were analyzed. The result, 0.677–0.827 accuracy, 0.023–0.523 precision, and 0.571–0.963 recall, is comparable with that with FAERS (Figure 2D, the full and raw data for the confusion matrix is available as [Table S3](#)), further supporting the effectiveness of our method.

Next, we evaluated the necessity of multiple organ transcriptome. Figure 3 shows the SAEs that the 24-organ transcriptome correctly predicts the outcome for each drug (Figure 3 and [Table S4](#)). The predictions of these SAEs by individual-organ transcriptome data in the cross-validation scheme show variable results among different organs (Figure 3). Although this result indicates that some of the individual-organ datasets are sufficient to predict the correct outcomes, which individual-organ dataset(s) is(are) necessary varies among which SAEs to be predicted (Figure 3). Hence, it appears beneficial to collect all 24-organ transcriptome data.

**AE Prediction by Random Forest Algorithm**

We tested another machine learning algorithm, random forest (RF), to predict the AE outcomes. We compared the prediction results for death event outcomes (Figure 4 and [Table S5](#)). The result shows that both SVM and RF provides the same predictions for all sex/age groups (Figure 4A). The calculations of accuracy, precision, recall, and F-measure scores by both algorithms show the similar results for all 15 drugs (Figure 4B). For all drugs except asenapine, empagliflozin, and lurasidone, all scores of accuracy/precision/recall are 0.5–1.0. Although the accuracy and the precision scores for asenapine, empagliflozin, and lurasidone are 0.2–0.4, the recall scores for these three drugs by both SVM and RF are 1.0 (asenapine, empagliflozin) and 0.667 (lurasidone), indicating that the occurrence of death event by these three drugs are efficiently predicted across all sex/age groups (i.e., not missing the possible occurrence of death event). The results indicate that both algorithms are equivalent in predicting AE outcomes.



**Figure 3. Continued**

data are shown. AdrenalG, adrenal gland; BM, bone marrow; WAT, gonadal white adipose tissue; ParotidG, parotid gland; PituitaryG, pituitary gland; SkMuscle, skeletal muscle; ThyroidG, thyroid gland. The age groups are 10: 10–19 years old, 20: 20–29 years old, 30: 30–39 years old, 40: 40–49 years old, 50: 50–59 years old, 60: 60–69 years old, 70: 70–79 years old, 80: 80–89 years old, 90: 90–99 years old. F, females; M, males; APAP, Acetaminophen; Dox, doxycycline; EMPA, empagliflozin; Repatha, evolocumab. The result is also included in [Table S4](#).

**AE Prediction by Majority Decision Framework**

Although the performance with two different algorithms (SVM and RF) were equivalent ([Figure 4](#)), we obtained differential results with the 24-organ and the individual-organ approaches ([Figures 2 and 3](#)). In the case of the individual-organ approach, which individual-organ dataset(s) to be used was critical for predicting the correct outcomes ([Figure 3](#)). Hence, it may be beneficial to conduct predictions with both the 24-organ and all individual-organs to maximize the effectiveness of the hMDB-i approach. For this purpose, we designed a majority decision framework and evaluated its effectiveness for predicting death event outcomes with alendronate and lenalidomide in female/20s group where the SVM approach with the 24-organ transcriptome dataset alone failed to predict its occurrence (see [Methods](#) for the detailed description, [Figure 4C](#)). The result shows that this majority decision framework effectively predicts the correct outcomes ([Figure 4C](#)).

**AE Prediction by Link-Prediction Framework**

Next, we tested the power of hMDB-i for the prediction of AE outcomes in a different framework. The relationship between the adverse effects and therapeutic indications has previously been exploited for clinical outcome predictions ([Zhang et al., 2013](#)). Hence, we examined whether such data could be utilized to enhance the prediction capability of hMDB-i. For this purpose, we devised a link-prediction (LP) framework as shown in [Figure 5A](#) (see [Transparent Methods](#) for the detailed description). Three training datasets are used: (1) the 24-organ transcriptome data of each of the 15 drugs, (2) AEs reported for each of all drugs in FAERS, and (3) all AEs reported in FAERS for each of all indications in FAERS ([Figure 5A](#)). In the LP framework, one-class SVM algorithm is used to train the models and the presence/absence of a link of an untrained drug (Drug Candidate X in [Figure 5A](#)), based on its 24-organ transcriptome pattern, to each AE is determined (see [Transparent Methods](#) for the detailed description).

The LP prediction of all AEs for three drugs, alendronate, clozapine, and evolocumab (Repatha), was conducted, and the result is summarized as a confusion matrix ([Figure 5B](#), [Table S6](#)). Although the accuracy and precision scores are slightly better with hMDB-i alone (SVM), the LP framework (LP) improves hMDB-i in the recall scores ([Figure 5B](#)). This result indicates two properties of these two approaches: (1) The AE outcomes predicted by hMDB-i alone is more likely to be observed than those with the LP framework; (2) The LP framework is superior to hMDB-i alone in not missing the occurrence of potential AEs. The examples of such superior property of the LP framework are shown in [Figure 5C](#) listing some of the reported SAEs that are missed by the hMDB-i alone but captured by the LP framework. The full list of such can be found in [Table S6](#).

**Biological Insights into the AE Mechanisms**

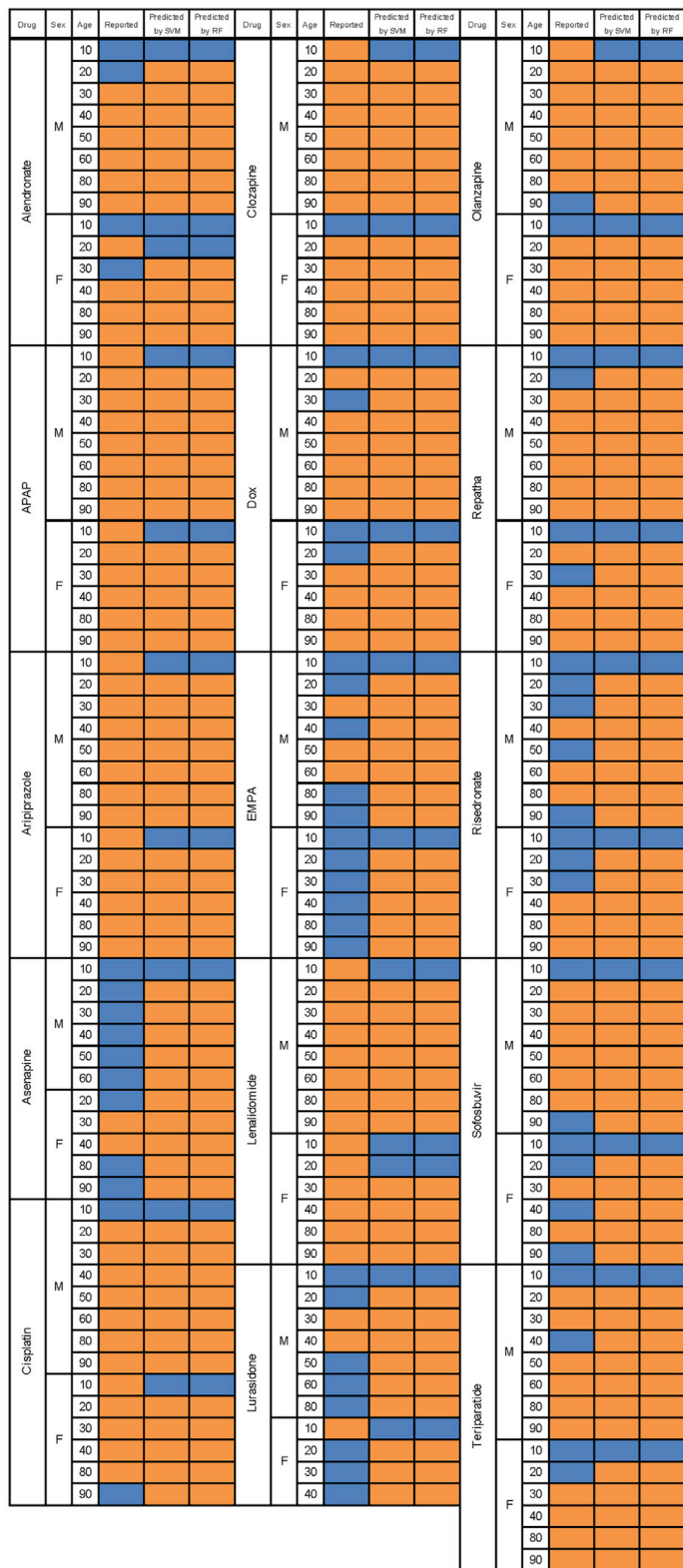
Insights into the biological mechanisms underlying AEs provide opportunities for designing the strategies to reduce the incidence of AEs during the drug development processes. For this purpose, RF algorithm is useful as it calculates feature importance of organ-gene datasets for the correct outcome predictions. [Tables 1 and 2](#) show organs and cellular/molecular pathways that were found important for predicting the correct outcomes of the death event for males at 50s for aripiprazole. Top 5 and Top 8 organ-pathway combinations identified by REACTOME ([Table 1](#)) and KEGG ([Table 2](#)), respectively, are shown. The results for this and other drugs and age/sex groups are summarized in [Tables S7 and S8](#). The result indicates a possibility that the hMDB-i is also useful for deducing mechanisms underlying the predicted AE outcomes.

**Predicting the Outcome Dynamics over Ages**

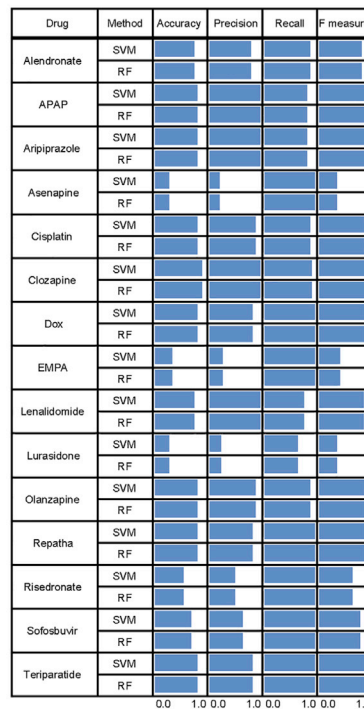
Information regarding the quantitative differences of the AE occurrence among the sex/age group(s) would be beneficial for designing efficient clinical trials. It allows for selecting target subjects with less chance of observing an AE in the trial. Hence, we next evaluated the performance for predicting quantitative changes for the outcome incidences of AEs (see [Transparent Methods](#) for the detailed description, [Figure 6](#), [Table S9](#)). For this purpose, we applied support vector regressor (SVR) and random forest regressor (RFR) algorithms to the hMDB-i framework. The number of death event reports changes over age to



A



B



C

Drug	Sex	Age	Reported	Predicted by SVM with all 24 organs	Predicted results	Predicted results	Majority decision
Alendronate	F	20	Orange	Blue	17	33	Orange
Lenalidomide	F	20	Orange	Blue	23	27	Orange

#### Figure 4. The Comparison of Death Event Prediction Results by SVM and RF Algorithms

(A) The prediction results of death event for each drug by SVM and RF algorithms are summarized for each age group (Age) and sex (Sex). The reported outcomes (Reported) are shown as orange and light-blue filled boxes indicating death event is reported and not-reported, respectively. The prediction results (Predicted by SVM/RF) are shown as orange and light-blue filled boxes indicating death event is predicted to occur and not-to-occur, respectively. The age groups are 10: 10–19 years old, 20: 20–29 years old, 30: 30–39 years old, 40: 40–49 years old, 50: 50–59 years old, 60: 60–69 years old, 70: 70–79 years old, 80: 80–89 years old, 90: 90–99 years old. F, females; M, males.

(B) The prediction results for each drug by SVM and RF are summarized as a confusion matrix. Scores of accuracy (Accuracy), precision (Precision), recall (Recall), and F-measure (F measure) for each drug (Drug) are indicated by bars (light blue) for each age group (Age) and sex (Sex). The scale of the scores (0.0–1.0) are indicated at the bottom. APAP, Acetaminophen; Dox, doxycycline; EMPA, empagliflozin; Repatha, evolocumab. The results of the RF prediction are also available as [Table S5](#).

(C) Majority decision approach for death event prediction. “Voting” (i.e., majority decision) by both SVM and RF methods using the 24-organ and all individual-organ transcriptome data are conducted. The reported outcomes (Reported) are shown as orange filled box indicating death event is reported. The prediction results with the 24-organ data (Predicted by SVM with all 24 organs) are shown as light-blue filled boxes indicating “death event” is predicted to not-to-occur. The majority decisions (Majority decision) by counting all votes are shown as orange filled box voting for death event to occur. The number of “votes” for “to-occur” (Predicted results) and “not-to-occur” (Predicted results) is indicated by the numbers.

variable degrees for each drug (“Reported” in [Figure 6](#)). There also appears that the sex difference exists (F versus M in [Figure 6](#)). The comparison of the predictions by both algorithms (SVR and RFR) to the reported (Reported) incidence (death/all reported AEs) shows that the predictions by both algorithms efficiently capture the general trends of the quantitative changes of death event incidence over age for drugs/sex such as alendronate/female, doxycycline/male, aripiprazole/female ([Figure 6](#), [Table S9](#)). On the other hand, the predictions for other drugs (e.g., clozapine: both sexes, empagliflozin:female, teriparatide:female) appear less efficient ([Figure 6](#), [Table S9](#)). Although further improvements are necessary, the result suggests that the approach could be exploited to select specific sex/age groups as target subjects in clinical trials of at least some drugs.

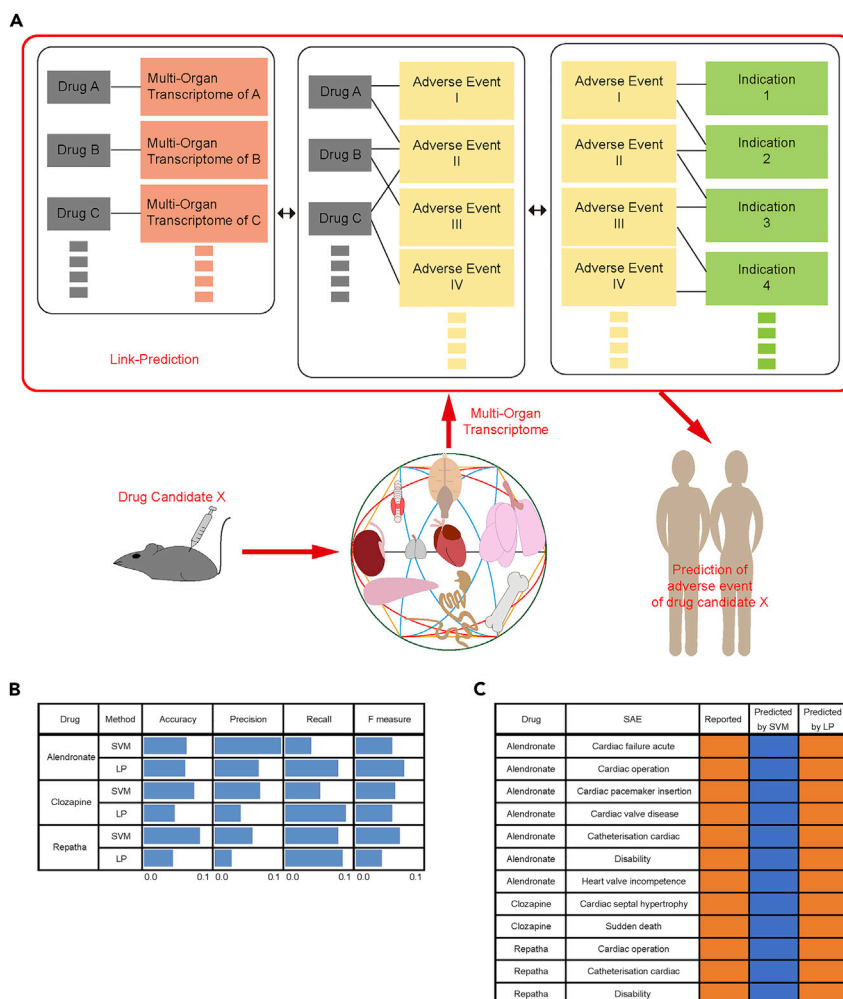
#### Predicting Therapeutic Indications

Identifying an appropriate therapeutic indication(s) (TIs) is also critical in drug development. Hence, we evaluated the utility of the multi-organ transcriptome datasets for predicting TIs. For this purpose, we adapted the LP framework ([Figure 7A](#)). Three training datasets are used: (1) the 24-organ transcriptome data of each drug of the 15 drugs, (2) all indications reported for each of the 15 drugs in FAERS, and (3) incidence of all AEs reported in FAERS for each of indications in FAERS ([Figure 7A](#)). In the LP framework, one-class SVM algorithm is used to train the models and the presence/absence of a link of an untrained drug (Drug Candidate X in [Figure 7A](#)), based on its 24-organ transcriptome data, to each TI is determined (see [Transparent Methods](#) for the detailed description, [Figure 7A](#)).

We evaluated the performance of this framework by omitting the data of one drug from all training datasets at a time and repeating this for all 15 drugs to evaluate the performance of predicting potential TIs of each omitted drug ([Figure 7B](#)). The result is shown as a confusion matrix ([Figure 7B](#)), and the full list is available as [Table S10](#). The accuracy scores are high (>0.78) for all 15 drugs ([Figure 7B](#)), indicating that more than 78% of the indications or non-indications predicted are indeed reported or not to be reported as the indications for each drug, respectively. The recall scores are also high (>0.8) for alendronate, aripiprazole, asenapine, clozapine, empagliflozin, lurasidone, olanzapine, evolocumab (Repatha), risedronate, sofosbuvir, and teriparatide ([Figure 7B](#)), indicating that over 80% of the reported indications are predicted for these 11 drugs by the method. The recall score of doxycycline is 0.527 ([Figure 7B](#)), indicating approximately 50% of the reported indications are predicted for this drug by the method. The recall scores are low for acetaminophen (0.141), cisplatin (0), and lenalidomide (0) ([Figure 7B](#)), indicating that this method fails to capture many reported indications for these drugs, as expressed by relatively smaller number of false-negatives (FNs) as compared with that of true-positives (TPs) for these drugs ([Figure 7B](#)). Only acetaminophen (APAP) shows high precision score (1.000), and all the others show low precision scores (<0.35). Both cisplatin and lenalidomide show 0 TP and 0 FN, thus the precision and the F-measure scores were unable to be calculated ([Figure 7B](#)). Such low precision scores for many drugs are mainly due to a large number of false-positives (FPs) as compared with that of TPs ([Figure 7B](#)). The result illustrates a possibly useful application of this LP framework to the prediction of drug TIs (see further in [Discussion](#)).

#### Application to Drug Repositioning

We also evaluated the utility of the multi-organ transcriptome datasets for drug repositioning. In the scheme described in [Figure 7B](#), the FP TIs (yellow highlight in [Figure 7B](#)) (the full list as [Table S10](#))



**Figure 5. AE Prediction by LP Framework**

(A) LP framework to predict AEs. See the text for the description.

(B) The comparison of all AEs prediction by SVM/hMDB-i and the LP approaches. The results are summarized as a confusion matrix. Scores of accuracy (Accuracy), precision (Precision), recall (Recall), and F-measure (F measure) for each drug (Drug) are indicated by bars (light blue) for each method (Methods). The scale of the scores (0.0–1.0) are indicated at the bottom. SVM, SVM/hMDB-i framework; LP, LP framework. The raw data are also available as [Table S6](#).

(C) The list of SAEs that are correctly predicted for alendronate, clozapine, and evolocumab (Repatha) by the LP but not by SVM/hMDB-i framework. The reported outcomes (Reported) are shown as orange and light-blue filled boxes indicating the SAE is reported and not-reported, respectively. The prediction results (Predicted by SVM/LP) are shown as orange and light-blue filled boxes indicating the SAE is predicted to occur and not-to-occur, respectively. The results are also available as [Table S6](#).

could include repositioning targets for the drugs (see further in [Discussion](#)). However, in drug repositioning, drug X does not exist. Hence, we used the same LP framework but did not omit the data of any drugs, i.e., the datasets of the drug of the prediction target is also included in the training datasets ([Figure 7C](#)).

The result is shown as a confusion matrix ([Figure 7C](#)) and the full list is available as [Table S11](#). The result shows the increased number of TPs and decreased number of FNs for all drugs, resulting in the improved recall scores ([Figure 7C](#)). In this scheme, both accuracy and recall scores for all drugs are 0.770–1.000 ([Figure 7C](#)), indicating the approach can capture over 77% of both reported and non-reported indications. The precision scores for all drugs remain relatively low due to the large number of FPs as compared with that of

Organ	Pathway	Entities FDR
Stomach	Detoxification of reactive oxygen species	0.00038
AdrenalG	PPARA activates gene expression	0.00062
AdrenalG	Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	0.00062
AdrenalG	Lipophagy	0.00282
WAT	Chaperone mediated autophagy	0.00282

**Table 1. Enriched Pathways in REACTOME**

The top 5-organ pathways for aripiprazole/male/50s are shown.

TPs (Figure 7C and Table S11). The indications found in FP (yellow highlight in Figure 7C) (the full list as Table S11) could include repositioning targets (see further in Discussion).

As the number of FPs are relatively large, calculating decision function values of each indication found in FP in both schemes could facilitate the ranking of each indication according to their likeliness as repositioning targets. Hence, the decision function values of all indications in both Figures 7B and 7C are shown in Tables S10 and S11. These tables could serve as a useful reference for selecting potential repositioning targets for further investigational evaluations (see further in Discussion).

We also validated our indication prediction with SIDER4.1 version. The result, 0.402–0.998 accuracy and 0.135–1 recall (omitting cisplatin and lenalidomide for the reason described above for the FAERS result in the leave-one-out cross-validation), is comparable with that with FAERS (Figures 7B and 7C, the full and raw data for the confusion matrix are available as Tables S12 and S13). The precision is 0.003–0.069, appears slightly lower than those with FAERS (Figures 7B and 7C), owing to the larger number of FPs. This may be caused by the larger total number of indications reported in FAERS (11,312 indications) as compared with that in SIDER (2,088 indications), contributing to the reduced precision with SIDER reference.

### Side-by-Side Comparisons of Our Methods to Other Multi-Features-Based Prediction Methods

Some other *in silico* approaches predicting SEs/AEs are previously reported (Pauwels et al., 2011; Wang et al., 2016). In particular, Wang et al. reported an SEs/AEs prediction method by combining multi-features (cell morphological features, MACCS chemical fragments, L1000 landmark genes) with adverse drug reactions labels (<http://maayanlab.net/SEP-L1000/index.html#>) (Wang et al., 2016). We compared the SEs/AEs prediction performance of our method with theirs side by side (Figure 8A, Table S14). Of 15 drugs that we analyzed, 9 drugs (alendronate, aripiprazole, acetaminophen, asenapine, cisplatin, clozapine, doxycycline, lenalidomide, olanzapine) are covered by their method. Their method is applicable to only small-molecule drugs, hence SEs/AEs of biologics/antibody drugs (evolocumab/Repatha in our study) or peptide drugs (teriparatide in our study) can be predicted only by our method, illustrating the advantage of our drug modality-independent approach.

We performed side-by-side comparisons of our method with theirs with these nine drugs represented by both methods (Figure 8A, Table S14). With seven drugs (alendronate, acetaminophen, asenapine, cisplatin, clozapine, doxycycline, olanzapine), our method detects far more SEs/AEs labeled in SIDER or the processed FAERS (Figure 8A, Table S14). Even with the other two drugs (aripiprazole, lenalidomide), our method identifies a large number of SEs/AEs (aripiprazole: 93 in SIDER and 37 in the processed FAERS, lenalidomide: 71 in SIDER and 50 in processed FAERS) that are missed by their method (Figure 8A, Table S14).

Prediction of TIs of small-molecule drugs on the basis of their structural information, targets, and interactions is previously reported (Li and Lu, 2012). We performed side-by-side comparisons of our method to theirs in predicting TIs (Figure 8B, Table S15). Of 15 drugs that we studied, 6 drugs (alendronate,

Organ	Pathway	P, Adjusted
Spleen	Glycosphingolipid biosynthesis—ganglio series	0.00174
Eye	Salivary secretion	0.00904
AdrenalG	PPAR signaling pathway	0.01008
Stomach	Glutathione metabolism	0.01032
Stomach	Thyroid hormone synthesis	0.01032
Stomach	Arachidonic acid metabolism	0.01032
ThyroidG	Proteasome	0.01066
ThyroidG	Epstein-Barr virus infection	0.02666

**Table 2. Enriched Pathways in KEGG**

The top 8-organ pathways for aripiprazole/male/50s are shown.

acetaminophen, cisplatin, clozapine, doxycycline, risedronate) are covered by their method. Like the SEs/AEs prediction, our method is modality free; hence, we can predict TIs of biologics/antibody drugs (evolocumab/Repatha) and peptide drugs (teriparatide), which their small-molecule-based prediction method fails. This again illustrates the advantage of our drug modality-independent approach. We performed side-by-side comparisons of our method to theirs with these six drugs represented by both methods (Figure 8B, Table S15) and show that our method detects far more TIs labeled in SIDER or FAERS with all six drugs (Figure 8B, Table S15).

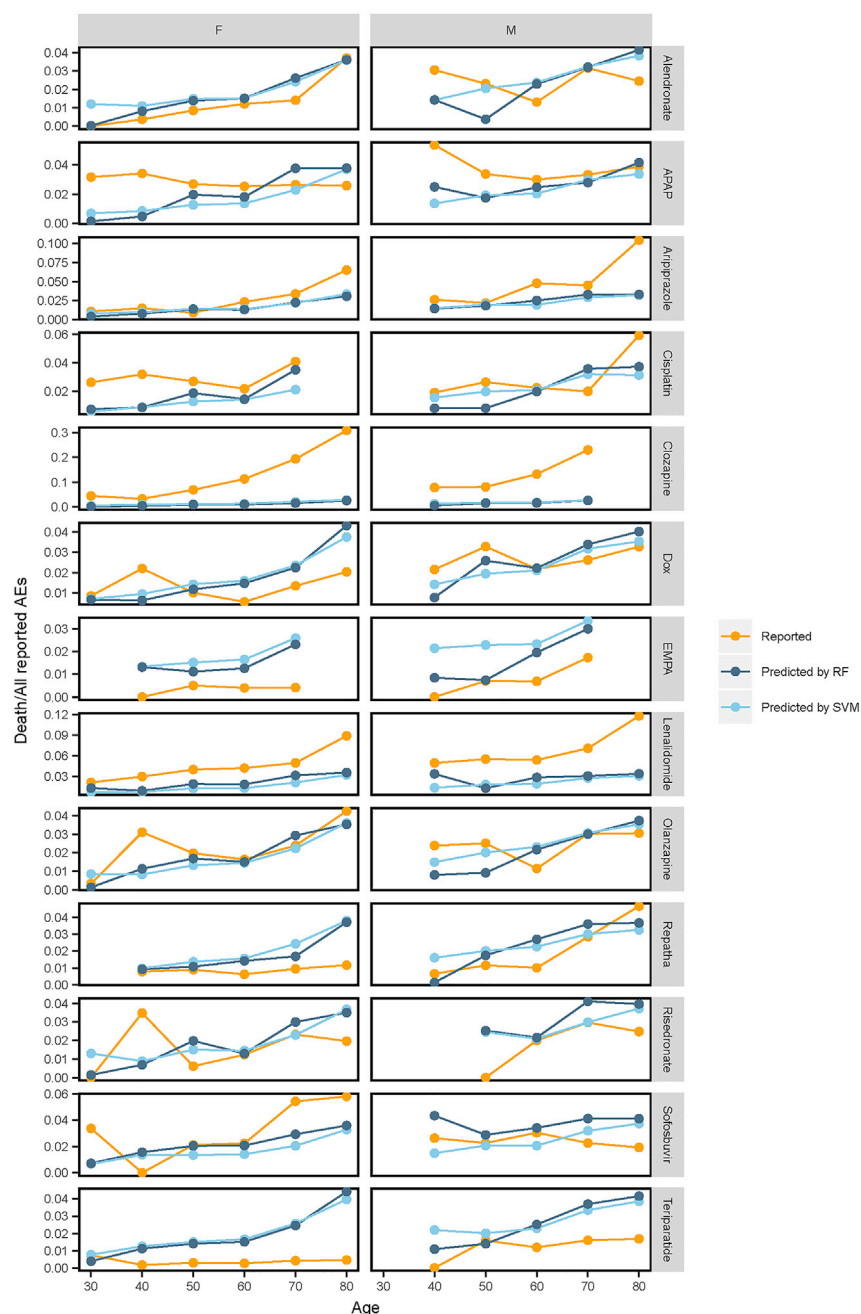
## DISCUSSION

The difficulty of predicting human clinical outcomes during pre-clinical studies is the major hindrance in drug development. This is mainly due to the animal models versus human difference in the outcomes. To overcome this problem, numerous experimental and computational approaches are currently used (Elitt et al., 2018; Ko and Gelb, 2014; Meseguer-Ripolles et al., 2018; Oleaga et al., 2016; Rezaei Kolahchi et al., 2016; Shah et al., 2019; Tateno et al., 2004; Vamathevan et al., 2019). Some approaches use transcriptome and/or other omics data to predict clinical outcomes (Ho et al., 2016). However, they use only cells and limited types of organs/tissues and none exploit the whole-body level biological features such as the body-wide multi-organ transcriptome data for clinical outcome prediction.

In the current study, we exploit multi-organ transcriptome datasets in mice to predict human clinical outcomes (Figure 1). We “humanize” the mouse data by training machine learning models with multi-organ transcriptome data derived from the mouse and human clinical datasets (Figure 1). Our evaluation illustrates advantages and effectiveness of this approach in several applications to the outcome prediction of both AEs and TIs (Figures 2, 3, 4, 5, 6, and 7).

Several other *in silico* approaches predicting SEs/AEs and TIs have been previously reported (Li and Lu, 2012; Pauwels et al., 2011; Wang et al., 2016). The side-by-side comparisons of our methods with theirs shows that our method detects far more SEs/AEs and TIs than theirs with most of the 15 drugs evaluated here (Figure 8). Even with the drugs that both methods are comparable (aripiprazole and lenalidomide in the SE/AE prediction), our method detected the outcomes that are missed by theirs (Figure 8A). These results illustrate a usefulness of exploiting multi-organ transcriptome datasets in predicting drug SEs/AEs and TIs.

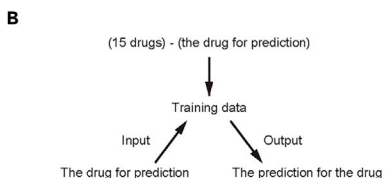
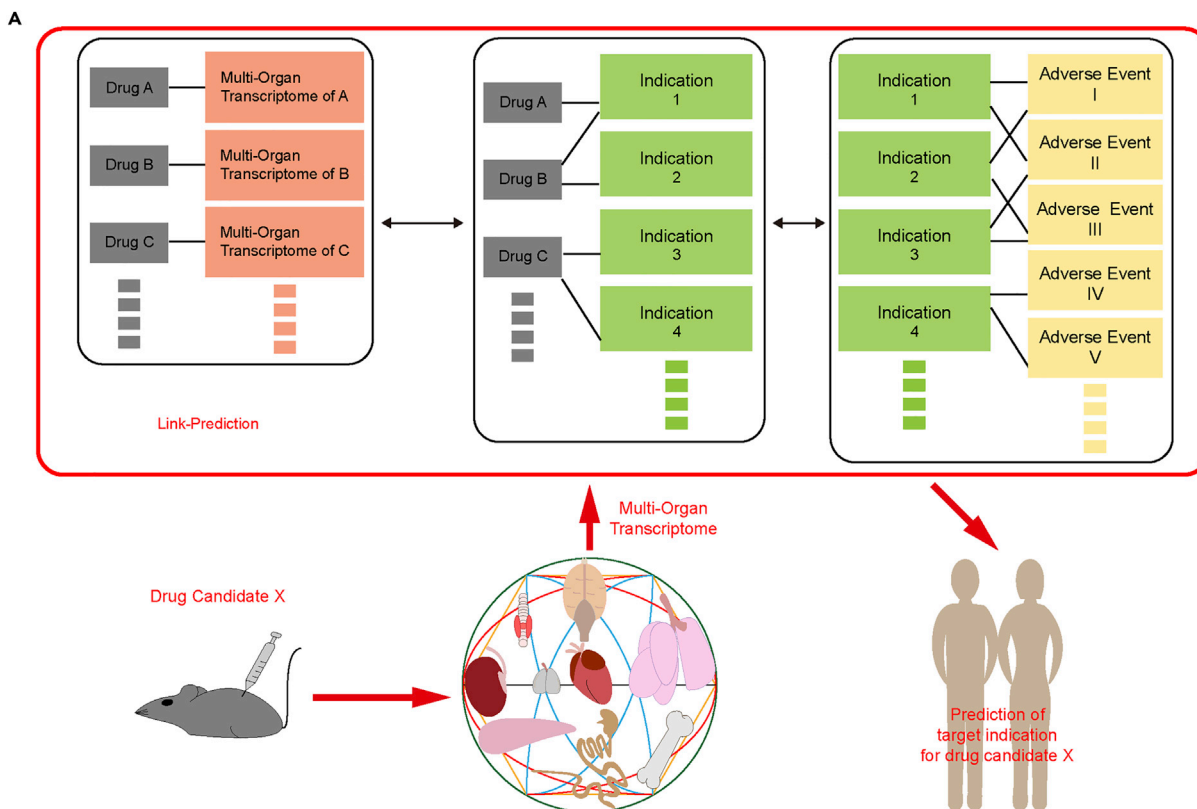
Potential associations of SEs with diseases are computed on the basis of drug fingerprints/targets and their known SEs/potential disease indications and reported as an integrated table (<https://astro.temple.edu/~tua87106/druganalysis.html>) (Zhang et al., 2013). Although, by using this table, they suggest that drug-repositioning candidates could be identified, they fail to explicitly tabulate specific drug-repositioning targets or validate them. In this paper, we exploit potential SE/AE-TI association and combined it with our body-wide multi-organ transcriptome approach in the LP framework to compute drug TIs and explicitly tabulate them (Figure 7). We, furthermore, evaluate the effectiveness and the efficiency of our method by cross-validation (Figure 7B).



**Figure 6. Prediction of Quantitative Dynamics of Death Events Over Age**

The results of predicting quantitative dynamics of death events at each age group for each drug are summarized. The results by SVM regressor (SVM) and RF regressor (RF) are shown. The age groups are 30: 30–39 years old, 40: 40–49 years old, 50: 50–59 years old, 60: 60–69 years old, 70: 70–79 years old, 80: 80–89 years old. The results are also available as [Table S9](#).

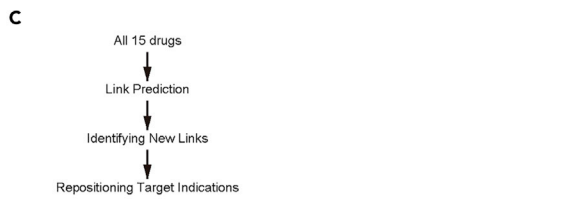
There are two major advantages of our approach over the other *in silico* prediction methods. The first is its drug-modality independence. This is illustrated by the effective predictions for small-molecule (alendronate, acetaminophen, aripiprazole, asenapine, cisplatin, clozapine, doxycycline, empagliflozin, lenalidomide, lurasidone, olanzapine, risedronate, sofosbuvir), antibody (evolocumab/Repatha), and peptide (teriparatide) drugs. In principle, our approach is expected to be as effective for other modalities such as nucleic acid, gene, and cell therapeutics and also for their mixtures, as the approach does not use



FAERS									
Drug	TP	TN	FP	FN	Accuracy	Precision	Recall	F measure	
Alendronate	392	9240	1638	33					
APAP	202	9875	0	1231					
Aniprazole	262	9495	1492	63					
Asenapine	8	9211	2093	0					
Cisplatin	0	10993	0	352					
Clozapine	152	8952	2188	20					
Dox	175	10633	345	157					
EMPA	68	9146	2092	6					
Lenalidomide	0	10903	0	409					
Lurasidone	6	8918	2358	0					
Olanzapine	260	9395	1599	57					
Repatha	54	9362	1891	5					
Risedronate	185	9281	1835	11					
Sofosbuvir	107	9838	1556	11					
Teriparadide	141	9162	2002	7					

SIDER									
Drug	TP	TN	FP	FN	Accuracy	Precision	Recall	F measure	
Alendronate	10	1068	1009	1					
APAP	10	1932	134	12					
Aniprazole	13	1157	916	3					
Asenapine	7	1051	1030	0					
Cisplatin	0	2059	0	29					
Clozapine	17	884	1095	2					
Dox	5	1717	334	32					
EMPA	3	1028	1057	0					
Lenalidomide	0	2094	0	4					
Lurasidone	6	972	1110	0					
Olanzapine	15	1126	947	0					
Risedronate	5	1077	1005	1					
Sofosbuvir	3	1182	903	0					
Teriparadide	4	1036	1049	0					



FAERS									
Drug	TP	TN	FP	FN	Accuracy	Precision	Recall	F measure	
Alendronate	407	9010	1871	18					
APAP	1298	9181	859	135					
Aniprazole	295	9045	1941	30					
Asenapine	8	9212	2091	0					
Cisplatin	271	9470	1430	81					
Clozapine	156	8829	2231	16					
Dox	288	9311	1690	48					
EMPA	70	9091	2147	4					
Lenalidomide	362	8810	1093	47					
Lurasidone	6	8919	2387	0					
Olanzapine	283	8991	2004	34					
Repatha	55	9306	1947	4					
Risedronate	188	9196	1918	8					
Sofosbuvir	109	9447	1747	9					
Teriparadide	143	9033	2081	5					

SIDER									
Drug	TP	TN	FP	FN	Accuracy	Precision	Recall	F measure	
Alendronate	10	992	1085	1					
APAP	21	818	1248	1					
Aniprazole	15	1003	1069	1					
Asenapine	7	1051	1030	0					
Cisplatin	27	1179	891	2					
Clozapine	18	947	1122	1					
Dox	12	1081	990	25					
EMPA	3	1016	1069	0					
Lenalidomide	4	1273	911	0					
Lurasidone	6	972	1110	0					
Olanzapine	15	992	1081	0					
Risedronate	5	1053	1029	1					
Sofosbuvir	3	1134	951	0					
Teriparadide	4	1017	1067	0					

### Figure 7. Prediction of Therapeutic Indications

(A) LP framework to predict therapeutic indications. See text for the detailed description.

(B) Cross-validation with FAERS and SIDER references. The hold-out scheme is shown at the top. The results are shown as a confusion matrix. The scores are indicated as light blue bars. Scores of precision and F-measure for cisplatin and lenalidomide are incalculable (black filled cells) as both TP and FP scores are 0. The results are also available as [Tables S10](#) (FAERS) and [S12](#) (SIDER).

(C) Repositioning prediction. The scheme is shown at the top. The results with FAERS and SIDER references are shown as a confusion matrix. The scores are indicated as light blue bars. TP, true positive; TN, true negative; FP, false positive; FN, false negative. The scale of the scores (0.0–1.0) are indicated at the bottom. FP as potential repositioning targets are yellow highlighted. The results are also available as [Tables S11](#) (FAERS) and [S13](#) (SIDER).

any structural features of the drugs for the predictions. The other advantage is the fact that the approach requires no prior knowledge about the molecular or cellular targets of the drugs or the mechanisms of their actions. Hence, the only requirement of our approach is that the test drug can be administered to the mouse.

In our approach, the effective prediction is critically dependent on identifying common transcriptome features between the drugs in the models and the target/test drug (i.e., drug X). The 15-drug group in the current study consists of those with diverse AE and indication patterns. Despite such complex clinical outcome patterns, the datasets from such a small number of drugs are surprisingly sufficient to effectively predict the outcomes of many AEs and TIs for most of the drugs in the cross-validation studies ([Figures 2, 3, 4, 5, 6, 7, and 8](#)). This may be due to the presence of many possibly useful features in the multi-organ transcriptome data for the predictions that could not be identified in the transcriptome or other omics data of the cells and the limited types of organs/tissues.

Although the use of the 24-organ transcriptome data (i.e., hMDB-i approach) is generally effective in predicting AEs, we find the differential performance by using all 24-organ data together or using them as individual-organ datasets ([Figure 3](#)). We also observed differential performance depending on which of the 24 individual-organ datasets is used ([Figure 3](#)). To address these issues, we suggest that majority decision approach could be beneficial and illustrate such examples ([Figure 4C](#)). Furthermore, the use of hMDB-i approach without and within the LP framework appears to influence the prediction consequence ([Figures 5B and 5C](#)). The accuracy and precision scores are better without the LP framework, but the recall scores are better with the LP framework. Hence, it may be beneficial to select the framework (i.e., with/without the LP framework) according to the objective of the prediction (e.g., to be more confident about the predicted AEs versus to capture all possible AEs).

Drug repositioning is a strategy to develop the existing drug for an additional therapeutic indication(s) ([Pushpakom et al., 2019](#)). The advantage is that the properties (such as pharmacokinetics) of such drugs are already determined. Furthermore, their safety in human is secured; hence, phase I trials could often be bypassed ([Pushpakom et al., 2019](#)). Such advantage of drug repositioning results in the overall cost- and time-saving in the development, as compared with new drug development ([Pushpakom et al., 2019](#)). We examined the potential utility of our approach to predict TIs for drug repositioning ([Figure 7](#)). The potential therapeutic targets for the repositioning of each drug are those that are labeled as “FP” in [Figures 7B and 7C](#) and [Tables S10 and S11](#). The examination of the list identified many common FP indications shared among the drugs of the same or related mechanisms of actions ([Tables S10 and S11](#)). For alendronate and risedronate, both bisphosphonate medications for treating osteoporosis, Addison’s disease and amyloidosis are among those predicated as FP targets with high decision function values. For clozapine and aripiprazole, both atypical anti-psychotic medications antagonizing related neurotransmitter receptors, coagulopathy and ulcer are examples of such FP targets. Although their investigational validations must await the future clinical trials, such targets shared among the related medications could be among the leading candidates for the repositioning.

In the study of predicting TIs, we show that both cisplatin and lenalidomide failed to identify true positive targets ([Figure 7B](#)). On the other hand, including these two drugs in the training data for the same prediction framework resulted in the improved identification of the true-positive targets ([Figure 7C](#)). These results may suggest that the gene expression patterns of these two drugs include distinguishing features that are unique to each of these two drugs; hence, such features could not be exploited by the scheme shown in [Figure 7B](#) (i.e., the data of these drugs are omitted from the training data for their predictions). This point is supported by the fact that the inclusion of the transcriptome data of these two drugs in the training datasets lead to the effective outcome predictions for the other drugs ([Figure 7B](#)). Furthermore, such unique





**Figure 8. Side-by-Side Comparisons of Our Methods with Other Multi-Features-Based Prediction Methods**

The comparisons are shown as Venn diagrams for SE/AE (A) and TI (B) predictions. The description of the diagrams is shown as “Legend” diagram located at the right bottom in each panel. The number of the SEs/AEs and TIs predicted only by our method (A: hMDB, B: hMDB/LP), only by the existing methods (A: multiple-features/L1000, B: fingerprints/targets/Interactions), and by both methods are indicated accordingly in the diagrams. Our methods (A: hMDB, B: hMDB/LP) are compared with multiple-features/L1000 method (A) and fingerprints/targets/Interactions method (B), respectively. The predictions for the drugs that are not represented or cannot be predicted by the other methods are indicated as 0. Of SEs/AEs and TIs predicted only by our methods, those labeled in SIDER, FAERS, and the processed FAERS databases are also indicated. SEs for evolocumab/Repatha do not exist in SIDER; hence, it is indicated as 0. The table forms of the Venn diagrams and their full searchable raw data are available as [Tables S14 \(A\)](#) and [S15 \(B\)](#).

features can be effectively exploited when included in the training data, resulting in the successful identification of the true-positives (Figure 7C).

In drug development and clinical trials, it would be important and beneficial to understand the mechanisms of the outcomes. Such mechanistic understanding could serve as a basis for further improvement on the drug designs and also designing the pre-clinical and clinical studies. Furthermore, availability of the mechanisms facilitates the regulatory approval processes and also improves ethical responsibilities providing an explanation to the study subjects and patients. Using RF algorithm allows us to identify

specific molecular/cellular pathways that contributed to the AE predictions of the drug, and the examples are shown in [Tables 1](#) and [2](#) and the full lists are available in [Tables S7](#) and [S8](#). Although such pathways are those modulated by the drugs in mice and may not necessarily be the same in human, many are conserved across animal species. Hence, they could serve as at least, if not at all, as clues for deducing possible mechanisms underlying the AEs and/or generating working hypothesis for further experimental and clinical studies for the validation.

The current study illustrates beneficial utilities of applying body-wide multi-organ transcriptome data obtained from mice to predict human clinical outcomes. The further inclusion of other drugs' multi-organ transcriptome and their clinical outcome data in hMDB-i and other relevant data, such as individuals' genotypes and lifestyle information, as training datasets may improve the current performance where some limitations are observed with some drugs and/or some outcomes. With such refinements in the training datasets, further improvements in the individualized predictions is expected in the future, contributing to the realization of the ultimate virtual precision medicine.

### Limitations of the Study

In the prediction study of therapeutic indications, we used all therapeutic indications associated with each drug in FAERS as training data. Thus, no distinction can be made between the effective and ineffective treatments. Furthermore, all indications for which each drug was used are included as the training data for that particular drug, i.e., the indications for which other drug(s) was/were used for the purpose of treatments and drug X as simply an accompanying drug were included as "reported" therapeutic indications for drug X in the training data. For example, acetaminophen is used to attenuate pain associated with osteoporosis; hence, both pain and osteoporosis were included as the indications for this drug. Furthermore, we also used all AE labels associated with each drug and/or indication. Hence, no distinctions were made between those of causative relationships and of simple association. Currently, using FAERS, it is difficult to make such distinctions for all drugs in a systematic manner. Some distinction may be possible by using other public database such as [clinicaltrials.gov](https://clinicaltrials.gov), but the number of reports is limited and not sufficient for our machine-learning approach. Despite such compromise, the fact that our approach predicts many human clinical outcomes according to the quality of the input training data indicates that, with further refinement in the input training datasets, it is reasonable to expect that our method generates more sophisticated predictions. In fact, we validated our predictions with a drug-label based reference, SIDER. The results show that both SE/AE and TI predictions with SIDER are comparable with those with FAERS ([Figures 2D, 7B, and 7C](#), [Tables S3, S12, and S13](#)). These results further support the effectiveness and usefulness of our approach.

Owing to the experimental cost necessary to perform the 24-organ transcriptome analyses, the current study is limited to the evaluation of 15 drugs. There are 3,732 approved drugs (<https://www.drugbank.ca/stats>), and it would be extremely expensive to perform the 24-organ transcriptome analyses with most, if not all, of these drugs. With the current cost for the transcriptome analyses, we estimate such cost would be US\$20–30 million. This is expensive for one laboratory scale but possibly realistic as a consortium effort. Considering the expensiveness of drug development, this might be a worthy investment. In fact, we may not need all drug data, as we show even the 15-drug datasets were sufficient for many of the outcome predictions, despite their diverse modalities and TIs ([Figures 2, 3, 4, 5, 6, 7, and 8](#)).

### METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

### DATA AND CODE AVAILABILITY

The accession number for the RNA-seq data reported in this paper is GEO: GSE142068. Hyperparameters are available at <https://www.hmdb.karydo-tx.com/>. The code for our algorithm is available for non-profit use with Material Transfer Agreement.

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2019.100791>.

## ACKNOWLEDGMENTS

We thank S. Keiko, R. Takahashi, and E. Kojima for administrative assistance. We are also grateful to the members of Karydo TherapeutiX, Inc. and Sato lab at ATR for advice and discussion throughout the course of this work. This work was supported by funds from Public Interest Incorporated Foundation KYOTO Industrial Support Organization 21 (T.N.S.) and JST ERATO Grant Number JPMJER1303 (T.N.S.).

## AUTHOR CONTRIBUTIONS

T.N.S. conceived the idea of the project, designed the overall experiments and *in silico* analyses, and supervised the overall research project. S.K. designed and performed the *in silico* analyses. F.S., S.E., G.M.D.A., and Y.M. contributed to designing the experiments and performed the experiments. T.N.S., S.K., F.S., and S.E. wrote the manuscript.

## DECLARATION OF INTERESTS

T.N.S. is the inventor of the patents filed on hMDB-i and is the scientific founder of and holds shares of Karydo TherapeutiX, Inc., which owns the right to this patent. S.K., S.F., S.E., G.M.D.A., and Y.M. are employees of Karydo TherapeutiX, Inc. Correspondence and requests for materials should be addressed to T.N.S. (island1005@gmail.com).

Received: August 31, 2019

Revised: October 17, 2019

Accepted: December 16, 2019

Published: January 9, 2020

## REFERENCES

- Berger, S.I., and Iyengar, R. (2011). Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 129–135.
- Droujinine, I.A., and Perrimon, N. (2013). Defining the interorgan communication network: systemic coordination of organismal cellular processes under homeostasis and localized stress. *Front. Cell. Infect. Microbiol.* 3, 82.
- Droujinine, I.A., and Perrimon, N. (2016). Interorgan communication pathways in physiology: focus on *Drosophila*. *Annu. Rev. Genet.* 50, 539–570.
- Elitt, M.S., Barbar, L., and Tesar, P.J. (2018). Drug screening for human genetic diseases using iPSC models. *Hum. Mol. Genet.* 27, R89–R98.
- Harpaz, R., DuMouchel, W., LePendou, P., Bauer-Mehren, A., Ryan, P., and Shah, N.H. (2013). Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin. Pharmacol. Ther.* 93, 539–546.
- Ho, T.B., Le, L., Thai, D.T., and Taewijit, S. (2016). Data-driven approach to detect and predict adverse drug reactions. *Curr. Pharm. Des.* 22, 3498–3526.
- Ko, H.C., and Gelb, B.D. (2014). Concise review: drug discovery in the age of the induced pluripotent stem cell. *Stem Cells Transl. Med.* 3, 500–509.
- Kozawa, S., Ueda, R., Urayama, K., Sagawa, F., Endo, S., Shiizaki, K., Kurosu, H., Maria de Almeida, G., Hasan, S.M., Nakazato, K., et al. (2018). The body-wide transcriptome landscape of disease models. *iScience* 2, 238–268.
- Li, J., and Lu, Z. (2012). A new method for computational drug repositioning using drug pairwise similarity. *Proceedings (IEEE Int. Conf. Bioinformatics Biomed.)* 2012, 1–4.
- Meseguer-Ripolles, J., Khetani, S.R., Blanco, J.G., Iredale, M., and Hay, D.C. (2018). Correction to: pluripotent stem cell-derived human tissue: platforms to evaluate drug metabolism and safety. *AAPS J.* 20, 30.
- Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C., and Greyson, D. (2011). The cost of drug development: a systematic review. *Health Policy* 100, 4–17.
- Oleaga, C., Bernabini, C., Smith, A.S., Srinivasan, B., Jackson, M., McLamb, W., Platt, V., Bridges, R., Cai, Y., Santhanam, N., et al. (2016). Multi-Organ toxicity demonstration in a functional human in vitro system composed of four organs. *Sci. Rep.* 6, 20030.
- Pauwels, E., Stoven, V., and Yamanishi, Y. (2011). Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinformatics* 12, 169.
- Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., et al. (2019). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 18, 41–58.
- Rezaei Kolahchi, A., Khadem Mohtaram, N., Pezeshgi Modarres, H., Mohammadi, M.H., Geraili, A., Jafari, P., Akbari, M., and Sanati-Nezhad, A. (2016). Microfluidic-based multi-organ platforms for drug discovery. *Micromachines (Basel)* 7, 162.
- Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., Ringel, M., and Schork, N. (2019). Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit. Med.* 2, 69.
- Sonawane, K.B., Cheng, N., and Hansen, R.A. (2018). Serious adverse drug events reported to the FDA: analysis of the FDA adverse event reporting system 2006–2014 database. *J. Manag. Care Spec. Pharm.* 24, 682–690.
- Tateno, C., Yoshizane, Y., Saito, N., Kataoka, M., Utoh, R., Yamasaki, C., Tachibana, A., Soeno, Y., Asahina, K., Hino, H., et al. (2004). Near completely humanized liver in mice shows human-type metabolic responses to drugs. *Am. J. Pathol.* 165, 901–912.
- Tatonetti, N.P., Ye, P.P., Daneshjou, R., and Altman, R.B. (2012). Data-driven prediction of drug effects and interactions. *Sci. Transl. Med.* 4, 125ra131.
- Vamathevan, J., Clark, D., Czdrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., et al. (2019). Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 18, 463–477.
- Wang, Z., Clark, N.R., and Ma'ayan, A. (2016). Drug-induced adverse events prediction with the LINCS L1000 data. *Bioinformatics* 32, 2338–2345.
- Wong, C.H., Siah, K.W., and Lo, A.W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics* 20, 273–286.
- Zhang, P., Wang, F., Hu, J., and Sorrentino, R. (2013). Exploring the relationship between drug side-effects and therapeutic indications. *AMIA Annu. Symp. Proc.* 2013, 1568–1577.

**iScience, Volume 23**

## **Supplemental Information**

### **Predicting Human Clinical Outcomes**

#### **Using Mouse Multi-Organ Transcriptome**

**Satoshi Kozawa, Fumihiko Sagawa, Satsuki Endo, Glicia Maria De Almeida, Yuto Mitsuishi, and Thomas N. Sato**

## Supplementary Figure

Figure S1



**Figure S1. The patterns of differentially expressed genes in each organ of the mice treated with each drug, Related to Figure 1.** Y-axis: (the number of differentially expressed genes for the drug) / (the number of differentially expressed genes for all drugs) for each organ. The differentially expressed genes are those with adjusted  $p < 0.0001$  by DESeq2 analysis. The DESeq2 data are available as Supplementary Table 1. The raw RNAseq sequence data are available at [https://www.hmdb.karyodt.com/hMDB\\_freeDL/](https://www.hmdb.karyodt.com/hMDB_freeDL/).

## **Supplementary Table Legends**

### **Table S1. The DESeq2 data for Figure S1, Related to Figure 1.**

DE gene: the number of differentially expressed genes.

### **Table S2. The raw data of Figure 2B**

The binary labelling system is applied to each AE: 1 when the AE is reported/predicted to occur and 0 when the AE is not-reported/predicted not-to-occur. The empty cells indicate that none of the AEs are reported for this drug in this sex/age-group.

### **Table S3. The raw data of Figure 2D**

The binary labelling system is applied to each side effect (SE): 1 when the SE is reported/predicted to occur and 0 when the SE is not-reported/predicted not-to-occur.

### **Table S4. The raw data of Figure 3**

The binary labelling system is applied to each AE: 1 when the AE is reported/predicted to occur and 0 when the AE is not-reported/predicted not-to-occur.

### **Table S5. The raw data of Figures 4A&B.**

The binary labelling system is applied to each AE: 1 when the AE is reported/predicted to occur and 0 when the AE is not-reported/predicted not-to-occur.

### **Table S6. The raw data of Figures 5B&C**

The binary labelling system is applied to each AE: 1 when the AE is reported/predicted to occur and 0 when the AE is not-reported/predicted not-to-occur.

### **Table S7. The raw data of Table 1**

All pathways of  $FDR < 0.05$  are included.

**Table S8. The raw data of Table 2**

All pathways of adjusted  $p < 0.05$  are included.

**Table S9. The raw data of Figure 6**

The actual number of reports in FAERS (Reported #) is shown for each drug, AE, sex, age-group in the table.

**Table S10. The raw data of Figure 7B (vs. FAERS)**

Decision function value for each indication is indicated.

**Table S11. The raw data of Figure 7C (vs. FAERS)**

Decision function value for each indication is indicated.

**Table S12. The raw data of Figure 7B (vs. SIDER)**

1 when the TI is reported/predicted to occur and 0 when the TI is not-reported/predicted not-to-occur.

**Table S13. The raw data of Figure 7C (vs. SIDER)**

1 when the TI is reported/predicted to occur and 0 when the TI is not-reported/predicted not-to-occur.

**Table S14. The raw data of Figure 8A**

SEs predicted only by our method (highlighted by orange), only by the multiple features/L1000 method (highlighted by sky blue) and by both methods (highlighted by light purple) are indicated.

**Table S15. The raw data of Figure 8B**

TIs predicted only by our method (highlighted by orange), only by the fingerprints/targets/interactions method (highlighted by sky blue) and by both methods (highlighted by light purple) are indicated.

## Transparent Methods

### *Mice and 24-organ transcriptome analyses*

The all mouse studies and the transcriptome analyses are performed as previously reported (Kozawa et al., 2018). The 24 organs are adrenal glands, aorta, bone marrow cells (BM), brain, colon, eyes, heart, ileum, jejunum, left kidney, liver, lung, pancreas, parotid glands, pituitary glands, skeletal muscle, skin, skull, spleen, stomach, left testis, thymus, thyroid glands, and gonadal white adipose tissue (WAT). All animal protocols were approved by the Animal Care and Use Committee of Karydo TherapeutiX, Inc. (Approved Number: AN20170001KTX, AN20170010KTX, AN20170014KTX, AN20180012KTX and AN20190006KTX). All animals were housed in a temperature-controlled room at around 25 °C with 12 h light/dark cycle and allowed free access to water and normal foods (CE-2, CLEA Japan, Inc., Tokyo, Japan). For the transcriptome analyses, we used QuantSeq 3'mRNA-Seq Library Prep Kit for Illumina (FWD) (cat#015.384, LEXOGEN) and Illumina NextSeq 500 (75bp single-read, ca. 400 million reads/run, NextSeq 500/550 High Output Kit v2.5, cat#20024906).

### *Drug administration and harvesting organs*

**Alendronate:** Male C57BL/6NCrSlc mice at 11 weeks old received subcutaneous injection of alendronate sodium salt trihydrate (Wako) in PBS (Nacalai Tesque) at 1.0 mg/kg dosage every 3 to 4 days for 8 days. The drug was prepared fresh everytime. Organs were harvested in the afternoon on the 8th day after the drug administration.

**Acetaminophen:** Male C57BL/6NCrSlc mice at 10 weeks old were fasted for 12 hours with free access to water during the whole time. Immediately after the fasting period, mice received single intraperitoneal injection of acetaminophen (Wako) in saline (Otsuka Pharmaceutical) at 300 mg/kg dosage. After the injection, mice were allowed the free access to the normal foods. Organs were harvested 2 hours after the drug administration before noon.

**Aripiprazole:** Male C57BL/6NJcl mice at 11 weeks old received single intraperitoneal injection of aripiprazole (Sigma-Aldrich) in 0.5%(w/v) carboxymethylcellulose 400 solution (Wako) at 0.3 mg/kg dosage. Organs were harvested 2 hours after the drug administration in the afternoon.



**Asenapine:** Male C57BL/6NCrSlc mice at 11 weeks old received single subcutaneous injection of asenapine maleate (Chemscene) in saline at 0.3 mg/kg dosage. Organs were harvested 2 hours after the drug administration in the afternoon.

**Cisplatin:** Male C57BL/6NJcl mice at 11 weeks old received single intraperitoneal injection of cisplatin (Bristol-Myers Squibb) at 20 mg/kg dosage. Organs were harvested in the afternoon on the 3<sup>rd</sup> day after the drug administration.

**Clozapine:** Male C57BL/6NJcl mice at 11 weeks old received single subcutaneous injection of clozapine (Sigma-Aldrich) at 0.3 mg/kg dosage. Clozapine was first dissolved in acetic acid before diluted in saline and the pH was adjusted to 6 with 1M NaOH. Organs were harvested 2 hours after the drug administration in the afternoon.

**Doxycycline:** Male C57BL/6NCrSlc mice at 9 weeks old were allowed to a free access to the doxycycline solution containing 2 mg/mL doxycycline hydrochloride n-hydrate (Wako) and 5% sucrose (Nacalai Tesque) in RO water or a vehicle solution containing 5% sucrose in RO water for 2 weeks. The new bottle was prepared weekly for 2 weeks. Organs were harvested in the afternoon on the 13<sup>th</sup> day after the drug administration.

**Empagliflozin:** Male C57BL/6NCrSlc mice at 10 weeks old received oral (gavage) administration of empagliflozin (Toronto research chemicals) in 0.5% carboxymethylcellulose at 10 mg/kg everyday for 2 weeks. The drug was prepared fresh at every administration. Organs were harvested 2 weeks after the drug administration in the afternoon.

**Lenalidomide:** Male C57BL/6NCrSlc mice at 8 weeks old received oral (gavage) administration of lenalidomide (Wako) in 0.5% carboxymethylcellulose and 0.25% Tween-80 (Nacalai Tesque) at 50 mg/kg or a vehicle solution (0.5% carboxymethylcellulose and 0.25% Tween-80) everyday for 69 days. The drug was prepared fresh at every administration. Organs were harvested 69 days after the drug administration in the afternoon.

**Lurasidone:** Male C57BL/6NCrSlc mice at 11 weeks old received single oral (gavage) administration of lurasidone hydrochloride (Medchemexpress) in 0.5% carboxymethylcellulose at 0.3 mg/kg. Organs were harvested 2 hours after the drug administration in the afternoon.

**Olanzapine:** Male C57BL/6NCrSlc mice at 11 weeks old received single oral (gavage) administration of olanzapine (Tokyo Chemical Industry) in 0.5% carboxymethylcellulose at 0.3 mg/kg. Organs were harvested 2 hours after the drug administration in the afternoon.

**Evolocumab (Repatha®):** Male C57BL/6NJcl mice at 11 weeks old received subcutaneous injection of Repatha® (Astellas) in saline at 10 mg/kg every 10 days for 4 weeks. Organs were harvested 4 weeks after the drug administration in the afternoon.

**Risedronate:** Male C57BL/6NCrSlc mice at 11 weeks old received oral (gavage) administration of risedronate sodium salt (Cayman Chemical Company) in PBS at 10 mg/kg dosage every other day for 8 days. The drug was prepared fresh at every administration. Organs were harvested 8 days after the drug administration in the afternoon.

**Sofosbuvir:** Male C57BL/6NCrSlc mice at 7 weeks old received intraperitoneal injection of sofosbuvir (LKT) at 20 mg/kg dosage everyday for 10 days. Sofosbuvir was first diluted in DMSO (Nacalai Tesque) and then further diluted x100 in PBS prior to the administration (the final concentration is 1.0% DMSO/PBS). Organs were harvested 10 days after the drug administration in the afternoon.

**Teriparatide:** Male C57BL/6NCrSlc mice at 10 weeks old received subcutaneous injection of parathyroid hormone fragment 1-34 human (teriparatide) (Sigma-Aldrich) in saline at 40 µg/kg dosage or a vehicle solution (saline) everyday for 4 weeks. Organs were harvested 4 weeks after the drug administration in the afternoon.

**Wild type (WT):** Organs of male C57BL/6NJcl mice at 11 weeks old were harvested in the afternoon.

#### *Differential gene expression data for the prediction*

Differential gene expression data in each organ obtained from the mice to which each drug was administered were used as the features of each drug in the machine learning frameworks. The RNA-seq data processing (mapping and counting transcripts) were conducted as previously described (Kozawa et al., 2018). The mapping on mouse genome was conducted on mm10 with TopHat2. The differentially expressed genes in each organ of the drug-administered vs. vehicle-control (doxycycline and lenalidomide) or WT (all the other drugs) mice were identified by DESeq2 (1.22.1). Biological

duplicates (i.e., two mice for drug-administered and for controls) were analyzed for all organs and the drug-administered models. We selected the differentially expressed genes of which adjusted p-values  $< 1e-4$  for each organ as the features of each drug. The combination of the organ and the  $\log_2$ fold values of all the selected genes combined from all organs (the 24-organ framework) or individual-organs (the individual-organ framework) are used as the input data for each drug. For SVM algorithm, we applied principle component analysis (PCA) to the organ-transcriptome data to reduce data dimensionality.

#### *Prediction of AEs by SVM and RF*

AEs were predicted by two machine learning algorithms, support vector machine (SVM) and random forest (RF). To predict an AE for a drug, we trained machine learning models by the transcriptome data for 14 drugs (i.e., the 15 drugs - the prediction-target drug) and the AE data reported for the 14 drugs in FAERS (Figure 2A). To obtain an AE prediction for the target drug, we input the organ-transcriptome data of the target drug into the trained machine learning models (Figure 2A). This training/prediction procedure was repeated for all drugs and all AEs. The transcriptome data are differential gene expression data described in “Differential gene expression data for the prediction” section. To obtain AE data, we downloaded raw AE reports from FAERS (<https://www.fda.gov/drugs/surveillance/fda-adverse-event-reporting-system-faers>) (2014 3<sup>rd</sup> quarter to 2017 4<sup>th</sup> quarter) and extracted the reports associated with the 15 drugs for each sex and age-groups. A total of 5,519 AEs is reported for one or more of the 15 drugs (Table S2). For each sex and age-group, the AEs reported in one or more the 15 drugs are used as all AEs for that particular sex and age-group (Table S2). The number of all AEs for each group are: 727 (M:10s), 1,147 (M:20s), 1,339 (M:30s), 1,747 (M:40s), 2,441 (M:50s), 2,618 (M:60s), 2,169 (M:70s), 1,189 (M:80s), 263 (M:90s), 784 (F:10s), 1,289 (F:20s), 1,687 (F:30s), 2,149 (F:40s), 2,828 (F:50s), 3,188 (F:60s), 2,807 (F:70s), 1,863 (F:80s), 573 (F:90s) (Supplementary Table 2). The binary labelling system was applied to each AE: 1 when the AE is reported/predicted to occur and 0 when the AE is not-reported/predicted not-to-occur. AEs that are 1 or 0 for all drugs in a sex/age-group were omitted from the training or prediction as their outcomes are naturally predictable. In addition, AEs that are 1 or 0 for all drugs except for the prediction-target drug in a sex/age-group were also omitted as their outcomes are, in principle, unpredictable. SVM was performed on R library ‘e1071’

with default parameters. RF was performed on Python package scikitlearn. RF parameter ‘n\_estimator’ was set to 10, and the other parameters (‘max\_feature’, ‘max\_depth’ and ‘min\_samples\_leaf’) were decided by leave-one-out cross-validation for each AE, sex- and sex-group.

#### *Majority decision approach for AE prediction*

This approach is applied to predicting the outcomes of death event for alendronate/female/20s and lenalidomide/female/20s where the prediction results by SVM with all 24-organs failed. For the SVM method, there are 25 “votes” – SVM with the 24-organs (1 vote) and SVM with each of the 24 individual-organs (24 votes). For the RF method, there are also 25 votes – RF with the 24-organs (1 vote) and RF with each of the 24 individual-organs (24 votes). The number of the prediction votes for “to-occur” and “not-to-occur” is counted and summed. The prediction that won the majority votes become the prediction.

#### *Prediction of AEs by LP*

The prediction of AEs was also performed by solving a LP problem using one-class SVM. Let  $\mathbf{g}_A, \mathbf{g}_B$  be the differential gene expression data described in “Differential gene expression data for the prediction” section without PCA of drug A and drug B respectively, and let  $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{iM})$ ,  $\mathbf{a}_{ii}$  be vectors in which each element represents the number of reports for indication 1, indication 2, ..., indication M that are co-reported with AE i and AE ii, respectively.

In the training phase of one-class SVM to predict AEs for drug X, we made combined vector for the drug A – AE i combination  $\mathbf{g}_A \mathbf{a}_i$  as a positive-link sample when more than 10 records exist for drug A inducing AE i in FAERS, and input all positive sample vectors without those of drug X to a one-class SVM. The kernel function of one-class SVM was designed as below,

$$k(\mathbf{g}_A \mathbf{a}_i, \mathbf{g}_B \mathbf{a}_{ii}) = \langle \mathbf{g}_A, \mathbf{g}_B \rangle \langle \mathbf{a}_i, \mathbf{a}_{ii} \rangle$$

where  $\langle \cdot, \cdot \rangle$  represents operator that scales each vector to be that l2-norm equals 1 and take inner product between both scaled vectors. In the test phase of one-class SVM to predict the AEs for drug X, we input a combined vector for the target drug-AE combination  $\mathbf{g} \mathbf{a}$  and ask whether the AE occurs with the drug

(i.e., the presence or the absence of the drug-AE link in the LP problem), and SVM returns 1 (occur) or -1 (not occur). We repeat this prediction all AEs reported at FAERS. One-class SVM was performed on Python package ‘scikit-learn’ with parameter  $\nu=0.1$ .

#### *REACTOME and KEGG analyses*

REACTOME and KEGG analyses were performed on the gene sets of each organ with the feature importance greater than 0 in the RF prediction of the death event. REACTOME pathway analyses was performed REACTOME web API (<https://reactome.org/dev/analysis>) and KEGG pathway analyses was performed on R library ‘clusterR’.

#### *Prediction of AEs by SVR and RFR*

For the quantitative prediction (Figure 7), the incidence of an AE for a drug and each sex/age-group was calculated as (the number of the reports of the AE for a sex/age-group)/(the number of the reports of all AEs for all sex/age-groups) for each drug. These values were used as target variables with two machine learning algorithms, support vector regressor (SVR) and random forest regressor (RFR) to train the models. The features of the two algorithms are differential gene expression data described in “Differential gene expression data for the prediction” section. While the actual incidence ranges from 0.0 to 1.0, the output values from the trained machine learning models range from  $-\infty$  to  $\infty$ . To address this issue, we used the output of logit function (inverse function of sigmoid function) when inputting the incidence values. SVM was performed on R library ‘e1071’ with default parameters. RF was performed on Python package scikit-learn. RF parameter ‘n\_estimators’ was set to 10, and other parameters (‘max\_feature’, ‘max\_depth’ and ‘min\_samples\_leaf’) were decided by leave-one-out cross-validation for each AE, sex and age-group.

#### *Prediction of therapeutic indications*

The LP of therapeutic indications was performed by one-class SVM as follows: Let  $\mathbf{g}_A, \mathbf{g}_B$  be the differential gene expression data described in “Differential gene expression data for the prediction” section without PCA for drug A and drug B, respectively, and let  $\mathbf{d}_1 = (d_{1i}, d_{1ii}, \dots, d_{1N}), \mathbf{d}_2$  be

vectors that element represents the number of reports AE i, AE ii, ..., AE N that be co-reported indication 1 and indication 2, respectively. In the training phase of one-class SVM to predict the indications for drug X, we made combined vector for the drug A - indication 1 combination  $\mathbf{g}_A \mathbf{d}_1$  as a positive sample when more than 10 records exist for drug A taken by patients with the indication 1 in FAERS, and input all positive sample vectors without those for drug X to a one-class SVM. The kernel function of the one-class SVM was designed as below,

$$k(\mathbf{g}_A \mathbf{d}_1, \mathbf{g}_B \mathbf{d}_2) = \langle \mathbf{g}_A, \mathbf{g}_B \rangle \langle \mathbf{d}_1, \mathbf{d}_2 \rangle$$

where  $\langle \cdot, \cdot \rangle$  represents operator that scales each vector to be that l2-norm equals 1 and takes inner product between both scaled vectors. In the test phase of one-class SVM to predict the indications for drug X, we input a combined vector for target drug-indication combination  $\mathbf{g} \mathbf{d}$  and ask whether the drug can treat the indication (i.e., the presence or the absence of the drug-indication link in the LP problem), and SVM returns 1 (can treat) or -1 (cannot treat). We repeat prediction for all indications. One-class SVM was performed on Python package 'scikit-learn' with parameter nu=0.1.

#### *Validation with SIDER*

The Side effects (SEs) and indication datasets reported in SIDER4.1 (<http://sideeffects.embl.de>) were compiled for each drug. The total number of SEs and indications reported (labeled) for the 14 drugs in this study (the 15 drugs minus evolocumab) are 3,061 and 2,088, respectively. SIDER does not classify the incidence of each SE according to sex/age-groups. Hence, all sex/age-group division-based predictions with FAERS are combined for each drug. When the outcomes of an AE in any one or more of the sex/age-groups show positive for a drug in the FAERS-prediction, we scored it as positive for that AE and for this drug, and scored the accuracy, precision and recall using the reported/labeled outcomes in SIDER4.1.

### **Supplementary References**

Kozawa, S., Ueda, R., Urayama, K., Sagawa, F., Endo, S., Shiizaki, K., Kurosu, H., Maria de Almeida, G., Hasan, S.M., Nakazato, K., *et al.* (2018). The Body-wide Transcriptome Landscape of Disease Models. *iScience* 2, 238-268.