

Electrically heterogeneous: discovering the dynamic electrical signatures of human breast cancers

Department of Bioengineering

Imperial College London

*A Project Report Submitted in
Partial Fulfilment of the*

MEng --- Degree

Supervisor:
Dr Amanda Foust

Department of Bioengineering
a.foust@imperial.ac.uk

April, 2024

Electrically heterogeneous: discovering the dynamic electrical signatures of human breast cancers

Ai Niimura, Ivy Ng, Tian Pan, Hongyu Rui, David Silcock, Trevi Yek

May 27, 2024

Abstract

Breast cancer cells exhibit dynamic membrane potential fluctuations compared to normal cells, offering insights into cancer dynamics. Quicke et al. used the electrical signatures of cells to characterise membrane potential dynamics in the MDA-MB-231 breast cancer cell line. This study builds upon this previous work by analysing nine additional cell lines, investigating the intercellular spatial and temporal patterns, and hypothesising the underlying biophysical mechanisms. Cell lines can then be categorised by comparing their electrical signatures. We investigate hyperpolarisation patterns within and between different breast cancer cell lines using data obtained from voltage-sensitive dye imaging, employing both spatial and temporal statistical analysis. Additionally, cell lines including MCF10A treated with TGF β , which induces cancerous qualities, and melanoma cells, were also analysed. To investigate temporal patterns, statistical methods such as the Pearson's Correlation Coefficient and Spike Time Tiling Coefficient were used. Subsequently, t-tests determine if each cell line's patterns significantly deviate from randomness, while the Kruskal-Wallis test and Analysis of Variances assess differences between cell lines. Extending the analysis to the spatial dimension, Pearson's Correlation Coefficient and density-space clustering were used to ascertain whether correlations are also dependent on distance. Temporal analysis reveals statistically significant cell lines, including MDA-MB-231, MDA-MB-468 and WM, suggesting non-random interactions likely mediated by underlying biological mechanisms. Further spatio-temporal analysis indicates that cancerous cells in close proximity exhibit stronger temporal correlations, potentially indicative of direct cell-to-cell communication. As intercellular communication is known to induce cancer proliferation, analysing bioelectrical dynamics could provide insight into cancer development and progression.

1. Introduction

1.1. Project Overview

Imaging human breast cancer cells with voltage-sensitive dyes revealed dynamic membrane potential (V_m) fluctuations not observed in non-cancerous cells [1]. While Quicke et al. quantified the rate of hyperpolarising events across cell lines [1], our study aims to further investigate the spatial and temporal correlations by quantifying the negative events occurring in different cells for each cell line. Understanding the characteristics of such correlations could reveal the underlying biophysical mechanisms driving electrical signalling among cancer cells [2]. This comprehension of the bioelectrical signatures provides insight into the behaviour of cancers and holds promise for discovering new diagnostic and therapeutic biomarkers, aiding breast cancer identification and characterisation. To accomplish this, we first expanded on previous temporal analyses [1], before performing novel statistical spatio-temporal analyses.

1.2. Objectives

The primary objectives of the project were:

- To determine the presence of spatial and temporal correlations between cells in human breast cancers
- To identify any spatio-temporal patterns within the cells, and evaluate whether it suggests the presence of distinct biophysical mechanisms within the cancer
- To analyse the similarities and differences between the cell lines, enabling the characterisation of cancers based on bioelectrical behaviour

1.3. Characterisation of Cell Lines

Breast cancer is a heterogeneous disease. Figure 1 presents the breast cancer cell lines we analysed along with their corresponding subtypes: Luminal, Basal A, and Basal B, which are categorised based on expressed receptors and aggressiveness [1]. Low-grade luminal tumours are oestrogen receptor (ER) positive and progesterone receptor (PR) positive, with the absence of human epidermal growth factor receptor 2 (HER2) [3]. Highly aggressive basal-like breast cancers, known as triple-negative breast cancers (TNBC), lack expression of ER, PR or HER2 receptors [4], making hormone or targeted therapies ineffective, posing treatment challenges. The MDA-MB-231 cell line, known for its invasive phenotype, has been extensively utilised in study-

ing the tumour microenvironment. It serves as a valuable model for identifying genes and signalling pathways implicated in cancer progression and metastasis [5].

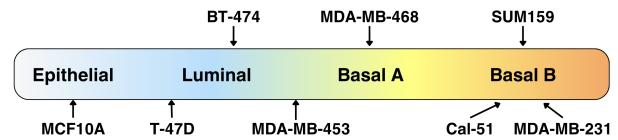


Figure 1: Cell line subtypes [1]

Additionally, this study includes analysis of MCF10A epithelial cells induced with $TGF\beta$ to exhibit cancerous traits (MCF10A- $TGF\beta$) and a melanoma cell line (WM), investigating whether similar characteristics extend beyond breast cancer to skin cancers. Spatial data was limited to breast cancer cells, thus constraining the spatio-temporal analysis.

The data analysed in this study, similar to Quicke et al.'s research [1], consisted of voltage time series for each cell line, with variations in the number of Field of Views (FOVs) and the number of cells within each FOV, as outlined in Table 1.

Table 1: Available Data for each cell lines

Cell Line	Number of FOVs	Number of cells
MCF10A	15	30
MCF10A- $TGF\beta$	25	69
T-47D	29	155
BT-474	33	371
MDA-MB-453	7	41
MDA-MB-468	26	858
Cal-51	32	128
SUM159	23	57
MDA-MB-231	39	412
WM	28	329

2. Background

Breast cancer, a leading cause of cancer mortality globally, accounted for an estimated 2.3 million new diagnoses in 2020 [6]. Metastasis is the major contributor to cancer mortality, clinically believed to be responsible for 90% of all cancer deaths [7].

2.1. Membrane Potential in Breast Cancer Progression

The resting membrane potential is established through the differential permeability of cell membrane to var-

ious ions, and is typically around -70mV in most excitable cells like nerves and muscles [1][8]. A change in V_m , facilitated by voltage-gated sodium and potassium channels, triggers the generation and propagation of action potentials. Electrophysiological recordings have shown that rapidly proliferating cells, including cancer cells, possess a more depolarised V_m . Cells with strong metastatic potentials have a more depolarised V_m compared to weakly or non-metastatic cells [2]. For example, in MDA-MB-231, V_m rests between -40 and -20mV [1].

The 'Celex hypothesis', proposed by Djamgoz, suggests that membrane excitability induces cellular hyperactivity, thereby promoting cell invasion and ultimately metastasis [9]. Hence, the study of cancer cells' distinct bioelectrical properties is of emerging significance. Ion channels are closely linked to cell V_m , playing a significant role in cell signalling - a complex process fundamental to cancer development.

2.2. Pathological Role of Ion Channels in Breast Cancer

Gene expression profiling identified 30 ion channel genes, significantly correlated with histological tumour grade [10]. Anomalous expression of ion channels is associated with tumorigenesis, including cell proliferation, adhesion, migration, and invasion. The 'Celex hypothesis' is hence rooted in voltage-activated membrane currents, predominantly from functional voltage-gated sodium channels (VGSCs) which have been shown to play a significant role in inducing metastasis in various cancers [11]. An upregulation of the neonatal splice form of $\text{Na}_v1.5$ is shown in MDA-MB-231 which facilitates epithelial-to-mesenchymal transition, promoting invasiveness and metastasis.

Membrane depolarisation and increased numbers of VGSCs are proposed biomarkers for cancer [12]. Recently, Payne et al. characterised hyperpolarisation-induced metastasis in the MDA-MB-231 cell line, which escalated cell migration and invasion, and attributed it to Kv1.5 and Kir 2.1 potassium channel overexpression [13]. Hyperpolarisation is necessary for the G1/S checkpoint and S phase transition [2], suggesting a higher proportion of cells in a transient hyperpolarised state in rapidly proliferating tumour cells compared to healthy tissues. This evidence of induced hyperpolarisation due to K^+ channel overexpression, presented by Payne et al, suggests potential heterogeneity in the bioelectrical signatures of metastatic breast cancer.

2.3. Intercellular Communication in Cancer

Cell-to-cell communication via paracrine signalling or through specialised structures plays an important role in cancer progression.

Paracrine signalling can influence breast cancer heterogeneity. For instance, interactions between cancer cells and fibroblasts enrich the tumour microenvironment, promoting tumour growth [14]. Additionally, abnormal calcium ion (Ca^{2+}) channel expression, along with intra- and extracellular messengers, can influence Ca^{2+} signalling, contributing to cancer progression and metastasis. In particular, IP3 receptors regulate cell proliferation via intercellular Ca^{2+} wave progression [15, 16].

Specialised gap junctions (GJs), shown in Figure 2, allow direct cell-to-cell communication through rapid passive exchange of molecules, ions and small metabolites. Composed of connexins, GJs are voltage sensitive, facilitating electrical coupling and cell synchronisation [17]. GJs act as tumour suppressors in early stages of tumorigenesis, with deregulation of connexins Cx26 and Cx43 observed in human breast cancer cell lines, including MDA-MB-231 [18, 19]. However, in later stages, overexpression of connexins are detected and GJs become promoters of invasion in migrating tumour cells [20]. The complex mechanism underlying connexins and GJs in cancer pathogenesis remains a major barrier to evolutionary Cxs-targeted therapy.

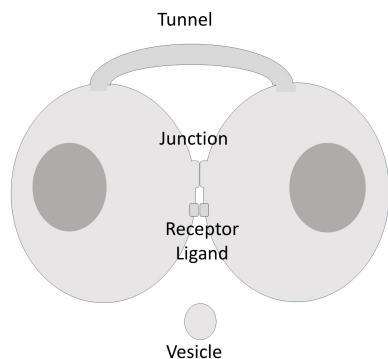


Figure 2: Specialised gap junctions, tunnel nanotubes and extracellular vesicles for cell-to-cell communication [17]

Furthermore, tunnel nanotubes (TNT) could facilitate long-range electrical coupling in cancer cells [17], with distances up to several microns [21]. Increased TNT levels and intercellular exchange of drug-resistant proteins via TNTs suggest their involvement in cell communication and potential as therapeutic targets to address chemoresistance [22]. Additionally, extracellular vesicles (EVs) contribute significantly to cell-to-cell communication in cancer metastasis with its capability of travelling short to long distance to share oncogenic moieties to metastatic niches.

2.4. Voltage Imaging and Time Series Acquisition

The electrochromic voltage-sensitive dye, di-4-AN(F)EP(F)PTEA, inserted into the outer membrane, was used for visualising V_m [1]. This dye exhibits spectral shifts, corresponding with changes in V_m , with an excitation peak at 465nm [23]. Blue and green LEDs, with wavelengths lying on either side of this peak, have excitation emissions which are inversely affected by V_m changes. By sequentially exciting the dye with both LEDs, this allowed calculation of V_m fluctuations, calculated from the ratio ΔR of fluorescence excited by each wavelength [1], represented by pixel intensity. The normalised pixel intensity, $\frac{\Delta R}{R_0}$, was then computed by dividing this ratio by the baseline ratio R_0 to ensure data comparability across images.

Utilising a wide FOV at cellular resolutions, hundreds of cells were imaged simultaneously. Sequences of 5000 images per FOV were acquired at 5Hz, followed by cell segmentation using CellPose and manual verification [1]. Each cell was designated as an individual region of interest (ROI). Cells were excluded if they were affected by imaging artefacts, including floating dust, mechanical vibrations, and illumination edge effects. Additionally, cells exhibiting signs of being dead or dying, where raw pixel values increased over 25% throughout the measurement duration, were also eliminated. To isolate cells with V_m fluctuations, ROIs where the pixel intensities deviated by less than 2.5 standard deviations from the median were also excluded. Time series for each cell were constructed from the median pixel intensities in successive images for each remaining ROI.

3. Methods

To fulfil the objectives of identifying temporal and spatial patterns in cell voltage fluctuations, we first convert these fluctuations into spike trains. Since hyperpolarisation spikes were predominantly observed, each spike represents a time-binned hyperpolarisation event. Using the spike trains for pairwise cells, correlation coefficients were calculated to quantitatively assess intercellular correlations, assuming that events in one cell may influence neighbouring cells.

The temporal analysis involves assessing whether the correlation coefficients deviate from randomness, suggesting structured interactions among cells. Additionally, it investigates whether distinct correlation patterns exist among different cell lines. Following this, spatio-temporal analysis examines the distances over which these temporal patterns occur, facilitating hy-

potheses regarding the biophysical mechanisms underlying these events.

3.1. Converting Time Series Data into Spike Trains

Hyperpolarisation events are detected by comparing the voltage time series to a threshold set at 2.5 standard deviations. Following this, the time series is divided into bins, with '1' indicating event presence (spike) and '0' indicating absence. For small time bins, an additional threshold ensures a spike is marked as '1' only if over 50% of the discrete values exceed the set threshold. This process aggregates segments into a spike train, converting continuous voltage fluctuations into discrete events.

3.2. Quantifying Temporal Correlations

Intercellular pairwise correlations in spike trains for each cell were quantified using two methodologies: Pearson's Correlation Coefficient (PCC) and Spike Time Tiling Coefficient (STTC) (see Appendices A.1 and A.2 respectively). Both methods were applied to the same cell pair configurations across each FOV to explore spatial and temporal patterns. PCC offers a broad analysis of linear relationships between spike trains, suitable for general correlation assessments. In contrast, STTC is tailored for analysing spike trains, capturing the non-linear dynamics and intricate temporal relationships that PCC may overlook.

3.3. Methods for Temporal Correlation Analysis

After calculating correlations, statistical tests like the t-test determine whether observed correlations within a cell line could arise by chance. Subsequently, Analysis of Variance (ANOVA), and the Kruskal-Wallis (KW) test, followed by post-hoc analysis, determine differences between cell lines, including those with non-random correlations. Further thresholding for moderate and strong correlations identified the percentage of pairwise cells expressing these fluctuations.

3.3.1. T-Tests to Identify Intercellular Communication

The central limit theorem dictates that with large sample sizes, sample means follow a normal distribution. Hence, a two-sample unpaired t-test (see Appendix

A.3) assessed differences in means between two normally distributed populations. The data was compared to the mean PCC value of a randomised set of obtained spike trains, considering each FOV's mean PCC value as an independent data point. Rejecting the null hypothesis suggests that any correlations between cells are statistically different from randomness, potentially indicative of biophysical mechanisms.

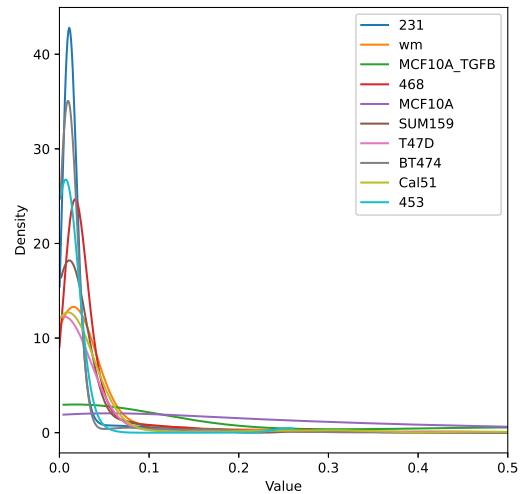
3.3.2. ANOVA and KW Tests with Post-Hoc Analysis to Determine Difference Between Cell Lines

Upon identifying cell lines whose correlations are not attributable to random chance, statistical methods used to analyse multiple samples, KW test and ANOVA, tested the null hypothesis that the intercellular temporal correlations across all cell lines are equivalent. From Figure 3, it is evident that the assumption of normality does not hold for the distribution of temporal correlations, therefore the KW test, a non parametric method, was initially utilised. In addition, ANOVA was conducted to validate the non-parametric analysis to ensure the robustness of results. To address ANOVA's normality assumption, bootstrapped was employed to approximate normally distributed data.

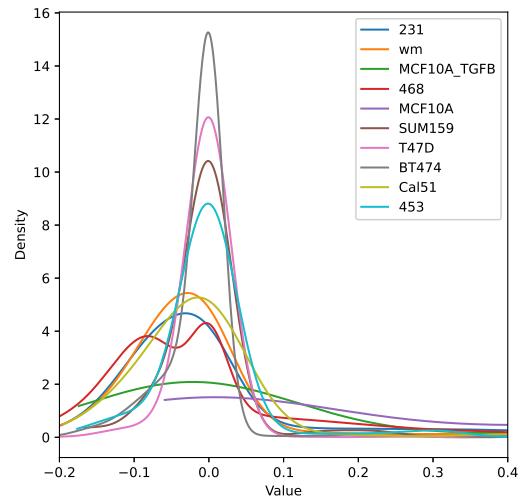
Since the null hypotheses for the KW test and ANOVA were rejected, indicating statistically different correlation means or medians among cell lines, a post-hoc analysis was conducted to determine the correlation across cell lines. The Bonferroni correction was used in conjunction with one-tailed t-tests for the post-hoc analysis. One-tailed t-tests provide directionality, determining which cell line contained more intercellular correlations. The Bonferroni correction divides the significance level (α) for each t-test by the number of pairwise comparisons, minimising Type I errors.

3.3.3. Evaluating Moderate and Strong Correlations Coefficients for Synchronous Events

Correlation coefficients between 0.4 and 0.8 indicate moderate temporal correlations, while values above 0.8 signify strong correlations [24]. These thresholds, coupled with analysing the percentage of cell pairs expressing temporal correlations, offers insight into underlying biophysical mechanisms. For the PCC, time bins of 1s and 10s are chosen, while for STTC, time windows of ± 1 s and ± 5 s capture biophysical mechanisms like gap junctions and calcium waves.



(a) Estimated probability density functions of the PCC values for all cell lines



(b) Estimated probability density function of the STTC values for all cell lines

Figure 3: Estimated probability density functions of PCC and STTC values in all cell lines showing non-normal distribution

3.4. Methods for Spatio-Temporal Analysis

The correlation coefficients obtained from temporal analysis were incorporated alongside statistical methods and density-based clustering to explore spatial and temporal correlations within a cell line. Plotting pairwise correlations against the intercellular distances facilitated the identification of potential relationships between cell proximity and temporal correlations, aiding in cell line characterisation. Additionally, density-based clustering was used to further explore potential

biophysical mechanisms; the additional spatial dimension aids in gaining a deeper understanding of the observed correlations in the former method.

3.4.1. Finding the Relationship Between Intercellular Distances and Pairwise Correlation

To explore the spatial patterns, the Euclidean distance (Equation 1), was calculated between the cells.

$$d = \sqrt{(x_2^2 - x_1^2)^2 + (y_2^2 - y_1^2)^2} \quad (1)$$

In the original images, each pixel between cells in an FOV corresponds to a distance of 1.04 microns, and it is assumed that cells maintained their positions during imaging. Scatter plots of the pairwise correlation coefficients against their corresponding distances were constructed to explore the possible link between cell proximity and temporal correlations. To refine the analysis, only correlation coefficients exceeding 0.4 were analysed, focusing solely on synchronous events. This approach aims to use the PCC to determine if a spatial relationship exists between the occurrence of synchronous events.

In previous temporal analysis, various bin sizes were computed for the PCC, however for this analysis, only bin sizes of 1s and 10s were selected. The selection of 1s bins reflects the instantaneous nature of patterns arising from GJs. Moreover, a bin size of 10s was chosen to capture the temporal dynamics associated with calcium waves as they range from microseconds to tens of seconds [25]. Similarly, for the STTC, time windows of ± 1 s and ± 5 s were examined to assess the degree of synchronisation between cells.

3.4.2. Spatio-Temporal Density-Based Clustering

To visualise spatio-temporal events, we used Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise algorithm (ST-DBSCAN), a variation of the traditional DBSCAN, tailored for our specific application [26, 27], (see Appendix B). This Python package plotted the cell positions within each field of view (FOV) and clustered the points based on spatial density and temporal events. It uses normalised cell coordinates and the timestamps for all hyperpolarisation events.

An example is depicted in Figure 4, where each cluster is represented by a distinct colour.

DBSCAN identifies core points and forms clusters around them using a fixed radius and considering the

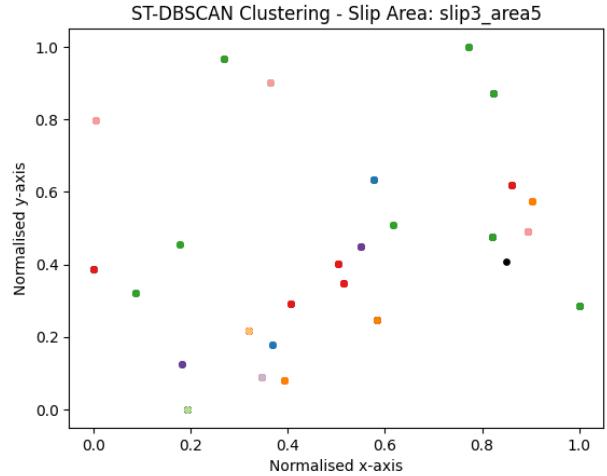


Figure 4: A scatter plot showing the spatial distribution of events detected within Slip 3 Area 5 for MDA-MB-468 cells. The parameters used are $\epsilon_1 = 0.9$, $\epsilon_2 = 1$ and Minimum Samples = 6. Events are clustered using ST-DBSCAN, with each cluster distinguished by a unique colour. Noise points, denoted in black, represent isolated data points.

density of neighbouring points [28]. ST-DBSCAN extends this by incorporating event timestamps as an additional dimension. Parameters for this analysis included spatial radii (ϵ_1), temporal radii (ϵ_2), and minimum samples - the minimum number of points required to be considered a core point [29].

4. Results

4.1. Temporal Correlation Analysis Results

4.1.1. Significant Correlations Found in the MDA-MB-231, MDA-MB-468, and WM Cell Lines

To evaluate temporal correlations between transient hyperpolarisations of cells within a FOV, pairwise PCC values were calculated, followed by t-tests for each time bin size and cell line. A subset of the results shown in Figure 5 for the MDA-MB-231, MDA-MB-468, Cal-51 and WM cell lines, see Appendix C for the full results. The green box represents $p < 0.05$ for that bin size, and a red box represents $p > 0.05$.

From the t-tests it is evident that MDA-MB-231 (Figure 5a), MDA-MB-468 (Figure 5b), and WM (Figure 5c) cell lines are statistically different from randomness at smaller bin sizes. The other cell lines, such as Cal-51 (Figure 5d) show no statistical differences at smaller

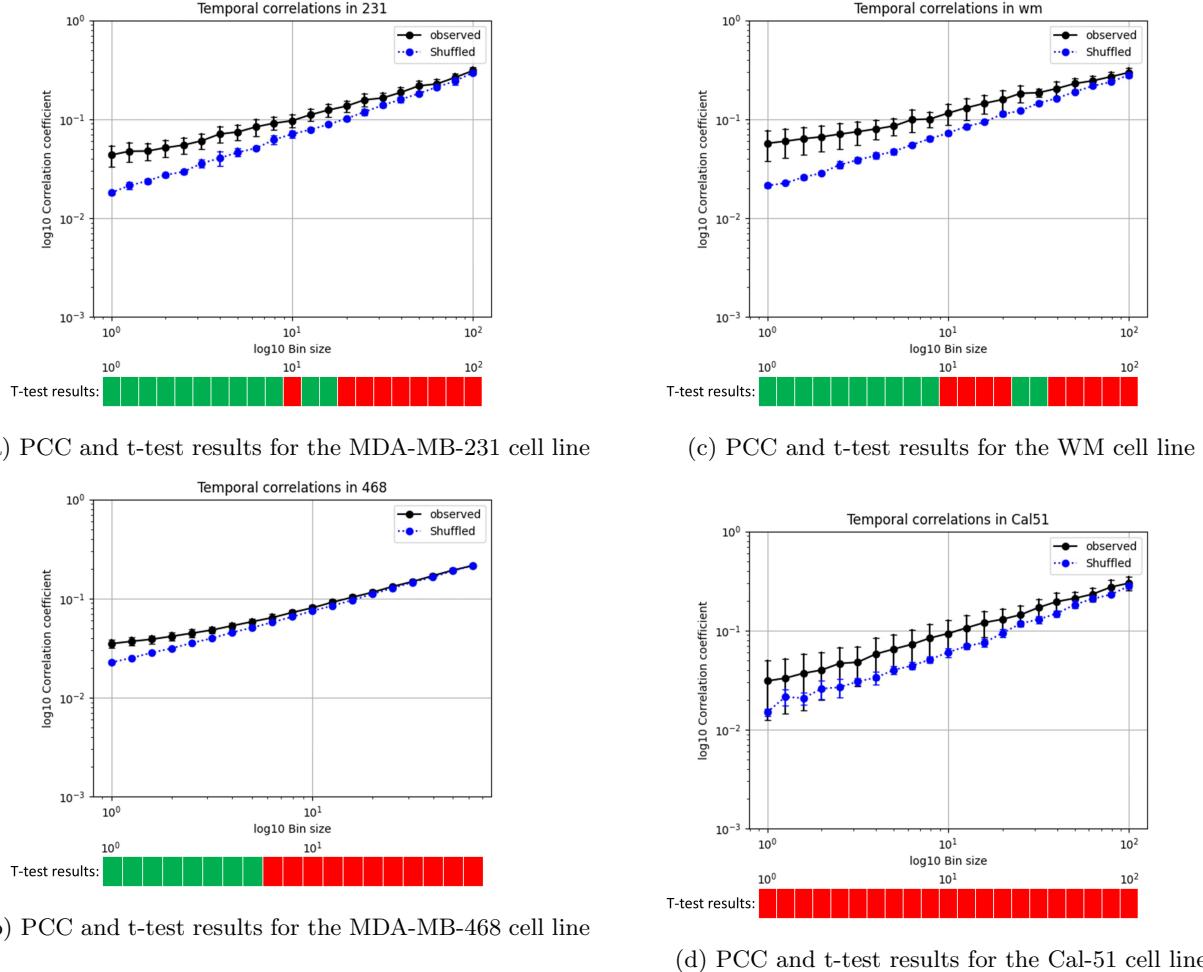


Figure 5: Graphs of logarithmic PCC values against bin sizes ranging from 10^0 to 10^2 in increments of \log_{10} . The black line represents PCC values calculated using the observed time series, while the blue line corresponds to PCC values calculated from randomised data. For each t-test, performed on each bin size, green markers indicate a statistical difference with a p-value of $p < 0.05$, while red markers indicate a p-value of $p \geq 0.05$.

bin sizes, largely due to a lack of data resulting in large standard deviations. At larger bin sizes, there are no statistical differences observed due to fewer data points available to randomise, resulting in more coincidental correlations and higher PCC values.

4.1.2. Differences Found Between the Cell Lines

To evaluate differences in temporal correlation among cell lines, both the KW test and ANOVA was used, focusing on all cell lines. Using both correlation coefficient indexes with a time bin of 10s, both the KW test and ANOVA rejected the null hypotheses (p values in Appendix D), indicating a statistical difference in correlation of at least one of the cell lines compared to others.

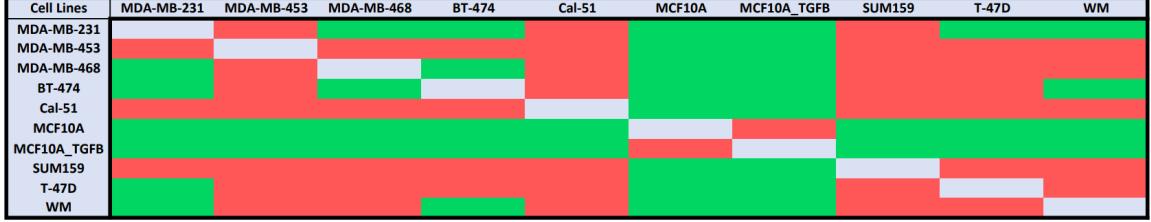
After rejecting the null hypothesis in the previous tests,

the post-hoc analysis enabled us to identify differences between the cell lines. Figure 6 shows the results, depicting non-rejections as red and rejections as green. The analysis focused on 1s time frames to capture synchronous events occurring instantaneously. Post-hoc analysis highlighted that MCF10A and MCF10A-TGF β exhibit statistical differences from all cancerous cell lines, but not each other. Furthermore, both STTC and PCC indicated statistical differences between MDA-MB-231, MDA-MB-468, and BT-474. For the PCC values (Figure 6a), WM is shown to be statistically different to all breast cancer cell lines except Cal-51 and SUM159. However, the STTC (Figure 6b) indicated that WM is also statistically similar to MDA-MB-453, MDA-MB-468, and T-47D.

Table 2 shows the percentages of the moderate and strong correlations for the PCC and STTC values across different time bins. Noticeably, as the bin size



(a) Post-hoc test results for PCC with a 1s time bin



(b) Post-hoc test results for STTC with a 1s time bin

Figure 6: Post-hoc test results for PCC and STTC with a 1s time bin. The red indicates the null hypothesis is not rejected while the green indicates a rejection.

Table 2: Percentages of moderate and strong correlations for PCC and STTC values for different bin sizes for all breast cancer cell lines

(a) Percentages of moderate and strong correlations with PCC values using different bin sizes

Cell line	Percentage of PCC >0.4		Percentage of PCC >0.8		Total Pairs
	Time bin 1s	Time bin 10s	Time bin 1s	Time bin 10s	
MCF10A	7.1429	14.2857	0.0000	0.0000	28
MCF10A-TGFB	15.6250	21.8750	3.1250	7.8125	128
BT-474	0.0854	0.8366	0.0000	0.0341	5857
MDA-MB-231	0.7613	2.8820	0.0816	0.3535	3678
MDA-MB-468	0.4302	0.9565	0.0962	0.1670	19760
T-47D	1.1312	1.1312	0.0000	0.0000	884
MDA-MB-453	0.0000	0.6098	0.0000	0.0000	164
Cal-51	0.6369	2.5478	0.0000	0.3185	314
SUM159	0.0000	0.0000	0.0000	0.0000	79

(b) Percentages of moderate and strong correlations with STTC values using different time windows ($\pm \Delta t$)

Cell line	Percentage of STTC >0.4			Percentage of STTC >0.8			Total Pairs
	$\pm 1s$	$\pm 5s$	$\pm 10s$	$\pm 1s$	$\pm 5s$	$\pm 10s$	
MCF10A	21.4286	32.1429	32.1429	0.0000	0.0000	7.1429	28
MCF10A-TGFB	19.5313	23.4375	24.2188	8.5938	13.2813	13.2813	128
BT-474	0.5805	1.4342	2.3220	0.0341	0.0854	0.3585	5857
MDA-MB-231	2.3382	6.9603	10.5492	0.1903	0.6525	1.3594	3678
MDA-MB-468	1.5891	3.8057	6.5334	0.1771	0.2328	0.3289	19760
T-47D	2.1493	2.4887	2.8281	0.0000	0.3394	0.6787	884
MDA-MB-453	0.6098	0.6098	1.2195	0.0000	0.0000	0.0000	164
Cal-51	2.5478	4.4586	5.7325	0.0000	0.6369	1.2739	314
SUM159	0.0000	0.0000	1.2658	0.0000	0.0000	0.0000	79

increases for MDA-MB-231, the percentage of values with moderate or strong correlations increases more compared to MDA-MB-468 and BT-474. When observing the 1s time bin to investigate synchronous events, the order of correlations, from highest to lowest, is observed as follows: MDA-MB-231, MDA-MB-468 and then BT-474. It is also noteworthy that cell

lines including MCF10A, SUM159, T-47D, Cal-51 and MDA-MB-453 have no strong correlations for both PCC and STTC.

4.2. Spatio-Temporal Analysis Results

4.2.1. Neighbouring Cells are More Temporally Correlated

The moderately strong correlation coefficient indexes are plotted against intercellular distances for all cell lines to explore whether there is a higher frequency of events between cells that were in close proximity (see Appendix E). Figure 7 displays the plots for the MDA-MB-231, MDA-MB-468, and BT-474 cell lines, which had a larger number of data points. Moderate and strong correlations are indicated by blue and grey shading respectively. Generally, the plots show that cells in closer proximity have higher temporal correlations, evidenced by the top left corner having a higher point density compared to the top right corner across various time frames for PCC and STTC.

4.2.2. Evaluation of PCC and STTC Plot with Statistical Methods

Figure 8 illustrates the PCC between temporal correlation coefficients greater than 0.4 and distances, with 95% confidence intervals. It is notable that the correlation coefficient is negative for most cell lines, suggesting that cells closer together tend to display stronger correlations in time compared to more distant cells. Confidence intervals not including zero indicate significant results, notably for MDA-MB-468 and MDA-MB-231. However, other cell lines, particularly MCF10A, Cal-51, and T-47D, exhibit very large confidence intervals resulting from limited data availability.

4.2.3. Change in Cluster Size for Different ϵ_1 and ϵ_2 Values

Three graphs were produced to investigate the biophysical mechanisms of different cell lines (Figure 9). Figure 9a shows a plot of average cluster size against spatial radii for $\epsilon_2 = 1\text{s}$, aiming to evaluate the effect of distance within a short time, implying instantaneous signalling. The graph shows an increase in cluster numbers for MDA-MB-231, MDA-MB-468, and BT-474, reaching a plateau at higher ϵ_1 values. Notably, MDA-MB-468 exhibits the highest number of clusters across all ϵ_1 values of just below 35, followed by MDA-MB-231 and BT-474. Both MDA-MB-468 and MDA-MB-231 reach a plateau in cluster count at $\epsilon_1 = 0.4$, indicating minimal impact on clustering with further increases in ϵ_1 . When ϵ_2 was increased to 10s (Figure 9b), the graph produced an exponential decrease with MDA-MB-468 having the highest number of clusters.

Subsequently, Figure 9c was produced by maintaining a

constant spatial radius of 0.3, and varying time intervals to investigate whether hyperpolarisations occurring within a time interval also occur in neighbouring cells. Across all cell lines, there is a sharp rise in clusters, followed by an exponential decrease, leading to a plateau for higher ϵ_2 values. MDA-MB-468 exhibits a steep decrease in clusters, especially between 1s and 20s bin sizes, while MDA-MB-231 and BT-474 show a shallower decline. Among the cell lines, MDA-MB-231 demonstrates the highest cluster count of 40 and the most rapid decline across all ϵ_2 values, followed by MDA-MB-468. Conversely, BT-474 exhibits a consistent decrease. For all three graphs, the rest of the cell lines did not indicate any significant changes for all ϵ values.

5. Discussion

5.1. Temporal Correlation Analysis

To quantify temporal behaviours, T-tests confirmed non-random correlation patterns, with significant results observed in MDA-MB-231, MDA-MB-468, and WM cell lines, suggesting that correlation in voltage fluctuations were caused by intercellular interactions. This correlation may be attributed to communication via GJs, which have been shown to regulate cancer metabolism by exchanging metabolites between hypoxic and oxygenated cells to sustain growth in MDA-MB-231 cells [30].

Further post-hoc analysis and percentages in temporal correlations from Table 2, revealed that MDA-MB-231, MDA-MB-468, and BT-474 cell lines have distinct hyperpolarisation patterns, indicating differing characteristics within these cell lines. This is likely related to factors such as tumour histological grade, heterogeneity in GJs, and gating potential of GJs, altered through post-translational modification of connexins [17].

When examining strong temporal correlations, MDA-MB-231 cells show a significant increase in percentage of synchronised events with larger time bins as shown in Figure 10b. This suggests heightened cell-to-cell communication or synchronisation of biological processes over extended periods, aligning with the more aggressive cancer phenotype of MDA-MB-231 [31]. Conversely, the percentage of strongly correlated MDA-MB-468 cells increased less with larger time bins, indicating sporadic synchronisations as shown in Figure 10d. This scattered pattern implies a different mode of cellular communication or behaviour, reflecting a more heterogeneous distribution of electrical events in MDA-MB-468 cells. The observed differences offer valuable insights into breast cancer behaviour and the underlying cellular dynamics of these distinct phenotypes.

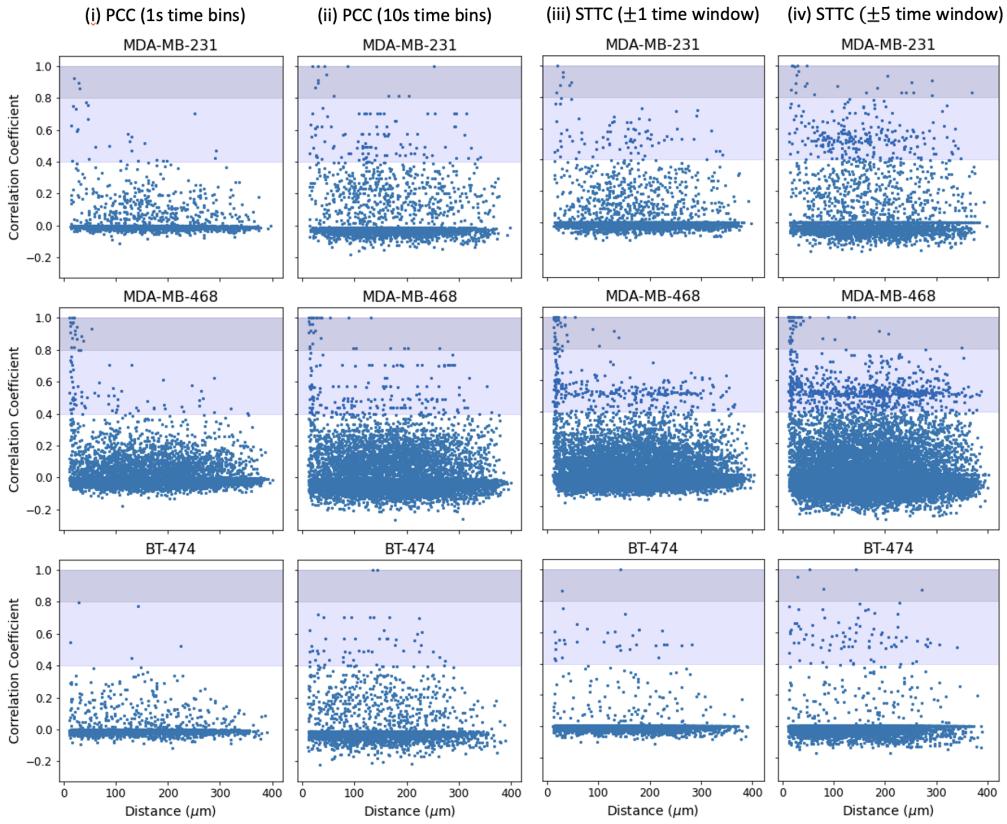
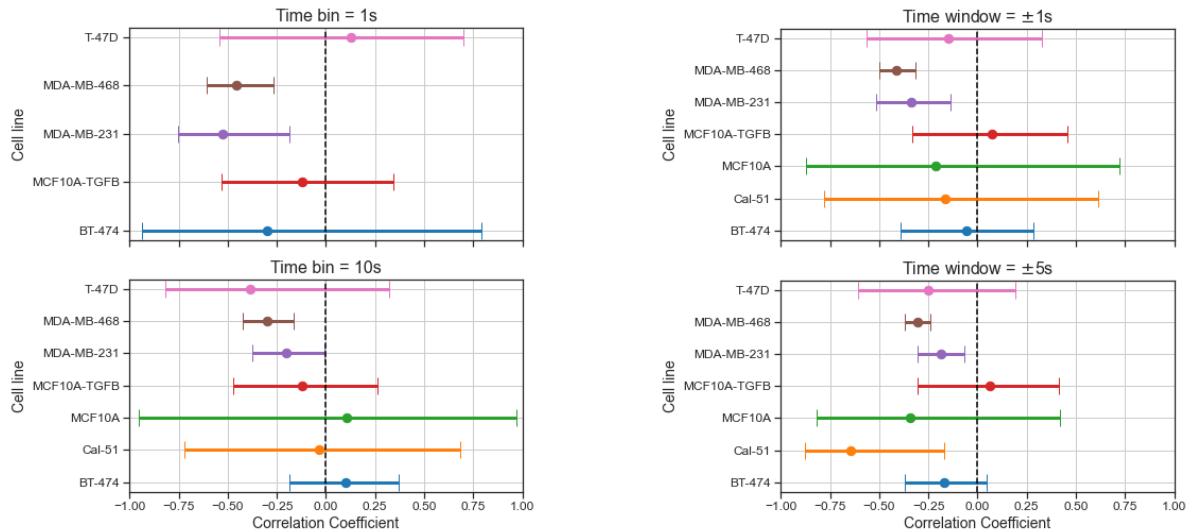


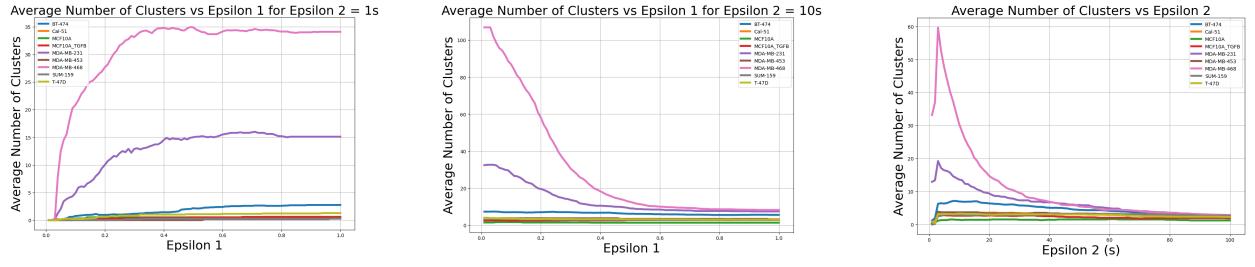
Figure 7: Scatter plots of the PCC and the STTC values against distance for all hyperpolarisation events in cell lines MDA-MB-231, MDA-MB-468 and BT-474. Cells in close proximity exhibit higher temporal correlation, shown by the points in the top left corner. The blue shading represents moderate correlations, while the grey shading indicates strong temporal correlations.



(a) Correlation coefficients for PCC and distance for different time bins

(b) Correlation coefficients for STTC and distance for different time windows

Figure 8: Plots of correlation coefficients, with a 95% confidence interval, for different cell lines. Most cells show a negative correlation between distance and temporal correlations. Significant results for MDA-MB-231 and MDA-MB-468 due to a small confidence interval excluding 0.



(a) Plot of average cluster number against ϵ_1 values ranging from 0.01 to 1 in increments of 0.01 at $\epsilon_2 = 1\text{s}$.

(b) Plot of average cluster number against ϵ_1 values ranging from 0.01 to 1 in increments of 0.01 at $\epsilon_2 = 10\text{s}$.

(c) Plot of average cluster number against ϵ_2 values, ranging from 1s to 100s in 1s increments, with $\epsilon_1 = 0.3$.

Figure 9: ST-DBSCAN was applied to nine different cell lines, each represented by a different colour. Notably, MDA-MB-468 is the pink line, MDA-MB-231 is purple and BT-474 is blue. The Average Cluster Number is plotted against a range of ϵ_1 and ϵ_2 values for a Minimum Sample value of 6.

Furthermore, with 1s time bins, indicative of synchronous spikes, there is a higher occurrence of strong correlations in MDA-MB-231 compared to MDA-MB-468 cells. This suggests stronger intercellular communication, potentially influenced by the dynamic extracellular environment. Connexin, particularly Cx43, integral to GJ formation, is known to be a tumour suppressor and is down-regulated in TNBC tissues [32]. However, under low-glucose conditions, upregulation of Cx43 with increased cell-to-cell contact and GJ intercellular communication are shown in MDA-MB-231 [33]. Microenvironmental factors can affect cellular excitability and synchronisation, while differences in signalling pathways, including ion channels and receptors, may contribute to variations in spike occurrences [33]. Additionally, inherent cellular heterogeneity among cell lines may also influence GJ intercellular communication and signalling pathways. For instance, MDA-MB-468's epithelial-like phenotype [34], and less aggressive nature [31], may require fewer rapid signalling pathways, hence displaying a lower number of instantaneous spikes.

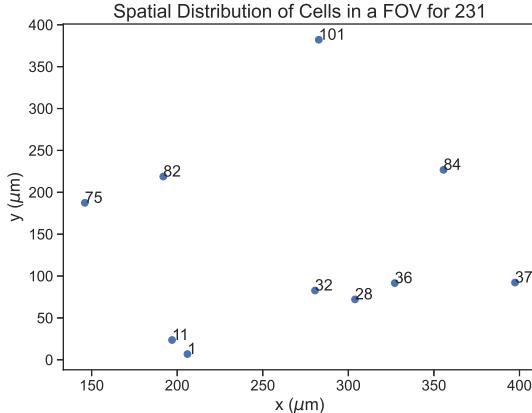
Post-hoc analysis further reveals statistical similarity between non-cancerous MCF10A and MCF10A-TGF β but differences between the cancerous cell lines. TGF β is known for its roles in tissue homeostasis and cancer progression, affecting immune evasion, cell invasion, and the tumour microenvironment [35, 36]. Despite its ability to induce changes in cellular behaviour and signalling pathways, MCF10A-TGF β cells are not inherently cancerous and do not exhibit the same bioelectrical activities as cancer cell lines. Since MCF10A-TGF β does not statistically differ from MCF-10A, a non-tumorigenic breast epithelial cell line [37], this underscores the limitation of solely relying on TGF β to model all facets of cancer biology, emphasising the need to consider the intricate molecular context and mi-

croenvironmental influences in cellular studies.

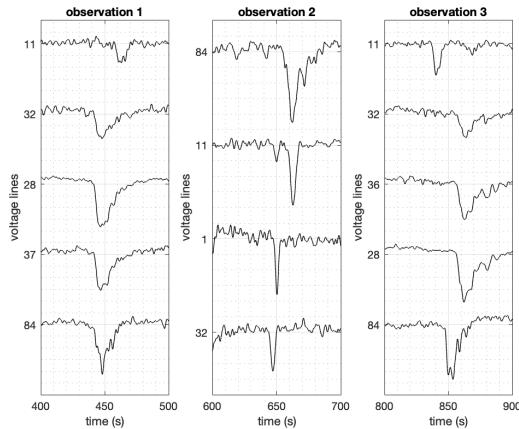
Moreover, considering WM, and excluding cell lines with small sample sizes, it can be inferred that the hyperpolarisations in MDA-MB-468 are similar. The commonalities in hyperpolarisation patterns observed could be influenced by various factors, including genetic mutations i.e. BRCA2 [38], signalling pathways, microenvironmental cues, and common intercellular communication mechanisms.

The absence of strong temporal correlations in MCF10A, SUM159, T-47D, Cal-51 and MDA-MB-453 cell lines in Table 2, can be attributed to several factors. These include cell line heterogeneity, diversity in signalling pathways, and microenvironmental influences, all of which can affect cellular responses and signalling dynamics. Moreover, significant epigenetic heterogeneity reported in basal-like breast cancer cell lines [39] underscores the role of genetic and epigenetic differences in contributing to cell line-specific variations in gene expression and protein function, leading to distinct cell behaviours.

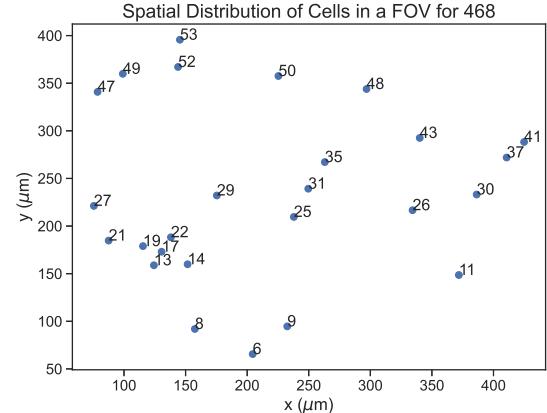
While limited data may partially explain the weak temporal correlations within these cell lines, it highlights the inherent complexity and diversity of cellular systems. Breast cancer genome sequencing and integrative epigenomic studies could further elucidate these differences, identifying specific phenotypes and underlying mechanisms. Additionally, the selection criterion during the time-series extraction stage, favouring cells with fluctuations exceeding 2.5 standard deviations, might disproportionately affect MCF10A, an epithelial cell line known for its stable membrane potential and reduced instances of hyperpolarisations. Consequently, many normal cells may have been excluded, leading to many FOVs containing one cell and resulting



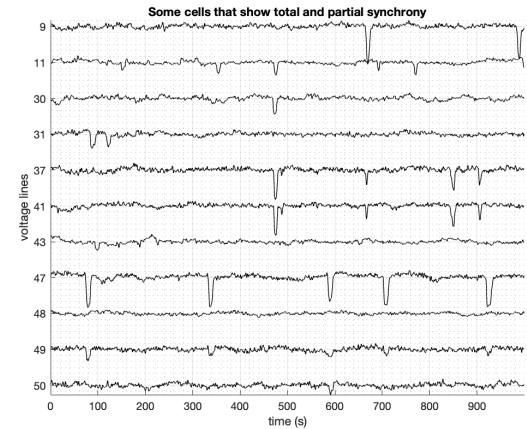
(a) Spatial distribution of cells for a FOV for MDA-MB-231



(b) Time series for some cells in a FOV for MDA-MB-231



(c) Spatial distribution of cells for a FOV for MDA-MB-468



(d) Time series for some cells in a FOV for MDA-MB-468

Figure 10: Spatial and time series plots for an FOV for MDA-MB-231 and MDA-MB-468 highlighting the relationship in synchronous events and space between these two cell lines

in a skewed dataset representing the electrical activity of the cell line, as shown in correlation percentages in Tables 2a and 2b.

Despite the promising discoveries regarding temporal relationships in cancer cell lines, further investigation is necessary to validate these observations and elucidate the underlying mechanisms driving the hyperpolarisation expression patterns among the studied cell lines. This may unveil specific therapeutic targets aimed at disrupting cell interactions in each cell line. Additionally, considering the higher volume of data available for certain cell lines, including MDA-MB-231, MDA-MB-468, BT-474, and WM, it remains unclear whether the characterised hyperpolarisations occur in other cell lines.

5.2. Spatio-Temporal Analysis

To understand the behaviour of different cell lines over time and space, we used statistical and density-based clustering methods to highlight a connection between neighbouring cells and synchronised hyperpolarisation expression. Negative correlation for most cell lines in Figure 8, suggests that cells closer together tend to have stronger temporal correlations, with MDA-MB-231 and MDA-MB-468 being statistically significant.

Scatter plots in Figure 7 reveal a distinct cluster in MDA-MB-468 with an STTC of 0.5, indicating partial synchronisation between cell pairs. This unique behaviour is unique to the MDA-MB-468 cell line, where approximately half of the spikes synchronise within a specified time window. This can be observed in Figures 10c and 10d, where the partial synchronisations between cells may be attributed to the clustering of circulating tumour cells upon cell communication, con-

sistent with the adaptive nature observed in metastatic cancer [40].

As for in Figure 9c, the sharp rise in cluster numbers between approximately 1s and 10s suggests a high sensitivity for the different peak times which may be caused by the instantaneous GJ communications, calcium signalling or other long-distance intercellular interactions. However, beyond a certain distance, there is a plateau due to the merging of smaller clusters. Conversely, in Figure 9a and 9b, the number of clusters respectively increases and decreases to a plateau with increasing distance parameter (ϵ_1), suggesting that intercellular distances play a crucial and diminishing role in spike correlations, up to about 0.4 normalised distance.

Similarly, the MDA-MB-231 cell line demonstrated a comparable trend with a fixed time width and varying distance, albeit with a lower cluster count compared to MDA-MB-468. This difference suggests a potentially higher level of interconnectivity among MDA-MB-468 cells. Furthermore, with varying time widths, the decrease in cluster count was less pronounced than in MDA-MB-468, potentially indicating a similar involvement of calcium channels but to a lesser extent, as shown in Figures 10a and 10b.

Although both MDA-MB-468 and MDA-MB-231 cell lines demonstrate more significant results compared to other breast cancer cell lines, it is important to note that this may potentially be due to the insufficient data for other cell lines, as shown in Table 1. The unique phenomena observed in the MDA-MB-468 cell line may potentially occur in other cell lines with additional data. With insufficient data for some cell lines, it is difficult to suggest whether cell lines from the same cancer cell type exhibit similar behaviours.

5.3. Drawbacks of PCC for Spike Train Analysis

Two different correlation coefficient indexes have been utilised in order to compare cell lines and enable us to identify any patterns in time or space. However the assumptions and differences in the methods have lead to some disparity in results in the post-hoc and spatio-temporal analysis.

Although PCC is widely used for the analysis of spike trains in neuroscience, there are multiple disadvantages to this method. It disregards non-stationary activities like changing firing rates in the cell membrane potential. Additionally, it assumes linear pairwise relationships, neglecting potential nonlinear dynamics like synchronisation. Temporal resolution is limited as the PCC measures correlation between spike trains aver-

aged over a specific time window. Finally the analysis is very vulnerable to noise, leading to unintended correlations from the non-spiking regions.

Conversely, the STTC is considered a better alternative as it offers high temporal resolution by capturing precise spike timing between cell voltages within small time windows. These windows can be adjusted to study diverse cell interactions, including short-term and long-term correlations, non-linear dynamics and precise spike timing relationships. Unlike PCC, STTC does not assume stationarity, enabling it to capture dynamic changes in electrical activity patterns. Additionally, it is more robust against noise due to reliance on spike timings rather than binary spike trains.

5.4. Possible Biological Mechanisms Causing Hyperpolarisations

While valuable insights emerged from analysing the correlation between spatial location and voltage fluctuations, definitive links to biophysical mechanisms remain challenging without further laboratory experimentation. Due to the complexity and randomness of cancer, our analysis can only speculate on similar behaviours across different cell lines. Therefore, it is important to acknowledge the diversity in intercellular communication, as well as the influence of additional factors including hormonal influences and environmental factors. For instance, oestrogen has been shown to stimulate breast cancer stem cells, promoting cell differentiation [41].

In this analysis, hyperpolarisation is observed in a minority of cells, with simultaneous spike correlation occurring in 0.2% pair of cells as shown in Table 2. Potential causes include GJ communication, crucial for maintaining cell homeostasis and disruption could lead to impaired immune response [42]. Despite their minority status, these abnormal voltage fluctuations may significantly influence cancer cell propagation. Notably, even in cell lines with limited data, a small percentage of cells exhibit these distinctive behaviours, highlighting their potential role in cancer progression. Further fluorescent dye imaging with flow cytometry can determine and quantify GJ cell communication in all breast cancer cell lines, offering a potential therapeutic target [42].

6. Conclusion

6.1. Summary

This study investigates the dynamic voltage fluctuations of human breast cancer cells to understand the underlying biophysical mechanisms behind these bioelectric behaviours. Firstly, temporal analysis reveals significant correlations among various cancer cell lines, notably MDA-MB-231, MDA-MB-468 and WM, suggesting potential interactions mediated by mechanisms such as GJs. Moreover, the most aggressive phenotype, MDA-MB-231, exhibits higher percentages of synchronised events for both 1s and 10s time bins, indicating increased cell-to-cell communication, forming more clusters. Further spatio-temporal analysis reveals increased synchronisation among nearby cancer cells, particularly in cell lines MDA-MB-231 and MDA-MB-468. While these findings offer insights into understanding breast cancer cell behaviour, the functional implications of membrane potential dynamics in cancer progression and metastasis are still unclear. Further investigations are required to uncover the underlying biophysical mechanisms accurately. Given the complexity of cancer, understanding the interactions between V_m dynamics, intercellular communication, and ion channels could yield valuable insights into innovative biomarkers for diagnosis and treatment, highlighting the importance of ongoing research in this field.

6.2. Future Work

A potential alternative method for accurately obtaining spike trains could involve thresholding with the gradient (see Appendix F). Additionally, integrating calcium studies and dye imaging studies for GJs alongside spatio-temporal bioelectricity analysis is crucial to gain a better understanding of dynamic cellular interactions underlying cancer progression. Moreover, integrating genome sequencing would enable the identification of specific mutations, and highlight the genetic landscape associated with observed bioelectric behaviours across different cancer cell lines. In addition to breast cancer, other cancers could be analysed to further characterise these behaviours. By combining these analyses, correlations between observations and underlying genetic or biophysical causes can be established, offering valuable insights for future research and clinical applications, thus advancing precision oncology strategies.

Appendices

A. Statistical methods

A.1. Pearson's Correlation Coefficient (PCC)

For each FOV, the time series were converted into spike trains using a time binning approach, with bin sizes ranging from 1s to 100s. To evaluate the correlation between pairs of cells within a given FOV, PCC is calculated using Equation 2 on the pairs of spike trains.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \quad (2)$$

$r = 1$ indicates a positive correlation, where one cell consistently hyperpolarises within the same time bin when the other cell hyperpolarises. And $r = -1$, where one cell is inhibited from hyper-polarising within the same time bin when the other cell hyperpolarises.

A.2. Spike Time Tiling Coefficient (STTC)

The STTC [43] is a measure of synchrony between two spikes. To quantify the correlation of spike times between the same configuration of spike train pairs as the previous section, A and B, we consider the spikes from A that falls within $\pm\Delta t$ of each spikes in B and vice versa.

$$STTC = \frac{1}{2} \left(\frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A} \right) \quad (3)$$

where T_A and T_B is the the proportion of total recording time that lies within $\pm\Delta t$ of the spikes in A and B, respectively. P_A is the proportion of spike from A that lie between $\pm\Delta t$ of any spikes in B, and P_A is calculated similarly using spikes in B and $\pm\Delta t$ of any spikes in A. $STTC = 1$ means that the cells hyperpolarises within the time window in perfect synchrony, while $STTC = -1$ indicates anti-synchrony.

A.3. Two Sample Unpaired T-Test

The t-statistic T is calculated using the difference between the sample means, \bar{Y} , the sample sizes n , and the sample variances s^2 as shown below.

$$T = \frac{(\bar{Y} - \mu_0)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \quad (4)$$

With an alpha value of 0.05, where the null hypothesis is accepted when the obtained value is below this value.

B. DBSCAN and ST-DBSCAN

DBSCAN is a clustering algorithm based on the density of surrounding points [28], as opposed to k-means clustering which clusters solely on distance. This allows DBSCAN to form arbitrarily shaped clusters and makes it more robust against noise. For two dimensions, two parameters must be specified: the epsilon (ϵ) value (representing the radius of the shape drawn around each point) and the minimum number of points required to consider a point as 'core'.

Each point is characterised as either core, boundary or noise. A point is considered core if there is at least the minimum number of points within its radius, otherwise it is labelled as a boundary point. Points with no neighbouring points are categorised as noise points. Once all points have been to either category, the clustering stage begins.

For each core point, all other core points within the ϵ radius are assigned to the same cluster. Following this, the boundary points within the radius of a cluster are also assigned to that cluster. Points that are not assigned to any core point remain unclustered. See Figure 11 [44] for a flow diagram of each stage in the algorithm, performed on an example with a minimum of 3 points.

ST-DBSCAN, an extension of DBSCAN, introduces a temporal dimension, allowing clustering based on both spatial and temporal data [26]. This introduces another parameter; ϵ_2 , which represents the radius of the shape in the time dimension. ST-DBSCAN is thus capable of clustering points not only based on their spatial proximity but also considering their temporal relations; it enables the identification of spatio-temporal clusters, making ST-DBSCAN suitable for our research which looks at both attributes.

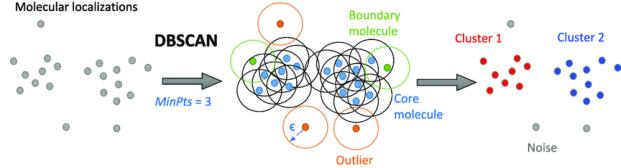


Figure 11: A flow diagram showing the stages of the DBSCAN algorithm. The diagram uses a minimum point of 3. [44]

C. Complete PCC and T-Test Results

The complete PCC and T-test results for the 10 cell lines are as follows in Figure 12 and 13.

D. P-Value of ANOVA and the KW Test

Regarding the quantitative outcomes of those two tests, both statistical tests reject the null hypotheses. For the PCC values, the KW test indicated a rejection of the null hypothesis with a p-value of 5.75×10^{-25} , while the ANOVA reported a p-value of approximately 7×10^{-150} . Furthermore, for the STTC values, the p-value for the KW test was 7×10^{-9} , and the ANOVA resulted in a p-value of about 4.99×10^{-56} .

E. Complete Correlation Coefficients Against Distance Plots

Figure 14 displays all the scatter plots for all breast cancer cell lines and for different time bins.

F. Spike Train Detection

The common method for spike train detection within neuroscience is utilising a threshold. Due to the nature of the cell membrane potential, where some cell lines have negative events lasting for long periods, the threshold will incorrectly register it as a spike. Since for our research we have used this method, it has led to occasional incorrect registration of hyperpolarisations impacting all of our statistical analysis.

An additional method that could be implemented is by setting a threshold on the first derivative, similar to the algorithm used to detect the QRS complex from ECG [45]. If the gradient is steep over some period of time, we can be certain that there is a hyperpolarisation spike. Therefore both methods could be used in conjunction to create the spike train.

G. Code Availability

Here is a link to our Github with the code we have used to perform the statistical analysis and plots for this report : [project repositories](#).

H. Project Management Assessment

Several setbacks occurred within our project, leading to schedule delays and deviation from our original plan. Initially, based on our background research, we identified various potential methods for exploring the spatial and temporal relationships within cell lines, including cross-correlation, dynamic time warping, spatial auto-correlation and neural networks. However, upon implementation of our dataset, it was evident that certain techniques were not suitable for our purposes. For instance, neural networks excel in classification tasks, which is unsuitable for hypothesis testing, necessitating a redirection of our efforts. Considerable time was invested in testing these analytical approaches, inadvertently delaying our project timeline. Additionally, the presence of concurrent coursework and other commitments among team members contributed to the slow pace of our data analysis. During this period, our teamwork faced a rigorous test, requiring us to redistribute tasks and prioritise project responsibilities within the constraints of our limited time frame. To mitigate the impact of these delays, we were compelled to adapt and explore alternative methodologies and promptly recalibrate subsequent project milestones to ensure continued progress within the revised timeline.

Overall, this adaptive approach shows our group's resilience and commitment to navigating challenges while maintaining momentum towards our project objectives.

From our project the three key project management lessons we learnt were:

1. Teamwork and effective communication

Our team has embraced these principles as fundamental to our project's success. Given the project's open-ended and extended duration, we maintained weekly meetings to collaborate on ideas and track progress. We assigned distinct roles and tasks to each team member in the beginning, leveraging their unique interests and expertise to optimise project efficiency. During our project, we encountered challenges in data analysis that transcended individual capabilities. These instances underscored the invaluable role of our weekly meetings for discussion, collaboration and ensuring everyone is aligned on project goals and timelines. Through collective brainstorming, we were able to confront these challenges, emphasising the potency of team synergy over individual endeavours.

2. Adaptability

This was critical in propelling our project forward. From our comprehensive background research, we have identified numerous methods for conducting temporal and spatial analysis of cell lines. However, not all methods align with our dataset's intricacies. Thus, approaching our data with a flexible and adaptable mindset proves pivotal. Additionally, recognising that each member juggles diverse commitments, occasionally lagging in progress, underscores the importance of flexibility and mutual support. There have been instances where setbacks, be it exams or challenges in data analysis, momentarily delayed our project timeline. Nevertheless, our team showed resilience by swiftly adapting and supporting one another to persevere through tough times.

3. Feedback and reflection

This was integral to the success of our project. We scheduled weekly meetings with our supervisor to share our progress and seek constructive feedback. Her insights and expertise not only inspired us but also provided invaluable guidance on navigating complex project challenges. Following each meeting, we analysed the feedback received, identifying areas for improvement and refinement, allowing us to continuously elevate the quality of our work and refine our methodologies. Moreover, regular reflection guided us to evaluate our progress against the overarching objectives of our project. This provided us with the opportunity to pause, assess, and re-calibrate our strategies as needed, ensuring that we remained firmly on course to achieve our project aims.

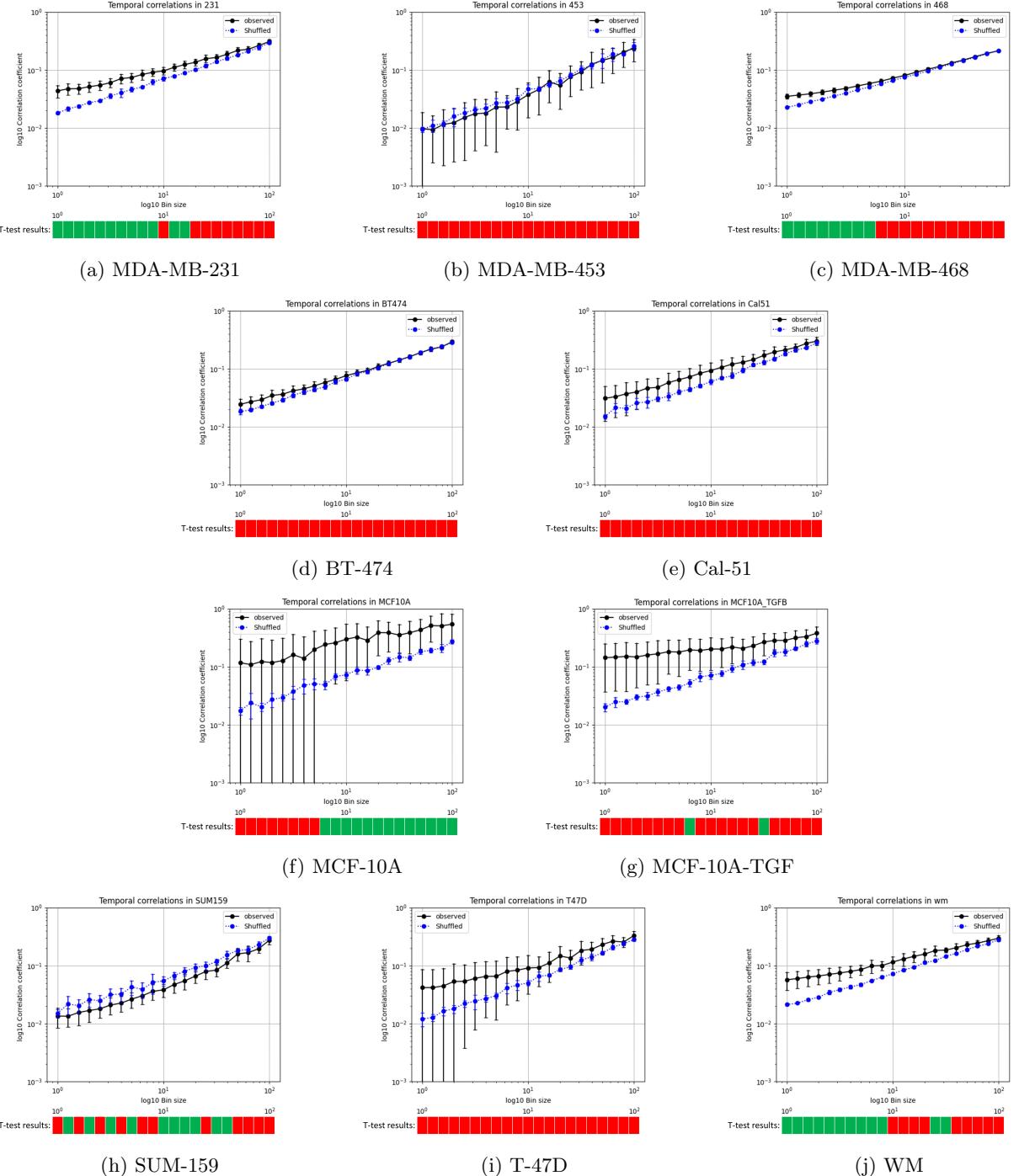
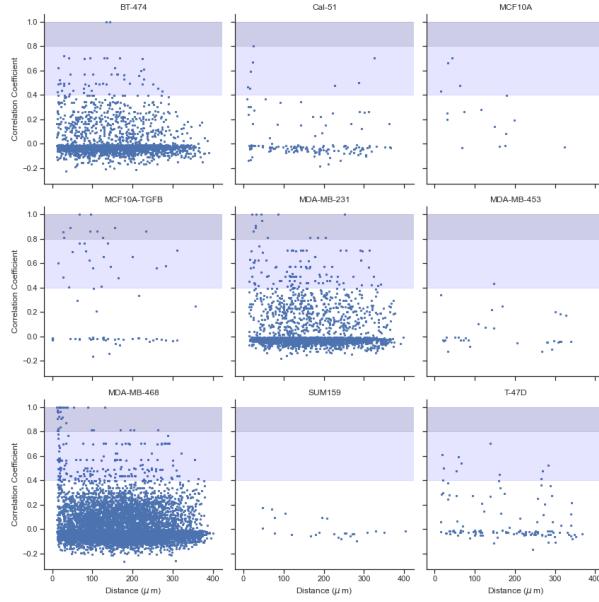


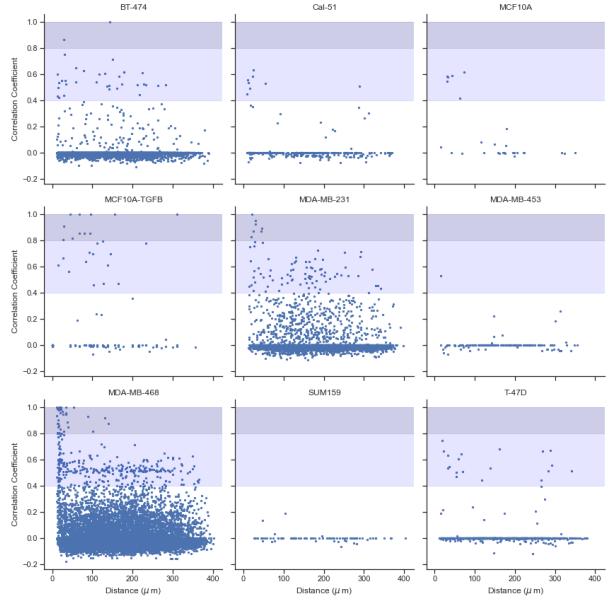
Figure 12: PCC values against bin size for all cell lines. The black line represents PCC values calculated using the observed time series, while the blue line corresponds to PCC values calculated from randomised data. T-tests were performed on bins in increments of \log_{10} , green markers indicate a p-value of $p < 0.05$, while red markers indicate a p-value of $p > 0.05$

Cell lines	Bin sizes (seconds)																				
	1.000	1.259	1.585	1.995	2.512	3.162	3.981	5.012	6.310	7.943	10.000	12.589	15.849	19.953	25.119	31.623	39.811	50.119	63.096	79.433	100.000
231	0.011	0.010	0.010	0.012	0.014	0.022	0.017	0.023	0.033	0.030	0.056	0.029	0.046	0.056	0.093	0.184	0.193	0.173	0.510	0.417	0.560
453	0.900	0.463	0.870	0.407	0.444	0.447	0.363	0.504	0.366	0.601	0.224	0.884	0.474	0.285	0.690	0.435	0.772	0.912	0.486	0.529	0.580
468	0.000	0.000	0.001	0.001	0.006	0.010	0.024	0.035	0.115	0.077	0.183	0.120	0.165	0.365	0.281	0.729	0.299	0.774	0.864	NaN	NaN
BT474	0.207	0.139	0.201	0.188	0.241	0.376	0.432	0.427	0.235	0.485	0.338	0.751	0.666	0.634	0.894	0.968	0.824	0.781	0.931	0.785	0.427
Cal51	0.260	0.415	0.322	0.361	0.219	0.269	0.225	0.221	0.204	0.172	0.206	0.174	0.097	0.186	0.291	0.155	0.276	0.412	0.241	0.584	0.584
MCF10A	0.126	0.146	0.120	0.139	0.127	0.087	0.177	0.058	0.008	0.016	0.014	0.009	0.009	0.002	0.002	0.006	0.005	0.006	0.001	0.014	0.009
MCF10A-TGFB	0.083	0.091	0.096	0.090	0.083	0.092	0.052	0.071	0.049	0.081	0.089	0.073	0.073	0.107	0.093	0.049	0.077	0.137	0.119	0.229	0.155
SUM159	0.587	0.041	0.239	0.040	0.085	0.041	0.056	0.017	0.189	0.109	0.023	0.022	0.019	0.017	0.132	0.007	0.011	0.297	0.463	0.068	0.370
T47D	0.348	0.356	0.373	0.353	0.377	0.338	0.309	0.360	0.363	0.377	0.331	0.435	0.311	0.201	0.253	0.314	0.266	0.164	0.205	0.658	0.246
wm	0.033	0.025	0.025	0.019	0.038	0.037	0.015	0.009	0.044	0.019	0.064	0.089	0.050	0.147	0.048	0.012	0.158	0.100	0.277	0.258	0.585

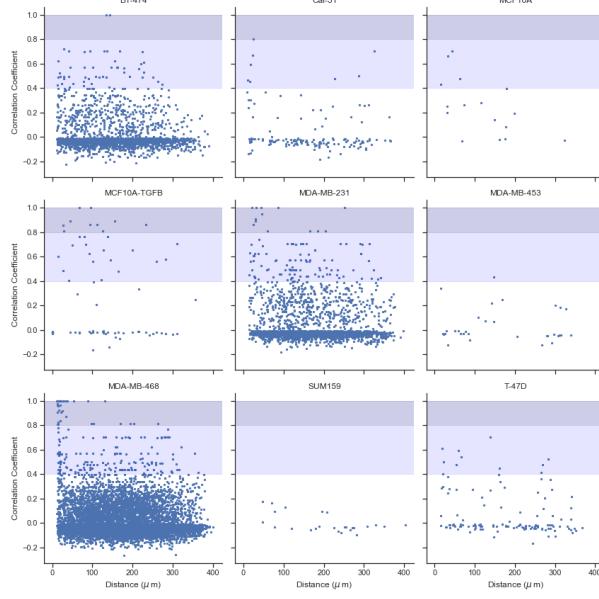
Figure 13: T-test results with p-values for all cell lines and time bin sizes from 1s to 100s. The colour indicate how low the p-values are; from red indicating a high p-value to green indicating a low p-value.



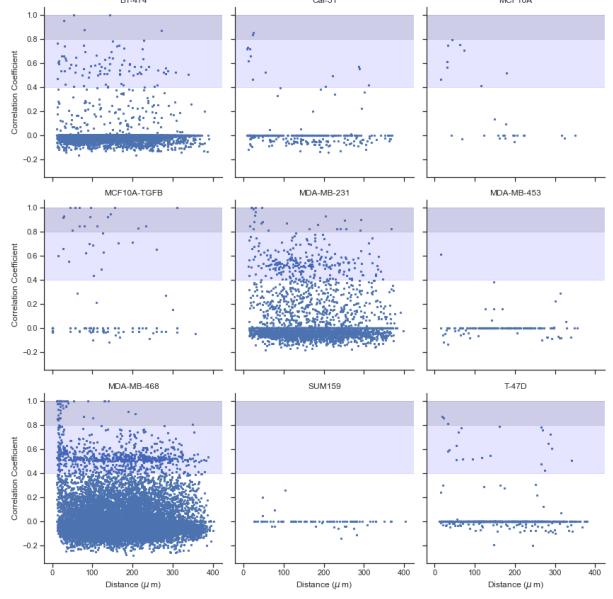
(a) PCC with 1s time bin against distance



(c) STTC with ±1s time window against distance



(b) PCC with 10s time bin against distance



(d) STTC with ±5s time window against distance

Figure 14: Scatter plots of the correlation coefficients against distance for different times for all cell lines. The blue shading represents values greater than 0.4, while the grey shading represents values greater than 0.8.

Nomenclature

ANOVA Analysis of Variance

DBSCAN Density-based Spatial Clustering of Applications with Noise

ER Oestrogen receptor

EV Extracellular vesicles

FOV Field of View

GJ Gap junction

HER2 Human epidermal growth factor receptor 2

KW test Kruskal-Wallis test

PCC Pearson's Correlation Coefficient

PR Progesterone receptor

ROI Region of Interest

STTC Spike time tiling coefficient

TM Tumour microtubules

TNBC Triple negative breast cancer

TNT Tunnel nanotubes

Glossary

Action Potential: Rapid sequence of changes in the voltage across a membrane.

Basal-Like Breast Cancer: An aggressive molecular subtype of breast cancer which do not express oestrogen receptor (ER) and progesterone receptor (PR) positive with the absence of human epidermal growth factor receptor 2 (HER2).

Basal A Tumour: Subtype of triple negative breast cancer that display epithelial characteristics and are associated with BRCA1 gene signatures.

Basal B Tumour: Subtype of triple negative breast cancer that display mesenchymal and stem/progenitor-like characteristics.

Bioelectrical: relating to electrical current generated by biological activity.

Cancer Heterogeneity: Existence of subpopulations of cells, with distinct genotypes and phenotypes that may harbour divergent biological behaviours, within a primary tumour and its metastases, or between tumours of the same histopathological subtype (intra- and inter-tumour, respectively). **Cell Line:** Cell cultures originating from one cell

Cellpose: A single-class instance segmentation algorithm which detects all cells of one type in an image.

Chemoresistance: The ability of cancer cells to evade or to cope with the presence of therapeutics.

Connexins: Gap junction proteins that assemble into intercellular channels connecting adjacent cells.

Depolarisation: Movement of a cell's membrane potential to a more positive value.

Epithelial to Mesenchymal Transition: Process by which epithelial cells lose polarity and cell-cell adhesion and gain migratory and invasive properties to become mesenchymal stem cells.

Extracellular Vesicles: Cell-derived membrane-surrounded vesicles that carry bioactive molecules and deliver them to recipient cells.

Field of Views: The viewable area that can be imaged by a lens system.

Gap Junctions: Membrane proteins between neighbouring cells that allow direct cytoplasmic exchange of ions and metabolites.

Gene Expression Profiling: Measures the expression level of mRNAs (transcripts) in a cell population at a certain time. A gene expression profile may be used to help diagnose a disease or condition, such as cancer.

G1 Phase: First growth phase of the cell between mitosis and initiation of replication of DNA.

Histological Tumour Grade: The appearance of a tumour cell under the microscope. They are assessed by comparing the appearance of tumour cells to normal cells and evaluating the degree of abnormality.

Hyperpolarisation: Movement of a cell's membrane potential to a more negative value.

IP₃ Receptor: Inositol trisphosphate receptor (InsP₃R) is a membrane glycoprotein complex acting as a Ca channel activated by inositol trisphosphate (InsP₃).

Luminal Tumours: Subtype of breast cancer defined as ER-positive and/or PR-positive tumors with negative HER2.

Melanoma: An aggressive type of skin cancer that originates from melanocytes (melanin-forming cells).

Membrane Excitability: A property of a cell, allowing it to respond to stimulation by rapid changes in membrane potential produced by ion fluxes across the plasma membrane.

Membrane Potential: Difference in voltage (or electrical potential) between the inside and outside of a cell, a weighted mean of the equilibrium potentials of the different permeant ions.

Metastasis: Spread of cancer from an initial or primary site to a different or secondary site within the host's body.

Niches: Anatomically distinct microenvironment within the overall tumour microenvironment.

Oncogenic Moieties: Fragment of molecules or signalling molecules/genes that can cause cancer.

Paracrine Signalling: A form of cell communication in which cells produce small molecules (signal) to induce changes in nearby cells, altering the behaviour of the nearby cells.

Phenotype: Observable traits associated with the rearrangement of genetic information and the altered expression of many gene products,

S Phase: Occur between the G1 and G2 phase of the cell cycle and is when DNA replication and much of the DNA repair activity in cells take place.

Spatial: Relating to, or involved in the perception of relationships (as of objects) in space.

Spatio-Temporal: Relating to both space and time.

Temporal: Relating to time.

TGF- β : Transforming Growth Factor-beta are growth factors that regulate cellular processes, including cell division, differentiation, motility, adhesion, and death.

Time Bins: The process of dividing time, a continuous set of data, into discrete divisions called bins.

Time Series: A sequence taken at successive equally spaced points in time.

Tunnel Nanotubes: Long and thin protrusions of the cytoskeleton and plasma membrane which connect two different cells.

Type-1 Error: An incorrectly rejected null hypothesis, also known as a false positive.

References

- [1] Quicke P, Sun Y, Foust AJ. Voltage imaging reveals the dynamic electrical signatures of human breast cancer cells. Communications Biology. 2022;5(1):1178. Available from: <https://doi.org/10.1038/s42003-022-04077-2>.
- [2] Yang M, Brackenbury WJ. Membrane potential and cancer progression. Frontiers in Physiology. 2013;4. Available from: <https://doi.org/10.3389/fphys.2013.00185>.
- [3] Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol. 2014;5(3):412-24. Available from: <https://doi.org/10.5306/wjco.v5.i3.412>.
- [4] Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. Journal of Clinical Investigation. 2011;121(7):2750–2767. Available from: <https://doi.org/10.1172/jci45014>.
- [5] Dai X, Cheng H, Bai Z, Li J. Breast Cancer Cell Line classification and its relevance with breast tumor subtyping. Journal of Cancer. 2017;8(16):3131–3141. Available from: <https://doi.org/10.7150/jca.18457>.
- [6] World Health Organization. Breast cancer; 2024. Available from: <https://www.who.int/news-room/detail/breast-cancer#:~:text=Scope%20of%20the%20problem,the%20world's%20most%20prevalent%20cancer>. [Accessed: 29th January 2024].
- [7] Dillekås H, Rogers MS, Straume O. Are 90 Cancer Medicine. 2019;8(12):5574-6. Available from: <https://doi.org/10.1002/cam4.2474>.
- [8] Hannah R, Joo E. Animal Physiology. Pressbooks; 2022.
- [9] Djamgoz MBA. Biophysics of cancer: Cellular excitability (“Celex”) hypothesis of metastasis. Journal of Clinical & Experimental Oncology. 2014;s1(01). Available from: <https://doi.org/10.4172/2324-9110.s1-005>.
- [10] Ko JH, Ko EA, Gu W, Inja Lim HB, Zhou T. Expression profiling of ion channel genes predicts clinical outcome in breast cancer. Molecular Cancer. 2013;12. Available from: <https://doi.org/10.1186/1476-4598-12-106>.

- [11] Djamgoz MBA, Fraser SP, Brackenbury WJ. In vivo evidence for voltage-gated sodium channel expression in carcinomas and potentiation of metastasis. *Cancers.* 2017;8(16):3131–3141. Available from: <https://doi.org/10.3390/cancers11111675>.
- [12] Ribeiro M, Elghajiji A, Fraser SP, Burke ZD, Tosh D, Djamgoz MBA, et al. Human breast cancer cells demonstrate electrical excitability. *Frontiers in Neuroscience.* 2020;14. Available from: <https://doi.org/10.3389/fnins.2020.00404>.
- [13] Payne SL, Ram P, Srinivasan DH, Le TT, Levin M, Oudin MJ. Potassium channel-driven bioelectric signalling regulates metastasis in triple-negative breast cancer. *eBioMedicine.* 2022;75. Available from: <https://doi.org/10.1016/j.ebiom.2021.103767>.
- [14] Chatterjee S, Bhat V, Berdnikov A, Liu J, Zhang G, Buchel E, et al. Paracrine Crosstalk between Fibroblasts and ER+ Breast Cancer Cells Creates an IL1 β -Enriched Niche that Promotes Tumor Growth. *iScience.* 2019;19:388-401. Available from: <https://doi.org/10.1016/j.isci.2019.07.034>.
- [15] Bortolin A, Neto E, Lamghari M. Calcium Signalling in Breast Cancer Associated Bone Pain. *International Journal of Molecular Sciences.* 2022;23(3). Available from: <https://www.mdpi.com/1422-0067/23/3/1902>.
- [16] Leybaert L, Sanderson MJ. Intercellular Ca²⁺ Waves: Mechanisms and Function. *Physiological Reviews.* 2012;92(3):1359–1392. Available from: <https://doi.org/10.1152/physrev.00029.2011>.
- [17] Ebrahim T, Ebrahim AS, Kandouz M. Diversity of Intercellular Communication Modes: A Cancer Biology Perspective. *Cells.* 2024;13(6). Available from: <https://www.mdpi.com/2073-4409/13/6/495>.
- [18] Banerjee D. Connexin's connection in breast cancer growth and progression. *International Journal of Cell Biology.* 2016;2016:1–11. Available from: <https://www.hindawi.com/journals/ijcb/2016/9025905/>.
- [19] Sinha G, Ferrer AI, Moore CA, Naaldijk Y, Rameshwar P. Gap Junctions and Breast Cancer Dormancy. *Trends in Cancer.* 2020;6(4):348-57. Available from: <https://www.sciencedirect.com/science/article/pii/S2405803320300297>.
- [20] Dominiak A, Chelstowska B, Olejarz W, Nowicka G. Communication in the Cancer Microenvironment as a Target for Therapeutic Interventions. *Cancers.* 2020;12(5). Available from: <https://doi.org/10.3390/cancers12051232>.
- [21] Ottonelli I, Caraffi R, Tosi G, Vandelli MA, Duskey JT, Ruozzi B. Tunneling Nanotubes: A New Target for Nanomedicine? *Int J Mol Sci.* 2022;23(4):2237. Available from: <https://doi.org/10.3390/ijms23042237>.
- [22] Kato K, Nguyen KT, Decker CW, Silkwood KH, Eck SM, Hernandez JB, et al. Tunneling nanotube formation promotes survival against 5-fluorouracil in MCF-7 breast cancer cells. *FEBS open bio.* 2022;12(1):203-10. Available from: <https://pubmed.ncbi.nlm.nih.gov/34738322/>.
- [23] Yan P, Acker CD, Zhou WL, Lee P, Bollensdorff C, Negrean A, et al. Palette of fluorinated voltage-sensitive hemicyanine dyes. *Proc Natl Acad Sci USA.* 2012. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3528613/>.
- [24] Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med.* 2018;18(3):91–93. Available from: <https://doi.org/10.1016/j.tjem.2018.08.001>.
- [25] Boulware MJ, Marchant JS. Timing in Cellular Ca²⁺ signaling. *Current Biology.* 2008;18(17). Available from: <https://doi.org/10.1016/j.cub.2008.07.018>.
- [26] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering.* 2007;60(1):208–21. Intelligent Data Mining. Available from: <https://www.sciencedirect.com/science/article/pii/S0169023X06000218>.
- [27] scikit learn. Sklearn.cluster.DBSCAN;. [Accessed: 7th April 2024]. Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>.
- [28] StatQuest with Josh Starmer. Clustering with DBSCAN, Clearly Explained!!!; 2022. Available from: <https://youtu.be/RDZUDRSD0ok?si=3zic8R9uUbS94Z2S>. [Accessed: 6th April 2024].
- [29] Cakmak E, Plank M, Calovi DS, Jordan A, Keim D. Spatio-Temporal Clustering Benchmark for Collective Animal Behavior. In: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility. HANIMOB '21. New York, NY, USA: Association for Computing Machinery; 2021. p. 5–8. Available from: <https://doi.org/10.1145/3486637.3489487>.

- [30] Zhou M, Zheng M, Zhou X, Tian S, Yang X, Ning Y, et al. The roles of connexins and gap junctions in the progression of cancer. *Cell Commun Signal.* 2023;21(1):8. Available from: <https://doi.org/10.1186/s12964-022-01009-9>.
- [31] Thompson EW, Paik S, Brünner N, Sommers CL, Zugmaier G, Clarke R, et al. Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. *J Cell Physiol.* 1992;150(3):534-44. Available from: <https://doi.org/10.1002/jcp.1041500314>.
- [32] Kazan JM, El-Saghir J, Saliba J, Shaito A, Jalaleddine N, El-Hajjar L, et al. Cx43 Expression Correlates with Breast Cancer Metastasis in MDA-MB-231 Cells In Vitro, In a Mouse Xenograft Model and in Human Breast Cancer Tissues. *Cancers (Basel).* 2019;11(4):460. Available from: <https://doi.org/10.3390/cancers11040460>.
- [33] Jones JC, Miceli AM, Chaudhry MM, Kaunitz CS, Jai MA, Pancho RN, et al. Glucose-limiting conditions induce an invasive population of MDA-MB-231 breast cancer cells with increased connexin 43 expression and membrane localization. *Journal of Cell Communication and Signaling.* 2021;15:223-36.
- [34] Isert L, Mehta A, Loiudice G, Oliva A, Roidl A, Merkel OM. An In Vitro Approach to Model EMT in Breast Cancer. *International Journal of Molecular Sciences Mol.* 2023;24(9):7757. Available from: <https://doi.org/10.3390/ijms24097757>.
- [35] Massagué J. TGF β in Cancer. *Cell.* 2012;134(2):215-30. Available from: <https://doi.org/10.1016/j.cell.2008.07.001>.
- [36] Yang Y, Ye WL, Zhang RN, He XS, Wang JR, Liu YX, et al. The Role of TGF- β Signaling Pathways in Cancer and Its Potential as a Therapeutic Target. *Evid Based Complement Alternat Med.* 2021. Available from: <https://doi.org/10.1155/2021/6675208>.
- [37] Qu Y, Han B, Yu Y, Yao W, Bose S, Karlan BY, et al. Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. *PLoS One.* 2015;6(10):7. Available from: <https://doi.org/10.1371/journal.pone.0131285>.
- [38] Ginsburg OM, Kim-Sing C, Foulkes WD, Ghadirian P, Lynch HT, Sun P, et al. BRCA1 and BRCA2 families and the risk of skin cancer. *Fam Cancer.* 2010;9(4):489-93. Available from: <https://doi.org/10.1007/s10689-010-9377-y>.
- [39] Su Y, Subedee A, Bloushtain-Qimron N, Savova V, Krzystanek M, Li L, et al. Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. *Cell Reports.* 2015;11(10):1549-63. Available from: <https://doi.org/10.1016/j.celrep.2015.05.011>.
- [40] Schuster E, Taftaf R, Reduzzi C, Albert MK, Romero-Calvo I, Liu H. Better together: circulating tumor cell clustering in metastatic cancer. *Trends Cancer.* 2021;7(11):1020-32. Available from: <https://doi.org/10.1016/j.trecan.2021.07.001>.
- [41] Rosen JM, Roarty K. Paracrine signaling in mammary gland development: what can we learn about intratumoral heterogeneity? *Breast Cancer Research.* 2014;16. Available from: <https://doi.org/10.1186/bcr3610>.
- [42] Warawdekar U. An Assay to Assess Gap Junction Communication in Cell Lines. *J Biomol Tech.* 2019;30(1):1-6. Available from: <https://doi.org/10.7171/jbt.19-3001-001>.
- [43] Cutts CS, Eglen SJ. Detecting Pairwise Correlations in Spike Trains: An Objective Comparison of Methods and Application to the Study of Retinal Waves. *J Neurosci.* 2014;34(43). Available from: <https://doi.org/10.1523/JNEUROSCI.2767-14.2014>.
- [44] Khater IM, Nabi IR, Hamarneh G. A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods. *Patterns.* 2020;1(3):100038. Available from: <https://www.sciencedirect.com/science/article/pii/S266638992030043X>.
- [45] Arzeno NM, Deng ZD, Poon CS. Analysis of first-derivative based QRS detection algorithms. *IEEE Trans Biomed Eng.* 2008. Available from: <https://doi.org/10.1109%2FTBME.2007.912658>.