

# The implementation of a novel assay along with the development of an image analysis tool to quantify localization and gating function of CFTR

Catherine Hastings

25th August 2017

## **Abstract**

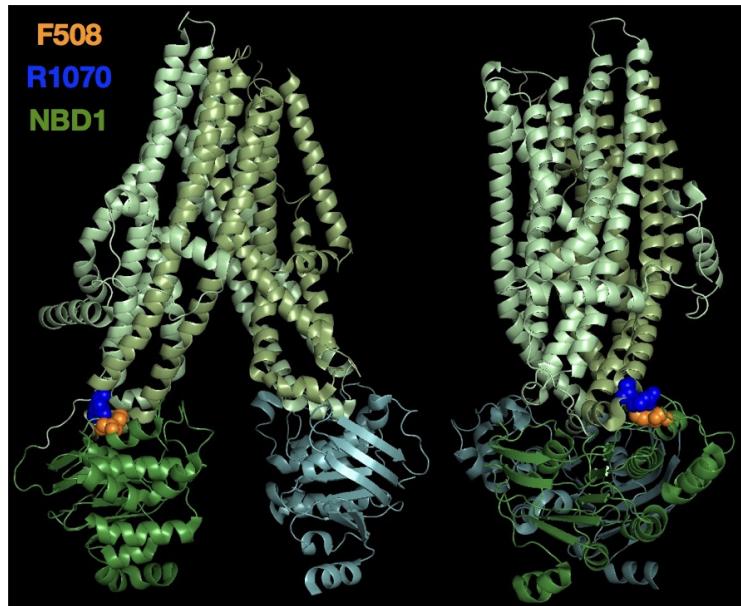
Defects in the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) protein result in its loss of function as an ion channel, causing cystic fibrosis (CF). The F508del mutation is the most common CF-causing mutation and leads to reduced intra-cell trafficking, increased degradation in the endoplasmic reticulum, loss of stability at the plasma membrane and decreased gating function. Transfection of cells *in vitro* with CFTR bound to a yellow fluorescent protein (YFP) alongside a red fluorescent protein, mCherry, allows for location of CFTR in reference to the cell membrane using wide-field imaging. With image analysis tools, a quantification of the CFTR density at the membrane can be calculated. In addition, the sensitivity of YFP (with H148Q and I152L mutations) to iodide ions can be used to quantify the gating function of CFTR at the membrane. With the introduction of this single assay, high-content data is collected for wild type (WT) CFTR, CFTR-F508del and CFTR-F508del/R1070W. Results show that the introduction of the R1070W mutation into CFTR-F508del rescues gating function but does not significantly increase the density of CFTR at the membrane.

# 1 Introduction

## 1.1 Cystic fibrosis, CFTR and the F508del mutation

Cystic fibrosis (CF) is an autosomal recessive disorder affecting multiple organ system. Secretions throughout the body are abnormally viscous, causing blockages, inflammation and tissue damage (Cutting, 2015) and resulting in decreased life expectancy. CF is caused by mutations in the Cystic Fibrosis Transmembrane conductance Regulator (*CFTR*) gene (Riordan et al., 1989). There are currently 281 known CF-causing mutations (CFTR2, 2017), the most common of which results in the deletion of phenylalanine from location 508 (F508del). The CFTR-F508del mutations accounts for approximately two-thirds of cases world-wide (Bobadilla et al., 2002).

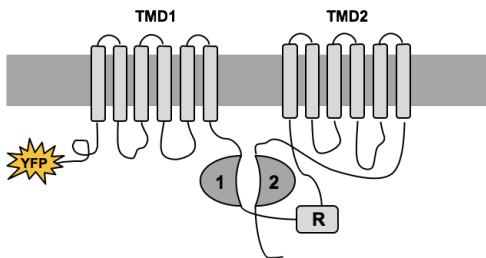
The CFTR protein acts as a plasma membrane ion channel. The loss of function of the CFTR ion channel observed when the F508del mutation is present, is thought to be due to a combination of factors (Lukacs and Verkman, 2012). After CFTR-F508del is translated in the endoplasmic reticulum (ER) it is misfolded (Qu and Thomas, 1996). A reduced quantity of protein is trafficked to the Golgi meaning very little of the protein is complex-glycosylated and ready to be trafficked to the membrane (Cheng et al., 1990). As a result, a large proportion of CFTR-F508del undergoes ER associated degradation. It has also been shown that CFTR-F508del has decreased stability on the membrane (Lukacs et al., 1993) and that it suffers from poor gating function (Dalemans et al. (1991) and Langron et al. (2017)). These features combined result in severely decreased ion transport in and out of the cell when CFTR-F508del is expressed.



**Figure 1.1:** A representation of the CFTR protein. The locations of F508 and R1070 are highlighted in orange and blue respectively. NDB1 is one of the nucleotide binding domains and is shown in green. Protein structure from Liu et al. (2017).

## 1.2 Second-site rescue mutations of CFTR

The CFTR ion channel has two transmembrane domains (TMD1 and TMD2), two nucleotide binding domains (NBD1 and NBD2) which bind ATP and a regulatory domain (R) which requires phosphorylation for the channel to open (Lieberman et al., 2009). F508 is located on NBD1 very close to an intracellular loop (specifically ICL4) of TMD2 (see figures 1.1 and 1.2). It has been shown by Thibodeau et al. (2010) that the introduction of mutation R1070W on ICL4 partially rescues the loss of function caused by the F508del mutation. It is suggested by Thibodeau et al. (2010) that this restoration of function is due to a larger quantity of CFTR reaching the membrane.



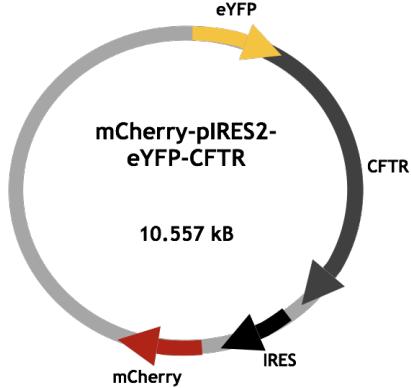
**Figure 1.2:** YFP-CFTR on the membrane. TMD1 and TMD2 are the transmembrane domains, R is the regulatory domain, and the 2 nucleotide-binding domains (NBD) are numbered.

## 1.3 Techniques for studying CFTR

Here introduced is an assay which, alongside novel image analysis techniques, aims to quantify both the density of CFTR protein at the membrane and its gating function. This work further builds on that of Langron et al. (2017) who implement a similar technique studying the gating function of CFTR. Three genotypes of *CFTR* are compared: wild type (WT) CFTR, CFTR-F508del and CFTR-F508del/R1070W, with the aim of discovering more about the mechanism by which the second-site mutation R1070W rescues the function of CFTR-F508del. The unification of the image analysis techniques into a single piece of software also allows large amounts of data to be gathered in a single experiment.

The assay involves transfecting a human cell line *in vitro* with a plasmid (shown in figure 1.3) encoding for the CFTR gene. A yellow fluorescent protein (YFP) is tagged to the N-terminal of CFTR (see figure 1.2), allowing the location of the protein to be pinpointed within the cell. In addition, the plasmid encodes for a red fluorescent protein, mCherry. mCherry is soluble in the cytosol and consequently provides an approximate outline for each cell. In tandem, the two fluorescent proteins allow for the location of the membrane and then the study of the position of YFP-CFTR in reference to that membrane.

It is worth noting that YFP-CFTR and mCherry are separated on the plasmid by pIRES2. pIRES2 is an example of an internal ribosome entry site (IRES), which were first discovered in the polio virus (Pelletier and Sonenberg, 1988). Normally, a strand of mRNA encodes for a single gene and is translated from the 5' cap to the end. However, an IRES provides a second entry point for translation. This means that two proteins can be expressed from a single mRNA plasmid. It is also implies that the ratio of expression for



**Figure 1.3:** A representation of the plasmid used in these experiments, which encodes for YFP-CFTR, pIRES2 and mCherry. The DNA on the plasmid is 10,557 bases long.

the two proteins remains constant (see Langron et al. (2017), supplementary material and appendix B).

The tagging of CFTR with YFP allows for its location to be studied, but additional mutations have been introduced into YFP, imbuing it with properties making it suitable for the study of the gating function of CFTR. As first described by Galietta et al. (2001), the two mutations H148Q and I152L have been introduced into YFP (though it shall be referred to only as YFP throughout). The addition of the H148Q mutation increases the sensitivity of YFP to halides, meaning that its fluorescence is quenched in the presence of chloride and iodide ions (Jayaraman et al., 2000). With both mutations H148Q and I152L, fluorescence is quenched in the presence of iodide ions, but quenching no longer occurs at physiological levels of chloride ions (Galietta et al., 2001). As the level of iodide within cells is naturally very low, the introduction of extracellular iodide and the observation of consequent fluorescence quenching can be used to determine the in-flux of iodide via the CFTR ion channel.

## 2 Method

### 2.1 Experimental procedure

The experimental procedure described here is derived from the method of Langron et al. (2017). Unless otherwise stated, chemicals were purchased from Sigma Aldrich. HEK293 cells were maintained in DMEM with additions of 2 mM L-glutamine,  $100 \text{ U} \cdot \text{mL}^{-1}$  penicillin and streptomycin, and 10% FBS (all Life Technologies). The cells were split every 2 to 3 days. When this was done, 5 mL HBSS medium was used to wash the cells and 1.5 mL of the enzyme trypsin-EDTA was added to detach cells from the previous container. 8 mL of DMEM medium was added, and cells were then separated out using a centrifuge spinning at 1000rpm for 2 minutes. This process was also used when cells were transferred into a 96-well blacked-sided plate used for fluorescence imaging.

Before the cells were transfected with DNA, the concentrations of DNA were verified using gel electrophoresis. The plasmids were linearized and then pushed through 0.8% agarose gel by an electric field. Longer strands of DNA are heavier and move less quickly through

the gel. Using a reference sample and knowing the lengths of the DNA means that the concentration of the DNA can be estimated. This is important for ensuring different conditions are transfected with the same DNA concentration.

Each 96-well plate was coated with poly-D-lysine and, once cells were added, incubated overnight at 37°C. Lipofectamine 2000 (Life Technologies) was used to transiently transfect the cells with the appropriate YFP-CFTR plasmid. Lipofectamine aids the transport of the DNA into the cells by forming a liposome around the plasmid. This negates the effect of the opposing charges of DNA and the cell membrane (Dalby et al., 2004). After transfection, the cells were incubated at 37°C for a further 24 hours before imaging.

## 2.2 Image collection

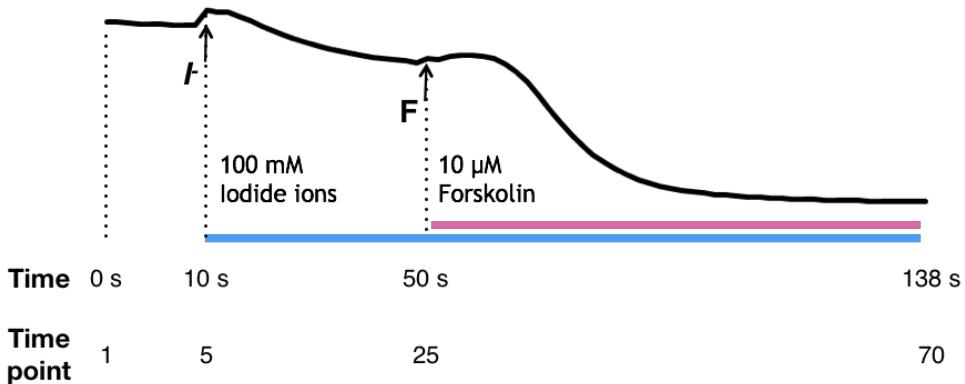
Cells were washed twice before imaging with 100 µL standard buffer (140 mM NaCl, 4.7 mM KCl, 1.2 mM MgCl<sub>2</sub>, 5 mM HEPES, 2.5 mM CaCl<sub>2</sub>, 1 mM glucose, pH 7.4). All images were acquired with ImageExpress (ImageXpress Micro XLS, Molecular Devices), an inverted wide-field fluorescence microscope with a CMOS camera and an automated fluidics system. During imaging, the 96-well plated was maintained at 37°C. Two wavelengths of light were used, one for obtaining images of YFP-CFTR and one for mCherry. For YFP, excitation/emission filters of 472 ± 30 nm and 520 ± 35 nm were employed and for mCherry the excitation/emission filters were 562 ± 40 nm and 624 ± 40 nm.

### Localization of YFP-CFTR

For images intended to be used for the localization of YFP-CFTR, a 60x objective was used with no binning of pixels. Images were collected from 9 adjacent sites in each well. Laser intensity and exposure times were optimized for each experiment so as to maximize the fluorescence without damaging the cells (typically an intensity of 100cd and an exposure of 700ms for YFP, and an intensity of 150cd and an exposure of 150ms for mCherry).

### YFP-CFTR gating function

For images intended to be used for the analysis of CFTR gating function, a 20x objective was used along with 5x5 binning of pixels (i.e. 25 pixels were reduced to a single pixel) in order to decrease the size of data collected and consequently image processing time. For each well, images were collected from a single site at 70 time points at an interval of 2 seconds. Before the fifth timepoint (i.e. after 10 seconds), 50 µL of iodide in buffer solution were added to create an extracellular iodide concentration, [I<sup>-</sup>], of 100 mM. The iodide buffer composition is as the standard buffer but with chloride ions replaced as iodide. After a further 40 seconds, before time point 25, 50 µL of iodised standard buffer containing forskolin was added, creating an extracellular forskolin concentration of 10 µM whilst the extracellular iodide concentration was maintained at 100 µM. See figure 2.1 for a depiction of the experiment time course. Forskolin is an activation compound for CFTR and is dissolved in a DMSO vehicle.



**Figure 2.1:** A diagram representing the time course of the experiment. An image was collected at every time point. The presence of iodide in extracellular solution is represented in blue and the presence of forskolin is represented in pink. The black line shows an example variation in normalized YFP that might be observed from YFP-CFTR-WT.

## 2.3 Data processing

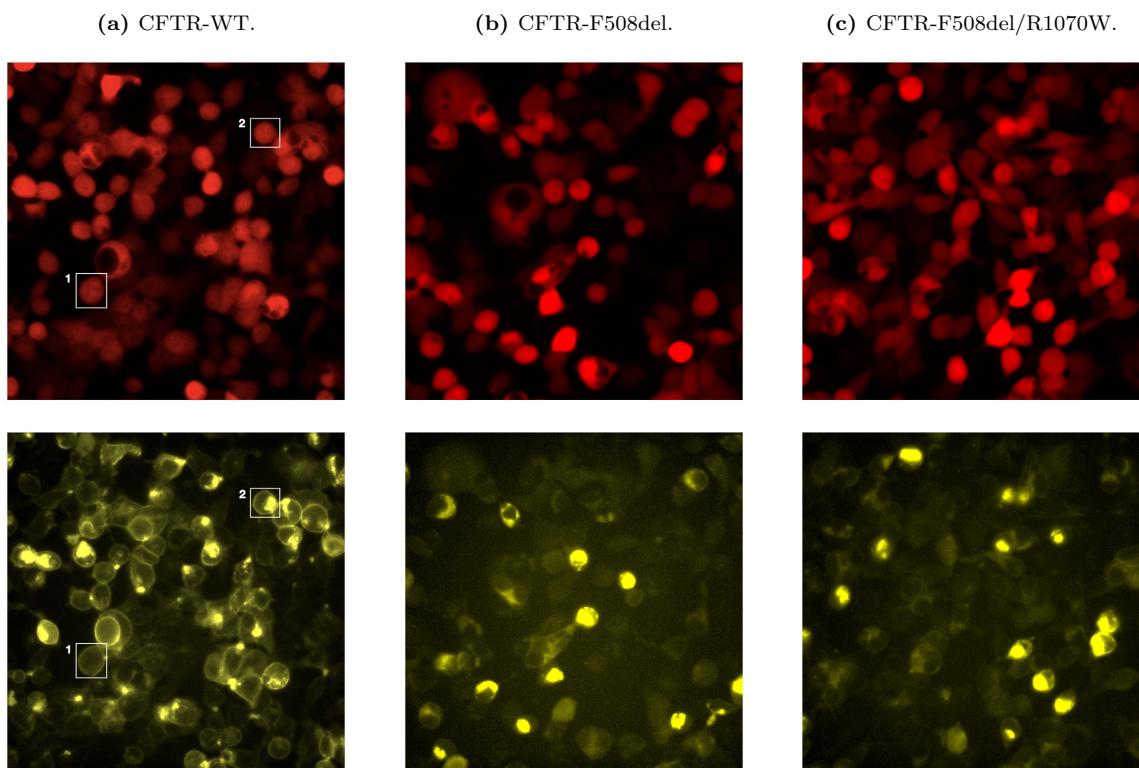
### Localization of YFP-CFTR

Raw images are collected as 2160x2160 pixel images, with grayscale values between 0 and 65535. The images in figure 2.2 have been coloured appropriately to represent the red mCherry fluorescence and the yellow YFP fluorescence.

Images are segmented to locate individual cells, because in order to be able to confidently state conclusions about the position of CFTR in relation to the membrane, the cell membrane must first be accurately located. Image segmentation is carried out on the mCherry images, and it should be noted that there is no significant difference in mCherry images between conditions (see appendix A).

The software implements marker-controlled watershed segmentation, following a procedure described in the MATLAB documentation (Mathworks, 2017). Watershed segmentation identifies local minima in intensity data and defines these as catchment basins. These basins are then extended until neighbouring basins meet, at which point a segmentation line is drawn between the two regions containing the local minima. This is analogous to a watershed line between two rain catchments on a mountainside. The benefit of this method is that neighbouring cells can be separated despite having similar intensity values.

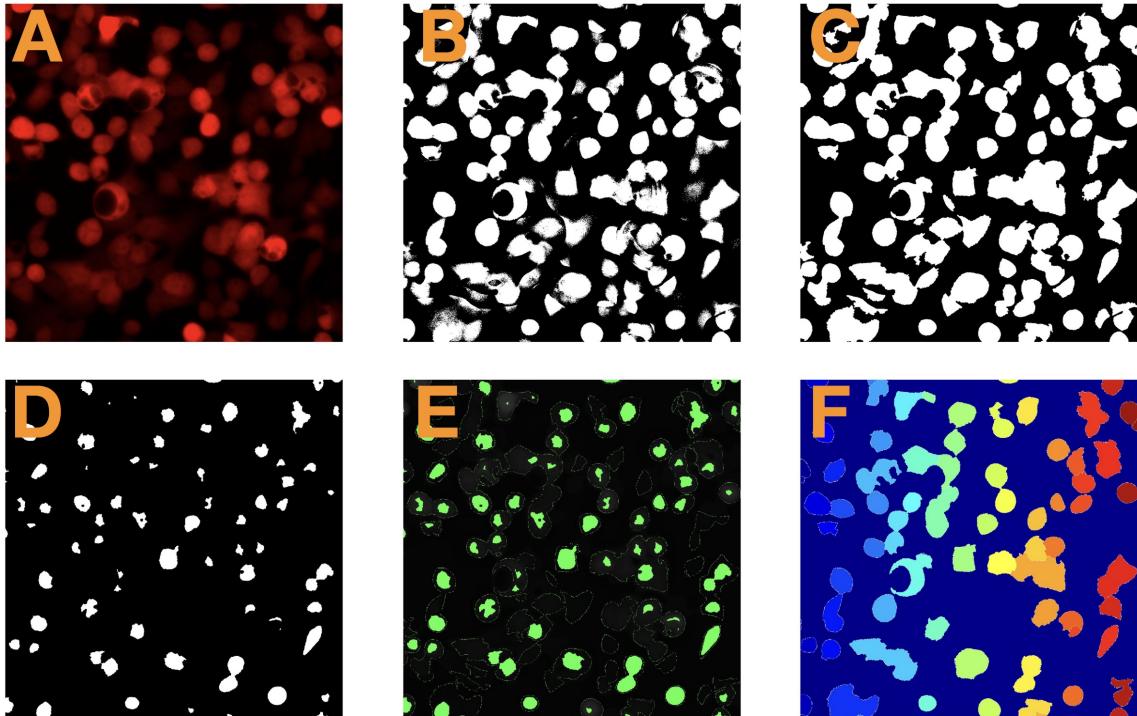
The first stage of the segmentation process is to increase the contrast of an image, before it is then binarized using MATLAB's in-built `graythresh` function to find an appropriate threshold. The binarization locates the majority of the area filled with cells (see panel (B) of figure 2.3). The rest of the area is found by detecting local maxima, and morphologically dilating these areas. Both the threshold-located region and the local maxima are smoothed and filled with a series of morphological erosions and dilations. The union of these two areas forms the cell background (panel (C), figure 2.3). Small cell markers are found by



**Figure 2.2:** Example raw images. Each column represents one of the three genotypes of CFTR tested. The top row shows images displaying mCherry fluorescence, and the bottom row shows images displaying YFP fluorescence. The cells marked in the CFTR-WT images will later be used to demonstrate image processing methods.

identifying local maxima in the image intensity (figure 2.3, panel (D)).

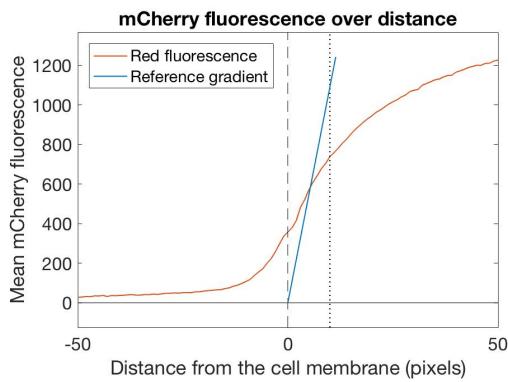
Images are inverted, converting local maxima to minima, and watershed segmentation is performed using the MATLAB `watershed` function. The small cell markers provide starting points for the watershed segmentation and the cell background imposes an artificial limit to the expansion of catchment basins. Panel (E) of figure 2.3 shows the small cell markers superimposed over the outline of the cell background. The final result of the segmentation process is shown in figure 2.3, panel (F).



**Figure 2.3:** Demonstration of the stages of the automated image segmentation into cell components. **A**, an unprocessed mCherry fluorescence image coloured red. **B**, the result of binarizing an image after increasing contrast. **C**, the elements of image B after smoothing and being added to by local maxima. **D**, small cell markers providing a starting point for watershed segmentation, located as local maxima. **E**, the small cells markers overlaid on the outline of the cell background from panel C. **F**, the final result of image segmentation.

After a full image has been segmented into individual cells, these cells are then identified by the bounding box enclosing the cell area. Each cell bounding box undergoes processing to identify a more precise boundary of the cell. Creating a component boundary which accurately mirrors that of the cell is important due to the narrow thickness of the membrane in relation to the cell size and the subtle differences in CFTR density between the membrane and the cell interior. The bounding box area is binarized using a newly identified threshold, ensuring that a similar proportion of the cell area is located for all cells despite different levels of mCherry fluorescence (caused by varying expression). The boundary of the cell component is smoothed and filled using a series of morphological erosions and dilations (see panel (B) of figure 2.5).

Once a good approximation of the cell boundary has been located, checks are performed on each cell component and any cells failing these tests are discarded as being unsuitable for analysis. Analysing poorly segmented cells risks obscuring the results. Firstly, any bounding boxes abutting the edge of the full image are omitted, as this suggests that the cell component is incomplete. Secondly, cells with such low mCherry expression that a second binarizing cannot distinguish the cell from the background are discarded. In these cases the automated segmentation has not located any cell area for analysis. Thirdly, cells are selected based upon their dimensions. Limits are imposed so that cell components must be between 100 and 300 pixels in length and between 80 and 250 pixels in width. Additionally components with a high perimeter to area ratio ( $> 0.05$  pixels) are discarded. This is because components suitable for analysis are convex in shape and therefore have a low perimeter relative to their area.

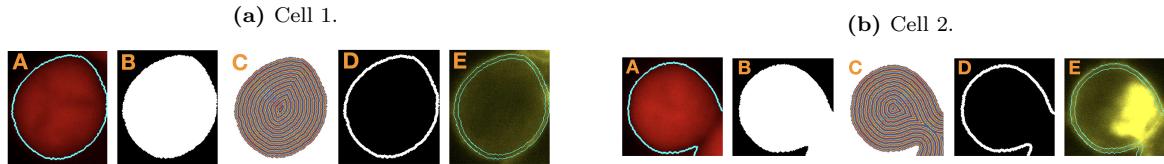


**Figure 2.4:** Plot demonstrating the reference gradient in relation to the change in mCherry fluorescence over distance from the cell boundary for cell 1 in figures 2.2 and 2.5. The maximum increase in mCherry fluorescence occurs at 4 pixels within the cell boundary (dashed line). The dotted line shows the inner border of the cell component membrane.

well segmented. Any cell components failing this test are discarded from the subsequent analysis.

In order to quantify the amount of YFP-CFTR at the membrane, a membrane region must first be defined in terms of the segmented cells components. Here we take the membrane of the cell to mean the area 10 pixels inside of the cell component border (relative to its approximate 200 pixel diameter). This definition of the cell membrane is illustrated in panel (D) of figure 2.5.

Before analysis is performed upon intensity values, the background fluorescence is calculated, with a single background fluorescence value being taken for an full 2160x2160 pixel image. The image is binarized with a threshold half way between the minimum value



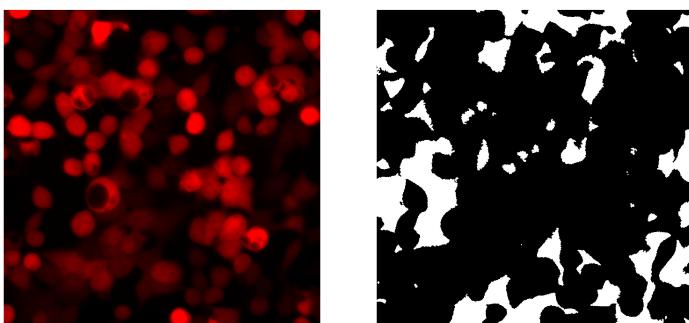
**Figure 2.5:** Demonstration of the process of creating a minimum distance map from each pixel inside the cell to the boundary. The two cells are those labelled in figure 2.2. **A**, an mCherry image cropped to the size of the bounding box of a single cell component. The blue line marks the edge of the cell component shown in panel **B**. **B**, binarized smoothed cell component. **C**, distance map showing rings of pixels at an equal distance from the cell boundary. **D**, an illustration of the area defined to be the cell component membrane. **E**, a cropped YFP image with the edges of the membrane area marked in blue.

and the lower quartile of the fluorescence intensity data for the entire image. This area is smoothed and filled, and the background fluorescence is taken to be the average intensity value over this area. The background intensity value is subtracted from all fluorescence values before analysis for both YFP and mCherry images.

### YFP-CFTR gating function

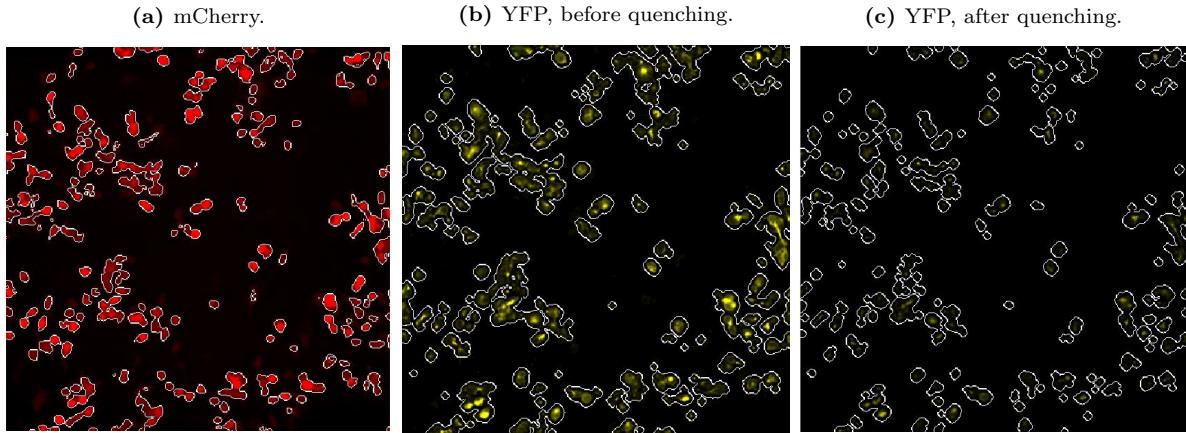
Raw images are collected as 432x432 pixel images, with grayscale values between 0 and 65535. All images in figure 2.7 have been coloured appropriately to demonstrate either mCherry or YFP fluorescence. YFP fluorescence images are collected at every 2 seconds throughout the experiment, which consists of 70 timepoints in total as described in section 2.2 and demonstrated in figure 2.1. Images showing mCherry fluorescence are only collected at the first and last time point. Fluid additions are made just before time points 5 and 25.

The first stage of image processing is to select cells. The mCherry image collected at the start of the experiment (before any fluid additions) is binarized using MATLAB's in-built `imbinarize` function. As described above, the mCherry is used because it shows least variation between genotypes (appendix A). The cell mask is then dilated by 4 pixels in every direction and overlaid on each YFP image, as shown in the second panel of 2.7. The mean YFP signal is then calculated to be the average YFP fluorescence intensity within the cell area minus the mean YFP intensity outside the cell area. This calculation is performed on the YFP image from every time point, and when plotted against time gives



**Figure 2.6:** Displaying an mCherry image and the corresponding area (shown in white) used to calculate background fluorescence.

a graph as shown in figure 3.2. Calculating the YFP signal in this way reduces variation in results due to background fluorescence.



**Figure 2.7:** Demonstration of the image processing on images collected for the study of the gating function of CFTR. **(a)**, the result of cell selection by binarization of an mCherry image. The edges of the cell selection are shown in white, overlaid on the mCherry image. **(b)**, the dilated cell selection on the corresponding YFP image before quenching. **(c)**, the dilated cell selection on the YFP image after quenching for YFP-CFTR-WT.

## 2.4 Data analysis

Before image analysis is performed it should be noted that fluorescence images vary hugely in intensity by both additive and multiplicative factors, particularly between experiments performed on different days.

Differences in fluorescence intensities can be caused by a range of factors, and care is taken throughout the experimental and image gathering processes to minimize these. Firstly, it is possible that the 96-well plates are coated with slightly different thicknesses of poly-D-lysine. When the cells are transferred into wells, the cell density is fixed but this can vary between experiments. Equally, if left to divide for slightly different lengths of time, more or less cells will be present per well in different experiments. Differences between the success of the transfection occur between cells, between wells and between experiments. Finally, when gathering images, attempts are made to keep the fluorescence settings as constant as possible, but there are minor differences between exposure times and laser intensities.

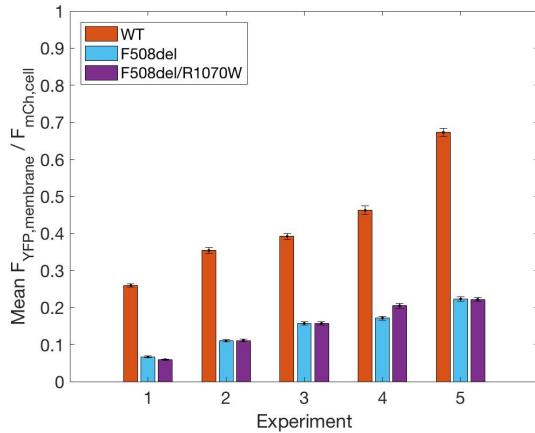
### Localization of YFP-CFTR

Cells have now been isolated and the membrane of the cell component has been defined. The aim of localising the YFP-CFTR is to compare the quantity of CFTR at the membrane of the cell for different genotypes of CFTR. A metric is required which quantifies this difference in CFTR membrane density while also minimizing the effects of experimental variation.

Let

$$\Phi = \frac{F_{\text{YFP,membrane}}}{F_{\text{mCherry,cell}}} , \quad (1)$$

with  $F_{\text{protein,region}}$  representing the mean fluorescence intensity per pixel due to the given protein within the given region. Due to the construction of the plasmid (and more particularly the presence of pIRE2), it can be assumed that the ratio between mCherry expression and YFP-CFTR expression is constant across cells and across conditions. The metric,  $\Phi$ , can be calculated for each cell component. Dividing YFP membrane fluorescence by the average mCherry fluorescence corrects for the different levels of YFP-CFTR expression between cells.



**Figure 2.8:** A bar chart showing the mean  $\Phi$  value across all cells, before normalizing. The results are shown for each *CFTR* genotype, and for each experiment. The error bars represent the standard error from the mean (SEM).

Figure 2.8 does show that, for every experiment, the mean values of  $\Phi$  for CFTR-F508del and CFTR-F508del/R1070W are approximately one third the mean value of  $\Phi$  for CFTR-WT. Therefore, to adjust for differences between experiments, a normalization constant is calculated with the WT condition acting as a control. For each experiment, a value is calculated which multiplicatively adjusts the mean value of  $\Phi$  for CFTR-WT to 1. Then, for every cell across all conditions, the mean YFP fluorescence value within the membrane is multiplied by this normalization constant. Normalizing the mean YFP membrane fluorescence, results in the value for  $\Phi$  being similarly normalized.

### YFP-CFTR gating function

The analysis methods used here are those described by Langron et al. (2017). Once the YFP signal has been calculated for each time point, the YFP fluorescence values are normalized against the YFP signal value just before the first addition of iodide. It is important to normalize our data here as it compensates for variability in fluorescence

intensities over different experiments. Normalized YFP fluorescence can then be plotted against time as shown in figure 3.2.

In order to calculate a single value to quantify the effect of mutations upon the gating function of the CFTR ion channel, further analysis must be performed. It can be seen in figure 2.7 that after fluid additions are made, the YFP fluorescence quenches. It can be assumed that when CFTR is functioning correctly, high extracellular concentrations of iodide ions,  $[I^-]_{out}$ , result in the accumulation of iodide ions within the cell, increasing  $[I^-]_{in}$ . Iodide ions within the cell bind to the chromophore and quench the fluorescence, meaning that a Hill–Langmuir-type equation can be applied, which states

$$\frac{F}{F_{max}} = 1 - \frac{[I^-]_{in}}{K_1 + [I^-]_{in}} . \quad (2)$$

Here  $F/F_{max}$  is the normalized YFP fluorescence,  $[I^-]_{in}$  is the concentration of iodide ions within the cell mM and  $K_1$  is the binding affinity of iodide ions to YFP. A value of 1.9 mM is used (Galietta et al., 2001) which represents the binding affinity for this specific variant of YFP.

Rearranging equation (2) for the concentration of iodide gives

$$[I^-]_{in} = K_1 \frac{1 - F/F_{max}}{F/F_{max}} .$$

Plotting the graph of  $[I^-]_{in}$  against time allows for the estimation of the maximum value of

$$\frac{d[I^-]_{in}}{dt} ,$$

the maximum rate of flow of iodine ions into the cell. Seeing how this value varies between conditions gives a simple quantification of the gating function of YFP-CFTR.

## 2.5 Statistical analysis

The threshold for statistical significance used is 0.05 throughout, unless corrected due to performing multiple comparisons. Where this is the case the type of correction is stated. The type of error bars shown are either the standard error from the mean or the standard deviation, and are labelled in individual plots. Field (2009) was used a reference for all statistical tests.

For studying the localization of CFTR, values of the normalized metric  $\Phi$  are not normally distributed (see appendix C). As a result, and because it is necessary to compare more than two conditions, the Kruskal–Wallis test is used to test for significance. The Kruskal–Wallis test ranks data values and compares mean rank values. Post-hoc tests are performed upon these mean ranks implementing a Dunn–Šidák significance correction for multiple tests.

For studying the gating function of YFP-CFTR, values for mean maximum rate of iodide in-flow were tested with the Kolmogorov–Smirnov test for normality, and it was found that the null hypothesis that the samples are normally distributed could be rejected (see

appendix D). As a result, non-parametric tests were used to test for significant differences between conditions. A Kruskal–Wallis test was used to test significance differences between genotypes for cases where forskolin was present. In addition Mann–Whitney  $U$ -tests were performed to test for significant differences between cases where forskolin was used and where it was not, within a single genotypic condition.

## 3 Results

### 3.1 Localization of YFP-CFTR

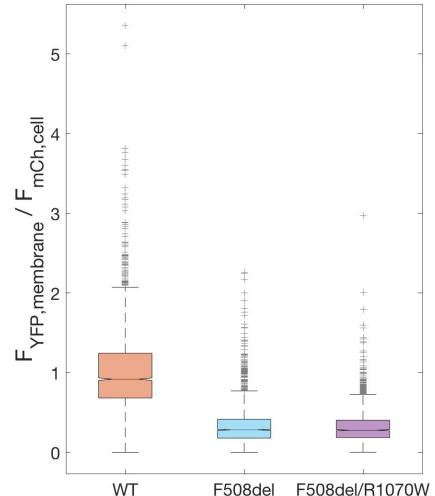
270 images were analysed per condition, collected over five experiments. After processing, a total of 8503 cells were located with 3202, 2795 and 2506 cells collected for conditions CFTR-WT, F508del and F508del/R1070W respectively.

Calculating the value for our normalized metric ( $\Phi = F_{\text{YFP,mem}} / F_{\text{mCherry,cell}}$ ) gives an estimate for the relative number of CFTR ion channels at the membrane between conditions. A value is calculated for each cell as displayed in figure 3.1. These values show significant skew (as can be seen in appendix C), and therefore the non-parametric Kruskal–Wallis test is used to test for significant differences between the metric values for each condition.

The Kruskal–Wallis  $H$ -test gives a test statistic of  $H = 4527.54$  with 2 degrees of freedom and a  $p$ -value much less than 0.001<sup>1</sup>. This means that there is a statistical significance between conditions, with mean ranks of 6557, 2877 and 2840 for CFTR-WT, F508del and F508del/R1070W respectively.

Performing post-hoc Dunn–Šidák tests on the mean ranks show that there is a significant difference between CFTR-WT and both CFTR-F508del and CFTR-F508del/R1070W (with significance levels less 0.001). However, we cannot reject the null hypothesis that the mean ranks of CFTR-F508del and CFTR-F508del/R1070W are the same ( $p=0.928$ ).

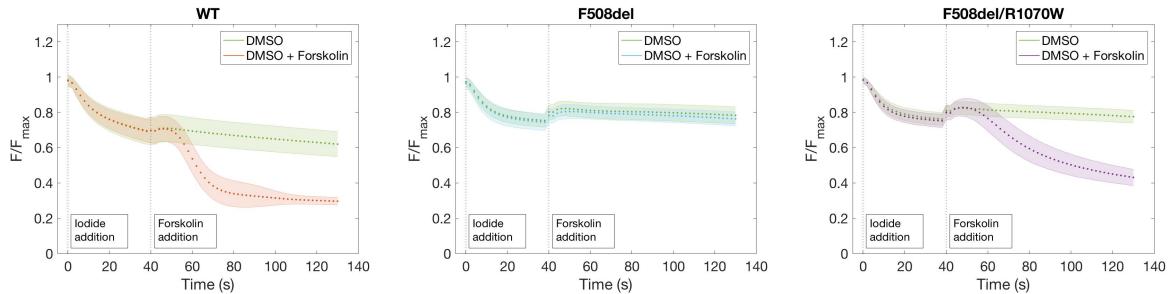
**Figure 3.1:** A box plot of the normalized  $\Phi$  value for each cell component in each condition. The maximum whisker length permitted is 1.5 times that of the interquartile range. Each point outside of that range is considered an outlier and marked with a +.



<sup>1</sup>The test statistic  $H$  approximates a  $\chi^2$  distribution, and  $P(\chi^2(2) \geq 13.8155) = 0.001$ .

### 3.2 YFP-CFTR gating function

For each well, time-series data comprising of 70 YFP fluorescence images were collected, one every two seconds. In addition, mCherry images were collected at the first and last timepoint. There were six conditions: a DMSO control and a forskolin condition for each of the three CFTR variants, WT, F508del and F508del/R1070W. For each condition, 12 time-series were collected (4 plates, with 3 replicates per plate). Mean time-series data for the normalized YFP fluorescence is show in figure 3.2.



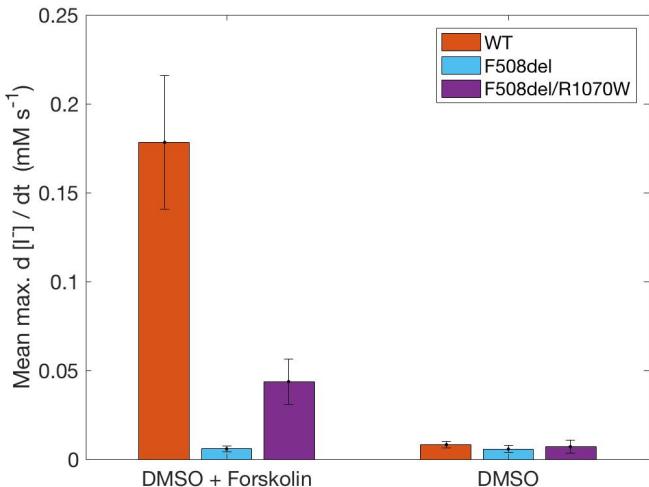
**Figure 3.2:** Plots showing how normalized YFP fluorescence changes over time, and with respect the two to fluid additions. The variant of CFTR is given in the plot title. The vertical dotted lines show the times of the fluid additions. Data was collected from 4 experiments giving a total of  $n = 12$  wells. The error here shows the standard deviation within a sample.

The maximum rate of in-flow of iodide ions was calculated for each time-series. The mean maxima of the rates are shown in figure 3.3. In testing for differences between conditions where forskolin was used, the Kruskal–Wallis statistic  $H$  was 31.14 with 2 degrees of freedom. The corresponding significance value was  $1.73 \times 10^{-7}$ . The mean ranks resulting from the Kruskal–Wallis test are 30.6, 6.5 and 18.5 for conditions CFTR-WT, F508del and F508del/R1070W respectively. Mann–Whitney  $U$ -tests comparing the control condition with the forskolin (test) condition for each genotype reveal that there are significant differences between test and control conditions for both CFTR-WT and CFTR-F508del/R1070W. The  $p$ -values were  $3.66 \times 10^{-5}$  for both genotypes. For CFTR-F508del, the Mann–Whitney  $U$ -test revealed no statistical difference between the control and forskolin cases (with  $p$ -value 0.977).

## 4 Discussion

A dual-purpose assay has been introduced, along with image processing methods, to collect and analyse high-content data quantifying the effects of mutations on the function of CFTR. This assay has here been used to compare the density at the membrane and the gating function of CFTR-WT, CFTR-F508del and CFTR-F508del/R1070W.

Figure 3.1 has shown that the presence of the mutation F508del results in a significant reduction in the density of CFTR at the membrane, confirming the results of Cheng et al.



**Figure 3.3:** Bar chart of mean maximum rate of iodide ion inflow, for each condition in both test and control cases. DMSO is used as a vehicle for the CFTR activation compound forskolin, and in the control condition only DMSO was added. Error bars show the standard deviation within the sample.

(1990). It can also be concluded that the presence of mutation R1070W in addition to F508del does not significantly rescue the reduction in intra-cell trafficking or the increase in degradation experienced by CFTR-F508del. This is a significant result, and sheds further light on the work of Thibodeau et al. (2010) and their discovery of the rescue mutation R1070W.

Figures 3.2 and 3.3 show that the presence of the mutation F508del significantly reduces the gating function of CFTR, which is in line with the findings of Dalemans et al. (1991) and Langron et al. (2017). The plots also show that the introduction of R1070W into CFTR-F508del significantly improves the function of CFTR at the membrane, though the amount of quenching observed is not entirely restored to that of CFTR-WT.

It is known that CFTR-F508del has a very low open probability ( $P_o$ ) at the membrane (Dalemans et al., 1991). It is clear that CFTR-F508del/R1070W has an increased  $P_o$ , which results in some quenching despite the low concentration of CFTR at the membrane. There are multiple factors which could result in the difference in the rate of in-flow of iodide ions observed between WT and F508del/R1070W conditions, including the respective  $P_o$  values, the density of the CFTR at the membrane and the rate of removal from the membrane. Given that figure 3.1 has shown that the density of protein at the membrane is significantly lower for CFTR-F508del/R1070W, it is possible that this is the main reason for the difference in quenching observed, and that the gating function of CFTR-F508del has been restored by the introduction of R1070W. Further study could be done to ascertain whether this is case.

## 4.1 Comments on methodology

The development of a single assay to study two parts of the function of CFTR simultaneously is extremely useful. It allows for the collection of large amounts of data in a relatively short time frame, and at lower cost. As a result the assay can be used to study any CF-causing mutations, the effect of second-site mutations and to investigate drugs

aiming to fix the function of a mutated CFTR.

The automation of image gathering by ImageXpress occasionally failed, resulting in some out-of-focus images were collected. This will not have greatly affected the results as the segmentation process only selected a single cell from more than twenty blurred images (when between 5 and 20 cells are collected for focussed images). However, it would be useful to create a function which automatically filters out-of-focus images. Equally, this explains the disparity in the number if cells analysed per condition.

The formulation of this assay allowing for the normalization of the YFP fluorescence by the corresponding mCherry fluorescence is a significant advantage. Additionally, the ability to normalize *CFTR* mutation genotypes conditions against WT provides a very simple method of reducing error in the results due to experimental variability. However, there are still differences between mean  $\Phi$  values which cannot be explained by standard error, even after normalization, as shown in figure 2.8. It is possible that variation in fluorescence images is one of the causes of this variability, and it could be investigated whether a method for calibrating fluorescence intensity could be used, for example by using fluorescent micro-spheres.

There are many ways that the quality of the segmentation could be further tested, for example a manual segmentation could be performed against which to compare the automated results. However, the additional filtering of cell components by calculating a reference gradient in mCherry fluorescence provides a reasonable quality control. It should be noted that the automated segmentation occasionally analyses abutting cells, as shown in figure 2.5b. However, this is not likely to have a large impact upon results as the membrane is still accurately located.

The Kruskal–Wallis test was used to identify significant variation in the rate of iodide ions in-flow between genotypes. This was due to the data failing the Kolmogorov–Smirnov test as shown in appendix D. However, one might expect the mean maximum in-flow to be normally distributed. It is possible that with a small sample size of  $n = 12$ , outliers have affected the distribution. However, the bar chart 3.3 displaying the mean maximum gradient of iodide ions shows such large differences between conditions that the results are still persuasive.

## Acknowledgements

I would like to thank my supervisors Paola Vergani and Lewis Griffin. I would also like to thank Emily Langron and Stella Prins for their significant help and encouragement. Their work has formed the basis of the work shown here, including construction of the plasmids. In addition, figures 1.1, 1.2, 1.3 and 2.1 were made by Stella Prins, with only minor adjustments.

## References

- Joseph L. Bobadilla, Milan Macek, Jason P. Fine, and Philip M. Farrell. Cystic fibrosis: A worldwide analysis of CFTR mutations—correlation with incidence data and application to screening. *Human Mutation*, 19(6):575–606, 2002.
- CFTR2. The Clinical and Functional TRanslation of CFTR.  
<http://cftr2.org>, 2017. [Accessed 24-August-2017].
- Seng H. Cheng, Richard J. Gregory, John Marshall, Sucharita Paul, David W. Souza, Gary A. White, Catherine R. O’Riordan, and Alan E. Smith. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*, 63(4):827 – 834, 1990.
- Garry R. Cutting. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet*, 16(1):45–56, 2015.
- Brian Dalby, Sharon Cates, Adam Harris, Elise C. Ohki, Mary L. Tilkins, Paul J. Price, and Valentina C. Ciccarone. Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods*, 33(2):95 – 103, 2004.
- Wilfried Dalemans, Pascal Barbry, Guy Champigny, Sophie Jallat, Karin Dott, Dominique Dreyer, Ronald G Crystal, Andréa Pavirani, Jean-Pierre Lecocq, and Michel Lazdunski. Altered chloride ion channel kinetics associated with the ΔF508 cystic fibrosis mutation. *Nature*, 354(6354):526–528, 1991.
- Andy Field. *Discovering statistics using SPSS*. Sage publications, 2009.
- Luis J.V Galietta, Peter M Haggie, and A.S Verkman. Green fluorescent protein-based halide indicators with improved chloride and iodide affinities. *FEBS Letters*, 499(3): 220–224, 2001.
- Sujatha Jayaraman, Peter Haggie, Rebekka M. Wachter, S. James Remington, and A. S. Verkman. Mechanism and cellular applications of a green fluorescent protein-based halide sensor. *Journal of Biological Chemistry*, 275(9):6047–6050, 2000.
- Emily Langron, Michela I Simone, Clémence Delalande, Jean-Louis Reymond, David L Selwood, and Paola Vergani. Improved fluorescence assays to measure the defects

associated with F508del-CFTR allow identification of new active compounds. *British journal of pharmacology*, 174(7):525–539, 2017.

Michael Lieberman, Allan D. Marks, and Matthew Chansky. *Marks' basic medical biochemistry: a clinical approach*. Wolters Kluwer/Lippincott Williams Wilkins Health, third edition, 2009.

Fangyu Liu, Zhe Zhang, László Csanády, David C. Gadsby, and Jue Chen. Molecular structure of the human CFTR ion channel. *Cell*, 169(1):85 – 95.e8, 2017.

G L Lukacs, X B Chang, C Bear, N Kartner, A Mohamed, J R Riordan, and S Grinstein. The delta F508 mutation decreases the stability of cystic fibrosis transmembrane conductance regulator in the plasma membrane. determination of functional half-lives on transfected cells. *Journal of Biological Chemistry*, 268(29):21592–21598, 1993.

Gergely L. Lukacs and A.S. Verkman. CFTR: folding, misfolding and correcting the  $\Delta F508$  conformational defect. *Trends in Molecular Medicine*, 18(2):81 – 91, 2012.

Mathworks. Marker-controlled watershed segmentation, MATLAB documentation.  
<http://bit.ly/2v01liI>, 2017. [Accessed 21-August-2017].

Jerry Pelletier and Nahum Sonenberg. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, 334(6180):320–325, 1988.

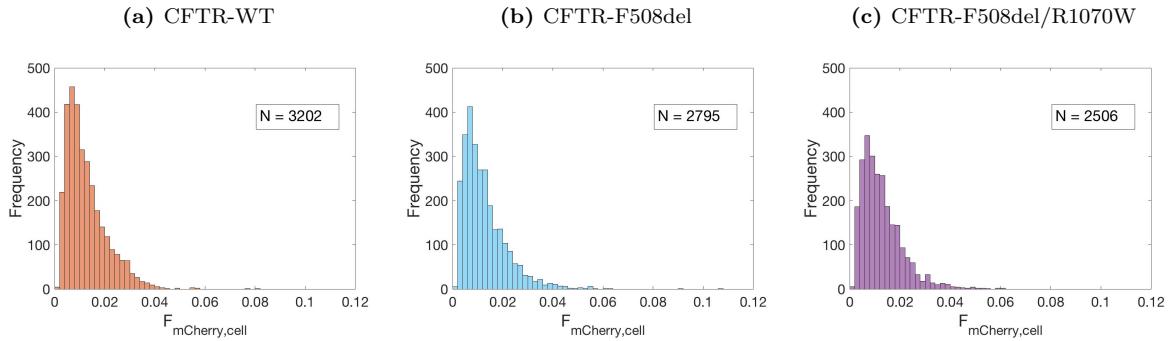
Bao-He Qu and Philip J. Thomas. Alteration of the Cystic Fibrosis Transmembrane Conductance Regulator folding pathway: Effects of the  $\Delta F508$  mutation on the thermodynamic stability and folding yield of NBD1. *Journal of Biological Chemistry*, 271 (13):7261–7264, 1996.

JR Riordan, JM Rommens, B Kerem, N Alon, R Rozmahel, Z Grzelczak, J Zielenski, S Lok, N Plavsic, JL Chou, and et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245(4922):1066–1073, 1989.

Patrick H Thibodeau, John M Richardson, Wei Wang, Linda Millen, Jarod Watson, Juan L Mendoza, Kai Du, Sharon Fischman, Hanoch Senderowitz, Gergely L Lukacs, et al. The cystic fibrosis-causing mutation  $\Delta F508$  affects multiple steps in cystic fibrosis transmembrane conductance regulator biogenesis. *Journal of Biological Chemistry*, 285 (46):35825–35835, 2010.

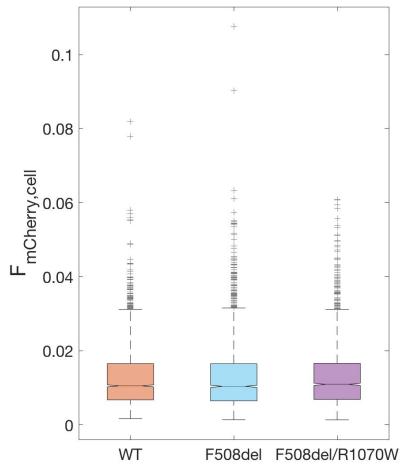
## A Testing differences in mCherry data across conditions

In equation (1), the normalized YFP fluorescence intensity in the membrane is normalized by the mCherry fluorescence inside the cell in order to correct for differences in expression levels between cells. In order for this to be fair, it is necessary to check that the mCherry values do not vary significantly between conditions. It is first shown that the data is not normally distributed (figure A.1), and then a Kruskal–Wallis test is performed (figure A.2).



**Figure A.1:** Histograms for each condition, which aim graphically to show that the values of  $F_{\text{mCherry},\text{cell}}$  are not normally distributed.

The Kruskal–Wallis test showed that there was no statistical difference between conditions. The  $H$  statistic was 3.09 with 2 degrees of freedom, resulting in a  $p$ -value of 0.2132. The mean ranks were 4305, 4266 and 4189 for conditions CFTR-WT, CFTR-F508del and CFTR-F508del/R1070W respectively.

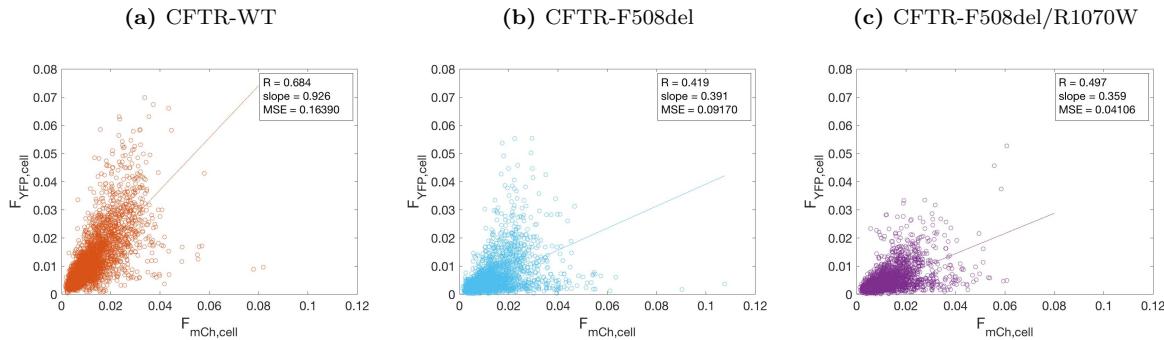


**Figure A.2:** A box plot of the mean mCherry fluorescence values inside cells for each cell component in each condition. The maximum whisker length permitted is 1.5 times that of the inter-quartile range. Each point outside of that range is considered an outlier and marked with a +.

## B Correlation between mCherry and YFP-CFTR expression

The plasmid was constructed with the insertion of pIRES2 so that two proteins could be expressed on a single mRNA. The mechanism by which an internal ribosome entry site (IRES) works, implies that there should be a correlation between the rates of expression of mCherry and YFP.

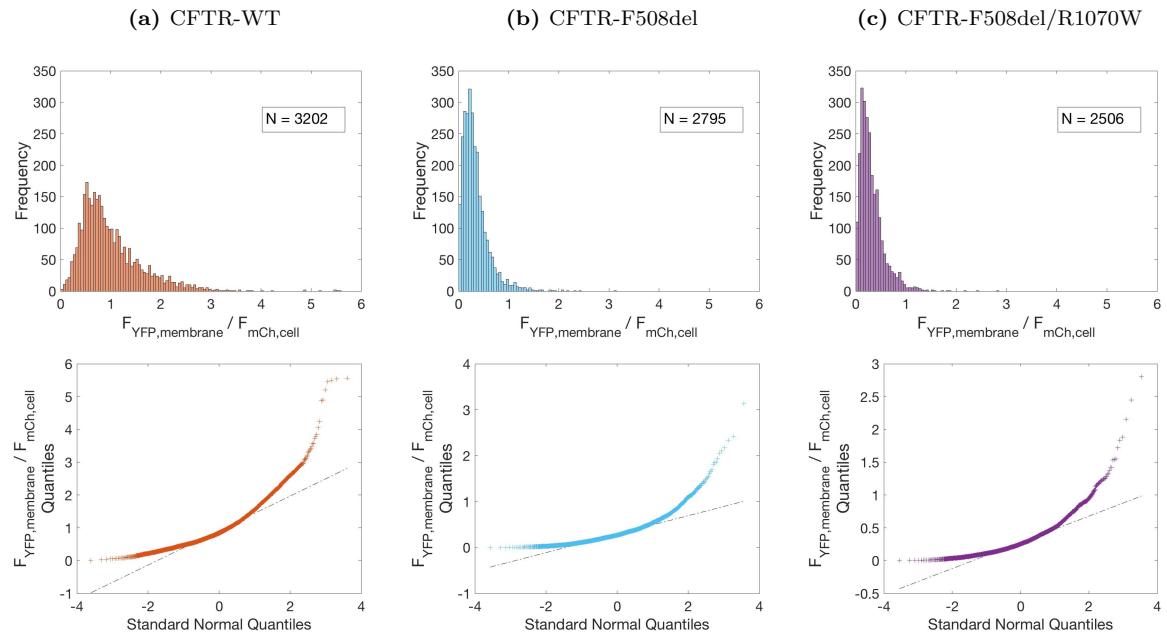
The plots in figure B.1 demonstrate the mean mCherry and YFP fluorescence intensity is correlated. This also ensures the validity of using the mCherry intensity within the cell as a means of normalizing for differences in expression. The differences in slope represent between conditions is most likely due to the increased degradation of YFP at the ER.



**Figure B.1:** Scatter diagram plotting the mean mCherry fluorescence within a cell against the mean YFP fluorescence. This figure demonstrates the relationship between the expression of mCherry and YFP, which can be seen to be roughly linear. Each point represents a single cell. The line is the result of performing a linear regression via calculation of the least squares error with the intercept forced through the origin.  $R$  is the Pearson correlation coefficient and MSE is the mean squared error.

## C Tests for normality of localization data

In order to justify the use of the Kruskal–Wallis test for testing whether the different genotypic conditions for CFTR show a statistically significant difference in normalized values for  $\Phi$ , it is first necessary to demonstrate that the data is not normally distributed. This is demonstrated graphically in figure C.1. For this data, it is not worth performing a Kolmogorov–Smirnov test for normality as the number of observations in each group are so large that a significant result is almost guaranteed.



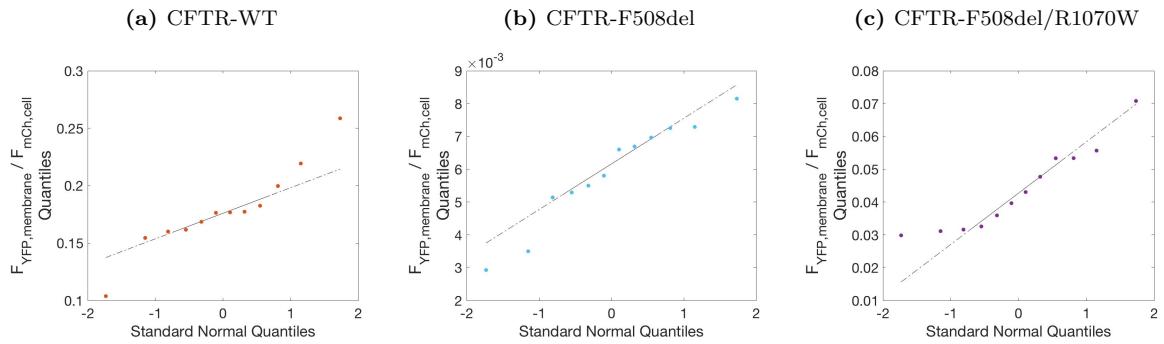
**Figure C.1:** Histograms and QQ-plots for each condition, which aim graphically to show that the values for the normalized metric,  $\Phi$ , are not normally distributed.

## D Tests for normality of gating function data

In order to test whether the results for the maximum gradient of the iodide inflow were normally distributed, a Kolmogorov–Smirnov test was performed for each condition. The results are given in table D.1. In addition, figure D.1 gives a graphical representation of the normality of the data.

Genotype	WT	F508del	F508del/R1070W
K–S statistic	0.5413	0.5012	0.5119
<i>p</i> -value	0.0008	0.0026	0.0019

**Table D.1:** A table giving the test statistics from the Kolmogorov–Smirnov test and corresponding *p*-values when performed on the maxima of gradient in-flow of iodide ions. Results are provided for each genotype of *CFTR*.  $n = 12$  for all conditions.



**Figure D.1:** QQ-plots for each condition, which aim graphically to show that the values for the maxima of the rate of iodide in-flow are not normally distributed.