Intrinsic structure of model-derived metrics for *in silico* proarrhytmic risk assessment identified by global sensitivity analysis

Jaimit Parikh[1], Paolo Di Achille[1], James Kozloski[1], Viatcheslav Gurev[1*]

1 IBM T. J. Watson Research Center, Yorktown Heights, NY, United States

* vgurev@us.ibm.com

Abstract

Multiscale computational models of heart are being extensively investigated for improved assessment of drug-induced Torsades de Pointes (TdP) risk, a fatal side effect of many drugs. Model-derived metrics (features) such as action potential duration, net charge carried by ionic currents ($qNet$) and others have been proposed in the past as potential candidates for classifying TdP risk. However, the criteria for selection of new risk metrics are still poorly justified, and they are often trained/tested only on small datasets. Moreover, classifiers built on derived features have thus far not consistently provided increased prediction accuracies compared to classifiers based on *in vitro* measurements of drug effects on ion channels (direct features). In this paper, we analyze a large population of virtual drugs to examine systematically the sensitivity of several model-derived features. The influence of different ion channels in regulation of the model-derived features is identified using global sensitivity analysis (GSA). Specifically, the analysis points to key differences in the input parameters that affect several model-derived features and the generation of early afterdepolarizations (EAD), thus opposing the idea that these features and sensitivity to EAD might be strongly correlated. We also demonstrate that previously proposed model-derived features could be well fitted by a linear combination of direct features. This well explains the observed comparable performances of classifiers built on direct features and model-derived features. Combining GSA and simple probability analysis, we also show that the odds of any linear metric constructed from direct features to perform as well as $qNet$ is very low. Nevertheless, despite high predictive power of $qNet$ to separate drugs into correct categories of TdP risk, the GSA results suggest that the actual mechanistic interpretation for $qNet's$ improved performance deserves further investigation. In conclusion, analyses like ours can provide more robust feature selection/construction. Improved experimental designs with increased focus on the critical model parameters indicated by GSA can potentially reduce the uncertainties of key model components and result in increased confidence of TdP risk predicted by *in silico* models.

Author Summary

Biophysical models often have extremely involved intrinsic structure. In the majority of research, either complex methods of non-linear dynamics and empirical analysis are employed to explore the underlying structure of cell processes such as transmembrane ionic currents. Global sensitivity analysis (GSA) could be considered as a brute force alternative to study the model relationships between physical processes, discovering the mechanisms responsible for phenomena of interest. As we demonstrated here, GSA application could be extended to explore the structure of features derived from outputs of

biophysical models and used in statistical models to build regressions or classifiers. In particular, GSA seems to be valuable to formalize the methods of feature selection/construction that are used for classification of drugs with respect to their cardiotoxicity.

Introduction

Drug-induced Torsades de Pointes (TdP) is a specific form of polymorphic ventricular tachycardia that leads to ventricular fibrillation and sudden cardiac death [1]. Several drugs have been withdrawn from the market in the past due to TdP risk [2]. Although the current clinical safety guidelines are successfully preventing drugs with torsadogenic risk from reaching the market [3], safe drugs may be potentially excluded due to the low specificity of the screening process, which targets only hERG channels. The Comprehensive *in vitro* Proarrhythmia Assay (CiPA) is a global initiative to provide revised guidelines for better evaluation of the proarrhythmic risk of drugs [4]. *In silico* evaluation of proarrhythmic action for different compounds constitutes an important foundation under the CiPA initiative to link data from *in vitro* assays to changes in cell behavior [4, 5].

The main component of the *in silico* evaluation are classifiers that are based on the so-called "derived features", input variables for the classifiers that are extracted from the outputs of biophysical models. The term "direct features" refers instead to the original feature set estimated from experiments investigating how drugs affect ion channel kinetics. Biophysical models serve as complex transformations that generate feature sets conditioned to the prior knowledge used in creating the model, thus potentially improving the efficacy of linear classifiers in inferring TdP risk. Diverse sets of derived features have been suggested as potential candidates for TdP risk classification (Table 1). In one of the earliest works on the use of the myocyte models for TdP risk prediction, simulated action potential duration at 90 % repolarization ($APD90$) was shown to provide the best discrimination of torsadogenic and non-torsadogenic drugs [6]. Other derived features extracted from the action potential (e.g., early after depolarization (EAD) and transmural dispersion of repolarization (TDR)), have also been suggested as possible candidate metrics for TdP risk prediction [7, 8]. Considering derived features from calcium transient in addition to features of the action potential have been shown to improve TdP risk discrimination [9]. Recently, tertiary TdP risk classifiers trained on a set of 12 drugs categorized into 3 clinical TdP risk groups (high, intermediate, and low/no risk) have been developed at FDA as a part of the CiPA initiative [10, 11]. Finally, two new derived features $cqInward$ [10] and $qNet$ [11] have been proposed to separate the 12 training drugs into desired target groups. The $qNet$ metric was further validated on 16 test compounds [12].

Criteria for selection of derived features for TdP classification are usually weakly justified, since the mechanisms of TdP are still not fully understood, and the initial feature selection and construction are usually performed either by testing on a small set of drugs or in an unsupervised fashion. Methods such as hierarchical clustering and principal component analysis are often employed for feature construction from derived metrics [9]. Nevertheless, the benefits of using derived features in TdP risk classifiers have so far not been convincingly demonstrated [13–15]. In this paper, we analyze a large virtual population of drugs and provide an explanation of the equivalent performance of direct and derived features. The main idea presented here is that results derived by exploring the structure of a cellular model using global sensitivity analysis (GSA) can be used to guide feature construction for statistical models. GSA allows one to examine sensitivity of derived features to the input parameters and to their interactions. In addition, we use GSA to understand differences in sensitivity of various model-derived features such as $APD90$, $qNet$ etc. to the blocks of different channels. We demonstrate that, in spite of previously claimed ties between $qNet$ and EAD generation, the parameters that affect $qNet$ are different than those influencing cell sensitivity to EAD. At the same time, we show that for the 28 drugs tested, the probability that any metric obtained by linearly combining direct features performs as well as the $qNet$ is very low. Finally, we also examine EAD generation in the model and its relationship with different direct and derived features.

Table 1. Previously proposed derived features.

Feature	*In silico* model	# compounds tested	Reference
APD_{90}	Ventricular myocyte models of rabbit, rat and human	31	[6]
$\frac{C_{drug,EAD}}{EFTPC}$	Human ventricular myocyte model	31 from [6]	[7]
TDR	Human ventricular myocyte model	55 from [16]	[8]
$\frac{C_{Drug,Arrhythmia}}{EFTPC}$	3D FEM model of human heart	12	[17]
$APD_{50}\&DiastolicCa^{2+}$	Human ventricular myocyte model	86 from [6, 16]	[9]
$cqInward$	Human ventricular myocyte model	12	[10]
$TdP_{population,score}$	Human ventricular myocyte model	62 (55 from [16])	[18]
$qNet$	Human ventricular myocyte model	12	[11]

$C_{drug,EAD}$ - concentration of the drug that produces EAD. $C_{drug,Arrhythmia}$ - concentration of the drug that produces arrhythmia in the model, TDR - Transmural dispersion, $cqInward$ - metric that that quantifies the change in the amount of charge carried by INaL and ICaL, $APD90$ - Action potential duration at 90% amplitude, $APD50$ - Action potential duration at 50% amplitude, $DiastolicCa^{2+}$ - Diastolic calcium concentration and $TdP_{population,score}$ - The fraction of models developing repolarization abnormalities(RA) multiplied by a factor inversely related to the drug concentration at which those RA occur

Methods

To perform GSA, we generated a large set of virtual drugs. A virtual drug comprises a random vector of changes to parameters of ion channels of the model. Responses to the virtual drugs were examined through evaluation of several model-derived features such as APD_{90}, $qNet$, and peak calcium concentration ($peakCa$). The CiPAORd model and input parameters section describes the *in silico* model used in the paper. The details of the input parameters considered for generating the virtual drug population are presented in Sampling virtual drug population. The section *In silico* simulations and derived features presents details on the derived features extracted from the *in silico* models. The virtual drugs were also tested for their ability to induce EAD in the model. In the section EAD protocols we describe the protocols used to test for EAD generation in the model. The methods used for GSA are described in the Global sensitivity analysis. The methods used for classifying virtual drugs based on direct and derived features with respect to EAD are listed in Binary risk stratification of virtual drug population. Finally, we describe the methods for classifying the 28 drugs selected under the CiPA initiative, which we refer in the manuscript as "CiPA drugs", with respect to their defined torsadogenic risk in the section Tertiary risk stratification of "CiPA drugs".

CiPAORd model and input parameters

In this study, we perform GSA on the CiPAORd model [11]. The CiPAORd model was developed at FDA by introducing several modifications to the original O'Hara-Rudy ventricular myocyte model [19] to improve proarrhythmic risk assessment.

Several input parameters have been used for simulation of drug effects. For the hERG channels, we used the concentration response of the drug, E_{max}, the unbinding reaction rate, K_u, and the membrane voltage at which half of drug-bound channels are open, V_{half}, as input parameters for the model. Note that the E_{max} parameter is calculated from other inputs to the original model. In this paper, we refer to the E_{max} parameter that represents the static component of the hERG block as $sbIKr$. For the other channel currents (i.e., fast sodium current INa, late sodium current $INaL$, L-type calcium channel current $ICaL$, slow-rectifying potassium channel current IKs, inward rectifier potassium current $IK1$, transient-outward current Ito) we used the general Hill equation of channel block,

$$b_{current,drug} = 100\% \times \frac{C_{drug}^h}{IC_{50,current} + C_{drug}^h}, \tag{1}$$

where current = $\{INa, INaL, ICaL, IKs, IK1, Ito\}$, $IC_{50,current}$ is the drug concentration at which a current is

reduced by half, C_{drug} is the drug concentration, and h is the Hill coefficient. The drug-induced blocks of channel currents $b_{current,drug}$ are used to scale the maximum conductance of the current $g_{current}$ in the *in silico* model calculated as

$$g_{current,drug} = \frac{100\% - b_{current,drug}}{100\%} \times g_{current}. \tag{2}$$

We perform GSA explicitly with respect to $b_{current,drug}$ rather than $IC_{50,current}$, C_{drug}, and h. In this study, we refer to the parameters of the block of INa, $INaL$, $ICaL$, IKs, $IK1$ and Ito as $bINa$, $bINaL$, $bICaL$, $bIKs$, $bIK1$, and $bIto$, respectively. Equation (1) is used in classification of real compounds.

Sampling virtual drug population

The population of virtual drugs is created through Monte Carlo sampling from a high-dimensional (10-D) input parametric space. The parametric space represent changes in model parameters used to describe drug binding and blocks of ionic currents. Basic cycle length (BCL) of cell pacing in the simulations was also considered as a parameter for GSA. The input parameters and their examined ranges is provided in (Table 2). In some cases, GSA was performed on metrics

Table 2. Ranges of input parameters

	endo cell		M cell		
Parameter	Min	Max	Min	Max	Description
$bINa$, %	0	90	0	90	percent block of fast sodium current
$bINaL$, %	0	90	0	90	percent block of late sodium current
$bIto$, %	0	90	0	90	percent block of transient outward current
$bIKs$, %	0	90	0	90	percent block of slowly activating delayed rectifier potassium current
$bICaL$, %	0	90	0	90	percent block of L-type calcium channel current
$bIK1$, %	0	90	0	90	percent block of inward rectifier potassium current
$sbIKr$	0	8	0	1.5	static component of the hERG channel current
V_{half}, mV	-200	-1	-200	-1	degree of drug trapping for the hERG channel
Ku, ms^{-1}	0	0.1	0	0.1	unbinding reaction rate for the hERG channel
BCL, ms	500	2000	500	2000	Basic cycle length for the simulations

derived from outputs of a mid-myocardial cell (defined as M cell in [19]) model. M cells are very sensitive to blocks of repolarization currents and produce EAD more easily. Sensitivity to EADs makes the analysis more complicated, and the range of hERG block for M cells had to be accordingly reduced.

In silico simulations and derived features

The action potential and calcium transients of the cells were simulated for the large virtual population of drug dataset (>10000 drugs) generated for GSA, and, separately, for the CiPA training (12 drugs) and validation (16 drugs) datasets (manual patch clamp data) [10, 11] using the CiPAORd model. Model simulations were run for 200 beats to achieve steady state. Simulations were carried out at different pacing rates (a parameter in GSA) for each of the endocardial (endo), mid-myocardial (M), and epicardial (epi) cell types. Several standard metrics explored previously for TdP risk discrimination were calculated from the action potential and Ca^{2+} transients. The metrics obtained from the *in silico* models are listed in the Table 3. Note that the metric $qNet$ was calculated as the area under the curve traced by the net current ($Inet = ICaL + INaL + IKr + IKs + IK1 + Ito$) from the beginning to the end of the last simulated beat as defined in [11].

Table 3. Derived features extracted from the CiPAORd model

Derived Feature	Description
$qNet$	Net electronic charge carried by IKr, $INaL$, $ICaL$, IKs, $IK1$, Ito currents
$APD90$	Action potential duration at 90% repolarization
$APD50$	Action potential duration at 50% repolarization
$peakVm$	Peak voltage
$diastolicCa$	Diastolic calcium level
$peakCa$	Peak value of intracellular calcium
$CaTD50$	Calcium transient duration at 50% return to baseline
$CaTD90$	Calcium transient duration at 90% return to baseline

EAD protocols

Drug-induced EAD risk (sensitivity of a cell against EAD generation) for both the virtual drugs and the CiPA compounds was examined in the endo and M cell types using two separate protocols. The M cell type in the CiPAORd model was more prone to EAD generation than the endo cell type. We tested generation of pause-induced EADs (that are implicated as triggers of TdP [20–22]) in the M cell type as in our previous study [14]. Briefly, the cell was stimulated 200 times at a constant cycle length. After 200 stimuli, an additional stimulus was applied following a pause equal to the basic cycle length. In the endo cells, pause-induced EADs occurred rarely, and we examined drug-induced EAD risk in presence of an added perturbation by reducing the maximum conductance of hERG channel current (IKr) as in [11]. The cell was stimulated for 200 beats with additional block of maximum conductance of IKr by 85%. The 85% block was selected since almost half of the population of virtual drugs (across the entire parametric space observed) resulted in EAD development in the model simulations.

Global sensitivity analysis

GSA was performed using a difference-based Morris sensitivity method [23], a variance-based sensitivity method [24, 25]), and Monte Carlo filtering [24, 26].

Morris method

The Morris-method [23] provides the simplest form of GSA and can be considered as an extension of the local sensitivity methods. Similar to other local sensitivity analysis methods, the input parameters are varied one at a time (OAT) to obtain the partial derivatives of output metrics with respect to individual input parameters. A k-dimensional unit cube (representing k independent normalized input parameters) is discretized into the p-level grid Ω. Unlike for local sensitivity analysis, however, the reference values of each parameter are selected randomly from the set {0, 1/(p-1), 2/(p-1) ... , 1 - Δ} rather than perturbing some nominal values. The fixed increment $\Delta = p/2(p-1)$ is added to each parameter one at a time in a random order to compute the elementary effect (EE) of the ith input parameter X_i,

$$EE_i = \frac{Y(X_1, X_2, \cdots, X_{i-1}, X_i + \Delta, \cdots, X_k) - Y(X_1, X_2, \cdots, X_{i-1}, X_i, \cdots, X_k)}{\Delta}. \quad (3)$$

$k+1$ simulations are required to compute EE_i for k parameters. Multiple runs (paths) are performed from different start points to cover the entire parameteric space and obtain ensemble of EEs for each input parameter. The total number of simulations is $r \times (k+1)$, where r is the number of paths. The mean of absolute EE_i (denoted as μ^*) and standard deviation of EE_i (denoted as σ_{EE}) are computed for each input parameter. The μ^* statistic allows to rank the input

parameters based on their influence [24] and σ_{EE} helps identify nonlinear and/or interaction effects. The Python SALib package was used to perform the elementary effect analysis [27]. In this study, we considered $r = 1000$ unless otherwise specified, resulting in 9000 and 11000 input samples (virtual drugs) for $k = 8$ and 10, respectively.

Variance-based global sensitivity analysis

Sobol sensitivity analysis method [25] is a model-independent GSA method that is based on variance decomposition. It relies on an all-at-time (AAT) sampling strategy where output variations are induced by varying all the input factors simultaneously. Let a derived-metric Y from a computational model be represented by a function f of the input parameters,

$$Y = f(\mathbf{X}) = f(X_1, X_2, \cdots, X_k), \tag{4}$$

where $\mathbf{X} = \{X_1, X_2 \cdots X_k\}$ is the input parameter set. The function can then be decomposed into a sum of elementary functions of increasing dimensions,

$$Y = f_0 + \sum_i f_i(X_i) + \sum_i \sum_{j>i} f_{ij}(X_i, X_j) + \cdots + f_{12\cdots k}(X_1, \cdots, X_k). \tag{5}$$

The input parameters are assumed to be random variables that uncorrelated and mutually independent. The functional decomposition can be translated into a variance decomposition. This allows to quantify the variance contribution to the total output of individual parameters and the parameter interactions,

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{123\cdots k}, \tag{6}$$

where $V_i = V_{X_i}[E_{X_{\sim i}}(Y|X_i)]$ is the first-order effect for a given model input X_i, $V_{ij} = V_{X_i,X_j}[E_{X_{\sim ij}}(Y|X_i, X_j)] - V_{X_i}[E_{X_{\sim i}}(Y|X_i)] - V_{X_j}[E_{X_{\sim j}}(Y|X_j)]$ and so on are the higher-order effects due to interactions of model inputs. Here, E_{X_i}, V_{X_i} are expectation and variance taken over X_i; $X_{\sim i}$ denotes all factors but X_i. The Sobol sensitivity indices are obtained as the ratio of partial variance to the total output variance,

$$S1_i = \frac{V_i}{V(Y)}, \quad S2_{ij} = \frac{V_{ij}}{V(Y)} \cdots \quad . \tag{7}$$

The number of sensitivity indices in (7) grow exponentially with k and typically only sensitivity indices of up to order two ($S1_i$ and $S2_i$) and the total-effect indices (ST_i) are estimated [28]. The total-effect index

$$ST_i = \frac{E_{X\sim i}[V_{X_i}(Y|X_{\sim i})]}{V(Y)} = 1 - \frac{V_{X\sim i}[E_{X_i}(Y|X_{\sim i})]}{V(Y)} \tag{8}$$

measures the impact of main effect of X_i and all its higher-order interaction effects with the other parameters [29]. The Python SALib package was employed to perform the variance-based sensitivity analysis [27]. The calculations of $S1_i$, ST_i and $S2_{ij}$ require $n \times (2k + 2)$ model evaluations using Saltelli's sampling scheme [30] where n is the sample size and k is the number of input parameters. In this study, we considered $n = 1000$ unless otherwise specified, resulting in 18000 and 22000 Monte Carlo samples (virtual drugs) for $k = 8$ and 10, respectively.

Monte Carlo filtering

Monte Carlo filtering (MCF) is used generally in factor-mapping tasks to identify key input parameters responsible for driving model outputs within or outside predefined target regions (refer to [24] for a detailed methodology). A brief overview of the MCF technique in the context of EAD sensitivity analysis of the CiPAORd model is presented here. Model simulations were carried out for n Monte Carlo samples (virtual drugs) generated for the Sobol sensitivity analysis in presence of additional perturbations (see section EAD protocols). Each input parameter X_i of the Monte Carlo input sample set with size n is categorized into the "Behavioral" subset $(X_i|EAD-)$ and the "Non-behavioral" $(X_i|EAD+)$ with sizes n_1 and $n_2 = n - n_1$, respectively, based on the absence and presence of EADs in simulated output. Empirical cumulative density functions (CDF) $F_{n_1}(X_i|EAD-)$ and $F_{n_2}(X_i|EAD+)$ are estimated for both the subsets of random input samples. The distance between the two empirical CDFs provides an estimate of sensitivity EAD feature to the input parameter X_i. Kolmogorov-Smirnov two-sample statistic test was used to quantify the difference between the two CDFs. Kolmogorov-Smirnov test is characterized by a D-statistic and a p-value. The D-statistic is defined as [24]

$$d_{n_1,n_2} = sup||F_{n_1}(X_i|EAD-) - F_{n_2}(X_i|EAD+)||. \tag{9}$$

The larger the D-statistic (or equivalently the smaller the p-value), the more important the input parameter is in driving the behavior of the model to EAD [24].

Binary risk stratification of virtual drug population

Logistic regression classifiers were constructed with both the direct, Table 2, and derived features, Table 3, individually or in combinations as inputs for binary classification of the virtual population of drugs into $EAD+$ and $EAD-$ categories. To determine the targets for risk discrimination, the virtual population of drugs (n = 22000) generated for Sobol sensitivity analysis was tested in the presence of additional perturbations (see section EAD protocols). Model outputs were examined for EADs, and targets for EAD risk classification were assigned based on presence (target = 1) or absence (target = 0) of EADs. The derived features used as input for the classifier were obtained for the same set of virtual drugs in the absence of added perturbations in the endo cell type. The equation for the generalized classifier model is described as

$$logit(EAD) = \frac{1}{1 + exp(-(\beta_0 + \sum_i^n \beta_i M_i))}, \tag{10}$$

where M_i represents the ith input metric to the model (which could be either a direct or a derived feature), n is the number of input features used to train/test the model, and β_i $(i = 1, 2, ..n)$ are parameters to be determined. The performance of the classifier was evaluated by K-fold cross validation and was reported in terms of Area Under the Receiver Operating Characteristic Curve (AUC ROC) and Jaccard index (ratio of correctly classified samples to the total number of samples). Python's scikit-learn package [31] was used to train/test the logistic-regression classifiers.

Tertiary risk stratification of "CiPA drugs"

In silico simulations of blocks with the 28 "CiPA drugs" were carried out using the *in vivo* manual patch clamp measurements collected on the pharmacological effects of these compounds reported in [10, 12]. The effective therapeutic concentrations, the IC_{50} values, the Hill coefficient values, the drug binding parameters, and the defined torsadogenic risk of the "CiPA drugs" are listed in the Supplemental Material. One-versus-rest multiclass logistic regression classifiers were constructed for classification of the "CiPA drugs" into low, medium, and high risk categories according to their defined

torsadogenic risk based on the model-derived metric $qNet$. A multiclass logistic regression classifier was also constructed using a surrogate of $qNet$ ($qNet_s$) that was estimated via linear regression from the direct features. The "CiPA drugs" were also classified by metrics constructed from the linear combination of input metrics with random coefficients β_i as described by

$$M_l = \sum_i^n \beta_i F_i, \tag{11}$$

where F_i represents the ith direct feature normalized by their ranges in the merged CiPA training and validation set, n is the total number of direct features, and β_i $(i = 1, 2, ..., n)$ are coefficients of the ith feature. Several such linear metrics, M_l, were obtained by selecting sets of coefficients β_i via Latin hypercube sampling. The Python SALib package was used for Latin Hypercube sampling [27].

The generalized classifier model equation is

$$logit(TdP) = \frac{1}{1 + exp(-(\gamma_0 + \gamma_1 m))}, \tag{12}$$

where $m = M_l$, $qNet$, or $qNet_s$ is the input variable to the classifier model and γ_0 and γ_1 are parameters that are determined from training the classifier. Note that unlike multiclass logistic regression constructed directly on the metrics M_i, where the coefficients β_i may vary between separating hyperplane of individual classes, the metrics obtained using (11) present an additional constraint for tertiary risk classification by not allowing the coefficients β_i to change when performing separation of one class or another. The Python scikit-learn package [31] was used to train/test the multiclass (one-versus-rest) logistic-regression classifiers and perform linear regression. The trained logistic regression model was used to estimate the number of correctly classified drugs and reported as a confusion matrix.

Finally, the "CiPA drugs" were also classified based on the concentration (normalized to effective free therapeutic plasma concentration ((EFTPC)) necessary to induce EADs in the CiPAOrd model. The effects of "CiPA drugs" were simulated using protocols described in the EAD protocols section at progressively increasing drug concentrations until EADs were observed. A maximum concentration of 70xEFTPC was tested for each drug. Drugs that did not result in EADs in the tested concentration range were classified as low risk drugs. Drugs that instead resulted in EADs at concentrations smaller than 8xEFTPC were labeled as high risk, while the remaining drugs were labeled as intermediate risk drugs.

The amount of drug-induced block of the channels depends on the EFTPC of the compound. Unfortunately, reported EFTPC values are highly variable [32]. In addition, the actual free plasma concentrations of drugs in subjects could also differ because of inter-individual variations, impaired metabolism, and interactions with other drugs. In fact, drug concentrations could potentially be much larger than reported maximum EFTPC values. We have recently proposed a strategy [14] to partially resolve the problem of EFTPC variability by examining the drug-induced effects at concentrations that result in identical block of hERG channel current for each drug. Therefore, in this study, in addition to classification by EAD risk due to drug-induced changes in all seven ion-channel currents typically measured in *in vitro* assays, we also examine the EAD model-derived metric that were constructed considering concentrations at a fixed value $sbIKr = 4$. Also the parameter BCL was fixed to 1000.

Results

Analysis of global sensitivity for $APD90$, $qNet$, $peakCa$, and EAD

Analysis of elementary effects

First, GSA was performed using the Morris method (see section Morris method) with $r = 1000$ repetitions and $k = 10$ input parameters, which require $n = r(k+1) = 11000$ model runs per cell type. Figure 1 shows relative impact of different input parameters on the variations of the $qNet$, $APD90$, and $peakCa$ output metrics for the endo cell. The X-axis represents μ_j^*, a measure of the influence of the jth input variable on the output response. Larger μ^* values indicate greater influence of the input parameter on the output response. The Y-axis represents the standard deviation of the elementary effect σ_j. Higher σ values for an input parameter indicate the presence of possible interactions with other variables affecting the output. The condition $\sigma/\mu^* > 1$ is an evidence of strong non-linear effects due to interactions between parameters. The input parameters with $\sigma/\mu^* < 0.1$ are instead considered to have minimal interactions. The analysis of elementary effects reveals that $bINaL$, $sbIKr$, $bINa$, and $bICaL$ have the strongest influence on the $qNet$ metric for the endo cell. Moreover, the ratio $\sigma/\mu^* < 1$ for these input parameters suggests that although they are not linearly correlated with $qNet$, they show a monotonic effect. On the contrary, the most influential input parameters affecting $APD90$ in the model was found to be $sbIKr$, the static component of the block of hERG channel. $peakCa$ concentration was most sensitive to the $bICaL$ parameter that represents the block of the L-type calcium channel current. The results suggest that blocks of different cardiac ion-channels (input parameters) have distinct influence on the derived metrics $APD90$, $qNet$, and $peakCa$ in the CiPAORd model. The Supplemental Material reports estimated elementary effect values for all the derived features listed in Table 3.

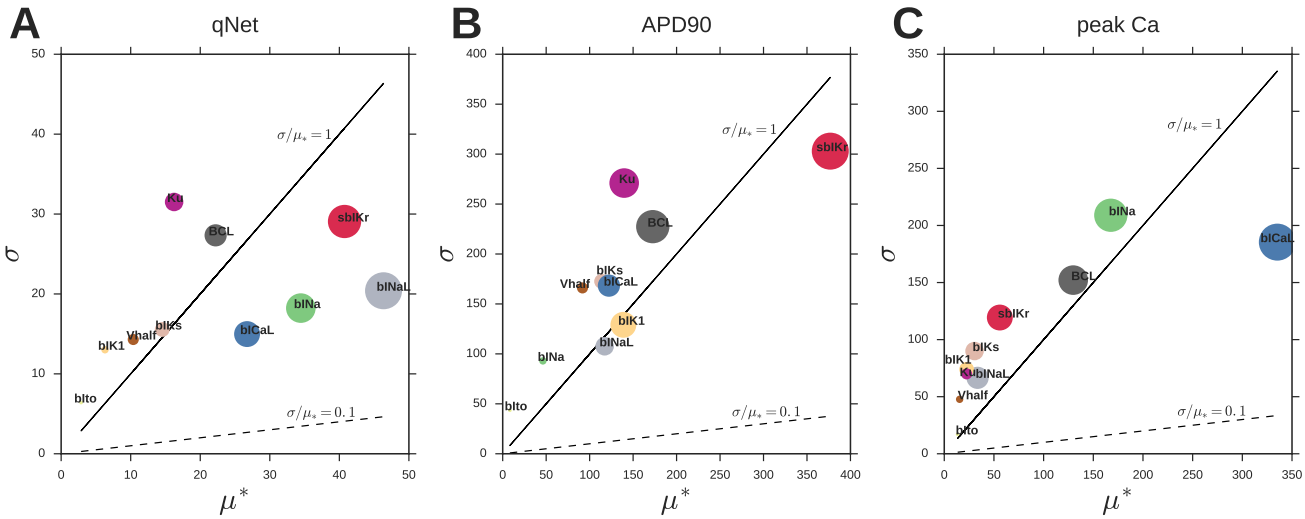

Fig 1. Estimators σ and μ^* providing information on influence of blocks of different cardiac ion-channels, drug-binding parameters and BCL on CiPAORd endo cell model-derived metrics. **A**: $qNet$, **B**: $APD90$, and **C**: $peakCa$. The circle size for the input parameters correspond to their relative importance, with the largest diameter circle denoting the most influential input and the smallest diameter circle denoting the least influential inputs.

Variance-based analysis

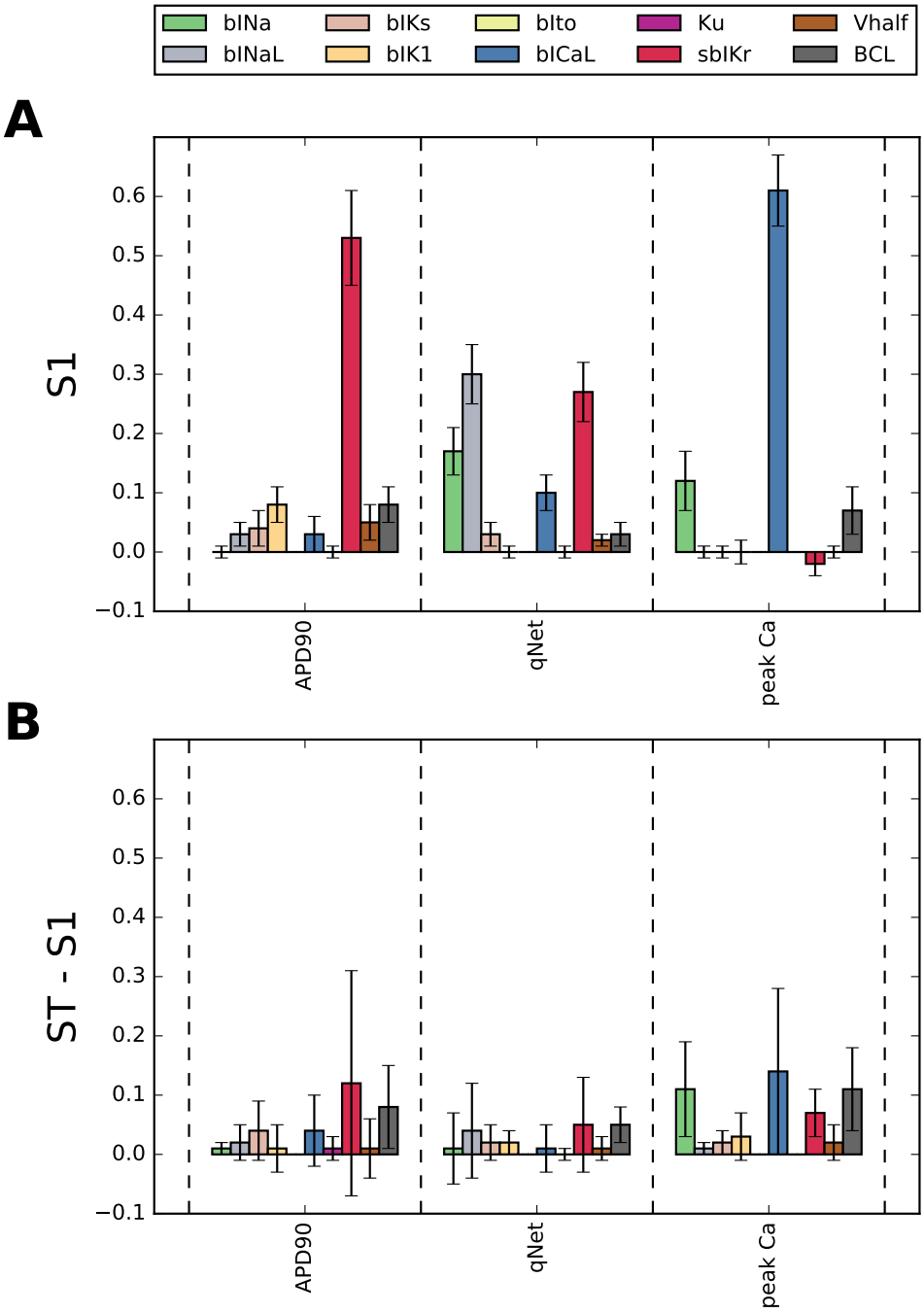

Fig 2. Assessment of sensitivity to blocks of different cardiac ion-channels, drug-binding parameters and BCL for $APD90$, $qNet$, and $peakCa$ output responses in the CiPAORd endo cell model using the Sobol sensitivity indices. **A**: First-order sensitivity Sobol index, $S1$. **B**: Total sensitivity Sobol index, ST. The algorithm to calculate Sobol indices can produce negative values that could be eliminated by increasing sampling size.

Figure 2 shows values of the first-order Sobol sensitivity indices ($S1$) and differences between total sensitivity indices

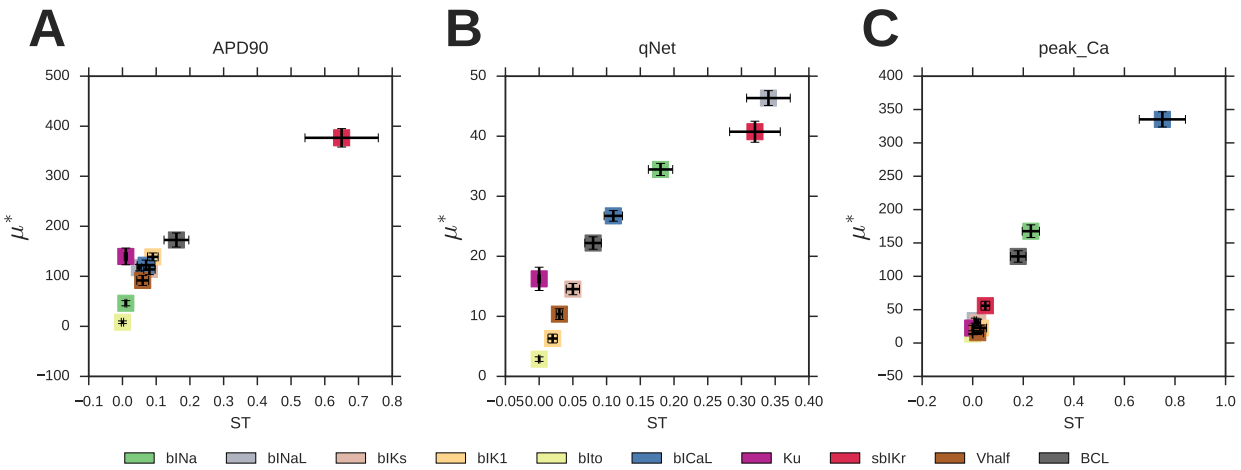

Fig 3. Plot of μ^* index calculated by Morris method versus ST estimated in the Sobol sensitivity analysis. **A**: $qNet$, **B**: $APD90$, and **C**: $peakCa$.

(ST) and $S1$ for three output responses $APD90$, $qNet$, and $peakCa$. Similar to the Morris method results, the Sobol sensitivity indices indicate that $APD90$ is the most sensitive to $sbIKr$ block, $qNet$ to $sbINaL$, and $peakCa$ to $bICaL$. Unlike elementary effects, Sobol indices quantify the contribution of the input parameter to the metrics in isolation as well as in presence of interaction with other parameters. The effect of $sbIKr$ on $APD90$ as quantified by $S1$ indicates that $sbIKr$ contributes to $> 50\%$ of the variation observed in $APD90$ across the observed input space. $qNet$ was found to be most sensitive to $bINaL$, $sbIKr$, $bINa$, and $bICaL$ with contributions to the output variation of 30%, 26%, 17%, and 10%, respectively. $bICaL$ had the strongest impact on the variability of $peakCa$ concentrations with an $S1$ index of around 0.6. The difference between the ST and $S1$ indices in Figure 2**B** shows the impact of higher-order indices. Small differences between $S1$ and ST for several derived metrics such as $APD90$, $qNet$, $peakCa$ suggest minor influence of parameter interactions. The $S1$ and ST sensitivity indices obtained for all the derived features (Table 3) across different cell types are reported in the Supplemental Material.

The Figure 3 shows a plot of the ST index against the μ^* statistic from the Morris method. The observed strong correlation is in agreement with previous analyses comparing elementary effects and variance-based methods [33]. Given the results, the Morris method was able to rank the relative influence of parameters with similar accuracy of the more computationally expensive variance-based method. However, as discussed earlier, the variance-based sensitivity measures provide also additional information on parameter interactions.

Derived features predicted as linear combination of direct features

The estimated sum, $\sum S1_i$, of the first-order Sobol indices $S1_i$ for the i direct features (Table 2) for different model-derived metrics are listed in Table 4. $\sum S1_i$ represents the contribution of all individual input parameters, i.e., the main effects, to the total variation of an output metric of the model without accounting for the influence of parameter interactions. The observed values (Table 4) indicate that >85% of the variance in $qNet$, $APD90$, $CaTD90$, and $diastolicCa$ can be attributed to the individual input parameters for the endo cell model. The parameter interactions explain less than 15% of the variance of these derived metrics. 20-30% variance in $APD50$, $peakCa$, $peakVm$ and $APD50$ in the endo cell model is from the interactions in parameters. A large effect of parameter interaction is observed for $CaTD50$ feature ($\sim$ 47% variance can be attributed to interactions). Note that although $S1$ does not account for the effects of eventual interactions between parameters, contributions to the output variance from higher-order nonlinear effects of individual inputs are indeed captured. For example, let us consider the simple model $Y = x_1 + x_2^2$, for which 100% of the output variance is captured by the sum $S1_{x1} + S1_{x2}$. The $S1_{x2}$ index comprises an estimate of the variance in Y due to the second-order term x_2^2. Hence, even in the absence of parameter interactions, the model-derived metrics can be nonlinear with respect to some inputs. We further examined the fitting of linear regressions of direct features (assuming absence of nonlinearity and parameter interactions) to approximate the different model-derived metrics. The goodness of fit was reported as R^2 values (Table 4). The R^2 value represents the portion of variation in output response that can be explained by the regressors of the model. R^2 values between 0.79 and 0.86 for $qNet$, $APD90$, $CaTD90$, and $diastolicCa$ features (Table 4) for the endo cell model indicate that the particular derived features can be well captured as linear combinations of the inputs, i.e., of the direct features. Moreover, the small differences between the $\sum S1_i$ and the R^2 indicate minor contributions of higher-order terms of individual inputs on variation of most of the derived features except for $diastolicCa$ and $peakVm$. We also report the $\sum S1_i$ metric and R^2 values obtained for the M cell model in Table 4 for comparision.

Table 4. Total contribution of the main effects on the variance of derived metrics estimated by the sum of the Sobol index and the fit of linear regression of direct features

	endo cell		M cell	
	$\sum S1_i$ (Sobol analysis)	R^2 (Linear regression)	$\sum S1_i$ (Sobol analysis)	R^2 (Linear regression)
$qNet$	0.92	0.86	0.94	0.91
$APD90$	0.84	0.83	0.93	0.86
$CaTD90$	0.89	0.82	0.96	0.80
$diastolicCa$	0.93	0.79	0.94	0.78
$peakCa$	0.78	0.76	0.90	0.86
$APD50$	0.70	0.72	0.85	0.80
$peakVm$	0.73	0.61	0.85	0.67
$CaTD50$	0.53	0.56	0.89	0.65

Regional sensitivity analysis

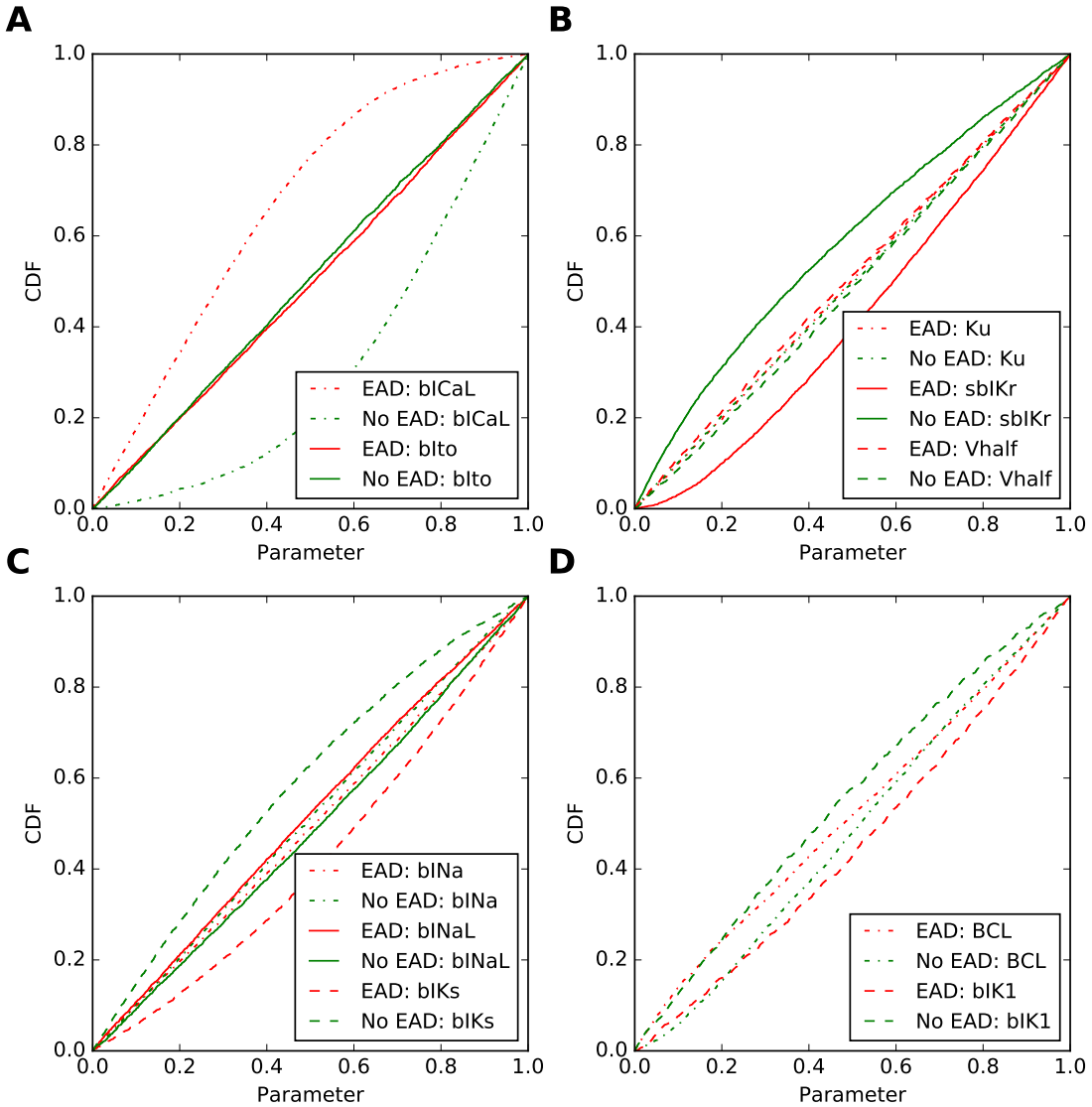

Fig 4. Ranking the most influential model parameters regulating EAD generation in the CiPAORd endo cell model using Monte Carlo filtering analysis. Empirical CDF for each of the 10 input parameters conditional to the presence or absence of EADs: **A**: $bIto$, $bICaL$; **B**: Ku, $sbIKr$, V_{half} **C**: $bINa$, $bINaL$, $bIKs$; and **D**: $bIK1$ and BCL.

For Monte Carlo filtering, the target space was partitioned into realizations with either presence or absence of EADs in simulated action potentials under the action of a virtual drug population ($n = 22000$). Figure 4 shows the empirical CDFs for each of the 10 input parameters. The dotted and solid lines represent the estimated CDFs for the behavioral $F_{n_2}(X_i|EAD+)$ and the non-behavioral $F_{n_1}(X_i|EAD-)$ subsets, respectively. The behavioral and non-behavioral subsets comprised 10479 and 11521 samples. The bigger the distance between the empirical CDFs for the behavioral and non-behavioral subsets, the greater is the influence of the parameter to development of EADs. The figure shows that the parameter $bICaL$ (Figure 4**A**, dashed-dotted line) has the strongest influence on susceptibility of the model to EADs. The parameters $sbIKr$ (Figure 4**B**, solid line), $bIKs$ (Figure 4**C**, dashed line), and $bIK1$ (Figure 4**D**, dashed line) have

the next highest contribution to model sensitivity to EADs. Table 5 provides estimates of the differences in the behavioral and non-behavioral empirical CDFs using a Kolmogorov-Smirnov two-sample test. The results show that the parameter $bICaL$ regulating the block of the L-type calcium channel current had the highest influence on EAD in models with both endo and M cell types. In the M cell, the parameter $sbIKr$ and $bCaL$ appear to be equally important for regulation of EADs in the model. In contrast, since the additional block of the hERG channel current was required to trigger EAD in the endo cell type, the parameter $sbIKr$ had moderate influence in regulation of EADs. The parameter $bIKs$ had moderate influence in both endo and M cell types. The block $bIK1$ appears to have higher influence in the endo cell compared to the M cell. The parameters Ku, Ito and $bINa$ were the least important. Monte Carlo filtering demonstrates that $sbIKr$ and $bICaL$ parameters contribute the most to generation of EADs.

Table 5. Monte Carlo filtering statistics

Parameter	M cell	endo cell
$bICaL$	**0.50**	**0.58**
$sbIKr$	**0.49**	**0.24**
$bIKs$	0.12	**0.25**
$bIK1$	0.08	0.16
V_{half}	0.08	0.05
$bINaL$	0.07	0.05
BCL	0.04	0.10
$bINa$	0.03	0.03
$bIto$	0.02	0.03
Ku	0.02	0.02

Classification of virtual and CiPA training/validation drug datasets

Classification of CiPA compounds based on $qNet$ and a surrogate of $qNet$

Here, we wanted to examine how the findings from the global sensitivity analysis on the virtual drug population would translate for a set of actual drugs with channel blocks covering only a relatively small portion of the parametric space examined previously. Our GSA results on the virtual set of drugs (Figure 2 and Table 4) demonstrate minimal contribution of input parameter interactions to the variance of $qNet$ and a good linear approximation of $qNet$ from direct features. To corroborate this finding also for actual drugs, we constructed a surrogate metric of $qNet$ as the linear combination of the 5 direct features $bICaL$, $bINaL$, $bINa$, $sbIKr$, and V_{half}. A second surrogate accounted for only 4 direct features neglecting the least influential parameter V_{half}. Multiclass logistic regression was then employed to fit TdP risk for the CiPA training (n = 12) and validation (n = 16) compounds at 2x EFTPC drug concentrations based on actual $qNet$ and its surrogates. The values of $qNet$ and its surrogates are plotted in Figure 5**A**, **B**, **C**. The performance measure of the classifiers reported as confusion matrices in Figure 5**D**, **E**, **F** indicates that classifiers based on surrogates of $qNet$ fit the data similar to the original $qNet$, although with a slightly worse accuracy due to small perturbations in the $qNet$ surrogates with respect to the original metric obtained from the model.

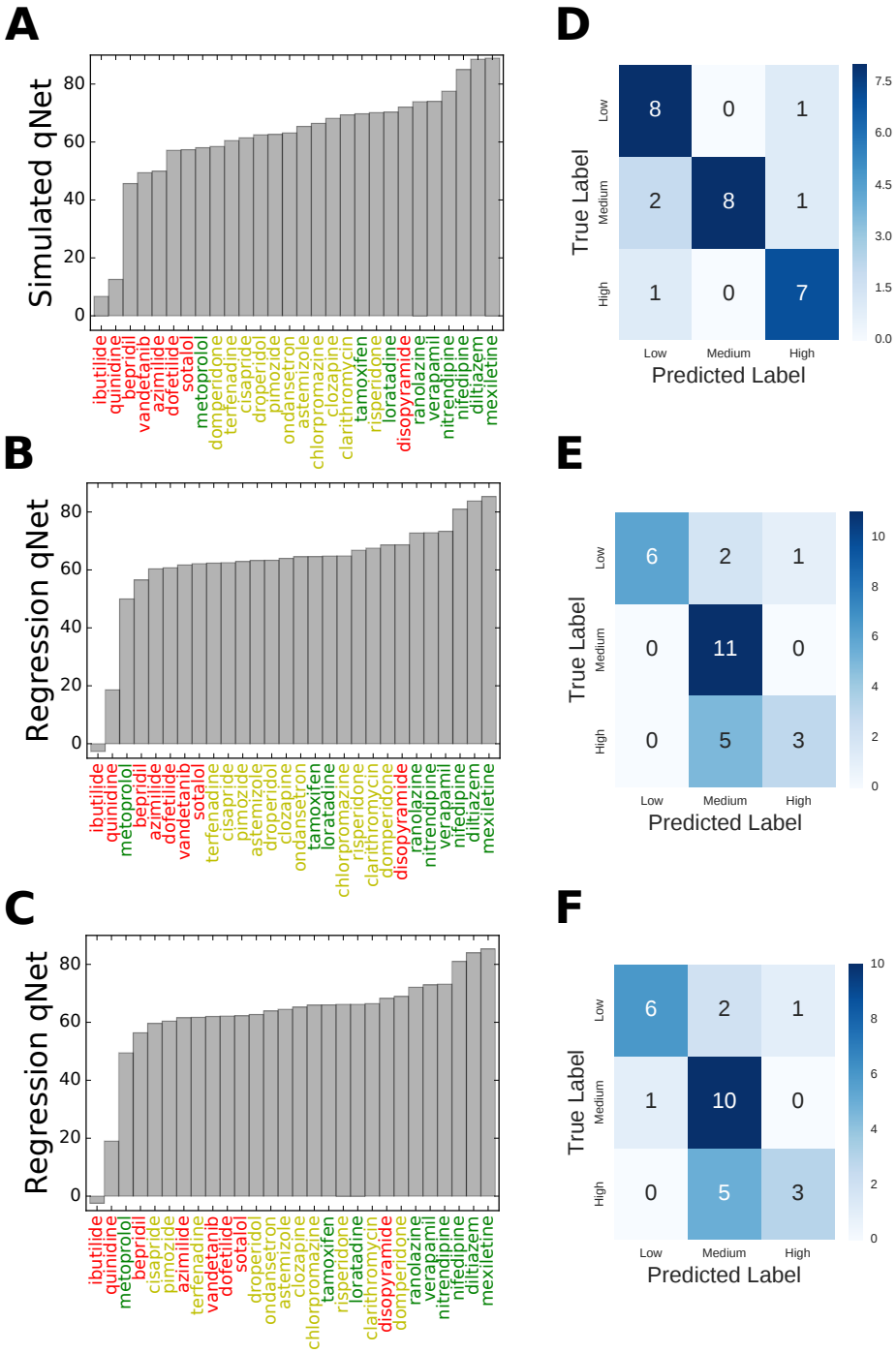

Fig 5. Classification based on $qNet$ and its surrogates for the 28 CiPA training and validation compounds at 2x EFTPC drug concentrations. **A**: $qNet$ obtained from simulations of the drugs at pacing rate of 2000 ms in the CiPAOrd endo cell model. **B**: $qNet$ surrogate calculated using linear regression of the 5 direct features ($bICaL$, $bINaL$, $bINa$, $sbIKr$, V_{half}). **C**: $qNet$ surrogate calculated using linear regression of the 4 direct features ($bICaL$, $bINaL$, $bINa$, $sbIKr$). **D**: Concordance analysis of $qNet$ based risk categorization of the 28 drugs against the actual torsadogenic risk of the compounds reported as a confusion matrix. **E**: Confusion matrix based on $qNet$ surrogate from linear regression of 5 direct features. **F**: Confusion matrix based on $qNet$ surrogate from linear regression of 4 direct features.

Randomly generated linear metrics

In the previous sections, we demonstrated that popular derived metrics for classification (i.e., $APD90$, $qNet$) could be well recapitulated by linear regressions constructed on direct features. In this section, we further explore the matter by tackling the following question: What are the odds that a randomly generated metric based on the linear combination of direct features would perform well in classification of TdP risk?

To address the problem, we automatically generated several classification metrics for the CiPA datasets as linear combinations of the 5 normalized direct features $\overline{sbIKr}$, $\overline{bINaL}$, $\overline{bINa}$, $\overline{bICaL}$, and $\overline{V_{half}}$. Note that the parameters for IKs, $IK1$, and Ito channels were not accounted for developing the metric as most of the CiPA compounds had minor effects on these channels.

$$metric = \beta_1\overline{sbIKr} + \beta_2\overline{bINaL} + \beta_3\overline{bINa} + \beta_4\overline{bICaL} + \beta_5\overline{V_{half}}. \tag{13}$$

In the first step of our analysis, two groups of 40,000 sets of coefficients β_i, $i = 1, ..., 5$ were randomly extracted via Latin Hypercube Sampling (LHS) from within the prescribed bounds reported in the Table 6. The first group of coefficient sets (group A) was selected to ensure a certain degree of correlation to APD, as exhibited by many derived metrics. More specifically, the coefficient of $sbIKr$ was prescribed to have opposite sign with respect to the coefficients of $bINaL$ and $bICaL$. The second group of coefficient sets (group B) was more inclusive and allowed the same admissible range ([-1, 1]) for all coefficients. All the linear combination metrics generated by (13) according to the coefficients in group A and group B were then used in multiclass logistic regression for tertiary classification of the 12 CiPA training drugs. Out of the two groups of random sets of coefficients, approximately 1% (n=400) and 0.25% (n=100) of the metrics generated from group A and B, respectively, resulted in perfect discrimination of the drugs into low, intermediate, and high risk categories. In other words, the probability that a randomly generated metric would perform as well as $qNet$, which also achieves perfect accuracy, is quite low. Figure 6**A** shows frequency of the drugs that are correctly classified by metrics generated from the coefficient sets of group A. We observe that a classifier built on a metric with random coefficients have highest frequency for correctly predicting 7 out of the 12 training drugs. For visual comparison, the performance of logistic regression classifiers based on $qNet$, $APD90$, and $peakCa$ metrics are labeled in the figure as solid vertical lines. The low probability of having randomly generated metrics that would perform at least as well as $APD90$ and $qNet$ highlights the strong link between perturbed APD in an isolated cell and TdP risk.

The second step of the analysis targeted the small portion of metrics that were able to classify correctly at least 11 out of the 12 drugs of the training set. The selected metrics were further tested for classifying the merged training-validation set from CiPA, again via multiclass logistic regression. Note that we refit logistic regression (12) to the merged dataset instead of reusing coefficients γ_i obtained for the training set to quantify the ability of the metrics to separate full set of drugs into correct categories. Figure 6**B** and **C** show the frequencies of correctly predicted drugs from the merged training-validation dataset, but when accounting only for the selected metrics generated from group A that were able to classify correctly at least 11 of the 12 training drugs. Note that the $peakCa$ and $APD90$ based metrics classified correctly less than 11 training drugs, but were included in these subpanels just for comparision. Out of the 1% (n=400) of the randomly generated metrics regressions that resulted in 12 out of 12 correctly classified training compounds, approximately 0.2-0.5% (n=1-2) of such regressions resulted in >23 correctly classified compounds (accuracy of $qNet$), and about 11% (n=46) coefficients resulted in >20 correctly classified compounds. Therefore, in spite of not being directly related to EADs in the model as shown in this paper, $qNet$ demonstrates a remarkable score even among metrics that perform perfectly on the training set.

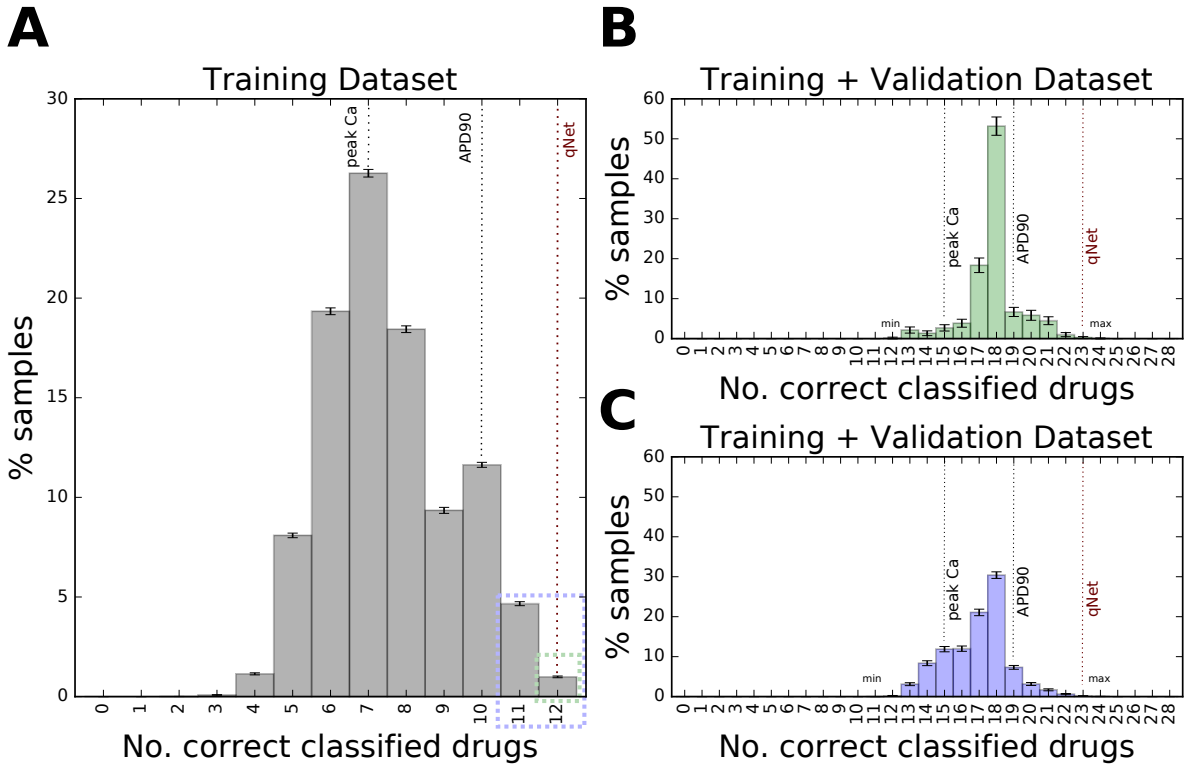

Fig 6. Histogram of number of correctly classified drugs by regressions with random coefficients within the ranges in the column A of Table 6. **A**: Frequency distribution of the randomly generated metrics that classify a given number of CiPA training drugs correctly. **B**: Frequency distribution of the subset of randomly generated metrics that classify a given number of CiPA training and validation drugs correctly and all of the CiPA training drugs. **C**: Frequency distribution of the subset of randomly generated metrics that classify a given number of CiPA training and validation drugs correctly and at least 11 of the 12 CiPA training drugs. Labels $qNet$, $APD90$, and $peakCa$ point to accuracy of these metrics for comparison. In order to estimate the confidence intervals, LHS was repeated for 100 different random seeds.

Table 6. Ranges for the coefficients

Coefficient	Parameter	group A	group B
β_1	$sbIKr$	[-1, 0]	[-1, 1]
β_2	$bINa$	[-1, 1]	[-1, 1]
β_3	$bINaL$	[0, 1]	[-1, 1]
β_4	$bICaL$	[0, 1]	[-1, 1]
β_5	V_{half}	[-1, 1]	[-1, 1]

group A Ranges that force metrics to have different signs between $sbIKr$ and $bINaL$ with $bICaL$.
group B General ranges.

Classification of CiPA training/validation drugs based on EADs

We also examined the classification of CiPA compounds based on the drug concentration (normalized to EFTPC) necessary to induce EADs in the CiPAORd model, defined here as threshold of EAD. Since the Monte Carlo filtering results point to $sbIKr$ and $bICaL$ being the most critical parameters regulating EAD development, only drug-induced changes of these two parameters were taken into consideration. We also compared the obtained thresholds for $EADs$ to the thresholds considering drug-induced changes of all the seven ion channel currents measured from the *in vitro* assays and characterized by the 9 parameters reported in Table 2. The results are summarized in the Table 7. Our EAD analysis show that the drugs in the high-risk category had consistently a threshold value of less than 7, even for the threshold obtained when considering only the drug effects on two parameters, static block of hERG channel current $sbIKr$ and the block of L-type calcium channel current $bICaL$. Addition of dynamic hERG channel current parameters as well as other input parameters resulted in no significant changes in the observed thresholds for EAD generation. A high risk drug disopyramide from the CiPA validation dataset did not induce EAD in the model under all tested conditions. The intermediate risk drugs Cisapride and Ondansetron resulted in EADs in the model at a threshold of less than 7 similar to the drugs in high risk category under all tested conditions. Similarly, Ranolazine and Metoprolol drugs that are defined as low risk under the CiPA initiative had a threshold value of less than 7 for all the conditions tested. The low-risk drugs Diltiazem, Mexiletine, Verapamil, Loratadine, Nifedipine, and Nitrendipine did not produce EADs in the model under all the tested conditions. Chlorpromazine resulted in EADs at relatively high threshold compared to the high risk drugs with a threshold of > 14 under all conditions. Low risk drug Tamoxifen consistently resulted in EADs in the model at thresholds values greater than 15 similar to intermediate risk drugs. Terfenadine, Pimozide, Clozapine were the only few drug with significant changes in the observed threshold and switching to a different risk category when not considering the drug-induced changes of parameters other than $sbIKr$ and $bICaL$.

Although EADs are thought to be cellular precursors of TdP, the classifier based on EADs alone ranks correctly only 17-19 drugs, thus performing worse than $qNet$ and some other metrics. Among plausible reasons for the poor performance, we note that only isolated cell types were tested to generate EADs, models might not yet captures the phenomenon correctly, and in general TdP risk assessment should not rely on EAD evaluation alone.

Table 7. Threshold (xEFTPC) of EAD for CiPA training (12 drugs) and validation (16 drugs) datasets

Threshold of EAD					
	endo cell (CiPAORd)		M cell (CiPAORd)		TdP risk
Drug	C1	C2	C1	C2	C1
Quinidine	1	1	1	1	High
Ibutilide	1	1	1	1	High
Azimilide	2	4	3	4	High
Bepridil	3	2	3	2	High
Dofetilide	3	3	6	4	High
Vandetanib	3	>70	3	4	High
Sotalol	4	5	5	6	High
Ranolazine	4	3	5	3	Low
Cisapride	5	2	7	3	Medium
Metoprolol	6	5	6	6	Low
Ondansetron	6	7	7	7	Medium
Astemizole	9	9	22	12	Medium
Droperidol	10	7	25	10	Medium
Chlorpromazine	14	35	21	27	Medium
Terfenadine	17	4	15	4	Medium
Tamoxifen	17	31	20	27	Low
Nitrendipine	>70	>70	>70	>70	Low
Pimozide	>70	4	20	3	Medium
Clozapine	>70	>70	23	>70	Medium
Risperidone	>70	>70	>70	>70	Medium
Clarithromycin	>70	>70	>70	>70	Medium
Diltiazem	>70	>70	>70	>70	Low
Mexiletine	>70	>70	>70	>70	Low
Verapamil	>70	>70	>70	>70	Low
Disopyramide	>70	>70	>70	>70	High
Loratadine	>70	>70	>70	>70	Low
Domperidone	>70	>70	>70	>70	Medium
Nifedipine	>70	>70	>70	>70	Low

TdP risk classification summary					
	Number of correctly classified drugs				Total number of Drugs
Category	C1	C2	C1	C2	
High	7 (4, 3)	6 (4, 2)	7 (4, 3)	7 (4, 3)	8 (4, 4)
Intermediate	4 (2, 2)	2 (1, 1)	6 (2, 4)	3 (1, 2)	11 (4, 7)
Low	6 (3, 3)	6 (3, 3)	6 (3, 3)	6 (3, 3)	9 (4, 5)
Total	17 (9, 8)	14 (8, 6)	19 (9, 10)	16 (8, 8)	28 (12, 16)

[C1] Drug-induced modulation of 9 parameters ($sbIKr$, Ku, V_{half}, $bINa$, $bINaL$, $bICaL$, $bIKs$, $bIK1$ and $bIto$) is considered.

[C2] Only drug-induced changes in $sbIKr$ and $bICaL$ is considered ($V_{half} = -100$, $Ku = 0.05$).

[Note] Numbers in parentheses are number of drug from training and validation set.

[Note] BCL fixed to 2000 for endo cell type and 700 for M cell type model under all tested conditions.

Logistic regression classifiers for EAD risk

Table 8. EAD risk discrimination accuracies for individual direct and derived features. Classification was performed on the virtual population of drugs generated for Sobol sensitivity analysis (n=22000 drugs). The virtual population of drugs considering fixed $sbIKr$ and BCL was also analyzed (n=18000 drugs). The numbers in the parentheses are for drug population generated at fixed $sbIKr$ and BCL.

	Predictive power				Feature Rank			
	endo cell		M cell		M cell	endo cell		
Parameter	Accuracy	AUC	Accuracy	AUC	Direct	Direct	Derived	Merged
$bICaL$	**0.78** (**0.81**)	**0.85** (**0.89**)	**0.74** (**0.87**)	**0.82** (**0.95**)	1 (1)	1 (1)	-	1 1 (1 1)
$bIKs$	0.62 (**0.67**)	**0.66** (**0.72**)	0.56 (0.59)	0.57 (0.62)	3 (2)	2 (2)	-	4 4 (3 4)
$sbIKr$	0.61 (-)	**0.65** (-)	**0.72** (-)	**0.80** (-)	2 (-)	3 (-)	-	10 2 (-)
$bIK1$	0.57 (0.6)	0.59 (0.63)	0.54 (0.55)	0.54 (0.57)	5 (3)	4 (3)	-	12 4 (3 4)
$bINaL$	0.53 (0.55)	0.53 (0.56)	0.54 (0.55)	0.55 (0.57)	6 (4)	5 (4)	-	13 4 (12 2)
V_{half}	0.52 (0.54)	0.52 (0.53)	0.54 (0.55)	0.54 (0.57)	4 (5)	6 (6)	-	14 4 (14 4)
$bIto$	0.52 (0.53)	0.50 (0.50)	0.53 (0.51)	0.51 (0.51)	10 (7)	7 (7)	-	15 4 (17 4)
Ku	0.52 (0.53)	0.49 (0.51)	0.53 (0.50)	0.51 (0.51)	9 (8)	8 (8)	-	16 4 (16 4)
$bINa$	0.52 (0.54)	0.52 (0.52)	0.53 (0.51)	0.52 (0.53)	7 (6)	9 (5)	-	18 3 (13 2)
BCL	0.51 (-)	0.53 (-)	0.53 (-)	0.52 (-)	8 (-)	10 (-)	-	19 1 (-)
$peakCa$	**0.73** (**0.76**)	**0.82** (**0.85**)	-	-	-	-	1 2 (5 1)	5 1
$qNet$	**0.68** (**0.68**)	**0.75** (**0.76**)	-	-	-	-	2 1 (8 3)	2 2 (10 2)
$peakVm$	**0.68** (**0.72**)	**0.76** (**0.80**)	-	-	-	-	3 3 (3 2)	3 3 (8 1)
$diastolicCa$	**0.70** (**0.81**)	**0.77** (**0.89**)	-	-	-	-	4 2 (1 1)	6 1 (5 1)
$APD90$	**0.68** (**0.76**)	**0.75** (**0.84**)	-	-	-	-	5 1 (2 3)	7 2 (2 2)
$CaTD50$	**0.65** (**0.74**)	**0.69** (**0.81**)	-	-	-	-	6 2 (6 1)	8 1 (7 1)
$APD50$	**0.65** (**0.69**)	**0.72** (**0.76**)	-	-	-	-	7 1 (7 3)	9 2 (9 2)
$CaTD90$	0.59 (**0.68**)	0.60 (**0.75**)	-	-	-	-	8 2 (9 1)	11 1 (11 1)
$restVm$	0.52 (0.52)	0.64 (0.56)	-	-	-	-	9 3 (4 4)	17 4 (15 3)

We tested the predictive capabilities of several of the previously proposed direct and derived TdP risk metrics for classifying the virtual drug population (generated for Sobol sensitivity analysis) into low and high risk categories for EAD generation. We used EADs as surrogate markers for TdP risk, as the actual torsadogenic risk associated with virtual drugs is unknown. We built logistic regression classifiers with the defined binary targets based on EAD risk. The accuracy of EAD risk classification was evaluated for the individual features by 10-fold cross-validation and are reported in terms of Jaccard index and area under the ROC curve (ROC AUC) (Table 8). The EAD risk classification results show that $bICaL$ had the highest predictive power for classifiers built on individual direct and derived features. This was in agreement with the Monte Carlo filtering analysis (see, section Regional sensitivity analysis) that predicted $bICaL$ to be the one of the most influential parameter for EAD regulation in the model. Surprisingly, the classifiers based on the derived features that are sensitive to drug-induced effects of combination of input parameters did not improve EAD risk classification performance. Among the derived features, $peakCa$ performed the best. Several derived features such as $APD90$, $qNet$, $peakVm$, and $diastolicCa$ had similar predictive power.

Feature ranking was implemented using an approach similar to [9]. Figure 7 shows the clustering map obtained using unsupervised hierarchical clustering on the correlation coefficient matrix of the features. The feature with the best ROC AUC when tested individually was ranked the highest (rank=1). Subsequent features were chosen with the next best ROC AUC score but were conditioned to being picked from outside of the cluster of the previously chosen features. Once a feature from each cluster was picked, the remaining features were ranked based on their ROC AUC scores. The obtained feature ranks are listed in Table 8. The cluster number of the feature is reported in blue in the feature rank

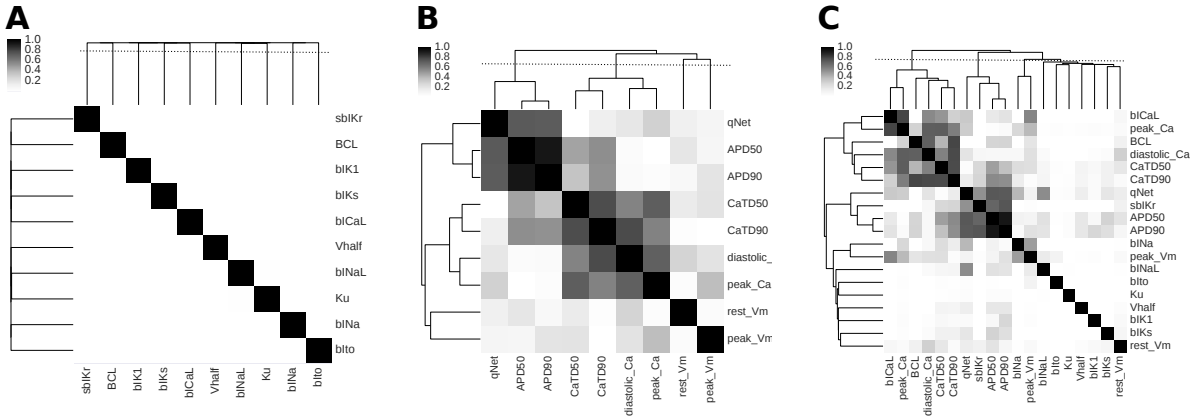

Fig 7. Hierarchical clustering based on correlation coefficient of the features. **A**: Direct features. **B**: Derived features. **C**: Merge set of direct and derived features.

column Table 8. The direct features are independent and uncorrelated to each other. Hence in the hierarchical clustering map of direct features, each feature appears in a separate group (Figure 7 A). The cluster number is the same as the feature rank when only the direct features are considered, and hence were not reported in the Table 8.

Discussion

Here we presented a study that applied GSA within the context of *in silico* prediction of pharmacological toxicity. The target of GSA is the latest version of the *in silico* model of an isolated cardiac cell [11], CiPAORd, which was developed under the CiPA initiative and incorporates dynamic hERG-drug interactions [10]. Our analysis explored the effects of a large population of virtual drugs on the seven major cardiac ion-channel currents thought to be important in regulation of TdP. GSA provided a systematic understanding of the model input/output relationships and allowed for the identification of the most influential parameters that regulate model-derived features of interest. In the future, the knowledge gained from GSA could help improve the model structure and potentially guide the design of novel effective experiments to increase reliability of model-driven predicted risk.

GSA of APD_{90}, $qNet$, $peakCa$, and $EADs$

Different methods and tools, each with their own advantages and disadvantages, allow for the analysis of the sensitivity of complex model systems to the input parameters (e.g, refer to [24, 28, 34] for thorough reviews on the subject). Unlike local sensitivity analysis, which quantifies the impact on the model outputs of small (i.e., local) input perturbations around nominal values, GSA measures the effects of global variations over the entire input parameter space. In this work, we used two GSA methods, the difference-based Morris sensitivity method [23], and the Sobol's variance-based sensitivity method [24, 25], to rank cardiac ion-channel currents based on their relative contributions to output variability in the model. We also performed regional sensitivity analysis [24, 26] to determine the cardiac ion-channels responsible for regulation of the EAD generation in the CiPAORd models.

Both Sobol's variance analysis and Morris method results indicate that individual model-derived features are sensitive to different inputs (see Figure 1, 2, and data in the Supplemental Material). More specifically, features derived from calcium transients are most sensitive to the $bICaL$ parameter, while the $qNet$ metric is most sensitive to the modulations in sodium currents and to the $sbIKr$ parameter. $sbIKr$ is the most influential parameter regulating $APD90$. In spite of these observed dissimilarities, different derived metrics have been reported to perform well on particular *in vitro* drug effects datasets. $APD90$ [6], a metric based on $APD50$ and $diastolicCa$ [9], and a metric based on EADs [7], have been shown to provide good risk discrimination of drugs considering *in vitro* measurements reported in [6]. We have shown previously that different derived features extracted from the original ORd model [19] show similar performance in TdP risk discrimination [14] when tested on a combination of several datasets. The similarity in performance of different metrics might be due to the presence of estimates of drug effects on only three channel currents (i.e., fast sodium current, L-type calcium channel current, and hERG current) in the majority of the datasets, the small size of the datasets, and the differences in structure of the myocyte models used for obtaining the derived feature. Different derived metrics, such as $APD50$, $APD90$, $peakCa$, and $CaD90$ have been shown to provide the best classification when varying the computational model of interest [6].

It should be noted that GSA results are highly dependent on explored parametric space. Here, we evaluate the sensitivity over the 10-D input space comprising parameters of seven major cardiac channel currents that are thought to play an important role in determining the risk of TdP. However, the actual drugs might lie within a very small subspace of the explored hyperspace. Visualization of the blocks of different ion channel currents for the 28 CiPA drugs (Figure 8) reveals that majority of the drugs do not result in block of $IK1$, INa, and IKs currents. Constraining the input parameter space based on prior information and performing systematic sensitivity analysis across different models can allow for better understanding of performance of particular derived features. Certain derived metrics may perform superior to others in the presence of data on drug effects on multiple ion-channels. For example, Figure 8 demonstrates

that accurate classification of the Ranolazine drug to low risk category requires a feature that is at least moderately sensitive to variations in block of late sodium current, since the drug appears to be a pure hERG and sodium channel blocker. Our GSA results (Figures 1 and 2) point to $qNet$ as a candidate, as it is the only feature among the tested derived features that is highly sensitive to block of late sodium current and block of hERG. $qNet$ has been observed to outperform other standard derived features on the 28 CiPA drugs [10–12].

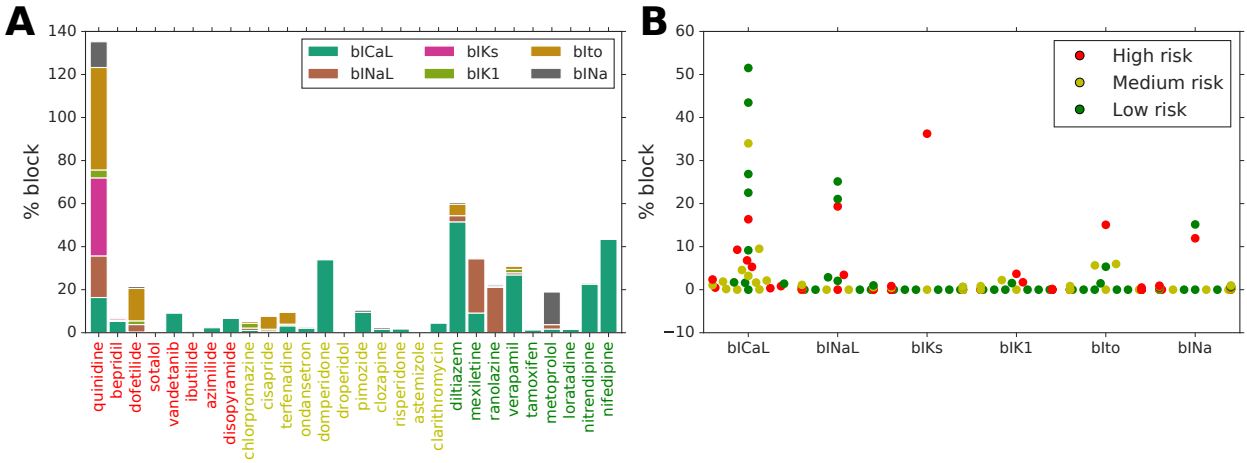

Fig 8. Drug-induced block of non-hERG ion channels for 28 CiPA compounds at their EFTPC based on the measurements from the *in vitro* assay [12, 35]. **A**: Stacked bar chart of six ion channel current block values for each of the 28 drugs.**B**: A swarm plot of block values of six ion channel currents categorized into high, medium, and low risk groups.

The results of our Monte Carlo filtering analysis (Table 5 and Figure 4) indicate that the generation of EADs, which are thought to be cellular precursors of TdP genesis [20], are most sensitive to variations in block of $ICaL$ and to the static component of the hERG channel current in the CiPAORd model. Variations in blocks of the IKs and $IK1$ currents have a moderate influence on the genesis of EADs. Drug effects on the Ito, INa have the least influence on EAD generation (Table 5, Figure 4). Our analysis shows that several of the derived features proposed previously are sensitive to different model inputs than those for EADs (see Figure 1). For example, the $qNet$ metric is sensitive to variations in block of sodium currents and block of $sbIKr$ for the endo cell model. In contrast, the $bICaL$ and $sbIKr$ parameters are the most influential for regulating EADs.

Surprisingly, the recently introduced dynamic-hERG block parameters V_{half} and Ku [10], which are measured using challenging experimental protocols [36, 37], show minor influence on EAD genesis (Table 5) when compared to other tested inputs. These parameters have relatively small contribution to the variance of all the tested derived metrics compared to other influential input parameters (Figure 1, 2, and data in the Supplemental Material). In cases where the majority of the primary regulating parameters are similar between drugs, accounting for changes in the modestly influential parameters can allow for improved predictions. We observe that a surrogate of $qNet$ estimated by inclusion of the V_{half} parameter allows for improved prediction of few intermediate risk drugs (Figure 5). Similarly, when classifying CiPA drugs based on EADs, we observe that prediction improves by correctly classifying 3 more drugs when accounting for drug-induced effects of other parameters in addition to the $sbIKr$ and $bICaL$ parameters (Table 7). However, our results also point towards the important consideration that errors in measuring the most influential parameters regulating a particular metric have a bigger impact on the predicted classification compared to neglecting some of the less influential parameters. GSA allows us to determine and rank most of the critcal model components.

GSA vs. uncertainty analysis

Uncertainties in *in vitro* measurements of drug-induced effects on ionic currents present an important concern in evaluating the torsadogenic risk of compounds by interrogating *in silico* biophysical models. Discrepancies in estimates for model parameters based on available *in vitro* assay data have been recently highlighted in uncertainty quantification studies [38, 39]. High uncertainty in model parameters leads to low confidence in model predicted risk, and thus, not surprisingly, risk stratification of the CiPA training drugs proved to be unreliable especially at high drug concentrations [39], where model parameter estimates are inherently less accurate. However, it is important to emphasize that the relative contributions of drug-induced modulation of ion-channels on output features differ significantly. Uncertainties in model input parameters that are higly influential (e.g., as revealed by GSA) result, therefore, in lower confidence in the predicted risk, while errors in estimating less influential model parameters are better tolerated by risk measures [40].

GSA vs. linear regression

The link between GSA and feature selection is well illustrated in the following example. We consider that some model-derived metric M can be well approximated by a multivariatie quadratic polynomial, with direct features as input variables of the polynomial. The equation of the polynomial is

$$M = c + \sum_i a_i x_i + \sum_{ij} b_{ij} x_i x_j, \tag{14}$$

where $\{x_i\}$ is a vector of direct features; $\{a_i\}$ and $\{b_{ij}\}$ are a vector and a symmetric matrix of coefficients for the linear and quadratic terms of the poylnomial, respectively, and c is a constant. The components of $\{x_i\}$, i.e., the individual direct features, are assumed to be uniformly distributed with expectation e_i and variance v_i. In this case, the effect of each direct feature x_i and the joint effect of direct features $x_i x_j$ on variance in M are given by

$$V_i = (a_i + \sum_{\substack{k \\ k \neq i}} 2b_{ik} e_i)^2 v_i + b_{ii}^2 var(x_i^2) + 2b_{ii}(a_i + \sum_{\substack{k \\ k \neq i}} 2b_{ik} e_i) cov(x_i, x_i^2), \tag{15}$$

$$V_{ij} = 4b_{ij}^2 \left(v_i v_j + e_j^2 v_i + v_j e_i^2 - cov(x_i x_j, x_i) e_j - cov(x_j x_i, x_j) e_i \right). \tag{16}$$

Typically, to rank the features based on their importance, feature scaling is carried out on the inputs before performing the regression. If each of the uniformly distributed direct features x_i are centered and normalized to have $e_i = 0$ and $v_i = 1$, then

$$V_i = a_i^2 + \frac{4}{5} b_{ii}^2, \tag{17}$$

$$V_{ij} = 4b_{ij}^2. \tag{18}$$

Hence, the first- and second-order Sobol indices can be represented as

$$S1_i = \frac{a_i^2 + \frac{4}{5}b_{ii}^2}{var(M)}, \tag{19}$$

$$S2_{ij} = \frac{4b_{ij}^2}{var(M)}. \tag{20}$$

Expressions in (19)-(20) provide a link between calculated variances in the Sobol sensitivity method and the coefficients of the polynomial. Indeed, to demonstrate these relationships, we fitted a linear regression to approximate model-derived metrics ($APD90$, $qNet$, and $peakCa$) accounting for only the first-order effects of the direct features, i.e., $M = c + \sum a_i x_i$. Then, the squared coefficients of the regression are normalized by the variance of derived features and compared with Sobol $S1$ index. As expected, in the presence of minor interactions and nonlinear effects for these derived metrics (Table 4), we observe a good match between indices estimated from the regression coefficients and the Sobol $S1$ indices as shown in Figure 9, even when considering only the first-order effects to construct the regression.

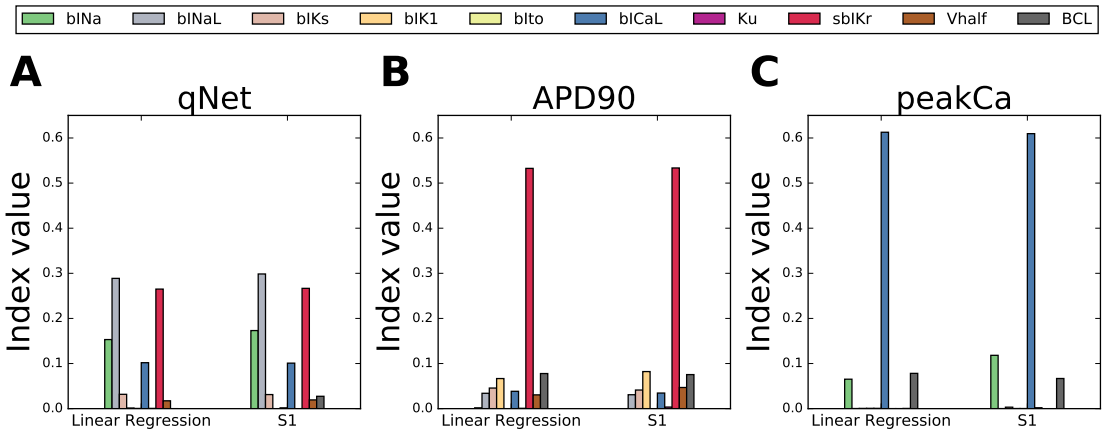

Fig 9. Comparision of index estimated from linear regression coefficients and Sobol first-order index $S1$ for **A:** $qNet$, **B:** $APD90$, and **C:** $peakCa$.

Classification from direct vs. derived features

Model-derived features attempt to capture aspects of the underlying physiological phenomena such as APD prolongation, increase in calcium levels, etc. Being biophysically motivated, classifiers of risk based on model-derived features are thought to allow generalizable assessments also in cases where the training datasets are small and hence the effects on targets of interest might need to be extrapolated. Different model-derived features models have shown promising performance as candidate metrics for risk classification depending on the training datset [6–11, 17, 18]. However, while biophysical models provide a mechanistic understanding, the equivalent predictive capability of direct features that do not require modeling have raised concerns on whether the added complexity introduced by modeling is in fact necessary for risk classification purposes [15]. We have recently shown comparable accuracies for TdP risk classification from direct features in tests against several of the other published methods [14]. Few previous studies [13, 15, 16] have also reported similar findings for particular datasets where metrics based on direct features provided high predictive power.

Classifiers of TdP risk can be categorized into two classes according to the basic metrics underlying their definition. A first class of metrics aims at achieving enhanced performance by combining (e.g., via PCA) several model-derived metrics,

each of which provides a biophysical explanation for its associated TdP risk. While promising in principle, such an approach cannot yield, however, metrics that outperform significantly a direct model-agnostic classifiers. As indicated by our GSA, in fact, the variance of popular model-derived features is only slightly affected by the effects of model input interactions and nonlinearities (Figure 2). Subsequently, model-derived features and, hence, classifier metrics constructed as their combinations can be well recapitulated by surrogates that account only for linear transformations of the model inputs. It is not surprising, then, that the classifications performance of combining derived metrics is very similar to what can achieved by direct classifiers. A second class of metrics optimally designed to perform classification based on a single model-derived feature, transforming direct measurements from *in vitro* assays. Feature selection occurs on the basis of a (potentially automized) trial and error process, in which several "randomly designed" metrics are tested. In section Randomly generated linear metrics, we reproduce such process for linear combinations of model inputs to demonstrate that the probability of designing a metric that could perform at least as well as $qNet$ on the training set is about 1% over reasonable bounds for the admissible combination space. Therefore, if 100 metrics in total are tested, there is an expected probability to obtain one that performs as well as $qNet$. However, the probability that a randomly generated metric would also perform satisfactorily on the 28-drug merged dataset from the CiPA initiative is very low, with only 0.2% of the tested metrics being able to classify correctly more than 23 compounds. Our overall analysis thus supports $qNet$ as a metric for TdP risk classification.

Classification of EAD risk, and TDP risk based on EADs

Several derived metrics show similar performance in classification of virtual drugs into low and high risk categories, based on their EAD risk. However, none of the derived features succeeded in perfectly discriminating the set of virtual drugs into a group of EAD generating drugs and a group of drugs that do not produce EADs (Table 8). This is in agreement with our GSA results, with none of the tested derived features demonstrating identical sensitivity profile to EAD (Figure 1, 2 and the Supplemental material). $qNet$ has been shown to provide excellent classification of drugs in the CiPA training and validation data, a result thought to be linked to EAD generation [11, 12]. However, we observe that the categorization of CiPA drugs based on analysis of EADs was not as predictive (Table 7) as model-derived metric such as $qNet$ (Figure 5). Our GSA indicates that $qNet$ and EAD do not fully share sensitivity profiles (Figure 1, 2). Hence, the previously reported correlation between $qNet$ and EAD generation [11] seems to be not justified. The efficacy of $qNet$ seem to rather arise from its dependence on both the blocks of late sodium current and of the hERG channel. One possible speculation for improved predictive power of $qNet$ for drugs like Ranolazine might be the reduced transmural dispersion, which can be accounted for by the block of late sodium current, and hence $qNet$. There can be several other possible explanations for poor performance of EAD metric compared to $qNet$ such as inaccurate reproduction of EADs in the current model, small size of the tested datasets, biases in the target, and the need to test EADs on coupled cells/tissue models etc.

Conclusion/Summary

The proarrhythmic risk assessment based on simulated drug responses in isolated cell model [6–12, 18], tissue models [41, 42] or organ level computational models [17, 43] can provide physiological/mechanistic insights and interpretability. Moreover, *in silico* models serve as an excellent tool for evaluation of drug safety in diseased conditions [41, 42]. Surprisingly, the present analysis reveals, however, that several of currently proposed model-derived metrics for TdP risk classification can be well captured as linear combination of the set of direct features. The analysis also highlights the key differences in the most influential cardiac ion-channels affecting the different model-derived metrics

and pointing towards a need for better mechanistic understanding of a promising metric such as $qNet$. In summary, the study emphasizes the role of tools such as GSA for systematic understanding of the model structure for improved proarrythmic risk assessment. Sensitivity analysis can help identify and focus on the critical model components that drive the major variations in the desired output metrics. This can ultimately allow for refinement of the biophysical models, guidance of future experiments and improvement in model-driven TdP risk predictions.

Supplementary Material

CiPA training and validations compounds dataset

IC_{50} values and Hill coefficients describing the effects of the CiPA training drugs on non-hERG channels were extracted from [10]. Similarly, median values from the manual patch clamp experiments on the CiPA validation drugs were taken from [12]. The values are reported in Table 9.

Table 9. IC_{50} values (nM) and Hill coefficients for non-hERG channels

	ICaL_IC50	ICaL_h	IK1_IC50	IK1_h	IKs_IC50	IKs_h	INa_IC50	INa_h	INaL_IC50	INaL_h	Ito_IC50	Ito_h	hERG_IC50	hERG_h
astemizole	553.0	1.2000	-	-	-	-	5.410000e+03	0.7600	10300.0	2.3000	-	-	-	-
azimilide	13200.0	0.7100	-	-	-	-	3.630000e+05	0.7200	2940000.0	0.4700	-	-	-	-
bepridil	2808.0	0.6486	-	-	28630.0	0.7061	2.929000e+03	1.1640	1814.0	1.4160	8.594000e+03	3.5410	51.340	0.9293
chlorpromazine	8192.0	0.8441	9270.0	0.6878	-	-	4.536000e+03	1.9950	4560.0	0.9379	1.762000e+07	0.3654	975.200	0.8281
cisapride	9267000.0	0.4261	29480.0	0.5133	81170000.0	0.2921	-	-	-	-	2.191000e+05	0.2430	11.270	0.6210
clarithromycin	38100.0	0.8800	-	-	-	-	1.090000e+06	0.8900	1810000.0	3.0000	-	-	-	-
clozapine	5490.0	0.9400	-	-	-	-	2.570000e+05	0.5900	73600.0	2.0000	-	-	-	-
diltiazem	112.1	0.7142	-	-	-	-	1.109000e+05	0.7022	21870.0	0.6779	2.822000e+09	0.1696	13140.000	0.9119
disopyramide	32900.0	0.6900	-	-	-	-	1.920000e+05	1.3000	377000.0	2.1000	-	-	-	-
dofetilide	260.3	1.1630	394.4	0.7650	-	-	3.805000e+02	0.8920	753100.0	0.2597	1.882000e+01	0.7712	6.125	1.0790
domperidone	73.6	0.4900	-	-	-	-	4.190000e+04	1.5000	225000.0	2.1000	-	-	-	-
droperidol	3230.0	1.2000	-	-	-	-	3.660000e+04	2.5000	33800.0	2.8000	-	-	-	-
ibutilide	37000.0	0.8600	-	-	-	-	2.410000e+04	2.3000	287000.0	2.4000	-	-	-	-
loratadine	703.0	0.5600	-	-	-	-	1.130000e+05	1.4000	192000.0	2.1000	-	-	-	-
metoprolol	3280000.0	0.5400	-	-	-	-	3.030000e+04	0.6100	630000.0	0.6600	-	-	-	-
mexiletine	38240.0	1.0310	-	-	-	-	-	-	8957.0	1.4090	-	-	29070.000	0.8928
nifedipine	11.4	0.6700	-	-	-	-	2.760000e+04	1.1000	45600.0	4.5000	-	-	-	-
nitrendipine	35.7	0.5000	-	-	-	-	2.240000e+04	0.5800	70700.0	3.2000	-	-	-	-
ondansetron	22550.0	0.7526	-	-	569800.0	0.6535	5.767000e+04	1.0200	19180.0	1.0350	1.023000e+06	0.9891	1325.000	0.9210
pimozide	64.5	0.4500	-	-	-	-	1.020000e+04	0.4600	1910.0	1.9100	-	-	-	-
quinidine	51590.0	0.5892	39590000.0	0.3468	4899.0	1.3630	1.233000e+04	1.4940	9417.0	1.3370	3.487000e+03	1.2820	986.300	0.8404
ranolazine	-	-	-	-	36160000.0	0.5191	6.877000e+04	1.4250	7884.0	0.9450	-	-	8208.000	0.8576
risperidone	1470.0	0.5900	-	-	-	-	5.340000e+05	0.7100	11400000.0	5.8000	-	-	-	-
sotalol	7062000.0	0.8651	3050000.0	1.2040	4222000.0	1.1670	1.144000e+09	0.5089	-	-	4.314000e+07	0.6632	107100.000	0.7850
tamoxifen	5720.0	0.7600	-	-	-	-	8.400000e+04	0.8200	3640000.0	4.0000	-	-	-	-
terfenadine	700.4	0.6601	-	-	399800.0	0.5430	4.803000e+03	1.0150	20060.0	0.6011	2.400000e+05	0.2559	20.400	0.6118
vandetanib	6060.0	0.7200	-	-	-	-	8.090000e+04	1.9000	3790000.0	0.8200	-	-	-	-
verapamil	201.8	1.0970	348800000.0	0.2728	-	-	-	-	7028.0	1.0310	1.343000e+04	0.8222	295.600	0.9378

hERG dynamic data was extracted from [10] and [12] for the training and validation compounds, respectively, as listed in Table 10. In the simulations, the reported median values of parameters were used.

Table 10. Drug-hERG binding dynamic parameters.

	Kmax	Ku	n	halfmax	Vhalf
astemizole	2.420	0.000033	1.4490	4.883000e+00	-6.110
azimilide	654000.000	0.008250	0.6028	1.413000e+07	-8.821
bepridil	5594000.000	0.000172	0.9374	1.472000e+08	-61.340
chlorpromazine	157900.000	0.046710	0.8871	4.351000e+07	-14.450
cisapride	10.220	0.000416	0.9615	4.232000e+01	-167.400
clarithromycin	92.890	0.012160	0.7867	2.229000e+05	-102.800
clozapine	7.486	0.029890	1.3670	9.048000e+04	-8.810
diltiazem	182500.000	0.282000	0.9382	6.677000e+08	-90.650
disopyramide	3.685	0.121600	0.7894	4.473000e+04	-78.110
dofetilide	35.100	0.000018	1.0800	2.166000e+02	-1.000
domperidone	3.339	0.000356	0.7026	1.609000e+01	-65.650
droperidol	14.210	0.001256	0.5780	1.165000e+02	-78.680
ibutilide	14.570	0.000061	0.9231	3.863000e+01	-9.771
loratadine	376500.000	0.009642	0.8368	4.770000e+08	-1.000
metoprolol	31790.000	0.850800	0.8110	4.223000e+07	-89.350
mexiletine	15.000	0.071140	1.1390	7.230000e+05	-87.510
nifedipine	4.748	0.991600	1.2350	9.752000e+08	-87.370
nitrendipine	1.713	0.981900	1.9230	5.669000e+08	-61.580
ondansetron	172000.000	0.023240	0.8910	5.224000e+07	-82.200
pimozide	10.070	0.000046	0.8714	5.601000e+00	-158.500
quinidine	275.700	0.004103	0.8488	5.383000e+04	-61.350
ranolazine	52.840	0.020350	0.9532	1.430000e+05	-94.990
risperidone	3.930	0.001151	1.1220	7.528000e+02	-80.430
sotalol	96190.000	0.022250	0.7513	3.856000e+08	-51.500
tamoxifen	3.900	0.011750	2.0000	4.067000e+05	-2.036
terfenadine	102200.000	0.000078	0.6502	4.095000e+05	-81.630
vandetanib	36.280	0.019740	0.7126	2.223000e+03	-48.550
verapamil	1694000.000	0.000816	1.0430	3.356000e+08	-97.080

The known TdP categories and maximum effective free therapeutic concentrations for the validation drugs ([10, 12]) are listed in Table 11.

Table 11. Drug Therapeutic concentrations and risk

Drug	EFTPC (nM)	Risk
astemizole	0.260	Medium
azimilide	70.000	High
bepridil	33.000	High
chlorpromazine	38.000	Medium
cisapride	2.600	Medium
clarithromycin	1206.000	Medium
clozapine	71.000	Medium
diltiazem	122.000	Low
disopyramide	742.000	High
dofetilide	2.000	High
domperidone	19.000	Medium
droperidol	6.330	Medium
ibutilide	140.000	High
loratadine	0.450	Low
metoprolol	1800.000	Low
mexiletine	4129.000	Low
nifedipine	7.700	Low
nitrendipine	3.020	Low
ondansetron	139.000	Medium
pimozide	0.431	Medium
quinidine	3237.000	High
ranolazine	1948.200	Low
risperidone	1.810	Medium
sotalol	14690.000	High
tamoxifen	21.000	Low
terfenadine	4.000	Medium
vandetanib	255.400	High
verapamil	81.000	Low

Global sensitivity analysis results

Visulation of changes in the model-derived metrics against individual input parameters

The figures (Figure 10, 11, 12) show values of the derived features extracted from the simulation of virtual drugs generated for the Sobol sensitivity analysis against individual input parameters.

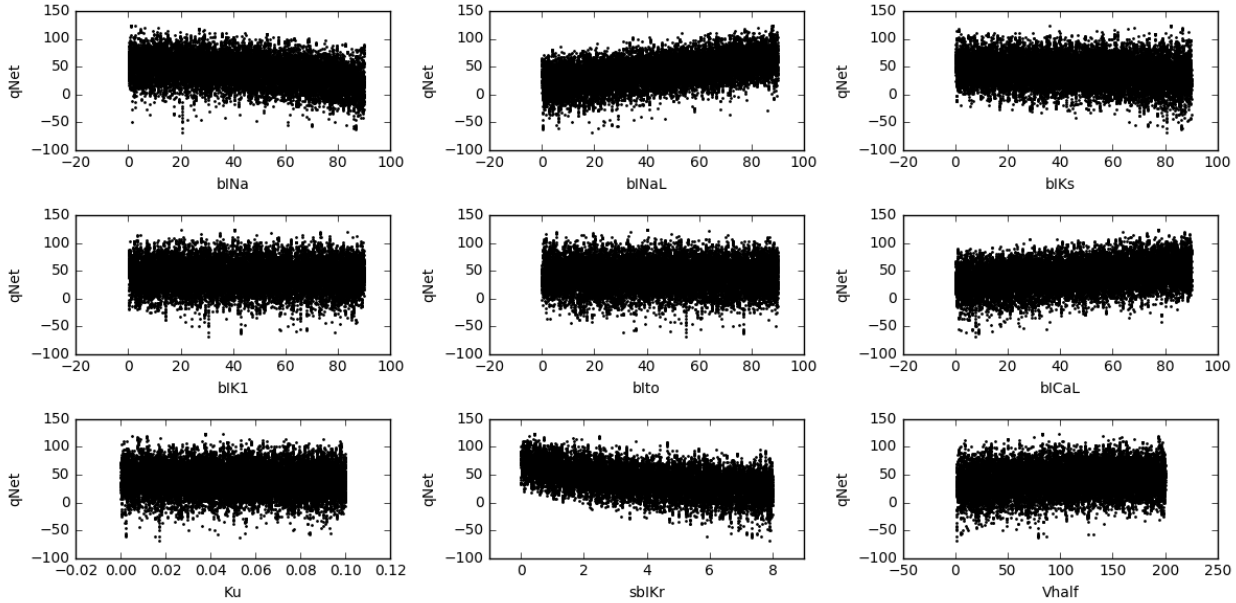

Fig 10. Scatter plot of $qNet$ verus different direct features for the 22000 simulated virtual drugs (endo cell model)

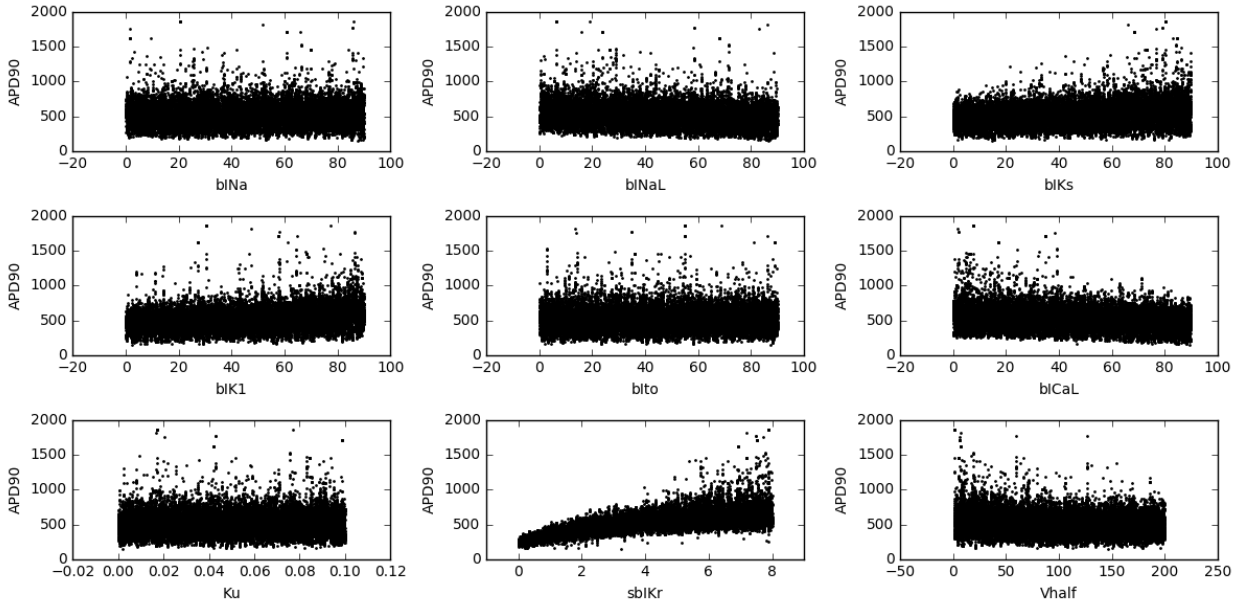

Fig 11. Scatter plot of $APD90$ versus different direct features for the 22000 simulated virtual drugs (endo cell model)

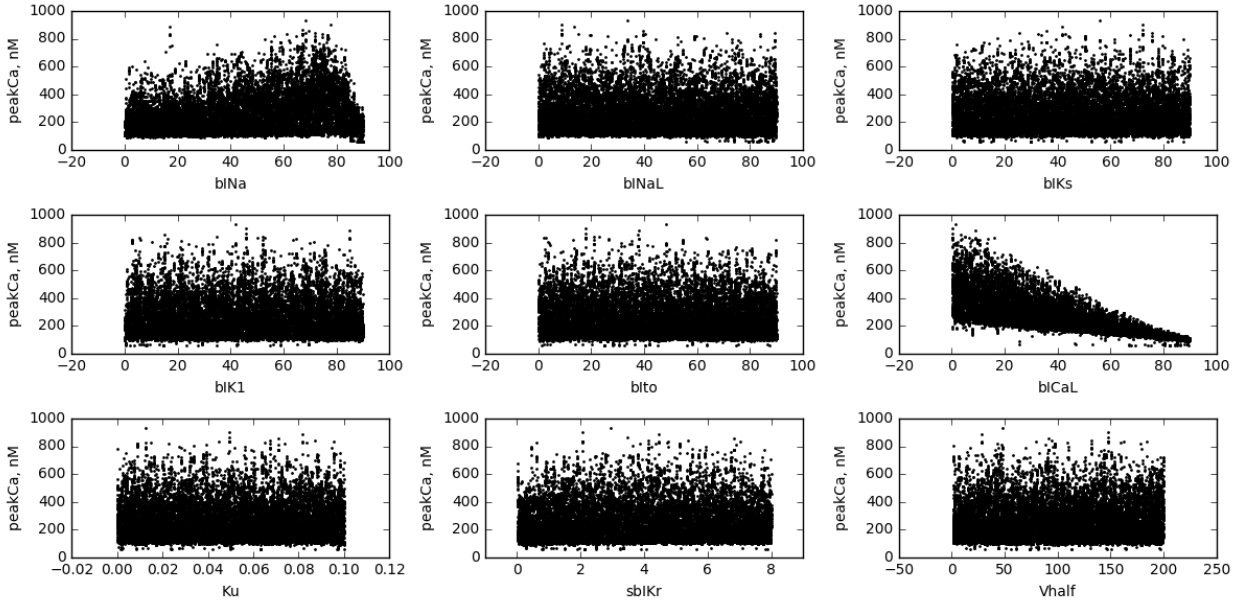

Fig 12. Scatter plot of *peakCa* versus different direct features for the 22000 simulated virtual drugs (endo cell model)

Morris method indices

The $\mu*$ and σ statistics were calcuated using Morris method to assess the influence of the blocks of different cardiac ion-channels on various model-derived features. The manuscript reports statistics for $qNet$, $APD90$ and $peakCa$ output metrics extracted from simulations of the endo cell model. Figures 13, 14, and 15 show the obtained values of $\mu*$ and σ for eight derived features including $qNet$, $APD90$ and $peakCa$ extracted from the endo, M, and epi cell type models, respectively. The μ^* statistic allows to rank the input parameters based on their influence, and the σ helps identify nonlinear and/or interaction effects.

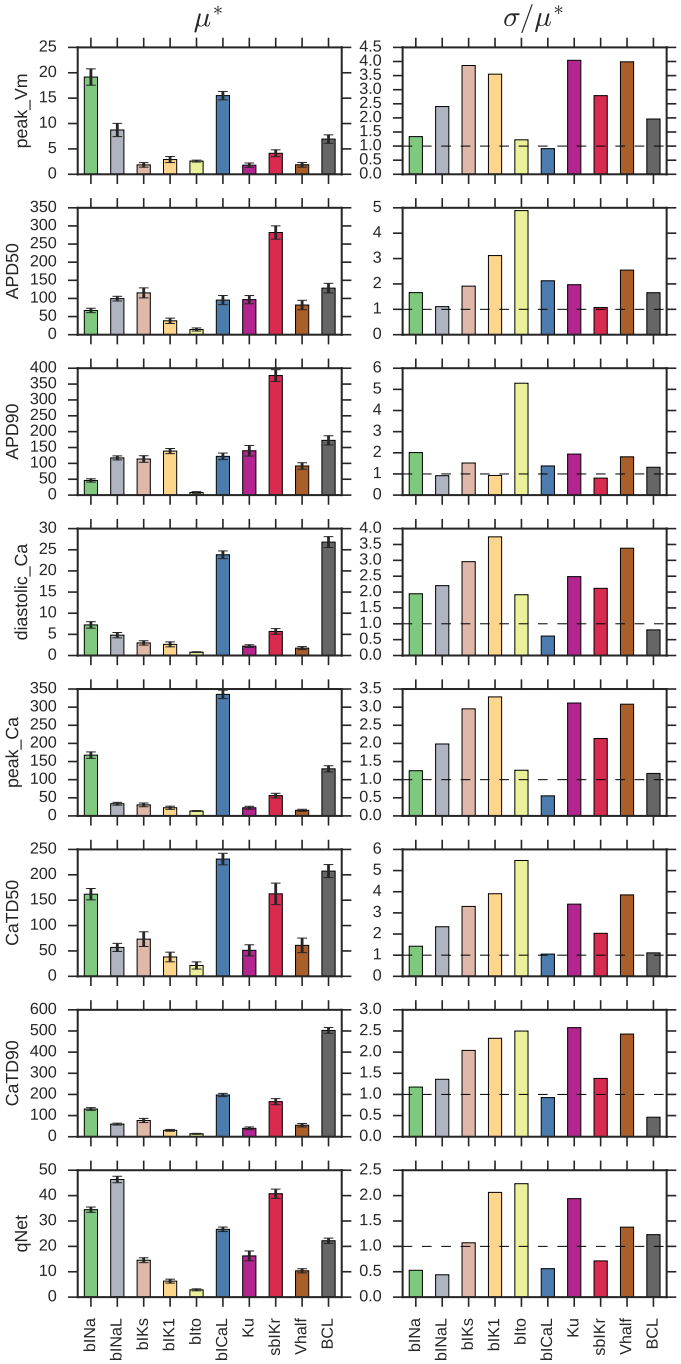

Fig 13. Plot of estimators $\mu*$ and σ, providing information on influence of input parameters on eight model-derived features (endo cell model)

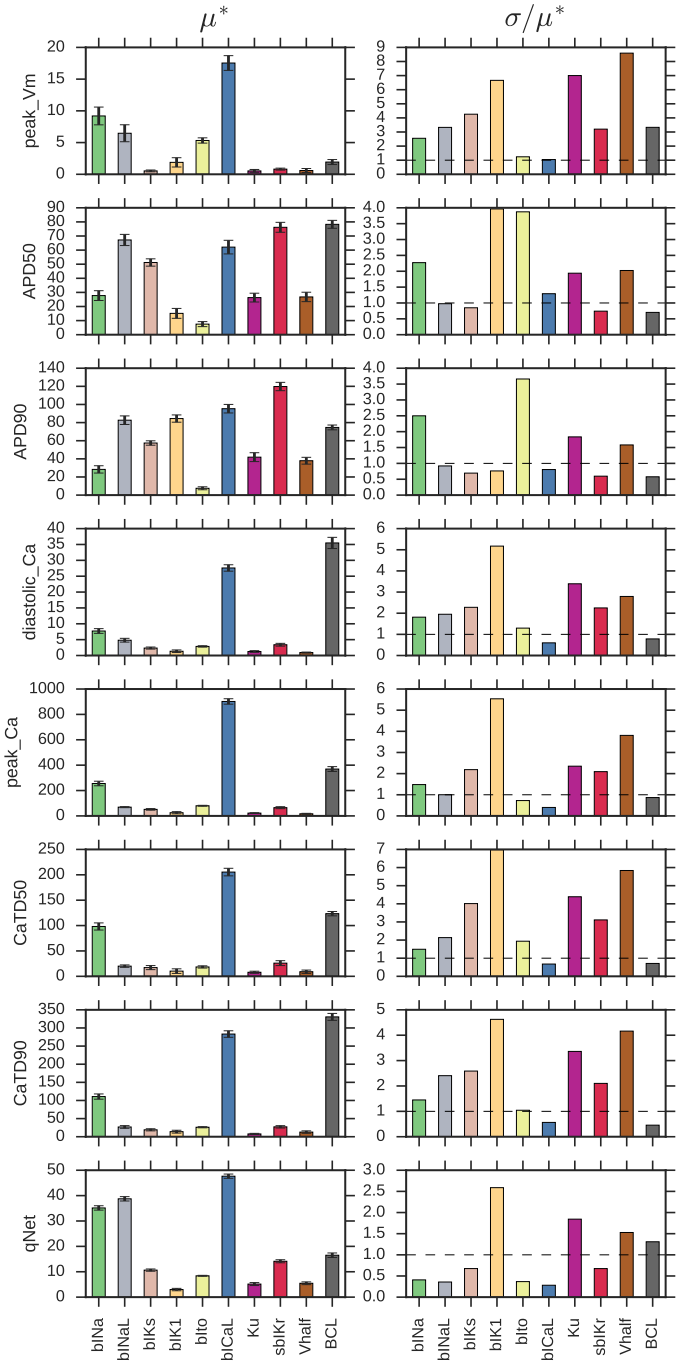

Fig 14. Plot of estimators $\mu*$ and σ, providing information on influence of input parameters on eight model-derived features (M cell model)

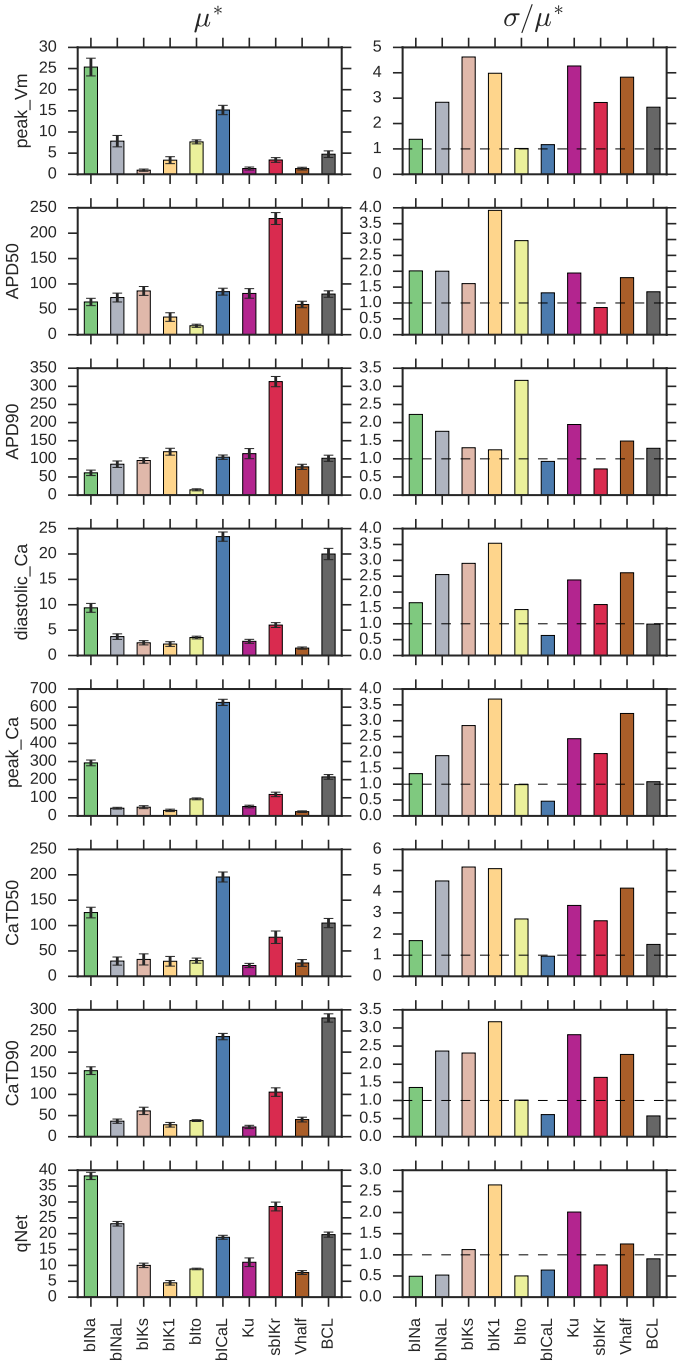

Fig 15. Plot of estimators $\mu*$ and σ, providing information on influence of input parameters on eight model-derived features (epi cell model)

Sobol sensitivity indices

Similar to the Morris method analysis, the manuscript reports first-order ($S1$) and total sensitivity indices (ST) for the $qNet$, $APD90$ and $peakCa$ output metrics extracted from the endo cell model. Figures 16, 17, and 18 show the estimated values of the Sobol indices ($S1$ and ST) for eight derived features extracted from the endo, M, and epi cell type models, respectively.

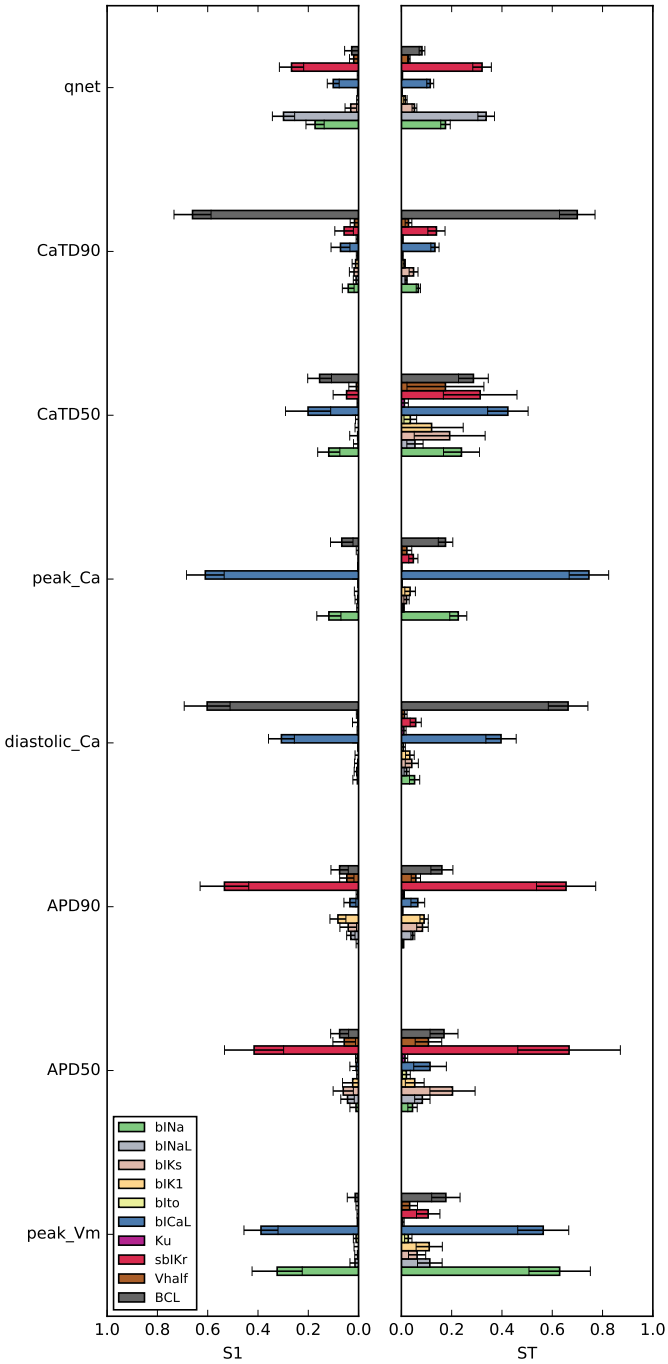

Fig 16. Plot of estimated first-order and total Sobol sensitivity indices (S1 and ST) to evaluate relative contributions of different input parameters on output variablity of 8 model-derived metrics (endo cell model)

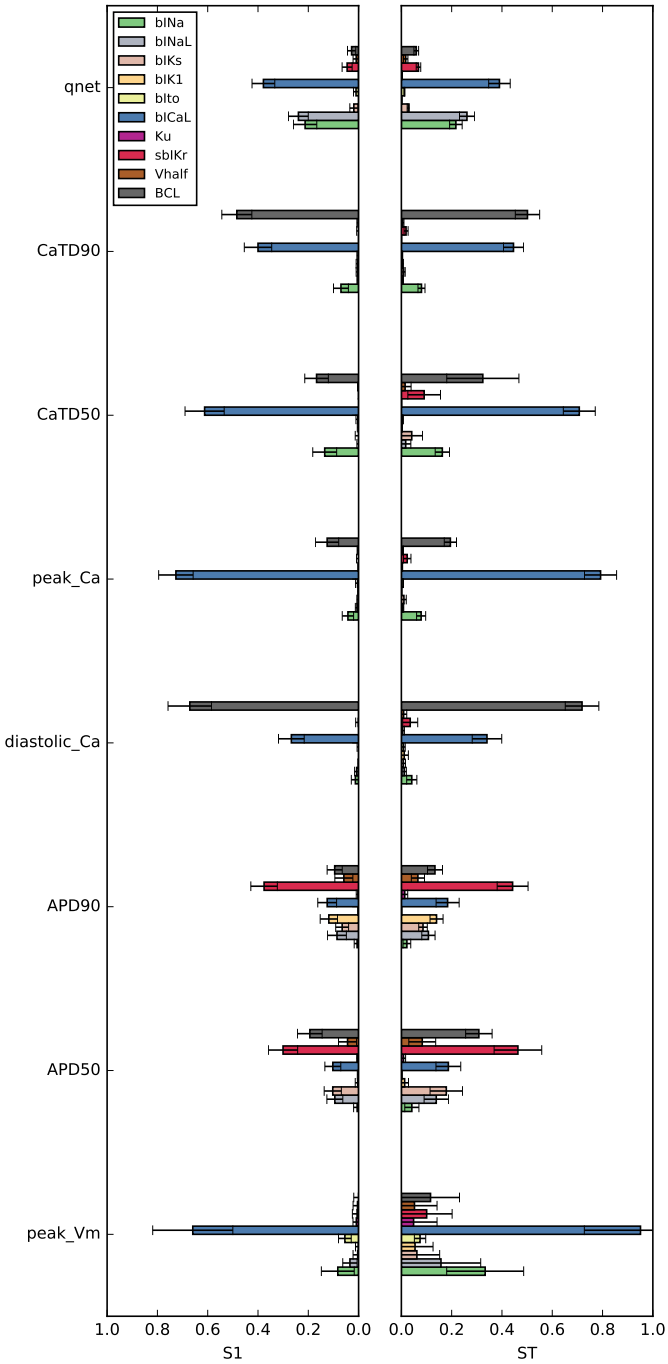

Fig 17. Plot of estimated first-order and total Sobol sensitivity indices (S1 and ST) to evaluate relative contribution of different input parameters on output variablity of 8 model-derived metrics (M cell model)

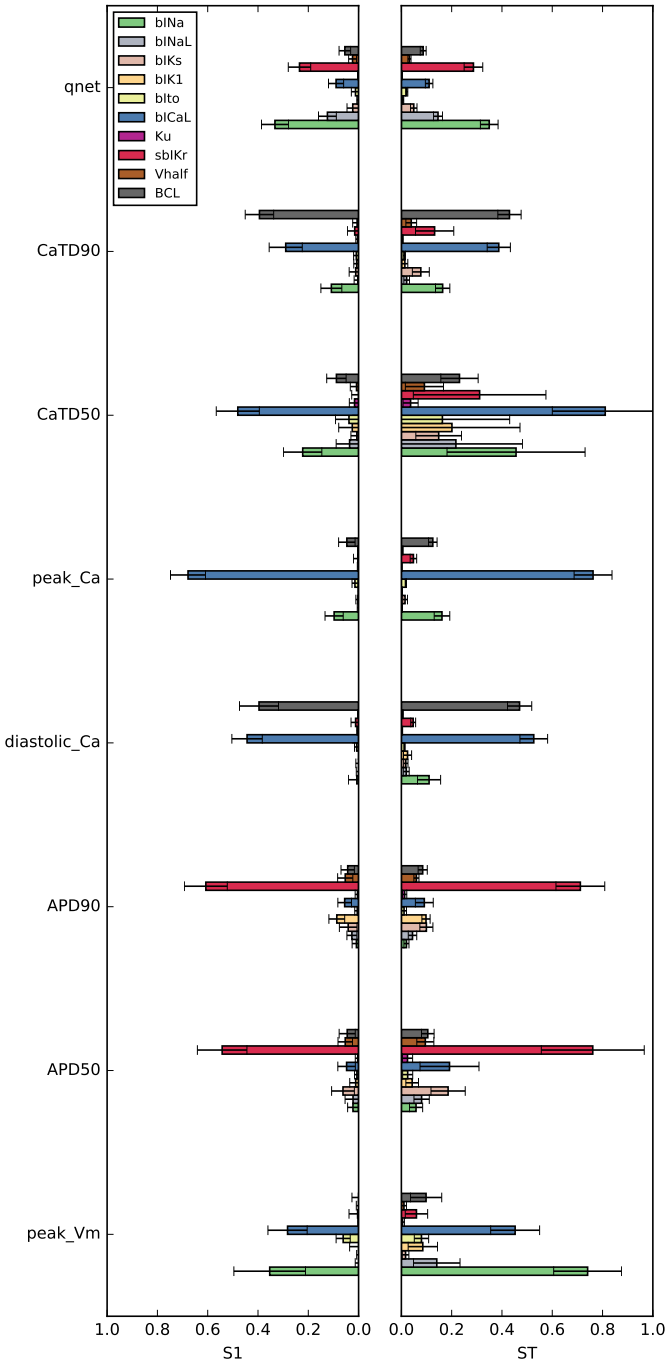

Fig 18. Plot of estimated first-order and total Sobol sensitivity indices (S1 and ST) to evaluate relative contribution of different input parameters on output variablity of 8 model-derived metrics (epi cell model)

TdP risk classification of "CiPA drugs" based on transmural dispersion of repolarization

The "CiPA drugs" were also calssified based on the transmural dispersion of repolarizations as shown in Figure 19.

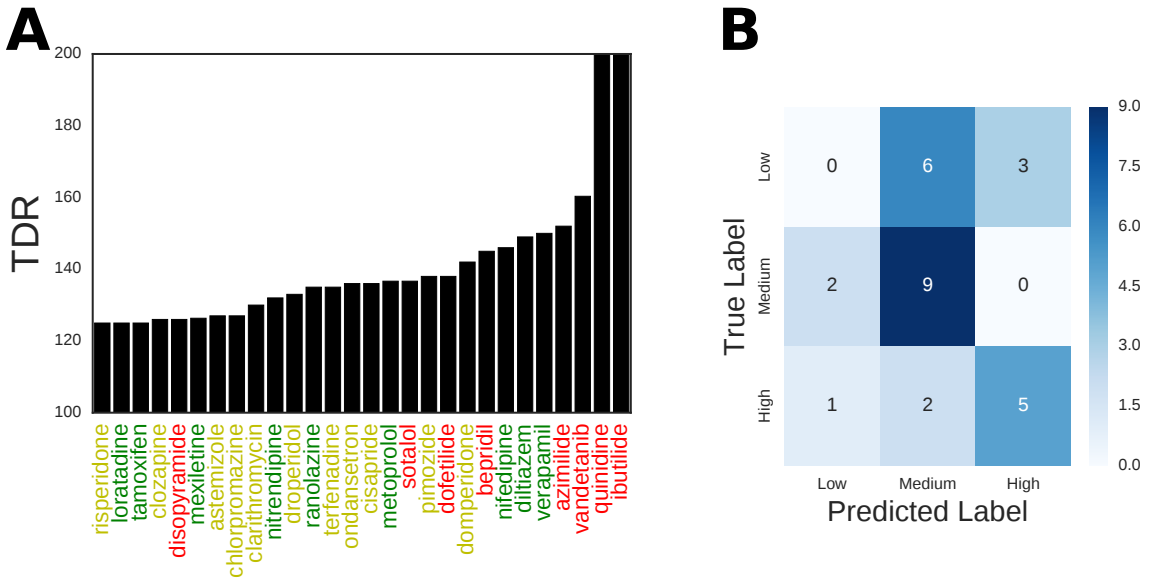

Fig 19. Risk categorization of "CiPA drugs" based on transmular dispersion of repolarization (TDR). **A:** TDR values for the 28 CiPA training and validations compound at (2x Cmax drug concentrations) are estimated from simulations of the drugs at pacing rate of 2000 ms in the CiPAOrd model as difference between the $APD90$ values between the epi and the M cell type. **B:** Concordance analysis of TDR-based risk categorization of the 28 drugs against the actual torsadogenic risk of the compounds reported as confusion matrix.

Examination of pause-induced EADs in the ORd and CiPAORd models

A 3-D input parametric space was generated considering variations in the block of late sodium current ($INaL$), block of slow rectifitying potassium channel current (IKs) and L-type calcium channel currents ($ICaL$) at fixed value of static component of the hERG channel current ($sbIKr = 7$), as in our previous study [14]. Uniformly spaced samples in this 3D space were selected to generate a set of virtual drugs. The virtual drugs were simulated and outputs of the *in silico* models were visualized for presence or absence of EADs (Figure 20). The region in green in the Figure 20 denotes the region of the examined parameteric space where EADs were not observed in the CiPAOrd model, while the region in red denotes the region with presence of EADs in the model. The surface separating the $EAD+$ and $EAD-$ regions is shown in yellow for the CiPAORd model [10]. For comparision the estimated surface separating the $EAD+$ and $EAD-$ for the original ORd model is shown in blue. [19]. The CiPA training compounds were plotted as red (High risk drugs), yellow (Intermediate risk drugs) and green (Low risk drugs) dots.

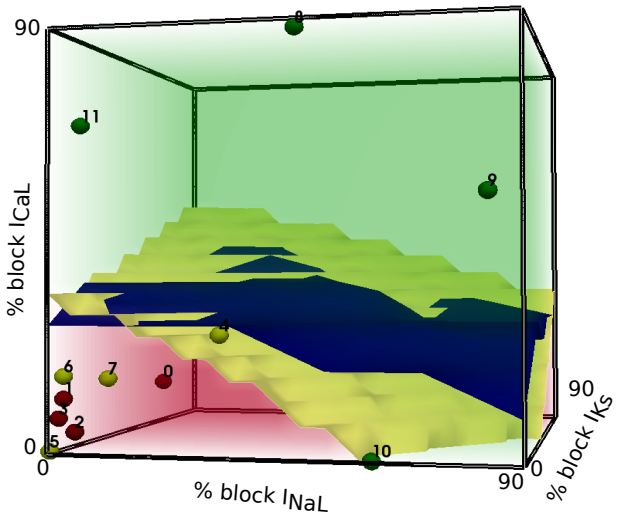

Fig 20. Surface separating the region of the input parametric space where EADs are observed in the simulated action potential from the region of the parameteric space where no EADs are observed. Blue surface: original ORd model and Yellow surface: CiPAORd model.

Acknowledgments

563

References

1. Yap YG, Camm AJ. Drug Induced QT Prolongation and Torsades de Pointes. Heart. 2003;89(11):1363–1372. doi:10.1136/heart.89.11.1363.

2. Gintant GA. Preclinical Torsades-de-Pointes Screens: Advantages and Limitations of Surrogate and Direct Approaches in Evaluating Proarrhythmic Risk. Pharmacology & Therapeutics. 2008;119(2):199–209. doi:10.1016/j.pharmthera.2008.04.010.

3. Sager PT, Gintant G, Turner JR, Pettit S, Stockbridge N. Rechanneling the Cardiac Proarrhythmia Safety Paradigm: A Meeting Report from the Cardiac Safety Research Consortium. American Heart Journal. 2014;167(3):292–300. doi:10.1016/j.ahj.2013.11.004.

4. Fermini B, Hancox JC, Abi-Gerges N, Bridgland-Taylor M, Chaudhary KW, Colatsky T, et al. A New Perspective in the Field of Cardiac Safety Testing through the Comprehensive In Vitro Proarrhythmia Assay Paradigm , A New Perspective in the Field of Cardiac Safety Testing through the Comprehensive In Vitro Proarrhythmia Assay Paradigm. Journal of Biomolecular Screening. 2016;21(1):1–11. doi:10.1177/1087057115594589.

5. Colatsky T, Fermini B, Gintant G, Pierson JB, Sager P, Sekino Y, et al. The Comprehensive in Vitro Proarrhythmia Assay (CiPA) Initiative — Update on Progress. Journal of Pharmacological and Toxicological Methods. 2016;81:15–20. doi:10.1016/j.vascn.2016.06.002.

6. Mirams GR, Cui Y, Sher A, Fink M, Cooper J, Heath BM, et al. Simulation of Multiple Ion Channel Block Provides Improved Early Prediction of Compounds' Clinical Torsadogenic Risk. Cardiovascular Research. 2011;91(1):53–61. doi:10.1093/cvr/cvr044.

7. Christophe B. Simulation of Early After-Depolarisation in Non-Failing Human Ventricular Myocytes: Can This Help Cardiac Safety Pharmacology? Pharmacological Reports. 2013;65(5):1281–1293. doi:10.1016/S1734-1140(13)71486-5.

8. Christophe B. In Silico Study of Transmural Dispersion of Repolarization in Non-Failing Human Ventricular Myocytes: Contribution to Cardiac Safety Pharmacology. British Journal of Pharmaceutical Research. 2015;7(2):88–101.

9. Lancaster MC, Sobie EA. Improved Prediction of Drug-Induced Torsades de Pointes Through Simulations of Dynamics and Machine Learning Algorithms. Clinical Pharmacology and Therapeutics. 2016;100(4):371–379. doi:10.1002/cpt.367.

10. Li Z, Dutta S, Sheng J, Tran PN, Wu W, Chang K, et al. Improving the In Silico Assessment of Proarrhythmia Risk by Combining hERG (Human Ether-à-Go-Go-Related Gene) Channel-Drug Binding Kinetics and Multichannel Pharmacology. Circulation Arrhythmia and Electrophysiology. 2017;10(2):e004628. doi:10.1161/CIRCEP.116.004628.

11. Dutta S, Chang KC, Beattie KA, Sheng J, Tran PN, Wu WW, et al. Optimization of an In Silico Cardiac Cell Model for Proarrhythmia Risk Assessment. Frontiers in Physiology. 2017;8. doi:10.3389/fphys.2017.00616.

12. Li Z, Ridder BJ, Han X, Wu WW, Sheng J, Tran PN, et al. Assessment of an In Silico Mechanistic Model for Proarrhythmia Risk Prediction Under the CiPA Initiative. Clinical Pharmacology & Therapeutics;0(0). doi:10.1002/cpt.1184.

13. Mistry HB, Davies MR, Di Veroli GY. A New Classifier-Based Strategy for in-Silico Ion-Channel Cardiac Drug Safety Assessment. Frontiers in Pharmacology. 2015;6. doi:10.3389/fphar.2015.00059.

14. Parikh J, Gurev V, Rice JJ. Novel Two-Step Classifier for Torsades de Pointes Risk Stratification from Direct Features. Frontiers in Pharmacology. 2017;8. doi:10.3389/fphar.2017.00816.

15. Mistry HB. Complex versus Simple Models: Ion-Channel Cardiac Toxicity Prediction. PeerJ. 2018;6. doi:10.7717/peerj.4352.

16. Kramer J, Obejero-Paz CA, Myatt G, Kuryshev YA, Bruening-Wright A, Verducci JS, et al. MICE Models: Superior to the HERG Model in Predicting Torsade de Pointes. Scientific Reports. 2013;3:2100. doi:10.1038/srep02100.

17. Okada Ji, Yoshinaga T, Kurokawa J, Washio T, Furukawa T, Sawada K, et al. Screening System for Drug-Induced Arrhythmogenic Risk Combining a Patch Clamp and Heart Simulator. Science Advances. 2015;1(4):e1400142. doi:10.1126/sciadv.1400142.

18. Passini E, Britton OJ, Lu HR, Rohrbacher J, Hermans AN, Gallacher DJ, et al. Human In Silico Drug Trials Demonstrate Higher Accuracy than Animal Models in Predicting Clinical Pro-Arrhythmic Cardiotoxicity. Frontiers in Physiology. 2017;8:668. doi:10.3389/fphys.2017.00668.

19. O'Hara T, Virág L, Varró A, Rudy Y. Simulation of the Undiseased Human Cardiac Ventricular Action Potential: Model Formulation and Experimental Validation. PLoS computational biology. 2011;7(5):e1002061. doi:10.1371/journal.pcbi.1002061.

20. Yan GX, Wu Y, Liu T, Wang J, Marinchak RA, Kowey PR. Phase 2 Early Afterdepolarization as a Trigger of Polymorphic Ventricular Tachycardia in Acquired Long-QT Syndrome : Direct Evidence from Intracellular Recordings in the Intact Left Ventricular Wall. Circulation. 2001;103(23):2851–2856.

21. Liu J, Laurita KR. The Mechanism of Pause-Induced Torsade de Pointes in Long QT Syndrome. Journal of Cardiovascular Electrophysiology. 2005;16(9):981–987. doi:10.1111/j.1540-8167.2005.40677.x.

22. Viswanathan PC, Rudy Y. Pause Induced Early Afterdepolarizations in the Long QT Syndrome: A Simulation Study. Cardiovascular Research. 1999;42(2):530–542.

23. Morris MD. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics. 1991;33(2):161–174. doi:10.2307/1269043.

24. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global Sensitivity Analysis: The Primer. Wiley; 2008.

25. Sobol′ IM. Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates. Mathematics and Computers in Simulation. 2001;55(1):271–280. doi:10.1016/S0378-4754(00)00270-6.

26. Hornberger GMUoV, Spear RC. Approach to the Preliminary Analysis of Environmental Systems. J Environ Manage; (United States). 1981;12:1.

27. Herman J, Usher W. SALib: An Open-Source Python Library for Sensitivity Analysis. The Journal of Open Source Software. 2017;2(9). doi:10.21105/joss.00097.

28. Iooss B, Lemaître P. A Review on Global Sensitivity Analysis Methods. arXiv:14042405 [math, stat]. 2014;.

29. Homma T, Saltelli A. Importance Measures in Global Sensitivity Analysis of Nonlinear Models. Reliability Engineering & System Safety. 1996;52(1):1–17. doi:10.1016/0951-8320(96)00002-6.

30. Saltelli A. Making Best Use of Model Evaluations to Compute Sensitivity Indices. Computer Physics Communications. 2002;145(2):280–297. doi:10.1016/S0010-4655(02)00280-1.

31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.

32. Redfern WS, Carlsson L, Davis AS, Lynch WG, MacKenzie I, Palethorpe S, et al. Relationships between Preclinical Cardiac Electrophysiology, Clinical QT Interval Prolongation and Torsade de Pointes for a Broad Range of Drugs: Evidence for a Provisional Safety Margin in Drug Development. Cardiovascular Research. 2003;58(1):32–45.

33. Campolongo F, Cariboni J, Saltelli A. An Effective Screening Design for Sensitivity Analysis of Large Models. Environmental Modelling & Software. 2007;22(10):1509–1518. doi:10.1016/j.envsoft.2006.10.004.

34. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, et al. Sensitivity Analysis of Environmental Models: A Systematic Review with Practical Workflow. Environmental Modelling & Software. 2016;79:214–232. doi:10.1016/j.envsoft.2016.02.008.

35. Crumb WJ, Vicente J, Johannesen L, Strauss DG. An Evaluation of 30 Clinical Drugs against the Comprehensive in Vitro Proarrhythmia Assay (CiPA) Proposed Ion Channel Panel. Journal of Pharmacological and Toxicological Methods. 2016 Sep-Oct;81:251–262. doi:10.1016/j.vascn.2016.03.009.

36. Milnes JT, Witchel HJ, Leaney JL, Leishman DJ, Hancox JC. Investigating Dynamic Protocol-Dependence of hERG Potassium Channel Inhibition at 37 Degrees C: Cisapride versus Dofetilide. Journal of Pharmacological and Toxicological Methods. 2010 Mar-Apr;61(2):178–191. doi:10.1016/j.vascn.2010.02.007.

37. Veroli GYD, Davies MR, Zhang H, Abi-Gerges N, Boyett MR. hERG Inhibitors with Similar Potency But Different Binding Kinetics Do Not Pose the Same Proarrhythmic Risk: Implications for Drug Safety Assessment. Journal of Cardiovascular Electrophysiology. 2014;25(2):197–207. doi:10.1111/jce.12289.

38. Johnstone RH, Bardenet R, Gavaghan DJ, Mirams GR. Hierarchical Bayesian Inference for Ion Channel Screening Dose-Response Data. Wellcome Open Research. 2016;1:6. doi:10.12688/wellcomeopenres.9945.2.

39. Chang KC, Dutta S, Mirams GR, Beattie KA, Sheng J, Tran PN, et al. Uncertainty Quantification Reveals the Importance of Data Variability and Experimental Design Considerations for in Silico Proarrhythmia Risk Assessment. Frontiers in Physiology. 2017;8. doi:10.3389/fphys.2017.00917.

40. Loucks DP, van Beek E, Stedinger JR, Dijkman JPM, Villars MT. Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications. Deltares, UNESCO-IHE, Springer; 2017.

41. Kubo T, Ashihara T, Tsubouchi T, Horie M. Significance of Integrated in Silico Transmural Ventricular Wedge Preparation Models of Human Non-Failing and Failing Hearts for Safety Evaluation of Drug Candidates. Journal of Pharmacological and Toxicological Methods. 2017 Jan - Feb;83:30–41. doi:10.1016/j.vascn.2016.08.007.

42. Trenor B, Gomis-Tena J, Cardona K, Romero L, Rajamani S, Belardinelli L, et al. In Silico Assessment of Drug Safety in Human Heart Applied to Late Sodium Current Blockers. Channels (Austin, Tex). 2013 Jul-Aug;7(4):249–262.

43. Costabal FS, Yao J, Kuhl E. Predicting the Cardiac Toxicity of Drugs Using a Novel Multiscale Exposure–Response Simulator. Computer Methods in Biomechanics and Biomedical Engineering. 2018;21(3):232–246. doi:10.1080/10255842.2018.1439479.