Machine learning provides insight into models of heterogeneous electrical activity in human beta-cells

Daniele Andrean and Morten Gram Pedersen*

[a]*Department of Information Engineering, University of Padova, Via Gradenigo 6/b, I-35131 Padova, Italy*

Abstract

Understanding how heterogeneous cellular responses emerge from cell-to-cell variations in expression and function of subcellular components is of general interest. Here, we focus on human insulin-secreting beta-cells, which are believed to constitute a population in which heterogeneity is of physiological importance. We exploit recent single-cell electrophysiological data that allow biologically realistic population modeling of human beta-cells that accounts for cellular heterogeneity and correlation between ion channel parameters. To investigate how ion channels influence the dynamics of our updated mathematical model of human pancreatic beta-cells, we explore several machine learning techniques to determine which model parameters are important for determining the qualitative patterns of electrical activity of the model cells. As expected, K^+ channels promote absence of activity, but once a cell is active, they increase the likelihood of having action potential firing. HERG channels were of great importance for determining cell behavior in most of the investigated scenarios. Fast bursting is influenced by the time scales of ion channel activation and, interestingly, by the type of Ca^{2+} channels coupled to BK channels in BK-CaV complexes. Slow, metabolically driven oscillations are promoted mostly by K(ATP) channels. In summary, combining population modelling with machine learning analysis provides insight into the model and generates new hypotheses to be investigated both experimentally, via simulations and through mathematical analysis.

Keywords: Electrical activity, action potential, bursting, classification trees, logistic regression, random forest, mathematical modeling

*Corresponding author: mortengram.pedersen@unipd.it

1. Introduction

Biological heterogeneity is a fundamental fact, which is well recognized and thought to underlie, e.g., robustness and flexibility of biological systems [1, 2]. Nonetheless, most mathematical modeling of cellular dynamics choose a "typical" set (or a few sets) of parameters representing an "average" cell (cells), see e.g. [3–5]. The robustness of the model can then be investigated, e.g., by performing bifurcation analyses of the model with respect to a few number of parameters thought to be important [4–7].

However, many recent models of cellular electrophysiology contain a large number of parameters that preclude a traditional bifurcation analysis, and therefore it has been suggested to vary the parameters randomly by extracting them from an experimentally well-described distribution [8–13]. Moreover, cells may co-regulate different mechanistic components such as ion channels to obtain or maintain a certain behavior, which mathematically correspond to parameters being correlated within the cell population [14–19].

If the model and its parameters are realistic, the analysis of its behavior for different combinations of parameters then provide insight into the mechanistic control of cellular dynamics. Such insight can be obtained by direct exploration of parameter space, as has been done for models of neurons [8, 9], cardiac cells [10], pituitary cells [20] and mouse beta-cells [12]. Further understanding can be obtained with statistical analyses of the simulation results. For example, Sobie and colleagues [21, 22] investigated how the duration of the cardiac action potential depends on parameters by performing multivariate linear regression. Montefusco et al. [13] used multinomial logistic regression to investigate which ion currents determine the responses within the very heterogeneous pancreatic alpha-cell population under different conditions. Machine learning provides an alternative toolbox and could provide further insight in addition to the results obtained from the statistical methods described above [11, 23, 24].

Here we explore several machine learning techniques and population modeling based on high-quality single-cell electrophysiological data [25] to investigate how different ion channels influence the dynamics of an updated model of human pancreatic beta-cells. Pancreatic beta-cells release insulin in response to glucose following a cascade of events culminating with electrical activity, calcium influx and exocytosis of insulin containing granules [26, 27]. The heterogeneity of the beta-cell population is well established and believed to contribute to the refined glucose-sensing capabilities of the endocrine pancreas [28–33]. Mathematical modeling of electrical activity

in mouse beta-cells has a 40-year long history [34–36], and within the last decade electrical activity in human beta-cells has been modeled [37–41]. Mathematical models of exocytosis and cellular insulin release have also been developed over several decades [42–47]. However, a systematic investigation of the contributions of the different ion currents and their dynamics to causing and shaping electrical activity in human beta-cells is still lacking.

2. Methods

2.1. Human beta-cell mathematical model

We used an updated version of our previous model of human-cells [39] that includes the description of BK-CaV complexes [41]. The model is composed of a metabolic component [48] that drives a model of electrical activity [37, 39, 41], which was developed from electrophysiological data from human beta-cells [49–52]. In brief, the glycolytic enzyme phosphofructokinase (PFK) can generate metabolic oscillations due to positive feedback by its product fructobisphosphate (FBP). In the glycolytic submodel, FBP is the output that is transformed into ATP, which provides the link with the electrophysiological submodel since ATP closes ATP-sensitive potassium K(ATP)-channels. This permits other currents to depolarize the cell, which leads to opening of voltage-gated Ca^{2+} and Na^{+} channels (CaVs and NaVs, respectively) that cause action potentials (APs). The APs are ended by voltage- and calcium-sensitive K^{+} currents through Kv, SK, BK and HERG potassium channels that open as the cell depolarizes during the APs. Equations are given in the Supplementary Material.

2.2. Simulated population of human beta-cells

The model was used to simulate a population of beta-cells by generating random values for ion channel conductances and time scales. For a given cell, some of the parameters were extracted from probabilistic distributions estimated from real electrophysiological data from human beta-cells from donors without diabetes [25] (Fig. S1 in the Supplementary Material), whereas others non-measured parameters were generated randomly as explained in the next subsection.

The measures from [25] used to extract parameters directly from estimated distributions are: whole-cell capacitance, sodium channel conductance, early calcium conductance and late calcium conductance. The experimental histograms of each of these quantities were fitted by parametric

distributions, taking into consideration correlation between the electrophysiological parameters, as reported in the Supplementary Material (Figs. S1 and S2).

Calcium channel conductances (L- and P/Q-type) are calculated from the extracted early and late calcium conductances, respectively. Considering that late calcium conductance reflects almost only P/Q contribution [25], the P/Q-type CaV conductance is calculated as 90% of the late calcium conductance. L-type CaV conductance is taken as 75% of the early calcium conductance as the latter reflects mixed contributions between L- and P/Q-type Ca^{2+} channels. All conductances are converted from pS to nS and normalized by the extracted membrane capacitance.

2.2.1. Randomization of non-measured parameters

The remaining non-measured parameters were generated randomly based on the assumption that non-measured parameters have a variability that is comparable to the variability of the measured parameters. More precisely, the non-measured parameters that have been randomized are the maximal whole-cell conductances g_{SK}, g_{BK}, g_{Kv}, g_{KATP}, g_{HERG}, g_{leak}, g_{CaT} and the time scales τ_{mBK}, τ_{mKv0}, τ_{mHERG}, τ_{hHERG}, τ_{hNa}, τ_{hCaL}, τ_{hCaT}. However, since no real cell measurements on time scales are available in the exploited data, these parameters were extracted from a uniform distribution on the interval $[0.9\bar{\tau}_X, 1.1\bar{\tau}_X]$, where $\bar{\tau}_X$ denotes the default model value of the time scale τ_X. The following description of the methods therefore only refers to the non-measured channel conductances.

Our method takes the assumption that, in a single cell, the various whole-cell conductances have similar variability from their default model values. The first step consists in calculating the ratios between the parameters extracted from the experimentally estimated distributions and their corresponding default values in the single cell model. The next step is to find the maximum and minimum of these ratios to assess the variation range for each of the measured parameters. The same variation range is then used for the non-measured parameters, which is obtained by extracting a realization of a uniformly distributed random variable in this range for each of the parameters to randomize. The value of the parameter, for this cell, is then obtained as the product of the default parameter value and the realization of the uniform random variable.

This algorithm generates randomized parameters that follow lognormal distributions and are highly correlated (Figs. S3). They are also highly correlated with the measured parameters.

2.2.2. Cellular activity simulation

Using the estimation and randomization methods described above, 2000 sets of random electrophysiological parameters are obtained, representing a population of virtual cells. Each set is used to perform a simulation of cellular electrical activity with and without the metabolic subsystem. These two different conditions will be used to infer on how the different parameters control the cellular electrical behaviour. The subset of parameters in common between conditions with and without the metabolic system, e.g., ion channel conductances and timescales, are identical for the two conditions. However, the virtual cell populations are not equal due to the absence or presence of oscillatory metabolism.

2.3. Activity classification

Simulated beta-cell membrane potential traces were classified into four classes (silent, depolarized, spiking, bursting) based on the pattern of electrical activity. A fifth class named 'Other' was added for simulations with metabolic oscillations where peculiar behaviours – much different from any of the above – arose. An overview of the classification algorithm is given in Fig. S4.

Silent cells are the ones in which electrical activity does not exceed -40 mV. Depolarized cells' membrane potential is similar to the silent cells potential in shape except that it remains stable or almost stable (oscillations lower than 10 mV peak-to-peak) at a value above -40 mV. Spiking cells fire action potentials (with height >10 mV) constantly during the simulation period.

When metabolism is not simulated, cellular activity is classified as (rapid) bursting if the membrane potential oscillates multiple times above the repolarization threshold before going back below the threshold. When metabolism is oscillatory, cells may show a much slower bursting pattern driven by oscillations in K(ATP) channel conductance. During the period when K(ATP) channels conductance is low the membrane potential fires action potentials, while it remains hyperpolarized close to the resting potential when the channel conductance is higher in value.

The classification algorithm has been implemented in MATLAB R2021a (The MathWorks, Inc.) and returns a label containing the activity class of the analyzed cell, which is saved together with the cell's parameters. Details and are given in the Supplementary Material with example traces and their classification, see Figs. S5 and S6.

2.4. Machine learning analysis

The next step was to analyse the classified simulation results in order to investigate the contributions of the electrophysiological parameters on the cells' electrical behaviour.

First, the cell population was split into a modelling set of 75% and a validation set of 25% of the simulated cells. The modelling set was used for training and testing the classifiers described in the following, i.e., to obtain insight into the role of the electrophysiological parameters. During model building, this subset was further divided in training and test sets keeping 70% of the modelling set as training set and the remaining 30% as test set. The validation set was completely ignored during model construction and was used to verify that the obtained model was able to predict the classes of completely unseen data with high accuracy as measured by the area under the ROC curve (AUC), which we interpret as the model being a good representation of the data, and hence that the conclusions drawn from it are reliable.

The data analysis of each case was performed in three steps. All cells were separated into active and silent cells, where the active cells are all the cells that are not silent as defined above, i.e., depolarized, spiking or bursting (but also 'Other' in simulations with metabolic oscillations), and the machine learning classifiers described in the following were trained to classify the cells into these two subgroups. Similarly, active cells were divided into depolarized and oscillating (spiking, bursting or 'Other') cells and classification was performed. Finally, oscillating cells were analyzed by considering spiking versus bursting cells ('Others' were neglected here as they cannot be reduced to any of the two behaviours).

All analyses were performed using R/RStudio software [53, 54] on centered and scaled data.

2.4.1. Logistic regression

The data was analyzed by logistic regression with an automatic feature selection method based on elastic net regularization [55] to recognize case-by-case the most relevant features, including possible feature interactions, i.e., the products of parameters, in determining the cellular activity. The R package *glmnet* [56] was used to perform the elastic net penalized logistic regression, which has two tuning parameters to be optimized: α and λ. The glmnet package tune automatically the λ parameter using k-fold cross-validation for a given value of α. This procedure is repeated on a grid of α-values to find the optimal value of α. Further details of the implementation as well as detailed results are given in the Supplementary Material.

2.4.2. Decision trees

Decision tree analyses were performed in R using the package *rpart* [57]. Trees are built taking into consideration a user-defined complexity parameter *cp*, which controls the minimum improvement in terms of misclassification error at each node, so fixing a target complexity is therefore a stopping criteria to the tree's expansion. Low values of *cp* can lead to overfitting while higher values could reduce the classification capabilities of the tree.

Complexity has been hand-tuned case by case to obtain reasonable trees starting from the value $cp = 0.01$. The unbalance in some of the datasets could in fact lead to splits that only separated a single cell from the others, which is a clear sign of overfitting and led to higher misclassification errors in the test and left-out datasets. The opposite problem could also arise so that in some cases (mostly when dealing with the "oscillating cells" datasets) the resulting tree was too simple with a single split, and hence the complexity value had to be lowered to enhance the classification performances. All the *cp* values are around 0.01 in the $0.005-0.03$ range. Maximum tree depth has been set to 30 but this value was never reached (all the resulting trees had depth ≤ 7). This solution has been preferred over other (semi-)automatic pruning approaches since, in many cases, they resulted in uninformative trees for the oscillating cells subgroups that simply identified a single cell as belonging to the least frequent class and all the others as belonging to the most frequent one, which might be correct in terms of low error in the training and test sets but completely failed when considering the validation data untouched during model construction.

The classification trees should be read as follows. Starting from the top (the "root" of the tree) and for a given cell (set of parameters), if the shown condition is fulfilled ("yes") the cell is sent to the left, otherwise it is sent to the right. This is repeated until the cell reaches the final, lowest row (containing the "leaves" of the tree), where the prediction of the class (e.g., Active/Silent) of the cell is done. For each node of the tree, the rows indicate the predicted class, the fraction, for the two classes, of cells in the training set that reached the node, and the percentage of all cells in the training set that reached the node, respectively. The fraction of cells in the two classes can be interpreted as the probabilities that a cell that reached a certain node, e.g., a leave, belongs to the corresponding class. The predictions of the nodes are color coded with intensity reflecting the predicted probability of belonging to the predicted class.

2.4.3. Random Forests

Random forests were used specifically with the intent to determine the importance of the various physiological parameters in the activity classification process. The implementation has been done in R using the *randomForest* package [58].

For all the datasets the number of trees *nTree* was set to 2000. No limit to the trees size has been given to the *randomForest* function so by default they grow to their maximum depth possible. Pruning is not required because overfitting is avoided by the random sampling of the features. The number of parameters randomly sampled at each split to grow the trees was tuned using 4-10 fold cross validation.

By default *randomForest* calculates variable importance based on two indices: the Mean Decrease Accuracy and the Mean Decrease Gini index. The first one operates on the classification error on the Out-Of-Bag data for each tree and expresses how much accuracy the model loses by excluding a certain variable. The more accuracy suffers from removing a variable, i.e., the higher this index is, the more important that variable is. The second measure is the decrease in node impurities resulting from splitting with a certain variable, averaged over all the trees, with impurity represented by the Gini index. It is a measure of how much each variable contributes to the node homogeneity in the random forest. The higher its value, the higher the importance of the variable. Importance plots (see Supplementary Material) show that both indices agree on the most important features (usually the first 5-10) even if there are minor differences in the importance order among them. Importance of the least important variables can differ quite a lot between the two methods, but since the importance plots are L shaped, the differences in importance of these variables is negligible most of the times.

3. Results

We simulated heterogeneous populations of human beta-cells as explained in the Methods and in greater details in the Supplementary Material. The results were analyzed with logistic regression models, classification trees and random forests. A summary of the results are provided in this section (Table 1). Detailed results are presented for All Cells without the metabolic oscillator in Fig. 1, and for all cases in Figs. S7-S12 in the Supplementary Material. The performances of the classifiers for the different scenarios, as measured by the AUC metric, are presented in Table 2.

3.1. Which parameters cause electrical activity?

Our random forest analyses indicated that the maximal whole-cell conductance of the HERG K^+ channels was consistently among the most important parameters for distinguishing silent from active cells (Table 1 and Fig. 1C). This observation was confirmed by logistic regression (Fig. 1A), where it was found that an increase in g_{hERG} reduces the probability of a cell being active, in agreement with the fact that HERG channels carry an outward K^+ current. The decision trees further confirmed these observations, as g_{hERG} was located near the root of the tree (Figs. 1B and S10) and promoted the absence of cellular activity. These results agree with the findings by Rosati et al. [50] who found that inhibition of HERG channels increase insulin release from human islets.

Similarly to the HERG channels, increased K(ATP) channel conductance was found to reduce the probability of a cell showing activity. This fact is confirmed by the well-known role of these channels in beta-cell biology, which makes them a target of widely used anti-diabetic drugs that reduce K(ATP)-channel activity [59]. Other K^+ channels (Kv, SK and BK channels) appear to play no prominent role in controlling activation of human beta-cells, probably because they do not activate until the cell membrane is depolarized and Ca^{2+} flows into the cell.

Calcium and sodium channels tended to promote activation as expected due to the inward currents that they carry. In the absence of a metabolic oscillator, P/Q-type Ca^{2+} channels were among the most important parameters according to the random forest, and were found to increase the probability of observing electrical activity. With oscillatory metabolism, conflicting results were found for P/Q-type Ca^{2+} channels, which were found to promote activity according to the logistic models, whereas the decision trees found the opposite effect of P/Q-type Ca^{2+} channels. However, it should be noted that the tree had a substantially lower predictive ability as measured by the AUC (Table 2). Control simulations (Fig. 2) showed that increased g_{CaPQ} lowers the number of silent cells, confirming the prediction based on logistic regression. T- and L-type Ca^{2+} channels were found to promote electrical activity, and in the presence of metabolism L-type Ca^{2+} channels were among the most influential according to the random forest results.

Leak channels also played an important role. We found that they promote activation according to both the decision trees and logistic regression. Finally, time constants and the configuration of BK-CaV complexes do not appear to influence the onset of electrical activity.

Overall, in our model presence or absence of electrical activity is controlled mainly by a balance between hyperpolarizing K(ATP) and HERG

currents and depolarizing leak, Na^+ and Ca^{2+} currents.

3.2. Which parameters cause oscillations rather than depolarization?

HERG, Kv, and, in the presence of metabolism, K(ATP) channel conductances were the most important parameters according to the random forests, and they consistently augment the probability of oscillatory electrical activity rather than permanent depolarization, which can be explained by their repolarizing effect once the cell has been activated. For the simulations without metabolic oscillations, Ca^{2+}-sensitive SK channels also played an important role in causing membrane oscillations, consistent with the fact that they activate when the cell is depolarized, Ca^{2+} channels open and Ca^{2+} enters the cell. These channels have less effect in the presence of metabolic oscillations. Somewhat surprisingly, in our analyses BK channels have no influence on whether active cells oscillate or remain depolarized, but coupling of BK channels to T-type, rather than to L- or P/Q-type, Ca^{2+} channels was found to promote oscillatory behavior.

L-type Ca^{2+} channel conductance was among the most important parameters related to inward currents, and it tended – maybe unexpectedly – to promote depolarization, although in the absence of oscillatory metabolism the tree analysis predicted that increasing g_{CaL} raised the likelihood of obtaining oscillations. In the presence of metabolism, the decision tree found that g_{Na} promotes oscillations, whereas leak currents promote steady depolarization. The random forest analysis indicated that in this case g_{Na} was the second most important parameter. The effect of P/Q Ca^{2+} channels was less clear since it was found to depend on the value of g_{CaL}, and moreover the prediction depends non-monotonuously on g_{CaPQ} for some values of g_{CaL}. Notably, the AUC for the decision tree was substantially lower than the AUC of the logistic model (Table 2), and control simulations (Fig. 2) suggested that indeed increased g_{CaPQ} promote depolarization although the effect is less marked than in the absence of metabolism.

Overall, K^+ currents promoted oscillatory behavior of active cells due to their hyperpolarizing effect. Inward currents on the other hand increased the probability of a cell being depolarized, with some exceptions in the case of metabolic oscillations.

3.3. Which parameters cause fast bursting?

Fast bursting has a period of a few seconds and consists of a few small oscillations occurring on a depolarized plateau [37, 39, 49]. It is believed to be of physiological relevance since bursting has been proposed to cause more hormone release than simple action potential firing [60, 61].

	ALL CELLS (silent vs active)		ACTIVE CELLS (depol. vs oscillating)		OSCILLATING CELLS (spiking vs bursting)	
Metabolism	No	Yes	No	Yes	No	Yes
Cm			Osc	Osc		
			4	13		
Na	Act	Act				
				Osc		
	6	7		2		
CaL	Act		Dep	Dep		
		Act	Osc	Dep	Bur	Spk
	4	1	3	3	3	3
CaT	Act	Act				
						Spk
	8	6				8
$CaPQ$	Act	Act	Dep			
	Act	Sil		Dep/Osc		
	2	5	8	7		
$KATP$	Sil	Sil		Osc		Bur
	Sil	Sil		Osc		Bur
	5	4		1		1
$HERG$	Sil	Sil	Osc	Osc	Spk	
	Sil	Sil	Osc		Spk	
	1	2	1	8	1	
Kv	Sil	Sil	Osc	Osc		
			Osc		Spk	
	9	8	2	5	4	
SK	Act		Osc			
	7		7			
BK						
$leak$	Act	Act				Spk
	Act	Act		Dep		
	3	3		4		7
τ_{mKv0}					Spk	
					Spk	
					5	
τ_{mHERG}		Act			Bur	
		14			14	
τ_{hHERG}						
τ_{hNa}				Dep		
				19		
τ_{hCaL}			Osc	Dep		
			15	20		
τ_{hCaT}	Act					
	18					
$Coupling$		Act-T,PQ	Osc-T			
					Spk-T	
		17	5		2	
$nCaVs$						Spk
			Osc			
			13			15

Table 1: Effects of each parameter on the cellular activity grouped by subgroups (All, Active, Oscillating, respectively) and absence/presence of metabolism. Parameters cause the cells to be active/silent (Act/Sil), depolarized/oscillating (Dep/Osc), or spiking/bursting (Spk/Bur), respectively. Blue rows indicate logistic regression results while the red ones refer to the decision tree analyses. The numbers in black are the importance ranks according to random forests. Only parameters found by logistic regression or decision trees are reported, and interaction terms are not shown for logistic regression. Complete results including statistically significant interaction terms for the logistic models can be found in the Supplementary Material.

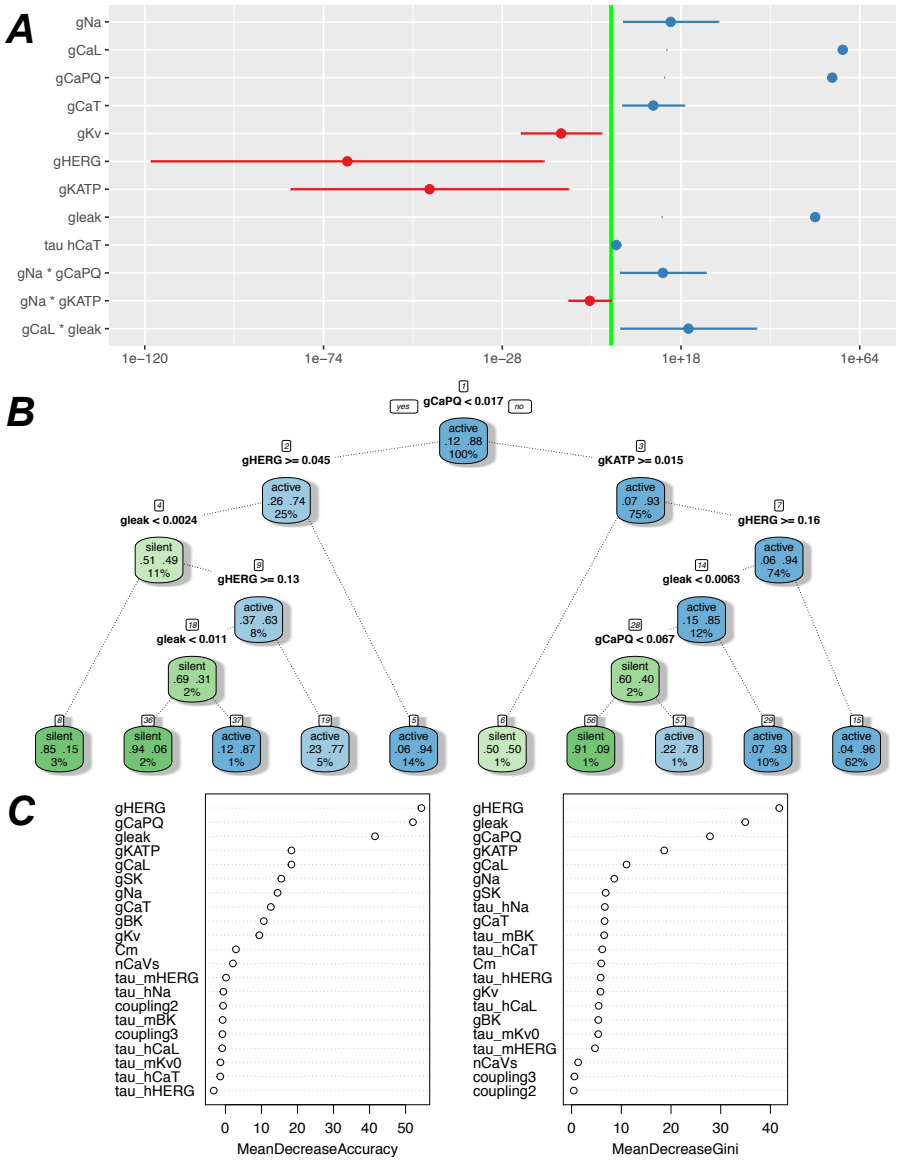

Figure 1: Graphical representation of the results for All Cells without the metabolic oscillations. *A*: Odds Ratio plot for the logistic regression analysis. Terms with asterisks are interaction terms involving the two parameters. Odds ratios greater than one (blue) indicate that larger values of the parameter increases the probability of cells being active. For further details on Odds ratios, see Section 5 in the Supplementary Material. *B*: The obtained tree. For a description of how to read the tree, see Section 2.4.2 in the Methods and Section 5 in the Supplementary Material. *C*: Random Forest variable importance plots. The parameters are ordered so that the uppermost is the most important for classification by the given forest. See Section 2.4.3 in the Methods for further description.

	NO METABOLISM	WITH METABOLISM	
ALL CELLS	0.96	0.98	LOGISTIC REGRESSION
ACTIVE	0.97	0.94	
OSCILLATING	0.75	0.99	
ALL CELLS	0.78	0.76	TREES
ACTIVE	0.90	0.86	
OSCILLATING	0.72	0.77	
ALL CELLS	0.90	0.93	FOREST
ACTIVE	0.96	0.92	
OSCILLATING	0.82	0.92	

Table 2: AUC values obtained on the test set for each method in the various scenarios.

In the absence of metabolic oscillations, HERG conductance was the most important parameter and promoted spiking electrical activity rather than bursting, in agreement with previous studies showing that bursting can occur in the model when g_{hERG} is reduced [62]. Similar effects were found for Kv potassium currents, again in agreement with previous results [37, 39]. L-type Ca^{2+} channels were found to increase the probability of observing bursting, whereas no other inward currents were found to be of importance for controlling fast bursting vs. spiking.

In contrast to the analysis of the other data subsets, the time constants of gating variables showed up among the important parameters. In addition, the interaction between BK-channels and the different types of Ca^{2+} channels was found to have an important control on the appearance of bursting versus spiking electrical activity in the case of local randomization. In particular, the presence of complexes of BK and T-type Ca^{2+} channels was found by the decision tree to increase the probability of observing spiking electrical activity, compared to the baseline of complexes with L-type Ca^{2+} channels.

These results suggest that the detailed, dynamic and nonlinear interaction between the different currents underlie fast bursting. Indeed, it has been suggested that this kind of electrical activity is a manifestation of so-called mixed-mode oscillations [62], which are well-known to depend critically of the dynamic details of model variables [63].

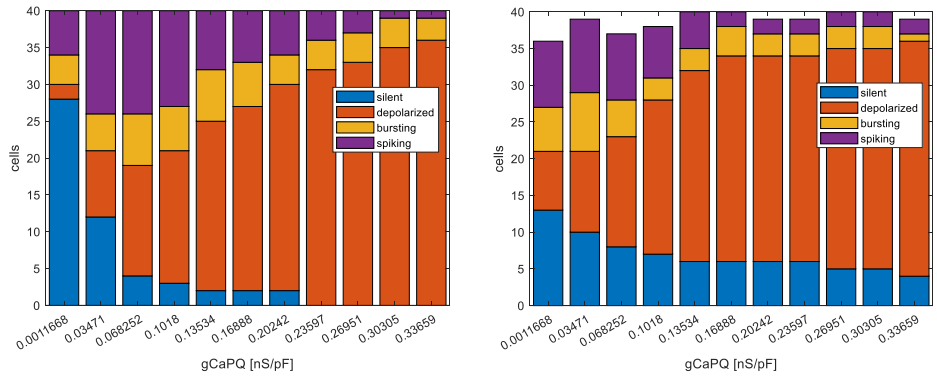

Figure 2: Control simulations showing that P/Q Ca^{2+} channels promote cellular activity. For each scenario without (left panel) or with (right panel) metabolism, 10 cells from each class were randomly extracted and simulations were performed for these 40 cells by varying the value of g_{PQ} along the grid of values shown in the figure. The white space in the right panel corresponds to cells classified as "Other".

3.4. Which parameters cause slow bursting?

Slow bursting is qualitatively different from fast bursting. It has a period of several minutes [39] and is believed to be driven by the slow oscillations in metabolism, which create periodic surges of ATP that act to close K(ATP) channels [39, 64]. The pattern is generally composed of active phases with action potential firing separated by silent phases of hyperpolarization [39].

K(ATP)-channels were found to be the main parameter distinguishing slow bursting from spiking in the presence of oscillatory metabolism. When $\bar{g}_{KATP}$ is relatively large, the K(ATP) current keeps the cell hyperpolarized during the phase when ATP levels are low. In contrast, with low $\bar{g}_{KATP}$ the changes in ATP levels cause fluctuations in K(ATP) current of insufficient magnitude to switch off electrical activity. Since all cells in this data subset have oscillatory electrical activity by definition, this means that action potentials fire continuously, producing spiking electrical activity.

The second-most important parameter was L-type Ca^{2+} channel conductance, which promoted spiking activity by causing action potentials also when ATP levels are low. A similar mechanism can explain the fact that T-type $Ca^{2+}$and leak currents were found by logistic regression to increase the probability of having spiking rather than bursting.

4. Discussion

In the present work, we have provided a first analysis of electrical activity in heterogeneous human beta-cells using a combination of mechanistic population modeling and various machine learning methods. Our analysis revealed both expected results as well as novel findings that merit further investigation.

In most of the cases, the dynamics of the gating variables, as modelled via the time constants τ_X, did not influence the results much, which to a large degree justifies the typical approach of population modeling that varies maximal channel conductances only [9, 12, 13, 20]. However, fast bursting appears to be influenced by the dynamic behavior of the gating variables. Indeed, we [62] recently showed that fast bursting in the human beta-cell model can be understood from slow-fast analysis, folded singularities and singular Hopf bifurcation, which are highly sensitive to the time-scales of the involved parameters.

Interestingly, the Random Forest analysis revealed that the type of CaV in the BK-CaV complexes is of great importance for triggering and determining the type of oscillating electrical activity (spiking or rapid bursting). The results suggested that coupling of BK channels with T-type Ca^{2+} channels promote oscillations rather than steady depolarization, and spiking rather than bursting, compared to coupling with L- or P/Q-type CaVs. This provides an example of how machine learning analyses can provide a result to be understood with further mathematical analysis of the model such as slow-fast analysis [62].

Somewhat surprisingly, the BK conductance did not appear as a highly important parameter in any of the analyses, suggesting that this parameter is not important for determining the qualitative pattern. We performed additional simulations that confirmed these findings (Fig. S13). Experimentally, it was found that BK channels are involved in controlling AP amplitude [51], and this was reproduced in simulations [37]. A limitation of the present study is that we did not investigate how quantitative characteristics of electrical activity, such as action potential amplitude, spike frequency or the number of spikes per burst, are controlled by model parameters. Future studies could investigate these issues, e.g., following the work of Sobie [21]. Such an approach also addresses drawbacks of our our labelling algorithm, which uses continuous characteristics to map the simulated cells into discrete classes, and thus has inherent difficulties in "gray zones" between classes. For example, some cells labeled as spikers have action potentials of long duration, and their patterns are similar to cases of fast bursting with

very small and few small-amplitude oscillations (Fig. S5). The difficulties of distinguishing spiking from bursting likely underlie the relatively low performance of the machine learning classifiers on the set of "oscillating cells" in the case with no metabolism (Table 2).

For some of simulated datasets, the results regarding L- and P/Q-type Ca^{2+} channels obtained from logistic regression were in contrast to the results obtained with the corresponding decision tree. We believe this fact showcases the need for applying several analytical methodologies to the synthetic cell population as different methods can reveal different aspects of the nonlinear interactions between parameters in the model. Such discrepancies indicate "where to look", i.e., which parameters control model behavior nontrivially in some parts of parameter space where methods from dynamical systems theory can be applied ("how to look") [12].

Overall, combining population modelling with machine learning analysis provides insight into the model - and thus into human beta-cell physiology - that can not be obtained by traditional population-averaged modelling. The approach generates new hypotheses to be investigated both experimentally, via simulations and through mathematical analysis. Finally, we encourage that electrophysiological studies are published at the single-cell level as in the dataset exploited here [25] in order to generate biologically realistic *in silico* cell populations, which take into account realistic parameter correlation.

Acknowlegments

This work was financially supported by MIUR (Italian Minister for Education) under the initiative 'Departments of Excellence' (Law 232/2016) and by a grant from the University of Padova (SEED2020).

Conflict of interest

These authors declare that they have no conflict of interest.

References

[1] S. J. Altschuler, L. F. Wu, Cellular heterogeneity: do differences make a difference?, Cell 141 (2010) 559–63. doi:10.1016/j.cell.2010.04.033.

[2] N. Komin, A. Skupin, How to address cellular heterogeneity by distribution biology, Current Opinion in Systems Biology 3 (2017) 154–160.

[3] A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J Physiol 117 (1952) 500–44.

[4] C. P. Fall, E. S. Marland, J. M. Wagner, J. J. Tyson, Computational cell biology, New York (2002).

[5] J. Keener, J. Sneyd, Mathematical physiology, Springer, 2009.

[6] A. Sherman, Dynamical systems theory in physiology, J Gen Physiol 138 (2011) 13–9. doi:10.1085/jgp.201110668.

[7] E. M. Izhikevich, Dynamical systems in neuroscience, MIT press, 2007.

[8] M. S. Goldman, J. Golowasch, E. Marder, L. F. Abbott, Global structure, robustness, and modulation of neuronal models, J Neurosci 21 (2001) 5229–38.

[9] A. A. Prinz, C. P. Billimoria, E. Marder, Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons, J Neurophysiol 90 (2003) 3998–4015. doi:10.1152/jn.00641.2003.

[10] O. J. Britton, A. Bueno-Orovio, K. Van Ammel, H. R. Lu, R. Towart, D. J. Gallacher, B. Rodriguez, Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology, Proc Natl Acad Sci U S A 110 (2013) E2098–105. doi:10.1073/pnas.1304382110.

[11] H. Ori, E. Marder, S. Marom, Cellular function given parametric variation in the Hodgkin and Huxley model of excitability, Proc Natl Acad Sci U S A 115 (2018) E8211–E8218. doi:10.1073/pnas.1808552115.

[12] M. Aggarwal, N. Cogan, R. Bertram, Where to look and how to look: Combining global sensitivity analysis with fast/slow analysis to study multi-timescale oscillations, Math Biosci 314 (2019) 1–12. doi:10.1016/j.mbs.2019.05.004.

[13] F. Montefusco, G. Cortese, M. G. Pedersen, Heterogeneous alpha-cell population modeling of glucose-induced inhibition of electrical activity, J Theor Biol 485 (2020) 110036. doi:10.1016/j.jtbi.2019.110036.

[14] G. LeMasson, E. Marder, L. F. Abbott, Activity-dependent regulation of conductances in model neurons, Science 259 (1993) 1915–7. doi:10.1126/science.8456317.

[15] E. Marder, A. A. Prinz, Modeling stability in neuron and network function: the role of activity in homeostasis, Bioessays 24 (2002) 1145–54. doi:10.1002/bies.10185.

[16] J. N. MacLean, Y. Zhang, B. R. Johnson, R. M. Harris-Warrick, Activity-independent homeostasis in rhythmically active neurons, Neuron 37 (2003) 109–20. doi:10.1016/s0896-6273(02)01104-2.

[17] B. Rosati, D. McKinnon, Regulation of ion channel expression, Circ Res 94 (2004) 874–83. doi:10.1161/01.RES.0000124921.81025.1F.

[18] A. M. Swensen, B. P. Bean, Robustness of burst firing in dissociated purkinje neurons with acute or long-term reductions in sodium conductance, J Neurosci 25 (2005) 3509–20. doi:10.1523/JNEUROSCI.3929-04.2005.

[19] J.-M. Goaillard, E. Marder, Ion channel degeneracy, variability, and covariation in neuron and circuit resilience, Annu Rev Neurosci 44 (2021) 335–357. doi:10.1146/annurev-neuro-092920-121538.

[20] P. Fletcher, R. Bertram, J. Tabak, From global to local: exploring the relationship between parameters and behaviors in models of electrical excitability, J Comput Neurosci 40 (2016) 331–45. doi:10.1007/s10827-016-0600-1.

[21] E. A. Sobie, Parameter sensitivity analysis in electrophysiological models using multivariable regression, Biophys J 96 (2009) 1264–74. doi:10.1016/j.bpj.2008.10.056.

[22] A. X. Sarkar, D. J. Christini, E. A. Sobie, Exploiting mathematical models to illuminate electrophysiological variability between individuals, J Physiol 590 (2012) 2555–67. doi:10.1113/jphysiol.2011.223313.

[23] M. C. Lancaster, E. A. Sobie, Improved prediction of drug-induced torsades de pointes through simulations of dynamics and machine learning algorithms, Clin Pharmacol Ther 100 (2016) 371–9. doi:10.1002/cpt.367.

[24] R. E. Baker, J.-M. Peña, J. Jayamohan, A. Jérusalem, Mechanistic models versus machine learning, a fight worth fighting for the biological community?, Biol Lett 14 (2018). doi:10.1098/rsbl.2017.0660.

[25] J. Camunas-Soler, X.-Q. Dai, Y. Hang, A. Bautista, J. Lyon, K. Suzuki, S. K. Kim, S. R. Quake, P. E. MacDonald, Patch-seq links single-cell transcriptomes to human islet dysfunction in diabetes, Cell Metab 31 (2020) 1017–1031.e4. doi:10.1016/j.cmet.2020.04.005.

[26] J. C. Henquin, Regulation of insulin secretion: a matter of phase control and amplitude modulation, Diabetologia 52 (2009) 739–51. doi:10.1007/s00125-009-1314-y.

[27] P. Rorsman, F. M. Ashcroft, Pancreatic β-cell electrical activity and insulin secretion: Of mice and men, Physiol Rev 98 (2018) 117–214. doi:10.1152/physrev.00008.2017.

[28] C. F. H. Van Schravendijk, R. Kiekens, D. G. Pipeleers, Pancreatic β cell heterogeneity in glucose-induced insulin secretion, J. Biol. Chem. 267 (1992) 21344–21348.

[29] D. Pipeleers, R. Kiekens, Z. Ling, A. Wilikens, F. Schuit, Physiologic relevance of heterogeneity in the pancreatic beta-cell population, Diabetologia 37 Suppl 2 (1994) S57–64. doi:10.1007/BF00400827.

[30] D. Pipeleers, I. De Mesmaeker, T. Robert, F. Van Hulle, Heterogeneity in the beta-cell population: a guided search into its significance in pancreas and in implants, Curr Diab Rep 17 (2017) 86. doi:10.1007/s11892-017-0925-9.

[31] R. K. P. Benninger, D. W. Piston, Cellular communication and heterogeneity in pancreatic islet insulin secretion dynamics, Trends Endocrinol Metab 25 (2014) 399–406. doi:10.1016/j.tem.2014.02.005.

[32] A. Stožer, M. Skelin Klemen, M. Gosak, L. Križančić Bombek, V. Pohorec, M. Slak Rupnik, J. Dolenšek, Glucose-dependent activation, activity, and deactivation of beta cell networks in acute mouse pancreas tissue slices, Am J Physiol Endocrinol Metab 321 (2021) E305–E323. doi:10.1152/ajpendo.00043.2021.

[33] G. Cappon, M. G. Pedersen, Heterogeneity and nearest-neighbor coupling can explain small-worldness and wave properties in pancreatic islets, Chaos 26 (2016) 053103. doi:10.1063/1.4949020.

[34] T. R. Chay, J. Keizer, Minimal model for membrane oscillations in the pancreatic beta-cell., Biophys J 42 (1983) 181–190. doi:10.1016/S0006-3495(83)84384-7.

[35] M. G. Pedersen, Contributions of mathematical modeling of beta cells to the understanding of beta-cell oscillations and insulin secretion., J Diabetes Sci Technol 3 (2009) 12–20. doi:10.1177/193229680900300103.

[36] I. Marinelli, P. A. Fletcher, A. S. Sherman, L. S. Satin, R. Bertram, Symbiosis of electrical and metabolic oscillations in pancreatic β-cells, Front Physiol 12 (2021) 781581. doi:10.3389/fphys.2021.781581.

[37] M. G. Pedersen, A biophysical model of electrical activity in human β-cells., Biophys J 99 (2010) 3200–3207. doi:10.1016/j.bpj.2010.09.004.

[38] L. E. Fridlyand, D. A. Jacobson, L. H. Philipson, Ion channels and regulation of insulin secretion in human β-cells: a computational systems analysis, Islets 5 (2013) 1–15. doi:10.4161/isl.24166.

[39] M. Riz, M. Braun, M. G. Pedersen, Mathematical modeling of heterogeneous electrophysiological responses in human β-cells, PLoS Comput Biol 10 (2014) e1003389. doi:10.1371/journal.pcbi.1003389.

[40] A. Loppini, M. Braun, S. Filippi, M. G. Pedersen, Mathematical modeling of gap junction coupling and electrical activity in human β-cells, Phys Biol 12 (2015) 066002. doi:10.1088/1478-3975/12/6/066002.

[41] F. Montefusco, A. Tagliavini, M. Ferrante, M. G. Pedersen, Concise whole-cell modeling of BKCa-CaV activity controlled by local coupling and stoichiometry, Biophys J 112 (2017) 2387–2396. doi:10.1016/j.bpj.2017.04.035.

[42] G. M. Grodsky, A threshold distribution hypothesis for packet storage of insulin and its mathematical modeling., J. Clin. Invest. 51 (1972) 2047–2059. doi:10.1172/JCI107011.

[43] A. Bertuzzi, S. Salinari, G. Mingrone, Insulin granule trafficking in beta-cells: mathematical model of glucose-induced insulin secretion., Am J Physiol Endocrinol Metab 293 (2007) E396–E409. doi:10.1152/ajpendo.00647.2006.

[44] M. G. Pedersen, A. Sherman, Newcomer insulin secretory granules as a highly calcium-sensitive pool., Proc Natl Acad Sci U S A 106 (2009) 7432–7436. doi:10.1073/pnas.0901202106.

[45] F. Montefusco, M. G. Pedersen, Explicit theoretical analysis of how the rate of exocytosis depends on local control by Ca2+ channels, Comput Math Methods Med 2018 (2018) 5721097. doi:10.1155/2018/5721097.

[46] F. Montefusco, M. G. Pedersen, From local to global modeling for characterizing calcium dynamics and their effects on electrical activity and exocytosis in excitable cells, Int J Mol Sci 20 (2019). doi:10.3390/ijms20236057.

[47] M. G. Pedersen, A. Tagliavini, J.-C. Henquin, Calcium signaling and secretory granule pool dynamics underlie biphasic insulin secretion and its amplification by glucose: experiments and modeling, Am J Physiol Endocrinol Metab 316 (2019) E475–E486. doi:10.1152/ajpendo.00380.2018.

[48] P. O. Westermark, A. Lansner, A model of phosphofructokinase and glycolytic oscillations in the pancreatic beta-cell, Biophys J 85 (2003) 126–39. doi:10.1016/S0006-3495(03)74460-9.

[49] S. Misler, D. W. Barnett, K. D. Gillis, D. M. Pressel, Electrophysiology of stimulus-secretion coupling in human beta-cells., Diabetes 41 (1992) 1221–1228.

[50] B. Rosati, P. Marchetti, O. Crociani, M. Lecchi, R. Lupi, A. Arcangeli, M. Olivotto, E. Wanke, Glucose- and arginine-induced insulin secretion by human pancreatic beta-cells: the role of HERG K(+) channels in firing and release., FASEB J 14 (2000) 2601–2610. doi:10.1096/fj.00-0077com.

[51] M. Braun, R. Ramracheya, M. Bengtsson, Q. Zhang, J. Karanauskaite, C. Partridge, P. R. Johnson, P. Rorsman, Voltage-gated ion channels in human pancreatic beta-cells: electrophysiological characterization and role in insulin secretion., Diabetes 57 (2008) 1618–1628. doi:10.2337/db07-0991.

[52] J. Herrington, M. Sanchez, D. Wunderler, L. Yan, R. M. Bugianesi, I. E. Dick, S. A. Clark, R. M. Brochu, B. T. Priest, M. G. Kohler, O. B. McManus, Biophysical and pharmacological properties of the voltage-gated potassium current of human pancreatic beta-cells., J Physiol 567 (2005) 159–175. doi:10.1113/jphysiol.2005.089375.

[53] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. URL: `https://www.R-project.org/`.

[54] RStudio Team, RStudio: Integrated Development for R, RStudio, PBC, Boston, MA, USA, 2020.

[55] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Methodological) 67 (2005) 301–320. doi:10.1111/j.1467- 9868.2005.00503.x.

[56] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, Journal of statistical software 33 (2010) 1.

[57] T. Therneau, B. Atkinson, rpart: Recursive Partitioning and Regression Trees, 2019. URL: `https://CRAN.R-project.org/package=rpart`, r package version 4.1-15.

[58] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (2002) 18–22. URL: `https://CRAN.R-project.org/doc/Rnews/`.

[59] F. M. Ashcroft, P. Rorsman, K(ATP) channels and islet hormone secretion: new insights and controversies, Nat Rev Endocrinol 9 (2013) 660–9. doi:10.1038/nrendo.2013.166.

[60] A. Tagliavini, J. Tabak, R. Bertram, M. G. Pedersen, Is bursting more effective than spiking in evoking pituitary hormone secretion? A spatiotemporal simulation study of calcium and granule dynamics, Am J Physiol Endocrinol Metab 310 (2016) E515–25. doi:10.1152/ajpendo.00500.2015.

[61] M. G. Pedersen, G. Cortese, L. Eliasson, Mathematical modeling and statistical analysis of calcium-regulated insulin granule exocytosis in β-cells from mice and humans, Prog Biophys Mol Biol 107 (2011) 257–64. doi:10.1016/j.pbiomolbio.2011.07.012.

[62] S. Battaglin, M. G. Pedersen, Geometric analysis of mixed-mode oscillations in a model of electrical activity in human beta-cells, Nonlinear Dynamics 104 (2021) 4445–4457.

[63] M. Desroches, J. Guckenheimer, B. Krauskopf, C. Kuehn, H. M. Osinga, M. Wechselberger, Mixed-mode oscillations with multiple time scales, SIAM Review 54 (2012) 211–288. doi:10.1137/100791233.

[64] I. Marinelli, B. M. Thompson, V. S. Parekh, P. A. Fletcher, L. Gerardo-Giorda, A. S. Sherman, L. S. Satin, R. Bertram, Oscillations in K(ATP) conductance drive slow calcium oscillations in pancreatic β-cells, Biophys J (2022). doi:10.1016/j.bpj.2022.03.015.