Stem Cell Reports
Article

OPEN ACCESS

Machine Learning of Human Pluripotent Stem Cell-Derived Engineered Cardiac Tissue Contractility for Automated Drug Classification

Eugene K. Lee,[1,2,6] David D. Tran,[2,6] Wendy Keung,[3,4] Patrick Chan,[3,4] Gabriel Wong,[2] Camie W. Chan,[2] Kevin D. Costa,[2,5] Ronald A. Li,[2,3,4] and Michelle Khine[1,2,*]

[1]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, USA
[2]Novoheart LTD, Shatin, Hong Kong
[3]Dr. Li Dak-Sum Research Centre, The University of Hong Kong – Karolinska Institutet Collaboration in Regenerative Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong
[4]Ming-Wai Lau Centre for Reparative Medicine, Karolinska Institutet, Stockholm 17177, Sweden
[5]Cardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York City, NY 10029, USA
[6]Co-first author
*Correspondence: mkhine@uci.edu
https://doi.org/10.1016/j.stemcr.2017.09.008

SUMMARY

Accurately predicting cardioactive effects of new molecular entities for therapeutics remains a daunting challenge. Immense research effort has been focused toward creating new screening platforms that utilize human pluripotent stem cell (hPSC)-derived cardiomyocytes and three-dimensional engineered cardiac tissue constructs to better recapitulate human heart function and drug responses. As these new platforms become increasingly sophisticated and high throughput, the drug screens result in larger multidimensional datasets. Improved automated analysis methods must therefore be developed in parallel to fully comprehend the cellular response across a multidimensional parameter space. Here, we describe the use of machine learning to comprehensively analyze 17 functional parameters derived from force readouts of hPSC-derived ventricular cardiac tissue strips (hvCTS) electrically paced at a range of frequencies and exposed to a library of compounds. A generated metric is effective for then determining the cardioactivity of a given drug. Furthermore, we demonstrate a classification model that can automatically predict the mechanistic action of an unknown cardioactive drug.

INTRODUCTION

Drugs in late-stage development as well as those with market approval are often withdrawn due to previously undetected drug-induced cardiotoxicity. Unpredicted drug-induced cardiotoxicity jeopardizes patients' lives, erodes public trust in the regulatory process, and financially burdens the pharmaceutical industry. For example, cisapride, a gastrointestinal drug intended to treat heartburn, was reported to have caused serious ventricular arrhythmias and sudden deaths prior to withdrawal (Ferriman, 2000). The associated pharmaceutical company agreed to settle lawsuits for a total of $90 million US dollars for 300 deaths and 16,000 injuries (Harris and Koli, 2005). Progress has been made to facilitate better safety through the adoption of US Food and Drug Administration guidelines that recommend screening new drugs with the human ether-a-go-go-related gene (hERG) inhibition assay (FDA, 2005). Nonetheless, within the past decade, an assortment of market-approved therapeutics (e.g., clobutinol, sibutramine, and tegaserod) had to be withdrawn due to unpredicted drug-induced cardiotoxicity (FDA, 2016; Li et al., 2016).

The need for more accurate and faster pre-clinical detection methods has led to the emergence of a variety of screening platforms in recent years. A majority of these detection systems have begun to utilize human pluripotent stem cell-derived cardiomyocytes (hPSC-CMs). These cells are more physiologically similar to human myocardium than laboratory animals or genetically transformed non-cardiac cell lines (e.g., human embryonic kidney cells) (Dick et al., 2010). Aside from the commonality of using hPSC-CMs, these platforms vary greatly in their setup, including differences in tissue geometry. Certain platforms employ hPSC-CMs in a two-dimensional manner (e.g., monolayer), while others attempt to recapitulate aspects of the three-dimensional environment of native myocardium by modeling the cells as cardiac muscle fibers or heart chambers (Chen et al., 2014; Huebsch et al., 2016; Lee et al., 2008). These systems can also differ drastically in their methodology of quantifying changes in hPSC-CMs exposed to cardioactive compounds. Some systems examine the electrophysiological properties while others focus on the calcium transients or measurements of contractility (e.g., shortening, force, pressure) generated by the cardiomyocytes (Lu et al., 2015; Maddah et al., 2015; Navarrete et al., 2013; Zhang et al., 2014).

While the experimental platforms may vary, quantitative readouts generally characterize individual contractile events. Combining this with the number of experimental conditions (e.g., various pacing frequencies or drug concentrations) can yield large multidimensional datasets that make it difficult to draw thorough conclusions. Researchers often simplify the raw data by preselecting a limited number of parameters as an attempt to

comprehend the complexity of the data, albeit losing information in the process. Without full interpretation of such a rich dataset, there is a risk of not detecting information that differentiates the behavior of normal hPSC-CMs from those exposed to cardioactive compounds. In addition, as these platforms are meant for high-content screens, the analysis of the datasets needs to be automated.

Holistic approaches must be developed to optimize the utility of datasets generated from screening platforms. Machine learning has been shown to handle such multidimensional datasets in an automated fashion (Lee et al., 2015). We previously demonstrated that support vector machine (SVM), a supervised learning algorithm, can be used to consolidate 12 parameters, which characterize the contractile behavior of hPSC-CMs exposed to cardioactive compounds, into a singular quantitative index that expresses the level of induced cardioactivity (Lee et al., 2015). Machine learning can be further leveraged into a suite of tools that provide more in-depth details of hPSC-CM behavior when exposed to cardioactive compounds. In this study, we hypothesize that multiclassification algorithms can be implemented to create a model to define drug classes and subsequently predict an unknown compound's mechanistic action. Such information would assist in streamlining the drug discovery pipeline, allowing for the rapid identification of select compounds for more in-depth follow-up assays. In addition, this information coupled with knowledge of predicted class can guide scientists to efficiently and selectively screen for specific drug-to-drug interactions that prompt cardiotoxicity (e.g., disruption of Ca^{2+} handling when sofosbuvir and amiodarone are combined) instead of relying on the traditional brute force approach (Millard et al., 2016). Furthermore, drug-response relationships between the unknown compound and the library can be determined.

To test our hypothesis, we examined a database (not yet published) containing drug screens of various compounds on twitch force measurements from human ventricular cardiac tissue strips (hvCTS) engineered from hPSC-CMs embedded in a 3D collagen-based matrix (Turnbull et al., 2014). A unique aspect of these screens was that the hvCTSs were electrically paced at four different frequencies from 0.5 to 2.0 Hz, spanning a physiological range. These measurements interrogated the influence of cardioactive compounds on the hvCTS force-frequency relationship and contributed to a multidimensional dataset. We selected a total of 12 compounds with acute cardiac effects that represented five drug classes (1, Ca^{2+} channel blockers; 2, adrenergic agonists; 3, cardiac glycosides; 4, hERG K^+ channel blockers; and 5, angiotensin-converting enzyme [ACE] inhibitors) along with one reference compound (aspirin). In this study, we report the use of machine learning to establish a drug classification model based on hvCTS contractile behavior (using half of the selected compounds) and subsequently demonstrate predictive capabilities by having the model classify unknown compounds, which were withheld from the machine during training.

RESULTS

Formation of the Drug Classification Model

To form the drug classification model, the screens of 12 compounds (Table 1) acquired on the hvCTS platform were used. Each of the compounds, with the exception of aspirin, belonged to one of five classes with each class comprising a minimum of two compounds. Aspirin functioned as a reference for a cardiac-neutral compound. To quantify the cardioactive effects of these compounds, a total of 17 parameters were derived from each contractile event recorded in the hvCTS twitch force versus time tracings (Figure S1). Once the parameters characterizing each contraction were calculated, establishing the library with machine learning consisted of two primary steps. The first step was determining the degree of cardioactivity for each compound at a given dosage by calculating a singular quantitative index generated by a binary SVM approach (Figure 1A). The second step involved multiclass SVM to establish the boundaries that represent each drug class (Figure 1B).

The binary SVM is capable of summarizing all parameters and providing a simple metric that expresses a compound's degree of cardioactivity at a given dosage (Lee et al., 2015). Specifically, the machine is tasked with creating a decision boundary that separates two groups (data from untreated hvCTSs and those from hvCTSs exposed to a concentration of a compound) as seen in Figure 1A. The decision boundary is evaluated for generalizability by classifying withheld data, referred to as a test set. As a result, an SVM accuracy metric is calculated to reflect the machine's ability to identify a cardioactive effect (e.g., 75% accuracy means that out of 100 data points, the machine could correctly classify 75 of them). To account for variations within the dataset, multiple runs are performed to calculate the mean SVM accuracy. A value of approximately 50% mean SVM accuracy suggests non-cardioactivity as the machine cannot distinguish between treated hvCTS data from control data, and the classification becomes random. Prominent cardioactive effects allow the space between the data from treated and control conditions to become more distinguishable, leading to a higher SVM accuracy with 100% being the maximum.

The second step was the utilization of multiclass SVM to create and evaluate a model. The 11 compounds (excluding aspirin) were divided into two groups (Figure 1B). The first group was referred to as a "drug library" and used to train

Table 1. Library Compounds

Compound Name	Class	Description	Test Range (M)	Number of hvCTS (*n*)
Nifedipine	Ca^{2+} channel blocker	an L-type Ca^{2+} channel blocker known to shorten action potential duration (Harris et al., 2013)	10^{-8} to 3.0×10^{-5}	10
Mibefradil	Ca^{2+} channel blocker	a tetralol derivative that blocks both L- and T-type Ca^{2+} channels with higher affinity for T-type (Martin et al., 2000)	10^{-9} to 3.0×10^{-6}	6
Isoproterenol	adrenergic agonist	a mixed β-adrenergic agonists; compound is non-selective in terms of βreceptors (Steinberg, 1999)	10^{-8} to 10^{-4}	10
Norepinephrine	adrenergic agonist	mixed adrenergic agonist that stimulates both α and βreceptors (Yang et al., 2014)	10^{-9} to 10^{-5}	8
Digoxin	cardiac glycoside	a cardiac glycoside that inhibits Na^+/K^+-ATPase, resulting in higher intracellular Na^+; higher Na^+ concentration suppresses the Na^+/Ca^{2+} exchanger causing the accumulation of intracellular Ca^{2+} (Katz et al., 2010)	10^{-8} to 10^{-4}	9
Ouabain	cardiac glycoside	a cardiac glycoside that affects Na^+/K^+-ATPase, which consists of both α and β subunits; has a lower affinity for α subunits than digoxin (Katz et al., 2010).	10^{-8} to 10^{-4}	10
Flecainide	hERG K^+ channel blocker	a mixed hERG K^+ blocker that also inhibits Na^+ channels, causing effects on action potential repolarization and conduction (Harris et al., 2013)	10^{-8} to 10^{-4}	8
E-4031	hERG K^+ channel blocker	a pure hERG K^+ channel blocker known for its pro-arrhythmic potential (Ziupa et al., 2014)	10^{-8} to 10^{-4}	8
Cisapride	hERG K^+ channel blocker	a serotonin ($5\text{-}HT_4$) receptor agonist that also inhibits the hERG K^+ channel (Wong et al., 2010)	10^{-8} to 10^{-4}	9
Lisinopril	ACE inhibitor	an ACE inhibitor, which reduces vasoconstriction and lowers blood pressure in patients (Williams, 1988)	10^{-8} to 10^{-4}	8
Ramipril	ACE inhibitor	an ACE inhibitor; it does not block ACE until it is converted by the liver (Williams, 1988)	10^{-9} to 10^{-5}	7
Aspirin	non-cardioactive reference	non-steroidal anti-inflammatory drug that has been shown to have no cardioactive effects in screening platforms (Maddah et al., 2015)	10^{-8} to 10^{-4}	6

n refers to independent biological replicates. hvCTS, human ventricular cardiac tissue strips; hERG, human ether-a-go-go-related gene; ACE, angiotensin-converting enzyme.

a model that defined the boundaries of each drug class. The second group was completely withheld from the computer throughout the entire training and tuning of the model and was referred to as the "unknown" compounds. To normalize for the varying degrees of cardioactivity among compounds, the concentration of a compound that achieved a metric closest to 85% mean SVM accuracy (see Supplemental Experimental Procedures) was used in the formation and evaluation of the model. To ensure that the library was generalizable or had the capability to classify itself, a subset of the first group's data, a test set, was randomly withheld from the computer prior to training. The library was evaluated on its performance to accurately identify this test set. Afterward, the library was asked to predict the unknown compounds, and its performance was evaluated for predictive capabilities. To account for the variation and random selection of training and test sets, the creation and evaluation of the models were performed 50 times.

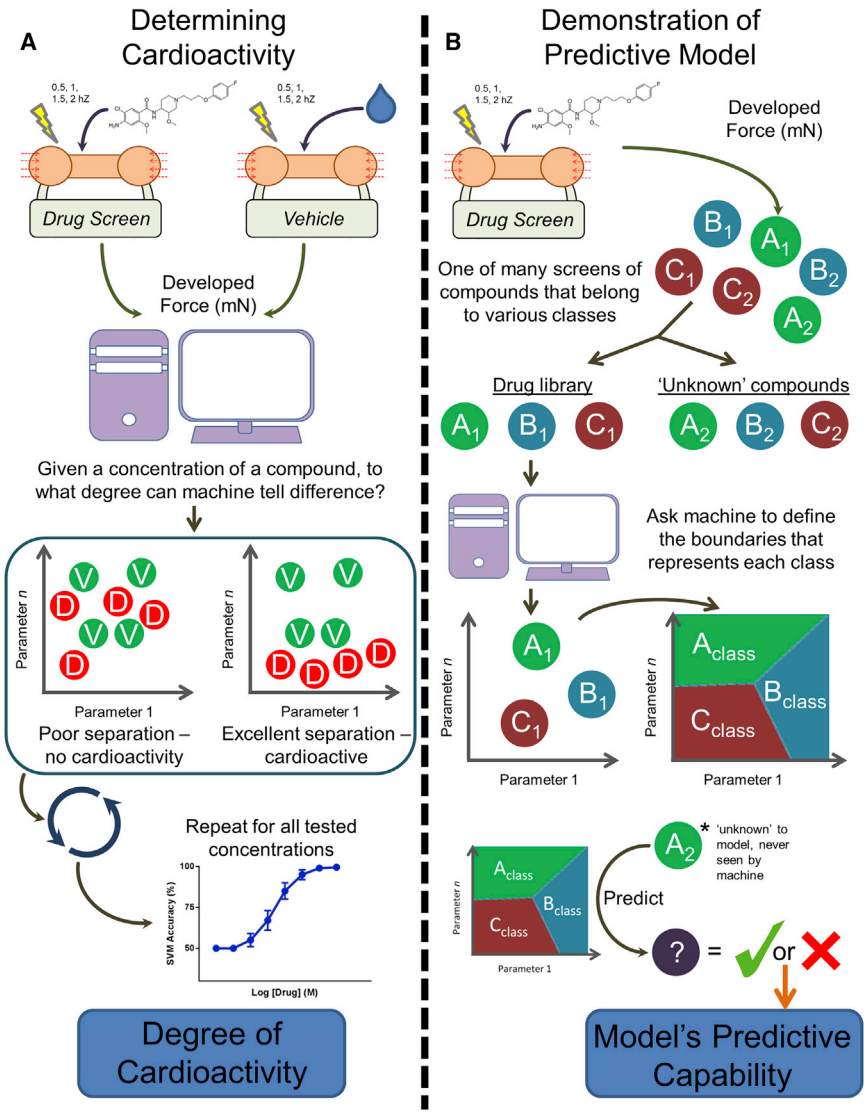

Figure 1. Determining Cardioactivity and Predicting Mechanistic Action

(A) To determine if a concentration of a compound was cardioactive, parameters describing the force waveforms were compared with those of vehicle study with the binary SVM approach. If the compound does not modulate the contractile behavior of hvCTSs, the generated data points would be similar to those of vehicle, yielding poor separation and an SVM accuracy of approximately 50%. However, if a compound's cardioactive effects become more distinguishable, separation between the two groups becomes more feasible and results in a higher SVM accuracy (100% as maximum distinguishability). This binary SVM approach generates a singular quantitative index that describes the degree of cardioactivity of a given concentration for a drug compound.

(B) To create a model for the prediction of mechanistic action, data from screened compounds were divided into two groups: library and unknown. Using data from the library group, the machine defined boundaries that represent various drug families. The model was evaluated for its predictive capabilities by having the machine classify the unknown compounds.

Control Experiments

Although the hvCTSs (Figure 2A) were examined under temperature-controlled conditions, there was an observable drift in contractile behavior of the vehicle-treated hvCTSs. For example, the relative measured maximum developed force increased for all pacing frequencies by a cumulative average of 16.96% ± 0.83% upon the ninth serial addition (Figure 2B). To account and normalize for baseline drift, each drug condition was compared with its respective vehicle condition via binary SVM (e.g., measurements of the seventh serial drug addition were compared with those of vehicle-treated hvCTSs at the seventh serial addition). To establish a benchmark of non-cardioactivity, a subset of the vehicle-treated hvCTSs was randomly selected to model a non-cardioactive compound. Binary SVM was then performed between the subset and a corresponding control group of equal size (n).

To ensure that the number of hvCTSs in the subset had no effect, the calculations were performed with the sample size, n, equal to 6–10, which matched the range of numbers of strips used in each drug study of the 12 tested compounds. As expected, the SVM accuracy, regardless of the size of n, was approximately 50% for all serial additions (Figure 2C). These results indicated that there were no consistent and distinguishable trends within each serial addition, and therefore a reference of non-cardioactivity was created.

To validate this reference of non-cardioactivity, drug screens of aspirin from the database were used as negative controls. Aspirin is known to have no cardioactive effects on hPSC-CMs (Lu et al., 2015; Maddah et al., 2015; Scott

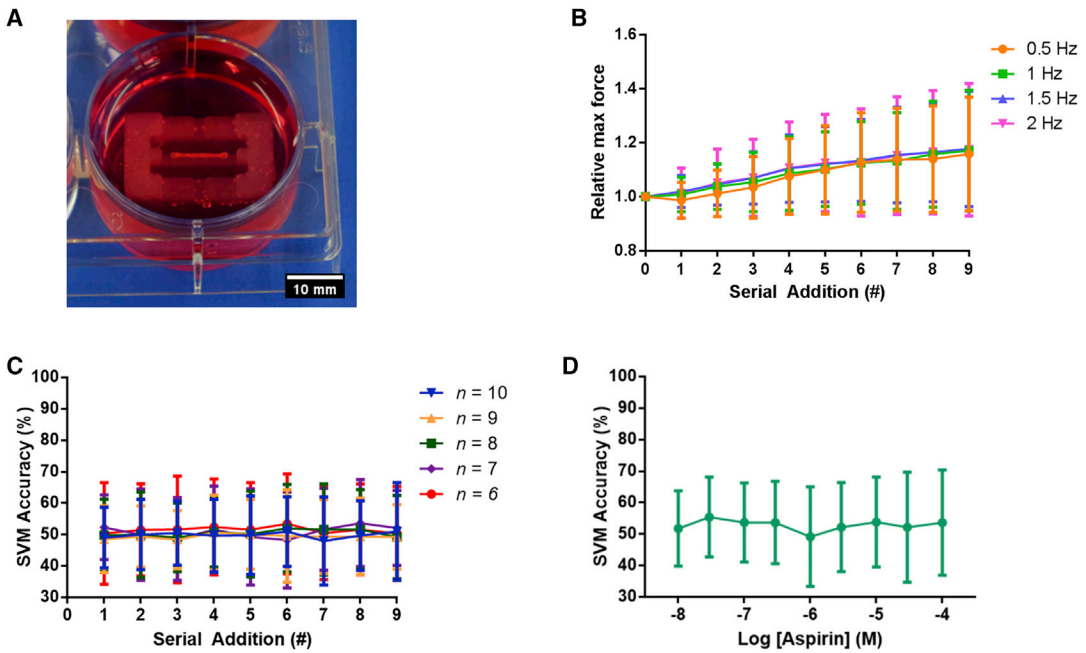

Figure 2. Control Experiments and Screen of Non-cardioactive Compounds

(A) hvCTSs were formed by compaction of hPSC-CMS, fibroblasts, and extracellular matrix solution around two PDMS posts. Force was derived from the deflection of the posts. Scale bar, 10 mm.

(B) hvCTSs of the vehicle study exhibited a drift in contractile behavior after serial additions of water (n = 28).

(C) To normalize for shift, data from each drug condition were compared with its respective vehicle condition (matching serial addition number). With this, a benchmark of non-cardioactivity was created with subsets of vehicle-treated strips being modeled as strips exposed to non-cardioactive compounds and compared with other vehicle-treated strips with binary SVM. The number of tissue strips in the subset (n = 6–10) had no effect as SVM accuracies were approximately 50% for all conditions.

(D) To assess the non-cardioactivity benchmark, the data of aspirin-treated hvCTSs indicated no statistical difference from the vehicle study over the tested range (n = 6). n refers to independent biological replicates.

All results are presented as means ± SD.

et al., 2014). The SVM accuracies of the aspirin drug screens (n = 6) had an average of 52.85% ± 1.77% among all serial additions (10 nM to 100 μM). None of the conditions were statistically different from vehicle counterparts, indicating non-cardioactivity by aspirin (Figure 2D).

Generalizability of the Drug Classification Model

In setting up the drug classification model, the 11 non-reference compounds were compared with vehicle-treated tissue strips with the aforementioned binary SVM approach. At one or more of the tested concentrations, all but two compounds, lisinopril and ramipril, had SVM accuracies that were significantly greater than those of the respective vehicle studies (Figure 3). The ACE inhibitors, lisinopril and ramipril, did not have detectable cardioactive effects on hvCTS contractility, consistent with the results of other platforms (Harmer et al., 2012; Scott et al., 2014). Therefore, the ACE inhibitor class was removed from the library resulting in a four-class system. Mibefradil, norepinephrine, ouabain, and cisapride were chosen to represent the unknown compound group on which the model would make *de novo* predictions. Mibefradil and cisapride were of particular interest because both compounds had received market approval and were subsequently withdrawn (Li et al., 2016). Nifedipine, isoproterenol, and digoxin were chosen to represent the Ca^{2+} channel blocker, adrenergic agonist, and cardiac glycoside classes, respectively. As both flecainide and E-4031 have known hERG K^+ channel-blocking capabilities, the impact of having either compound represent the hERG K^+ blocker family was evaluated by generating the multiclass model under three different conditions: (1) flecainide only, (2) E-4031 only, and (3) both flecainide and E-4031.

A subset of the data was always withheld from the machine prior to training in each of the runs. This withheld set quantified the generalizability in the models and ensured that overfitting had not occurred. Upon asking the machine to classify these test sets, the multiclass models demonstrated good generalizability by being able

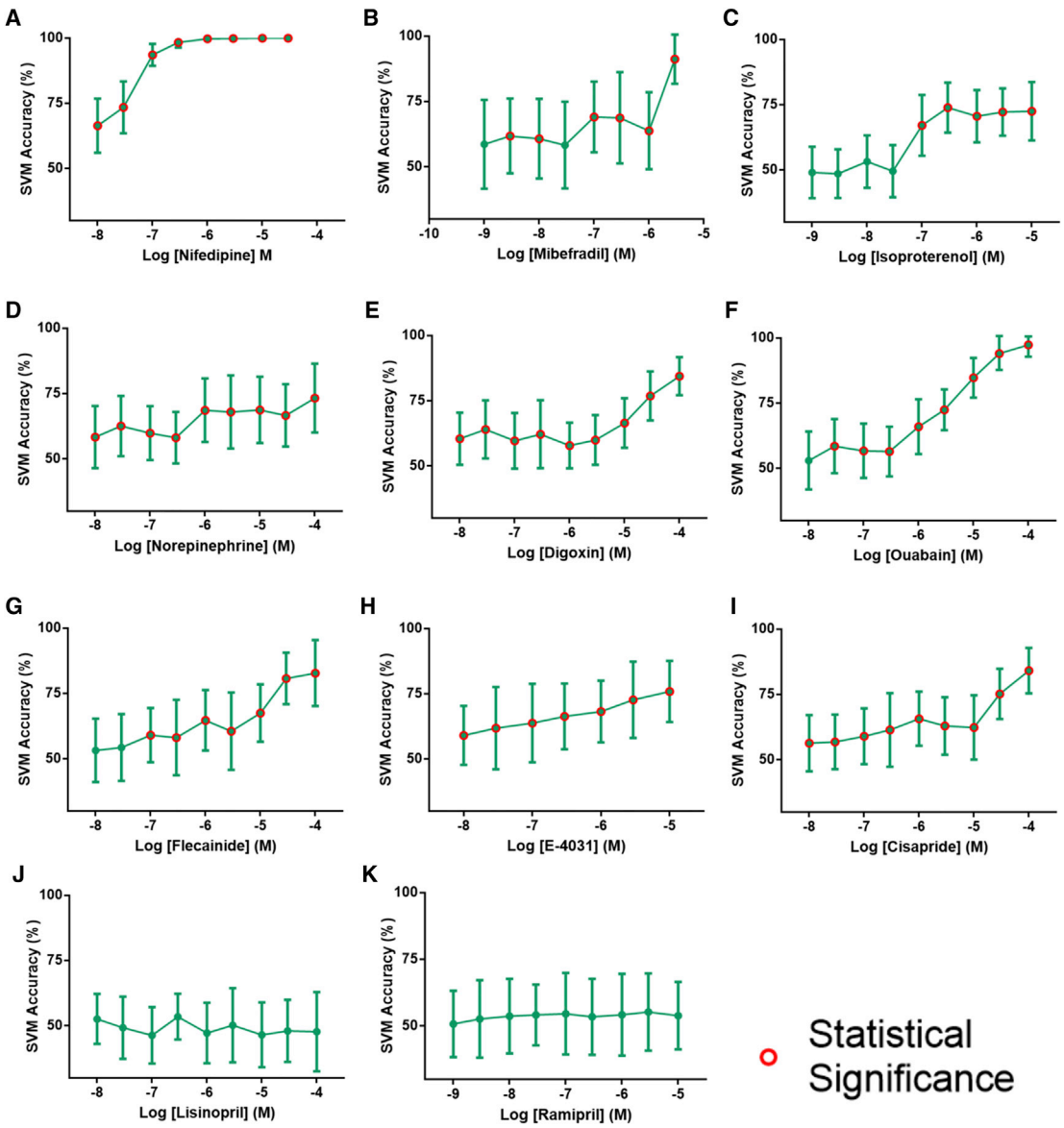

Figure 3. Applying Cardioactive Index to Drug Responses

Implementation of binary SVM to determine cardioactive effects of 11 compounds, including: (A) nifedipine ($p \leq 0.0063$; $n = 10$), (B) mibefradil ($p \leq 0.0063$; $n = 6$), (C) isoproterenol ($p \leq 0.0055$; $n = 10$), (D) norepinephrine ($p \leq 0.0055$; $n = 8$), (E) digoxin ($p \leq 0.0055$; $n = 9$), (F) ouabain ($p \leq 0.0055$; $n = 10$), (G) flecainide ($p \leq 0.0055$; $n = 8$), (H) E-4031 ($p \leq 0.0071$; $n = 8$), (I) cisapride ($p \leq 0.0055$; $n = 9$), (J) lisinopril ($p \leq 0.0055$; $n = 8$), and (K) ramipril ($p \leq 0.0055$; $n = 7$). Red circles indicate statistical significance in comparison with vehicle-treated hvCTSs. p values are adjusted with a Bonferroni correction. *n* refers to independent biological replicates. All results are presented as means ± SD.

to correctly classify itself at an average accuracy rate of 76.09% ± 6.43%, 78.29% ± 5.34%, and 73.61% ± 5.19% for the flecainide only, E-4031 only, and flecainide and E-4031 conditions, respectively (Figure 4).

In all three conditions, the multiclass models behaved similarly in that both the nifedipine and isoproterenol classifiers performed the best by always achieving the highest F_1 score values, a metric that ranges from 0 to 1 with 1 representing perfection in the model's classification. This performance indicates that the data points of the nifedipine and isoproterenol compounds occupied very distinct boundaries compared with the other two classes, allowing for the binary learners to more accurately separate these compounds from others. For perspective on the quality of model performance, if there were no discernable differences among the four compounds for the machine to use,

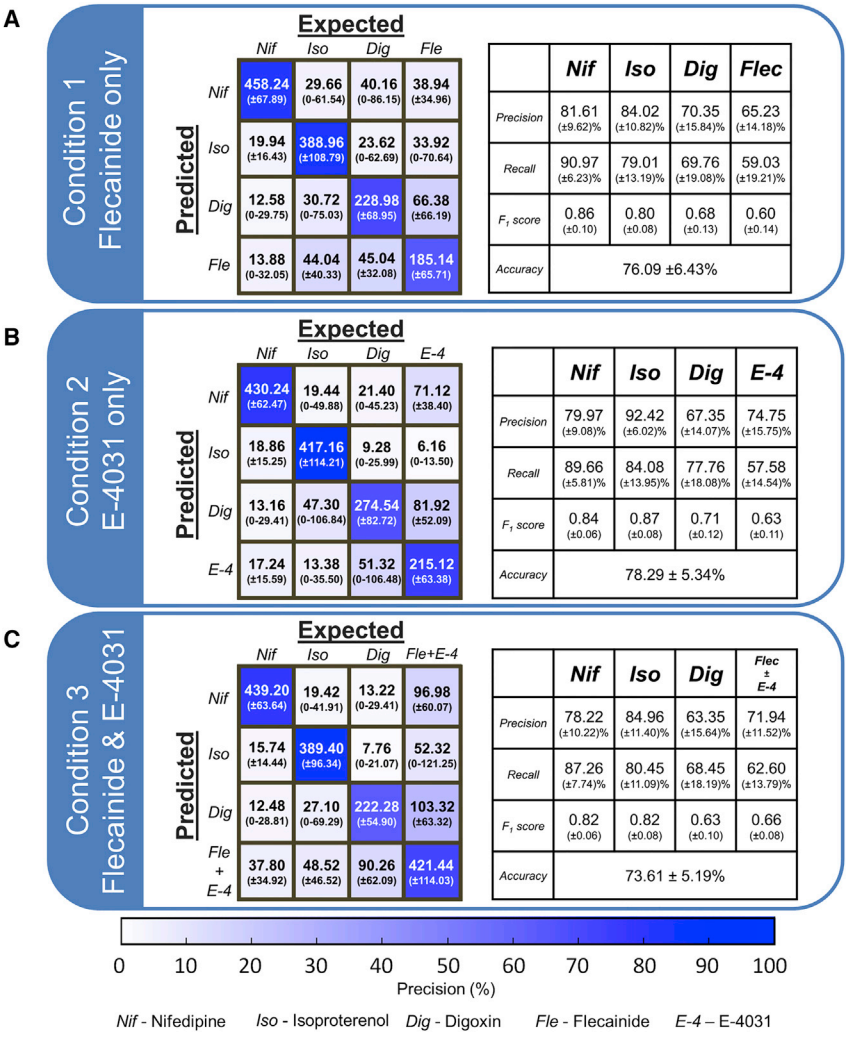

Figure 4. Generalizability of Drug Classification Models with Varying Representative Compound(s) for the hERG K^+ Channel Blocker Family

(A) Condition 1: flecainide. Generalizability was evaluated by having the model classify a withheld test set composed of compounds used to define the library. (Left) The confusion matrix displays the average number of classified contractile events (individual twitches in acquired waveform) over 50 runs and is imposed with a color scale that indicates precision rate. (Right) Summary of the metrics evaluating performance. F_1 scores are above 0.6 for all classifiers, indicating good predictability.

(B) Condition 2: E-4031. (Left) Confusion matrix indicates that all compounds were being correctly classified as themselves as the diagonal of the matrix had the highest precision rate. (Right) Metrics indicate a similar performance between conditions 1 and 2.

(C) Condition 3: flecainide and E-4031. (Left) Confusion matrix indicates no major effect on generalizability by having two compounds define a class. (Right) Similar to conditions 1 and 2, F_1 scores are all above 0.6, indicating good generalizability of the model.

All results are presented as means ± SD.

the expected values for precision (i.e., positive predictive value), recall (i.e., sensitivity), and accuracy would be a rate of 25% with an F_1 score of 0.25 (see Experimental Procedures). As all three multiclass models demonstrated good generalizability with average accuracy rates exceeding 70%, these results suggest the setup of the model was robust to the choice of compound representing the hERG K^+ channel blocker family.

Prediction on Unknown Compounds

With each condition's model established and evaluated, the machine was then asked to predict the data from the unknown compounds group. In the first scenario with flecainide as the only hERG K^+ channel blocker representative, the multiclass model was able to correctly assign the four unknown compounds to their corresponding counterparts with an average accuracy of 71.69% ± 1.96% (Figure 5A). The four binary classifiers of this model demonstrated overall predictability by all having an average F_1 score above 0.6.

When the second drug class model (only E-4031 defining the hERG K^+ channel blocker class) was used to predict the unknown compounds, the average accuracy diminished to 65.37% ± 2.33% (Figure 5B). This decrease was mainly the result of the contractile events from hvCTSs exposed to cisapride being miscategorized. Of the 779 contractile events from hvCTSs exposed to cisapride, on average, 46.76% ± 7.00% of the events were incorrectly labeled as hvCTSs affected by a cardiac glycoside. These misclassifications yielded a poor precision rate, 20.51% ± 11.20%, and subsequently, a low F_1 score, 0.27 ± 0.11, for the hERG K^+ channel blocker classifier. As for the classifiers of the other three families, they performed comparably with those of the first condition, and all had an average F_1 score over 0.6. These results indicate that using only data of hvCTSs exposed to E-4031 is not sufficient for defining the hERG K^+ channel

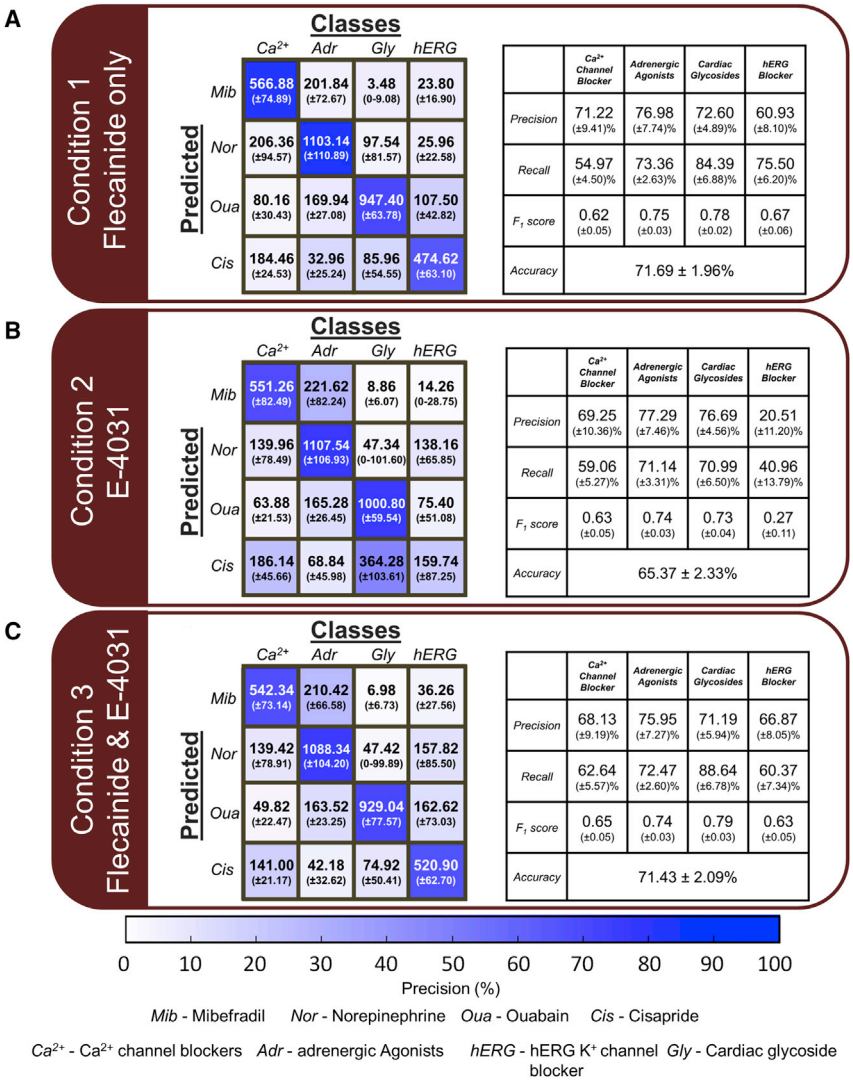

Figure 5. Predictability of Drug Classification Libraries with Varying Representative Compound(s) for the hERG K^+ Channel Blocker Family

(A) Condition 1: flecainide. Libraries of all three conditions were evaluated for capabilities to predict drug families of compounds previously unseen by the machine. (Left) Confusion matrix displays the average number of predicted contractile events (individual twitches in acquired waveform) from the unknown group over 50 runs. Unknown compounds were classified to the correct drug family as the precision rate was highest along the diagonal of the matrix. (Right) The model demonstrated good predictability as the macro-average of the F_1 score was 0.71.

(B) Condition 2: E-4031. (Left) Cisapride was predominantly mislabeled as a member of the cardiac glycoside family, represented by digoxin. The other three compounds were classified correctly to their respective drug families. (Right) The classifier of the hERG K^+ channel blocker class had an F_1 score of 0.27. As the other three classifiers maintained an F_1 score above 0.6, E-4031 did not define boundaries similar to that of cisapride.

(C) Condition 3: flecainide and E-4031. (Left) Precision rates are highest along the diagonal of the matrix. (Right) The model yielded good predictability and was not affected by the inclusion of E-4031 data to those of flecainide.

All results are presented as means ± SD.

blocker family within a multiclass model and enabling that model to correctly predict cisapride's mechanistic action on cardiac tissue contractility.

In the last condition where the machine was trained with both flecainide and E-4031 representing the hERG K^+ channel blocker family, the average accuracy was 71.43% ± 2.09% (Figure 5C). All four classifiers had an average F_1 score higher than 0.6. Unlike the second condition, the hERG K^+ channel blocker classifier had a precision rate higher than 25% at 66.87% ± 8.05%. Similar to that of the first condition, the macro-average of the F_1 scores for this multiclass model was 0.71, demonstrating good predictability. These results suggest that the data points from the hvCTSs exposed to flecainide designated boundaries within the model that were similar to those of cisapride-treated hvCTSs. Finally, the inclusion of E-4031 data with the flecainide data did not negatively affect the multiclass model's ability to correctly predict cisapride's mechanistic action as neither the precision rate nor the recall rate fell below 41% as they did in the second condition.

Class Relationship Metrics

Once the drug classes were predicted, the concentrations of library compounds that induced the most similar cardioactive effects as the unknown compounds were computed (Figure S2; protocol in Supplemental Experimental Procedures). These relationship metrics are summarized in Table S1. For example, an estimated 5.35×10^{-5} M digoxin would be needed to evoke a level of cardioactivity that matches ouabain tested at 1.0×10^{-5} M. Such relationships could provide insights about drug potency. In the aforementioned example, ouabain would be considered the more potent compound as it requires approximately 5-fold lower concentration to achieve the same level of

cardioactivity. Ouabain's higher potency has been observed in other *in vitro* studies (Guo et al., 2011; Katz et al., 2010).

Decoupling Force-Frequency Relationship

While the concept of examining multiple parameters from waveforms has been pursued lately, some studies have suggested that only a few select parameters (e.g., peak count) are necessary in assessing a compound's cardioactivity as other parameters provide no further mechanistic insight (Lu et al., 2015; Pointon et al., 2016; Sirenko et al., 2013). This is primarily true when the hPSC-CMs are spontaneously beating, meaning the force generated is linked to beating frequency. This study's dataset affirmed the importance of decoupling this force-frequency relationship through the pacing of the tissues. By setting a fixed pacing frequency, any changes to the force waveform can be truly accredited to a compound's inotropic and lusitropic effects. For example, if the nifedipine-treated strips were allowed to spontaneously beat, a positive chronotropic effect would have most likely been observed (Guo et al., 2011; Harris et al., 2013; Pillekamp et al., 2012). As the hvCTSs displayed a negative force-frequency relationship (Figure S3A), a decrease in maximum developed force could not be directly linked to either chronotropic or inotropic effects. When 0.3 μM nifedipine-treated hvCTSs were paced at 1 Hz, the captured frequency of the tissue was 0.99 ± 0.01 Hz and the maximum developed force decreased by 45.69% ± 10.42% (Figure S3B). This paced data confidently confirmed nifedipine had a negative inotropic effect.

Examination of Cardioactive Effects

The data were further examined on an individual parameter basis to better comprehend and confirm the performance of the multiclass models and their ability to differentiate between different mechanistic actions (Figure S4). The adrenergic agonists and cardiac glycosides were expected to induce a positive inotropic response in the hvCTSs, while the Ca^{2+} and hERG K^+ channel blockers would induce a negative inotropic response. The negative inotropic agents prompted distinct decreases in the maximum force generated among hvCTSs; however, the hvCTS sensitivity to positive inotropic agents was not very apparent. For example, hvCTSs exposed to 10 μM isoproterenol and paced at 0.5 Hz had a similar increase in maximum developed force to that of respective vehicle-treated strips (10.42% ± 16.23% and 15.76% ± 21.05%, respectively), suggesting the compound had negligible inotropic effects (Figures S5A and S5B) (Turnbull et al., 2014). Rather, isoproterenol's cardioactive effects manifested in other parameters. For example, the duration of the relaxation phase or time from maximum developed force to 95% cutoff decreased by 22.19% ± 19.35% for the strips exposed to 10 μM isoproterenol, while those of the vehicle-treated strips experienced essentially no change. Thus, through the various parameters, machine learning was able to leverage the positive lusitropic effects in both binary and multiclass SVM to distinguish the adrenergic agonist class.

Cardiac glycoside-treated hvCTSs also demonstrated the system's sensitivity to positive inotropes. Typically, these compounds increase the Ca^{2+} transient amplitude and amplitude of contractility transients; however, in this dataset, the hvCTSs decreased in maximum developed force as the concentration increased (Figure S5C) (Dempsey et al., 2016; Ravenscroft et al., 2016). Such results could be attributed to cardiac glycoside toxicity or immature phenotype. Studies have shown that above 3 μM digoxin or ouabain, monolayers of hPSC-CMs stopped beating (Dempsey et al., 2016; Guo et al., 2013; Sirenko et al., 2013). In this dataset, the highest concentration applied for both cardiac glycosides was 100 μM. At 100 μM ouabain, 4 of the 10 treated strips stopped beating at all pacing frequencies.

Like those of isoproterenol, cardioactive effects of the cardiac glycosides at lower concentrations appeared in other parameters. When either the concentration of cisapride or digoxin increased, the maximum developed force decreased while the duration of the relaxation phase increased (Figures S5C and S5D). If limited to observation of only these two parameters, one might conclude that cisapride and digoxin are related. But upon visual inspection of the force traces, it is evident that the compounds have distinct effects on hvCTSs and mechanistic actions (Figures S3E and S3F). Changes in these parameters were clearly unique to the cardiac glycoside family as the ouabain-treated strips were correctly predicted from the other three classes.

DISCUSSION

In recognition of the need for better detection of drug-induced cardiotoxicity, numerous methodologies have emerged to capture and quantify the attributes of hPSC-CMs when exposed to cardioactive compounds, ranging from calcium transients to contractile force. The nature of the output data becomes multidimensional when multiple experimental conditions are present or a multiplex system is used (Dempsey et al., 2016). In this study, we present the use of supervised machine learning to exploit multidimensional data and provide relevant information in an automated manner. Besides indicating if a compound was cardioactive, the machine constructed a multiclass drug model that accurately classified cardioactive compounds that it had never previously encountered. This comprehensive

approach can be readily applied to other screening platforms to more fully utilize generated datasets and enhance evidence-based decision-making for drug development.

With multiclass SVM, drug classification libraries were established under various conditions to examine the effects on predictive performance. The conditions that yielded the best performance in predicting mechanistic action were the two libraries that included flecainide as a representative of the hERG K^+ channel blocker family. In both libraries, the macro-averages of F_1 scores were 0.71 (the macro-average of F_1 scores would be 0.25 if random classifiers were used). While this clear difference in F_1 scores indicates that the models have the capability to predict a compound's mechanistic action, there are opportunities to further improve model performance and obtain F_1 scores closer to 1, indicating reduction in errors.

One method to improve model performance is to define each drug family with multiple compounds. By having only one compound define a class, there is a risk of only defining a partial region of space that the drug class truly encompasses. The data of E-4031 exemplified this when it was tasked with defining the hERG K^+ channel blocker family. E-4031's defined boundaries did not match or include those of cisapride, another hERG K^+ channel blocker, causing classification of cisapride to be closer to that of the cardiac glycoside family. The inclusion of flecainide, a mixed hERG K^+ channel blocker, with E-4031 in the definition of the class allowed for the correct prediction of cisapride without adversely affecting the predictive capability of the remaining classes. Although the addition of E-4031 to the hERG K^+ channel blocker definition does not necessarily improve the predictive capability with respect to cisapride classification, establishing a more expansive region of space to define the hERG K^+ channel blocker class may improve prediction of other unknown hERG K^+ channel blockers that have effects more similar to E-4031 than flecainide. These results also suggest the potential of having subgroups within classes of the model, which can be achieved through a series of multiclass classifications. For instance, a compound can be predicted as a Ca^{2+} channel blocker in the first classification; within this family, the compound can be subsequently categorized into a subgroup (e.g., defined by frequency-dependent cardioactivity). As machine learning does not define drug classes with *a priori* knowledge (e.g., guidelines on how parameters are expected to change), the number of drug families and subclasses that can be defined within a model are not limited. The unbiased and automated nature of machine learning is also advantageous when a new drug family needs to be added, because no rubric needs to be manually amended and re-evaluated.

This study demonstrates the potential of machine learning for providing insights in the detection of cardioactivity using hPSC-CMs. The basis of this study's libraries was an error-correcting output codes approach with binary learners being SVM. Different binary learners, such as decision trees, should be explored alongside completely different approaches (e.g., neural networks). The ideal machine learning technique should balance predictive capabilities and use of computational resources. In this study, all models were generated with a standard desktop. Each calculated instance of a model took approximately 4 hr. However, once all models were formed, the predictions made on unknown compounds were on the timescale of seconds.

Improvements of the multiclass drug libraries can also be achieved from enhancements of the hvCTSs and acquisition system. In particular, the sensitivity of this system to positive inotropic compounds can be increased by addressing two issues: the maturity of stem cell-derived cardiomyocytes and the drifting baseline of vehicle-treated strips. Studies have shown that hPSC-CMs elicit a minimal to non-existent response to certain positive inotropic compounds, such as β-adrenergic agonists, because of immature intracellular structures (Lundy et al., 2013; Pillekamp et al., 2012). When these diminished responses are paired with a baseline that has increasing contractility over time, positive inotropic effects of a compound can get masked and become harder to detect as seen in the aforementioned isoproterenol drug screen. While hvCTSs were arranged in an aligned manner and co-cultured with fibroblasts, they can be further matured through additional techniques, such as conditioning by electrical stimulation, a cellular tri-culture including endothelial cells, or forced expression of selected proteins (Eng et al., 2016; Liu et al., 2009; Ravenscroft et al., 2016). As for stabilization of the baseline, different components of the setup, ranging from pH to CO_2 levels in an ambient environment, should be re-evaluated to minimize overall drift during serial additions. Increasing the system's sensitivity to positive inotropic agents would yield even more distinct boundaries and subsequently better predictability in the drug classification libraries.

While we demonstrated the importance of pacing hvCTSs, a limitation of this study is the absence of responses of spontaneously beating hvCTSs to cardioactive compounds. These data would provide understanding of a compound's chronotropic effects. Even though the model has already demonstrated good predictability on inotropic- and lusitropic-related data alone, the model would benefit from chronotropic-related data as it would provide more dimensions in which compounds can further distinguish themselves and yield better predictability.

In summary, we present the implementation of supervised machine learning on multidimensional data of

hvCTSs exposed to drugs while paced at various frequencies. In an automated fashion, this machine learning approach is able to not only determine if a compound is cardioactive but it can predict the mechanistic action along with other metrics. Furthermore, this approach can be adapted to state-of-the-art tissue-engineered cardiac models, including different forms of signals (e.g., calcium transients, micro-electrode array, and optical recordings), and has the potential to integrate diverse output data of multiplex systems or even those across platforms. Along with analyses of compounds with acute cardioactive effects, machine learning can be readily applied with non-invasive techniques to longitudinal studies to inspect a compound's chronic effects. Moreover, in the future, machine learning may be utilized on a grander scale by incorporating past clinical data to determine the optimal combination of *in vitro* and *in silico* data for the prediction of drug-induced cardiotoxicity in patients.

EXPERIMENTAL PROCEDURES

hvCTS Formation

Human ventricular cardiomyocytes were differentiated from an hES2 stem cell line with a Wnt inhibitor-based protocol as described previously (Weng et al., 2014). Human ventricular cardiac tissue strips (hvCTS) were then formed by mixing cardiomyocytes (100,000 cells per strip) at 14–16 days post differentiation with a solution of bovine collagen I (2 mg/mL), Matrigel (0.9 mg/mL), and human foreskin fibroblasts (100,000 cells per strip) as previously described (Turnbull et al., 2014). The cell-matrix solution (100 μL per tissue strip) was injected into a custom polydimethylsiloxane (PDMS) force-sensing bioreactor device and placed in an incubator (37°C and 5% CO_2). Formed hvCTSs were fed DMEM with 10% newborn calf serum, 1% penicillin-streptomycin, and 0.1% amphotericin B. The PDMS device contains two flexible vertical end posts to which the tissue anchors, causing the posts to deflect as the tissue beats. Contractile force measurements were captured with a high-speed (100 fps) CCD camera while custom LabVIEW software tracked the centroid movement of the flexible post tips. Force was converted from the deflection of the PDMS posts by an elastic beam-bending equation (Serrao et al., 2012). A custom MATLAB script was used to calculate 17 parameters that described the overall shape of the force traces for each contractile event (Figure S1). Each contraction was regarded as an individual data point for the machine learning analysis.

Drug Treatment

Seven to eight days post tissue formation, hvCTS were exposed to drugs for pharmacodynamic analysis. Flecainide, lisinopril, norepinephrine, and ramipril were provided by Pfizer, while all other compounds were purchased form Sigma-Aldrich. Compounds were initially resuspended in DMSO and subsequently diluted in water for final concentrations of less than 0.1% (vol/vol) DMSO. The PDMS device containing the hvCTS was placed onto a heated stage (37°C) under a dissecting microscope. Before either vehicle or drug addition, the medium was replaced with DMEM containing high glucose (4.5 g/L) and HEPES without phenol red. Drug doses were added to a tissue in a consecutively increasing manner up to 10 concentrations with 3 min between measurements. Vehicle doses containing only water were applied similarly. A pulse stimulator (AMPI Master-9) connected to platinum wires electrically paced the hvCTSs with a monophasic electric field of 5 V/cm and a 10 ms pulse duration.

Machine Learning

To establish the drug class model, we identified individual compounds that respectively represented our defined classes. The compounds and the corresponding tested concentrations are listed in Table 1. To determine which concentration of a chosen compound to add to the model, we first gauged each compound's level of cardioactivity by utilizing binary SVM (Figure S6) (Lee et al., 2015). To normalize for the increasing variation seen in hvCTS contractile behavior during the later serial additions of the vehicle studies, the SVM was performed between each concentration of a compound and the vehicle data from the corresponding serial addition number in which the concentration was achieved (Figure 1A). Specific details and rationale regarding the optimization of SVM classifiers and other implemented machine learning approaches are in the Supplemental Experimental Procedures.

For multiclass classification, we then selected the compound concentration that met two criteria: (1) a binary SVM accuracy closest to 85% and (2) at least six of all screened tissue strips were still responsive to electrical stimulation (see Supplemental Experimental Procedures). As seen in Figure 1B, we then divided the compounds into two groups: one used to train the multiclass model and one to represent unknown compounds for predictions. During this separation, it was ensured that each class was at least represented by one compound within the two groups. For training of the model, we used an error-correcting output codes approach with the binary learners being SVM (Dietterich, 1995). To confirm generalizability of the generated models, we randomly pre-allocated a third of the data as a test set prior to the training. Finally, we evaluated the multiclass model with this test set and then asked the machine to predict the classes of the contractile beats derived from the unknown compounds group.

Statistics

SVM accuracies of strips exposed to a drug condition were compared with those of the non-cardioactive benchmark by using Student's t test (desired α value of 0.05) with a Bonferroni correction (m, number of tests or hypotheses, was dependent on the number of drug additions in a screen). If the adjusted p value was statistically significant, the drug condition was considered to have incited irregular behavior in hvCTSs and was labeled as cardioactive. The Bonferroni correction was also applied when examining changes in specific parameters.

To analyze the performance of the multiclass models, confusion matrices were generated for each of the 50 runs. In a

confusion matrix, M, the precision and recall rates were defined as follows:

$$Precision_i = \frac{M_{ii}}{\sum_{j=1}^{n} M_{ij}}, \quad Recall_j = \frac{M_{ii}}{\sum_{i=1}^{n} M_{ij}}.$$

The precision and recall rates were calculated for each of the classifiers. To further summarize these metrics, the F_1 score, the harmonic mean of precision and recall, was computed and defined as follows:

$$F_1\ score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i}.$$

A model that is perfect would achieve an F_1 score of 1. If a model were composed of s number of classes and had random classifiers, the expected F_1score would be $1/s$. To assess the model as a whole, accuracy, defined as

$$Accuracy_{model} = \frac{\sum_{i=1}^{n} M_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij}}.$$

was calculated. In summarizing the 50 runs of each model, all calculated metrics were averaged and a confusion matrix containing the average number of contractile events over all runs was provided. Contractile events were defined as the individual twitches in the force readouts acquired from the tissue strips. All reported sample sizes (n) refer to independent tissue strips (biological replicates). All descriptive statistics are in the format of mean ± SD.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, one table, and one data file and can be found with this article online at https://doi.org/10.1016/j.stemcr.2017.09.008.

AUTHOR CONTRIBUTIONS

E.K.L. and D.D.T. initiated and conceived project. G.W., C.W.C., K.D.C., R.A.L., and M.K. supervised the project. W.K. and P.C. collected biological data. E.K.L. implemented the machine learning concepts. E.K.L., D.D.T., W.K., K.D.C., and M.K. analyzed data. E.K.L. and D.D.T. wrote the manuscript. W.K., K.D.C., R.A.L., and M.K. contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

Funding for this research was in part by Novoheart LTD. R.L., K.D.C., and M.K. have equity interest in Novoheart Ltd, a company that may potentially benefit from the research results. During the period between initial manuscript submission and acceptance, E.K.L. has become an employee at Novoheart Ltd. D.D.T., G.W., and C.W.C. are employees of Novoheart Ltd. The terms of this arrangement have been reviewed and approved by the University of California in accordance with its conflict of interest policies. This work was also supported by the Innovation Technology Fund, HKSAR (ITS/131/13FX). We would like to thank Dr. Bernard Fermini for providing the compounds with known inotropic classification.

Received: June 1, 2017
Revised: September 11, 2017
Accepted: September 12, 2017
Published: October 12, 2017

REFERENCES

Chen, A., Lee, E., Tu, R., Santiago, K., Grosberg, A., Fowlkes, C., and Khine, M. (2014). Integrated platform for functional monitoring of biomimetic heart sheets derived from human pluripotent stem cells. Biomaterials *35*, 675–683.

Dempsey, G.T., Chaudhary, K.W., Atwater, N., Nguyen, C., Brown, B.S., McNeish, J.D., Cohen, A.E., and Kralj, J.M. (2016). Cardiotoxicity screening with simultaneous optogenetic pacing, voltage imaging and calcium imaging. J. Pharmacol. Toxicol. Methods *81*, 240–250.

Dick, E., Rajamohan, D., Ronksley, J., and Denning, C. (2010). Evaluating the utility of cardiomyocytes from human pluripotent stem cells for drug screening. Biochem. Soc. Trans. *38*, 1037–1045.

Dietterich, T.G. (1995). Solving multiclass learning problems via error-correcting output codes. J. Artif. Intell. Res. *2*, 263–286.

Eng, G., Lee, B.W., Protas, L., Gagliardi, M., Brown, K., Kass, R.S., Keller, G., Robinson, R.B., and Vunjak-novakovic, G. (2016). Autonomous beating rate adaptation in human stem cell-derived cardiomyocytes. Nat. Commun. *7*, 1–10.

FDA. (2005). Guidance for Industry: S7B Nonclinical Evaluation of by Human Pharmaceuticals Guidance for Industry (U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER)). https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm074963.pdf.

FDA (2016). Zelnorm (tegaserod maleate) Information. https://www.fda.gov/Drugs/DrugSafety/ucm103223.htm.

Ferriman, A. (2000). UK licence for cisapride suspended cancer drug may cause heart failure. Br. Med. J. *321*, 2000.

Guo, L., Qian, J., Abrams, R., Tang, H., Weiser, T., Sanders, M.J., and Kolaja, K.L. (2011). The electrophysiological effects of cardiac glycosides in cardiomyocytes and in guinea pig isolated hearts. Cell. Physiol. Biochem. *27*, 453–462.

Guo, L., Coyle, L., Abrams, R.M.C., Kemper, R., Chiao, E.T., and Kolaja, K.L. (2013). Refining the human iPSC-cardiomyocyte arrhythmic risk assessment model. Toxicol. Sci. *136*, 581–594.

Harmer, A.R., Abi-Gerges, N., Morton, M.J., Pullen, G.F., Valentin, J.P., and Pollard, C.E. (2012). Validation of an in vitro contractility assay using canine ventricular myocytes. Toxicol. Appl. Pharmacol. *260*, 162–172.

Harris, G., and Koli, E. (2005). Lucrative Drug, Danger Signals and the F.D.A (New York Times), pp. 1–9.

Harris, K., Aylott, M., Cui, Y., Louttit, J.B., McMahon, N.C., and Sridhar, A. (2013). Comparison of electrophysiological data from human-induced pluripotent stem cell-derived cardiomyocytes to functional preclinical safety assays. Toxicol. Sci. *134*, 412–426.

Huebsch, N., Loskill, P., Deveshwar, N., Spencer, C.I., Judge, L.M., Mandegar, M.A., Fox, C.B., Mohamed, T.M.A., Ma, Z., Mathur, A.,

et al. (2016). Miniaturized iPS-cell-derived cardiac muscles for physiologically relevant drug response analyses. Sci. Rep. *6*, 24726.

Katz, A., Lifshitz, Y., Bab-Dinitz, E., Kapri-Pardes, E., Goldshleger, R., Tal, D.M., and Karlish, S.J.D. (2010). Selectivity of digitalis glycosides for isoforms of human Na,K-ATPase. J. Biol. Chem. *285*, 19582–19592.

Lee, E.J., Kim, D.E., Azeloglu, E.U., and Costa, K.D. (2008). Engineered cardiac organoid chambers: toward a functional biological model ventricle. Tissue Eng. Part A *14*, 215–225.

Lee, E.K., Kurokawa, Y.K., Tu, R., George, S.C., and Khine, M. (2015). Machine learning plus optical flow: a simple and sensitive method to detect cardioactive drugs. Sci. Rep. *5*, 11817.

Li, X., Zhang, R., Zhao, B., Lossin, C., and Cao, Z. (2016). Cardiotoxicity screening: a review of rapid-throughput in vitro approaches. Arch. Toxicol. *90*, 1803–1816.

Liu, J., Lieu, D.K., Siu, C.W., Fu, J.D., Tse, H.F., and Li, R.A. (2009). Facilitated maturation of Ca 2 + handling properties of human embryonic stem cell-derived cardiomyocytes by calsequestrin expression. Am. J. Physiol. Cell Physiol. *297*, 152–159.

Lu, H.R., Whittaker, R., Price, J.H., Vega, R., Pfeiffer, E.R., Cerignoli, F., Towart, R., and Gallacher, D.J. (2015). High throughput measurement of Ca++ dynamics in human stem cell-derived cardiomyocytes by kinetic image cytometery: a cardiac risk assessment characterization using a large panel of cardioactive and inactive compounds. Toxicol. Sci. *148*, 503–516.

Lundy, S.D., Zhu, W.-Z., Regnier, M., and Laflamme, M.A. (2013). Structural and functional maturation of cardiomyocytes derived from human pluripotent stem cells. Stem Cells Dev. *22*, 1991–2002.

Maddah, M., Heidmann, J.D., Mandegar, M.A., Walker, C.D., Bolouki, S., Conklin, B.R., and Loewke, K.E. (2015). A non-invasive platform for functional characterization of stem cell-derived cardiomyocytes with applications in cardiotoxicity testing. Stem Cell Reports *4*, 621–631.

Martin, R.L., Lee, J., Cribbs, L.L., Perez-reyes, E., and Hanck, D.A. (2000). Mibefradil block of cloned T-type calcium channels 1. J. Pharmacol. Exp. Ther. *295*, 302–308.

Millard, D.C., Strock, C.J., Carlson, C.B., Aoyama, N., Juhasz, K., Goetze, T.A., Stoelzle-Feix, S., Becker, N., Fertig, N., January, C.T., et al. (2016). Identification of drug–drug interactions *in vitro*: a case study evaluating the effects of sofosbuvir and amiodarone on hiPSC-derived cardiomyocytes. Toxicol. Sci. *154*, 174–182.

Navarrete, E.G., Liang, P., Lan, F., Sanchez-Freire, V., Simmons, C., Gong, T., Sharma, a., Burridge, P.W., Patlolla, B., Lee, A.S., et al. (2013). Screening drug-induced arrhythmia events using human induced pluripotent stem cell-derived cardiomyocytes and low-impedance microelectrode arrays. Circulation *128*, S3–S13.

Pillekamp, F., Haustein, M., Khalil, M., Emmelheinz, M., Nazzal, R., Adelmann, R., Nguemo, F., Rubenchyk, O., Pfannkuche, K., Matzkies, M., et al. (2012). Contractile properties of early human embryonic stem cell-derived cardiomyocytes: beta-adrenergic stimulation induces positive chronotropy and lusitropy but not inotropy. Stem Cells Dev. *21*, 2111–2121.

Pointon, A., Pilling, J., Dorval, T., Wang, Y., Archer, C., and Pollard, C. (2016). High-throughput imaging of cardiac microtissues for the assessment of cardiac contraction during drug discovery. Toxicol. Sci. *155*, 444–457.

Ravenscroft, S.M., Pointon, A., Williams, A.W., Cross, M.J., and Sidaway, J.E. (2016). Cardiac non-myocyte cells show enhanced pharmacological function suggestive of contractile maturity in stem cell derived cardiomyocyte microtissues. Toxicol. Sci. *152*, 99–112.

Scott, C.W., Zhang, X., Abi-Gerges, N., Lamore, S.D., Abassi, Y.A., and Peters, M.F. (2014). An impedance-based cellular assay using human iPSC-derived cardiomyocytes to quantify modulators of cardiac contractility. Toxicol. Sci. *142*, 331–338.

Serrao, G.W., Turnbull, I.C., Ancukiewicz, D., Kim, D.E., Kao, E., Cashman, T.J., Hadri, L., Hajjar, R.J., and Costa, K.D. (2012). Myocyte-depleted engineered cardiac tissues support therapeutic potential of mesenchymal stem cells. Tissue Eng. Part A *18*, 1322–1333.

Sirenko, O., Crittenden, C., Callamaras, N., Hesley, J., Chen, Y.-W., Funes, C., Rusyn, I., Anson, B., and Cromwell, E.F. (2013). Multiparameter in vitro assessment of compound effects on cardiomyocyte physiology using iPSC cells. J. Biomol. Screen. *18*, 39–53.

Steinberg, S.F. (1999). The molecular basis for distinct B-adrenergic receptor subtype actions in cardiomyocytes. Circ. Res. *85*, 1101–1111.

Turnbull, I.C., Karakikes, I., Serrao, G.W., Backeris, P., Lee, J., Xie, C., Senyei, G., Gordon, R.E., Li, R.A., Akar, F.G., et al. (2014). Advancing functional engineered cardiac tissues toward a preclinical model of human myocardium. FASEB J. *28*, 644–654.

Weng, Z., Kong, C.-W., Ren, L., Karakikes, I., Geng, L., He, J., Chow, M.Z.Y., Mok, C.F., Chan, H.Y.S., Webb, S.E., et al. (2014). A simple, cost-effective but highly efficient system for deriving ventricular cardiomyocytes from human pluripotent stem cells. Stem Cells Dev. *23*, 1704–1716.

Williams, G.H. (1988). Converting-enzyme inhibitors in the treatment of hypertension. N. Engl. J. Med. *319*, 1517–1525.

Wong, B.S., Manabe, N., and Camilleri, M. (2010). Role of prucalopride, a serotonin (5-HT4) receptor agonist, for the treatment of chronic constipation. Clin. Exp. Gastroenterol. *3*, 49–56.

Yang, X., Pabon, L., and Murry, C.E. (2014). Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. Circ. Res. *114*, 549–561.

Zhang, D., Shardin, I., Lam, J., Xian, H.-Q., Snodgrass, R., and Bursac, N. (2014). Tissue-engineered cardiac patch for advanced functional maturation of human ESC-derived cardiomyocytes. Biomaterials *34*, 5813–5820.

Ziupa, D., Beck, J., Franke, G., Feliz, S.P., Hartmann, M., Koren, G., Zehender, M., Bode, C., Brunner, M., and Odening, K.E. (2014). Pronounced effects of HERG-blockers E-4031 and erythromycin on APD, spatial APD dispersion and triangulation in transgenic long-QT type 1 rabbits. PLoS One *9*, e107210.

Stem Cell Reports, Volume *9*

Supplemental Information

Machine Learning of Human Pluripotent Stem Cell-Derived Engineered Cardiac Tissue Contractility for Automated Drug Classification

Eugene K. Lee, David D. Tran, Wendy Keung, Patrick Chan, Gabriel Wong, Camie W. Chan, Kevin D. Costa, Ronald A. Li, and Michelle Khine

Supplemental Figures

Examined Parameters

1. Desired pacing frequency	7. Area under the curve of decline from max force to 95% cutoff	13. Area under the curve of decline from max force to 50% cutoff
2. Captured pacing frequency	8. Max change of force over time ($\Delta F/\Delta t$) of contraction phase	14. Duration of rise from 25% cutoff to max force
3. Max force generated (Amplitude)	9. Max change of force over time ($\Delta F/\Delta t$) of relaxation phase	15. Duration of decline from max force to 25% cutoff
4. Duration of rise from 95% cutoff to max force (contraction phase)	10. Duration of rise from 50% cutoff to max force	16. Area under the curve of rise from 50% cutoff to max force
5. Duration of decline from max force to 95% cutoff (relaxation phase)	11. Duration of decline from max force to 50% cutoff	17. Area under the curve of decline from max force to 50% cutoff
6. Area under the curve of rise from 95% cutoff to max force	12. Area under the curve of rise from 50% cutoff to max force	

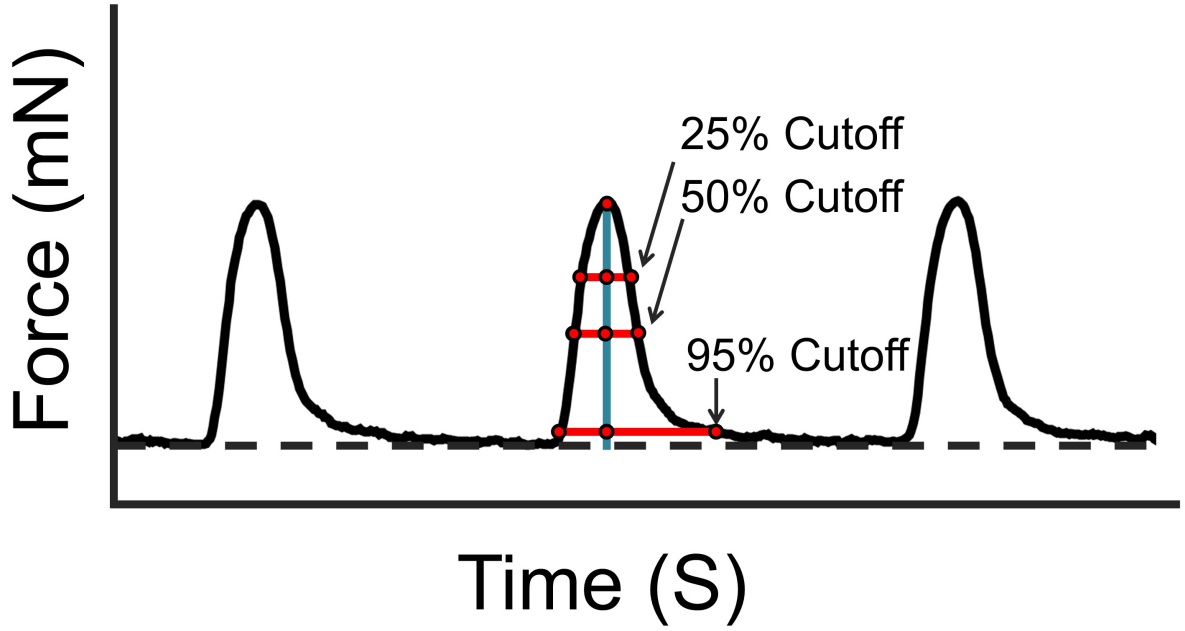

Figure S1 related to Figure 1: Parameters derived from recorded force waveforms (Top panel) List of the 17 parameters input into binary SVM. The desired pacing frequency parameter is an experimental stimulation parameter. All other parameters are extracted from recorded force waveforms. (Bottom Panel) Three cutoff points (25, 50, and 95% of the max force generated) were calculated to summarize changes to the overall shape of the force waveform. Blue line indicates the time at which max force is generated.

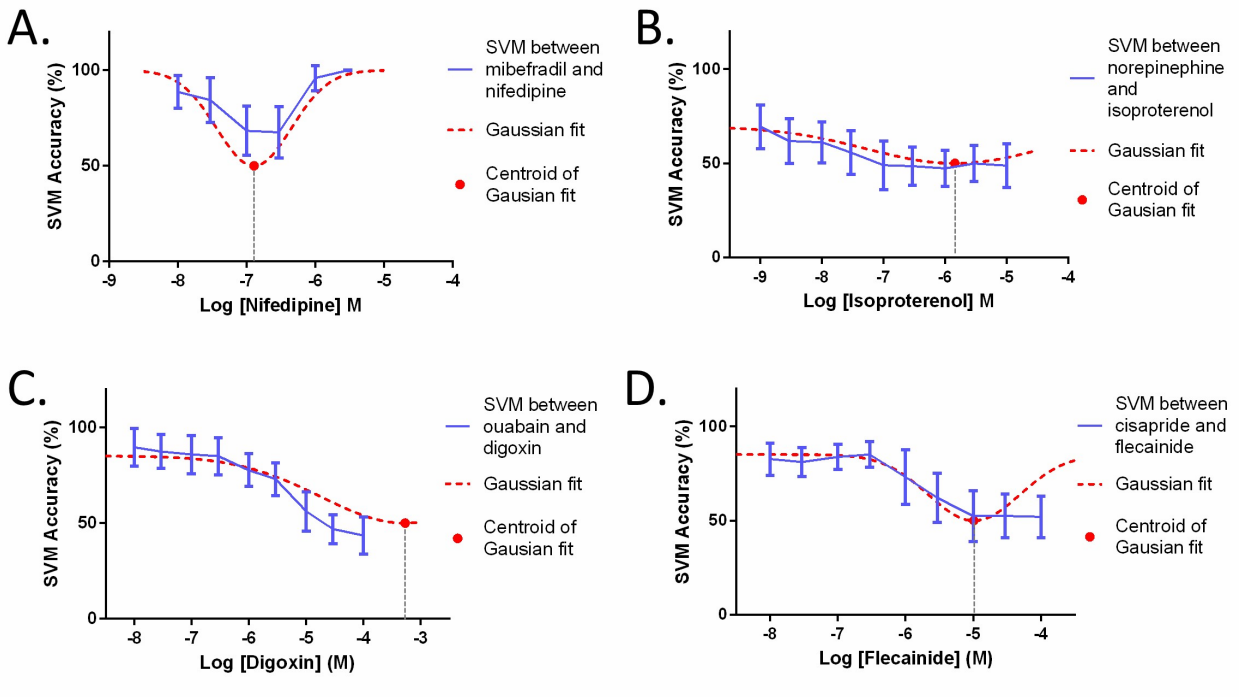

Figure S2 related to Figure 5: Drug response relationships of 'unknown' compounds to compounds representing predicted classes using a binary SVM approach. Solid blue line indicates the binary SVM accuracy between a concentration of the 'unknown' compound and each tested concentration of the representative compound. Dotted red line indicates a fitted Gaussian curve where the centroid represents the concentration at which the representative compound elicits the most similar response in the hPSC-CMs as the condition of the 'unknown' compound. (A) Ca^{2+} blockers: mibefradil (n = 6) and nifedipine (n = 10). (B) Adrenergic agonist: norepinephrine (n = 8) and isoproterenol (n = 10). (C) Cardiac glycoside: ouabain (n = 10) and digoxin (n = 9). (D) hERG K^+ channel blocker: cisapride (n = 9) and flecainide (n = 8). n refers to independent biological replicates. All results are presented as mean ± standard deviation.

Compound (M)	Predicted Class	Estimated Similar Concentration (M)
Mibefradil ($3.0x10^{-6}$)	Ca^{2+} channel blocker	Nifedipine ($1.28x10^{-7}$)
Norepinephrine ($1.0x10^{-5}$)	Adrenergic agonist	Isoproterenol ($1.44x10^{-6}$)
Ouabain ($1.0x10^{-5}$)	Cardiac glycoside	Digoxin ($5.35x10^{-5}$)
Cisapride ($1.0x10^{-4}$)	hERG K^+ channel blocker	Flecainide ($1.03x10^{-5}$)

Table S1 related to Figure 5: Summary of estimated drug response relationships of all four classes.

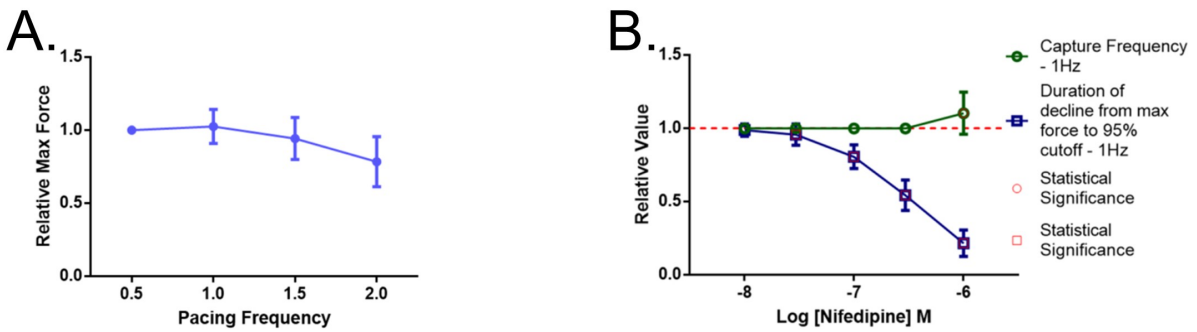

Figure S3 related to Figure 5: Confirmation of nifedipine's inotropic effects. (A) Force-frequency relationship of vehicle-treated hvCTSs (n = 28). (B) Nifedipine-treated hvCTSs had capturing frequencies that matched the pacing frequency of 1 Hz from 1.0×10^{-8} to 3.0×10^{-7} M. During this range, the max force generated decreased by 45.69 ± 10.42%. Since the force-frequency relationship was decoupled by pacing the strips, it can be concluded that the decrease in force is attributed to the compound's negative inotropic effects ($p < 0.0063$; n = 10). n refers to independent biological replicates. All results are presented as mean ± standard deviation.

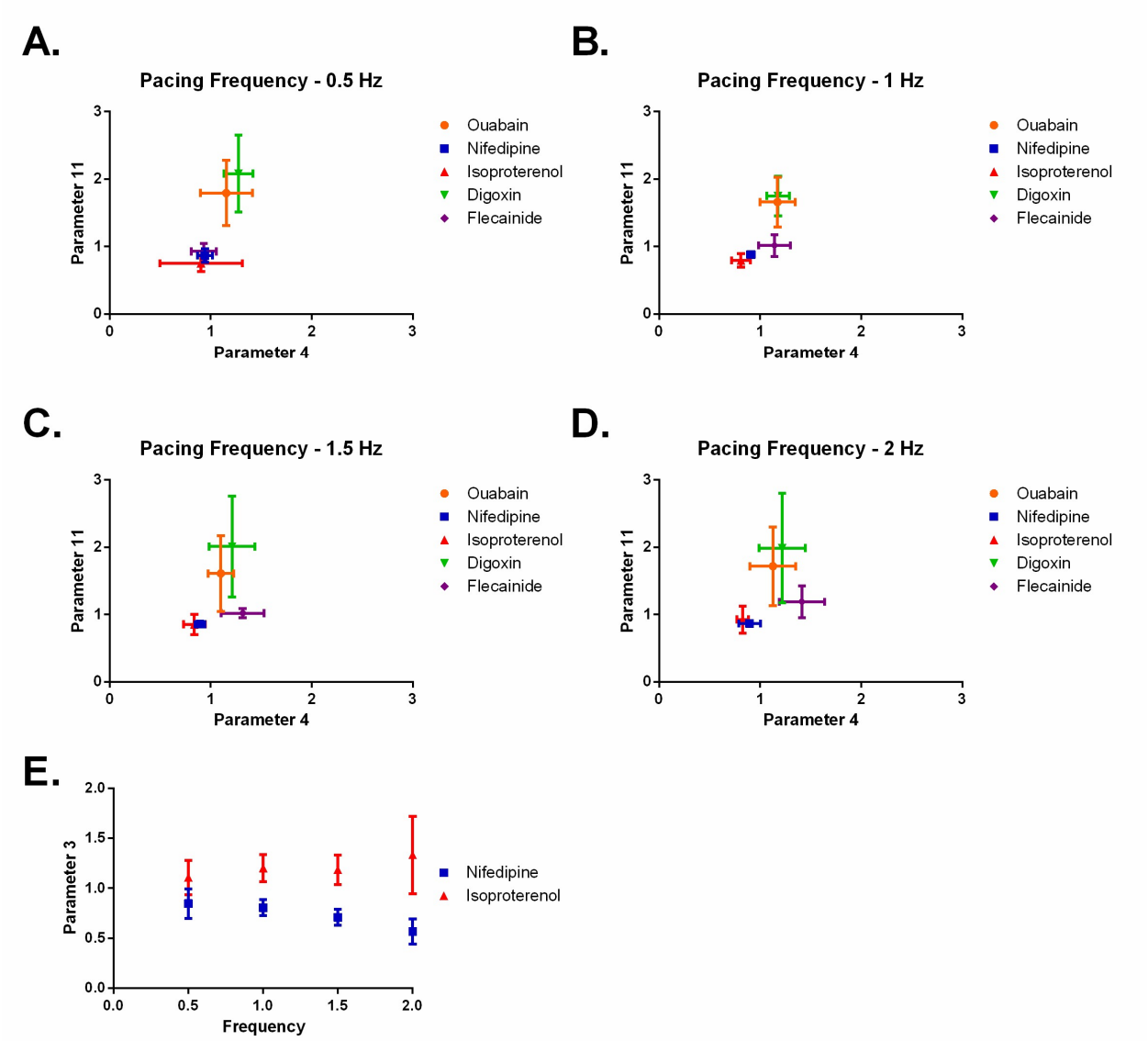

Figure S4 related to Figure 5: Confirmation of multi-class model predictive performance on an individual parameter basis. The mean and standard deviation of parameter 4 (duration of rise from 95% cutoff to max force) and parameter 11 (duration of decline from max force to 50% cutoff) were plotted for test compound ouabain and the drug library compounds (nifedipine, isoproterenol, digoxin, and flecainide) at the relevant concentration used in the multiclass model at paced rate of 0.5 Hz (A), 1.0 Hz (B), 1.5 Hz (C), and 2.0 Hz (D). In this given parameter space, it is visibly evident that ouabain (orange), a cardiac glycoside, is most similar to digoxin, the compound defining the cardiac glycoside family, for all pacing frequencies. Although nifedipine (Ca^{2+} channel blocker) and isoproterenol (beta-adrenergic agonist) occupy a similar space when evaluating Parameter 4 and 11 (A-D), the two drug families are distinguishable from each other when analyzing the Frequency-Parameter 3 (max force generated) relationship (E). For all frequencies, ouabain: *n*=10; nifedipine: *n*=10; isoproterenol: *n*=9; digoxin: *n*=9; flecainide: *n*=7, except at 2 Hz, where isoproterenol: *n*=8 and digoxin *n*=8. *n* refers to independent biological replicates. All results are presented as mean ± standard deviation.

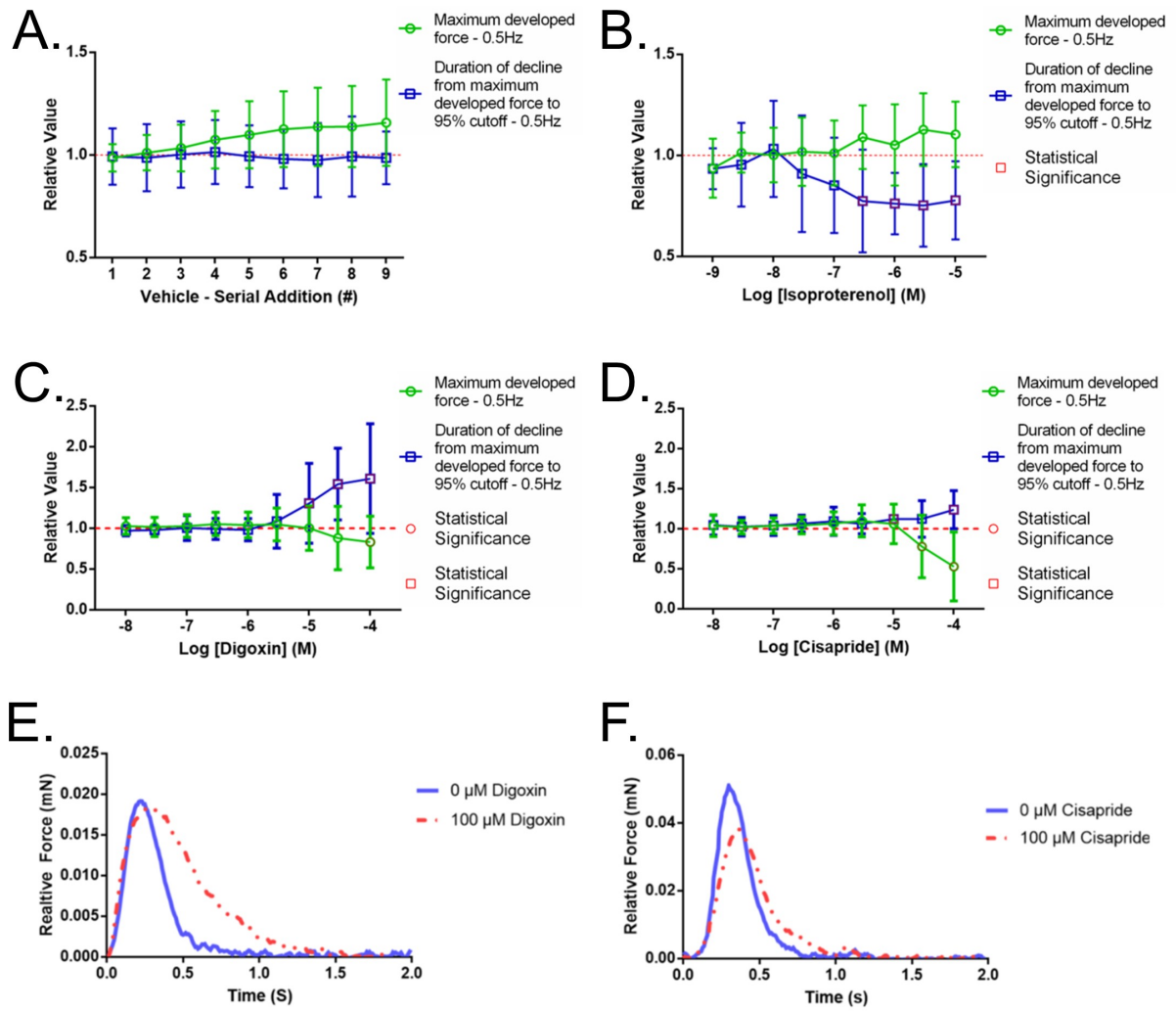

Figure S5 related to Figure 5: Examination of various compounds' cardioactive effects (A) In the vehicle study, hvCTSs paced at 0.5 Hz continually increased in maximum developed force through subsequent serial additions. The duration of decline from max force to 95% cutoff, relaxation phase, was unaffected by the number of serial additions (n = 28). (B) Isoproterenol-treated hvCTSs experienced an increase in max force similar to the vehicle-treated strips, suggesting no inotropic effects. Isoproterenol's lusitropic effects were apparent as duration of the relaxation phase decreased significantly versus the vehicle study over a concentration range of 3.0×10^{-7} to 10^{-5} M ($p \leq 0.0055$; n = 10). (C) Digoxin-treated hvCTSs (paced at 0.5 Hz) exhibited a decrease in max force and an increase in duration of relaxation phase by the highest concentration, 10^{-4} M ($p \leq 0.0055$; n = 9). (D) Cisapride-treated hvCTSs (paced at 0.5 Hz) exhibited similar trends in max force and duration of relaxation phase by the highest concentration, 100 µM ($p < 0.0055$; n = 9). (E) Representative force tracing of a contractile event from a digoxin-treated hvCTS (paced at 0.5 Hz) shows the distinct cardioactive effects prolonging the duration while increasing the area under the curve of the relaxation phase. (F) Representative force tracing of a contractile event from a 100 µM cisapride-treated hvCTS (paced at 0.5 Hz) shows that while cisapride and digoxin have similar trends in certain parameters, the cardioactive effects of the two are visibly distinguishable. All listed p-values are adjusted with a Bonferroni correction. n refers to independent biological replicates. All results are presented as mean ± standard deviation.

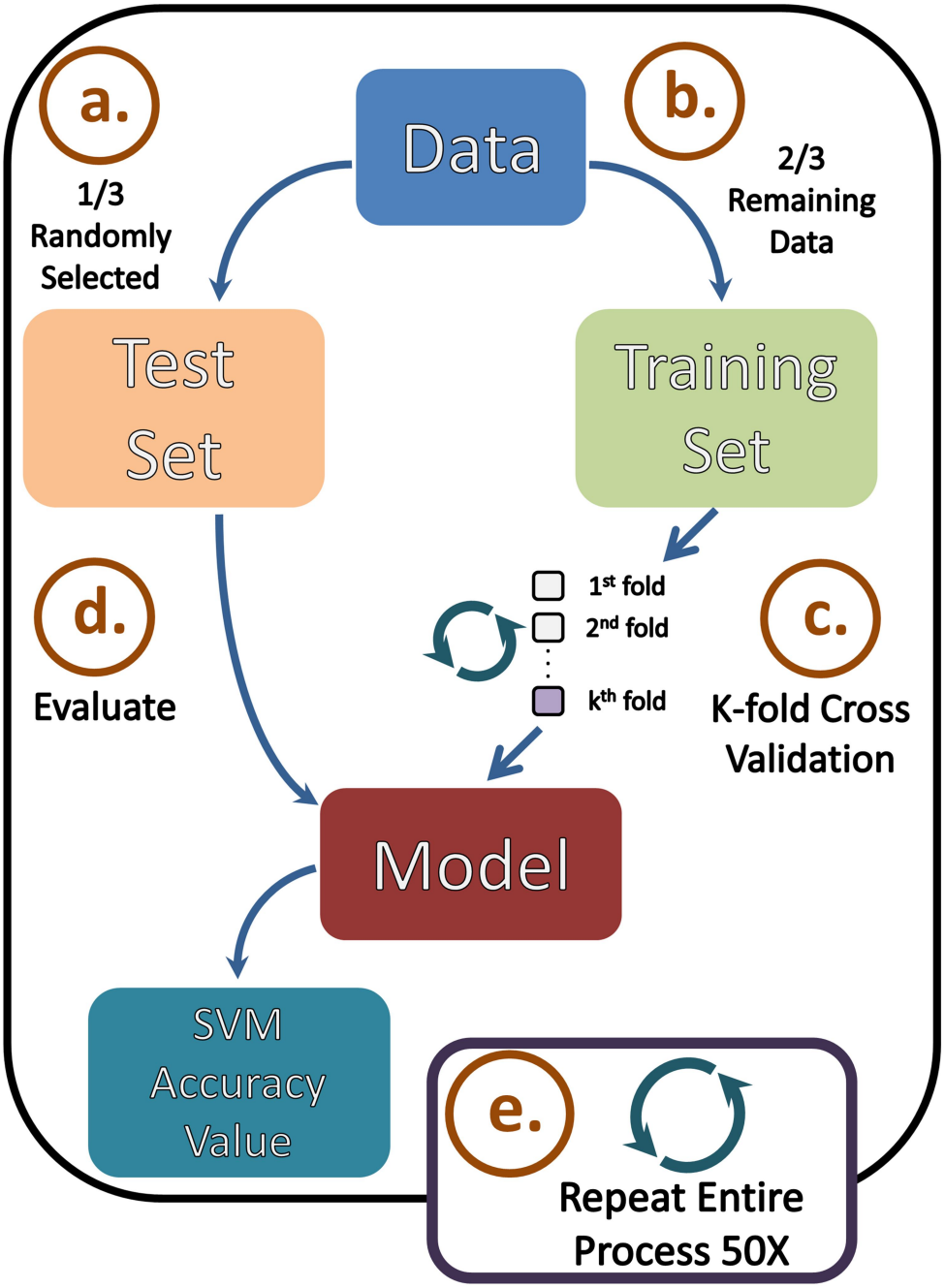

Figure S6 related to Figure 1: Process flow of how SVM accuracy value is generated. (A) Prior to any model training or formation, one third of the data is randomly withheld as the test set. (B) The remaining data is referred to as the training set and used to create a model. (C) To prevent overfitting, a K-fold cross validation is performed. (D) The model is than evaluated by having it label the data of the withheld test set. This evaluation leads to a SVM accuracy value. (E) As the data is randomly selected for the test set, the process is repeated 50 times to account for any variation.

Supplemental Experimental Procedures

Class relationship metrics

The concentration of a library compound that induced the most similar cardioactive effects as the compound of interest was determined. This class relationship metric was computed by first selecting the compound of interest at a desired concentration and performing a series of binary SVM among the tested range of a library compound. For each concentration of the library compound, the closer the SVM accuracy was to 50%, the more defined boundaries of the compounds overlapped and the more similar the cardioactive effects were. This relationship between SVM accuracy and tested concentration range was presumed to behave in a Gaussian manner with the centroid representing the concentration that would elicit the most similar effects. The Gaussian fit was set with 50% as the lower limit and the highest achieved SVM accuracy as the upper limit. If the original SVM accuracies reached the 50% mark and remained around this value for subsequent concentrations, only the first concentration to reach the 50% was included in the fit to accurately model one side of the Gaussian curve.

Optimization of Binary and Multi-class SVM classifiers

In this study, the machine learning techniques employed were predominantly based on Support Vector Machines (SVM). SVM can be described as an algorithm that defines the boundaries of two or more groups by maximizing the distance between such boundaries (Noble, 2006). As biological data like most real world data cannot be linearly separated, a radial basis function (RBF) kernel, a non-linear kernel, was implemented for the binary SVM classifiers (Tarca et al., 2007). Furthermore, SVM, in general, has parameters that influence the performance of the models and thus, need to be tuned. A cross validation approach can be used to estimate the prediction error of a model with a certain set of parameters. The optimal combination of parameter values can be determined by choosing the model that minimizes the error estimate. However that model's error is not a good estimate of the true error that is calculated with an independent dataset. Rather a nested cross validation approach, an additional cross validation step wrapped around the original cross validation step used to tune parameters, should be employed as this has been shown to give a better unbiased estimate of the true error (Varma and Simon, 2006).

In this study, a nested cross validation approach was used with the outer cross validation being a Monte Carlo cross validation (also referred to as repeated random subsampling) (Simon, 2007). The hvCTS data was allocated with approximately one third representing the test set and the remainder serving as the training set. Within each run, the random partitioning of the data was done on the hvCTS level. For example, if a tissue strip were randomly assigned to the test set, all of the individual contractile events recorded from the tissue strip were included in that test set. The approximate 2 to 1 ratio between the sizes of the training and test sets was chosen such that on average, data from at least two tissue strips would be used in the evaluation process. Furthermore, the number of contractile events among each tissue strip measurement was fairly consistent and within the same magnitude as the tissue strips were paced. At each serial addition or dosing of tissue strip, a range of 130 to 170 tissue twitches were typically present. A total of 50 SVM runs were performed for each concentration to account for the variation and random selection of datasets. We maintained a balanced number between the vehicle-treated strips and those exposed to a cardioactive compound of the model (n = 6, 7, 8, 9, or 10). Since the number of vehicle strips (n = 28) always outweighed those treated with drugs, we randomly selected a subset of the vehicle-treated tissue strips that equaled the sample size for each SVM run. As the performance of a SVM classifier with a RBF kernel is dependent on the values of the box constraint and sigma parameter, a geometric progression approach or grid search was used to determine the hyperparameters, the optimal combination of the two parameter values (Ben-hur and Weston, 2009). To prevent overfitting, we performed an inner cross validation, a 5-fold cross validation, within the training set for this tuning task (Simon, 2007). The initial search ranged from 10^{-5} to 10^{6} in 10-fold increments for both parameters. Once the optimal set of parameter values was identified, another search was implemented with 5 smaller intervals in both directions. For example, using this approach, the average sigma and box constraint parameter for Condition 3 (flecainide and E-4031) were 2.61 ± 0.91 and $4.67 \times 10^{4} \pm 1.16 \times 10^{4}$. Thus for a given binary SVM run, the distribution of the data could be estimated as 53.33% for training set, 13.33% for validation set, and 33.33% for test set. It should be noted that if more than half of the tissue strips became unresponsive to electrical stimulation after being given increasing drug concentrations, the SVM value for that condition was automatically designated as 100% accuracy (i.e., no uncertainty that the drug is cardioactive), and binary SVM was not performed. Under such conditions, it was unnecessary to use SVM or any machine learning to differentiate between such drastic states (i.e., beating vs. non-beating). For example, the drug nifedipine at 10 uM induced 7 of 10 hvCTSs to cease contraction, and was therefore designated cardioactive (SVM value of 100%) at that concentration.

For the multi-class models, a criterion of 85% binary SVM accuracy was used to determine the specific concentration of a compound that would be included in the library. This criterion was chosen as it was as a reference

point where the cardioactive effects of a compound would be prominent, but can still define generalizable boundaries from those of other compounds. The value of 85% was specifically chosen as it was approximately the midpoint between the maximum achievable separation (100%) and a minimum bound that would ensure cardioactivity. We defined the minimum bound as the largest sum of mean SVM accuracy and one standard deviation across all vehicle studies, resulting in a bound of 69.34% (mean SVM accuracy of 53.45% and standard deviation of 15.89%). The criterion of at least 6 responsive tissue strips was to ensure that within the test sets there were data from at least two strips for all runs.

For the creation and optimization of the multi-class models, a one-vs.-one strategy with binary SVM learners was used. In a one-vs.-one strategy, a binary classification is performed for each pair of classes (Rocha and Goldenstein, 2014). In our study, where the model consisted of four drug classes, a total of 6 binary learners would be used. An error-correcting output codes approach was used to summarize the results of the 6 classifiers. The same Monte Carlo cross validation approach with one third of the data partitioned into the test set was used. Similarly, the multi-class classification and prediction process was repeated a total of 50 times. The box constraint and sigma parameters were again tuned with a geometric progression approach; however, a 10-fold cross validation was performed within the training set as the inclusion of tissue strips exposed to compounds belonging to all 4 drug classes formed a larger dataset. Using a 10-fold cross validation, the distribution of the data could be estimated as 74.07% for training set, 6.66% for validation set, and 33.33% for test set. In terms of the size of these partitions, the tests sets (evaluation criteria for generalizability as seen in Fig 4) of the multi-class models averaged 1660.20 ± 204.94, 1707.64 ± 165.45, 1995.24 ± 183.94 contractile events for Condition 1, 2, and 3 respectively.

Supplemental References

Ben-hur, A., and Weston, J. (2009). A User's Guide to Support Vector Machines. In Methods in Molecular Biology: Data Mining Techniques for the Life Sciences, O. Carugo, and F. Eisenhaber, eds. (Springer Protocols), pp. 223–239.

Noble, W.S. (2006). What is a support vector machine? Nat. Biotechnol. *24*, 1565–1567.

Rocha, A., and Goldenstein, S.K. (2014). Multiclass from Binary: Expanding One-Versus-All, One-Versus-One and ECOC-Based Approaches. IEEE Trans. Neural Networks Learn. Syst. *25*, 289–302.

Simon, R. (2007). Resampling Strategies for Model Assessment and Selection. In Fundamentals of Data Mining in Genomics and Proteomics, W. Dubitzky, M. Granzow, and D.P. Berrar, eds. (Springer), pp. 178–179.

Tarca, A.L., Carey, V.J., Chen, X., Romero, R., and Dra, S. (2007). Machine Learning and Its Applications to Biology. PloS Comput. Biol. *3*.

Varma, S., and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics *7*.