Universiteit Leiden
The Netherlands

Opleiding Bioinformatica

Novel Way for the Calculation of Sarcomere Length in Human Pluripotent Stem Cell Derived Cardiomyocytes

Tijmen van Wel(s2637618)

Supervisors:
Lu Cao & Rohola Hosseini

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

19/08/2023

Abstract

Current day high throughput image analysis measurement of sarcomere almost exclusively makes use of FFT or a variant. This paper shows the limitations of FFT calculations at low data intensity. It proposes and alternative way of calculating the average length of sarcomeres in single cell live images of human pluripotent stem cell derived cardiomyocytes. Furthermore it shows the potential effects of Doxorubicin on the growth of sarcomeres.

Contents

1 Introduction

The development of medicine is subject to rigorous screening and safety protocols in order to ensure both their effectiveness as well as limit their harmfulness [EMA]. The key word that must be noted here is limit as only rarely a medicine is completely harmless. A group where the harmfulness of the medicine is infamous is cancer medication. While hair loss is amongst the most visible side effects, many treatments are cardio-toxic. This can lead to maladies ranging from mild blood pressure changes to heart failure[PN00].

1.1 Cardio-toxicity Testing

The common way for testing the cardio-toxicity of cancer medicine is thru animal testing. However, [CS23] proposed a new way of testing, by employing high throughput imaging and analysis of cardiomyocytes from human pluripotent stem cells (hPSC-CMs). This method uses a combination of image processing and myofibril tracing in order to determine myofibril density and organization. It provides a pipeline for sarcomere detection using a 2D-FFT bandpass filter, as well as a rule-based myofibril detection method. These are sufficient to show the the effects of Doxorubicin exposure in hPSC-CMs.

1.2 Cardiomyocytes

Cardiomyocytes are the muscle cells in our heart, responsible for the contractions of the atriums and ventricles. They are highly specialized cells, not only do they function 24 hours a day, they also slowly regenerate, meaning that only about 50% of the cardiomyocytes in our heart ever get renewed [PE09]. Because of this, any form of heart damage can take a very long time to repair. Cardiomyocytes can be grown from pluripotent stem cells, which can be genetically edited by using cas9 to include required reporter strains [RS20].

1.2.1 Myofibrils

The main functional component of cardiomyocytes are the myofibrils. An often linear structure of various proteins, it consists mainly of α-actinin, titin and myosin. Myofibrils begin formation as premyofibrils. Premyofibrils are highly unstructured, with actin, small pockets of α-actinin and the non muscle version of myosin II [SW10]. When growth progresses the α-actinin first forms a bead like structure, and then combines to form the linear z-disks. The non muscle myosin is lost and the muscle variant is incorporated into the myofibrils. In this stage the titin is incorparated as well. The fully grown myofibrils consist of a linear structure of z-disks between which the sarcomeres enable contraction.

1.2.2 Z-disk labeling

α-actinin is a protein located at these z-disks, and as such a prime target for labeling when researching muscle tissue. The gene responsible for the expression of α-actinin in cardiac muscle is α-actinin2 and is situated on chromosome 1. While predominantly and majorly expressed in the heart, it is also slightly expressed in the brain, esophagus and prostate [NCB]. This gene can be

labeled with fluorescent protein in order to study the z-disk structure under a capable microscope [CS23].

1.3 High Throughput Screening

The increasing understanding of molecular biology and genetics in the 90s revealed the need for methods of generating and processing high amounts of data. Together with the ever increasing capabilities of computers this resulted in the advent of high throughput screening. A term that covers a variety of methods, it has taken a prominent place in biological research ever since the latter half of the 90s [HP00]. Since then methods have kept developing across a multitude of fields, leading to various discoveries. Current software allows processing of images in under one second. Furthermore an important aspect of high throughput processing is that it is automated. This allows all images to be processed without any human input.

1.4 Research Question

The images resulting from z-disk labeling can be analyzed in high throughput image analysis in order to determine several traits of the myofibrils in analyzed cells [CS23]. But the length calculation of sarcomeres has only been done in well developing cardiomyocytes using FFT methods [PK13], [PS15]. It has never been studied in regards to treatment with Doxorubicin. This led to the following research question; How can sarcomere lengths in HPSC-CMs be calculated at low data counts and what effect does Doxorubicin have on sarcomere length?

1.5 Thesis overview

This bachelor thesis, written at LIACS and supervised by Lu Cao and Rohola Hosseini, consists of a number of sections. Above one can find the introduction. Any important definitions can be found in section 2. Work related to the thesis can be found in section 3. This includes the source of the images used in the research. The method, as well as the experiments and results can be found in section 4. Lastly section 5 will talk about conclusions and any future work.

2 Definitions

Before diving deeper into this paper it is important to get a few terms worked out. Firstly this research makes use of human pluripotent stem cell derived cardiomyocytes. These will from here on out be referred to as hPSC-CMs. Furthermore two methods of calculation will be talked about. Firstly originating from a binary mask(BM) representation. The other is a fourier frequency transform (FFT) representation. Sometimes referred to as 2D-FFT as the transformed images are in 2D. A third calculation which is talked about is a representation where imagej skeletonization is used. This will be referred to as Skel. Lastly the term N-dependent accuracy refers to accuracy of the software relative to the number of z-disks found.

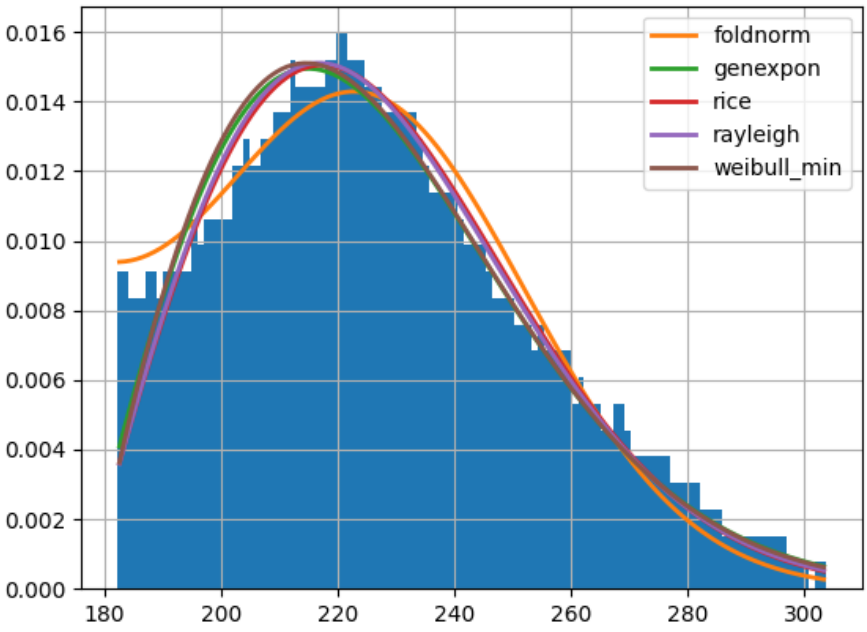

Figure 1: Various Distributions Fitted to a Dataset

3 Related Work

A lot of groundwork has already been completed in previous research at liacs. Firstly [CS23] has shown the potential for high throughput image analysis of sarcomere structures in hPSC-CMs. Together with [CvdM20] this showed the potential for using high throughput image analysis for the determining cardiotoxicity of various cancer treatments. Other essential work this project builds on is the development of the fitter package. This is a package which is able to fit distributions to data. It was developed by T. Cokelaer and various other contributors on GitHub [FIT]. The imagej application was also vital for this project [AM04]. It provides the means to process images for a vast variety of applications. Lastly the measurement of sarcomeres in stimulated cardiomyocytes using fourier transform calculations was previously done by [PG16].

3.1 Fitter

The workings of the FFT method are based on fitting a distribution to the found data. In order to do this the Fitter module has been used [FIT]. Developed primarily by T. Cokelear, it is a python module designed in order to provide a clear interface for the estimation of paramaters for a number of distributions. It uses the scipy stats module, but with additional streamlining and clearness. It allows the fitting of all available distributions to data. It then provides the best fits in both graph form for easy visualization, see Figure 1, as well as in data able to be processed by python. Scipy is an open source python module, developed on GitHub. It has a wide variety of features, ranging from linear algebra functions, to mathematical constants and signal processing.

3.2 Sarcomere Length Calculation

Sarcomere length calculation can be divided up into two distinct groups. Firstly the measurement of sarcomeres in adult animal heart muscle, and that in HPSC-CMs. One of the software packages developed for sarcomere measurements in adult cells is the SarcOptiM package developed for imageJ [PG16]. It is an addon for the calculation of sarcomere length which employs FFT transformations. There is an online mode for live calculations, although this is limited to single axis FFT transformations due to lack of processing power. However, it is possible to use this package with 2D-FFT on prerecorded video's. The development and accompanying research showed the viability of the FFT method in calculating sarcomere lengths. Length calculation in HPSC-CMs has been done amongst others by [DP18], which used a unknown method of custom programming based on mathlab in order to find distances in growing HPSC-CMs. Research by [PS15] used a variant of FFT calculation in combination with orientation to calculate sarcomere lengths.

3.3 Sarcomere Contractions

As discussed previously sarcomeres function as contractile units, which provides additional challenges when measuring their length. Research has been done by [RT15] in order to determine specifications for this contraction. This was done by recording aligned self pacing cardiomyocytes. Displacement was about 0.2 μm between contracted and extended. This was also shown in research by [MB19]. This research showed a rhythmic contraction pattern with again a contraction distance of about 0.2 μm. For both study's the total sarcomere length in its extended state equaled around 1.8 μm. It must however be noted that the first study used HPSC-CMs and the second study used isolated heart cells.

3.4 Material Gathering

The data used for testing was taken from the research done by [CS23]. This data included images processed to various degrees by methods explained below. A full detailed description of how these images were obtained can be found in [CS23]

3.4.1 Cell Preparation

Firstly hPSC-CMs were differentiated from double reporter mRubyII-ACTN2 and DRRAGN hPSC. This was done using growth factor differentiation, described by [RS20]. After 14 days of differentiation, these cells were sorted and seeded at a density of 50,000 cells per well. 10 to 12 days after seeding the cells were treated with either DMSO for the control group or Doxorubicin, an anticancer drug, for a period of 5 days. The drug was only applied on day zero.

3.4.2 Screening & Segmentation of Z-lines

Images of individual cells were acquired by the EVOS FL auto 2 (Thermofisher) microscope which can be used for high throughput imaging. The segmentation is then done by using the pipeline seen in figure 2. The images were segmented by firstly preforming background subtraction and contrast enhancement. Following this, a 2D-FFT operator with a custom bandwidth filter of 1.2-2.0 μm was applied. Firstly, the upper and lower bounds are calculated using the equation:

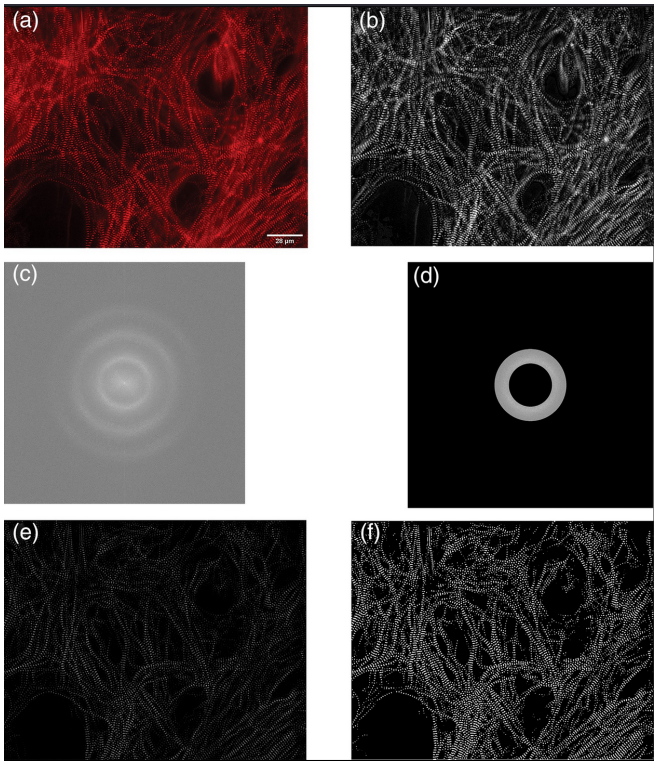

Figure 2: Pipeline for the creation of binary mask

$Frequency = N/(R * SL)$ with N being the lowest power of 2 that is higher than the max(width, height) of the original image, R the pixel to μm ratio, which in this case is 5.6, and SL being the sarcomere length, which in this case is the two decided bounds. Calculation of these two bounds leads to boundaries in the frequency domain of 180 to 300 pixels away from the origin. Once all data outside these bounds was removed, a reverse FFT operation is applied to restore the image with just the signals in the predefined frequency band. Lastly, a hard threshold is applied to filter out any trivial signals. The result is the binary mask representation. For the FFT images the raw red channel was converted by 2D-FFT (imagej).

4 Experiments

A variety of experiments were done in order to determine the accuracy of the methods developed. These tests included both the optimization of parameters as well as the gathering of data to determine runtime and differences in results between the two methods. These results are used to prove the accuracy of the BM method.

4.1 Calculation Methods

This paper tests two different methods of calculation. The first method use skeletonization of the binary mask as the basis. This method has been chosen because it has the ability to calculate sarcomere distances in cells with little information or paternization. This is useful because the

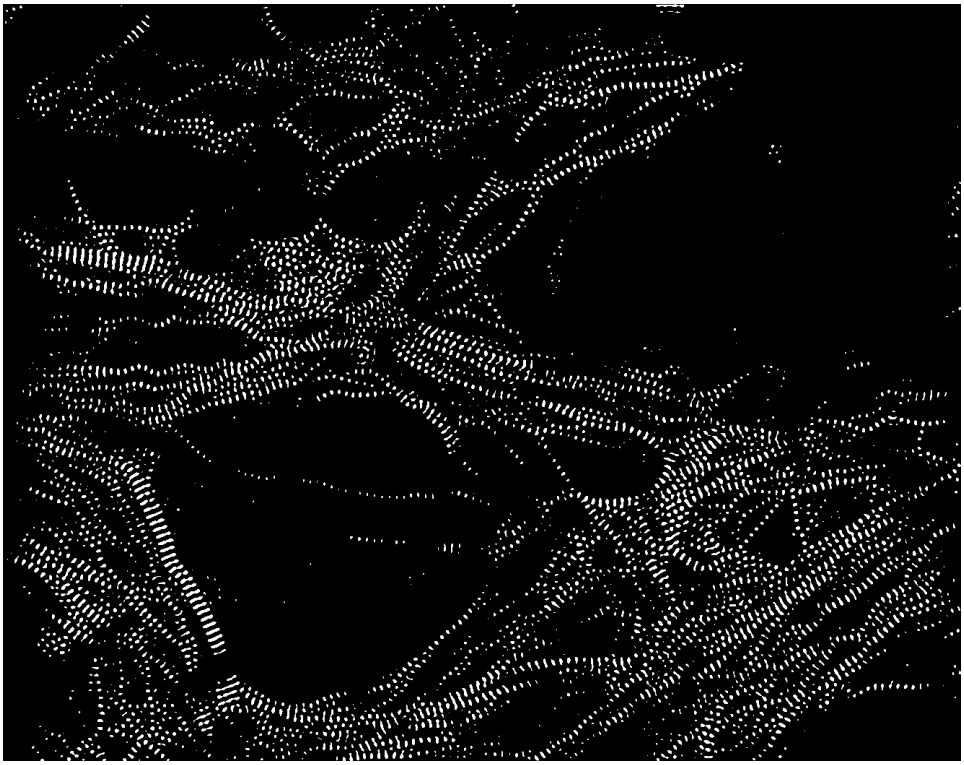

Figure 3: Image after processing explained in Section 3

HPSC-CMs are in an early stage of development in which the structure is not yet properly striated. This is something that the second method relies on. This method is length calculation using FFT conversions of the image. A tried and tested method, especially on fully grown cells, it is fast and accurate. The two main criteria when searching for methods useful in high throughput imaging.

4.1.1 Binary Mask

The basic principle for the calculation using the binary mask images is coordinate systems. However, while this sounds simple, implementing it requires more steps than one might initially expect. Firstly, the images are skeletonized using the skeletonization function of sci-img [sci], resulting in an image alike to Figure 4. A double loop is used to find an edge-point of each z-line skeleton, and a recursive function is then able to find the other edge point. These two points form a coordinate pair, and for each the length and angle in respect to the x-axis is calculated. Here an optional filter for length can also be applied. The second part of the program checks whether two pairs of lines are neighbouring z-lines. This is done amongst other things by checking whether the two pairs adhere to two paramaters. A maximum distance they can be apart, and a maximum difference in angle. If both parameters are satisfied, the program calculates the crossing points of the two perpendiculars, with starting points at either end and the other z-line, as can be seen in figure 11. The reason for only calculating from either edge point, instead of for example, a middle point, is that the current calculation in most cases yields two lengths, one shorter and one longer. This will always result in an average during the final calculation, while a distance calculated from the middle point might be skewed. After finding the intersection points, the corresponding tuples all get saved to a list. From

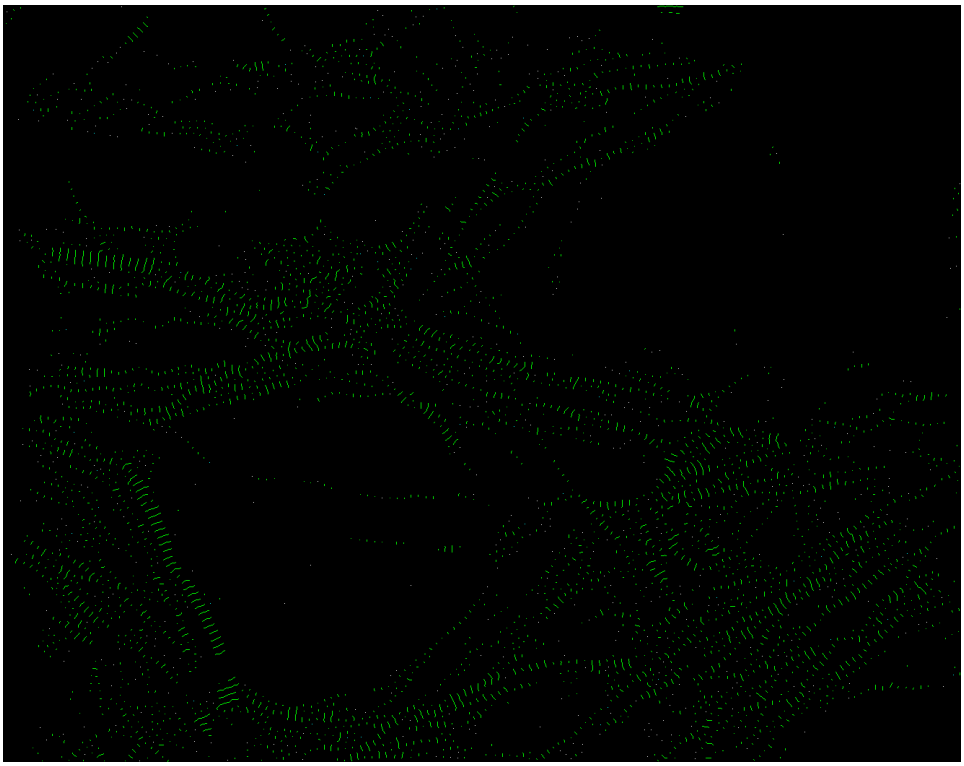

Figure 4: BM skeletonized image of Figure 3

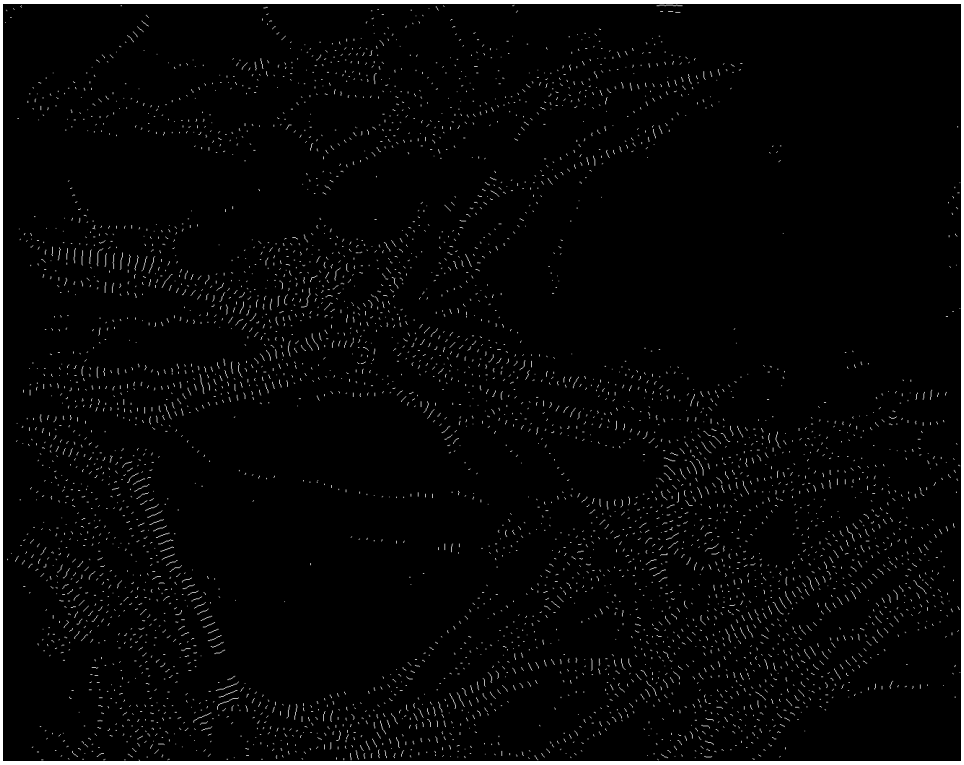

Figure 5: ImageJ skeletonized version of Figure 3

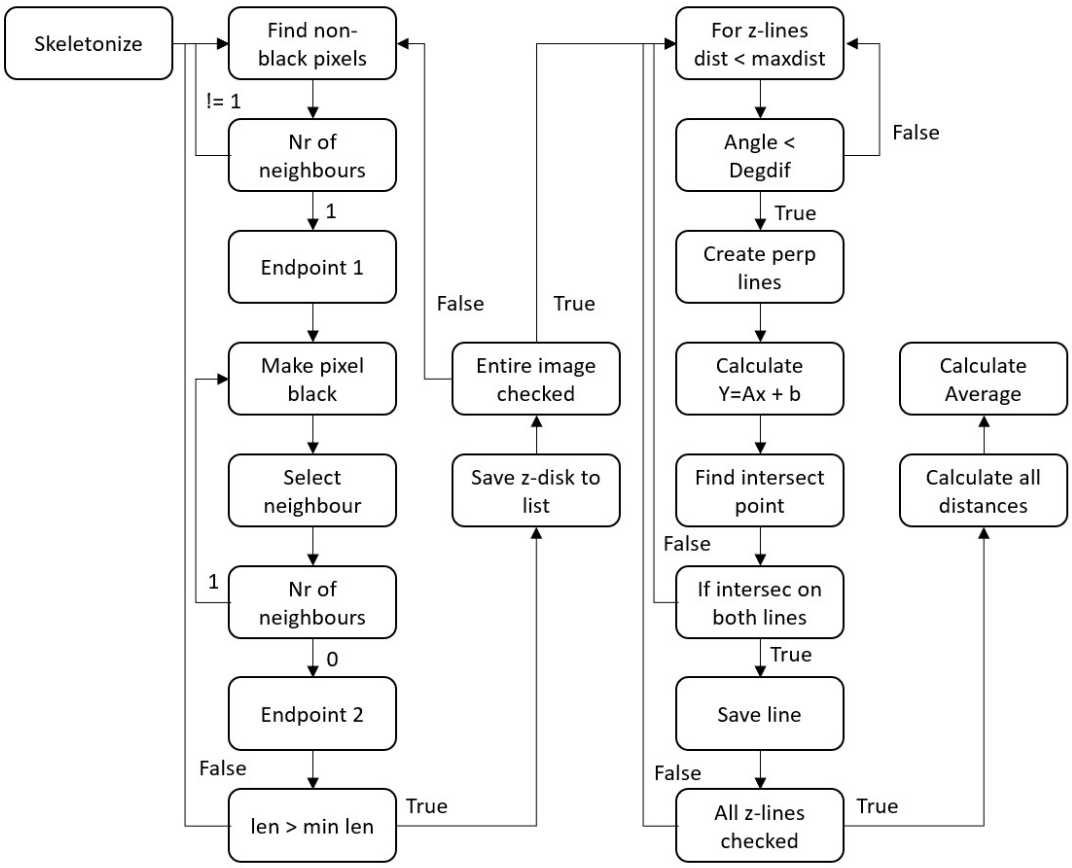

Figure 6: Flowchart BM process

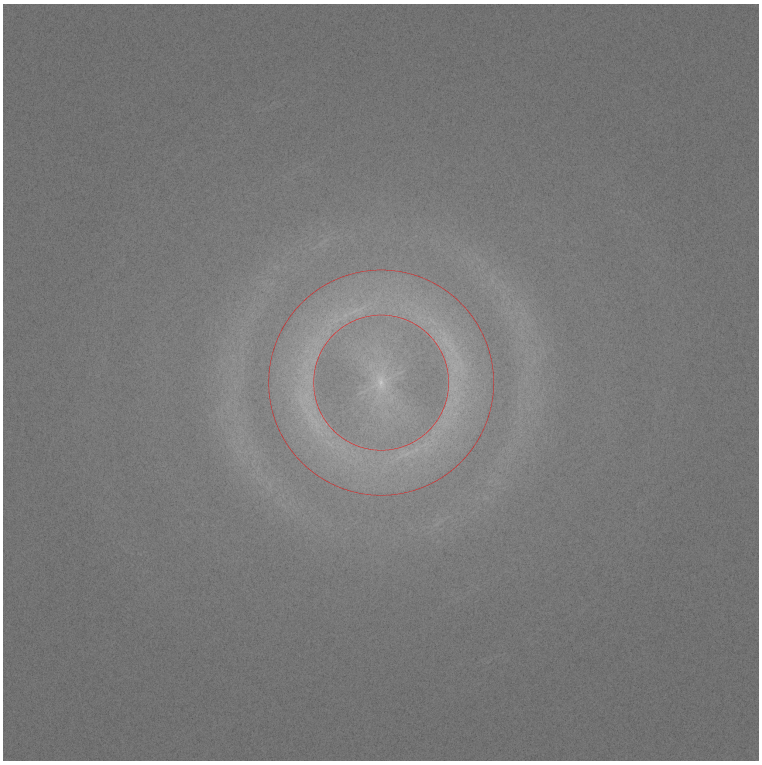

Figure 7: 2D-FFT conversion of an original image

this list all distances are calculated, which are finally averaged and divided by the pixel to μm ratio.

4.1.2 FFT

The FFT method, which is commonly used for the calculation of sarcomere lengths, both in fully grown cardiomyocytes [PK13] and developing HPSC-CMs [PS15], uses the 2D-FFT conversion of the original image. This yields an image alike to that in figure 7. Here the red lines were added later, to show the boundaries of 180 to 300 pixels that were also talked about in section 3.4.2. The program only reports on the data between these two lines. The brightness of a pixel indicates a peak in frequency. With a lighter tint implying a higher frequency. This can be used to create a distribution, with its peak being equal to the average sarcomere length. This is calculated in the program by firstly recording the total brightness as well as the number of pixels at every distance, using a dictionary to keep track of the distances. After this the average brightness per pixel gets calculated, resulting in a distribution alike to that seen in figure 10. The main focus is calculating the peak, and as such the next part of the program isolates the peak. This leftover data is then trimmed by subtracting the lowest value from all brightness values, and adding four to all. This way a proper normal distribution is created as can be seen in figure 9. Using the Fitter module [FIT], the matching normal distribution is calculated. The peak of this is equal to the average length, after converting it back into μm.

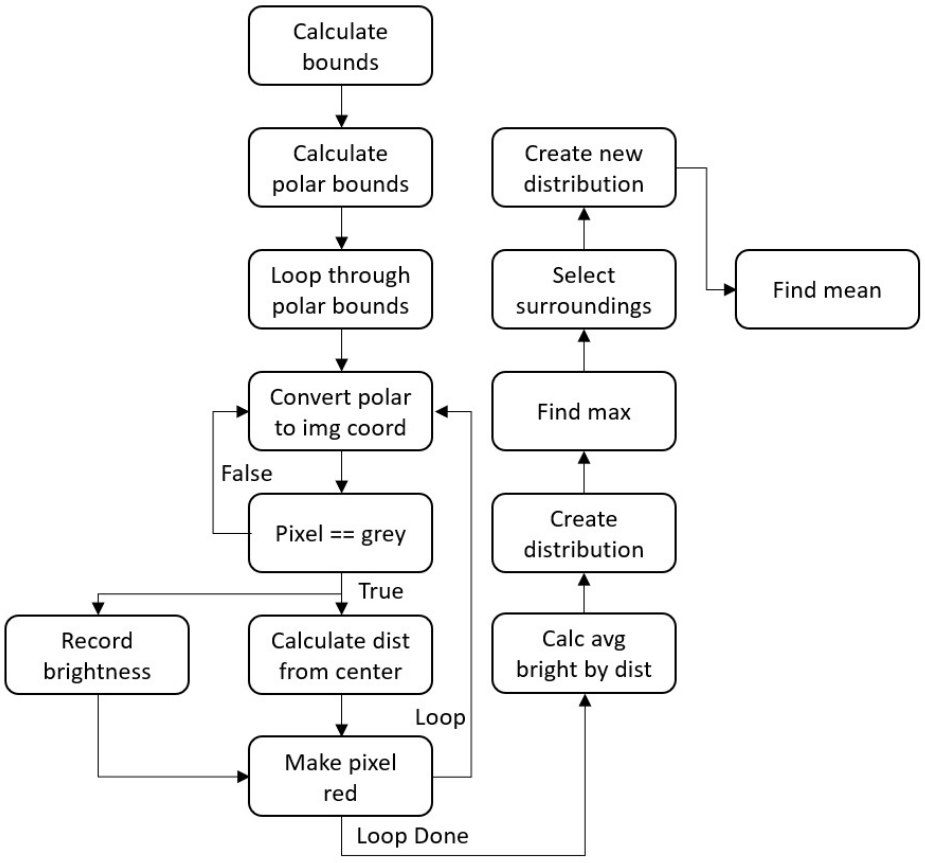

Figure 8: Flowchart for the FFT process

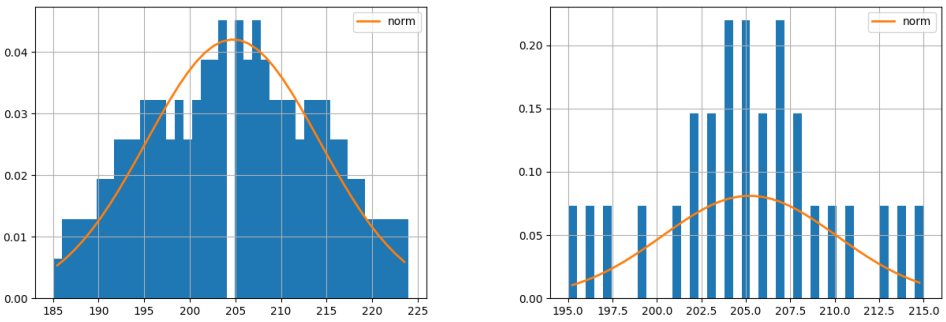

Figure 9: Distributions fitting the FFT data with 4 added on the left and without this on the right

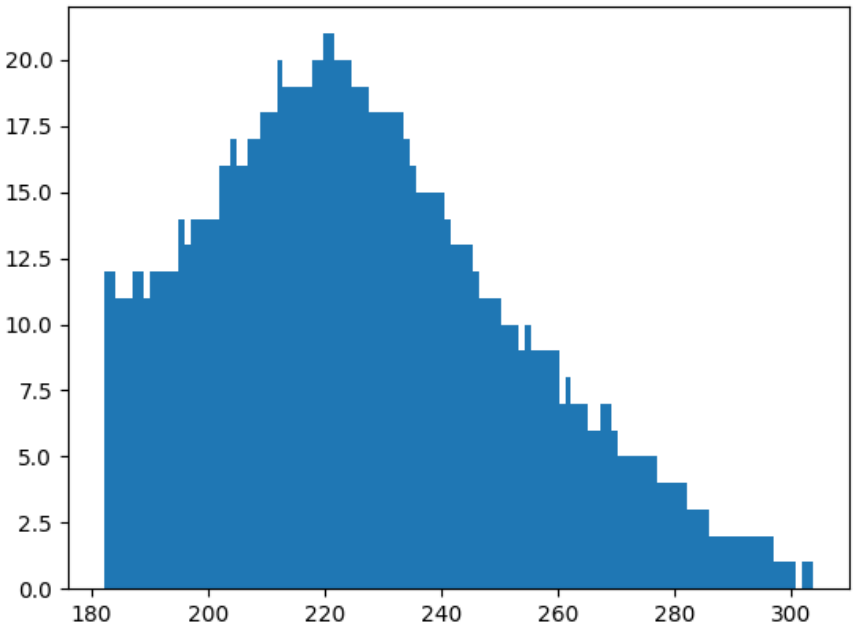

Figure 10: Distribution in FFT image

4.2 Experiments

The software for both methods was developed using two representative images from the data set which was used in later experiments. Parameters were first of all optimized using these images. For the BM method this meant a focus on including as many z lines as possible while keeping the average length within expected parameters. This also included visual inspection of the images generated, in order to confirm that all generated lines were correct. An example of the images checked can be seen in figure 11. For the FFT method it was only important to generate a proper distribution, as well as make sure all pixels in between the cirkles were checked. After determining the optimal parameters, the software was run on all data. This data was gathered previously by [CS23] as explained in section 3.4. The data consists of a directory with images sorted over 6 days, from day 0 to day 5. For each day the original red channel image was available, as well as the BM. For each day the BM images were situated in a separate folder. Using a macro, the red channel images were converted by imagej 2D-FFT to an FFT image, which were saved akin to the BM images in a separate folder. Another macro was used to convert the BM images to skeletonized images, using the imagej skeletonization function, resulting in images alike to Figure 5. These images are also saved to a separate folder. The resulting directory structure consisted of a folder per day, each with three folders with the three seperate input image sets. A text file with all the directories was then used in combination with a list parameter of the different folder names to enable running BM and FFT in one process. Because running the imagej skeletonized images required editing two parameters, this was not able to be run at the same time. Results were then written to csv files, ready for processing.

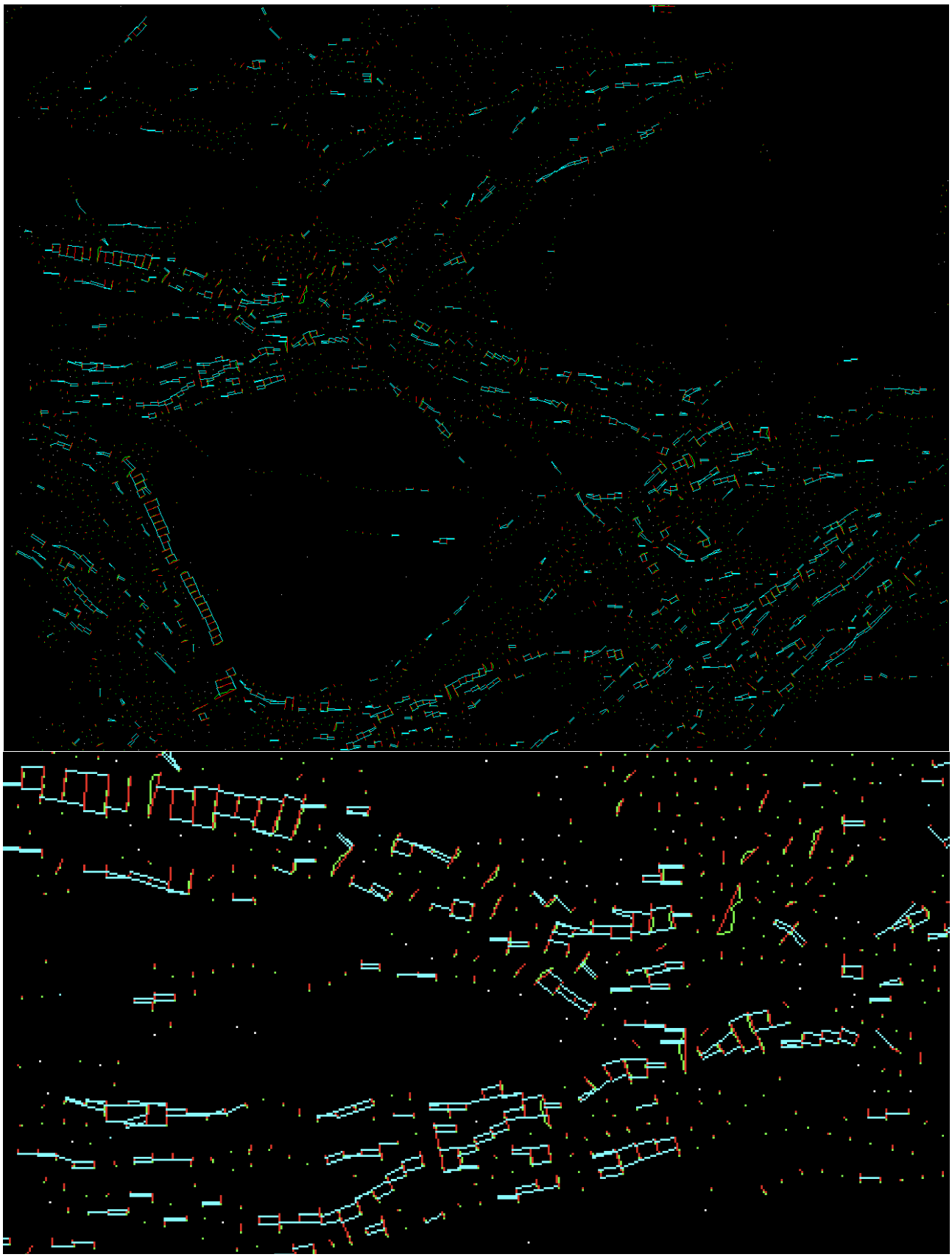

Figure 11: Line skeleton image generated, with red lines being the z-lines, green the skeleton, and light blue lines the perpendicular intersection lines, second image is a zoomed in part of the first

4.3 Result Processing

Results were saved in csv files, as they can be easily read into a pandas dataframe. The averages are calculated by date and condition. The data is then plotted using pyplot. The difference between the three methods is also calculated by matching days and calculating the delta, which was averaged. The results of which were also plotted. The calculation of average runtime was done similarly. Z-line counts were recorded during processing and again matched to their corresponding average in order to analyse performance on different levels of information.

4.4 Results

4.4.1 Parameter Finding BM

The first order of business is looking at the best parameters. The main values taken into account for this were z-line length, angle between z-line, and the max distance two lines could be apart and still be considered. For the z-line length it was initially thought to be important in order to also have an accurate orientation. As such this was checked between 4 and 2 pixels. For angle 10, 15 and 20 degrees were considered. Here the main concern was making sure unlikely matches did not occur. This was also a factor in the max distance parameter. Here it was especially about making sure that in a series of 3 sarcomeres, sarcomere 1 and 3 were not connected. As such the values considered for this were taken on the low side, resulting in 10, 13 and 15 being checked. Starting values for these were quite restrictive, with a max angle of 10 degrees, a max distance of 10 pixels, and a minimum z-line length of 4 pixels. This not only proved to yield a low number of z-lines, with only 40% matched. But as previously mentioned, with a pixel to μm ratio of 5.6, it might exclude measurement of existing z-lines. A test with 15 pixel max distance was then run, but images showed that this might find connections between z-lines which were not meant to be matched, in some cases even finding lines with another line in between. The max distance was then finally settled at 13 pixels, which resulted in a ratio of about 60% on the pictures used for development. Minimum length experiments showed that it was not disadvantageous to lower this parameter as far as possible, and as such it was pegged at 0.5. Finally, for the max angle it was decided on 20 degrees. As 10 degrees still proved to prevent a number of connections, while 20 degrees proved to not yield any unacceptable connections.

4.4.2 FFT Tuning

For the fourier transfer calculations, two of the main parameters were already calculated in the paper by [CS23]. These are the upper and lower bounds for the sarcomere length, of 1.2 and 2.0 μm. Another thing which had to be looked at was the distributions. As mention in section 4.1.2, the peak is isolated, in order to calculate said peak. This is because, as seen in figure 10, the distribution of all data is alike to a folded normal distribution. This distributions characteristics make it much harder to calculate its peak than that of a normal distribution. As such, when a visual inspection of the graph showed that the peak portion was akin to a normal distribution, whose peak is more easily calculated, the decision was made to isolate the peak and use a normal distribution to calculate the sarcomere length. It must be noted that adding four proved to yield a proper looking distribution as seen in Figure 9. As such it was decided to not further optimize this value.

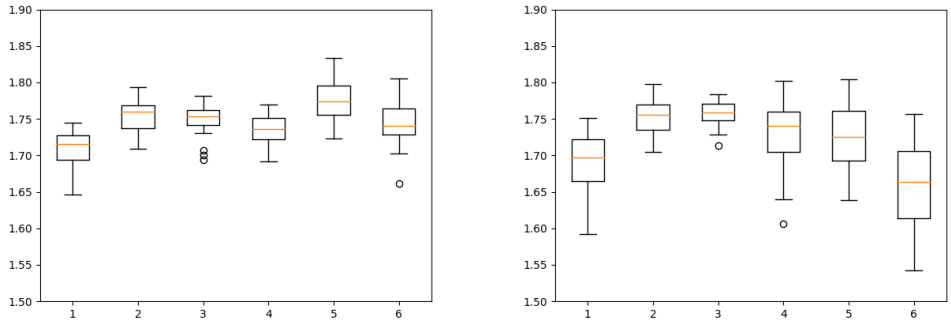

Figure 12: Graphs showing binary masks in control and test state with days on the x-axis and length in μm on the y-axis

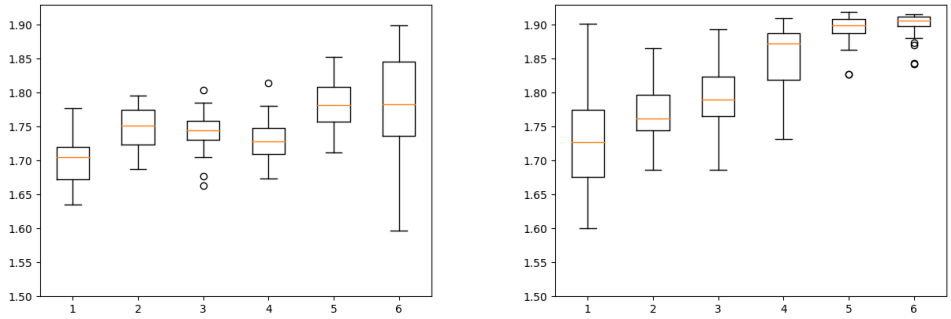

Figure 13: Graphs showing FFT in control and test state with days on the x-axis and length in μm on the y-axis

4.4.3 Average length

In figures 12 13 14 the calculated average sarcomere length is shown per method as well as per day and condition. Boxplots are used in order to show the deviation between different pictures. This shows that there exists deviations in sarcomere length between different cells which were similarly treated. Figure 15 shows the differences in measured length between the different methods. Each value represented in the boxplot is the difference on the same image. Here it can be seen that between the FFT and BM/Skel method there exist a sizeable difference of 0.3 μm in a number of pictures. The average difference is about 0.07 μm which equates to roughly 4%. For the Skel method on average 3353 z-lines are found, while the BM method with sci-img skeletonization only found 2230 z-lines on average.

Type	Average Runtime (s)	Total Runtime (s)
BM	11.8	4709
FFT	1.78	709
Skel	9.1	3647

Table 1: Runtimes of all methods

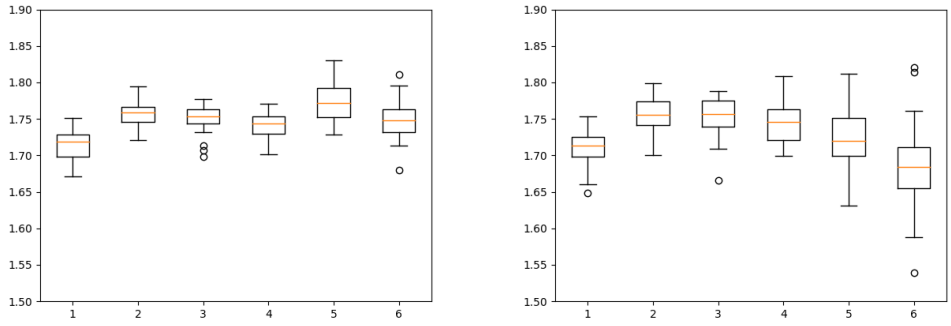

Figure 14: Graphs showing skeletonized binary masks in control and test state with days on the x-axis and length in μm on the y-axis

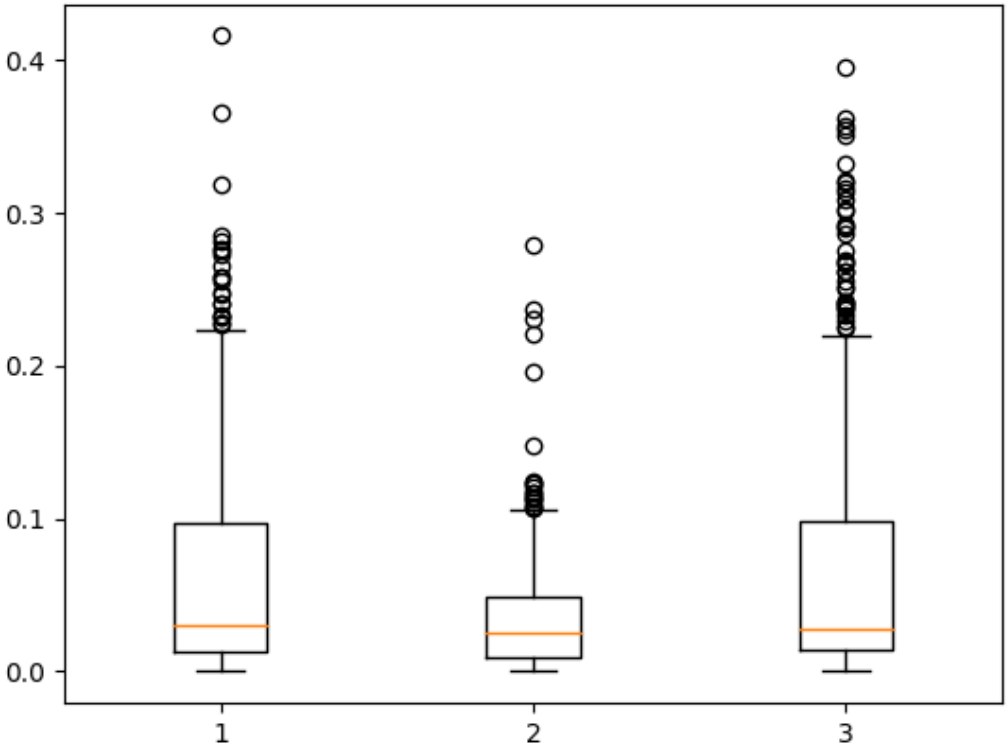

Figure 15: Difference in μm between the various method 1. FFT vs Skel 2. BM vs Skel 3. FFT vs BM

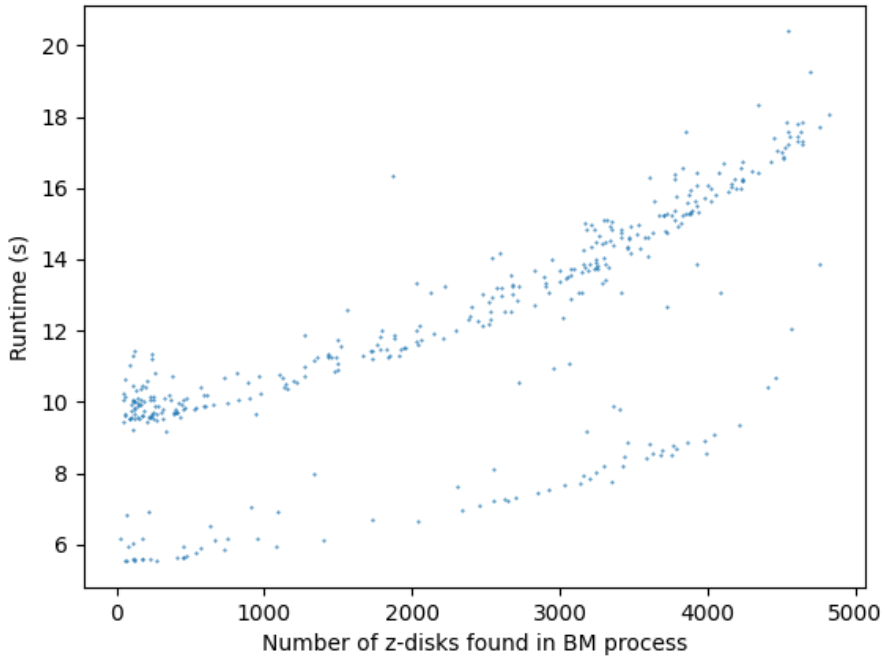

Figure 16: Runtime of BM against number of Z disks

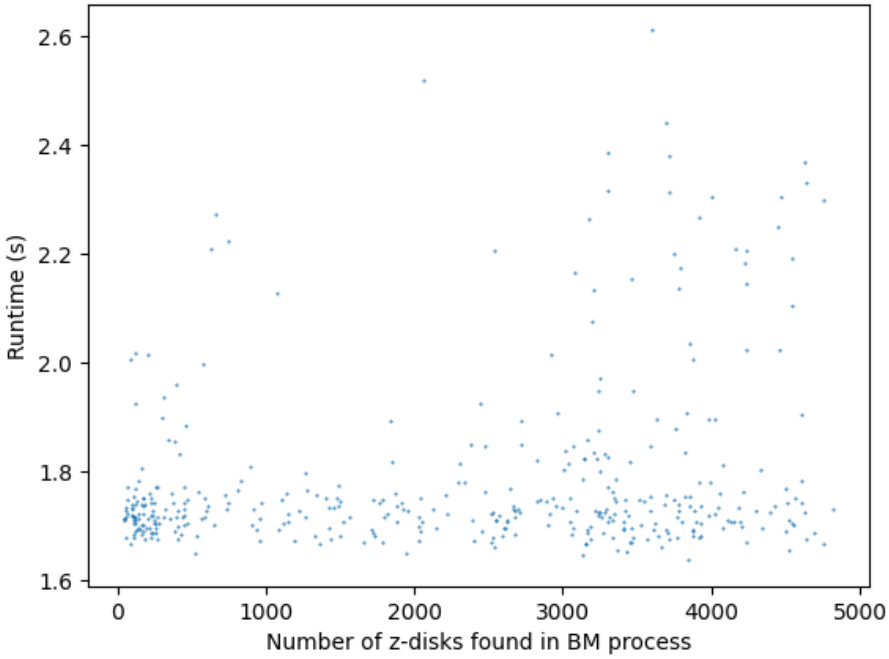

Figure 17: Runtime of FFT against number of Z disks

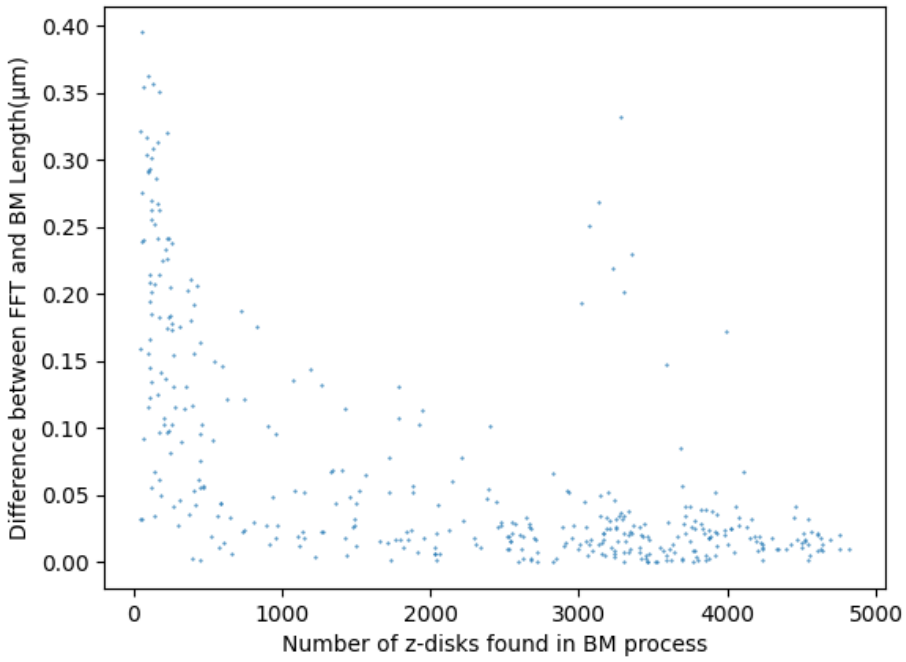

Figure 18: Difference between FFT and BM methods corresponding to Number of Z-disks

4.4.4 Runtime

The average runtimes of the methods can be seen in table 1. With about 11.8 seconds for a single image the runtime for BM is generally considered too high for high throughput purposes, as it crosses the 10 second boundary. The complexity for the BM calculation is roughly $p + 1/2(n + 1)$ with p the number of pixels in the image and n the number of z-lines. This is evident in Figure 16, however, the calculation time is lowered by around 4 seconds in some circumstances. It is unkown what leads to this, and can happen when running a single image multiple times. All together it leads to a runtime of about 4,700 seconds for 498 images. The Skel method is noticeably faster, with a total runtime of about 3647 seconds for the same number of images. The FFT method is 85% faster, with a total runtime of 709 seconds for 498 images. On top of this, as can be seen in Figure 17, the runtime of the FFT method is not dependent on the number of z-disks. This method also requires less processing with imagej, thus shortening the total pipeline time even further.

4.4.5 N dependent Accuracy

When processing the FFT data it was noticed that there was a significantly smaller spread in the later days compared to the initial days. This fell together with the calculated value approaching the upper boundary for sarcomere lenght. Additionally it was noted that this trend was not in any way visible in the BM method. The imaging gave an initial indication of where to look. This is because the FFT transformed images showed no clear bands as can be seen in Figure 7. Instead it showed just the center peak, and a gradient falling away from the center. Because this was due to a lack of z-disks it was possible to combine the z-disk data from the BM calculation with the

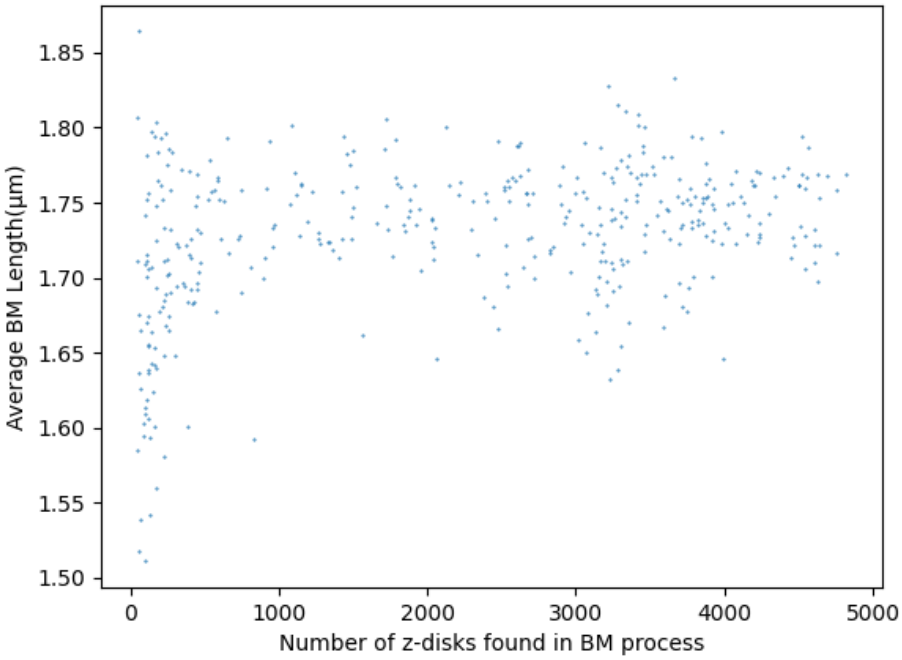

Figure 19: Average BM calculated sarcomere length compared to Number of Z-disks

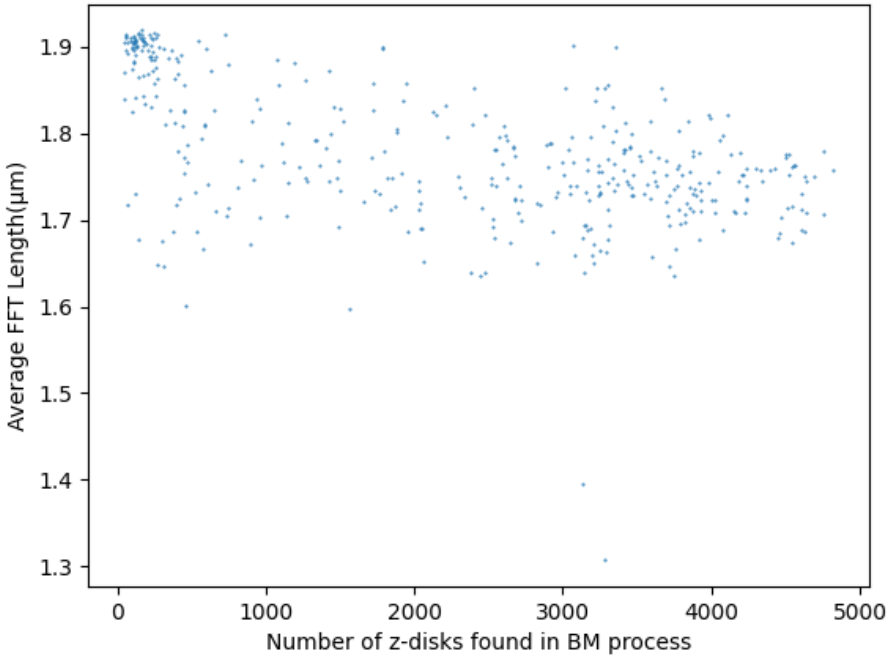

Figure 20: Average FFT calculated sarcomere length compared to Number of Z-disks

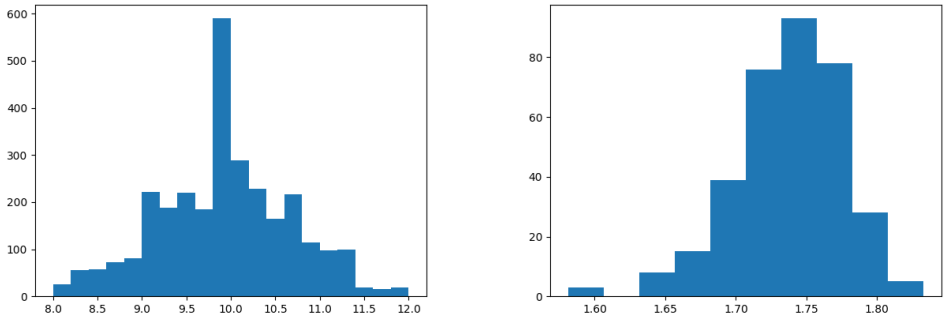

Figure 21: Distributions of sarcomere lengths in a single cell and overall

average length from the FFT calculation. This resulted in Figure 20, which shows a clear clumping around 1.9 μm and 0-200 z-disks. Meanwhile Figure 19 does not show this clumping. Instead it shows a very wide spread. This can be due to either inaccuracy, or the actual data, see Section 5. Another notable result of this, is that, as can be seen in Figure 18, the accuracy at higher z-disks is actually way lower than the 0.07 μm mentioned in Section 4.4.3. When calculated using only values of images with over 2000 z-disks the average difference is 0.028 μm. Though it must be said that there are still a small number of high difference results in the higher z-disk ranges, this is only a 1.6% difference between the two methods.

4.4.6 Contractions

Finally one more aspect must be considered when calculating sarcomere lengths. Because it functions as a contractile unit, it is possible that when measuring, the cells sarcomeres are either contracted or extended. This might skew the calculations of averages if not considered properly. In order to analyse this, a look was taken at two distributions, at the single cell level and overall. The data for single day proved inconclusive, with most graphs showing a single peak, while a number of others did show two peaks 21. The data for all days together showed only a single peak 21, which is also barely wide enough in order to facilitate the 0.2 μm contraction distance found by [RT15] and [MB19].

5 Conclusions and Further Research

The results show a number of interesting things. Most notably they show that FFT conversion is a favorite for high throughput image analysis because of its great advantage in runtime. The FFT program has to process a smaller number of pixels, where the BM program at the moment checks all pixels in the image, which makes up about 80% of the runtime. Additionally, as the FFT image used for this method is used to create the image for BM analysis, the creation of the FFT image is faster. As such the FFT calculation saves time at multiple aspects, making it especially viable for high throughput image analysis. While running the experiments it immediately became evident how much a few seconds of runtime per image can save, with the BM's total runtime of over 4700 seconds. However, some problems also arose with the FFT calculation. The bias at low

z-disk counts is not only a big issue for accurate calculation, it is also potentially very difficult to address, as further talked about in Section 5.3.

5.1 BM

The BM method faces issues of a different caliber. With it's runtime already being around 10 seconds at low z-disk counts, and increasing with their increase, the BM method at its current implementation is simply too slow. However it is currently not using any of the tricks python offers in order to speed up it's processes. With multi threading and better use of the NumPy module it can most likely get a fair bit faster. A more positive aspect to look at is the accuracy. It has shown that with an average difference of 4% and when only the values with over 2000 z-disks taken into account and average difference of only 1.6%, it can be stated that the BM method has a high degree of accuracy. The BM method is also more likely to be accurate at low z-disk ranges. This is sadly difficult to validate, but a logical conclusion due to the fact that the BM method calculates actually visible distances, as such it almost always has data for its calculations. Only one image in the dataset yielded no distance. Here it found 21 z-disk of which none were within the matching parameters. The biggest issue it thus faces with low data images is that any outliers are more impactful on the final average. As such it is unclear if the wide spread in distances at low z-disk counts is an actual representation of the data or an indication of inaccuracy in the software.

5.2 Skel

The Skel version, which makes use of the imagej skeletonization instead of that of sci-img, still leads to some difference between it and BM, it is however a smaller difference than that between FFT and Skel/BM, with only about 2% difference. This can be expected as the BM and Skel methods only differ in the skeletonization. However, the Skel method does find more z-lines. This is advantageous as it gives the software more data to work with, generally leading to a more accurate average. The method shows the same trend in avg length progression over the days as the BM method. The Skel method is preferable over BM because of this higher utilization, as well as it being easily implemented into the pipeline, with the BM processing already done in imagej, an extra step is easily added.

5.3 FFT

The FFT method has already been proven to work on fully grown cardiomyocytes, and can thus be used as a baseline for comparison. However, it does show some cracks in its effectiveness. The low data calculation bias proves once again that there is no one single software solution for all situations. The issue arises because of a combination of factors. Firstly the issue of low data means there is no distinct peak in the 2D-FFT conversion image. This results in the software not having a peak to center on. Because of the gradient in the picture is slowly decreasing from the center, the peak is taken at the boundary and after the calculation, which causes a slight shift, the value ends up around 1.9 μm. The FFT method can as such not be considered accurate at low data intensity. It becomes accurate at around 2000 z-disks. Here the peaks become prominent and can for sure return an accurate measurement. On higher z-disk counts FFT once again proves it proficiency at returning accurate values at high speed. The runtime of the FFT method is an enormous advantage

when processing large amounts of data and as such would stay a clear favourite if not for the bias at low values. A likely fix would be to decide which process to use on the basis of z-disk counts. However at the moment the bottleneck in BM speed is the finding of the z-disks.

5.4 Trends

The data shows a number of notable trends, which do not all match up with what is expected. Firstly, while it was expected to see linear growth in sarcomere length, none of the figures show this, except for the FFT B03 in figure 13. As discussed previously however, this is due to a calculation error and not due to the linear growth being present. Instead the trends visible show a decrease of sarcomere lengths at later days compared to previous days as well as between the control and treated cells. The effect is minimal and has a high spread, so it would require further research to ensure this trend is valid. It is however a good indication of where to look. Another trend visible in figure 20 shows that the accuracy of the FFT method depends on the number of sarcomeres present. Looking at data for days 5 and 6, with low sarcomere count in a number of images, it can be concluded that a higher difference exists between the FFT and BM/Skel methods. Because the FFT method is less accurate in this area, using BM for processing these low z-line count images would be advisable. For other images, where a band is clearly visible in 2D-FFT converted images, FFT is the advisable method in order to save on runtime, with a potentially as high or even higher accuracy. It must also be noted that the time gained on these high z-line images is higher, due to it requiring more processing power.

5.5 Research Conclusions

The collected data provides a few insights into answering the research question; How can sarcomere lengths in HPSC-CMs be calculated at low data intensity and what effect does Doxorubicin have on sarcomere length? Firstly the FFT method validated the BM method at high z-disk counts, showing it is a valid method for image analysis, however it requires a significant speed increase to make it viable for high throughput image analysis. This research has also shown the FFT method to be unreliable at low z-disk counts, and presents the BM method as a valid alternative. It is also shown that there is an indication for lower sarcomere lenghts and especially a length decrease when HPSC-CMs are treated with Doxorubicin.

5.6 Further Research

This software both leaves some gaps which need to be researched as well as opens up ample opportunities for research. Firstly, the main gap it leaves is verification with another source. For the data used for both development as well as verification and running the experiments, an accurate length statistic was not available. As such any calculations were only verified with the other calculation. It is debatable whether the program can be trusted 100% when such a form of verification has not been done. However when this is done a number of opportunities for research open up, the main area being research into the growth and development of cardiomyocytes. The experiments done in this paper already raise a big question when it comes to this because of the lack of linear growth. This is especially notable because it happens in both the control group as well as in the experimental group. Lastly, development can go into further automation of the pipeline.

It is possible to analyze the images before processing in order to decide both whether it would be viable, as well as which method might be best. Automation of this could enable both a quicker pipeline and require less input, which might make it possible for large datasets to be run overnight. Lastly a useful experiment would be using a different form of microscopy in order to not have both contracting and extended sarcomeres in the same dataset.

References

[AM04] M. D. Abramoff and P. J. Magalhaes. Image processing with imagej. *Biophotonics International*, 11:36–42, 2004.

[CS23] L. Cao and L. Schoenmaker. Automated image analysis system for studying cardiotoxicity in human pluripotent stem cell-derived cardiomyocytes. *Microscopy and Microanalysis*, 29:254–264, 2023.

[CvdM20] L. Cao and A. D. van der Meer. Cardiotoxicity and heart failure: Lessons from human-induced pluripotent stem cell-derived cardiomyocytes and anti-cancer drugs. *BMC Bioinformatics*, 21, 2020.

[DP18] T. P. Dias and S. N. Pinto. Biophysical study of human induced pluripotent stem cell-derived cardiomyocyte structural maturation during long-term culture. *Biochemical and Biophysical Research Communications*, 499:611–617, 2018.

[EMA] Clinical efficacy and safety guidelines. Date Accessed: 10/8/2023.

[FIT] Fitter documentation. Date Accessed: 9/8/2023.

[HP00] R. P. Hertzberg and A. J. Pope. High-throughput screening: new technology for the 21st century. *Current Opinion in Chemical Biology*, 4:445–451, 2000.

[MB19] T. A. Myachina and K. A. Butova. Development and program implementation of an algorithm to estimate the mean sarcomere length of a cardiomyocyte. *Biophysics*, 64:732–737, 2019.

[NCB] Actn2. Date Accessed: 7/8/2023.

[PE09] M. S. Parmacek and J. A. Epstein. Cardiomyocyte renewal. *New England Journal of Medicine*, 361:86–88, 2009.

[PG16] C. Pasqualin and F. Gannier. Sarcoptim for imagej: high-frequency online sarcomere length computing on stimulated cardiomyocytes. *American Journal of Physiology: Cell Physiology*, 311:C277–C283, 2016.

[PK13] P. Peterson and M. Kalda. Real-time determination of sarcomere length of a single cardiomyocyte during contraction. *American Journal of Physiology: Cell Physiology*, 304:C519–C531, 2013.

[PN00] V. B. Pai and M. C. Nahata. Cardiotoxicity of chemotherapeutic agents. *Drug Safety*, 22:263–302, 2000.

[PS15] F. S. Pasqualini and S. P. Sheehy. Structural phenotyping of stem cell-derived cardiomyocytes. *Stem Cell Reports*, 4:340–347, 2015.

[RS20] M. C. Ribeiro and R. H. Slaats. A cardiomyocyte show of force: A fluorescent alpha-actinin reporter line sheds light on human cardiomyocyte contractility versus substrate stiffness. *Journal of Molecular and Cellular Cardiology*, 141:54–64, 2020.

[RT15] M. C. Ribeiro and L. G. Tertoolen. Functional maturation of human pluripotent stem cell derived cardiomyocytes in vitro e correlation between contraction force and electrophysiology. *Biomaterials*, 51:138–150, 2015.

[sci] Scikit-image documentation. Date Accessed: 14/8/2023.

[SW10] J. W. Sanger and J. Wang. Assembly and dynamics of myofibrils. *Biomed Research International*, 2010:8, 2010.