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	

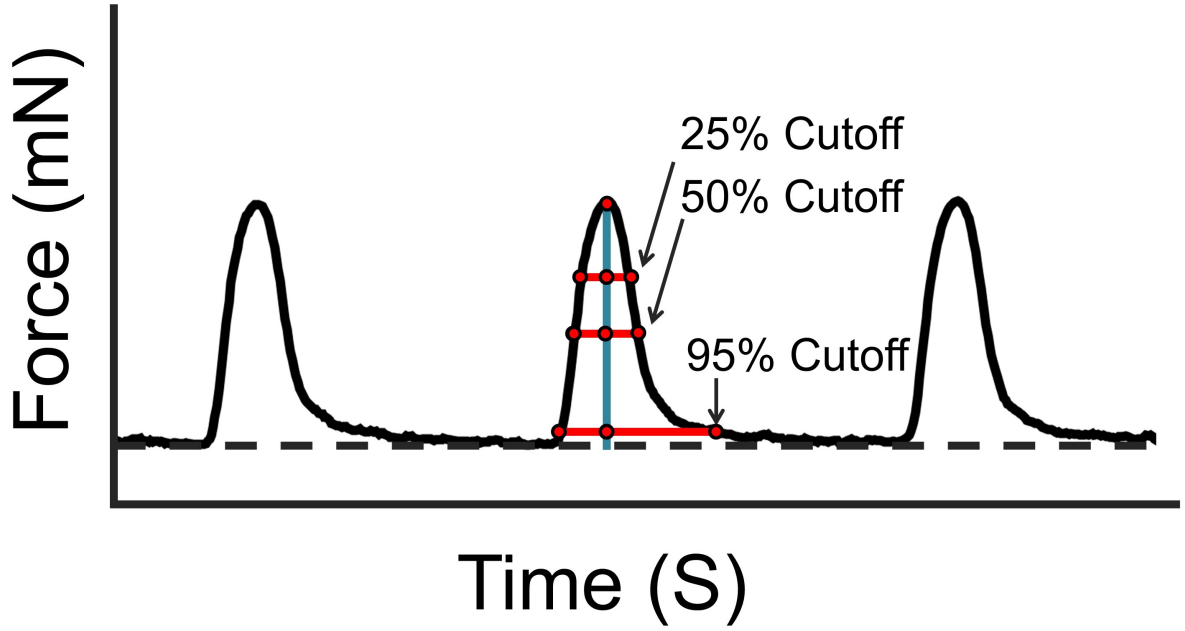

Figure S1 related to Figure 1: Parameters derived from recorded force waveforms (Top panel) List of the 17 parameters input into binary SVM. The desired pacing frequency parameter is an experimental stimulation parameter. All other parameters are extracted from recorded force waveforms. (Bottom Panel) 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.

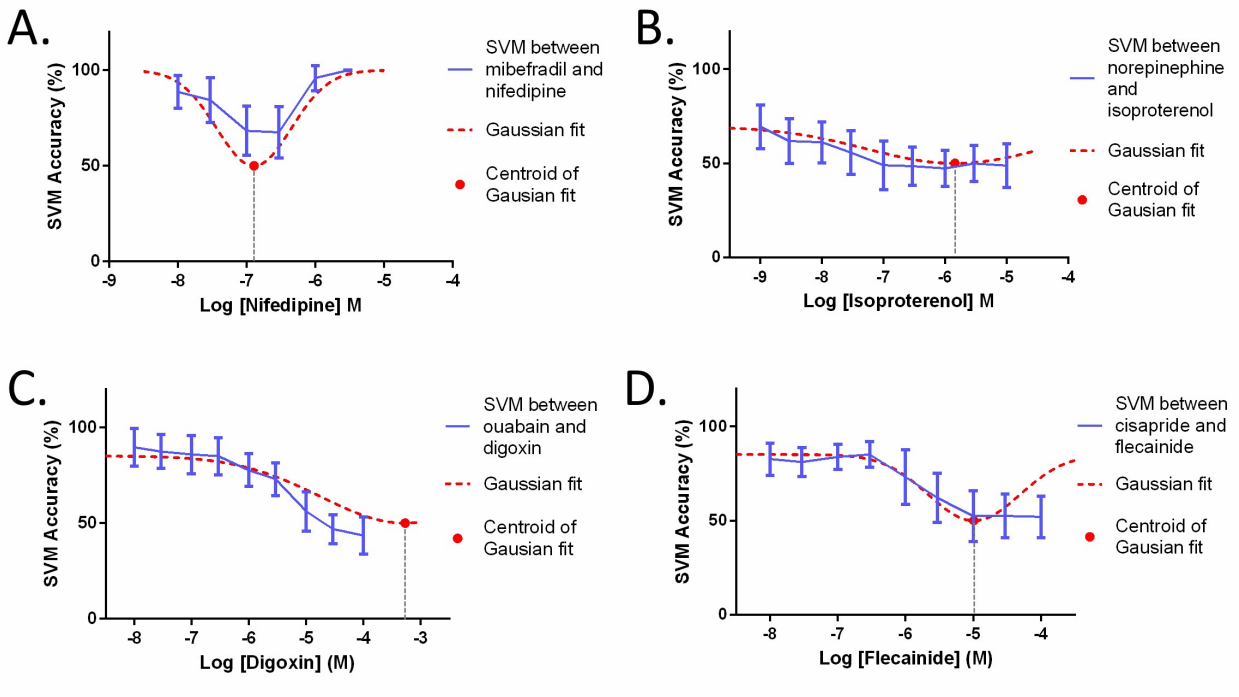

Figure S2 related to Figure 5: Drug response relationships of 'unknown' compounds to compounds representing predicted classes using a binary SVM approach. Solid blue line indicates the binary SVM accuracy between a concentration of the 'unknown' compound and each tested concentration of the representative 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.

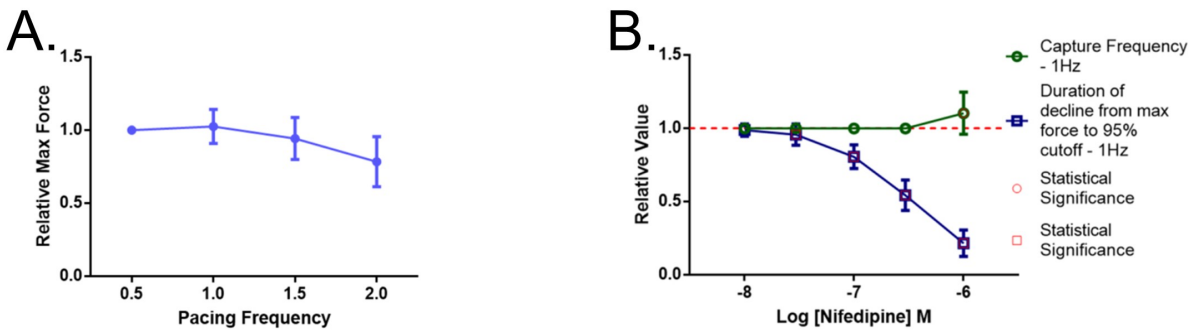

Figure S3 related to Figure 5: Confirmation of nifedipine's inotropic effects. (A) Force-frequency relationship of vehicle-treated hvCTSs (n = 28). (B) Nifedipine-treated hvCTSs had capturing frequencies that matched the pacing frequency of 1 Hz from 1.0×10^{-8} to 3.0×10^{-7} M. During this range, the 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.

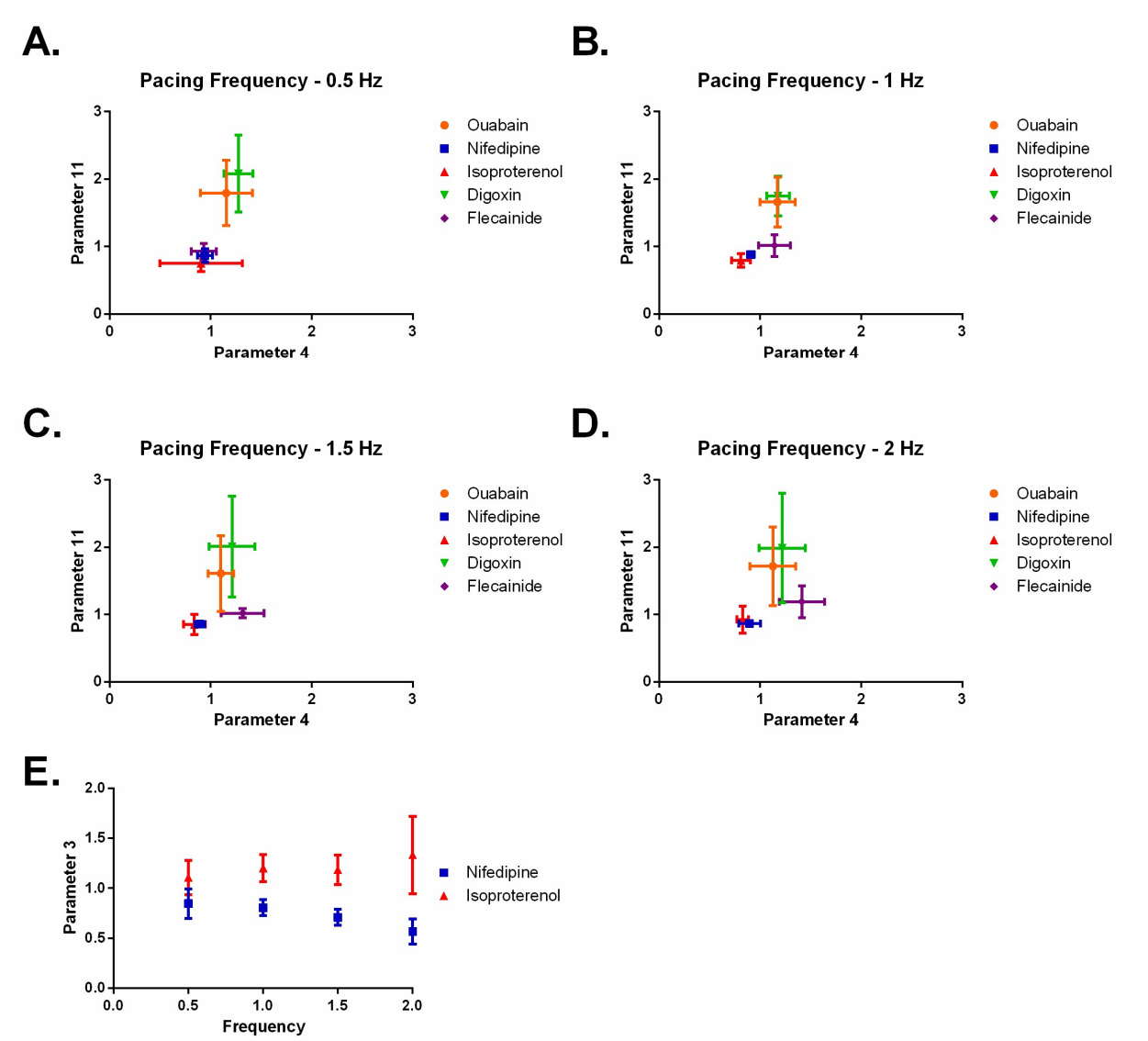

Figure S4 related to Figure 5: Confirmation of multi-class model predictive performance on an individual parameter basis. The mean and standard deviation of parameter 4 (duration of rise from 95% cutoff to max force) and parameter 11 (duration of decline from max force to 50% cutoff) were plotted for 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.

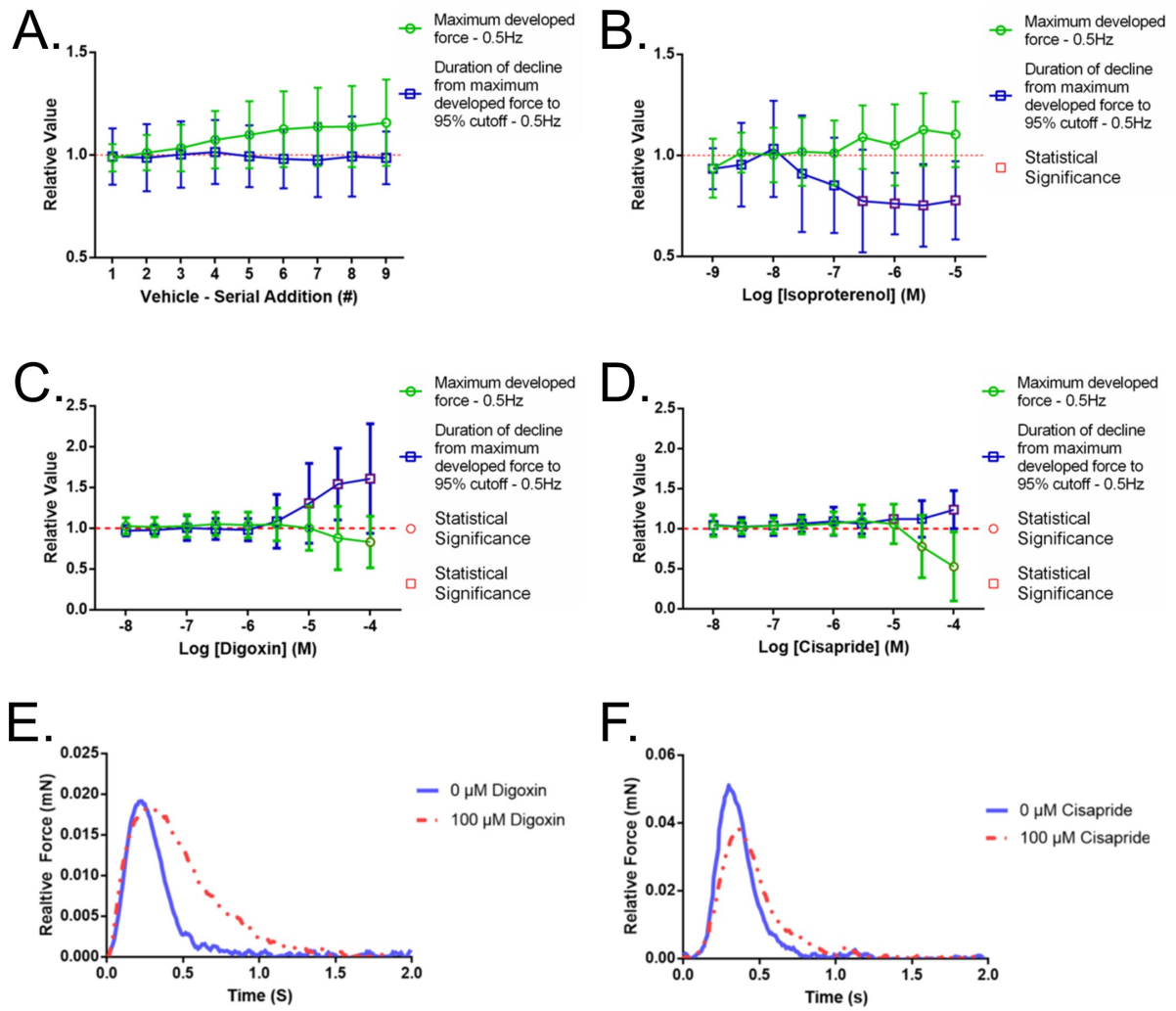

Figure S5 related to Figure 5: Examination of various compounds' cardioactive effects (A) In the vehicle study, hvCTSs paced at 0.5 Hz continually increased in maximum developed force through subsequent serial additions. The duration of decline from max force to 95% cutoff, relaxation phase, was unaffected 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.

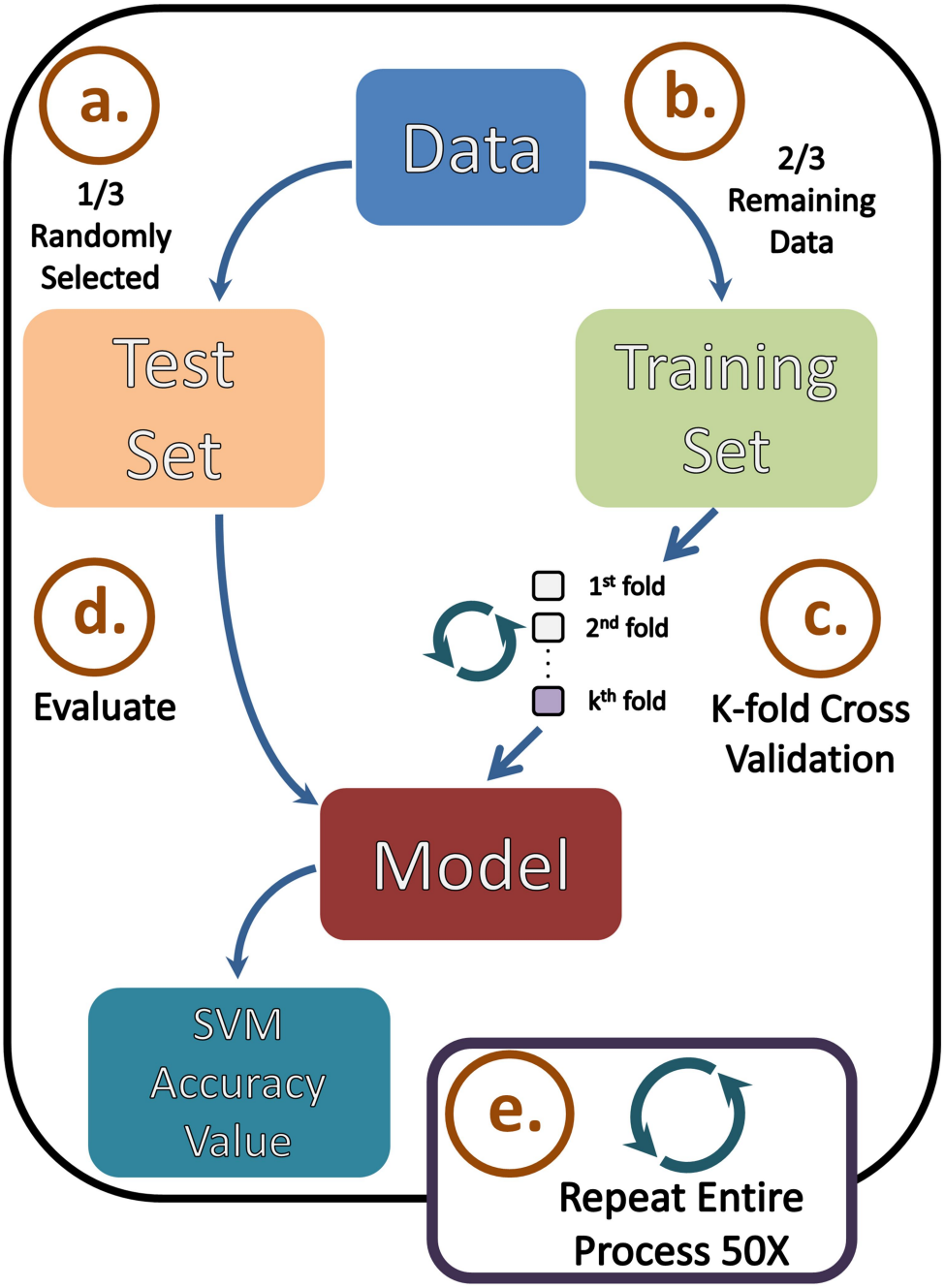

Figure S6 related to Figure 1: Process flow of how SVM accuracy value is generated. (A) Prior to any model training or formation, one third of the data is randomly withheld as the test set. (B) The remaining data is referred to as the training set and used to create a model. (C) To prevent overfitting, 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*.