HAL
open science

Composite biomarkers improve classification of drug-induced channel block

Eliott Tixier, Fabien Raphel, Damiano Lombardi, Jean-Frédéric Gerbeau

To cite this version:

Eliott Tixier, Fabien Raphel, Damiano Lombardi, Jean-Frédéric Gerbeau. Composite biomarkers improve classification of drug-induced channel block. 2017. hal-01570819v2

HAL Id: hal-01570819

https://hal.science/hal-01570819v2

Preprint submitted on 30 Oct 2017 (v2), last revised 13 Dec 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Composite biomarkers improve classification of drug-induced channel block

Eliott Tixier [2,1], Fabien Raphel [1,2], Damiano Lombardi [1,2], Jean-Frédéric Gerbeau[1,2]
[1]Inria Paris, 75012 Paris, France
[2]Sorbonne Universités, UPMC Univ Paris 6, UMR 7598 LJLL, 75005 Paris, France

Abstract

The Micro-Electrode Array device enables high-throughput electrophysiology measurements that are less labour-intensive than patch-clamp based techniques. Combined with human-induced pluripotent stem cells cardiomyocytes (hiPSC-CM), it represents a new and promising paradigm for automated and accurate *in vitro* drug safety evaluation. In this article, the following question is addressed: which features of the MEA signals should be measured to better classify the effects of drugs? A framework for the classification of drugs using MEA measurements is proposed. Such a classification is based on the drugs predicted ion channels blocks. It relies on an *in silico* electrophysiology model of the MEA, a feature selection algorithm and automatic classification tools. An *in silico* model of the MEA is developed and is used to generate synthetic measurements. An algorithm that extracts MEA measurements features designed to perform well in a classification context is described. These features are called composite biomarkers. A state-of-the-art machine learning program is used to carry out the classification of drugs using experimental MEA measurements. The experiments are carried out using five different drugs: Mexiletine, Flecainide, Diltiazem, Moxifloxacin and Dofetilide. We show that the composite biomarkers outperform the classical ones in different classification scenarios. We show that using both synthetic and experimental MEA measurements improves the robustness of the composite biomarkers and that the classification scores are increased.

Keywords: cardiac electrophysiology, numerical simulations, bidomain model, micro-electrode array, classification, drug safety evaluation

Introduction

One of the main goals of safety pharmacology studies is to anticipate how drugs affect cardiomyocytes. Among other adverse effects, it focuses on predicting arrhythmic behaviors which may lead to torsades de pointes (TdP). The most common risk factors under consideration are QT prolongation and hERG block. However these risk factors are now considered insufficient and the guidelines need to be improved [15]. For instance, an observed QT prolongation is not necessarily associated with TdP risk [3]. Several advances in technology and computational modeling may favor the emergence of new methods for more efficient drug safety evaluation. On the hardware side, the Micro-Electrode Array (MEA) technology [30, 22] enables high-throughput electrophysiology measurements that are less labour-intensive than patch-clamp based techniques. This device has been successfully used in large drug studies [5]. On the biological side, the use of human-induced pluripotent stem cells (hiPSC) has developed [29] and their recent large-scale

production makes it a viable human model replacement. The combined use of the MEA technology and hiPSC cardiomyocytes (hiPSC-CM) represents a new and promising paradigm for automated and accurate *in vitro* drug safety evaluation [11, 9]. The CIPA initiative [9, 15] promotes disruptive drug safety guidelines, in particular the use of hiPSC-CM and *in silico* modeling. In parallel of these technological breakthroughs, several efforts have been recently made towards promoting the use of computational tools in drug safety evaluation [14, 20]. In this context, a framework for drug safety evaluation using *in silico* models and experimental measurements using a MEA device is hereby presented. The device considered in the present work is a six-well nine-electrode MEA but, as shown in [27], the approach is general enough to be extended to other types of MEA.

The framework aims at predicting the effect of a drug onto the cardiomyocytes ion channels activities from the knowledge of MEA experimental recordings. More precisely, the goal is to determine which ion channels are affected by a given drug. Note that the aim of the present study is not to predict the drugs propensities to induce cardiac arrhythmias but rather to identify which ion channel is primarily blocked. This represents a first step towards the use of the MEA-hiPSC-CM platform in arrhythmogenicity studies. Even though patch-clamp experiments are the gold standard to assess drug-induced channel block, it was shown in a recent study [27] that it is possible to do so also using MEA field potential measurements. The approach is based on an *in silico* model of the MEA and the hiPSC-CM tissue, a feature selection algorithm and a classification model. The *in silico* model is based on a simple ionic model [8] for the cardiomyocytes electrical activity and on the bidomain equations [32] for the spatial propagation of the electrical potentials. The ionic model counts three different currents (fast inward, slow inward, slow outward), each being associated with an ionic species (respectively sodium, calcium, potassium). The activity of each current is controlled by a scaling parameter that is referred to as conductance in the following. In the present work, the drugs considered are assumed to affect one of those three currents. Thus, the inactivation of a current caused by a drug is modeled by a diminution of the corresponding conductance in the ionic model. The conductances and some other parameters of the model can be varied in order to replicate the variability observed in the experimental measurements. The *in silico* model is used to generate what is later referred to as synthetic MEA measurements. The experimental data set itself consists of MEA electrode recordings which come in the form of time series. Each recording is done in control conditions (no drug) and with different drug concentrations levels. The experimental data is also labelled, meaning the affected ionic channels are known for each drug.

As explained above, the MEA measurements, whether synthetic or experimental, come in the form of time series. For classification purposes, it is more efficient to extract features from these time series. Some features, also called biomarkers, are already widely used in the community such as the field potential duration [12] which may be associated with the QT segment in ECGs. These common features are referred to as classical biomarkers. We propose a way to automatically extract features from the MEA measurements that are designed to perform well in a classification context. First a set of biomarkers is built. The set is referred to as dictionary and each biomarker is referred to as an entry in the following. Then we define new features, referred to as composite biomarkers, as linear combinations of the dictionary entries. The weights of these linear combinations are found by solving a sparse optimization problem. The optimization procedure uses a data set which consists of experimental MEA measurements, simulated ones or a combination of both. To predict the effects of drugs onto channel block, we propose to adopt a Machine Learning approach. Machine Learning is a family of statistical methods whose aim is to build predictive models given a (ideally large) dataset. There exists a wide variety of such methods: neural networks [18], Support Vector Machine [17],

decision trees [4], etc. All these methods have proved their performances in many different scenarios of regression and classification, in particular when applied to biological data. In the present work, we propose to use Support Vector Classification (SVC) [6] which derives from Support Vector Machine. This method seeks a hyperplane that separates the data samples with a maximum margin. The samples are then classified according to their position with respect to the separating hyperplane.

The paper is organized as follows. First, the methods are described. The *in silico* model is presented and the generation of synthetic data is explained. The algorithm that computes the composite biomarkers is described and the classification tools are presented. Second, the performance of the composite biomarkers and of the classification tools are studied in different drug classification scenarios. The composite biomarkers are compared to the classical ones using two different classification strategies. Finally, composite biomarkers computed with experimental data only and with a mixed set of experimental and synthetic data are compared.

Methods

Equations

Bidomain equations and ionic model

Let Ω be the domain representing a MEA's well. The thickness of the layer of cells being small compared to the size of the well, the problem is assumed to be two-dimensional. We denote by A_{m}, C_{m}, z_{thick} the surface area of membrane per unit volume of tissue, the membrane capacitance, and the thickness of the cell layer, respectively. The intra and extra-cellular conductivity tensors σ_{i} and σ_{e} are assumed to be scalar. The propagation of the transmembrane potential V_{m} and the extracellular potential ϕ_{e} are modeled in Ω with the bidomain model [32]:

$$\begin{cases} A_{\mathrm{m}}C_{\mathrm{m}}\dfrac{\partial V_{\mathrm{m}}}{\partial t} + A_{\mathrm{m}}I_{\mathrm{ion}}(V_m, w) - \nabla\cdot(\sigma_{\mathrm{i}}\nabla V_{\mathrm{m}}) - \nabla\cdot(\sigma_{\mathrm{i}}\nabla\phi_{\mathrm{e}}) = A_{\mathrm{m}}I_{\mathrm{app}}, \\ -\nabla\cdot((\sigma_{\mathrm{i}}+\sigma_{\mathrm{e}})\nabla\phi_{\mathrm{e}}) - \nabla\cdot(\sigma_{\mathrm{i}}\nabla V_{\mathrm{m}}) = \frac{1}{z_{thick}}\displaystyle\sum_{e_k}\frac{I_{\mathrm{el}}^k}{|e_k|}\chi_{e_k}. \end{cases} \tag{1}$$

In the second equation, I_{el}^k is the electric current which goes through the electrode located at e_k, $|e_k|$ is the electrode surface and χ_{e_k} is the characteristic function of e_k (which takes the value 1 on the electrode and 0 elsewhere). An imperfect model for the electrode is used to compute I_{el}^k and described in the Supplementary Material. The activation is assumed to be triggered by a current I_{app} that is applied in an arbitrary region of the well with a cycle length of 1200 ms. The locations of the stimulations are randomized to model the uncertainties of the spontaneous stimulus locations in *in vitro* measurements. This is further explained in the Heterogeneity modeling sub-section. The computational domain Ω corresponds to one well of the MEA device as shown in Figure 1. Let $\boldsymbol{n}$ be the outward normal to the boundary of the domain Ω. Equations (1) are completed with the following boundary conditions: $\sigma_{\mathrm{i}}\nabla\phi_{\mathrm{i}}\cdot\boldsymbol{n} = 0$ (where $\phi_{\mathrm{i}} = V_{\mathrm{m}} + \phi_{\mathrm{e}}$), and either $\phi_e = 0$ on the region connected to the ground or $\sigma_{\mathrm{e}}\nabla\phi_{\mathrm{e}}\cdot\boldsymbol{n} = 0$ elsewhere. The ground location is indicated in Figure 1.

The transmembrane ionic current I_{ion} is described with the Minimal Ventricular (MV) model [8] which includes three currents: fast inward (fi), slow inward (si) and outward (so) currents. The reader is referred to the original publication for more details. Schematically, I_{ion} depends on V_{m} and on gating variables $\mathbf{w} = (w_j)_{1\leq j\leq 3}$, solution of a system of three nonlinear ordinary differential equations. A conductance

coefficient g_s, with $s = fi, si$ or so, controls the activity of the idealized channels associated with each of three currents of the model.

The partial differential equations are discretized in space by means of P1 finite elements, and in time by using backward differentiation formula (BDF) schemes with adaptive time steps and order provided by Sundials' CVODE library [16]. The quantity of interest is the extra-cellular potentiel, also referred to as field potential (FP). It is a function of time and recorded at the electrodes locations.

Synthetic measurements In the present work, the computational model is used to generate synthetic MEA measurements. The main idea is to enrich the experimental dataset with *in silico* measurements to make the classification more robust, in particular by exploring regions of the parametric space that are not covered by the experience. For a given set of conductances, the model is evaluated and the electrodes FPs are recorded. The conductances are chosen as to represent meaningful scenarios, as explained later in the Results section. To mimic experimental measurements, a zero-mean Gaussian noise of standard deviation 10 μV is added to the FPs (see Figure 2). A heterogeneity model of some ionic parameters is also considered to replicate the variability exhibited by the experimental measurements. This model is described later in this section. The stimulation location is also varied to model the uncertainty of the spontaneous stimulus location in the experiments. Figure 3 shows examples of synthetic recordings generated using the aforementioned *in silico* model. The FPs are simulated in control conditions and with five different concentration levels of a virtual potassium-antagonist drug. In Figure 1 of the Supplementary Material a simulated FP recorded on an electrode is shown with the simulated action potential recorded on the same electrode.

Steady-state regime Because the initial conditions of the ionic model do not correspond to those of a steady-state regime, several beats may need to be simulated before reaching a regime where there is negligible beat-to-beat variations. A numerical experiment was carried out to determine when this regime is reached. Figure 4 shows super-imposed consecutive simulated FPs and the normalized beat-to-beat variations in the FP. When considering noisy synthetic measurements as described above, the steady-state is assumed to be reached when the beat-to-beat variations are comparable to variations induced by noise only. The beat-to-beat variability observed after this beat may be imputed to the coarseness of the mesh, the time discretization errors and the fluctuations of the ionic model itself. In the present work, the steady-state is assumed to be reached at the second beat. Therefore, the simulations are run for two cardiac cycles and the second beat is recorded to be used as a synthetic measurement.

Drug modeling

We chose to model the action of drugs on the ion channels by the conductance-block formulation of the pore block model [23, 33, 7]. This simple approach, which relies on a small number of parameters, was shown in [1] to be able to reproduce the expected effects of several drugs on MEA signals. The conductance of a given channel s is given by:

$$g_s = g_{control,s} \left[1 + \left(\frac{[D]}{\mathrm{IC}_{50}} \right)^n \right]^{-1}, \tag{2}$$

where $g_{control,s}$ is the drug-free maximal conductance, $[D]$ is the drug concentration, IC_{50} is the value of the drug concentration at which the peak current is reduced of 50%, n is the Hill coefficient. In this work, n will be assumed to be equal to 1.

Heterogeneity modeling

A typical experimental MEA FP measurement exhibits both a depolarization spike and a repolarization wave (see Figure 2). Using the computational model described above, the repolarization wave is usually too small compared to what is observed in experiments. As noted in [1], the repolarization wave provided by this model is larger when the domain includes cells with different APDs. In [1], the cell heterogeneity was defined on a checkerboard arbitrarily chosen in the MEA's well. We propose here a different approach, based on a probabilistic description of the heterogeneity. The tissue is supposed to be a continuous mixture of two cell types: A and B. We make the assumption that the transition between these two types can be described by a single space dependent parameter $c(x, y)$ as follows:

$$\mathbf{p}(x, y) = (1 - c(x, y))\mathbf{p}^{(A)} + c(x, y)\mathbf{p}^{(B)}, \tag{3}$$

where c is a random process with values in $[0, 1]$ and $\mathbf{p}^{(A)}$ (resp. $\mathbf{p}^{(B)}$) the set of 19 parameters of the MV model corresponding to cell type A (resp. B). The values of $\mathbf{p}^{(A)}$ and $\mathbf{p}^{(B)}$ are given in Table 1 of the Supplementary Material. The APs corresponding to different realizations of c are shown in Figure 5. We make the hypothesis that the spatial variations of c are structured by a normal correlation function f_c:

$$f_c\left[\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x' \\ y' \end{pmatrix}\right] = \exp\left[-\frac{(x - x')^2 + (y - y')^2}{2l_c^2}\right], \tag{4}$$

where l_c is the correlation length, set to $l_c = 0.25$ mm in the present work. To discretize the random process c, we compute the correlation matrix on the finite element mesh used for the discretization of the bidomain equations. The correlation matrix $\mathbf{C} = [C_{i,j}] \in \mathbb{R}^{N_{\text{mesh}} \times N_{\text{mesh}}}$ reads:

$$C_{i,j} = f_c\left[\begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix}, \begin{pmatrix} \hat{x}_j \\ \hat{y}_j \end{pmatrix}\right], \tag{5}$$

where N_{mesh} is the total number of mesh nodes and $(\hat{x}_i, \hat{y}_i)$ are the coordinates of the ith node. The eigenpairs of $\mathbf{C}$ are denoted by (λ_i, Φ_i), and ordered by decreasing order of the eigenvalues λ_i. By a convenient abuse of notation, we denote by $(\hat{x}, \hat{y}) \to \Phi_i(\hat{x}, \hat{y})$ the function of the finite element space associated with the eigenmode Φ_i. Finally, the discretized heterogeneity field is approximated by the following truncated expansion:

$$c(\hat{x}, \hat{y}, \boldsymbol{\xi}) = \sum_{i=1}^{n_c} \xi_i \Phi_i(\hat{x}, \hat{y}) \tag{6}$$

where $\boldsymbol{\xi} = (\xi_i)_{i=1\ldots n_c}$ is a random vector and n_c a truncation index chosen so that the truncation explains at least 99% of the variance. In other words, n_c is the smallest index n such that the following criterion is verified:

$$\frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{N_{\text{mesh}}} \lambda_i} > 0.99 \ . \tag{7}$$

In our case, the choice of l_c and the domain geometry yields $n_c = 14$. Heterogeneity fields can now be generated simply by sampling the random variable $\boldsymbol{\xi}$. In the present work, $N_h = 128$ heterogeneity fields were generated by sampling $\boldsymbol{\xi}$ from an uncorrelated uniform distribution over $[-1, 1]^{n_c}$, and each sample is rescaled to range between 0 and 1. An example of heterogeneity field is presented in Figure 6.

The observed variations in the experimental MEA FP recordings are also attributable to fluctuations

in the stimulation location. In practice, the hiPSC-CM are not electrically stimulated: a stimulus arises spontaneously in the medium, probably due to the presence of pacemaker cells. The location of the spontaneous stimulation is not known to the experimentalist. We make the assumption that the location is random and therefore model it with a random unifom law over the square $[0.15, 0.85]^2$ where $\Omega = [0, 1]^2$ is the complete domain.

To conclude, in a given experimental setting, we do not know neither the stimulation position nor the cell distribution inside the well and we would like the classification method to be robust with respect to all these unknown, random elements. This is why, when generating synthetic MEA FPs using our *in silico* model, we introduce two sources of uncertainty: the heterogeneous CM field and the stimulation location.

Biomarkers

Biomarkers may be defined as quantities extracted from a signal that convey information about hidden quantities of interest. In our case, the biomarkers are features extracted from the MEA FP which would ideally provide information about the conductances of interest: g_{fi}, g_{so}, g_{si}. In this section, we present different choices of biomarkers to be used in a classification context.

"Classical" biomarkers

The MEA FP can be split into two regions of interest: the depolarization and the repolarization. The depolarization observed at one electrode corresponds to the local depolarization of the cardiomyocytes. The depolarization amplitude (DA, referred to as spike amplitude in [12]) may be qualitatively linked to the AP upstroke velocity. This biomarker is commonly associated with the activity of the fast sodium channel (g_{fi} for the MV model). The repolarization amplitude (RA) may be qualitatively linked to some extent to the AP repolarization slope and to a bigger extent to spatial heterogeneities in AP durations. Once the depolarization and repolarization have been detected, it is possible to measure the FP duration (FPD), simply as the difference between the repolarization and depolarization times. The FPD is a commonly used biomarker [12, 24] which may be seen as a surrogate for APD in patch clamp experiments and QT interval in electrocardiograms. Both biomarkers RA and FPD are associated with the activity of the potassium and calcium currents (g_{so} and g_{si} in the MV model). As explained above, each (real or numerical) experiment is performed both in drug-block conditions and in control condition. Because of the significant variability of measurements in MEA, it is important to consider the variations observed in the FP in drug block conditions with respect to the control conditions to isolate the effect of the drug from other sources of variability: tissue variability, stimulation protocol, *etc.* Therefore, as proposed in [27], the features of interest are the biomarkers in drug block condition divided by the biomarkers in control conditions. For instance, the depolarization amplitude is actually the following ratio:

$$\mathrm{DA}_{\mathrm{ratio}} = \frac{\mathrm{DA}_{\mathrm{drug}}}{\mathrm{DA}_{\mathrm{control}}} \tag{8}$$

For the sake of clarity in the notation, the subscript "ratio" is omitted in the following and any biomarker actually refers to a ratio with the control value. For each MEA measurement, the FP is recorded at each of the nine electrodes. Again, the important variability in the measurements motivates the use of robust features. Since the behavior of the FP may greatly vary from one electrode to another, the median of the biomarkers over all electrodes is in practice a good choice of features. In the following, the set of biomarkers

$\{\tilde{\mathrm{DA}}, \tilde{\mathrm{RA}}, \tilde{\mathrm{FPD}}\}$ is referred to as the classical biomarkers, where the $\tilde{}$ operator denotes the median over all nine electrodes.

Composite biomarkers

The rationale behind the choice of biomarkers described above is only qualitative and oftentimes does not represent the best set of features in a classification context. Here, we adopt a more automatic strategy to select the best set of biomarkers for a given experimental scenario. First, the set of features to be extracted from a given FP is enriched with other features.

It is indeed possible to extract additional quantities from the FP other than DA, RA and FPD. We propose to compute also, for each electrode of the MEA, the following features: the area under curve of the repolarization wave (AUCr), the repolarization center (RC), the repolarization width (RW), the FP notch (FPN) and the depolarization width (DW). The details on how to compute these additional biomarkers are described in Appendix A and illustrated in Figure 2. Ratios of these quantities are also added to the dictionary of features: RA/DA, DA/RA, RA/FPD, FPD/RA, DA/FPD, FPD/DA, RA/RW, RW/RA. Each feature is actually a ratio with its control counterpart as described in (8). To include the information of all nine electrodes, the median (denoted by the $\tilde{}$ operator), mean (denoted by the $<>$ operator) and maximum values (denoted by a max subscript) over the electrodes are retained in the dictionary. We finally add the conduction velocity (CV) which is not an electrode-wise quantity but defined using all nine electrodes signals as explained in Appendix A. This amounts to a total of $N_b = 41$ features reported in Table 2. The extended set of features is referred to as the dictionary or the biomarkers dictionary. Each biomarker is referred to as an entry, denoted by b_j, $1 \leq j \leq N_b$, in the following.

Before going into further details about the numerical methods, let us now explain the purpose of the composite biomarkers. The purpose of the method is to associate each conductance g_{fi}, g_{si}, g_{so} with a composite biomarker that is maximally correlated with it and minimally correlated with the others. For instance, the composite biomarker, denoted by y_1, associated with g_{fi} is maximally correlated with g_{fi} while being minimally correlated with g_{si} and g_{so}. The main idea is that by observing y_1 we have good information about the hidden variations of g_{fi} which is not tampered by simultaneous variations of g_{si} or g_{so}. The composite biomarkers are defined as weighted linear combinations of the dictionary entries. We also require that the weights are sparse, meaning there are a lot of zero weights. This makes the composite biomarkers more easily interpretable. Indeed, they can be seen as a combination of only a small subset of the dictionary entries, ideally including the classical biomarkers as seen in Figure 7.

The weights of such a combination are solution of an optimization problem. First, let us introduce some notation. We denote by y_1 (resp. y_2, y_3) the composite biomarker (to be determined) associated with g_{fi} (resp. g_{si}, g_{so}). From now on, the conductances (g_{fi}, g_{si}, g_{so}) are denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. Each dictionary entry is considered as a function of $\boldsymbol{\theta}$. The composite biomarkers are sought as a linear combination of the dictionary entries:

$$y_h(\boldsymbol{\theta}) = \sum_{j=1}^{N_b} w_j^{(h)} b_j(\boldsymbol{\theta}), \quad 1 \leq h \leq 3, \tag{9}$$

where the weights $\boldsymbol{w}^{(h)} = (w_j^{(h)}) \in \mathbb{R}^{N_b}$ are the unknowns of the problem. These weights are sought so that $y_h(\boldsymbol{\theta})$ is maximally correlated with θ_h and minimally correlated with θ_k, $\forall k \neq h$. This may be stated as

follows:

$$\forall h \in \{1, \dots, 3\}, \quad \begin{cases} \max\limits_{y_h} & \text{cov}\left(y_h(\boldsymbol{\theta}), \theta_h\right) & \text{(10a)} \\ \min\limits_{y_h} & \left|\text{cov}\left(y_h(\boldsymbol{\theta}), \theta_k\right)\right|, \quad \forall k \neq h & \text{(10b)} \\ \text{s.t.} & \text{var}\left(y_h(\boldsymbol{\theta})\right) = 1 & \text{(10c)} \end{cases}$$

where $\text{cov}(\cdot,\cdot)$ and $\text{var}(\cdot)$ are respectively the covariance and variance operators. In the following, we assume that each component of $\boldsymbol{\theta}$ is a zero-mean unit-variance random variable. This is achieved in practice by a simple rescaling of the conductances samples. We also adopt the following notation:

$$\tilde{b}_j(\boldsymbol{\theta}) = b_j(\boldsymbol{\theta}) - \mathbb{E}\left[b_j(\boldsymbol{\theta})\right], \tag{11}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. The problem may now be recast into an optimization problem where the cost function to be minimized reads:

$$\mathcal{J}(\boldsymbol{w}^{(h)}) = \mathcal{J}_C(\boldsymbol{w}^{(h)}) + \mathcal{J}_N(\boldsymbol{w}^{(h)}) + \mathcal{J}_P(\boldsymbol{w}^{(h)}), \tag{12}$$

where

$$\mathcal{J}_C(\boldsymbol{w}^{(h)}) = \frac{1}{2}\|\mathbf{C}\boldsymbol{w}^{(h)} - \mathbf{e}^{(h)}\|^2 \quad \text{where} \quad C_{kj} := \mathbb{E}(\theta_k \tilde{b}_j),\ e_k^{(h)} := \delta_{kh}, \tag{13a}$$

$$\mathcal{J}_N(\boldsymbol{w}^{(h)}) = \frac{\xi}{2}\left(\boldsymbol{w}^{(h)T}\mathbf{G}\boldsymbol{w}^{(h)} - 1\right)^2 \quad \text{where} \quad G_{ij} := \mathbb{E}(\tilde{b}_i \tilde{b}_j), \tag{13b}$$

$$\mathcal{J}_P(\boldsymbol{w}^{(h)}) = \frac{\lambda_h}{N_g}\|\boldsymbol{w}^{(h)}\|_1. \tag{13c}$$

Let us now explain each term of (13). $\mathcal{J}_C(\boldsymbol{w}^{(h)})$ corresponds to (10a) and (10b). It measures the discrepancy to the ideal situation where $\text{cov}\left(y_h(\boldsymbol{\theta}), \theta_h\right) = 1$ and $\text{cov}\left(y_h(\boldsymbol{\theta}), \theta_k\right) = 0$, $\forall k \neq h$.
$\mathcal{J}_N(\boldsymbol{w}^{(h)})$ is a relaxation of the constraint in (10c). ξ is a regularization parameter that is set to 1 in practice. $\mathcal{J}_P(\boldsymbol{w}^{(h)})$ is a regularization term by penalization of the 1–norm of $\boldsymbol{w}^{(h)}$. ℓ_1 penalized cost functions tend to promote sparse solutions [31]. Sparse solutions for $\boldsymbol{w}^{(h)}$ are interesting in that they offer a more interpretable decomposition onto the dictionary entries (since most weights are zero) than what an ℓ_2 penalization would yield.

We now discretize the problem by considering N samples of the parameters $\boldsymbol{\theta}$ drawn over a parameter space $\Theta \subset \mathbb{R}^3$. The expectation operator is approximated using a quasi-Monte-Carlo quadrature rule and the cost function in (12) is minimized using a Nesterov accelerated gradient descent [25]. The Monte-Carlo samples may come from synthetic or experimental measurements. For synthetic measurements, the conductances are known, but this is not the case for experimental measurements. In that case, an approximation of these conductances is computed using Equation (2). Note that the solution weights depend strongly on the choice of samples used for the Monte-Carlo approximations.

An example of the obtained weights is shown in Figure 7. Interestingly, the classical biomarkers are still among the most weighted features. The correlation between the conductances of interest and the composite biomarkers is compared to the correlation with the classical biomarkers in Figure 8. The correlation between two quantities u and v is defined as follows:

$$\text{cor}(u, v) = \frac{\text{cov}(u, v)}{\sqrt{\text{var}(u)\text{var}(v)}}. \tag{14}$$

As expected, each composite biomarker is well correlated with its associated conductance whereas uncorrelated with the others. This is not the case for the classical biomarkers. The results in the next section show that such a choice of features improves the classification performance.

Experimental data set

The MEA considered in the present work is a 6-well MEA with nine electrodes per well. Its geometry as well as the corresponding finite element mesh is shown in Figure 1. The MEA measurements come in the form of FP recordings corresponding to the different electrodes of the different wells of the MEA. These recordings come in the form of time series where several cardiac cycles, or beats, are recorded. We extracted several beats on each electrode from each well of the MEA. Data were provided by Janssen Pharmaceutica NV using MC_Rack (Multi Channel Systems GmbH) and post-processed by NOTOCORD Systems (NOTOCORD-FPS 3.0 software). The hiPSC-CM used in this study are a commercially available line of cells (iCell Cardiomyocytes) and were provided by the CDI (Cellular Dynamics International) company . As explained earlier the recordings were made in control conditions (no drug) and with different drugs at different concentrations levels. Figure 9 shows examples of experimental recordings in control conditions and with five different concentrations of moxifloxacin. The drugs used for the present study are summarized in Table 3. The corresponding concentrations are presented in Table 4. The IC_{50} values that were used in the study are also reported and are in the range of those reported in [13]. Note that the diltiazem was recorded in two different wells (A and B) since it was the only calcium-antagonist drug in the experimental data that were made available to the authors. The experimental process consists in adding five times a compound at increasing concentrations in a given well. Thus, including the control condition record, we finally obtain field potentials for six contexts in each well. Equation (2) was used to obtain an approximation of the conductances values associated with the experimental measurements which are needed for the composite biomarkers calculations. The Hill coefficients and IC_{50} values are given in the Supplementary Material of [19] and in [23]. Concerning the dictionary of features, a few adjustments need to be made in some cases. Indeed, it appears that at some high concentration levels of mexiletine, there is simply no action potential (because the sodium channels are too blocked) and therefore the field potential is a flat line. To take this into account, the values of dictionary entries are set to the ones at the last concentration where an action potential was observed. In addition, all features where DA is in the numerator position in a ratio are set to zero for this concentration.

Classification

Support Vector Classification

Support vector classification [6] (SVC) is an adaptation of the support vector machine (SVM) method in a classification setting. Classification generally consists in attributing labels to inputs. The available data set, comprising both inputs and labels, is generally split into a training set used to build the classifier and a validation set to test the classifier. The inputs are often multi-dimensional and in our case correspond to the biomarkers, whether classical or composite. The labels are integers that represent the classes to which the inputs are assigned. These classes are mutually exclusive, meaning one sample can only belong to a single class. SVC belongs to the so-called supervised methods since the labels are known, at least for the training set. The main idea behind SVC is to maximize the margin between the inputs and the decision boundary [6]. In the linear case, the decision boundary is a hyperplane of the input space. In general however, this is not

sufficient to properly separate the samples according to their classes. A common way to obtain more complex boundary decisions is to use a so-called "kernel trick" [28] which is based on a mapping from the input space to a higher-dimensional space where the existence of a separating hyperplane is more likely. In the present case, the labels are "sodium antagonist", "calcium antagonist" and "potassium antagonist", respectively associated with labels 0, 1 and 2. Among various possible choices of kernels, a Gaussian kernel is employed in this work.

We used a Python implementation of SVC through the Scikit-learn [26] machine learning library which itself uses the LIBSVM library [10]. For a given training set, a so-called classifier is built. The classifier is then called to predict the labels of the validation set samples. The predictions are finally compared to the true labels. There exist several metrics to quantify the prediction quality. Two different metrics are considered here: the Cohen's kappa and the receiver operating characteristic area under curve (AUC). The Cohen's kappa is a single scalar designed to measure the performance of multi-class classifiers. Its value ranges from -1 (worst possible classifier) to 1 (perfect classifier), 0 corresponding to a coin-flip classifier. The AUC is defined for each class and measures how a classifier performs with respect to a given class. Its value ranges from 0 (worst) to 1 (best), 0.5 being a coin-flip. Because the classification is repeated several times with different data set splittings, the classification metrics are summarized using their means and standard deviations (see Table 5 for instance). The "averaged AUC" corresponds to the average of all AUCs (one AUC per class).

Both metrics are described in detail in the Supplementary Material. We now present two different strategies to employ SVC in the context of drug classification.

3-versus-3 classification Since there are three distinct classes in the experimental set, those three classes need to be included in the training set, preferably in equal proportions. The strategy of 3-versus-3 (3v3) classification consists in dividing the experimental set into a training set and validation set that both include samples from the three classes. Each class is divided into two sub-classes. This is naturally done for the sodium and potassium antagonist classes since they are each comprised of data from two different drugs. For the calcium antagonist class, the diltiazem data is artificially split into two drugs "diltiazem A" and "diltiazem B" (see Table 3). Each sub-class is associated with an identification number (ID) from 0 to 5. Therefore, there are 8 possible choices for the training and validation set combinations as summarized in Table 1.

One-versus-All classification The One-versus-All (OvA) classification strategy consists in training one classifier for each class. For each class j, the training set labels are modified to take the value 1 for samples in class j and 0 otherwise and a classifier is trained on this relabeled training set. In other words, the classifier for class j is only trained to recognize whether or not a sample belongs to class j. For the validation step, the classifiers do not predict a class label but a probability for a given sample to be in their respective class. Each sample of the validation step goes through each of the three classifiers and the predicted class corresponds to the classifier returning the highest probability. The splitting between training and validation sets is done in the same way as in the 3-vs-3 classification strategy.

Results

Comparison between classical and composite biomarkers

Here the performance of the composite biomarkers in a classification context is compared to that of the classical biomarkers for two different classification strategies. The data set is composed of 880 experiments, each counting one control measurement and 5 measurements at different drug concentration levels. For each experiment, the conductances values and FP features are computed as explained in the Methods section and the labels are defined according to Table 3. The classification results are summarized in the following and more detailed results may be found in Tables 3, 4 and 5 of the Supplementary Material.

3v3 classification

The performance of the composite biomarkers compared to the classical ones is evaluated using the 3v3 classification strategy. The classification procedure is carried out for each different splitting of the data set as summarized in Table 1. First, the classification inputs are the 3 classical biomarkers for each drug concentration level:

$$\left\{\tilde{\mathrm{DA}}_{c1}, \tilde{\mathrm{RA}}_{c1}, \tilde{\mathrm{FPD}}_{c1}, \ldots, \tilde{\mathrm{DA}}_{c5}, \tilde{\mathrm{RA}}_{c5}, \tilde{\mathrm{FPD}}_{c5}\right\}, \tag{15}$$

where c_k is the k-th concentration level. Then, the classification inputs are the composite biomarkers for each concentration, computed as explained in the Methods section using the classification training set as samples for the Monte-Carlo approximations. The inputs now read:

$$\left\{y_{1,c1}, y_{2,c1}, y_{3,c_1}, \ldots, y_{1,c5}, y_{2,c5}, y_{3,c_5}\right\}. \tag{16}$$

In both cases, the inputs are therefore of dimension 15. Note that for each splitting of the data set, new weights for the composite biomarkers are computed. The classification procedure is carried out in both cases and the results are summarized in Table 5. Regardless of the chosen classification score, the results are better using the composite biomarkers as inputs.

OvA classification

The same procedure as in the 3v3 case is applied to the OvA strategy. The classification procedure is carried out with both classical and composite biomarkers as inputs and the results are summarized in Table 6. Again, the classification results are better using the composite biomarkers as input, regardless of the classification score considered. Furthermore, the results are overall better when using the OvA approach rather than the 3v3 one.

Using combined experimental and synthetic measurements for the composite biomarkers computation

Having established that composite biomarkers outperform classical ones in two different classification scenarios, we now investigate the addition of synthetic measurements for the computation of the composite biomarkers weights. To enrich the set of experimental samples used to compute the composite biomarkers, a set of synthetic measurements is built. First, conductances samples are chosen to mimic the effect of drugs as shown in Figure 10. Depending on the most affected conductance, these samples are associated with a synthetic

sodium (resp. calcium and potassium) antagonist drug called "synth A" (resp. B and C). 775 samples per drug are chosen which amounts to 155 experiments per drug. and their repartition is summarized in Table 3. This approximately corresponds to a 50%/50% split between experimental and synthetic measurements. For each conductances sample, the computational model described in the Methods section is evaluated and the dictionary entries are computed from the simulated FPs. For each experiment, the computational model is also evaluated in control conditions, *i.e.* with $g_{fi} = g_{si} = g_{so} = 1$ in order to compute the ratios as defined in (8). The *in silico* measurements are incorporated in the experimental set to create an augmented set. This augmented set is then used to compute the composite biomarkers weights. The same data set splitting procedure as described before is carried out. Note that the synthetic measurements are only used for the composite biomarkers computation and are not included neither in the training set nor in the validation set. Again, two classification strategies are explored.

Classification results

The classification is carried out using both 3v3 and OvA approaches. The results are summarized in Tables 7 and 8. The addition of synthetic measurements to compute the composite biomarkers improves the classifier predictions in both cases. In the 3v3 case, the Cohen's kappa is increased and its standard deviation decreased. The average AUC is slightly lower using mixed experiments and synthetic data than when using experiments only but the standard deviation is decreased which suggests the mixed approach is more robust. In the OvA case, the Cohen's kappa is also increased in average and its standard deviation decreased. The average AUC remains the same but its standard deviation in decreased.

Discussion

In this study, a framework for an automatic classification of drugs from MEA measurements has been presented. The framework relies on an *in silico* model of a MEA device, on a feature selection algorithm and on state-of-the-art machine learning tools. The *in silico* model is a PDE model (the bidomain equations) coupled with a ionic model that describes the transmembrane current of the cardiomyocytes. The ionic model is a phenomenological model consisting of a set of coupled non-linear ODEs. The feature selection algorithm proposes a way to compute a so-called composite biomarker for each conductance of interest, designed to perform better in a classification context than classical biomarkers. The composite biomarkers are linear combinations of the entries of a dictionary of features which is given. The calculation of the weights involves Monte-Carlo approximations which use experimental or synthetic (or both) conductances and FP samples. It has been applied to drug classification problems using experimental MEA recordings. The classification was carried out using the Scikit-Learn Python library [26] which includes several classification tools. In the present work a Support Vector Classification was used. The data used for the classification consist in FP features extracted from experimental measurements and their associated labels corresponding to the type of drug that is considered.

The purpose of the present work is twofold. First, it intends to establish that the classically used biomarkers may be improved, at least in a classification context, by using composite biomarkers instead. Second, it intends to show that the classification performance may benefit from the addition of synthetic measurements in the calculation of the composite biomarkers. More generally, the authors intend to show that numerical simulations are useful to cardiac electrophysiology in general, beyond the sole scope of drug classification.

First, a comparison between classical and composite biomarkers was carried out. The comparison consists in

classifying drugs from experimental measurements using two different strategies: 3v3 and OvA. For each strategy, the classification is performed using classical or composite biomarkers as inputs. As expected, the classification results in both cases are improved when using the composite biomarkers. The latter were indeed designed to be maximally correlated to their associated conductance and minimally correlated to the others. As a consequence, they are more revealing of the underlying conductances than the classical biomarkers. In the 3v3 case, the mean AUCs are increased when using the composite biomarkers as inputs and the standard deviations remain similar. The mean Cohen's kappa also increases when using composite biomarkers. However, its standard deviation is also higher but not enough to compromise the improvement of its mean. The same observations can be made for the OvA case except for the g_{fi} AUC which is reduced when using composite biomarkers.

Second, the use of combined experimental synthetic measurements to compute composite biomarkers is investigated. The composite biomarkers are computed using Monte-Carlo approximations that require conductances and FP features samples. In the previous case, these samples are experimental. The idea is to improve the robustness of the composite biomarkers by incorporating synthetic measurements that span better the parameters (*i.e.* conductances) space. This approach is meant to compensate the scarcity of experimental data and more generally the fact that experiments do not cover every possible drug block scenario. The *in silico* measurements allow for a more thorough exploration of the parameter space. Conductances samples were drawn and the computational model was evaluated to generate noisy FPs. From these FPs, the entries of the dictionary of features were computed. The composite biomarkers weights are then computed using a composite set of experimental and synthetic samples. These composite biomarkers are compared to the ones computed using only experimental data. The same two classification strategies as before are used to compare both approaches. In the 3v3 case, the Cohen's kappa is slightly improved in average but more importantly its standard deviation is divided by more than two. This suggests that the use of a mixed set in the composite biomarkers computation makes the classification more robust. As for the AUCs, they are approximately the same. The improvements for the OvA case are of the same nature: increase of the Cohen's kappa average, decrease in its standard deviation and no improvement in the average AUCs. Note however that the AUCs standard deviations decrease for g_{fi} and g_{si} while it increases for g_{so}. These results suggest that, for the classification scenarios envisioned here, the addition of synthetic measurements is always beneficial.

The use of FP features in a classification context is now discussed. In classification problems, and in machine learning in general, a large number of inputs tend to provoke an over-fitting of the model. This means that the classifier tends to have satisfactory training scores but generalizes poorly on a validation test. This is in part solved by the regularization used but the number of inputs still remains important. When dealing with experimentally recorded FPs, the different signals are often not perfectly synchronized, making timestep-wise comparisons meaningless. Furthermore, an important variability of the signal amplitudes is observed in practice, making even perfectly synchronized signals difficult to compare. Using features extracted from the FP that are do not depend on time shifts and amplitude variations are therefore more robust in a classification context.

Limitations

The limitations of the proposed approach are now discussed. First we discuss the heterogeneity modeling. In the present work, we make the assumption that the hiPSC-CM medium is a continuous mixture of two cell types ("A" and "B") based on a ventricular endocardium cell model, modified to match the action

potential duration of the experimental recordings. The actual nature of the hiPSC-CM types is still quite unknown, to the authors knowledge, even though some studies suggest it is a mixture of atrial, ventricular and pacemaker cells [21]. Even if the medium can be well characterized in a particular setting, it varies greatly from one cell line to another. In the present work, we propose a general method to generate heterogeneous media and for the sake of simplicity we restricted our study to a continuous mixture of two cell types. The approach is easily generalizable to more realistic heterogeneities, including for instance atrial, ventricular and pacemaker cells. Second, the conductances values associated with the experimental measurements are not known and are therefore approximated using Equation (2). This approximation is however subject to several sources of uncertainty such as the IC_{50} whose value for a given drug may vary according to the source considered [19, 23]. The uncertainties also come from the Hill's equation which is an imperfect model. Knowing the exact values for the conductances is however not critical since those values are only needed to derive the composite biomarkers and are not directly used during the classification procedure. Furthermore, the drugs studied in the present work are assumed to be single channel blockers. In reality, some drugs (*e.g.* diltiazem) are known to target more than one ion channel. In fact, it can be considered that any drug affects every ion channel with different IC_{50} values. In the present work, we make the strong assumption of single channel blocking as a first step towards a finer description of the drugs effects. This assumption is also motivated by the simplicity of the considered ionic model which only counts three different currents. Note also that mexiletine primarily blocks the late sodium channel current and not the fast one. In the MV model, there is no distinction between these two currents.

Another limitation comes from the computational model used in the present work. The sources of error are multiple: space and time discretizations, conductivities errors, modeling errors etc. These errors are not critical either since the computational model is only used to compute the composite biomarkers weights. This study shows that, despite the modeling errors, adding synthetic measurements simulated by the computational model leads to a better and more robust classification. In the present study, we based our *in silico* modeling on the MV ionic model. It is a very simplistic model which is not able to reproduce complex behaviors such as early after depolarizations for instance. Furthermore, the hiPSC-CM are spontaneously excitable cells in our case while the MV model is not sophisticated enough to reproduce such a behaviour. For this reason, it is not suited to the study of drug arrhythmogenicity. However, in the scope of the present work, we have established that it is suited to the characterization of drug-induced channel block, at least for a coarse description of it. Furthermore, it was also established in [27] that it is possible to identify which of the three main currents is affected by a drug using the MV model. Other limitations come from the classification strategies. Both classification strategies are non-exhaustive in that they do not explore every possible way of splitting the data set. Furthermore, the classification metrics used to compare the different approaches are not flawless. In some cases comparing AUCs for instance is not the best way to compare classifiers [2]. Other metrics exist, such as the mean squared error, but were not investigated in this work. Finally, the composite biomarkers derived in the present work are not optimal in the sense that their correlation with their associated conductances is not equal to one, as seen in Figure 8.

Finally, the limitations of the study arise from the MEA measurements themselves. The variations of the repolarization wave morphology and the depolarization amplitude from one experiment to another constitutes a technical challenge when one tries to extract meaningful information from the measurements. In the present study, we propose to model the heterogeneities of the experimental settings (CM cell types and stimulation location) to account for the observed variability in the data. Furthermore, considering ratios of biomarkers with their control counterparts makes the approach more robust and less dependent on fluctuations from one

experiment to another.

Perspectives

We now discuss some perspectives that could lead to interesting future works. Other classification methods than SVC exist, such as neural networks or random forests for instance. It would be interesting to assess whether the findings of this work are still valid when considering other classification tools. It would also be interesting to evaluate which classification tool generally performs best in the present drug classification context. Other perspectives concern the composite biomarkers computed using a mixed set of synthetic and experimental measurements. In the present work, the mixed set is roughly composed of half synthetic and half experimental measurements. However, other proportions could be investigated and an optimal proportion with respect to the classification score could be found. In the present work, only sodium, potassium and calcium antagonists drugs are considered but other types of drugs exist. Drugs that affect other ionic channels or even simultaneously several of them could be investigated. In parallel, more sophisticated ionic models including more current types would need to be used to model these new drugs. This would of course come at the expense or more computationally intensive simulations. Finally, training classifiers with only synthetic measurements instead of experimental ones could be considered. This would be very useful when experimental data are insufficient or even not available. The classifiers could also be trained with a mixed set of synthetic and experimental data just like it is done in this work for the computation of composite biomarkers. Finally, as explained earlier, the point of the present work is not the direct assessment of drugs arrhythmogeneicity but rather the identification of the main channel block induced by the drugs. This is, in the author's opinions, a necessary first step towards a better understanding of MEA measurements and *in fine* its use in drug safety evaluation. Considering a larger set of drugs and more realistic ionic models in order to perform drugs classification based on their arrhythmogenicity (or TdP risk) will be the purpose of future works.

References

[1] E. Abbate, M. Boulakia, Y. Coudière, J.-F. Gerbeau, P. Zitoun, and N. Zemzemi. In silico assessment of the effects of various compounds in MEA/hiPSC-CM assays: Modelling and numerical simulations. *Journal of Pharmacological and Toxicological Methods*, July In Press. `https://hal.inria.fr/hal-01562673`.

[2] N. M. Adams and D. J. Hand. Improving the practice of classifier performance assessment. *Neural computation*, 12(2):305–311, 2000.

[3] C. Antzelevitch, L. Belardinelli, A. C. Zygmunt, A. Burashnikov, J. M. Di Diego, J. M. Fish, J. M. Cordeiro, and G. Thomas. Electrophysiological effects of ranolazine, a novel antianginal agent with antiarrhythmic properties. *Circulation*, 110(8):904–910, 2004.

[4] S. Arikawa, S. Miyano, A. Shinohara, S. Kuhara, Y. Mukouchi, and T. Shinohara. A machine discovery from amino acid sequences by decision trees over regular patterns. *New Generation Computing*, 11(3):361–375, 1993.

[5] K. Blinova, J. Stohlman, J. Vicente, D. Chan, L. Johannesen, M. P. Hortigon-Vinagre, V. Zamora, G. Smith, W. J. Crumb, L. Pang, et al. Comprehensive translational assessment of human-induced pluripotent stem cell derived cardiomyocytes for evaluating drug-induced arrhythmias. *Toxicological Sciences*, 155(1):234–247, 2016.

[6] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[7] D. Bottino, R. C. Penland, A. Stamps, M. Traebert, B. Dumotier, A. Georgieva, G. Helmlinger, and G. S. Lett. Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart. *Progress in biophysics and molecular biology*, 90(1):414–443, 2006.

[8] A. Bueno-Orovio, E. M. Cherry, and F. H. Fenton. Minimal model for human ventricular action potentials in tissue. *Journal of theoretical biology*, 253(3):544–560, 2008.

[9] I. Cavero, J.-M. Guillon, V. Ballet, M. Clements, J.-F. Gerbeau, and H. Holzgrefe. Comprehensive in vitro proarrhythmia assay (cipa): Pending issues for successful validation and implementation. *Journal of pharmacological and toxicological methods*, 81:21–36, 2016.

[10] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.

[11] M. Clements and N. Thomas. High-throughput multi-parameter profiling of electrophysiological drug effects in human embryonic stem cell derived cardiomyocytes using multi-electrode arrays. *Toxicological Sciences*, 140(2):445–461, 2014.

[12] M. Clements and N. Thomas. High-throughput multi-parameter profiling of electrophysiological drug effects in human embryonic stem cell derived cardiomyocytes using multi-electrode arrays. *Toxicological Sciences*, 140(2):445–461, 2014.

[13] W. J. Crumb, J. Vicente, L. Johannesen, and D. G. Strauss. An evaluation of 30 clinical drugs against the comprehensive in vitro proarrhythmia assay (cipa) proposed ion channel panel. *Journal of pharmacological and toxicological methods*, 81:251–262, 2016.

[14] M. R. Davies, K. Wang, G. R. Mirams, A. Caruso, D. Noble, A. Walz, T. Lavé, F. Schuler, T. Singer, and L. Polonchuk. Recent developments in using mechanistic cardiac modelling for drug safety evaluation. *Drug Discovery Today*, 2016.

[15] B. Fermini, J. C. Hancox, N. Abi-Gerges, M. Bridgland-Taylor, K. W. Chaudhary, T. Colatsky, K. Correll, W. Crumb, B. Damiano, G. Erdemli, et al. A new perspective in the field of cardiac safety testing through the comprehensive in vitro proarrhythmia assay paradigm. *Journal of biomolecular screening*, 21(1):1–11, 2016.

[16] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, 31(3):363–396, 2005.

[17] S. Hua and Z. Sun. A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. *Journal of molecular biology*, 308(2):397–407, 2001.

[18] S. Kiranyaz, T. Ince, and M. Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2016.

[19] J. Kramer, C. A. Obejero-Paz, G. Myatt, Y. A. Kuryshev, A. Bruening-Wright, J. S. Verducci, and A. M. Brown. Mice models: superior to the herg model in predicting torsade de pointes. *Scientific reports*, 3, 2013.

[20] M. C. Lancaster and E. Sobie. Improved prediction of drug-induced torsades de pointes through simulations of dynamics and machine learning algorithms. *Clinical Pharmacology & Therapeutics*, 100(4):371–379, 2016.

[21] E. Matsa, D. Rajamohan, E. Dick, L. Young, I. Mellor, A. Staniforth, and C. Denning. Drug evaluation in cardiomyocytes derived from human induced pluripotent stem cells carrying a long qt syndrome type 2 mutation. *European heart journal*, 32(8):952–962, 2011.

[22] T. Meyer, K.-H. Boven, E. Günther, and M. Fejtl. Micro-electrode arrays in cardiac safety pharmacology. *Drug Safety*, 27(11):763–772, 2004.

[23] G. Mirams, Y. Cui, A. Sher, M. Fink, J. Cooper, B. Heath, N. McMahon, D. Gavaghan, and D. Noble. Simulation of multiple ion channel block provides improved early prediction of compounds clinical torsadogenic risk. *Cardiovascular research*, 91(1):53–61, 2011.

[24] E. G. Navarrete, P. Liang, F. Lan, V. Sanchez-Freire, C. Simmons, T. Gong, A. Sharma, P. W. Burridge, B. Patlolla, A. S. Lee, et al. Screening drug-induced arrhythmia using human induced pluripotent stem cell–derived cardiomyocytes and low-impedance microelectrode arrays. *Circulation*, 128(11 suppl 1):S3–S13, 2013.

[25] B. O'Donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15(3):715–732, 2015.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[27] F. Raphel, M. Boulakia, N. Zemzemi, Y. Coudière, J.-M. Guillon, P. Zitoun, and J.-F. Gerbeau. Identification of ion currents components generating field potential recorded in mea from hipsc-cm. *IEEE Transactions on Biomedical Engineering*, 2017.

[28] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[29] C. W. Scott, M. F. Peters, and Y. P. Dragan. Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. *Toxicology letters*, 219(1):49–58, 2013.

[30] M. Systems. Microelectrode array (mea) manual. `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/MEA_Manual.pdf`, -.

[31] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[32] L. Tung. *A bi-domain model for describing ischemic myocardial D-C potentials*. PhD thesis, MIT, 1978. `https://hdl.handle.net/1721.1/16177`.

[33] N. Zemzemi, M. Bernabeu, J. Saiz, J. Cooper, P. Pathmanathan, G. Mirams, J. Pitt-Francis, and B. Rodriguez. Computational assessment of drug-induced effects on the electrocardiogram: from ion channel to body surface potentials. *British journal of pharmacology*, 168(3):718–733, 2013.

Ethical concerns

The MEA recordings were made using a commercially available line of hiPSC-CM provided by the CDI (Cellular Dynamics International) company.

Data Accessibility

The implementation (in Python) of the composite biomarkers algorithm is available at this URL: `https://github.com/eltix/numbio`.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

All authors contributed equally to this work.

Funding

ET was funded by a doctoral grant from the French Ministry of Research and Higher Education. FR was funded by Instem. This work was partially supported by the Agency for Interaction in Mathematics with Business and Society (AMIES).

Acknowledgments

We would like to thank Janssen Pharmaceutica NV for providing us with the raw experimental data and Philippe Zitoun for many fruitful discussions.

Tables

Splitting index	0	1	2	3	4	5	6	7
training set IDs	$\{0,2,4\}$	$\{0,2,5\}$	$\{0,3,4\}$	$\{0,3,5\}$	$\{1,2,4\}$	$\{1,2,5\}$	$\{1,3,4\}$	$\{1,3,5\}$
validation set IDs	$\{1,3,5\}$	$\{1,3,4\}$	$\{1,2,5\}$	$\{1,2,4\}$	$\{0,3,5\}$	$\{0,3,4\}$	$\{0,2,5\}$	$\{0,2,4\}$

Table 1: Different possible splittings of the experimental data set.

Index (median)	Index (mean)	Index (max)	Entry
0	8		DA
1	9		RA
2	10		FPD
3	11		AUC_r
4	12		RC
5	13		RW
6	14		FPN
7	15		DW
16	24	32	RA/DA
17	25	33	DA/RA
18	26	34	RA/FPD
19	27	35	FPD/RA
20	28	36	DA/FPD
21	29	37	FPD/DA
22	30	38	RA/RW
23	31	39	RW/RA
40			CV

Table 2: Indices of the biomarkers dictionary entries.

Drug name	Blocked ionic channel	Associated conductance	ID	SVC class label	# experiments
Mexiletine	sodium	g_{fi}	0	0	160
Flecainide	sodium	g_{fi}	1	0	120
Diltiazem A	calcium	g_{si}	2	1	160
Diltiazem B	calcium	g_{si}	3	1	160
Moxifloxacin	potassium	g_{so}	4	2	120
Dofetilide	potassium	g_{so}	5	2	160
synth. A	sodium	g_{fi}	6	0	155
synth. B	calcium	g_{si}	7	1	155
synth. C	potassium	g_{so}	8	2	155

Table 3: Repartition of the available (experimental and synthetic) data set. The ID is used in the data set splitting (see Table 1). The SVC class label corresponds to the associated blocked channel conductance.

Concentration index	Mexiletine	Flecainide	Diltiazem	Moxifloxacin	Dofetilide
1	0.01 μM	0.1 μM	0.01 μM	10.0 μM	0.1 nM
2	0.1 μM	1.0 μM	0.1 μM	30.0 μM	1.0 nM
3	1.0 μM	10.0 μM	1.0 μM	100.0 μM	10.0 nM
4	10.0 μM	50.0 μM	5.0 μM	200.0 μM	50.0 nM
5	50.0 μM	100.0 μM	10.0 μM	300.0 μM	100.0 nM
Main channel blocked	Sodium	Sodium	Calcium	Potassium	Potassium
IC_{50}	43.0 μM	6.2 μM	0.75(or 0.45) μM	86.2 μM	30.0(or 5.0) nM

Table 4: Summary of the drugs informations constituting the experimental measurement set. Five concentrations were studied for each drug. The IC_{50} values are reported as well as the main channel blocked by each drug (in the scope of the single channel block assumption).

	classical biomarkers		composite biomarkers	
Score	mean	std.	mean	std.
Cohen's kappa	0.27	0.16	0.56	0.25
g_{fi} AUC	0.74	0.15	0.90	0.09
g_{si} AUC	0.98	0.01	1.00	0.00
g_{so} AUC	0.69	0.04	0.84	0.04
averaged AUC	0.80	-	0.92	-

Table 5: Comparison between classical and composite biomarkers with the 3v3 classification strategy.

	classical biomarkers		composite biomarkers	
Score	mean	std.	mean	std.
Cohen's kappa	0.44	0.24	0.54	0.24
g_{fi} AUC	0.83	0.10	0.74	0.24
g_{si} AUC	0.89	0.10	0.94	0.07
g_{so} AUC	0.77	0.13	0.92	0.08
averaged AUC	0.83	-	0.87	-

Table 6: Comparison between classical and composite biomarkers. Classification scores in the one-vs-all scenario.

	experiments only		experiments + synthetic	
Score	mean	std.	mean	std.
Cohen's kappa	0.56	0.25	0.59	0.10
g_{fi} AUC	0.90	0.09	0.89	0.06
g_{si} AUC	1.00	0.00	1.00	0.00
g_{so} AUC	0.84	0.04	0.85	0.06
averaged AUC	0.92	-	0.91	-

Table 7: Comparison between composite biomarkers computed from experiments only and combined experiments and synthetic measurements. 3v3 classification strategy.

	experiments only		experiments + synthetic	
Score	mean	std.	mean	std.
Cohen's kappa	0.54	0.24	0.63	0.19
g_{fi} AUC	0.74	0.24	0.81	0.14
g_{si} AUC	0.94	0.07	0.99	0.01
g_{so} AUC	0.92	0.08	0.81	0.17
averaged AUC	0.87	-	0.87	-

Table 8: Comparison between composite biomarkers computed from experiments only and combined experiments and synthetic measurements. OvA classification strategy

Figure captions

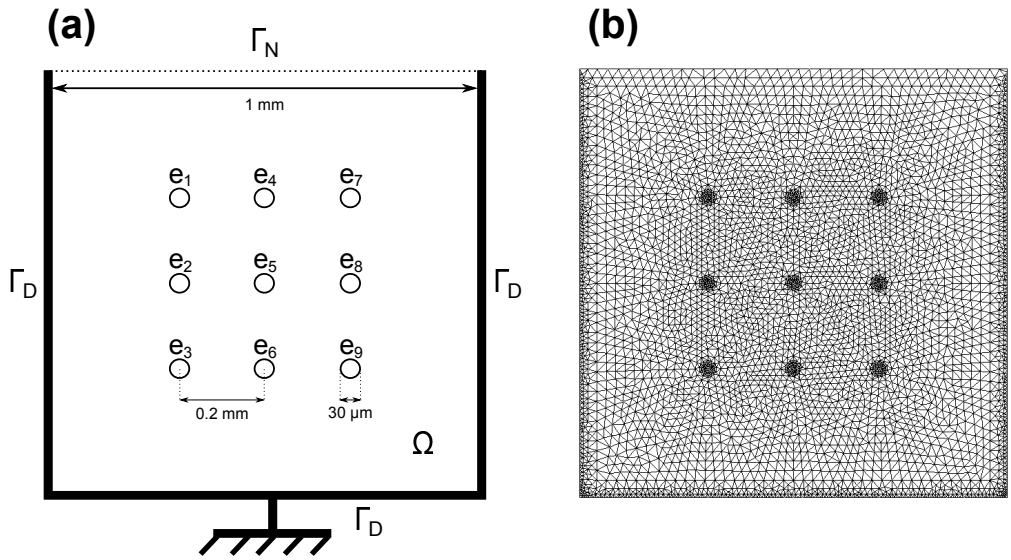

Figure 1: (a) Schematic of one well of the nine-electrode MEA device. The bidomain equations are solved in the domain Ω with homogeneous Neumann boundary conditions on Γ_N: $\nabla\phi_e \cdot \vec{n} = 0$ and homogeneous Dirichlet boundary conditions on Γ_D: $\phi_e = 0$ where the ground is located. (b) Corresponding finite element mesh. the locations of the nine electrodes.

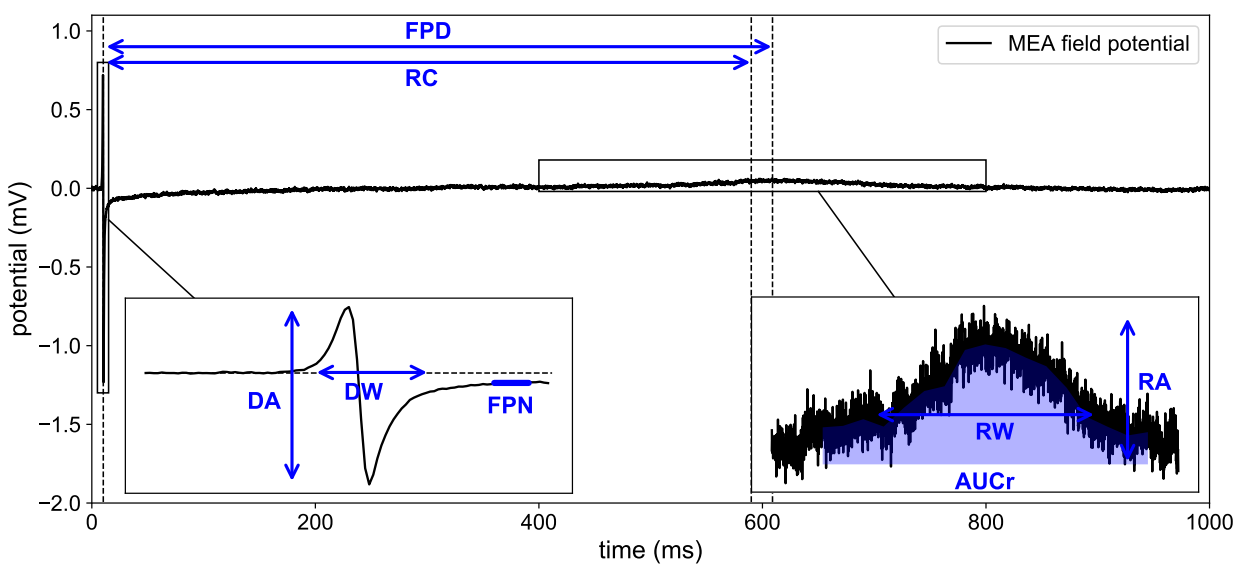

Figure 2: Experimental recording of MEA field potential. 8 biomarkers are extracted from the time series: DA (depolarization amplitude), DW (depolarization width), RA (repolarization amplitude), FPD (field potential duration), AUCr (area under repolarization curve), RC (repolarizarion center), RW (repolarization width) and FPN (field potential notch).

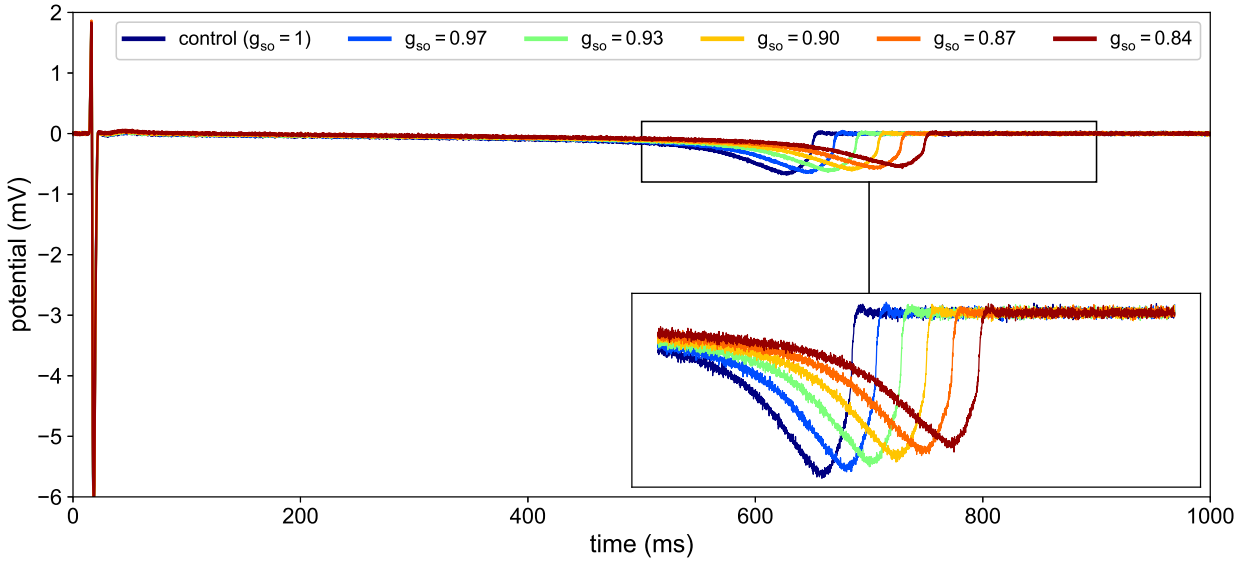

Figure 3: Effect of a virtual potassium antagonist drug onto simulated MEA FPs for five different drug concentrations and in control conditions. The drug effects are modeled using (2), which amounts to reducing the g_{so} conductance of the MV model. The observed effects (FPD elongation and RA reduction) are to be compared with experimental recordings using moxifloxacin in Fig 9.

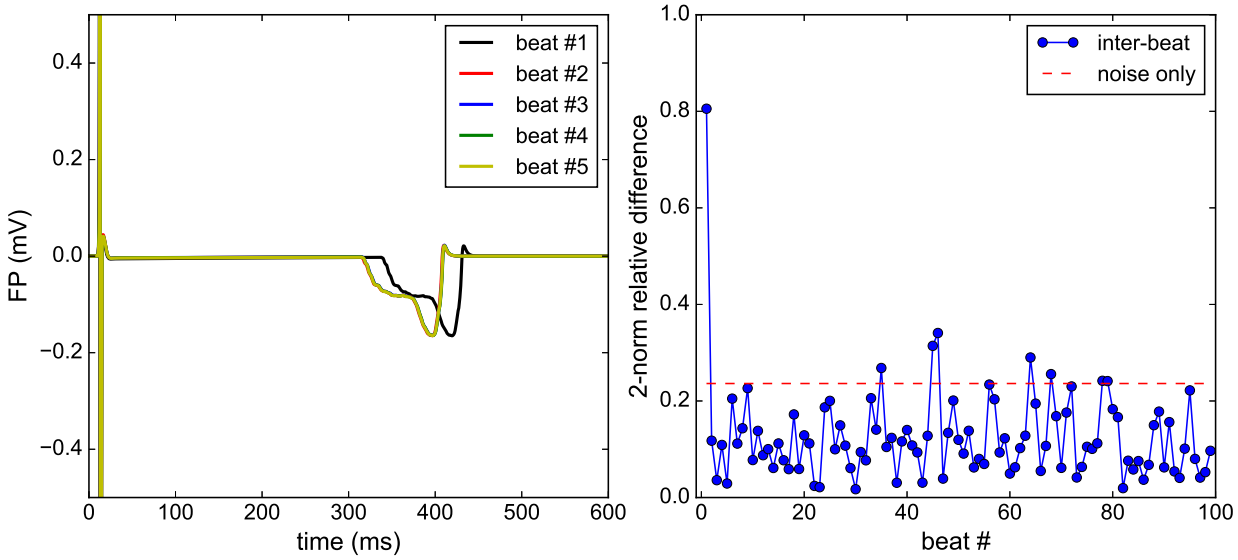

Figure 4: Steady-state analysis: the Bidomain equations are solved for 100 consecutive beats. Qualitatively, a satisfactory steady state is reached at the second beat (left). The beat-to-beat relative difference of the FP is monitored (right) and is to be compared to the relative difference between two identical solutions, each polluted by an independent noise (right).

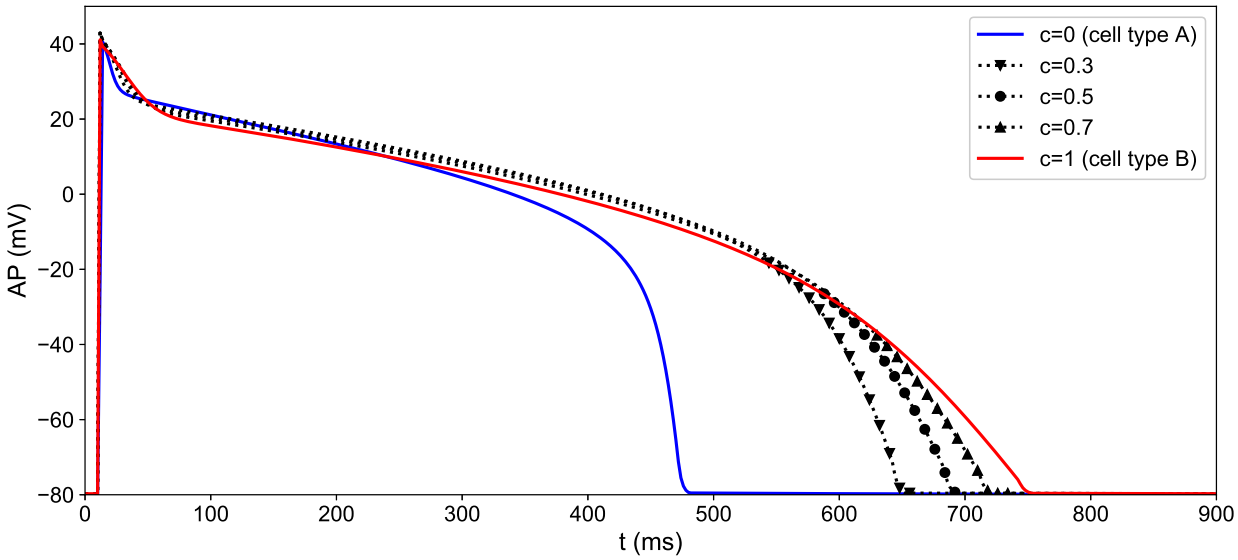

Figure 5: Heterogeneity modeling: different APs obtained by simulating the MV model with different values of the heterogeneity parameter c. The heterogeneity parameter is a function of space and its pattern differs from one well to another (see Figure 6).

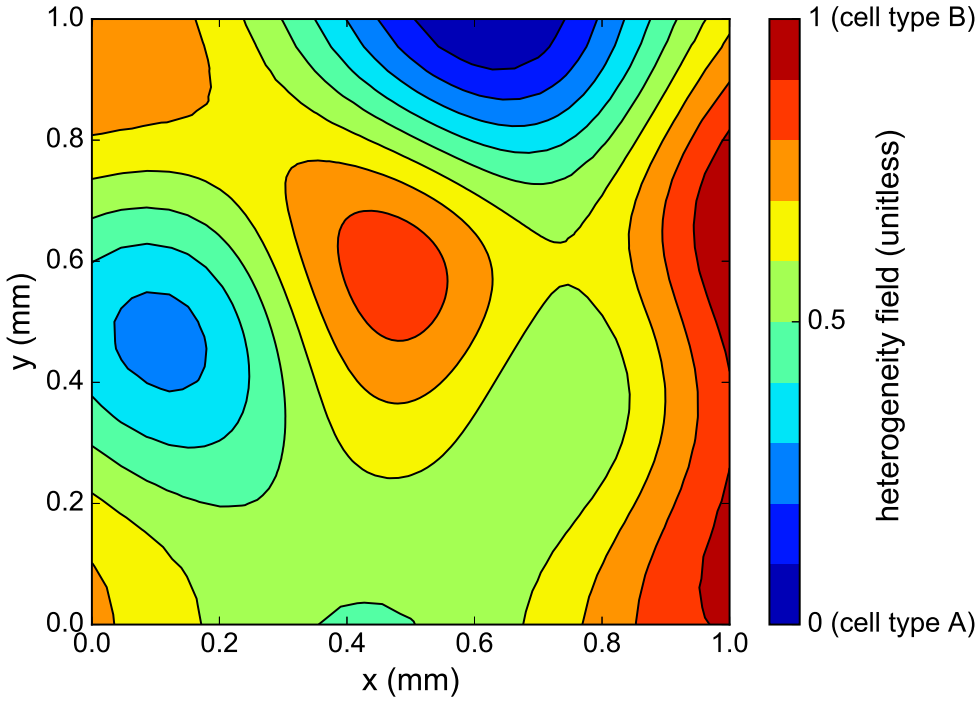

Figure 6: One sample of cell heterogeneity field $c(x, y)$ generated using the correlation matrix method. As c ranges from 0 to 1, the cell action potential varies from that of cell type "A" to cell type "B" (see Figure 5).

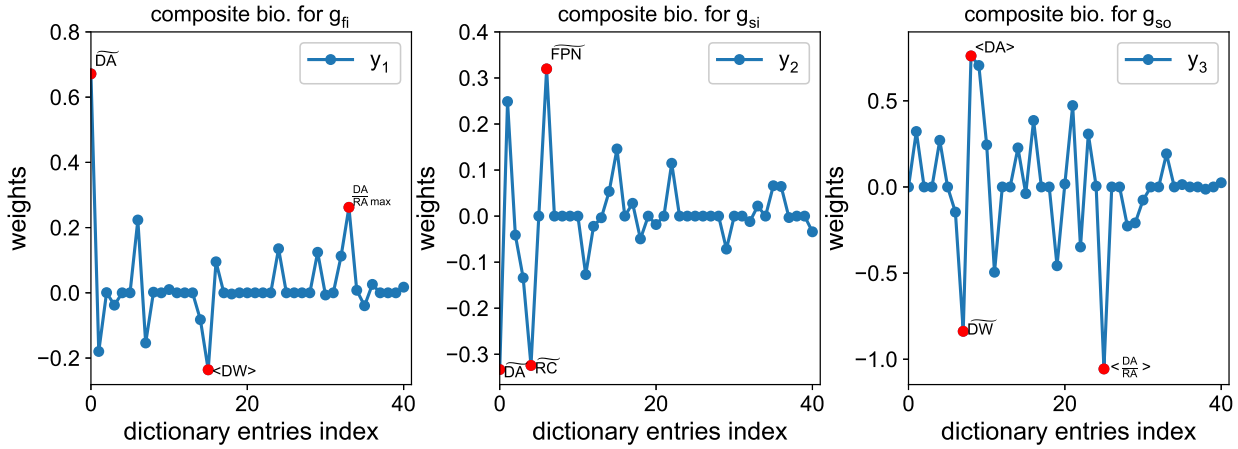

Figure 7: Example of composite biomarkers weights. The three highest weights (in absolute value) are highlighted by a red dot for each composite biomarker. Note that some classical biomarkers are selected by the method: $\tilde{\text{DA}}$ for g_{fi}, $\tilde{\text{RC}}$ (closely related to the FPD) for g_{si} and RA in the ratio (DA/RA) for g_{so}.

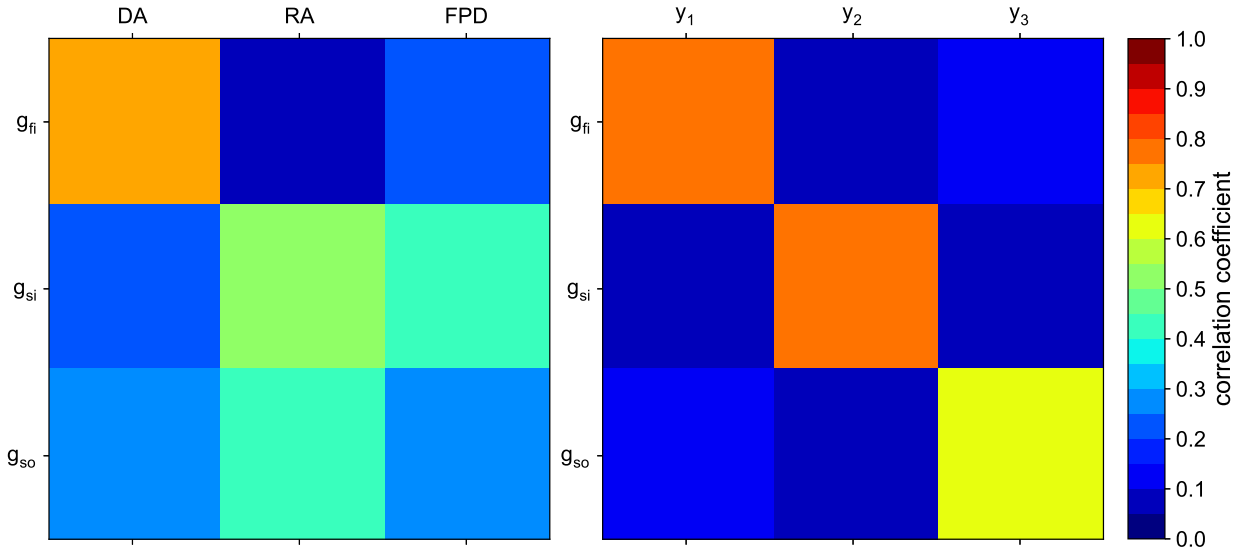

Figure 8: Correlation matrix of the conductances of interest with the "classical" biomarkers (left) and with the composite biomarkers (right)

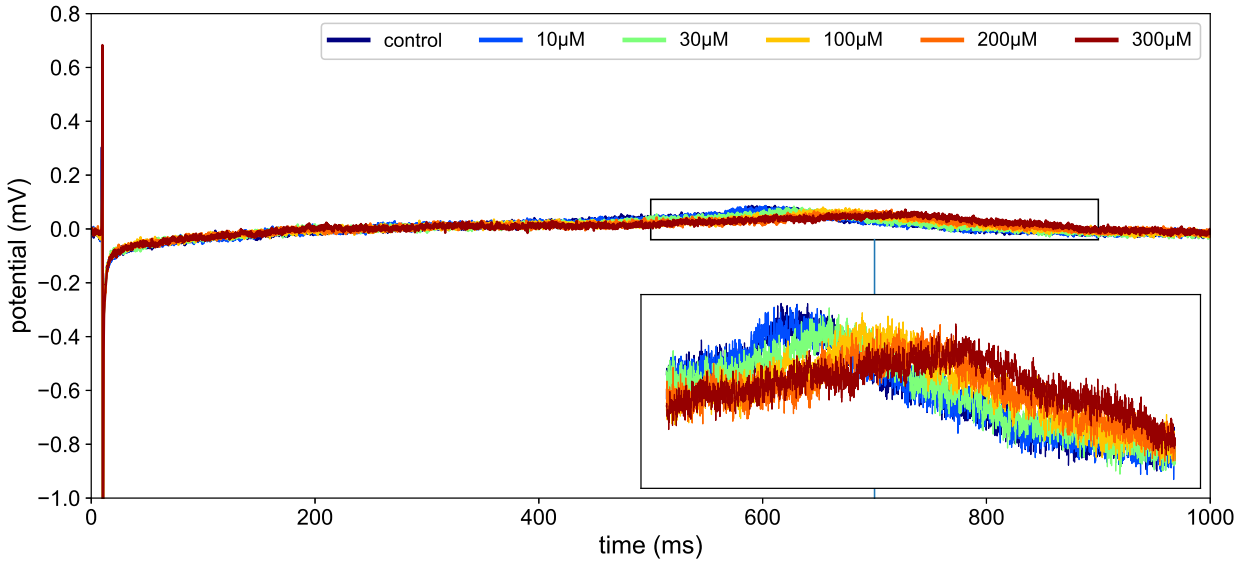

Figure 9: Effect of moxifloxacin (potassium antagonist drug) on experimental recordings of MEA FPs for five different drug concentrations and in control conditions. The main effect of moxifloxacin onto the MEA FP is the elongation of the field potential duration and the reduction of the repolarization amplitude.

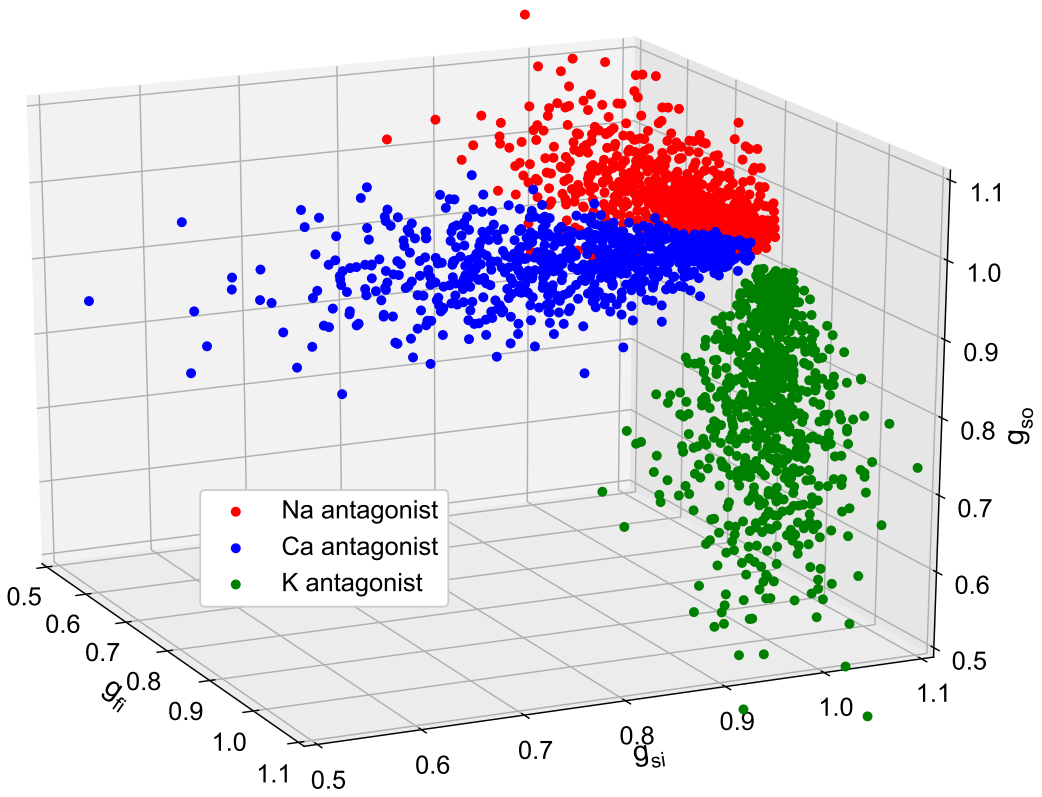

Figure 10: Plot of the 2325 *in silico* conductances samples. Three populations of 155 virtual drugs were generated according to their ion channel targets: sodium antagonist drugs (red), calcium antagonist (blue) and potassium antagonist (green). For each drug, 5 different concentrations are considered which correspond to 5 different set of conductances. These conductances are then used as inputs to generate *in silico* MEA measurements using the bidomain equations.

Appendix A: computation of the biomarkers

In this section, we provide details on how to compute the biomarkers from FP time series as shown in Figure 2. For a given signal, we denote by t the time vector and y the FP. Each signal is divided into two parts: the depolarization region (t_1, y_1) from $t = 0$ to $t = 100$ ms and the repolarization region (t_2, y_2) from $t = 0$ to $t = 1200$ ms.

Depolarization amplitude (DA) The DA is simply defined as the difference between the maximum and minimum values of the potential during the depolarization:

$$\mathrm{DA} = \max(y_1) - \min(y_1). \tag{17}$$

Depolarization width (DW) The DW is defined as the total duration during which the field potential is above, in absolute value, 10% of its peak value:

$$\mathrm{DW} = \int_{t_1 \in I} \mathrm{d}t_1, \tag{18}$$

where $I = \{t_1 \mid \quad |y_1(t_1)| \geq 0.1 \times \max(|y_1|)\}$

Repolarization amplitude (RA) The RA is defined as the maximum (in absolute value) of the repolarization.

$$\mathrm{RA} = \max(|y_2|). \tag{19}$$

Field potential duration (FPD) The FPD is defined as the time difference between the maximum (in absolute value) of the depolarization and the maximum (in absolute value) of the repolarization. Let $t_d = t\left[\underset{t}{\mathrm{argmax}}(|y_2(t)|)\right]$ and $t_r = t\left[\underset{t}{\mathrm{argmax}}(|y_1(t)|)\right]$. Then,

$$\mathrm{FPD} = t_r - t_d. \tag{20}$$

Area under the curve of the repolarization wave (AUCr) The AUCr is defined as the area under the curve of y_2 truncated around $\pm\Delta t$ of t_r. We used $\Delta t = 100$ms. The integral is approximated using the trapezoidal rule.

$$\mathrm{AUCr} = \left| \int_{t_r - \Delta t}^{t_r + \Delta t} y_2(t)\mathrm{d}t \right| \tag{21}$$

Repolarization center (RC) The RC is defined as the offset of the barycenter (with respect to time) of the repolarization wave.

$$\mathrm{RC} = \int_{t_r - \Delta t}^{t_r + \Delta t} t\bar{y}_2(t)\mathrm{d}t - t_d, \tag{22}$$

where $\bar{y}_2(t)$ is a non-dimensional rescaling of $y_2(t)$ so that it is strictly positive and integrates to 1 on $[t_r - \Delta t, t_r + \Delta t]$. $\bar{y}_2(t)$ may be seen as a probability density function over the domain $[t_r - \Delta t, t_r + \Delta t]$.

Repolarization width (RW) The RW is defined as the standard deviation of the repolarization wave.

$$\mathrm{RW} = \left[\int_{t_r-\Delta t}^{t_r+\Delta t} t^2 \bar{y}_2(t)\mathrm{d}t - \left(\int_{t_r-\Delta t}^{t_r+\Delta t} t\bar{y}_2(t)\mathrm{d}t \right)^2 \right]^{1/2}. \tag{23}$$

Field potential notch (FPN) The FPN is defined as the potential value 4ms after t_d. The FPN value is smoothed out by multiplying the signal with a test function and then integrat the product. This proves to be less sensitive to noise than just a point-wise evaluation. Let $\phi(t_1) = \exp\left[-\frac{(t_1-(t_d+4))^2}{.04}\right]$. Then,

$$\mathrm{FPN} = \int_{t_1} y_1(t_1)\phi(t_1)\mathrm{d}t_1. \tag{24}$$

Conduction velocity (CV) The CV is defined as the velocity of the depolarization wave. Since there are 9 electrodes in the studied MEA device, one can identify the first electrode, denoted by e_a, to detect the depolarization peak at time t_a and the last one, denoted by e_b, at time t_b. The velocity is then defined as follows:

$$\mathrm{CV} = \frac{\sqrt{(x_{e_b} - x_{e_a})^2 + (y_{e_b} - y_{e_a})^2}}{t_b - t_a}. \tag{25}$$

Note that this is the only biomarker for which we do not take statistics (median, mean, etc.) over the electrodes since there is only one value for all nine electrodes.