Computers in Biology and Medicine 61 (2015) 1–7

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Signal analysis and classification methods for the calcium transient data of stem cell-derived cardiomyocytes

Martti Juhola [a,*], Kirsi Penttinen [b], Henry Joutsijoki [a], Kirsi Varpa [a], Jyri Saarikoski [a], Jyrki Rasku [a], Harri Siirtola [a], Kati Iltanen [a], Jorma Laurikkala [a], Heikki Hyyrö [a], Jari Hyttinen [c], Katriina Aalto-Setälä [b,d]

[a] School of Information Sciences, University of Tampere, Tampere, Finland
[b] BioMediTech and School of Medicine, University of Tampere, Tampere, Finland
[c] Department of Electronics and Communication Engineering, Tampere University of Technology, Tampere, Finland
[d] Heart Hospital, Tampere University Hospital, Tampere, Finland

ARTICLE INFO

Article history:
Received 9 December 2014
Accepted 16 March 2015

Keywords:
Human pluripotent stem cells
Cardiomyocytes
Calcium cycling abnormalities
Biomedical data analysis

ABSTRACT

Calcium cycling is crucial in the excitation–contraction coupling of cardiomyocytes, and therefore has a key role in cardiac functionality. Cardiac disorders and different drugs alter the calcium transients of cardiomyocytes and can cause serious dysfunction of the heart. New insights into this biochemical phenomena can be achieved by studying and analyzing calcium transients. Calcium transients of spontaneously beating human induced pluripotent stem cell-derived cardiomyocytes were recorded for a data set of 280 signals. Our objective was to develop and program procedures: (1) to automatically detect cycling peaks from signals and to classify the peaks of signals as either normal or abnormal, and (2) on the basis of the preceding peak detection results, to classify the entire signals into either a normal class or an abnormal class. We obtained a classification accuracy of approximately 80% compared to class decisions made separately by an experienced researcher, which is promising for the further development of an automatic classification approach. Automated classification software would be beneficial in the future for analyzing cardiomyocyte functionality on a large scale when screening for the adverse cardiac effects of new potential compounds, and also in future clinical applications.

1. Introduction

Calcium (Ca^{2+}) cycling plays a critical role in the excitation–contraction coupling of cardiomyocytes, and it is the link between the electrical signaling in the cardiomyocyte and contraction. Changes and variability in Ca^{2+} transients can be seen because of cardiac diseases or different drugs, which can have profound consequences for the function and phenotype of cardiomyocytes. For example, when cardiac failure progresses, changes in Ca^{2+} regulation and flux are observed, and failing hearts are also characterized by more arrhythmic Ca^{2+} signals [1]. Characterization of Ca^{2+} cycling is crucial in cardiac research in order to facilitate investigations of cardiac disorders and dysfunction, and to study disease management with different compounds. Ca^{2+} imaging of cardiomyocytes is a widely used technique for monitoring their Ca^{2+} cycling activity in vitro. Intracellular Ca^{2+} cycling can be recorded with the help of fluorescent Ca^{2+} indicator dyes.

Cardiac functionality can be studied with the help of cardiomyocytes differentiated from human pluripotent stem cells [2–4]. Induced pluripotent stem cell (iPSC) technology – where pluripotent stem cells are generated by reprogramming differentiated cells into a pluripotent state – provides an especially useful tool for studying the pathophysiology of various disorders and drug responses in human cells. Human iPSCs can be differentiated into the desired cell type, retaining the original genotype. New insights into calcium handling in different cardiac diseases have been achieved after the invention of iPSCs [5–10].

To empower cardiologic investigations, we have developed a signal analysis procedure for the detection and classification of Ca^{2+} cycling or transients (peaks) in cardiomyocyte signal data, plus another procedure for the classification of entire signals into either a normal or abnormal class on the basis of results of the preceding procedure. These computational tasks are essential for the development of automatic tools for the selection of valid cell lines, the observation of abnormal Ca^{2+} transients, and the analysis of different drug responses.

* Correspondence to: School of Information Sciences, University of Tampere, Tampere 33014, Finland. Tel.: +358 40 1901716; fax: +358 3 2191001.
E-mail address: Martti.Juhola@sis.uta.fi (M. Juhola).

http://dx.doi.org/10.1016/j.compbiomed.2015.03.016
0010-4825/

Spontaneously beating cardiomyocytes were differentiated from iPSCs derived from patients with catecholaminergic polymorphic ventricular tachycardia (CPVT), a genetic cardiac disease. CPVT is an exercise-induced malignant arrhythmogenic disorder, which can cause increased calcium (Ca^{2+}) sensitivity and lead to spontaneous Ca^{2+} release from the sarcoplasmic reticulum, the generation of after-depolarizations, and triggered activity [11]. We were interested in the recognition of the Ca^{2+} transient abnormalities of these cardiomyocytes, as revealed by the frequency and especially the shape deformations being manifested within the peaks of the cycling signals. For this purpose, we developed signal analysis and classification procedures to enable the automatic processing of cardiomyocyte data [12]. In the present study, we have developed and extended our method to identify and classify Ca^{2+} transients efficiently. To the best of our knowledge, this kind of classification has so far been done only subjectively and visually. We have also developed a separate method for the interactive use of human experts to select and analyze Ca^{2+} transients that supports our efforts for the current method [13]. This analysis area will be important for the cardiology and pharmaceutical industry and, therefore, computational methods will be needed for future medical research and its applications.

2. Cell data

The study was approved by the Ethics Committee of Pirkanmaa Hospital District (R08070). Patient-specific iPSC lines were established with retroviruses encoding for OCT4, SOX2, KLF4, and MYC, as described earlier [2]. All the cell lines were characterized for their karyotypes, mutations, pluripotency by RT-PCR, immunocytochemistry, embryoid body (EB), and teratoma formation. The IPSCs were then co-cultured with murine visceral endoderm-like (END-2) cells (Humbrecht Institute, Utrecht, The Netherlands) to differentiate them into spontaneously beating cardiomyocytes. The beating areas of the cell colonies were dissociated mechanically and enzymatically with collagenase A (Roche Diagnostics) [14].

Ca^{2+} imaging was conducted in spontaneously beating, 4 μM Fura-2 AM (Invitrogen, Molecular Probes)—loaded dissociated cardiomyocytes as described earlier [7]. Cardiomyocytes were continuously perfused with 37 °C HEPES-based perfusate during measurements. The perfusate consisted of (in mM) 137 NaCl, 5 KCl, 0.44 KH_2PO_4, 20 HEPES, 4.2 $NaHCO_3$, 5 D-glucose, 2 $CaCl_2$, 1.2 $MgCl_2$, and 1 Na-pyruvate (the pH was adjusted to 7.4 with NaOH). Ca^{2+} measurements were conducted on an inverted IX70 microscope (Olympus Corporation, Hamburg, Germany), and cells were visualized with a UApo/340 × 20 air objective (Olympus). Images were taken with an ANDOR iXon 885 CCD camera (Andor Technology, Belfast, Northern Ireland) and synchronized with a Polychrome V light source by a real time DSP control unit and TILLvisION or Live Acquisition software (TILL Photonics, Munich, Germany). Fura 2-AM in the cardiomyocytes was excited at a light wavelength of 340 nm and 380 nm, and the emission was recorded at 505 nm. For Ca^{2+} analysis, regions of interest were selected for spontaneously beating cells and background noise was subtracted before further processing. Signals were acquired as the ratio of the emissions at 340/380 nm wavelengths.

3. Signal data and its preprocessing

The data were generated with two different software programs. Thus, various sampling frequencies were used to record signal data: frequencies of approximately 8, 10, and 11 Hz were used with one program and a frequency of 23 Hz was used with the other program. Cycling peaks and other properties varied remarkably in signals. We recorded 280 signals, the lengths of which varied from approximately 11 to 24 s. The recorded signals were fairly short since the Ca^{2+} imaging method can damage the cells by phototoxicity; therefore, this limits their exposure time. On the other hand, short signals contained scant information, such as peaks, making the computational decision-making (classification) tasks difficult.

At first, a linear descending trend was removed from each signal according to the best straight-line fit, because such a trend was present in all signals. These modified signals were only used up to the step of the peak detection, after which, the feature values of peaks detected were computed from the original signals. Removal of a trend was used in order to facilitate peak detection. Signals of the highest sampling frequency were also filtered with a median filter [15], including a filtering window of 3 samples. However, this was not done for the signals of the lower sampling frequencies of 8, 10, and 11 Hz, since the smallest pertinent peaks of these signals included only a few samples, i.e., they would have been too sensitive even to this light filtering. By using median filtering for the highest sampling frequency signals (23 Hz), the signals were smoothed out to more closely resemble the lower sampling frequencies' types of signals. Furthermore, the same constant thresholds for peak detection and some of the percentage bounds for peak classification employed for the low frequencies' signals also suited the higher frequency signals better. The signal's minimum was subtracted from all samples to set a zero minimum for simplicity, e.g., for visual exploration in later figures.

All samples were computed as amplitude values to explore the distribution of these values. In order to compare the average amplitude values at the beginning and the end, distribution estimate A was computed for an average amplitude of the large peaks (see Fig. 1). The purpose of A is simply to aid the subsequent peak detection as a rough estimate. The amplitudes of normal peaks are virtually always this kind of large peak. The amplitudes of abnormal peaks can be either smaller than the normal ones or also equally high. Of course, their shapes could also be affected (Figs. 1–3).

Ultimately, the first derivative signal was approximated from a preprocessed signal using linear regression by sliding it through the signal as a window of 3 or 7 (the latter for the highest sampling

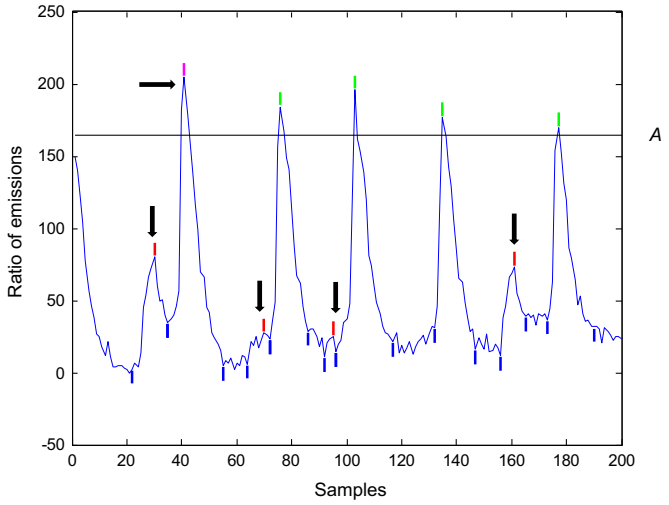

Fig. 1. A signal of 19.2 s sampled at 10.4 Hz, after the removal of a linear trend. A few peaks of small amplitudes with vertical black arrows were recognized as abnormal by the peak procedure since they were too small compared to amplitude estimate A. Therefore, the whole signal was determined to be abnormal by the procedure, and this was also judged to be the case by the human expert. In addition, the procedure assessed one larger peak (with the black horizontal arrow) as abnormal due to its asymmetry. There is no quantitative unit for the vertical axis since abscissa values are ratios of two measured values. The peak classification procedure recognized the peaks with green markers (without black errors) as normal and the others as abnormal.