

**An agent-based model
of stem and cancer cell interaction**

Marzieh A Tehrani

Supervisors: Dr Dawn C Walker

Dr Cecile M Perrault

May 2020

The University of Sheffield

Department of Computer Science

Through my undergraduate degree in BioMedical Engineering, I gained an appreciation of interdisciplinary research with its potential in computational biology and although I started this PhD with no experience in cell culture or Python programming and statistical knowledge limited to quoting the mean and standard deviation, I found the iterative process of research the biggest challenge. Whether conducting *in vitro* experiments or developing the computational model, the direct straight path was never an option and instead a layered spiral approach with many repetitions and minor alterations had to be taken. I learnt that there is so much more knowledge to be gained through a step-wise approach than a giant perfect leap aiming for the finish line!

Acknowledgements

I would like to express my deepest thanks to my supervisors, Dr Dawn Walker and Dr Cecile Perrault for giving me the opportunity to undertake this PhD. I will forever appreciate their invaluable supervision, both on a professional and personal level. The knowledge, experience and support they provided was key in tackling and breaking through problems that seemed otherwise unsolvable.

I am extremely grateful to Dr Mark Hepple and Dr Michael Croucher for their technical support. Our many discussions were not only insightful and interesting but gave me the confidence and knowledge needed to optimise and improve the quality of my written code. I would also like to thank Professor Gwendolen Reilly for kindly providing the cells used, without which this project would not have been possible.

I am very thankful to my friends, lab mates and colleagues, specially Dr Andre Castro, Dr Maryam Shariatzadeh, Dr Claudia Wittkowske, Dr Rob Owen, Dr Dhanak Gupta, Mrs Tugba Cebe and Mrs Julie Marshall for the moral support, countless discussions, memories and cherished time spent together. I would also like to thank members of the research groups, IMSB, CURE and ABM for the many thought-provoking and fascinating presentations in addition to their continuous feedback, critique and insightful comments and suggestions on my research.

And finally my greatest appreciation goes out to my dear parents and sister, Mrs Zahra Rezaee, Dr Hossein A Tehrani and Dr Razieh A Tehrani, to whom I dedicate this thesis to you. Thank you for your encouragement, support and patients. I could not have done it without you!

Abstract

Advancements in tissue engineering combined with the disease seeking nature of stem cells have provided new grounds for targeted therapy of cancer. However the discrepancies found in existing literature on the role of un-modified stem cells at tumour sites (Klopp et al. 2011), indicates the need for further research.

In vitro approaches provide an insight into actual cell behaviour under given conditions. However these methods are limited by factors such as cost, time and technological advancements in available protocols. *In silico* tools provide means for quantitative analysis of accumulated data in addition to exploring scenarios and queries otherwise impossible to create in the lab. However these tools can lack in accuracy and realistic correlation with actual biological behaviour. The combination of both *in vitro* and *in silico* methods results in a powerful tool that compensates for the limitations of both approaches.

An agent-based model (ABM) is a bottom-up approach that uses information regarding cell behaviour at the single cell level to generate emergent cell population results. Through the development of an agent-based model, the resulting effects of known and hypothesised rules regarding individual cell characteristics and cell-to-cell interactions will be simulated. Where possible, the model rules will be informed and the final model predictions validated using results and observations obtained from cell culture experiments run simultaneously, allowing for a one-to-one mapping of *in vitro* and *in silico* results.

Computational modelling coupled with cell culture experiments will provide an insight into the mechanisms behind stem cell and cancer cell interactions, taking us one step closer to using stem cells as a method of cancer treatment.

List of Contents

| | |
|--|------------|
| ACKNOWLEDGEMENTS | III |
| ABSTRACT | IV |
| LIST OF CONTENTS | V |
| LIST OF FIGURES | XI |
| LIST OF TABLES | 1 |
| CHAPTER 1 | 1 |
| INTRODUCTION | 1 |
| 1.1 Motivation | 1 |
| 1.2 Aims and Objectives | 2 |
| CHAPTER 2 | 4 |
| BACKGROUND AND LITERATURE REVIEW | 4 |
| 2.1 Biological background and literature review | 4 |
| 2.1.1 Cancer | 4 |
| 2.1.2 Current treatments | 6 |
| 2.1.2.1 Surgery | 6 |
| 2.1.2.2 Radiotherapy | 7 |
| 2.1.2.3 Chemotherapy | 7 |
| 2.1.2.4 Gene therapy | 8 |

| | | |
|------------------------------|--|-----------|
| 2.1.2.5 | Immunotherapy | 8 |
| 2.1.2.6 | Targeted therapy | 8 |
| 2.1.3 | Stem cell therapy of cancer | 9 |
| 2.1.4 | Stem cells as vehicles for targeted therapy | 12 |
| 2.1.5 | The effect of unmodified stem cells on tumour | 13 |
| 2.1.5.1 | Tumour promoting | 16 |
| 2.1.5.2 | Tumour inhibiting | 17 |
| 2.1.6 | Possible causes of discrepancies in the literature | 19 |
| 2.2 | Modelling background and literature review | 20 |
| 2.2.1 | Contrasting approaches to modelling | 20 |
| 2.2.1.1 | Bottom-up versus Top-down | 21 |
| 2.2.1.2 | Discrete versus Continuum | 23 |
| 2.2.2 | Agent-Based Modelling (ABM) | 24 |
| 2.2.2.1 | Advantages and limitations | 25 |
| 2.2.2.2 | Components of a biological ABM | 27 |
| 2.2.2.3 | Tools and methods for ABM implementation | 28 |
| 2.2.2.4 | Techniques for model verification and validation (V & V) | 32 |
| 2.2.2.5 | Examples of ABMs of biological tissue | 34 |
| 2.3 | Summary | 39 |
| CHAPTER 3 | | 40 |
| MATERIALS AND METHODS | | 40 |
| 3.1 | Materials | 41 |
| 3.2 | Experimental Methods | 44 |
| 3.2.1 | General cell culture | 44 |
| 3.2.1.1 | General cell culture conditions | 44 |
| 3.2.1.2 | MG-63 cell line | 45 |
| 3.2.1.3 | hES-MP cells | 45 |
| 3.2.2 | Co-culture | 46 |
| 3.2.2.1 | General co-culture conditions | 46 |
| 3.2.2.2 | Direct co-culture | 46 |
| 3.2.2.3 | Indirect co-culture | 46 |
| 3.2.2.4 | Seeding density and ratios | 47 |
| 3.2.3 | Cell staining and microscopic imaging | 48 |
| 3.2.3.1 | Hoechst 33342 | 48 |
| 3.2.3.2 | CellTracker | 49 |

| | | |
|------------------|--|-----------|
| 3.2.3.3 | Time-lapse | 51 |
| 3.2.4 | Cell viability | 52 |
| 3.2.4.1 | Resazurin reduction assay | 52 |
| 3.2.4.2 | Manual cell counting protocol | 53 |
| 3.2.5 | Statistical analysis of biological data | 55 |
| 3.3 | Computational Methods | 56 |
| 3.3.1 | Modelling process | 56 |
| 3.3.2 | Domain model | 57 |
| 3.3.2.1 | Analysis of single cell data | 57 |
| 3.3.2.2 | Analysis of multi-cell data | 58 |
| 3.3.3 | Model design and development | 58 |
| 3.3.3.1 | Model implementation | 58 |
| 3.3.3.2 | Model verification and testing | 59 |
| 3.3.3.3 | Simulation output | 59 |
| 3.3.4 | Model exploration | 60 |
| 3.3.4.1 | Aleatory uncertainty analysis | 60 |
| 3.3.4.2 | Epistemic uncertainty analysis (robustness analysis) | 61 |
| 3.3.4.3 | Global sensitivity analysis | 61 |
| 3.3.4.4 | Calibration | 62 |
| 3.3.4.5 | Model validation | 63 |
| CHAPTER 4 | | 64 |
| | MONO-CULTURE EXPERIMENTS | 64 |
| 4.1 | Single cell analysis | 65 |
| 4.1.1 | Cell shape and size | 65 |
| 4.1.2 | Time-lapse live cell imaging | 70 |
| 4.1.2.1 | Cell tracking | 70 |
| 4.1.2.2 | Analysis of cell displacement | 71 |
| 4.1.2.3 | Probability distribution model of displacement | 74 |
| 4.1.2.4 | Analysis of cell directionality | 75 |
| 4.1.2.5 | Relationship between displacement and turning angle | 78 |
| 4.1.2.6 | Lineage tracking | 80 |
| 4.1.2.7 | Death rate | 81 |
| 4.1.2.8 | Multi-nucleated cancer cells | 81 |
| 4.2 | Multi-cell analysis (population level analysis) | 82 |
| 4.2.1 | Rate of growth and seeding density | 82 |
| 4.2.2 | Mitotic rate model | 86 |

| | | |
|------------------|---|------------|
| 4.3 | Summary | 87 |
| | | |
| CHAPTER 5 | | 89 |
| | | |
| | AGENT-BASED ‘NULL’ MODEL | 89 |
| | | |
| 5.1 | Model development | 89 |
| 5.1.1 | CellABM model outline | 90 |
| 5.1.2 | The Agents | 92 |
| 5.1.2.1 | Cell shape | 93 |
| 5.1.2.2 | Cell Motility | 93 |
| 5.1.2.3 | Cell growth | 95 |
| 5.1.2.4 | Cell death | 96 |
| 5.1.2.5 | Cell to cell contact | 97 |
| 5.1.2.6 | Cancer clusters | 97 |
| 5.1.3 | The environment | 98 |
| 5.1.3.1 | Model initialisation | 98 |
| 5.1.3.2 | The grid | 98 |
| 5.1.3.3 | Overlap-correction model | 99 |
| 5.1.4 | Model parameters | 101 |
| 5.1.4.1 | Local parameters | 101 |
| 5.1.4.2 | External parameters | 102 |
| 5.2 | Model calibration and validation | 104 |
| 5.2.1 | Verification and calibration of OCM | 104 |
| 5.2.1.1 | Verification of optimization | 104 |
| 5.2.1.2 | Calibration | 105 |
| 5.2.2 | Aleatory uncertainty analysis | 106 |
| 5.2.2.1 | Preliminary uncertainty analysis | 106 |
| 5.2.2.2 | Uncertainty analysis with Vargha-Delaney A-Test | 107 |
| 5.2.2.3 | Effect of model scale on aleatory uncertainty | 109 |
| 5.2.3 | Calibration against experimental data | 110 |
| 5.2.3.1 | Cancer cells | 110 |
| 5.2.3.2 | Stem cells | 111 |
| 5.2.4 | Validation against experimental data | 112 |
| 5.2.4.1 | Face validity | 112 |
| 5.2.4.2 | Statistical analysis | 113 |
| 5.2.5 | Local sensitivity analysis | 115 |
| 5.2.6 | Global sensitivity analysis | 116 |
| 5.3 | Summary | 118 |

| | |
|--|------------|
| CHAPTER 6 | 120 |
| CO-CULTURE ANALYSIS | 120 |
| 6.1 Co-culture <i>in vitro</i> experiments | 120 |
| 6.1.1 Methods | 121 |
| 6.1.2 Results | 124 |
| 6.1.2.1 Microscopic co-culture images | 124 |
| 6.1.2.2 Verification of day 1 ratios | 125 |
| 6.1.2.3 Comparison of growth | 126 |
| 6.1.2.4 Statistical analysis of day 4 results | 128 |
| 6.2 Co-culture <i>in silico</i> | 129 |
| 6.3 Summary | 130 |
| CHAPTER 7 | 131 |
| DISCUSSION AND CONCLUSIONS | 131 |
| 7.1 Single cell analysis | 132 |
| 7.2 Null model | 134 |
| 7.3 Co-culture analysis | 137 |
| CHAPTER 8 | 140 |
| APPENDIX | 140 |
| 8.1 Statistical tests and sources | 141 |
| 8.2 Results of cell nuclei size and shape study | 142 |
| 8.3 Cell track diagrams | 143 |
| 8.4 Linear regression models of cell growth | 149 |
| 8.5 Aleatory uncertainty analysis results | 151 |
| 8.6 Model calibration process heatmaps | 156 |
| 8.7 Local sensitivity analysis results | 160 |

| | | |
|------------|---|------------|
| 8.8 | Global sensitivity analysis results | 166 |
| 8.9 | Global sensitivity analysis parameters | 173 |
| | REFERENCES | 174 |

List of Figures

Figure 2.1: (Left) Original six hallmarks proposed in 2000 and (right) the two new emerging hallmarks and enabling characteristics. These are used to break down the complexity of cancer into eight rational underlying biological capabilities (hallmarks) and two enabling characteristics, which come together in various combinations resulting in the multistep development of the many different types of tumours. [Images reproduced from (Hanahan and Weinberg, 2011) copyright clearance obtained via RightsLink order number: 4945050185156]

6

Figure 2.2: Classical hierarchical model of stem cell differentiation. HSC = haemopoietic stem cells, MSC = mesenchymal stem cells & NSC = neural stem cells. [Image reproduced from (Corsten & Shah 2008) copyright clearance obtained via RightsLink order number: 4945461063966]

10

*Figure 2.3: Potential transgene strategies for stem cell therapy of cancer. MMP2=matrix metalloproteinase 2. TRAIL=Tumour necrosis factor related apoptosis inducing ligand * TNF = tumour necrosis factor. [Image reproduced from (Corsten & Shah 2008) copyright clearance obtained via RightsLink order number: 4945461063966]*

12

Figure 2.4: The human cell cycle. The M phase refers to the cell division via mitosis and the G₁, S and G₂ phases are all classed as the interphase stage of the cell in which the cell grows and replicates its content ready for mitosis. The G₀ phase is technically outside of the cell cycle and is also referred to as the resting phase or quiescent stage. It includes cells that are no longer dividing.

16

Figure 2.5: The many spatial (top) and temporal (bottom) scales of the human body. [Reproduced from (Hunter, Robbins and Noble, 2002) copyright clearance obtained via RightsLink order number: 4946181068994]

21

Figure 2.6: At what level should modelling begin? Diagram to show how models can start at any scale and eventually 'reach' up, down or across to include other scales. [Image reproduced from (Noble 2002) copyright clearance obtained via RightsLink order number: 4946180268061] 22

Figure 2.7: Topologies for agent relationships and social interaction. [Image reproduced from (Macal and North, 2010) copyright clearance obtained via RightsLink order number: 4946500539234] 31

Figure 3.1: Diagram showing the experimental set up across all seeding conditions for seeding ratio configuration of 1:2.. MG-63 cells are depicted as red squares and hES-MPs as green circles. The PDMS barrier is depicted using a white line. All well plates are shown from a birds eye view. 48

Figure 3.2: Molecular structure and fluorescence spectra of the three stains used. Top: Hoechst 33342, middle: CellTracker Red CMTPX and bottom: CellTracker Green CMFDA. [Image taken from www.thermofisher.com]. 50

Figure 3.3: CoSMoS process of model development presented by Andrews et al., (2010). The process is centred around the research context which covers the overall scientific question, motivation and requirements for validation. The domain is the scientific field under study and the domain model involves scientific understanding of the domain prior to any simulation. The platform model is where the domain model is used to design and implement the model for the simulation platform which in turn is the encoding and building of the model. The result model is where the simulation output is explored, validated and compared against the initial research context. As shown in the diagram, the research context can be referred to at any stage of the modelling process (Andrews et al., 2010). 56

Figure 3.4: The ABM work flow designed and followed showing the integration of in vitro experimental data with in silico modelling approaches in order to both inform and validate the model. Cells are modelled as agents and the model design and implementation is carried out strictly at the single cell level whereas model validation is done at the multi-cell level. Two fold validation is done by coupling experimental one to one validation with predictive validation. 57

Figure 4.1: MG-63 cells (passage 10) seeded at 10, 000 cells per cm^2 and imaged on day 1, 4 and 7. The first column shows images taken under brightfield, the middle column, blue fluorescence to detect the cell nuclei stained with Hoechst and the final column is the merged image combining the first two. At high confluency, the MG-63 cells are no longer in single mono-layers but appear homogenously spread. The use of a nuclear stain allows for more accurate cell detection, specially at high confluency. 66

Figure 4.2: hES-MP cells (passage 9) seeded at 10, 000 cells per cm^2 and imaged on day 1, 4 and 7. The first column shows images taken under brightfield, the middle column, blue fluorescence to detect the cell nuclei stained with Hoechst and the final column is the merged image combining the first two. hES-MP cells appear elliptical in shape and at high confluency, align along their major axis. 67

Figure 4.3: Diagrams depicting eccentricity. Left, eccentricity of an elliptical shape is calculated as the ratio of the distance from the centre to the foci (a) and the distance from the centre to the edge along the shape's major axis (b). Therefore eccentricity, $e = a/b$. Right, from CellProfiler3 documentation, shows the relationship between eccentricity value (e) and object shape. 67

Figure 4.4: Comparison between cell nuclei eccentricity of hES-MP and MG-63 cells at day1, 4 and 7. Boxplot whiskers include fliers and show full range of values. hES-MP cells are skewed towards $e = 1$, indicating a more elliptical shape which increase as the cells grow (increase in confluency). Pair-wise comparison across all pairings showed significant differences ($p < 0.001$ between all pairings except day 1 and day 7 of MG-63 cells at $p < 0.05$) $n=12$. 68

Figure 4.5: Comparison between cell nuclei surface area of hES-MP and MG-63 cells at day1, 4 and 7. hES-MP cells were found to be significantly smaller in area to MG-63 cells ($p < 0.001$). Both cells show a statistically significant ($p < 0.001$) reduce in surface area as they continue to grow. $n=12$. 68

Figure 4.6: Performance of CellProfiler-IdentifyPrimaryObject module used to identify cell nuclei applied to images of MG-63 and hES-MP cells on day 7 at high confluency. In order to avoid cells clumps measured as single cells, boundary conditions were applied which would have contributed to the deviation from normal distribution of the data. 69

Figure 4.7: Examples of individual cell trajectory for hES-MP (left) and MG-63 (right) cells seeded at 10, 000 cells/cm². The start point is marked with a red '0' and endpoint or key events such as mitosis are marked with red numbers at the time they occur. Each black dot represents the cells position at a single time-point. MG-63 cells were highly proliferative, undergoing mitosis two times on average over the 42hours. Stem cell mitosis was observed at a very low rate (an average of 1 cell out of 100). The first generation cell is marked in black, 2nd generation in blue and 3rd generation in green. Over the same period of time, the hES-MPs covered a much larger displacement from start to final position in comparison to the MG-63 cells comparable by the range displayed on the axis. 71

Figure 4.8: Probability distributions of all 6 datasets, top: hES-MP and bottom, MG-63 cells. The hES-MP cells produced a larger range of displacements (0 – 80 μ m) across all seeding densities in comparison to the MG-63 cells (0 – 20 μ m). In all cases the distribution of displacements were non-parametric with a strong skew towards the lower displacements. A significant difference between all seeding density pairings of each cell were found (hES-MP medium and high $p < 0.05$ and all other pairings $p < 0.001$). 72

Figure 4.9: Top: shows the relationship between displacement and density. A weak negative correlation was found for MG-63 cells: $r = -0.0136$, $p < 0.001$ whereas for hES-MP cells, a weak positive correlation was observed: $r = 0.0521$, $p < 0.001$. Bottom: the relationship between displacement and time. Both MG-63 and hES-MP cells were found to have a weak negative correlation (MG-63 cells: $r = -0.237$, $p < 0.001$ and hES-MP cells: $r = -0.228$, $p < 0.001$). The correlation between displacement and time is stronger than the correlation between displacement and density. The sudden drop in displacement at time = 48hours in hES-MP cells seen, coincides with the exact time at which the media was changed. 73

Figure 4.10: The probability distribution of displacing for MG-63 (top) and hES-MP (bottom) cells was fitted to existing continuous probability functions, Weibull, Gamma and Lognormal. Both Man-Whitney U test and Wilcoxon rank sum tests found the lognormal fitted model statistically similar to the raw data for both cell types. The fitted lognormal model was used to generate random cell displacements in the final ABM. 74

Figure 4.11: A depiction of the definition of cell direction and turning angle: direction is the angle made by the cell velocity vector with respect to the x-axis whereas turning_angle is the angle between the cell's velocity vectors from one time frame to the next 75

Figure 4.12: Polar histogram of the 'direction' and 'turning angles' with 20 angular bins. No significant difference (Chi square = 1.04, $p = 0.60$) was found among the three seeding conditions for MG-63 direction angle and hES-MP turning angle. However for MG-63 turning angle and hES-MP direction angles, a significant difference was found, $p < 0.001$ for both. Man-Whitney test for MG-63 turning angles found that the high seeding density was significantly different to the low and medium densities ($p < 0.001$) and for hES-MP direction angles, significant difference was found across all pairings ($p < 0.001$). 77

Figure 4.13: The probability density function of a von Mises distribution for $-\pi < \theta < \pi$ centred on $\theta = 0$. For the smallest κ values (*approximately $\kappa = 0$) the distribution is very close to uniform and as κ is increased, the distribution becomes increasingly concentrated around the centre (0 here). 78

Figure 4.14: Top: the kernel density estimation (KDE) for the joint probability distribution of cell displacement and turning angle. Bottom two rows show the scatter plots or the (from left to right) raw experimental data, the data generated by $\text{vonMises}(\kappa = \text{displacement})$ and $\text{vonMises}(\kappa = \text{displacement} - \text{shift})$ respectively. For hES-MP cells the displacement shift was found to be $5 \mu\text{M}$ and for MG-63 cells $7 \mu\text{M}$. The final model using $\text{vonMises} + \text{displacement shift}$, is used to draw random turning angles for the cells in the final ABM. 79

Figure 4.15: An example of cell lineage tracking shown for 10 individual MG-63 cells over 2 days. Full set of track diagrams are included in appendix (8.3). Changes in colour represent change in cell generation. Circular markers are used to label start point, end point and key events. Open ended tracks indicate the cell migrated out of frame. Measuring the time elapsed between one mitosis event to the next can be used as an estimate of cell cycle duration. 80

Figure 4.16: Images from time-lapse analysis taken 10hours apart, showing the same cell (marked with arrow) that appears to have 4 nuclei and could be considered a cluster. 81

Figure 4.17: Growth curves for hES-MP cells (left) and MG-63 cells (right) seeding at four different initial densities. Cell numbers were counted every 24hours over 7 days and averaged from 4 well repeats and 3 experimental replicates. As evident by the range of cell density shown on each of the y-axis, the cancer cells were found to grow at a significantly higher rate than the stem cells. 84

Figure 4.18: Day 4 and day 7 cell density plotted as bar charts to contrast the final cell numbers for the same initial seeding number per cell type. MG-63 cell numbers (in blue) were found to be a minimum of 3.5 times larger than the hES-MP cells for the same seeding density (5 times larger in the 20,000 cells/cm² experiment). 84

Figure 4.19: Cell density on day 1-2 for both cells. MG-63 cell numbers are closer to the numbers of cells seeded, whereas the hES-MP cell numbers fall below, suggesting a lower attachment rate for hES-MP cells or a longer lag phase. 84

Figure 4.20: Cell growth modelled using linear regression with Box-Cox transformation ($\lambda = 0.3$). The slopes of the regression models, representing the increase in cell number per timepoint, are very similar across all conditions for MG-63 cells and only the smallest three seeding densities for the hES-MP cells. 85

Figure 4.21: The mitotic rate regression models for the relationship between cell density and percentage of mitotic cells in a 24 hour window. To the left of each model the residual plots for homoscedasticity and normality are included. These models were implemented in the ABM to simulate cell growth. 86

Figure 5.1: CellABM class diagram. There are two main classes, Environment and Cell which have a composition relationship. Each cell type, Cancer_cell (cc) and Stem_cell (sc) are inherited from the parent Cell class. Cancer_clusters have an aggregation relationship with Cancer_cells, where Cancer_cell instances can join or leave Cancer_clusters. 90

Figure 5.2: State and activity diagram depicting an overview of the CellABM. After model initialisation, the model is executed for each time step. At each time step the agent rules are applied asynchronously for each agent in the model. Overlap correction model is used to physically correct any unwanted overlapping of cells. 91

Figure 5.3: Diagram showing the cell motility rules implemented in CellABM. The displacement is randomly generated from a lognormal distribution and after applying boundary conditions it is used to randomly generate the turning angle from a von Mises distribution. 94

Figure 5.4: Diagram showing the cell growth rules implemented in CellABM. 96

Figure 5.5: Diagram showing the implementation of cell death in CellABM. 96

Figure 5.6: Diagram showing the methods implemented for 'null' model cell to cell contact check. 97

Figure 5.7: ERD showing the relationship between various elements of the ABM and the flow and storage of data. 98

Figure 5.8: The 'local' (left) and 'near'(right) grid formations. Local is used for cell to cell contact and correction of overlap, near is used for detection of overlapping pairs. Both allow for a significantly faster execution of rules involving cell to cell interactions. 99

Figure 5.9 Diagram showing the OCM process. 100

Figure 5.10: Left - bar chart showing the OCM execution time with and without the implementation of the grid. With the implementation of the grid, the execution time of the OCM decreased by 98.5%. Right – The number of overlapping cell pairs at each OCM iteration for both models used as verification that the model optimization did not alter the model output. The results are the average of 100 simulations with the random number generator seed fixed across both models. Errors shown are the standard deviation. 104

Figure 5.11: Top (bar chart) – The effect of alterations in *SD_CONSTANT* and *PERMITTED_OL* values, on the *MAX_OCM* required to solve the OCM. *SD-OL* pairing at 0.25 *SD* and -5*OL* resulted in the smallest *MAX_OCM* and was subsequently selected. Bottom – the behaviour of the OCM at the selected parameters, showing the changes in the number of overlapping pairs and amount of overlap with each iteration of the model. This was used to determine the *MAX_OCM*. 106

Figure 5.12: A visualisation of the effect of sample size on coefficient of variation and 95% confidence interval for cell number output at a fixed time step of the model. Each sample size plotted was selected randomly from 500 model simulations (at $n = 500$, the entire sample pool was used). 107

Figure 5.13: The maximum A-test score out of all 20 sets per sample size. Each sample size represents the number of repeat runs of the simulation from which the median responses are used in statistical comparison using the Vargha-Delaney A-Test. Horizontal lines labelled mark the magnitude of effect with a score of 0.5 exactly indicating no effect. Scores below 0.5 were assigned corresponding values above 0.5 as the focus of this analysis was

magnitude of effect only and not its direction. It was concluded that a sample size of 50 for mono-culture models, is sufficient whereas a sample size of 150 is required for co-culture models. 108

Figure 5.14: Maximum A-test scores for various model scales increasing in agent number while keeping the ratio of agents to environment surface area fixed. Increasing the number of agents, produces more statistically consistent outputs. 109

Figure 5.15: An example of the visualisation technique used in model calibration. Here the sum of all RMSE per scenario is displayed in a heatmap (left) and a filter is applied to identify the scenarios falling in the lower quartile (right). These scenarios were further analysed using ANCOVA (with time as the covariate). 111

Figure 5.16: Face validation of the ABM simulation outputs of MG-63 growth on day 4 (equivalent to 288 model time-steps) shown for each four different seeding densities. 112

Figure 5.17: Face validation of the ABM simulation outputs of hES-MP growth on day 5 (equivalent to 288 model time-steps) shown for four different seeding densities. 113

*Figure 5.18: Validation of the ABM for cancer cell mono-culture growth against four separate experimental data sets. The model was found statistically similar ($K-S p > 0.05$ and $\delta < 0.8$) for time points marked with * and hence statistically valid for medium to low seeding densities up to the 4 day time point equivalent to 288 model iterations. ** marks the dataset the model was calibrated against.* 114

*Figure 5.19: Validation of the ABM for stem cell mono-culture growth against four separate experimental data sets. The model was found statistically similar ($K-S p > 0.05$ and $\delta < 0.8$) for time points marked with * and hence statistically valid for medium to low seeding densities up to the 5 day time point equivalent to 288 model iterations. ** marks the dataset the model was calibrated against.* 114

Figure 5.20: Changes in the distribution of contact related outputs with changes in input parameter, gridsize. It was found that reducing the gridsize by more than 20% causes a relatively large change in all contact related output measures. However, the A-test score was found to be less than the medium effect mark. 116

Figure 5.21: Changes in the distribution of contact related outputs with changes in input parameter, *nuc_percent*. It was found that increasing the percentage of cell volume occupied by the nucleus, reduces the frequency of cell to cell contact and number of simultaneous contacts per cell, however it has very little effect on duration of contact. The effect size of the largest changes in parameter measured by A-test score are still less than the medium effect mark . 116

Figure 5.22: The first order (S_1) and total order (S_T) indices for all 31 model parameters for each of the six key model output measures. Also included is the difference between S_T and S_1 which indicates the amount of total interaction each parameter contributes to. 118

Figure 6.1: light microscope images (at 4x magnification) of the PDMS barrier placed at the centre of the plate with MG-63 cells seeded to the left and hES-MP cells on the right 121

Figure 6.2: Examples of fluorescent images of hES-MP and MG-63 cells in co-culture during preliminary experimental work. (a) MG-63 cells in indirect co-culture stained with CellTracker green. (b) hES-MP cells in indirect co-culture stained with CellTracker red. (c) Direct co-culture of MG-63 cells labelled with CellTracker red and hES-MP cells unlabelled. Both cells nuclei are stained with Hoechst. This was the final staining methodology selected. (d) Direct co-culture of hES-MP (red) and MG-63 (green). (e and f) Direct co-culture of Mg-63 (green) seeded in an island formation before addition of hES-MP (red) 4 and 48 hours after seeding respectively. 124

Figure 6.3: Results showing initial cell numbers at time 0 (hES-MP only, 24hours and 48hours post addition of MG-63 cells. Expected cell numbers are drawn with dotted lines per seeding ratio and solid lines represent actual cell numbers. MG-63 cell numbers on day 1 are close to the expected value whereas hES-MP cells were found to be much higher at high initial seeding densities (e.g. ratio 1:4 with MG-63 cells taking up only 20% of the total cells seeded) and/or when seeded in direct contact environment. 125

Figure 6.4: The change in ratio of MG-63 cells to hES-MPs across 4 different initial seeding ratios. Seeding ratio here means the percentage of MG-63 cells in the total number of cells seeded (therefore, seeding ratio of 1:1, 1:2, 1:3 and 1:4 or 50%, 33% 25% and 20% MG-63 cells). The data represented on the y-axis is normalised to day 1 data. The ratio of MG-63 to hES-MP cells increases across all seeding conditions, with the highest increase

observed at low initial MG-63 cell numbers. Within the low MG-63 to hES-MP seeding conditions it was also found that the increase in ratio occurs at a lower rate in direct co-culture in comparison to indirect or mono-culture. 127

Figure 6.5: Cell numbers of MG-63 normalised to day 1. The behaviour of MG-63 cells was found to be similar in indirect co-culture and mono-culture. This was also matched with direct co-culture at high MG-63 to hES-MP seeding ratios (1:1 and 1:2 or 50% and 33% MG-63s). At lower ratios, the MG-63 cells grew at a slow rate than their counterpart cells restricted to indirect co-culture or mono-culture. This suggests direct co-culture with hES-MP cells has a negative effect at high ratios of hES-MP to MG-63 cells. 127

Figure 6.6: Cell numbers of hES-MP cells normalised to day 1. The change in cell number of hES-MP in co-culture was found to be slower than in mono-culture, specially at experiments with higher initial number of MG-63 cells (lower hES-MP to MG-63 ratios). At high ratios of hES-MP to MG-63 cells, co-culture effects appeared to be minimised. 127

Figure 6.7: Bar charts showing the difference in MG-63 to hES-MP ratio on day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. The relationship between indirect co-culture and mono-culture found at seed ratios of 1:1 and 1:2 were found to be linked to hES-MP behaviour (figure6.9) and the significant difference observed in direct co-culture at initial seed ratio of 1:3 and 1:4 was found to be a result of changes in MG-63 growth (figure6.8). 128

Figure 6.8: : Bar charts showing the difference in MG-63 growth by day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. At low MG-63 to hES-MP ratios (1:3 and 1:4) cells in direct co-culture were found to grow significantly slower than their counter parts in mono-culture and indirect co-culture. No significant difference was found between mono-culture and indirect co-culture at these ratios, however at higher ratios of MG-63 to hES-MP cells (1:1 and 1:2), significant differences were found. 128

Figure 6.9: : Bar charts showing the difference in hES-MP growth by day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. The rate of growth of hES-MPs was found to be significantly slower in co-culture than mono-culture under initial seeding ratios of 1:1, 1:2 and 1:3. No difference was observed at ratio 1:4.

Significant difference between direct and indirect co-culture was only found at ratio 1:1.

128

Figure 6.10: The simulation results of no-contact ABM (presented in chapter 5), run under the same seeding conditions as in vitro co-culture experiments (results presented in figure 6.7). The results presented here are the day4 simulation results normalised to day 1 data in order to allow for direct comparison with the in vitro results.

129

*Figure 8.1: MG-63 cells in low seeding density (5, 000 cells/cm²) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by color change. *

143

Figure 8.2: MG-63 cells in medium seeding density (10, 000 cells/cm²) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by colour change.

144

Figure 8.3: MG-63 cells in high seeding density (20, 000 cells/cm²) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by colour change.

145

Figure 8.4: hES-MP cells in low seeding density (5, 000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

147

Figure 8.5: hES-MP cells in medium seeding density (10, 000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

148

Figure 8.6: hES-MP cells in high seeding density (20, 000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

148

Figure 8.7: A-test score of all sets per sample size of 1 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 151

*Figure 8.8: A-test score of all sets per sample size of 5 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines*5 151

Figure 8.9: A-test score of all sets per sample size of 10 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 152

Figure 8.10: A-test score of all sets per sample size of 25 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 152

Figure 8.11: A-test score of all sets per sample size of 50 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 153

Figure 8.12: A-test score of all sets per sample size of 100 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 153

Figure 8.13: A-test score of all sets per sample size of 150 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 154

Figure 8.14: A-test score of all sets per sample size of 200 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 154

Figure 8.15: A-test score of all sets per sample size of 250 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 155

Figure 8.16: Model calibration for cancer growth showing the RMSE between model output and experimental data (from top to bottom) at 24, 48, 72 and 96 hours post seeding. To the right is the lower quartile used to determine the closest fit scenarios which were further analysed using ANCOVA. 156

Figure 8.17: Model calibration for stem cell growth showing the total RMSE across all timepoints between model output and experimental data at cc.MAXSTAGE equal to (from top to bottom) 23, 24, 25, 26 and 27 hours, To the right is the lower quartile used to determine the closest fit scenarios which were further analysed using ANCOVA. 157

Figure 8.18: A-test score comparing the baseline model output against model outputs resulting from changes in gridsize parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 160

Figure 8.19: A-test score comparing the baseline model output against model outputs resulting from changes in nucl_percent parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 161

Figure 8.20: A-test score comparing the baseline model output against model outputs resulting from changes in precent_r_reduce parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 162

Figure 8.21: A-test score comparing the baseline model output against model outputs resulting from changes in stem cell disp_shift parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 163

Figure 8.22: A-test score comparing the baseline model output against model outputs resulting from changes in cancer cell disp_shift parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 164

Figure 8.23: A-test score comparing the baseline model output against model outputs resulting from changes in stem cell death rate parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 165

Figure 8.24: A-test score comparing the baseline model output against model outputs resulting from changes in cancer cell death rate parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines. 166

List of Tables

| | |
|--|----|
| <i>Table 2.1: Literature reporting MSC promotion of tumour growth. (MSC: Mesenchymal stem cell, CC: Cancer cell, ASC: Adipose stem cell, BMD: Bone marrow-derived).</i> | 14 |
| <i>Table 2.2: Literature reporting MSC inhibit tumour grow (MSC: Mesenchymal stem cell, CC: Cancer cell, ASC: Adipose stem cell, BMD: Bone marrow-derived)</i> | 15 |
| <i>Table 2.3: some of the advantages and limitations of ABM in comparison to other modelling paradigms such as in vivo, in vitro and in silico continuum models</i> | 26 |
| <i>Table 2.4: Common ABM toolkits available (including boiological applications)</i> | 29 |
| <i>Table 2.5: Comparison of various agent-based models of biological tissue.</i> | 38 |
| <i>Table 3.1: Seeding density of hES-MP and MG-63 cells under the 4 different seeding ratio configurations(1:1, 1:2, 1:3 and 1:4) and seeding conditions (mono-culture, direct and indirect co-culture).</i> | 48 |
| <i>Table 3.2: Volumes of Hoechst, trypsin and neutralising solution used in the each well plate for detachment and manual counting. Also shown are the final cell suspension volumes per timepoint. *for each side of the barrier.</i> | 54 |
| <i>Table 3.3: The relationship between A-test scores and various magnitudes of difference between the two populations (Vargha and Delaney, 2000; Read et al., 2011)</i> | 61 |
| <i>Table 4.1: Reduction in skew and kurtosis of data by applying filter to the CellProfiler pipeline to exclude clumped objects.</i> | 70 |
| <i>Table 4.2: Table listing the statistical tests results from comparison of each of the fitted models to the raw experimental data. p values in bold indicate non-significant results and</i> | |

hence statistically similarity between model and data. The Lognormal model was found to be the best fit by both tests for both cells. 75

Table 4.3: Table of linear regression assumptions for the mitotic rate models for hES-MP and MG-63 cells. 87

Table 5.1: Parameter values. Parameters prefixed by cc or sc refer to the cancer and stem class respectively. The symbols used in the parameter column are as follows: μ = mean, σ = standard deviation, β = linear regression coefficient, α = linear regression intercept, CI = confidence interval, s = lognormal shape, loc = lognormal local, and scale = lognormal scale 101

*Table 6.1: Table listing the reagents and volume per culture environment. All measurements are calculated per surface area.. Also shown are the final cell suspension volumes per timepoint post cell detachment. *values stated are per barrier side* 122

Table 6.2: Seeding density of MG-63 and hES-MP cells at each ratio configuration and environment condition 123

Table 8.1: Table listing all pythong libraries (or otherwise) utalised in statistical analysis. 141

Table 8.2: Cell nuclei size and shape results 142

Table 8.3: Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 2, 500 cells/cm² 149

Table 8.4: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 5, 000 cells/cm² 149

Table 8.5: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 10, 000 cells/cm² 150

Table 8.6: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 20, 000 cells/cm² 150

Table 8.7: Statistical results of the 'null' ABM of cancer growth validated against 4 seeding densities of experimental growth curves of MG-63 cells 158

Table 8.8: Statistical results of the 'null' ABM of stem growth validated against 4 seeding densities of experiemtnal hES-MP cells *159*

Table 8.9: Table of parameters used in global sensitivity analysis and their respective upper and lower bounds. *173*

Chapter 1

Introduction

1.1 Motivation

Trailing behind heart disease, cancer is mankind's second deadliest condition, responsible for an estimated 9.6 million deaths (WHO, 2020) and 12million new diagnosis a year (Hesketh 2013). Rising world population, increasing life-span and unhealthy lifestyle choices have resulted in an increase in morbidity, with an estimated forecast of over 13million deaths and 26 million new cases in 2030 (Boyle and Levine, 2008).

There is a strong need for specific tumour targeting treatment as the current commonly used methods (radiation and chemotherapy) fail to differentiate between

cancerous and healthy cells. Stem cells, with their potent pathotropic properties (resulting in their migration towards sites of disease) and abundance within the human body, combined with emerging tissue engineering techniques are an attractive candidate for targeted cancer therapy (Shah, 2013a). However discrepancies found in literature and our limited understanding of the role of stem cells at sites of inflammation and tumorigenesis, demand the need for further investigation (Klopp et al., 2011).

As *in vivo* methods are expensive, time consuming and sometimes impractical to carry out, in this PhD, both *in vitro* and *in silico* modelling will be used to investigate these discrepancies. Experimental methods will include cell culture techniques such as time-lapse imaging and co-culture of two cells and computationally agent based modelling paradigm constructed using the Python programming language.

The overall aim of this project is to develop an agent based model to explore interactions of stem cells with tumour cells, using *in vitro* obtained results to inform and validate the model. We hope this model will clarify the role of stem cells at tumour sites and ultimately refine and aid the development of these cells as therapeutic tools for the treatment of cancer.

1.2 Aims and Objectives

The main aim of this PhD is to test the following hypothesis:

Stem cells alone affect the growth of cancer cells in co-culture.

And further to this:

Stem cell and cancer cell interactions are governed by juxtacrine signalling.

By using a combination of *in vitro* and *in silico* modelling approaches.

In order to achieve this, the following objectives are set:

1. Qualitative and quantitative exploration of the biological domain via *in vitro* single cell culture analysis of cancer and stem cell behaviour.
2. Construction of a robust agent-based model (ABM) of the *in vitro* cell culture environment explored in objective 1 for each cell type. – the ‘null’ model

3. Verification of the 'null' ABM simulations response using in vitro mono-culture experimental data.
4. Design of *in vitro* co-culture experiments to differentiate between juxtacrine and paracrine signalling.
5. *in silico* exploration of cell to cell interaction in co-culture between cancer and stem cells within the agent-based modelling environment. – the 'contact' model
6. Validation of the 'contact' ABM simulation responses using in vitro co-culture experimental data

Chapter 2

Background and literature review

2.1 Biological background and literature review

2.1.1 Cancer

Cancer is the term used commonly for a group of diseases characterised by abnormal, uncontrollable and relatively autonomous growth of cells (referred to as a tumour or neoplasm) with the tendency to metastasise (migrate and spread from the site of origin) (Underwood, 2004).

Exposure to various environmental factors, such as radiation, carcinogenic chemicals and some infections can cause genetic alterations in DNA. The right combination of these mutations, affecting critical genes, has the potential to trigger neoplastic transformation in any nucleated cell of the human body (Underwood, 2004). Some cell types are more prone

than others, with epithelial cells having the highest number of instances, representing; 85% of all cancers (Cancer Research UK, 2020b). Cancers of epithelial cells are referred to as carcinomas. Breast, prostate, lung and bowel carcinomas are the most common types of cancer, accounting for more than 50% of overall incidences in the UK in 2017 (Cancer Research UK, 2020a).

In addition to histogenetic classification, tumours can also be grouped based on their behaviour; benign, malignant and borderline. Malignant tumours are invasive and have a relatively rapid growth rate in comparison to the non-invasive nature and slow growth rate of benign tumours. They also have a very high tendency to metastasis to secondary sites, destroying cells (e.g. blood and lymphatic cells) along the way. Morbidity and mortality rates for malignant tumours are considerably higher due to their invasive nature (Underwood, 2004).

A popular concept proposed by Douglas Hanahan and Robert A. Weinberg in 2000 (Hanahan, Weinberg and Francisco, 2000) and revisited in 2011 (Hanahan and Weinberg, 2011) breaks down the complexity of cancer into eight rational underlying biological capabilities (hallmarks) and two enabling characteristics, which come together in various combinations resulting in the multistep development of the many different types of tumours. These hallmarks, as shown in figure 2, are as follows: sustaining proliferative signalling (via deregulation of growth-promoting signals), evading growth suppressors (via inactivation of various tumour suppressor genes), resisting cell death (evasion of predominantly intrinsic programmed cell death or apoptosis), enabling replicative immortality (continuous cell division past stages of senescence and crisis that would otherwise limit the number of successive cell growth and division cycles in healthy cells), inducing angiogenesis (where normally quiescent vasculature sprouts new vessels that sustain the tumour microenvironment), activating invasion and metastasis (caused by alterations to key cell-to-cell and cell-to-ECM adhesion molecules), deregulating cellular energetics (in order to fuel cell growth and division) and avoiding immune destruction (in particular by T and B lymphocytes and natural killer cells). The two enabling characteristics are, genome instability and mutation (allowing for cancer cells that are able to mutate with genetic alterations that in turn drive tumour progression) and tumour-promoting inflammation (allowing for the recruitment of innate immune cells into the tumour microenvironment contributing to multiple hallmark capabilities).

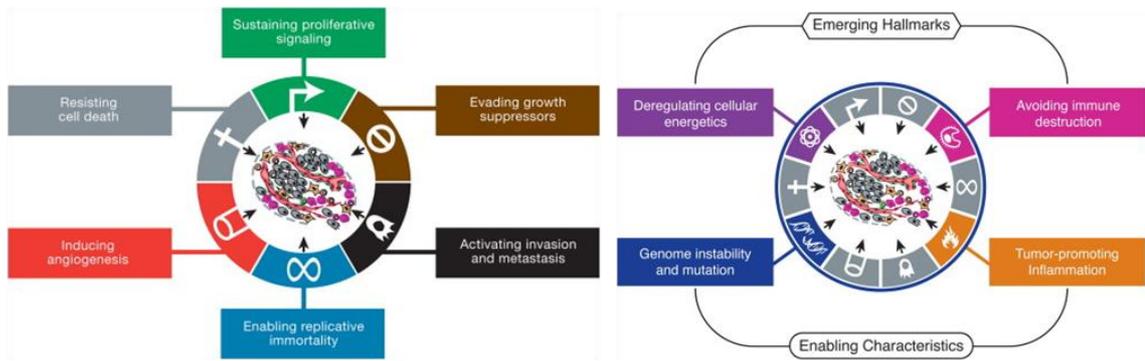


Figure 2.1: (Left) Original six hallmarks proposed in 2000 and (right) the two new emerging hallmarks and enabling characteristics. These are used to break down the complexity of cancer into eight rational underlying biological capabilities (hallmarks) and two enabling characteristics, which come together in various combinations resulting in the multistep development of the many different types of tumours. [Images reproduced from (Hanahan and Weinberg, 2011) copyright clearance obtained via RightsLink order number: 4945050185156]

2.1.2 Current treatments

An extensive international report looking at the cost of cancer in developed countries published by *The Lancet Oncology*, shows an increase in healthcare expenditure on cancer care in recent years (Coleman *et al.*, 2008). Another study looking at the economic burden of cancer across the European Union estimated the healthcare costs linked to cancer to be around £4.5bn with £900 million purely spent on drugs (Luengo-Fernandez *et al.*, 2013). The cost incurred due to loss of productivity as a result of cancer was calculated to be approximately 6bn (Luengo-Fernandez *et al.*, 2013).

There are different types of therapies to treat cancer and listed below are some of the most common forms of treatment available. Cancer patients are likely to receive a combination of these depending on their diagnosis and treatment plan.

2.1.2.1 Surgery

Surgery is the oldest method of cancer treatment and arguably still the treatment that produces the most cancer cures, especially when carried out on early stage tumours (Knowles and Selby, 2005).

Other than the obvious removal of the primary tumour and local extirpation of cancer, surgery can be used as a preventative (for example mastectomy in patients at high

risk of breast cancer (Lostumbo *et al.*, 2004)), reconstructive (via transplant or plastic surgery) or diagnostic tool (for example through lymphadenectomy) (Knowles and Selby, 2005; Bupa, 2014).

In most modern cancer care plans surgery is used in conjunction with adjuvant and neoadjuvant therapy such as radiotherapy, chemotherapy and various targeted therapies. This multidisciplinary approach maximises local control and minimizes occurrence of metastasis (Knowles and Selby, 2005).

2.1.2.2 Radiotherapy

Radiation oncology came into play subsequently after the discovery of X-rays in 1895, making it the second oldest form of cancer treatment (Robert A. Weinberg, 2013). Radiotherapy uses ionizing radiation to break single and double stranded DNA bonds and induce cell apoptosis (Kumar and Clark, 2005). It affects and damages both normal and cancerous cells and relies on the high multiplying rate of cancer cells and the ability of most normal cells to repair faster and more efficiently than cancer cells to deliver maximum damage to the tumour with minimal side effects (Baskar *et al.*, 2012). The side effects of radiotherapy depend on tissue sensitivity, fraction size and treatment volume and can be as serious as secondary malignancies appearing 10-20 years after treatment (Kumar and Clark, 2005).

2.1.2.3 Chemotherapy

Chemotherapy is a form of systemic cancer treatment in which cytotoxic drugs that target cellular DNA (e.g. DNA damaging or DNA repair and synthesis inhibiting) are administered. There is a narrow therapeutic window between achieving the desired anticancer effects and tissue toxicity (Kumar and Clark, 2005).

In addition to dose restrictions due to potentially serious side effects such as vomiting, hair loss, tiredness, mucositis and myelosuppression (Carey, 2003; Kumar and Clark, 2005), one of the major limiting factors of chemotherapy is drug resistance, the effect of which vary across different types of cancer (Hesketh, 2013). In order to overcome this, combinations of drugs (with differing mechanism and non-overlapping toxicities) are used (Kumar and Clark, 2005).

2.1.2.4 Gene therapy

This therapeutic method involves administration of exogenous genes (using a vector) into diseased cells and changing their phenotype through altered patterns of gene expression (Knowles and Selby, 2005). A popular method is the use of viral vectors (based on their ability to infect cells) however they are not tumour cell specific. An attempt to overcome this limitation is to engineer the virus to express ligands that are specific to cancer cell surface receptors (Hesketh, 2013).

2.1.2.5 Immunotherapy

The immune system detects and destroys abnormal cells via various mechanisms, many of which cancer cells are able to avoid. Immunotherapy is the referrers to the treatments that are designed to boost the body's immune response in order to fight cancer cells. These include, blocking antibodies to the immune checkpoint receptors, oncolytic virus therapy, T-cell therapy and cancer vaccines (Rosenberg, Yang and Restifo, 2004; Dougan, Dranoff and Dougan, 2019; Waldman, Fritz and Lenardo, 2020).

Although immunotherapies have proven to be highly effective in some select group of cancers and individuals, they have struggled to be consistently effective in majority of patients and cancer types. In addition to this unpredictability, the lack of known targetable tumour-specific antigens means that most immunotherapies are not cancer cell specific resulting in unwanted side effects to healthy cells (Lee Ventola, 2017).

2.1.2.6 Targeted therapy

Targeted therapy is the general term for treatments that target cancer cells more specifically. This means normal healthy cells aren't affected by the administered anti-cancer drug (or are affected to a much lesser extent). As a result, there should be less unwanted side-effects. These therapies range from molecular agents that target and disrupt cancer specific pathways to delivery agents that allow for local administration of anti-cancer drugs at the tumour site with minimal off-site accumulation. Listed below are some of the common forms of targeted therapy.

Intracellular signal inhibitors

These are small molecular agents (for example peptides) that inhibit products of oncogenes specifically responsible for transformation of cancer cells and their intracellular pathways (Green 2004; Kumar & Clark 2005).

Monoclonal antibodies (mAb)

mAbs can be used in different ways to treat cancer: specifically binding to target cancer cells in order to trigger the patient's own immune system to attack those cells (Knowles & Selby 2005); inducing cancer cell apoptosis by blocking cell surface receptors (Scott et al., 2012) or aid in the delivery of a drug or cytotoxic agent (for example attached to liposomal cytotoxic agent carriers (Immordino et al., 2006; Zamboni, 2005); and causing vascular and stromal cell ablation (Scott et al. 2012).

Endocrine therapy

Cancers derived from hormonally responsive tissue (such as breast, prostate and endometrium cancers) can be treated through manipulation of the hormonal environment. In these types of cancers, interactions between steroid hormones and their receptors causes cell growth and replication. This mechanism can be interfered with either via exogenous administration of specific hormones or drugs that inhibit hormone production to cause apoptosis and regression of the cancer (Kumar & Clark 2005).

Stem cell therapy

Stem cell therapy involves coupling the disease seeking nature of stem cells with advances in engineering in order to create targeted site specific therapeutic treatment.

The next section goes into more detail regarding what exactly stem cells and stem cell therapy of cancer are.

2.1.3 Stem cell therapy of cancer

Stem cells are the undifferentiated biological cells that allow for embryogenesis and the continued regeneration of tissue throughout adult life. They are defined by their ability to self-renew and differentiate and can be divided into two groups based on their site of origin, embryonic and adult stem cells.

Embryonic stem cells are pluripotent (have the ability to differentiate into many cell lines) and originate from the inner cell mass of blastocyst (found in the very early stages of the embryo) (J. a. Thomson, 1998). As outlined in figure 2.2 they differentiate into germ layer stem cells of either endoderm (going on to form the inner stomach lining, gastrointestinal tract and lung tissue), mesoderm (muscle, bone, blood and urogenital tissue) and ectoderm (epidermal tissue and the nervous system) class (Corsten and Shah, 2008).

Adult stem cells can be found in virtually every tissue and include mesenchymal and haemopoietic stem cells (found in the mesoderm) and neural stem cells (found in the ectoderm). They are undifferentiated stem cells that reside near their respective differentiated end organ tissues and play a vital role in maintaining and repairing it. They remain quiescent until they are needed in the repair and maintenance of tissue (Corsten and Shah, 2008).

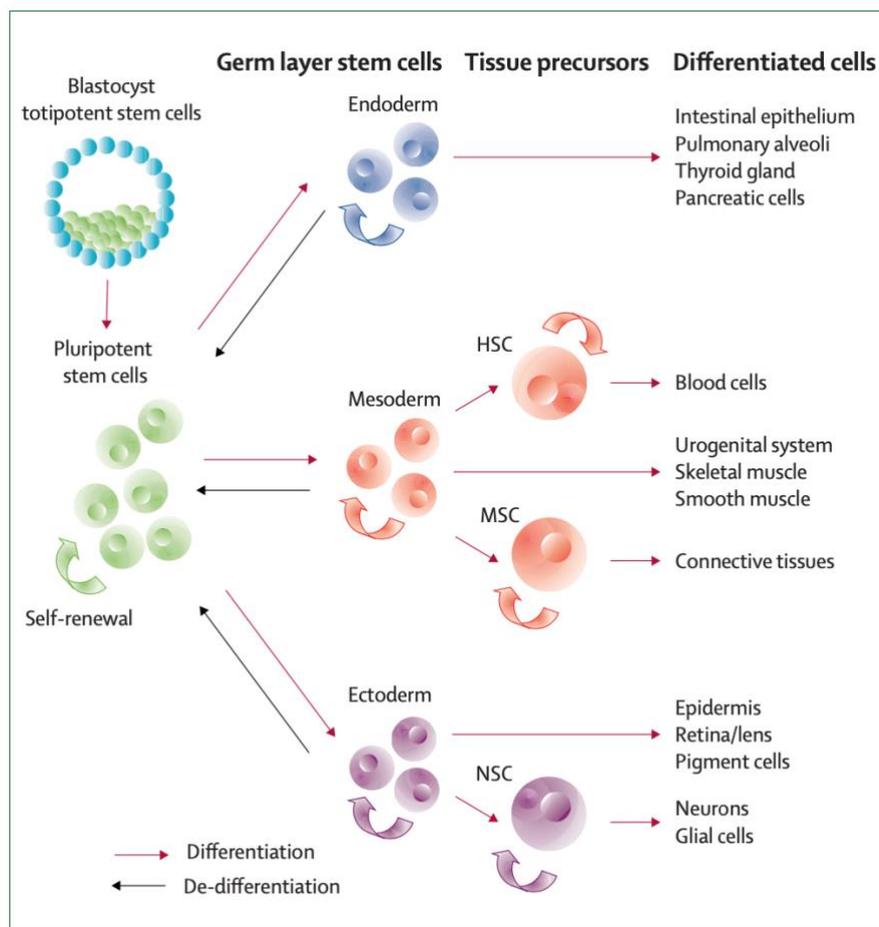


Figure 2.2: Classical hierarchical model of stem cell differentiation. HSC = haemopoietic stem cells, MSC = mesenchymal stem cells & NSC = neural stem cells. [Image reproduced from (Corsten & Shah 2008) copyright clearance obtained via RightsLink order number: 4945461063966]

Other types of stem cells worth mentioning are cancer stem cells (CSC) and induced pluripotent stem cells (iPSC). Cancer stem cells are a subpopulation of cancer cells that have a greater renewal and repair ability compared to other cancer cells. It is thought that they play a vital role in post therapy relapse, therapeutic drug resistance and providing stromal support. Little is known about their origin or even if they are present in all types of cancers. It may be that normal adult stem cells are being recruited and undergo oncogenic transformations to produce CSCs. (Hanahan and Weinberg, 2011).

Induced pluripotent stem cells (iPSC) are an engineered strain of stem cells that are achieved by inserting four specific genes of embryonic stem cells into fully matured cells. The result is a 'dedifferentiated' cell with pluripotent properties. These cells are very similar to embryonic stem cells, exhibiting the same morphology and growth properties and expressing the same cell marker genes. They were first engineered using cells from mice in 2006 and human cells in 2007 by Japanese researchers Takahashi and Yamanaka. (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007).

Stem cells are distinguishable from each other based on the tissue and developmental stage they were harvested from, their differentiation ability and the genes they express, giving them specific individual characteristics. However there are also some key universal properties of stem cells, referred to as 'stemness' that can be found in all stem cells with ongoing research being carried out to identify common biomarkers (Ivanova *et al.*, 2002; Cai, Weiss and Rao, 2004).

The stem cell niche refers to the microenvironment (*in vivo*) of each stem cell. The interaction (extrinsic signals and stimuli) between the stem cell and its environment determines its fate, with various studies that have shown that the characteristics of stem cells of one tissue can be altered to mimic the properties of another if subjected to environmental cues of the latter (Townsend and Jones, 2004; Ferraro, Lo Celso and Scadden, 2010).

In the context of cancer treatment, stem cells are mainly used in two ways; either post chemotherapy or radiotherapy transplantation to replace the patients lost stem cells (Cancer Research UK, 2014) or as delivery vehicles for targeted therapy. It is this latter use of stem cells for treatment of cancer that is referred to as stem cell therapy of cancer in this thesis and will be explored further.

2.1.4 Stem cells as vehicles for targeted therapy

Because of their inherent tumour tropism (Klopp et al., 2011; Shah, 2013a), stem cells are attractive candidates for targeted cancer therapy. This combined with new advances in gene therapy and tissue engineering means that stem cells can be used as highly specific cellular vehicles for delivery of effective treatment to the tumour site. Figure 2.3 lists some of the potential transgene strategies for stem cell therapy of cancer.

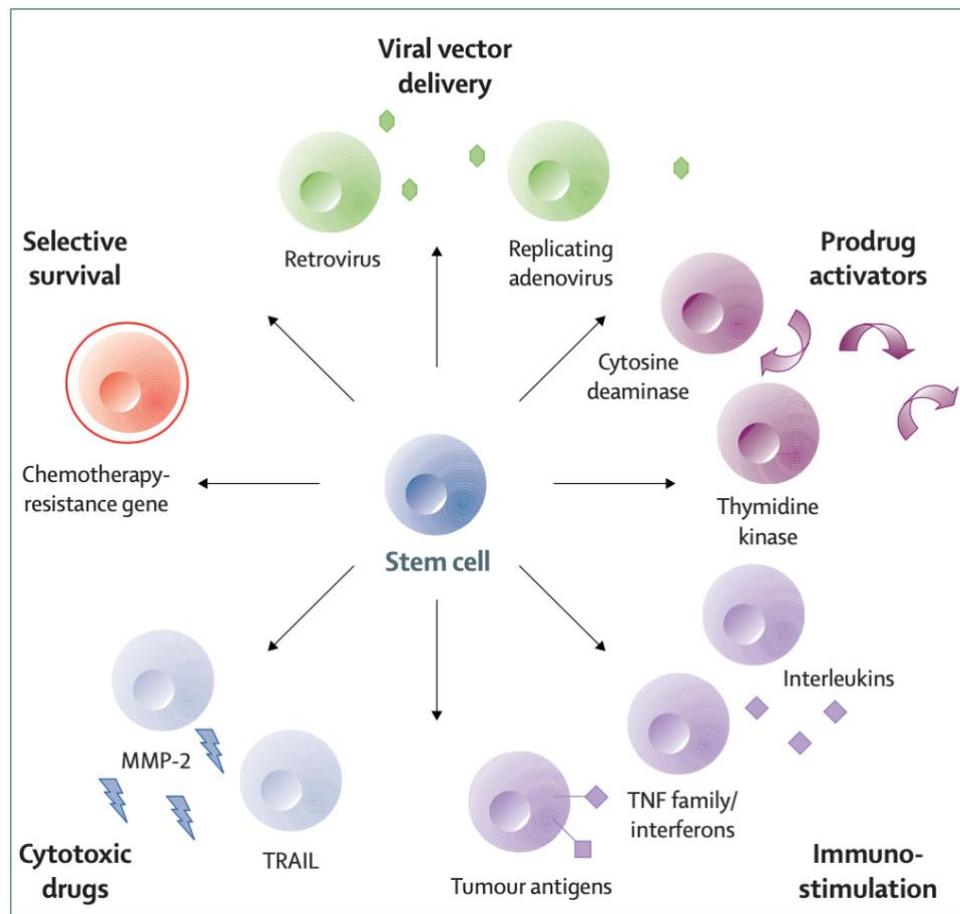


Figure 2.3: Potential transgene strategies for stem cell therapy of cancer. MMP2=matrix metalloproteinase 2. TRAIL=Tumour necrosis factor related apoptosis inducing ligand * TNF = tumour necrosis factor. [Image reproduced from (Corsten & Shah 2008) copyright clearance obtained via RightsLink order number: 4945461063966]

Neural stem cells were shown to have extensive tropism, seeking out intracranial microsatellite deposits of brain tumours (Aboody *et al.*, 2000). Similarly genetically modified embryonic endothelial progenitor cells were able to not only locate minor pulmonary tumour metastases but effectively treat them by expression of an engineered gene (Wei *et al.*, 2004). Another example of successful application of viral vector delivery of

tumour killing transgene is the work done by Studeny *et al.*, (2004) where human MSC, transduced with an adenoviral expression vector carrying the human IFN- β gene, were used to suppress the growth of pulmonary metastases in mouse xenograft model (Studeny *et al.*, 2004).

2.1.5 The effect of unmodified stem cells on tumour

Despite the promising advancements made in stem cell therapy of cancer, very little is known about the impact of unmodified stem cells on tumour progression (Klopp *et al.*, 2011). It can be argued that once stem cells are engineered as anti-tumour agents, any inherent tumour promoting qualities they may have, are overcome. However it is still very beneficial to improve our understanding of the mechanisms and conditions under which stem cells enhance or inhibit tumour growth and metastasis as this type of knowledge allows for the engineering of better suited therapies and furthers our understanding of the role of the tumour stroma in cancer progression. (Klopp *et al.*, 2011)

One specific type of stem cell that has been under investigation using various *in vivo* and *in vitro* models is mesenchymal stem cells. There is a clear discrepancy in the literature as to what the role of these cells in tumorigenesis is, with some papers indicating that they promote or support tumour growth (Djouad *et al.*, 2003; Zhu *et al.*, 2006; Karnoub *et al.*, 2007; Galiè *et al.*, 2008; Fabian L Muehlberg *et al.*, 2009; Kucerova *et al.*, 2010; Lin *et al.*, 2010; Prantl *et al.*, 2010; Shinagawa *et al.*, 2010; Huang *et al.*, 2012) while others suggest that they suppress it (Ohlsson *et al.*, 2003; Khakoo *et al.*, 2006; Lu *et al.*, 2008; Qiao, Z. Xu, *et al.*, 2008; Qiao, Z.-L. Xu, *et al.*, 2008; Cousin *et al.*, 2009; Zhu *et al.*, 2009; Otsu *et al.*, 2009; Dasari, Kaur, *et al.*, 2010; Dasari, Velpula, *et al.*, 2010; Secchiero *et al.*, 2010).

Tables 2.1 and 2.2 list some of these papers according to the models and cell types used. Across these papers different mechanisms have been proposed to be responsible for either the tumour inhibiting or promoting effect of the stem cells. Relevant to this project are the papers that refer to the effect of stem cells on regulating the cell cycle (and hence proliferation) and inducing or inhibiting apoptosis of the cancer cells.

Table 2.1: Literature reporting MSC promotion of tumour growth. (MSC: Mesenchymal stem cell, CC: Cancer cell, ASC: Adipose stem cell, BMD: Bone marrow-derived).

| Paper | MSC:CC (ratio) | Model | Findings | Mechanism |
|---|--|------------------------------|---|---|
| (Fabian L Muehlberg <i>et al.</i> , 2009) | Human ASC & Breast (10:1) | Syngeneic mouse model | Increased tumour size & a more rapidly forming tumour | Paracrine factor (SDF-1/CXCR secretion) |
| (Galiè <i>et al.</i> , 2008) | Mouse ASC & Breast | Syngeneic mouse model | Increased incidence & tumour size | Angiogenesis |
| (J. Yu <i>et al.</i> , 2008) | Human ASC & Lung or Glioma | Nude mouse model | Increased tumour size | Reduced apoptosis |
| (Karnoub <i>et al.</i> , 2007) | Human BMD & Breast cancer cells (3:1) | Immunocompromised mice model | Increased tumour size | Chemokine secretion |
| (Zhu <i>et al.</i> , 2006) | Foetal & adult BMD and colon cancer cells (10:1 & 1:1) | Murine xenograft model | Both ratio showed increased incidence & enhanced vascularity & necrosis | Enhanced proliferation and angiogenesis |
| (Djouad <i>et al.</i> , 2003) | Mouse BMD & Melanoma (1:1) | Allogenic mouse model | Increased incidence | Immunologic |

Table 2.2: Literature reporting MSC inhibit tumour grow (MSC: Mesenchymal stem cell, CC: Cancer cell, ASC: Adipose stem cell, BMD: Bone marrow-derived)

| Paper | MSC:CC (ratio) | Model | Findings | Mechanism |
|---|---|---|---|--|
| (Dasari, Velpula, <i>et al.</i> , 2010) | Umbilical cord blood-derived & Glioma (1:4) | <i>In vitro</i> co-culture <i>In vivo</i> U251 & 5310 nude mice brains | Reduced tumour size | Akt signalling (Apoptosis) |
| (Secchiero <i>et al.</i> , 2010) | BMD & Non-Hodgkins lymphoma (1:10 & 1:2) | Nude SCID mice model | Reduced tumour burden | Increase in apoptosis |
| (Cousin <i>et al.</i> , 2009) | Human ASC & Pancreatic cancer cells | <i>In vitro</i> co-culture & <i>in vivo</i> pancreatic cancer xenografts | Reduced tumour size | G1-phase arrest (cell cycle progression) |
| (Lu <i>et al.</i> , 2008) | Mouse BMD & Hepatoma or Lymphoma cells | <i>In vitro</i> co-culture <i>In vivo</i> BALB/c mice | Reduced tumour size Decreased ascites formation | G1-phase arrest Induced apoptosis |
| (Qiao, Z. Xu, <i>et al.</i> , 2008) | Human foetal skin & Hepatoma (1:1) or Breast cancer (1:100) | <i>In vitro</i> co-culture & <i>in vivo</i> SCID mice model | Reduced tumour size Reduced proliferation & colony formation Reduced metastasis | Wnt signalling |
| (Khakoo <i>et al.</i> , 2006) | Human BMD & Kaposi sarcoma | Athymic nude mice model | Reduced tumour size | AKT signalling (Apoptosis) |

The cell cycle is a sequence of stages that the cell goes through from one cell division to the next. Hence the collective progression of cells through the cell-cycle leads to cell proliferation. The five phases of the cell cycle are shown in figure 2.4.

The M phase refers to the cell division via mitosis and the G₁, S and G₂ phases are all classed as the interphase stage of the cell in which the cell grows and replicates its content ready for mitosis. This phase typically covers 90% of the cell cycle (Kent, 2000). The G₀ phase is technically outside of the cell cycle and is also referred to as the resting phase or quiescent stage. It includes cells that are no longer dividing.

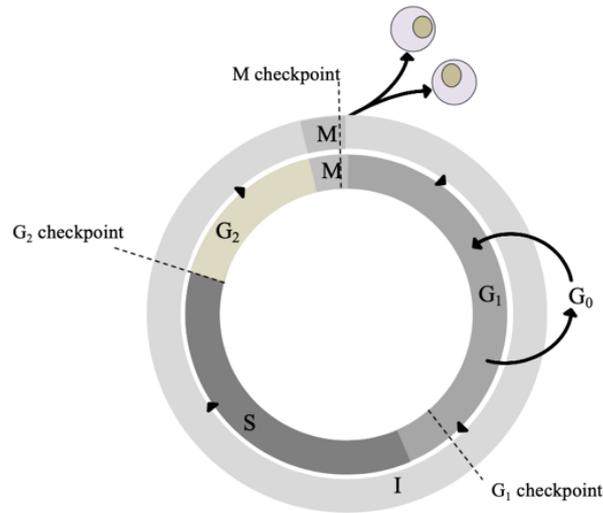


Figure 2.4: The human cell cycle. The M phase refers to the cell division via mitosis and the G₁, S and G₂ phases are all classed as the interphase stage of the cell in which the cell grows and replicates its content ready for mitosis. The G₀ phase is technically outside of the cell cycle and is also referred to as the resting phase or quiescent stage. It includes cells that are no longer dividing.

The regulation of the cell cycle is maintained by a series of checkpoints. Two of the most important check points are G₁/S phase, before DNA replication occurs and G₂/M phase and before the cell enters mitosis. These checkpoints are primarily controlled using cyclins and cyclin-dependent kinases (CDKs).

At the G₁/S phase checkpoint, also referred to as the restriction point, the cell size and environmental conditions are checked (Phillips *et al.*, 2013). If the cell meets these conditions, it can progress onto the S phase. However if it doesn't, the cell will either go into the G₀ phase (permanently or semi-permanently), or it will go into repair or apoptosis will be triggered. Apoptosis is the programmed cell death and can be triggered by many pathways and signals (not just checkpoint regulatory mechanisms). At the G₂/M phase the internal conditions of the cell are mainly checked, for example, the correct replication of the DNA).

Below some of the literature that highlight the cell cycle regulation and apoptosis mechanisms as the cause of the tumour promoting or inhibiting effect of MSCs on cancer cell interactions, are looked at in more detail.

2.1.5.1 Tumour promoting

Shinagawa *et al.*, (2010) compared the effect of injecting colon cancer cells alone into orthotopic nude mice model with injecting a mixture of colon cancer cells and MSCs. The

latter showed a higher cancer cell proliferation and significantly lower apoptotic index. (Shinagawa *et al.*, 2010)

Adipose stem cells (ASC) and lung tumour or glioma cells were transplanted subcutaneously or intracranially into nude mice in work done by Yu *et al.*, (2008). The result observed indicated an increase in the proliferation rate of the cancer cells in the presence of the stem cells or the stem cell's conditioned media. A reduction in apoptosis cell death was also observed. (J. M. Yu *et al.*, 2008)

Another group using adipose stem cells also noticed tumour promoting effects. Muehlberg *et al.*, (2009) cotransplanted ASCs with mammary breast cancer cells in a syngeneic mouse model. The results showed larger and more rapidly forming tumours. Both human and mouse ASCs showed this result. They concluded that the interaction of local tissue-resident stem cells with tumour stem cells plays an important role in tumour proliferation and metastasis. (Fabian L Muehlberg *et al.*, 2009)

Both adult and foetal bone-marrow derived (BMD) MSCs were coinjected with colon cancer cells in a murine xenograft model by Zhu *et al.*, (2006) They observed extensive necrosis and invasion of surrounding normal tissue. Tumour cells showed increased proliferation rates and enhanced vascularity when injected with MSCs in comparison to without. (Zhu *et al.*, 2006)

Karnoub *et al.*, (2007) coinjected BMD-MSCs with (four separate) breast cancer cell lines at a ratio of 3:1 into immunocompromised mice. One out of four cell lines used, showed accelerated tumour growth and all cell lines had increased incidences of metastasis. Metastasis was only seen when MSCs were coinjected with the tumour cells, suggesting they act only when in close proximity of the tumour. Based on these observations and co-culture in which the cytokines, chemokins and growth factors found in the conditioned media of MSC and breast cancer cells were screened, it was concluded that local paracrine MSC secretions of CCL5 is likely to have a prometastatic effect. (Karnoub *et al.*, 2007)

2.1.5.2 Tumour inhibiting

Cousin *et al.*, (2009) examined the effect of adipose-derived stem cells on pancreatic cancer cells both via *in vitro* co-culture and *in vivo* injection into established pancreatic cancer xenografts. In the co-culture experiments tumour the group concluded that these stem cells have the ability to interfere with the proliferation of tumour cells through altering cell cycle progression. They observed that the tumour cells were in a resting state (G_0) when coming

into contact with the stem cells and CDK4 and cyclinD1 (required for G₁/S phase progression) were down regulated. What was interesting though is that, in experiments where cancer cells were exposed to stem cell conditioned media, cell death by apoptosis was not identified despite observations of cancer cell necrosis. For the *in vivo* model, a “strong and long-lasting” tumour growth inhibition was observed resulting from a block at the G₁/S checkpoint. (Cousin *et al.*, 2009)

Lu *et al.*, (2008) also had similar findings. MSCs exhibited a ratio dependent growth inhibitory effect on murine tumour cell lines *in vitro* and inhibited the growth of ascitogenous hepatoma cells *in vivo* without host immunosuppression. They conclude via cell cycle analysis that MSCs caused an accumulation of cancer cells in the G₀ and G₁ phases. They also mention an increase in apoptotic cell death. (Lu *et al.*, 2008)

Investigations done by Khakoo *et al.*, (2006) show that human MSCs can inhibit the *in vitro* activation of the Akt protein kinase. This protein is a critical mediator of the Kaposi’s sarcoma growth and survival and it is part of an apoptosis blocking signalling pathway.(Khakoo *et al.* 2006)

Dasari *et al.*, (2010) have carried out a few experiments, looking at the effect of umbilical cord blood derived stem cells on glioma cells both *in vitro* and *in vivo*. They observed an upregulation of PTEN, a tumour suppressor gene that has a major role in the P13k/Akt signalling pathway. This upregulation was seen alongside a downregulation of Akt and P13k signalling pathway molecules and XIAP, an apoptosis inhibitor protein. The overall effect of this is an inhibition of tumour growth through increased cell apoptosis (Dasari, Kaur, *et al.*, 2010; Dasari, Velpula, *et al.*, 2010).

The work done by Secchiero *et al.*, (2010) also observed an increase in cell apoptosis. This group used two *in vivo* models of disseminated non-Hodgkin’s lymphomas in nude SCID mice (Secchiero *et al.*, 2010).

Qiao *et al.*, (2008) observed tumour inhibiting effects too. Here human foetal-derived MSCs were coinjected at a ratio of 1:1 with human liver cancer cells into an animal transplantation model (SCID Mice). *In vitro* investigations were also done by co-culturing the two cell types and by treatment of the cancer cells with the conditioned media from the MSCs. In all cases the MSCs showed to have tumour suppressing properties. The animal model showed a delay in tumour formation and smaller tumour size. The co-culture results were a decrease in cell proliferation, increase in apoptosis and a down regulation of the expression of Bcl-2, c-Myc, proliferating cell nuclear antigen (PCNA) and surviving. Decreased colony formation and proliferation were observed when treating the tumour cells

with the conditioned media of the MSCs (Qiao, Z. Xu, *et al.*, 2008) This group also carried out the same investigations on MCF-7 breast cancer cells and observed very similar results (Qiao, Z. Xu, *et al.* 2008) They conclude that the tumour inhibition is due to wnt signalling pathway (a mechanism crucial to stem cell self-renewal and differentiation as well as tumourigenesis).

2.1.6 Possible causes of discrepancies in the literature

The discrepancies found in the literature can be attributed to the differences in experimental design of each study. For example, the MSCs used across the studies were from various types of adult and foetal tissues. In addition to this, there is also variability (genetic, epigenetic or environmental) in MSCs from donor to donor. Isolation of a more homogenous population of MSCs using specific markers could increase the reproducibility and consistency of findings across the studies. Furthermore, isolated MSCs subjected to different culture media and growth conditions may experience alterations to their phenotype. This combined with use of cells from varying passages and confluencies could also contribute to the discrepancies found.

Another cause of inconsistency is the use of different methodologies and models. For example *in vitro* studies of co-culture done with conditioned media of stem cells versus transwell inserts (to allow both cells to grow in a shared environment). Or in the case of *in vivo* studies, the use of different tumour models. The timing at which MSCs were added to cancer cells or tumour microenvironment could also be a factor, with some studies introducing the MSCs to established tumours (generally reporting tumour inhibiting results) and others to early stage tumour models (reporting tumour promoting effects and increased vessel density). In addition to this, the protocol used to mix MSCs with cancer cells varied across studies. In some studies MSCs were injected into the tumour site, some used modifications such as implantation into gelatine matrix and some coinjected mixtures of MSCs and tumour cells. Finally there was a variety of dosages of MSCs used across the studies with ratios of MSCs to cancer cells ranging from 1:1 to 1:100. This could mean that the role MSCs play at the tumour microenvironment could be dose dependant with studies that used higher ratios of MSCs to cancer more likely to report tumour growth.

2.2 Modelling background and literature review

Model /^lmɒd(ə)l / •n. a simplified description of a system or process, to assist calculations and predictions. (Concise Oxford English Dictionary, 2009)

The above definition is the generic scientific meaning of the word ‘model’ found in the Oxford dictionary. It describes, in simple terms, what a model is and what it aims to achieve. A model can take many different forms, ranging from a physical prototype or schematic to a lab based experiment, aiming to create a simplified version of the real system (*in vitro*), to a computational simulation (*in silico*). This definition has been extended by Peter Kohl and colleagues to include any intellectual activity (or the product of) that attempts to reduce a system to its components and then predict the outcome of the interactions of those components. Hence, even classifying thinking as a modelling tool in itself (Kohl *et al.*, 2000).

So, in short, we model to understand the system (or problem) better. This is extremely valuable when trying to understand complex systems such as the human body that span across multiple spatial and temporal scales. For this purpose, biological studies have already established the use of *in vitro* (for example, cell culture assays) and *in vivo* (for example, nude mice models or the use of *E. coli*) modelling approaches (Genot *et al.*, 2013).

However, these experimental approaches are expensive, difficult to reproduce and to the most extent, unable to consider and explore all aspects of the data space methodically. As a result and encouraged by the recent exponential growth in computational processing power, an increased effort has been put on establishing and using *in silico* modelling tools for use in understanding physiological systems (Deisboeck and Stamatakos, 2011).

In this project *in silico* agent based modelling will be paired with *in vitro* models of interacting stem cells and cancer cells in order to gain a better insight into the interacting mechanisms of these two cell types.

2.2.1 Contrasting approaches to modelling

There are various *in silico* modelling methodologies, each seeking to answer a different type of problem. Here, two of the main independent opposing pairs of model classifications are compared and discussed in the context of the proposed problem of this PhD.

2.2.1.1 Bottom-up versus Top-down

As previously mentioned, the human body spans across a wide range of spatial and temporal scales (figure 2.5). The reductionist approach of the past 50 years has generated a vast amount of data at various levels, especially at the protein and gene level (a result of the major advancements made in proteomics and genomics) (Noble, 2002; Genot, Fujii and Rondelez, 2013). Thus one logical modelling approach would be to start at the smallest scale and model each of the components of a system in order to obtain an emergent (and potentially predictive) behaviour of the entire system. This is referred to as ‘bottom-up’ and agent-based modelling falls into this class of models.

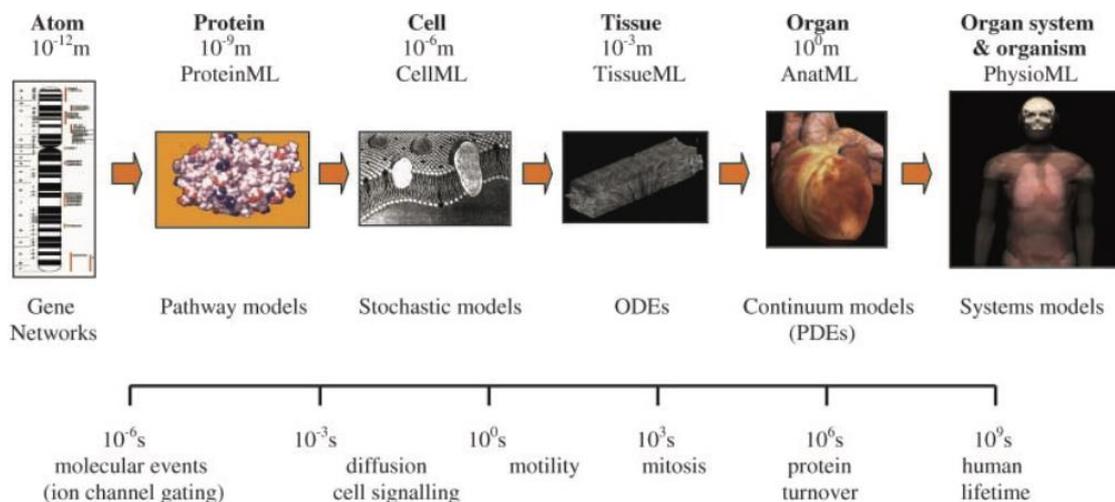


Figure 2.5: The many spatial (top) and temporal (bottom) scales of the human body. [Reproduced from (Hunter, Robbins and Noble, 2002) copyright clearance obtained via RightsLink order number: 4946181068994]

To apply this method successfully, the interactions, integrations and feedback across the scales need to be modelled too (for example, gene expression can be affected by factors at the environmental and organism scale) (Noble, 2002).

This type of model is also termed ‘mechanistic’ and although it is desirable in the sense that it could potentially lead to the creation of a virtual version of the human physiome (a goal set by projects such as and the physiome project (see <http://physiomeproject.org/>) and the Virtual Physiological Human (see <http://www.vph-institute.org/>)) which among many things, could be used as a simulated environment to test new medication or design patient specific treatment plans, it is limited by the lack of data available and level of model

abstraction. As mentioned by Denis Noble, for a model to teach us something new, it requires a sense of partial representation in order to show which features of the system are necessary and sufficient to understand it. And so a holistic model fails to have much interpretative value (Noble, 2002).

The opposing methodology to this is the ‘top-down’ approach. Here the model attempts at constructing theories to match and explain the observed system level characteristics (Noble, 2002). The limitation to this is that the final model may end up being ‘empty’ or incomplete due to the inability to find the right or complete set of lower level mechanisms (Noble, 2002).

A ‘top-down’ model can also be referred to as a phenomenological model and is commonly used in computer science. As the name suggests, a phenomenological model aims to explain the system based on the phenomena observed. The benefit of these models is that there is no need to include any explicit mechanistic details. The model is just required to fit the data available, so in a way very similar to a statistical model.

A new approach that has recently attracted a lot of interest is the ‘middle out’ method. This involves taking an intermediate scale and then gradually expanding it to include both smaller and larger scales a term coined by Sydney Brenner and Denis Noble (Noble, 2002; Walker and Southgate, 2009). This is demonstrated in figure 2.6, showing that the model can start at any scale and eventually ‘reach’ up, down or across to include other scales (Noble, 2002).

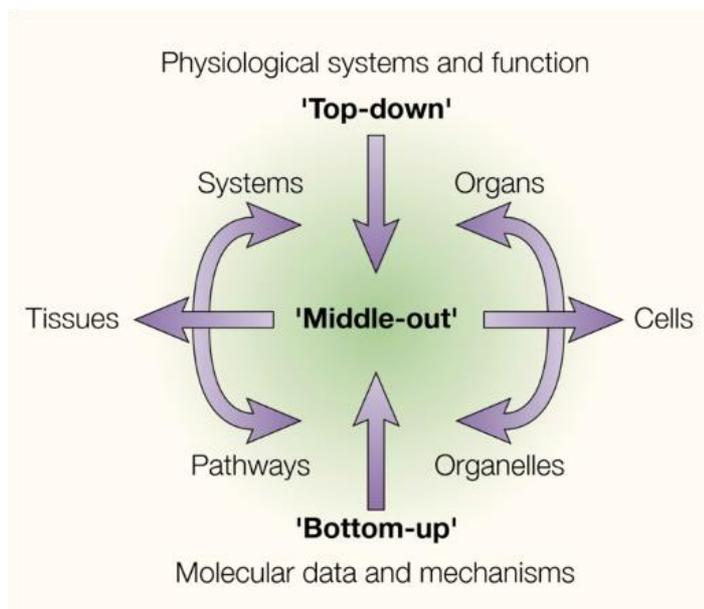


Figure 2.6: At what level should modelling begin? Diagram to show how models can start at any scale and eventually ‘reach’ up, down or across to include other scales. [Image reproduced from (Noble 2002) copyright clearance obtained via RightsLink order number: 4946180268061]

Two main factors decide the starting point, the availability of data and the range of scales of immediate interest (Noble, 2002; Walker and Southgate, 2009). For example in the context of this project starting at the cellular level seems the most natural. This is based on the fact that there is *in vivo* and *in vitro* generated data (Klopp *et al.*, 2011) and some of the missing data can be obtained in house through cell culture protocols. Also, the scales of interest span across cells (stem cells and cancer cells), pathways (interactions of these cells through signalling pathways) and tissue (the effect of these interactions on cell proliferation and hence cancerous tissue).

2.2.1.2 Discrete versus Continuum

Continuum models use fluid and continuum mechanics principles to represent an averaged behaviour of a system (mostly through the use of PDE and ODEs). In systems biology they can be used to describe subcellular components such as, cell volume fractions, density and cell substrate concentrations (e.g. nutrients and growth factors) (Deisboek *et al.*, 2011; Zhang *et al.*, 2009). Continuum models often lump multiple physical properties into one or two phenomenological parameters. Therefore where these models can be too ‘coarse-scaled’ to capture the intricacy of the system microarchitecture (e.g. biological tissue), discrete models can offer a higher degree of resolution and detail (Byrne and Drasdo, 2009).

A general definition of discretisation is the reduction of a system to its sub-components and it is important to note here that this is not unique to discrete models and is also employed in continuum modelling. However in discrete models, individual members of the system are directly defined at the model scale (e.g. modelling cells as discrete entities in tissue models) and rules and conditions are applied to each discrete body and can be tracked and updated as they progress through space and time and interact with each other. In other words they are autonomous, dynamic and adaptive.

There are two types of discrete models, those confined to a lattice-based environment (for example cellular automata models) or those left in a lattice-free setting in which they can take up arbitrary locations and interact in arbitrary directions (agent-based modelling) (Deisboek *et al.*, 2011; D. C. Walker & Southgate, 2009).

The ability to apply individualised rules to each agent type means that discrete models can produce a more accurate representation of the real system (specially with

reference to the systems heterogeneity). This is done through the integration of detailed biological findings into the rules sets of each agent which in some cases can be even be directly mapped. However, the larger the number of discrete elements and rules, the more computationally taxing the model becomes (Deisboek *et al.*, 2011)

Hybrid modelling is the middle point solution. It draws on the strengths of both continuum and discrete modelling through their integration. Hybrid models themselves can be grouped into composite and adaptive types. Composite hybrid modelling involves the discrete modelling of agents while still allowing them to interact with continuum fields. Adaptive hybrid models on the other hand, use both discrete and continuum modelling dynamically and adaptively where needed. (Deisboek *et al.*, 2011).

2.2.2 Agent-Based Modelling (ABM)

Agent-based modelling (ABM) can be defined as a computational approach to simulate the behaviour and interactions of autonomous entities with each other and their local environment in order to simulate emergent patterns (Thorne *et al.*, 2007). The individual entities are called *agents* and hence the model is referred to as an *agent-based model* (Srblijinovi and Skunca, 2003).

ABM is an *in silico* tool that is discrete (yet with potential to act as a hybrid model allowing integration with other modelling modalities), ‘bottom-up’ (yet also able to model ‘middle-out’ approaches and extend into other spatial and temporal scales) and capable of representing two or more scales across time and space. It is a popular modelling tool used in many different disciplines, from archaeology to politics and economics to more recently, biomedical sciences, for example, ABM of infection disease (Sarpe and Jacob, 2013).

As an interpretative tool, ABM can be used to observe and explore the consequences of specific individual-level rules on emergent population level results. And as a predictive tool, it can be used to predict emergent system level behaviour based on previously defined rule sets. ‘Emergence’ refers to the concept of macro-scale effects as a result of micro-scale interactions (Friesen, 2018).

2.2.2.1 Advantages and limitations

In comparison to other modelling paradigms, ABMs have many advantages and also some disadvantages. The main advantage of ABM, is the construction of a model that allows us to explore the emergent behaviour of a complex system such as the tumour microenvironment based on its (simpler) constituents in a virtual environment. This may not otherwise be possible in a real world setting due to the impracticable nature of some enquiries and time and money constraints. ABMs are also very flexible. As previously mentioned they can act as a hybrid model, integrating with other discrete and also continuum modelling approaches to model multiscale systems. Any system that can be reduced to smaller components can be modelled as agent-based.

The ability to directly link real and virtual agents (cells here) in order to use data directly from experiments is a huge advantage. This is beneficial in both the early model construction stages and final validation (Walker and Southgate, 2009).

ABMs reflect the real world system much more realistically (in terms of the way communication occurs) as agents only have access to local information. Agents won't interact with all (or just any random agent). They only have access to their neighbouring agents, which might change during the course of the simulation as most agents move spatially (Macal & North 2010). They are also autonomous and active, able to move independently without external influence. They can be heterogenous and if needed, they can be programmed to process information, interact and exchange information with other agents and to make independent decisions based on their rule sets or even in situations of incomplete knowledge and information (Srblijinovi and Skunca, 2003; Crooks and Heppenstall, 2012). Most ABMs are 'off-lattice' discrete models (Walker and Southgate, 2009), allowing the agents to take up arbitrary locations and interact in arbitrary directions (Deisboek *et al.*, 2011).

Like any model, agent based modelling also has some limitations and challenges. Some of these are generic and can affect most modelling techniques, for example, epistemic uncertainty (caused by too many assumptions) or finding the right balance between abstraction and model realism (Read *et al.*, 2011).

Table 2.3: some of the advantages and limitations of ABM in comparison to other modelling paradigms such as *in vivo*, *in vitro* and *in silico* continuum models

| In comparison with: | Advantages of ABM | Limitations of ABM |
|--------------------------------------|--|---|
| In vivo & in vitro models | Less time consuming | Less realistic (more abstract) |
| | More cost effective | Harder to communicate model with other researchers and replication of model by others is more difficult |
| | Free from experimental complications such as contamination | Subject to model implementation errors & bugs in a computational environment |
| | Able to explore areas that we don't have access to experimentally | |
| Continuum in silico models | Able to incorporate heterogeneity | Computationally taxing due to large number of agents |
| | Decentralised system | Lack of detailed data to inform model mechanisms |
| | Simulation of emergent behaviour | Validation much more challenging through abductive methods (as opposed to deductive or inductive) |
| | Direct one to one mapping of virtual agent directly and cellular agent | |
| | Able to represent multiple spatial and temporal scales if needed | |
| | Able to incorporate other modelling paradigms (Hybrid potential) | |

One of the major limiting factors of an ABM is computational power. The more complex an ABM is the more computationally taxing simulating it becomes as there are more agents and rule sets to deal with, not to mention longer simulation times mean higher number of time points required for the model to analyse. Currently performance can be improved through parallel processing on a multi-cored CPU or HPC cluster or by implementing the model on a GPU (Richmond *et al.*, 2010).

Due to the large number of model parameters, model-generated data (per time point and spatial position) and rule sets, it can be difficult to pin-point artefacts or mistakes made in the initial model configuration (Srblijinovi and Skunca, 2003).

A sensitivity analysis can be used to identify critical parameters as ABMs can be highly sensitive to certain given initial parameters and any variations in their rule sets

(Couclelis, 2002). This property, coupled with the large number of model parameters, model-generated data (per time point and spatial position) and rule sets of an agent based model, means the results become 'fragile' (Srblijinovi and Skunca, 2003). This makes drawing conclusions from unexpected results difficult as they could just be down to parameter artefacts. An effective way to overcome this is to start with the simplest possible form of the model (null model) and gradually add complexity (Thorne, Bailey and Peirce, 2007). This will also aid in communicating the final model with others.

2.2.2.2 Components of a biological ABM

Typically an ABM is built up of three components that must be identified by the developer, agents, rules and the environment.

Agents

The defining characteristic of an agent is its ability to act autonomously. This means it should be able to behave and respond to situations independently without any external influence. Therefore all the agents behaviours should be governed by decisions it makes based on the information it senses from other agents and/or its environment (Jennings, 2000). This characteristic is the only universally agreed on vital requirement of an agent, however Macal and North list some essential characteristics that should be included from a practical modelling standpoint (Macal and North, 2005, 2010). These are:

Self-Contained: This means that the agent should have clear, well defined and easily identifiable boundaries ensuring it can be distinguished as a unique agent by other elements of the model.

State: Very similar to how a system has a state (made up of its state variables) an agent should also have a state that is variable over time depending on its situation.

Social: This means it should have dynamic interactions with other agents that influence its behaviour. Protocols for interactions should be defined, such as methods of communication, movement and response to environment.

Heterogenous: This means an agent can have as many attributes as required to reflect the full range of agent diversity. The variations in defined agent attributes can also be used by the agent and its neighbours in their decision making process.

Rules

The primary goal of an agent based model is to simulate the resulting consequences of agent behaviours, relationships and interactions between agents. This is achieved by imposing rules that define the required mechanisms for these behaviours, relationships and interactions. These rules are an extremely vital component of an ABM as ABMs are decentralised (there is no central authority) and agents only have access to local information (Macal and North, 2010).

Agent rules can take various forms, from the use of direct conditionals (if-else statements) to integration with other models. These models can potentially take on many form, for example: implementation of genetic circuit that drive agent response to particular inputs (Gorochowski, 2016) or finite element mechanobiological models used to define the agents and their response to external stimuli (Zahedmanesh and Lally, 2012), and more recently, the application of machine learning frameworks to ABMs in order to design truly adaptive agents (Rand, 2006).

Environment

As the name suggests, this is the environment the agent is in. It can be used purely as a spatial location matrix for the agents to locate themselves and each other, or it can contain resources or physical environmental cues specific to the modelled system. Agents will communicate with the environment they are in, and make decisions based on the cues and information received from the environment. Spatial awareness provided by the environment allows agents to identify and communicate with their neighbouring agents. For example in an ABM of cells growing in an *in vitro* environment, the cells would be agents, rules could include conditions for cell cycle progression and proliferation and the model environment would be an abstract representation of the culture environment either as a purely spatial structure or with specific biological information regarding the culture media.

2.2.2.3 Tools and methods for ABM implementation

There are many different tools and methodologies for implementing an ABM. In this section some of the most popular approaches used in agent-based modelling of biological systems will be looked at.

Existing ABM platforms

Table 2.4 lists some of the ABM platforms available ranging from basic modelling tools (NetLogo) to more complex tools that allow multiscale modelling (FLAME).

Table 2.4: Common ABM toolkits available (including biological applications)

| Software | Language | Operating system | An Example | Website |
|--|---------------------|---|------------------------------|---|
|  NetLogo | NetLogo | Java virtual machine (Version 5 or above) | | https://ccl.northwestern.edu/netlogo/ |
|  CompuCell3D | C++ | Linux Windows Mac OS X | (Zhang <i>et al.</i> , 2011) | www.CompuCell3D.org |
|  FLAME | C | Systems that support C Pthreads needed to run in parallel | (X. Li <i>et al.</i> , 2013) | www.flame.ac.uk |
|  SPARK | Java | Windows Linux (Mac OS X) | (An <i>et al.</i> , 2009) | http://www.pitt.edu/~cirm/spark/ |
|  LINDSAY Composer | Java Objective-C | Windows Linux Mac OS X | (Sarpe and Jacob, 2013) | http://lindsayvirtualhuman.org/?q=node/39 |
|  AgentBuilder | Java | Windows | | http://www.agentbuilder.com/ |
| | C C++ | Linux | | |

Having so many different software toolkits available for ABM implementation is a good in the sense that the model developer has a lot of choice and can implement an ABM without much computational background (especially beneficial in an interdisciplinary environment). However one major drawback is that different ABM are constructed using different methods and limitations imposed by the various software available (Srblijinovi and Skunca, 2003).

As mentioned previously, model communication is an area that ABM struggles in. This is mainly due to the fact that ABMs require a computer for implementation and the results of the model are achieved through simulation and multiple lines of code. The many

different toolkits available add to this problem. Also there is very little consensus on the language and format to be used when publishing ABM work (Grimm *et al.*, 2006) and this is often made worse with the lack of space in journals for coding (Thorne, Bailey and Peirce, 2007).

Grimm and colleagues propose a standard method for describing agent based models to overcome explanatory opacity. This is through a three stage system they refer to as ODD, Overview, Design concepts and Details. ‘Overview’ is further divided into purpose, state variables and scales and process overview and scheduling and ‘Details’ is also divided into three subsections, initialisation, input and sub-models. The idea with this method is to make communicating ABM easier and to allow for model replication and comparisons (Grimm *et al.*, 2006).

Another disadvantage to using these toolkits is that the modeller has very little control over the code and is limited to functions of the software. Constructing an ABM from scratch has the benefit of giving full control and flexibility to the model developer. If needed other external bits of code can be integrated and the model is extensible to multiscale when required.

Coding an ABM

Agent-based models are typically programmed using object-orientated code. Each agent type should be given its own class with its own attributes and methods.

The attributes can be static (for example, agent name) or dynamic (e.g. age of agent, memory or state). The agent methods can be used to define the way the agent behaves (e.g. the way it moves) and also to update the state of the dynamic attributes. For more complex systems (adaptive or otherwise) agent methods can be used to define rules that modify lower level behaviours. (Macal and North, 2005, 2010).

Agents are also linked externally with the way they interact (interaction and relationship rules) with other agents and the way they interpret and interact with their environment (environment rules) (Jennings, 2000)

Since agent communication occurs locally, a definition of what is ‘local’ to each agent is required. The way an agent calculates its local range is dependent on the model’s topology. A model could have different topologies linking separate levels of hierarchy together or it could have multiple topologies for different rules. Figure 2.7, illustrates the different topological set ups that could be used in defining agent relationships, social interactions and their ‘local’ range.

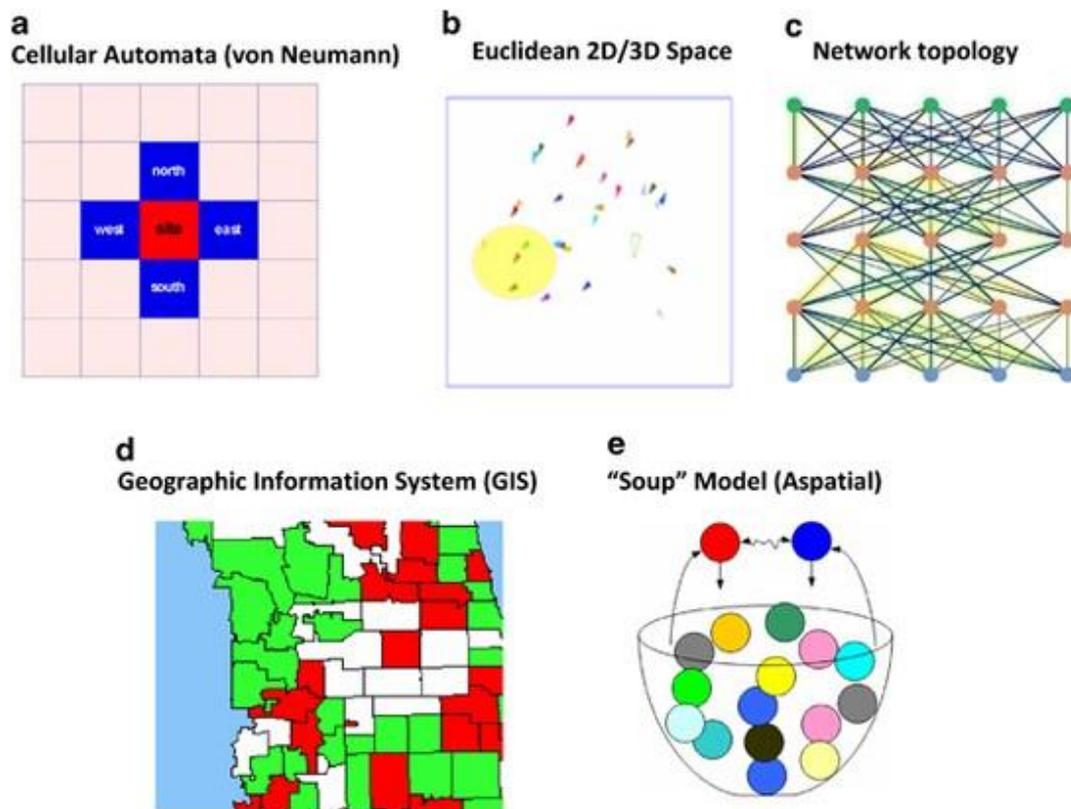


Figure 2.7: Topologies for agent relationships and social interaction. [Image reproduced from (Macal and North, 2010) copyright clearance obtained via RightsLink order number: 4946500539234]

Informing an ABM

One of the challenges of agent-based modelling is finding the right data to inform the model. ABMs rely on mechanistic models and so must be based on real biological system rules and mechanisms. Data like this can be obtained from relevant literature, via *in vitro* or *in vivo* experiments or clinically (Crooks and Heppenstall, 2012)

Finding suitable quantitative data amongst the sea of qualitative biological data is not easy. Some parameters may have to be derived from supplementary data or unpublished work, which can be done as long as the source is established reliable and accurate (Thorne, Bailey and Peirce, 2007). There might also be a need to adapt the model outline and settings in order to match the scope of the data available to use as input parameters.

In some cases the desired parameter requires experimental work that has not yet been carried out and has to be obtained from wet lab experiments (either by the modeller or through collaborations with experimentalists). If neither is possible or if the missing data is a result of insufficient experimental techniques, the missing data can be estimated and further

calibrated using appropriate techniques. In cases such as this sensitivity analysis should be used to assess the reliability and accuracy of any model predictions (Thorne, Bailey and Peirce, 2007; Read *et al.*, 2011)

Verifying an ABM

“Models of any type can only be as good as the data used to develop, calibrate and confirm them.” - (Bauer *et al.*, 2008)

Agent-based models don't just require data to inform them, experimental data is also required to validate the model. In order to avoid falling into a circular argument, results used for model validation should not be the same as those used to determine the initial model configurations. So, cellular-level-based experimental data should be used to inform the model (agent rules and interactions) and system-level, emergent experimental data can be used to validate the overall system level outcome of the model. The latter can be achieved either qualitatively, through visual validation or quantitative through dynamic results such as growth curves.

For ABMs, validation is done through abductive reasoning. Abduction, a term coined by Charles Peirce (Peirce, 1903), is a form of logical inference where probable conclusions are inferred from evident observations (i.e. inference of *a* as an explanation of observed *B*). Regardless of how likely the conclusion drawn is, these models are limited by their inability to fully exclude other probable causes. In comparison to deductive and inductive models, the falsification framework for abductive models is incomplete (Viceconti, 2011). There is no 100% validation proof. The current approach is to make comparisons between the model and the system it is modelling. For example, if a new element *a* is added to the model and *B* is observed and this *B* is also observed in the real system under very similar conditions, we suspect through abduction that *a* is responsible for observed effect *B* (Viceconti, 2011).

2.2.2.4 Techniques for model verification and validation (V & V)

Verification is the process of making sure the implemented model is solving the problem correctly, (matches its design) whereas validation is the process of making sure the implemented model is solving the correct problem (matches the real-world system) (Xiang *et al.*, 2005; Crooks and Heppenstall, 2012). Both are vital in determining whether a model and its results are “correct” (Sargent, 2004). “Correct” is used lightly here and in the context

of an ABM, refers to a model that can produce sufficiently credible solutions that can subsequently be used for predictions (Xiang *et al.*, 2005).

There are many different verification and validation techniques available (Balci, 1998) and often a combination of techniques is used in order to increase model confidence (Xiang *et al.*, 2005). However it is worth noting that obtaining absolute validity across all the possible domains of the model is costly, both in time and computational resources so instead the model developer should pick those techniques most relevant to their modelling paradigm, research question and intended model application (Sargent, 2004).

Verification can be sub-divided into static testing and dynamic testing (Fairley, 1976). Static testing is the steps taken to ensure the code is implemented correctly without executing the code and includes techniques such as code walkthrough and ensuring the structure and logic behind the code is correct (Sargent, 2004). Dynamic verification testing is checking the code through its execution. It includes techniques like trace analysis, input-output testing, pattern tests, boundary testing, debugging and calculation verification (Xiang *et al.*, 2005).

Listed below are some of the relevant V & V techniques that could be used for an ABM (Balci, 1998; Sargent, 2004; Xiang *et al.*, 2005)

Face validity

This is a subjective technique and involves consulting with the domain experts on whether the model is behaving reasonably at face value.

Tracing

Individual elements of the model can be followed to determine if the implemented logic is correct and resulting in an expected output.

Model-to-model comparison

As the name suggests, this is a comparison of the outputs of the developed model with other models. The models used here must be validated themselves and the modeller should be aware of any uncertainties or limitations in their validity.

Internal validity

Internal validity is relevant to stochastic models (often the case in biological models) and it involves running repeats of the simulation in order to determine the amount of random variability in the model. It is also worth noting that biological cells behaviour has an element of randomness and so wet-lab experiments should be internally validated too through repetitions of the experiment. This validity test accounts for aleatory uncertainty.

Historical data validation and Predictive validation

Comparison of the data collected at the system level with the ABM's emergent behaviour is common form of validation for agent-based modelling. This can be done in one or more stages. The minimum is to conduct the validation with historical data (here historical refers to the initial problem domain observations). This can be furthered using predictive validation which involves using the developed model to forecast an output or predict an emergent behaviour outside of the initial problem domain and then comparing this with the real system under the same conditions.

Uncertainty and Sensitivity analysis

Both these analysis are statistical techniques that relate variations in the models output to its inputs. They are important as the modelling process requires assumptions and abstractions to be made. This could be due to a lack of expert knowledge, performance concerns over including too much in the model or just simply down to some information being unnecessary (Xiang *et al.*, 2005).

Uncertainty analysis considers the effect of uncertainties in input values on the output (accounts for epistemic uncertainty) and sensitivity analysis investigates how sensitive the model outputs are to its inputs (Read *et al.*, 2011). Through these methods, parameters that have a significant effect in the model's behaviour can be highlighted. Extra care has to be taken to ensure these parameters are sufficiently accurate for use in the model. Also the relationship between the internal parameters and the model output should be the same as the real system.

2.2.2.5 Examples of ABMs of biological tissue

Some examples of ABMs of biological tissue include: ABM of sprouting angiogenesis (Shirinifard *et al.*, 2009), simulation of an immune response to influenza in lung tissue and

lymph node (Sarpe and Jacob, 2013), ABM of intracellular chemical interactions (Pogson *et al.*, 2006), simulating brain tumour heterogeneity (Zhang, Strouthos, *et al.*, 2009), cell sorting through various adhesion rules (Swat *et al.*, 2013), keratinocyte colony formation (Sun *et al.*, 2007, 2008; Adra *et al.*, 2010), simulating dynamic cluster patterns in brain tumour (Mansury *et al.*, 2002), epithelial tissue formation (Walker *et al.*, 2004, 2006) and many more. Table 2.5 lists some of these models according to the way they are built.

Epitheliome:

Developed by Walker *et al.*, (2004), this ABM predicts the emergent behaviour resulting from the interaction of epithelial tissue cells. The model is validated against a well-established *in vitro* epithelial cell culture system, showing a qualitative fit between the two. The model has been developed using object-orientated code in Matlab.

The model consists of two main agent classes, stem cells and transit amplifying (TA) cells. These classes each have 3 subsets, mitotic, post-mitotic and dead cells which stem and TA cells can change into. The main rules of the model dictate cell behaviour. This consists of rules defining the cell cycle, bonding, spreading, migration and apoptosis.

Communication between cells is handled via a global communication matrix that cells only access in order to gain information regarding their neighbouring cells. Cells wanting to transmit a certain message, post in the corresponding location of the communication matrix (row = agent sending message, column = agent receiving message). Once the message is posted, it is received by the relevant agent. The message is removed from the matrix and the receiving cell updates its internal parameters.

To compensate for any cell overlap due to cell movement and mitosis, a numerically-based physical correction model is applied after each iteration of the ABM. This artificially exerts a repulsive force proportional to the overlap of any pair of overlapping cells and an attractive force to correct any bonded cells that aren't touching.

This Epitheliome ABM was then extended to include an intercellular signalling mechanism (EGFR signalling pathway). A mathematical model was developed consisting of a number of equations representing aspects of signalling such as, ligand release and diffusion, ligand receptor binding and activated receptor trafficking and recycling. This model was incorporated into the existing Epitheliome model, assembled parallel to the cell monolayer. The cells are able to communicate with the mathematical model and the model is solved every iteration of the ABM.

The model has been used to examine the effect of plating density on tissue growth via autocrine/paracrine signalling and demonstrates a viable approach to combining ABM and mathematical modelling paradigms to model more complex cellular interactions. (Walker *et al.*, 2004, 2006)

Keratinocyte colony formation

Sun *et al.*, (2007) developed this ABM in order to explore some of the biological rules that govern keratinocyte self-organisation. The rules for the model are derived from literature and the model simulations were used to test hypothesis *in silico*. This was then shown to be supported by *in vitro* experimentation. They were able to identify parameters that were important for keratinocyte self-organisation and those that were not.

This model was implemented in FLAME (www.flame.ac.uk) and the model communication was based on X-machines framework (Coakley, Smallwood and Holcombe, 2006). In this framework agents communicate by reading and writing to message lists that are updated. Information such as the cell cycle, cell position, cell size and bonds are stored, communicated and updated this way.

Each cell is modelled as an agent and the culture dish and global factors like extracellular calcium are modelled as the environment. Rules included in the model are migration, proliferation (via cell cycle progression) and differentiation. The cell cycle rule involves the cells undergoing a “countdown” to when they can divide provided they are in the G₁ phase. There were two ways the cells will leave the cell cycle and enter G₀, contact inhibition or receiving differentiation signals.

Contact inhibition is implemented using the message lists. If the number of neighbouring cells is more than x , the cell will enter G₀. x depends on the cell type and exogenous culture environment. Cell motility rules are dependent on cell-substrate and cell-cell interactions. A physical rule is also implemented which is basically a movement applied to the agents corresponding to an attractive or repelling force between agents in the model to correct any overlap. At each time step (lasting 30mins) the agents first declare their location and type by outputting this information into the message lists. They then perform cell cycle related rules based on their internal conditions. Using the differentiation rules they decide if they want to change to another cell type or not and finally the migration and physical rules are executed based on the cells own internal state and immediate environment accessible through the message lists.

This ABM was extended by Sun *et al.*, (2008). This time it is designed to help understand the rules by which fibroblasts support keratinocyte colony formation. Two new agent types are added, proliferative human dermal fibroblast and differentiated human dermal fibroblast. Additional rules regarding the way these cells differentiate, proliferate and migrate are implemented. This model is validated against *in vitro* mono-culture and co-culture data.

3D Agent-based Tumour Model

Zhang *et al.*, (2007) developed a 3D multiscale ABM of a brain tumour. It is able to simulate gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. The environment is a 3D rectangular lattice that represents virtual brain tissue. Each site on the lattice has TGF α , glucose and oxygen tension values assigned. These represent the external chemical cues by normal distribution. There are two systems implemented in the molecular environment, EGFR gene-protein interaction network and cell cycle subsystem. Each agent has four layers, external space (with access to the lattice), cell membrane, cytoplasm and nucleus. Rules included in this model are migration and proliferation, quiescence and apoptosis and a biological on/off cell cycle switch.

3D Multiscale ABM of human epidermis

This model by Adra *et al.*, (2010) is very interesting in that it is built up of three interacting and integrated layers. Firstly a complex pathway simulator is used to simulate the expression and signalling of TGF- β 1 at the sub-cellular level. Then an adapted version of the ABM from the previous study done by sun *et al.*, (2008) captures the biological rules governing the cells in the human epidermis at the cellular level and finally a mechanical layer is used to implement the numerical physical solver which operates at the multi-cellular level. The model rules are informed from extensive research literature describing the signalling pathway and validated against literature again.

3D Multiscale ABM of clonal dynamics on epithelial homeostasis

Here, a 3D agent-based model constructed in FLAME by Lie *et al.*, (2013) is used to investigate the effect of three different hypotheses of clonal dynamics on epithelial homeostasis. The group were able to predict the long-term behaviour of colony dynamics

based on different hypotheses of stem cell generation of epithelium. They also identified the most reliable hypothesis out of the three.

The model is informed by literature gathered data. It includes biological rules such as cycle progression, taking into account contact-inhibited and confluence limited rules regarding mitosis using a checkpoint placed at the G1 stage of the cell cycle. Differentiation is modelled using mechanical interactions. Once the cells receive a net upward force (from neighbouring cells or the basement membrane), they are pushed into the next layer up which then initiates the differentiation process. This model was able to simulate growth and maintenance of the epidermis over three years and to trace every cell and its offspring. This is something that is impossible to achieve in the lab.

Table 2.5: Comparison of various agent-based models of biological tissue.

| | Agents | Environment | Mechanisms modelled: | Informed with: | Validated against: |
|--|--|--|---|--|---|
| (Mansury <i>et al.</i> , 2002) C++ & Repast | Tumour cell | 2D lattice contains nutrient supply levels | Migration Proliferation Apoptosis | Literature | 3D <i>in vitro</i> assay |
| (Walker <i>et al.</i> , 2004) MATLAB | Stem cells TA cells Dead cells Pre-mitotic Post-mitotic | 2D square substrate with modifiable calcium ion concentration levels | Cell-cycle Bonding Spreading Migration Apoptosis | Experiments and literature | <i>In vitro</i> epithelial cell culture |
| (Walker <i>et al.</i> , 2006) Multiscale in MATLAB | Stem cells TA cells Dead cells Pre-mitotic Post-mitotic | Same as Walker <i>et al.</i> , 2004 but integrated with a 3D numerical model | Same as Walker <i>et al.</i> , 2004 and EGFR signalling pathway | Literature | <i>In vitro</i> epithelial cell culture |
| (Sun <i>et al.</i> , 2007, 2008) FLAME using X-Machines framework | Stem cell TA cell committed cell Corneocyte P-HDF D-HDF | Culture dish including global factors like extracellular calcium levels | Cell cycle Migration Differentiation | Literature and experimental observations | <i>In vitro</i> scratch wound model <i>In vitro</i> single and co-culture Greens medium |

continued

| | | | | | |
|--|------------------------------|---|--|-------------------------------------|--|
| (Adra <i>et al.</i> , 2010) | Stem cell TA cell | 3D | Bonding Cell cycle | Literature | Literature |
| Multiscale Using Flame & COPASI | Committed cell Corneocyte | | Differentiation Migration TGF- β 1 Expression and signalling | | |
| (Zhang, Athale and Deisboeck, 2007) | Tumour cells | 3D rectangular lattice with assigned TGF $_{\alpha}$, glucose and oxygen tension levels | EGFR gene- protein interaction Cell cycle Migration Apoptosis | Literature | |
| (X. Li <i>et al.</i> , 2013) | Stem cell Progenitor cell | 3D | Cell cycle Differentiation | Literature derived hypotheses | <i>in vitro</i> tissue engineering human skin model |
| FLAME | | | | | |

2.3 Summary

Cancer is caused by abnormal, uncontrollable and relatively autonomous growth of cells with current popular treatments limited by their lack of specificity in only targeting cancerous cells. Stem cell therapy is an emerging solution and involves coupling the disease seeking nature of stem cells with advances in engineering in order to create targeted site specific therapeutic treatment. However exploration of the existing literature on the role of (unmodified) stem cells at the tumour site, highlights clear discrepancies regarding whether stem cells promote or inhibit tumour progression. These discrepancies can be attributed to the inconsistencies in experimental design across the various wet lab studies. Computational modelling can be used to explore the effect of these inconsistencies in addition to the validity of some of the proposed mechanisms of interaction between the two cell types. Agent-based modelling is bottom-up and discrete, making it ideal for this purpose as it simulates emergent cell population behaviour and patterns as a result of rules and characteristics defined at the single cell level. The aim of this this PhD is to explore the effect of stem cells alone on the growth of cancer cells in co-culture by using a combination of both experimental and computational methods.

Chapter 3

Materials and methods

A full list of materials, equipment and software utilised for both computational and experimental areas of this PhD is included in section 3.1.

Section 3.2 lists the experimental methods for cell culture and section 3.3 computational methods for the design, development and validation of a computational model.

3.1 Materials

1. MG-63 human osteosarcoma cell line (kindly donated by Dr G. Reilly, University of Sheffield)
2. Human embryonic stem cell derived mesenchymal progenitor hES-MP cells (kindly donated by Dr G. Reilly, University of Sheffield)
3. T75 and T25 cell culture treated polystyrene culture flasks (Thermo Fisher Scientific, UK)
4. Polystyrene 48 well plate (Starstedt Ltd)
5. Cell culture treated polystyrene 24 and 12 well plates (Thermo Fisher Scientific, UK)
6. PBS: Phosphate-buffered saline (PBS) tables, 1 tablet in 200 mL diH₂O at pH ~7.4 (Sigma, UK)
7. HBSS: Hanks Balanced Salt Solution without phenol red (Sigma, UK)
8. HEPES Solution: 1M at pH 7.0 – 7.6 (Sigma, UK)
9. FBS: Foetal bovine serum lot number 08F7674K (Gibco, UK)
10. MEM: Minimum Essential Medium Eagle (Sigma, UK)
11. Alpha MEM (α -MEM) with Desoxyribonucleotides and UltraGlutamin 1 (Lonza, Belgium)
12. PSG and PS: Penicillin (10,000 units), Streptomycin (10mg/mL) and L-Glutamine (200mM) solution (Sigma, UK)

13. Trypsin-EDTA solution (0.25%) containing: sterile-filtered, BioReagent, suitable for cell culture, 2.5 g porcine trypsin and 0.2 g EDTA and Hanks' Balanced Salt Solution with phenol red (Sigma, UK)
14. DMSO: Dimethyl sulfoxide (Sigma, UK)
15. Porcine gelatine type A (sigma, UK)
16. Fibroblastic growth factor basic recombinant human protein (FGF- β) 4ng (PeproTech, UK)
17. Resazurin sodium salt (Sigma, UK)
18. Trypan-blue (Sigma, UK)
19. Hoechst 33342 (Sigma, UK)
20. Celltracker Red (CMTPX) and Green (CMFDA) fluorescent probes (Thermo Fisher Scientific, UK)
21. PDMS SILGARD 184 silicone elastomer kit (Sigma, UK)
22. Everflex Aqua Mate aquarium silicon sealant (Sika Everbuild, UK)
23. Cell culture Thermo Forma 3311 incubator
24. Tecan infinite F200 PRO micro plate reader (Labtech, UK)
25. Nikon Ti-E microscope with incubation cabinet and the attachments:
 - Filter cube 49000: ET-Dapi
 - Filter cube 49002: ET-GFP
 - Filter cube 49008: ET-mCherry, Texas Red
 - Objectives: Plan Fluor 4x/0.13, 10x/0.30, 20x/0.45 and 40x/0.60

- Light: CoolLED pE-300 white

26. ImageJ (Image Processing and Analysis in Java) version 2 with the TrackMate (version 4) plugin

27. CellProfiler version 3.1.9 (cell image analysis software) (Carpenter lab at Broad Institute MIT, US)

28. MetaMorph acquisition analysis software (Molecular Devices, CA)

29. Lucidchart accessed from (<https://app.lucidchart.com/>) (Lucid, US)

30. Python version 3.7.4 with Spyder version 4 IDE and the following additional libraries:

- Numpy
- Scipy
- Matplotlib
- Seaborn
- Pandas
- Pickle
- Pingouin
- Statsmodels
- Uncertainties
- SciKit-learn (sklearn)
- SALib
- Time
- OS
- Copy

3.2 Experimental Methods

3.2.1 General cell culture

3.2.1.1 General cell culture conditions

All cell culture was performed under strict sterile conditions inside a cell culture hood under laminar flow. Cells were expanded in T75 cell culture flasks and kept in a humidified incubator at 37 °C with 5% CO₂. Basal media (BM) appropriate to each cell line (described in detail under section 3.2.1.2 and 3.2.1.3 for MG-63 and hES-MP cell lines respectively) was used and changed every 2-3 days. At approximately 90% confluency cells were either used or passaged.

To passage cells, the BM was discarded and cells were washed twice with PBS. 3 mL of Trypsin-EDTA solution (1 mL per 25cm² of culture surface area) was added and cells were incubated under standard conditions for 2-5 minutes (until they appeared visibly detached). Once detached, the trypsin and cell mixture was neutralised with 6 mL of BM warmed to 37°C. The mixture was then transferred to a 15 mL tube and centrifuged at 1,000 rpm for 5 minutes. The resulting supernatant was decanted carefully and cell pellet re-suspended in 1 mL of BM (or any other known volume), counted and seeded accordingly.

To manually count the cells, 20 µL of cell suspension was mixed with 20 µL of Trypan Blue in a 1.5 mL Eppendorf tube. Trypan blue stains dead cells dark blue, ensuring they're not counted. Using a haemocytometer under a light microscope, a cell count was obtained and then multiplied up (also taking into account any dilution factors) for an estimate on the total number of viable cells in the original cell suspension.

For seeding the required number of cells were added to the total known volume of media. The suspension was pipetted gently several times to ensure homogeneity and seeded slowly into desired well plate. Between every few wells, to re-homogenise, the suspension was pipetted again. The plate was then moved in a horizontal figure of 8 to evenly disperse the cells across the well plate and ensure homogenous seeding distribution. The cells were left to settle under the cell culture hood for 20 minutes before being transferred to the incubator.

To freeze the cells, the cell pellet was suspended in a solution of FBS with 10% DMSO with cell density of 1 x 10⁶ cells/mL. 1 mL vials of cells were frozen over-night (up to 24 hours) to -80°C in an isopropanol-jacket before being transferred to liquid nitrogen.

To recover, each vial was partially thawed in the water bath. 1 mL of fresh warmed (37°C) BM was slowly added to the partially thawed vial and the cells resuspended by gentle pipetting. The mixture was transferred to a 50 mL falcon tube and further diluted with 18 mL of warmed BM to diffuse out the DMSO. The cells were centrifuged, resuspended in 15 mL of BM and seeded in a T75 flask.

3.2.1.2 MG-63 cell line

MG-63 human osteosarcoma (bone cancer) cells from passage 8 to 12 were used as cancer models. They were passaged at a ratio of 1:4 every 5-6 days with a single media change in between. The basal media consisted of Sigma MEM (M4526) with 10% Labtech Foetal Bovine Serum (FBS) and 1% PSG (Penicillin, Streptomycin and L-Glutamine) warmed to 37°C prior to adding to cells.

3.2.1.3 hES-MP cells

Human embryonic stem cell derived mesenchymal progenitor cells (hES-MP) were used as a mesenchymal stem cell model. These cells, marketed as Cellartis, and derived from human embryonic stem cells were purchased from Takara Bio at passage 5 and used up to passage 10. They were seeded onto surfaces coated with 0.1% gelatine. To coat a surface with gelatine, the entire surface area was covered with 0.1% autoclaved gelatine solution (porcine gelatine type A) and left for 30 minutes. Excess gelatine was discarded and the surface was left to dry under the cell culture hood for a minimum of an hour before seeding.

Cells were seeding in a T75 flask at a density of 10,000 cells per cm² in 0.2 mL of BM per cm². The BM used for hES-MPs consisted of Lonza Alpha-MEM (α -MEM) with Desoxyribonucleotides and UltraGlutamine 1 supplemented with 10% Gibco HI Foetal Bovine Serum (FBS) and 1% PS (Penicillin and Streptomycin) with the addition of fibroblast growth factor-basic recombinant human (FGF- β) at a final concentration of 4ng/mL. Due to the long attachment and lag phase of these cells, initial media change after seeding was done at 48h-72hours and on alternate days thereafter. The cells were passaged or used after 6-7days.

3.2.2 Co-culture

3.2.2.1 General co-culture conditions

For co-culture both the MG-63 cells and the hES-MP cells were cultured in the same environment. Since the hES-MPs require more specialised conditions for growth, the shared growth environment used for both cells was that of the hES-MPs (mentioned above, 3.2.1.3). This included coating all wells used with 0.1% gelatine.

As preliminary experiments found the hES-MPs to have a longer lag phase than the MG-63s, they were seeded first. After 24hours the MG-63s were added. In order to visually distinguish between the two cell types, the MG-63 cells were stained with CellTracker Red (see 2.2.3.2) prior to seeding. Complete media changes were carried out at 48 and 96 hours after the addition of MG-63s.

3.2.2.2 Direct co-culture

For direct co-culture, both cells were seeded in the same well plate with shared BM. The hES-MPs were seeded first in 0.2 mL/cm² BM and incubated under standard conditions. After 24hours, MG-63s (stained with CellTracker Red) were suspended in 0.05 mL/cm² BM and added to the well containing hES-MPs (making a total of 0.25 mL/cm² of BM per well). At both seeding points, the well plate was gently swirled in a horizontal figure of 8 in order to ensure homogeneous seeding distribution. For consistency across both indirect and direct co-culture experiments, a further 0.125 mL/cm² of BM was added after 12-18hours.

3.2.2.3 Indirect co-culture

For indirect co-culture, the requirement was that the cells would share the same BM but be separated with a barrier so that they would not come into direct contact with each other. The barrier was created using polydimethylsiloxane (PDMS), a hydrophobic silicone elastomer that is bio-compatible.

The PDMS was made by combining the silicone elastomer and cross-linker at a ratio of 10:1. The mixture was mixed thoroughly for 2 minutes and placed under vacuum several times, each lasting 10-20 minutes in order to remove bubbles. Measurements of approximately 0.2 g/cm² were pipetted into the desired sized well plate in order to create

disks of equal thickness. This was vacuumed a further two times before being left to completely cure at room temperature for 48 hours.

The disks were removed and using a stencil, a thin 2mm strip was cut across the centre of the disk. This strip of PDMS was transferred to a fresh plate and sealed to the bottom of the plate using aquarium sealant (biocompatible) and left to harden for 48 hours. To sterilise the plates, they were washed twice with 70% IMS and again with sterile PBS and left under UV light inside the cell culture hood for 1 hour.

To seed hES-MP cells were suspended in 0.25 mL/cm^2 of BM and added to the left side of the barrier. The seeding was done slowly in drop by drop manner moving across the seeding area. After 24 hours the MG-63 cells stained with CellTracker Red were added in the same manner to the right side of the barrier. after 12-18 hours, once the MG-63 cells were fully attached, a further 0.125 mL/cm^2 of BM was added to the well in order to cross the barrier and allow mixing of BM across both sides of the barrier.

3.2.2.4 Seeding density and ratios

Initial cell numbers used for seeding were kept consistent across all conditions by using seeding density based on actual surface area and not cells per well (e.g. to account for reduced surface area due to barrier inserted for indirect co-culture experiments). Using the recommended seeding densities for both cell types and preliminary rate of growth experiments as a guide, a seeding density of $12,000 \text{ cells/cm}^2$ was used. For mono-culture experiments, this was calculated directly based on the selected well surface area and for indirect co-culture, the surface area across the bottom of the plate on each side of the barrier was used. For direct co-culture, $6,000 \text{ cells/cm}^2$ were seeded for each cell type to make up a total seeding density of $12,000 \text{ cells/cm}^2$.

To compare the effect of increasing the number of initial stem cells to cancer cells (and for use in predictive validation of ABM), in addition to the initial 1:1 ratio, 3 other ratio configurations were also seeded. Table 3.1 shows the 4 ratios and their resulting seeding density for each cell type. The total seeding density in co-culture was kept constant at $12,000 \text{ cells/cm}^2$ for all ratio configurations. Figure 3.1 is a visual representation of the experimental set up for one of the ratios (1:2) across the various conditions.

Table 3.1: Seeding density of hES-MP and MG-63 cells under the 4 different seeding ratio configurations (1:1, 1:2, 1:3 and 1:4) and seeding conditions (mono-culture, direct and indirect co-culture).

| Ratio MG63:hES-MP | Mono-culture ($\times 10^3$ cells/cm ²) | | Direct co-culture ($\times 10^3$ cells/cm ²) | | Indirect co-culture ($\times 10^3$ cells/cm ²) | |
|----------------------|---|--------|---|--------|---|--------|
| | MG-63 | hES-MP | MG-63 | hES-MP | MG-63 | hES-MP |
| 1:1 (R1) | 12 | 12 | 6 | 6 | 12 | 12 |
| 1:2 (R2) | 8 | 16 | 4 | 8 | 8 | 16 |
| 1:3 (R3) | 6 | 18 | 3 | 9 | 6 | 18 |
| 1:4 (R4) | 4.8 | 19.2 | 2.4 | 9.6 | 4.8 | 19.2 |

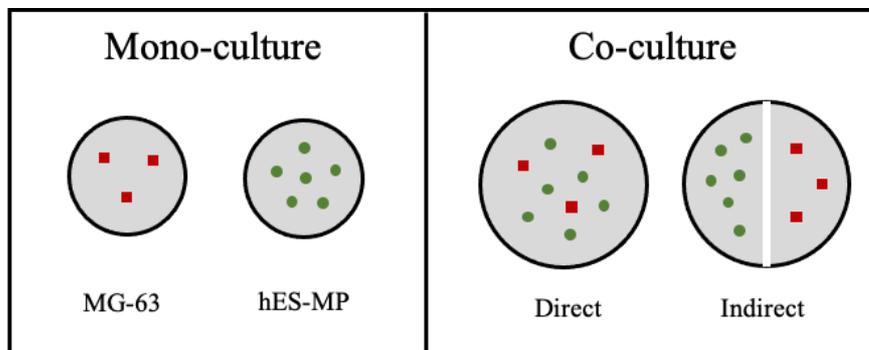


Figure 3.1: Diagram showing the experimental set up across all seeding conditions for seeding ratio configuration of 1:2. MG-63 cells are depicted as red squares and hES-MPs as green circles. The PDMS barrier is depicted using a white line. All well plates are shown from a birds eye view.

3.2.3 Cell staining and microscopic imaging

3.2.3.1 Hoechst 33342

Hoechst dyes are cell-permeable DNA stains that can be used in live cell imaging. Once bound to dsDNA, they emit bright blue fluorescence on excitation by ultraviolet light. The specific Hoechst stain used was Hoechst 33342, trihydrochloride, trihydrate, purchased from ThermoFisher (cat.no H3570). Figure 2.2 (top) shows its molecular structure and fluorescence spectra (excitation/emission: 350/461).

Staining solution was prepared by diluting the stock solution (10mg/mL) in sterile HBSS warmed to 37°C to a working concentration of 5 μ g/mL (e.g. 1 μ L stain in 2 mL buffer). The staining solution was added to adhered cells after aspirating the media and left on for 5 minutes incubated at 37°C. To image, staining solution was replaced with serum

free media and imaged with Nikon Ti-E fluorescence microscope using ET-DAPI filter cube 49000 (excitation 350nm; emission 460nm) exposed at 500ms.

3.2.3.2 CellTracker

In order to differentiate between MG-63 and hES-MP cells in co-culture, CellTracker fluorescent probes were used to label each cell population prior to seeding. CellTracker fluorescent probes are designed to freely pass through the cell-membrane. However once inside the cell, they are transformed into cell-impermeant products. At this point the stain cannot be transferred to neighbouring cells and is only passed onto daughter cells. This makes the stain ideal for tracking over time across several generations and for labelling separate cell populations in co-culture.

CellTracker Red (CMTPX) and CellTracker Green (CMFDA) from ThermoFisher were used. Figure 2.2 (middle and bottom) shows the respective molecular structure and fluorescence spectra of each of the stains.

The trackers were loaded onto adherent cells at equal to or less than 90% confluency, cultured in a separate T25 flask. To prepare the stains, each 50 µg vial of lyophilized powder was dissolved in sterile DMSO to a stock concentration of 10 mM (7.29 µL and 10.76 µL of DMSO for Red and Green respectively). This was then diluted in warmed serum free media to a final working concentration of 5 µM (2.5 µL per 5 mL of media). The flask culture media was replaced with 5 mL of CellTracker working solution and incubated at standard cell culture conditions. After one hour, the CellTracker working solution was removed and the cells gently washed with sterile PBS 5 times, swirling the flask for at least 30 seconds each time. Cells were detached and centrifuged twice, washing with PBS in between to ensure all residual working solution is drawn out. The appropriate seeding protocol was then followed to seed the cells in co-culture.

The cells were imaged separately with Nikon Ti-E fluorescence microscope using ET-mCherry filter cube 49008 (excitation 560nm; emission 630nm) for cells stained red and ET-GFP filter cube 49002 (excitation 470nm; emission 525nm) for cells stained green. The two grayscale images were colourised and merged using ImageJ (Fiji) (Schindelin *et al.*, 2012). Both stains were used mainly for qualitative imaging purposes, for quantitative analysis, in order to simplify the experiment and also minimise the number of manipulations carried out on the sensitive hES-MP cells, only the MG-63 cell population was stained.

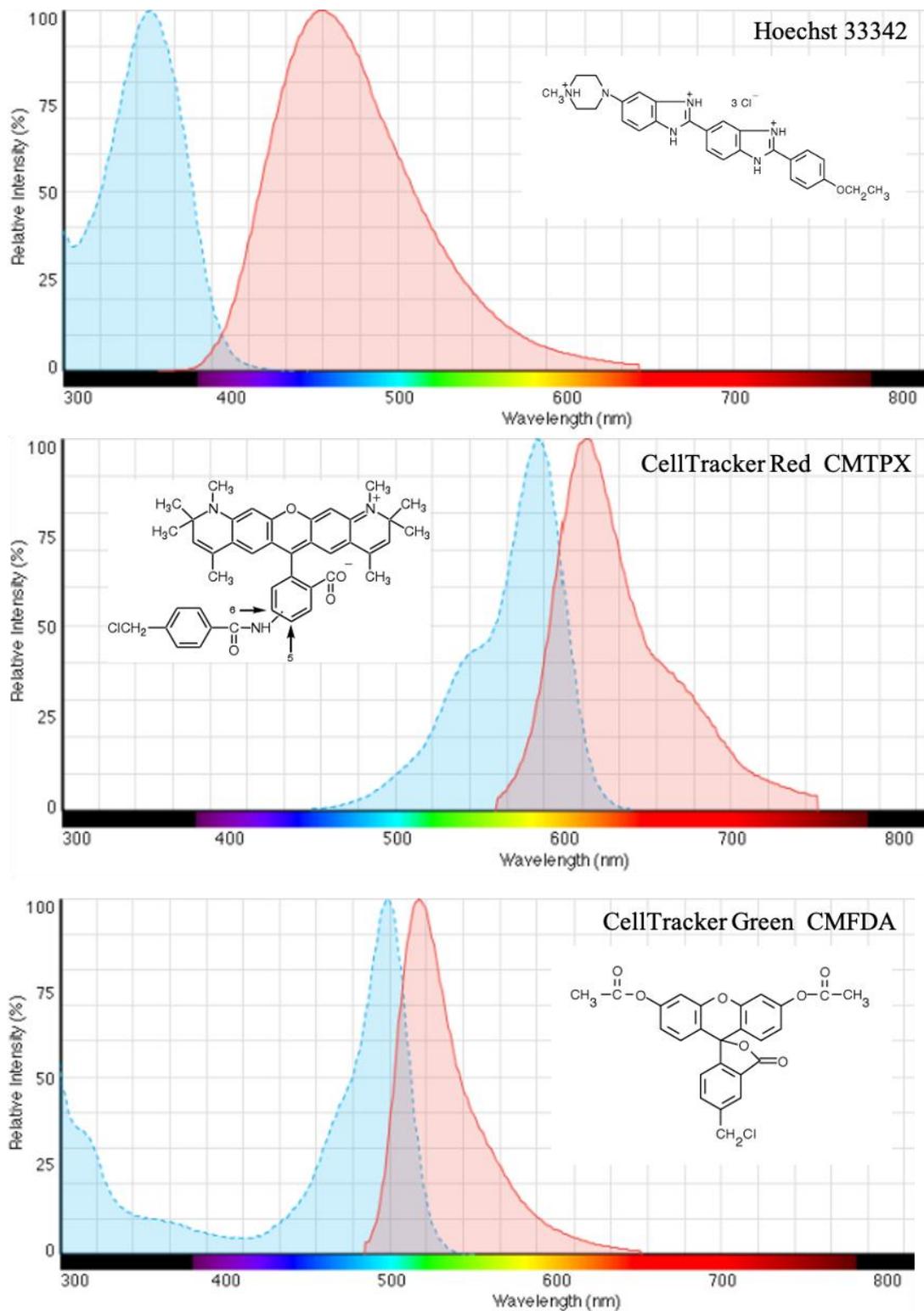


Figure 3.2: Molecular structure and fluorescence spectra of the three stains used. Top: Hoechst 33342, middle: CellTracker Red CMTPIX and bottom: CellTracker Green CMFDA. [Image taken from www.thermofisher.com].

3.2.3.3 Time-lapse

In order to obtain single cell data to inform the ABM, time-lapse images were taken of both cell types seeded in mono-culture for a duration of approximately 4 days. As cell behaviour can be affected by cell density, the cells were seeded in three different seeding densities, low (5,000 cells/cm²), medium (10,000 cells/cm²) and high (20,000 cells per cm²). Four separate wells were seeded for each seeding density and each cell type on a single 48 well plate. Blank media was added to the wells on all four edges of the well plate to reduce the effect of variations in evaporation and thermal changes across the plate while under the microscope for long periods of time. Cells were stained with nuclear stain Hoechst 3342 (see 3.2.3.1) prior to seeding. After seeding they were left in the incubator for 24 hours before setting up for time-lapse imaging.

To set up, the cell culture media was replaced with fresh BM containing HEPES buffer at a final working concentration of 25 mM (25 μ L per 1 mL). HEPES was used to maintain the pH of the culture media. The plate was secured and sealed inside the micro-environment chamber placed inside the incubation cabinet of the microscope. The incubation cabinet was maintained at 37°C with 5% CO₂, with humidity levels, temperature and CO₂ of the micro-environment also monitored.

Each well was imaged at two randomly picked x, y positions with the attached Nikon Ti-E fluorescence microscope in both brightfield and blue fluorescence channels. The stage positions were saved and continuous focus was maintained using the perfect focus system. Each saved position was subsequently imaged every 20 minutes. In between imaging, all channels were closed to keep the cells in the dark and reduce the effects of photo toxicity on the cells and photo bleaching on the stain. After 48 hours of imaging, the culture media was replaced with fresh BM and HEPES and imaging was resumed using the same stage positions.

To process the large amount of data produced, first the brightfield and blue fluorescence images were merged and the composite images stacked into smaller manageable groups. The TrakeMate plugin (Tinevez *et al.*, 2017) for ImageJ was used to extract the cartesian co-ordinates per cell per time point. To ensure accuracy and reliability of data, this process was semi-automated, whereby the initial cell detection and tracking was done by TrackMate and then subsequently checked and corrected manually for each cell.

This data was imported into Python for further organisation, analysis and visualisation. Object-orientated code was used to organise the data, linking each cell to itself

at timepoint -1 and timepoint +1. This was used to assess single cell motility and track lineage from the start of time-lapse to the end or when the cell was no longer in frame.

3.2.4 Cell viability

3.2.4.1 Resazurin reduction assay

Resazurin reduction assay (also known as alamar blue) is used as a means of assessing cell viability for cells grown in mono-culture. The resazurin sodium salt is reduced to resorufin by cell metabolic activity. The colour of the media containing the salt solution changes from a non-fluorescent blue to a fluorescent pink, the intensity of which correlates with metabolic activity.

To perform the assay, all culture media from cells were removed and a known volume of resazurin working solution was added. To make a working solution, 10% resazurin stock solution (1mM resazurin sodium salt in DiH₂O) was mixed into pre-warmed basal media appropriate for each cell line. In order to determine optimal volume of working solution, a preliminary run was carried out with cells at approximately 90% confluency. Various volumes of working solution was added and the smallest volume that didn't turn fully pink was selected as the volume to use. This was to ensure any plateauing of the results would not be due to not using enough of the resazurin solution. As this assay is light sensitive, the plates were then covered in aluminium foil and incubated under standard cell culture conditions for 4 hours. Then 200 μ L from each well was transferred to a 96-well plate in triplicates, ensuring a gap of at least a single well from either direction is left in order to prevent excitation of neighbouring wells. 200 μ L control sample of the working solution was also added. The 96-well plate was then read on a Tecan infinite 200-pro plate reader at excitation wavelength of 540 nm and emission wavelength of 590 nm. The final results were obtained by subtracting the control sample from the average of the triplicates.

As previously mentioned, the fluorescent intensity correlates with cell metabolic activity. The higher the intensity, the larger the number of cells. In order to determine quantifiable values for cell numbers a standard curve was produced. To do this, a range of known values of cells were seeded in separate plates (from 1,000 to 100,000 cells) and after 8-12hours resazurin reduction assay was carried out. Obtained readings were plotted against known cell numbers in order to produce a standard curve.

3.2.4.2 Manual cell counting protocol

Manual cell counting was used to obtain actual cell numbers across 5-7 days of cell growth in each of the mono-culture and co-culture experiments. Cells were imaged and counted before and after detachment. In order to aid the accuracy and automation of the counting process, the cells were stained with Hoechst 33342 (see 2.2.3.1) before imaging.

To detach cells, Trypsin-EDTA solution was applied to each well at 50 $\mu\text{L}/\text{cm}^2$ and incubated under standard conditions for 1-5 minutes. During this time, approximately every minute, wells were checked under the microscope for detachment and neutralising solution was added to inhibit the enzyme as soon as cells were seen to detach. Extra care was taken to not shake or vigorously move the wells during the trips from incubator to microscope and hood in order to prevent cell clumping.

The neutralising solution was made by adding 20% FBS to HBSS. HBSS was used in replacement of media containing phenol red to improved fluorescent imaging and stain intensity. FBS acts as a trypsin inhibitor and cell protection agent and was added at double the standard concentration commonly used, allowing for a smaller volume of neutralising solution to be added per well and in turn a smaller final cell suspension volume, increasing the accuracy of the cell counts at low cell densities. The resulting cell suspension was transferred to a 1.5 mL Eppendorf tube in preparation for counting. Wells were checked under the microscope to ensure all cells were detached and transferred. On day 4 and 5 of the experiment, due to high cell numbers, the cell suspension was diluted 2 and 3 fold respectively.

This entire process was done at end point which meant multiple repeats per well per timepoint were seeded for each of the conditions. At each timepoint there were 4 repeats per condition and all were detached first before any counting was done.

For each timepoint, once all the cells were transferred from the wells to Eppendorfs, counting began. For this, each sample was pipetted gently a few times against the Eppendorf wall to create a clump free homogenous cell suspension. 10 μL of this suspension was then added to one side of a clean dust free haemocytometer chamber and a separate 10 μL to the other side of the haemocytometer. A total of 8 areas from both sides of the haemocytometer were imaged using Nikon Ti-E fluorescence microscope in brightfield, blue fluorescence and red fluorescence.

Once all the samples were imaged, they were post processed and analysed using CellProfiler3 (Hyland, 2007). For this, 3 pipelines were created; mono-culture cell count,

red stain intensity measure and threshold calculation and co-culture cell count. In all 3, all cells stained with Hoechst were first identified by applying Otsu thresholding method to the blue fluorescence image. For standard mono-culture experiments, the output and object count from this alone was enough to obtain cell numbers.

*Table 3.2: Volumes of Hoechst, trypsin and neutralising solution used in the each well plate for detachment and manual counting. Also shown are the final cell suspension volumes per timepoint. *for each side of the barrier.*

| | Surface Area (cm ²) | Hoechst staining solution (uL) | Trypsin (uL) | Neutralising solution (uL) | Final cell suspension including dilution (μL) | | |
|-----------------------------|---------------------------------|--------------------------------|--------------|----------------------------|---|-------|-------|
| | | | | | Days 1-3 | Day 4 | Day 5 |
| 48 well plate | 0.64 | 80 | 32 | 32 | 64 | 128 | 192 |
| 24 well plate with barrier* | 0.80 | 100 | 40 | 40 | 80 | 160 | 240 |
| 24 well plate | 1.91 | 250 | 95 | 95 | 190 | 380 | 570 |
| 12 well plate with barrier* | 1.68 | 200 | 85 | 85 | 170 | 340 | 510 |
| 12 well plate | 3.80 | 500 | 190 | 190 | 380 | 760 | 1140 |

For co-culture a mask was created using the objects identified in the blue channel and applied to their respective red fluorescence images. The intensity distribution across the masked areas of the red fluorescence image was then measured for all mono-culture MG-63 images taken at each time-point. A minimum threshold value was determined using median minus the median absolute deviation (MAD). This was applied to the red fluorescent images of the co-culture experiments of the same timepoint to differentiate between the stained MG-63 cells and unstained hES-MP cells. It was important to calculate the minimum intensity threshold for the red fluorescence images for each timepoint in each experiment separately to account for variances in initial stain loading and stain fading due to photobleaching and cell division in addition to naturally over time.

All results were visually superimposed as outlines onto the brightfield image and manually checked, removing any incorrect objects identified and any objects crossing two of the image edges (as per standard haemocytometer protocol). For each sample, the average

of the 8 squares were taken and multiplied by the cell suspension volume and area of image measured (in mm) and 0.1 (haemocytometer chamber depth in mm) to obtain the total number of cells per sample.

3.2.5 Statistical analysis of biological data

All experiments were repeated 3-5 times with at least 3-4 replicate samples per condition. All plots show the mean \pm standard deviation unless otherwise stated. Statistical tools appropriate to each data set were used to analyse the biological data. Tests were done in python using various libraries (full list of all tests, python libraries or code utilised can be found in appendix (8.1)).

Grubbs test for outliers was used to identify and remove any significant outliers. For small data sets ($n < 20$) Shapiro-Wilk test was used to test for normality. Normality of larger data sets were assessed using D'Agostino and Pearson's omnibus test. Q-Q plots and histograms were also used as a visual tool for assessing normality. Levene's test was used to test for equality of variances.

Statistical difference between groups of normally distributed data was assessed using one or two way ANOVA followed by Tukey HSD post-hoc test for pairwise comparison. For non-parametric data, Kruskal-Wallis test was used, followed by pairwise analysis with Mann-Whitney and Wilcoxon rank sum tests. Unless otherwise stated, in all tests, significant difference was considered if $p < 0.05$.

Linear regression models were used for analysis of rate of growth. If needed, Box-Cox transformation was used in order for the model to meet the required assumptions. Residuals were tested for normality using Anderson-Darling test, equal variance using Goldfeld-Quandt test and linearity using Harvey-Collier multiplier test. For comparison between rate of growths, ANCOVA was used with time as the co-variate. Significant differences were first assessed for the slopes, then the intercepts.

3.3 Computational Methods

3.3.1 Modelling process

The steps taken to design, implement and develop the agent-based model were based on the CosMoS process (Andrews *et al.*, 2010). CosMoS (Complex Systems Modelling and Simulation), outlines a general work flow (for exploring, understanding and describing complex system behaviours and engineering them in simulation). Figure 3.4 depicts the relationship and flow of information between the various elements and highlights the iterative nature of the framework. For the purpose of designing an agent based model of interacting cells, two iterations were used, first the behaviour of individual cells at the single cell level and in mono-culture were explored and captured in an initial ‘null’ model. This model was developed, calibrated and validated, before moving onto the second iteration of the CosMoS process in order to further explore, model and test hypothesis regarding cell to cell contact.

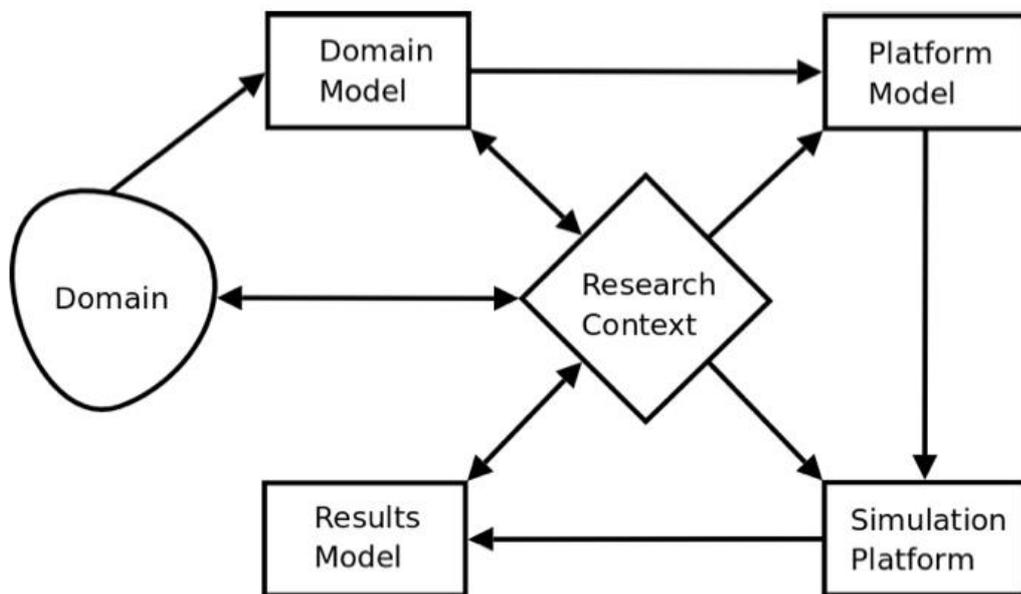


Figure 3.3: CoSMoS process of model development presented by Andrews *et al.*, (2010). The process is centred around the research context which covers the overall scientific question, motivation and requirements for validation. The domain is the scientific field under study and the domain model involves scientific understanding of the domain prior to any simulation. The platform model is where the domain model is used to design and implement the model for the simulation platform which in turn is the encoding and building of the model. The result model is where the simulation output is explored, validated and compared against the initial research context. As shown in the diagram, the research context can be referred to at any stage of the modelling process (Andrews *et al.*, 2010).

3.3.2 Domain model

The design of the domain model involves scoping of the biological domain to gain scientific understanding within the context of the research question. The level of abstraction, granularity and argumentation is decided here. A success criteria is also identified with which the simulation model can be validated against.

Figure 3.5 outlines the ABM work flow emphasising the need for two levels of domain exploration, single cell level to inform the model and multi-cell to validate the model emergent behaviour against.

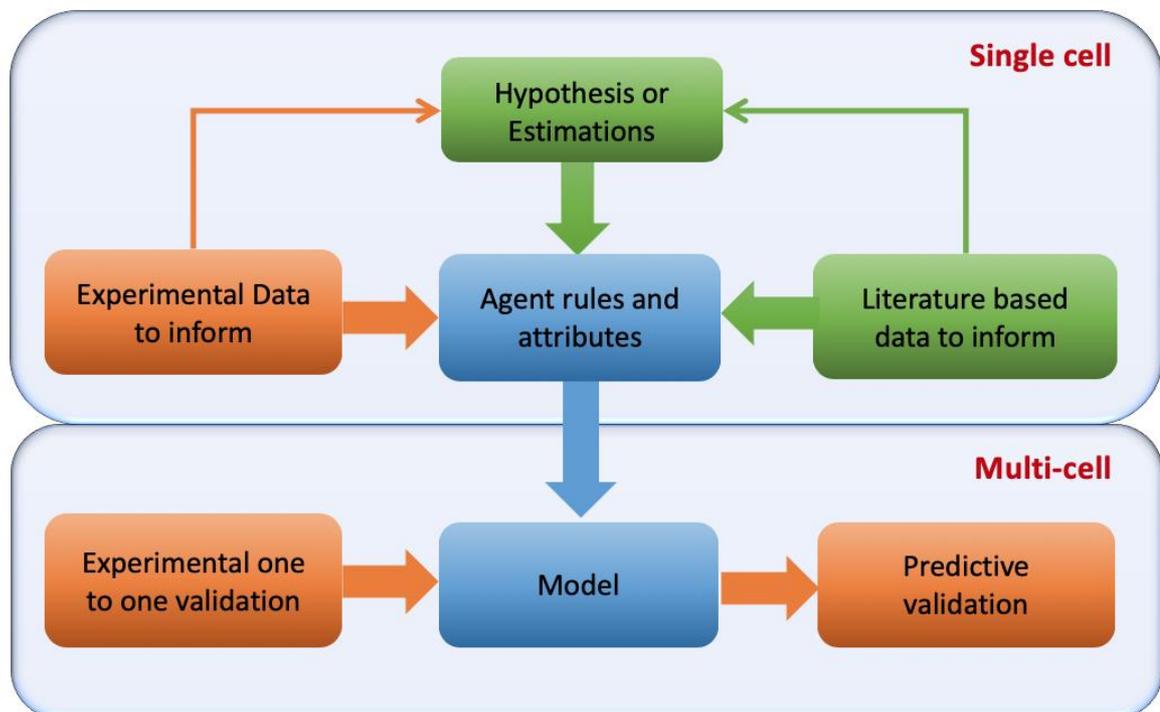


Figure 3.4: The ABM work flow designed and followed showing the integration of *in vitro* experimental data with *in silico* modelling approaches in order to both inform and validate the model. Cells are modelled as agents and the model design and implementation is carried out strictly at the single cell level whereas model validation is done at the multi-cell level. Two fold validation is done by coupling experimental one to one validation with predictive validation.

3.3.2.1 Analysis of single cell data

For the design of an ABM with cells modelled as agents, domain model scoping was at the level of individual cells. In order to achieve this, analysis was carried out on data gathered from time-lapse imaging (methodology explained in 3.2.3.3).

Statistical analysis was applied in order to explore relationships between the parameters of interest and identify possible parameter ranges. Where appropriate the data was fitted to existing probability distributions or regression models. Computationally the results gathered here were stored using the pickles library in python for direct transfer to the simulation platform and documented in a table of parameters.

Emphasis was made to gather model parameters directly from in house wet-lab experiments in order to establish continuity with multi-cell data gathered to validate the model. However when this was not possible, the parameter of interest was referenced from existing literature.

3.3.2.2 Analysis of multi-cell data

Lab derived growth curves of the cells grown in mono-culture and co-culture under various starting conditions were used to establish success criteria with which the emergent model output was validated against. Cell growth was measured by manual cell counting every 24 hours.

The data collected here was categorised into four separate groups, data used to calibrate the ‘null model’ (see 3.3.4.4), data to validate the ‘null model’, data to validate the ‘contact model’ and data for further predictive validation of the ‘contact model’ (see 3.3.4.5).

3.3.3 Model design and development

Continuing the CosMoS process, the translation from domain model to results model involves platform model and simulation platform, in other words, the design, implementation and encoding of the ABM based on the domain model and research context.

3.3.3.1 Model implementation

The model was designed using objected oriented code and key design elements were documented using standard UML notation. Python (ver. 3.7.4) accessed via the Anaconda package manager, was used to encode the ABM using Spyder IDE. A full list of libraries utilised can be found at the start of this chapter.

In the model, each cell type is defined by its own class. Individual agents are modelled as object instances of their respective class and are stored in python dictionaries (hash tables), identifiable by unique agent IDs. The dictionaries are stored using an environment class that is created on model initialisation. On each iteration of the model, agents are accessed in random order and prompted with a series of agent rules. Agent interactions with neighbouring cells were made practical by giving each agent a dictionary store of their local environment unique to them at any given point in time.

As mentioned earlier in this chapter, the model development was carried out in two separate phases referred to as ‘null’ and ‘contact’ model respectively. The focus of the former was to develop a baseline model of cell growth and motility under normal conditions (and in mono-culture). The latter involved implementation of various cell to cell contact rules and scenarios onto the baseline model, with each scenario documented separately as unique model version.

3.3.3.2 Model verification and testing

Verification or internal validation of the model was carried out at each stage of the development process. For changes made at the platform model, the simulation results were compared against the expected output for a fixed input (often manually calculated) using unit testing. For changes made at the simulation platform (e.g. model optimisation), direct comparison was made between the model output from before the change was made and after. This was made possible by fixing the random number generator seed. The comparison was done numerically and or visually where appropriate.

3.3.3.3 Simulation output

The emergent output of the ABM simulations was changes in agent number over time representing cell growth. The number of agents at each model iteration was stored in a 2D array and used to plot the growth of agents over time. Additionally, the environment, with a log of both alive and dead agents, was saved as a pickle object. This was reloaded for further analysis when needed. At the ‘null model’ stage, the number and nature of agent to agent contacts was also logged and produced as a means of analysis in preparation for the ‘contact model’.

Although the model was 2D, there was the option to visualise the data in both a 2D plot and or 3D projection. With graphical simulation as one of the more computationally taxing components of the model, an option to produce plots every n^{th} iteration or not at all was also available.

3.3.4 Model exploration

Model exploration was essential in initially establishing a reliable and robust ‘null’ model using calibration, uncertainty and sensitivity analysis techniques. This model was then used to explore the effect of various cell to cell contact hypothesis using validation techniques.

3.3.4.1 Aleatory uncertainty analysis

Aleatory uncertainty is one of two causes of uncertainty in a computational model. It is the result of inherent stochasticity in the system introduced by both the biological and simulation domain model (Read *et al.*, 2011). Therefore, in order establish representative and reliable simulation results, it is important to run replicates of the simulation. Intuitively, the larger the number of replicates (referred to as sample size from here on) the lower the effect of this uncertainty. However ABM with large numbers of agents can be very computationally taxing to run, therefore deciding on an appropriate sample size is a balancing act between the computational resources available and mitigating the effect of uncertainty analysis.

The following methodology was used to establish the relationship between sample size and aleatory uncertainty. First the coefficient of variation, defined as the ratio of standard deviation to mean, of a range of sample sizes (drawn randomly from a pool of samples) was calculated and plotted against their respective sample size, in order to visualise the sample number at which convergence occurs. This is used to identify a sensible range of sample sizes to base further analysis on as it requires a relatively low number of model runs (max run = max sample size).

Further analysis was based on the technique developed by (Read *et al.*, 2011). This was implemented by running 20 sets of simulation results per sample size are run (e.g. for a sample size of 50 this means 1000 simulations), the median response distribution per set was calculated and then the first set’s distribution was compared with the distributions of sets 2-20 in turn using the Vargha-Delaney A-Test (Vargha and Delaney, 2000). This is a non-parametric test that returns the probability that a randomly selected sample from one

population is larger than another with results ranging from 0-1. A result of 0.5 indicates no difference. Therefore the smallest sample size that settled around the 0.5 mark across all 20 sets was selected.

3.3.4.2 Epistemic uncertainty analysis (robustness analysis)

A lack of knowledge regarding certain parameters or abstraction of the biological domain results in epistemic uncertainty. Where possible model parameters and design were based on scoping done at the domain model stage, however this was not always practical. Single parameter analysis was carried out on decisions that could not be validated experimentally in order to understand the ABM’s robustness to any alterations in these parameters.

Single parameter analysis, also referred to as local sensitivity analysis, was carried out using a one-factor-at-a-time (OFAT) approach, with the factor being the parameter under investigation. The parameter was altered across a sensible range. The Vargha-Delaney A-Test (Vargha and Delaney, 2000) was used to contrast the altered model’s simulation distribution with that of the original model referred to as the ‘baseline’ model. A-test scores around the 0.5 mark indicated little to no difference caused by the change in parameter. Scores deviating in both the positive and negative direction from 0.5 indicated increasing differences in the two distributions and higher model sensitivity (table 3.4).

Table 3.3: The relationship between A-test scores and various magnitudes of difference between the two populations (Vargha and Delaney, 2000; Read et al., 2011)

| Difference | Large | Medium | Small | None | Small | Medium | Large |
|--------------|-------|--------|-------|------|-------|--------|-------|
| A test score | 0.29 | 0.36 | 0.44 | 0.50 | 0.56 | 0.64 | 0.71 |

High sensitivity of the model to parameter changes (within the plausible biological range of the parameter) indicates a need for further justification of the parameter value chosen. Robustness analysis was coupled with calibration in order to gain this justification and increase confidence in the model predictions.

3.3.4.3 Global sensitivity analysis

Sensitivity analysis is useful for understanding the relative sensitivity of the simulation’s output to its inputs. OFAT investigates the effect of changes in each parameter individually

and does not account for relationships between parameters and the compound effects that might occur. One method to account for this is variance based global sensitivity analysis, also referred to as Sobol's method (Sobol, 2001). Here the sensitivity of each input is represented by 3 types of sensitivity indices

1. First order index (S_1): A measure of a single input parameter's contribution to the model output variance.

2. Second order index (S_2): A measure of the interaction of the input paired with other inputs contributing to the model output variance.

3. Total order index (S_T): A measure of a model parameter's first order and all higher order interactions contributing to the model output variance.

The SALib python library was used to carry out the analysis. A list of parameters of interest was prepared including the upper and lower bounds. This was used to generate an input file containing a list of parameter combinations using the Saltelli sampling algorithm (Saltelli *et al.*, 2017). The number of items on this list and hence model simulations required is based on the sample size n (see 2.3.4.1) and the number of parameters under investigation d ($d = 31$ parameters in this study) using the formula: $n(2d + 2)$. The model simulation was run for each of the input combinations. The model outputs of interest for each of the input combinations were then used to generate the sensitivity indices using the analyse tool of SALib and visualised appropriately.

3.3.4.4 Calibration

Model calibration was used to align the model behaviour with the biological domain, specifically to pinpoint parameters that could not be verified experimentally or via existing literature. A systematic approach was used by establishing a list of scenarios with various parameter combinations based on their respective plausible biological ranges. Separate instances of the model were run for each of the scenarios a total of n times, with n as the established sample size needed in order to account for any uncertainty due to the stochastic nature of the model (see 3.3.4.1). As an example, if 30 scenarios were identified and n was established to be 100, the model was run a total of 3,000 times. The model output was then compared against the appropriate experimental data sets. Root mean squared error (RMSE) was used as a measure of goodness of fit and the parameter combination from the scenario producing the lowest RMSE value was selected.

The calibrated model was then validated against additional experimental data sets (collected independently from the calibration set) in order to increase confidence in the chosen model parameters.

3.3.4.5 Model validation

In addition to face validity and sensitivity analysis, the model output was validated against data gathered from wet-lab experiments. Two sample Kolmogorov-Smirnov test was used to compare the probability distributions of the simulation with that found experimentally at key time-points. ANCOVA was used for comparison of the predicted rate of growth simulated by the model with experimental growth curves after initially fitting the data to an appropriate linear regression model. Significant differences were first assessed for the slopes, then the intercepts. Unless otherwise stated, in all tests, significant difference was considered if $p < 0.05$.

The model validation against experimental data was done in three separate phases. Phase one occurring at the 'null' model stage and two and three at the 'contact' model stage. Phase one involved validating the model (prior to the implementation of contact rules but after calibration) against experimental growth curves. Phase two was the initial validation of the contact model against co-culture data and phase three, further predictive validation of the contact model against separate sets of co-culture data (gathered with varying seeding conditions).

Chapter 4

Mono-culture experiments

A prerequisite to the design and development of a computational model of a biological system, is to gain an understanding of the system behaviour. Through *in vitro* mono-culture experiments coupled with both qualitative and quantitative analysis, properties and behaviours of cancer and stem cells were explored. Human osteosarcoma cells (MG-63) and human embryonic stem cell derived mesenchymal progenitor cells (hES-MP), were used as *in vitro* models of cancer and stem cells respectively.

Osteosarcoma is the most common type of bone cancer and can be very aggressive with about 20% of patients presenting with metastases (Pautke *et al.*, 2004; Siclari and Qin, 2010). The MG-63 cell line was originally derived from the osteosarcoma of a 14-year male (Cellosaurus, 2012). This cell line is relatively well established, appearing in literature dating back to 1977 (Billiau *et al.*, 1977). In contrast, human embryonic stem (hES) cells are a much more recently established cell line group, first isolated in 1998 (J. A. Thomson, 1998). They provide a source of adult stem cells that do not lose their differentiation capacity when

expanded *in vitro* (De Peppo *et al.*, 2010). The hES-MP cell line is derived from hES cells and has been shown to be very similar to MSCs with regard to morphology and expression of markers while still allowing for higher experimental reproducibility in comparison to primary MSCs (Karlsson *et al.*, 2009; Bigdeli *et al.*, 2010; O. Li *et al.*, 2013).

4.1 Single cell analysis

Microscopic imaging was used for single cell analysis. Cell shape, size, motility, mitotic and apoptotic properties were studied and the results used as the basis for the design of the ABM.

4.1.1 Cell shape and size

Both cell types were seeded at a density of 10,000 cells per cm² and imaged on day 1, 4 and 7. Hoechst 33342, a blue fluorescent nuclear stain, was applied to the cells prior to imaging. Figures 4.1 and 4.2 are examples of the images taken of MG-63 and hES-MP cells respectively. On initial look, MG-63 cells appear more circular than the hES-MP cells, which appear elongated, specially at the high cell confluency seen on day 7. The shape of the cell nuclei of both cells also appear to follow this pattern. MG-63 cells grow at a faster rate than stem cells and are no longer a mono-layer of cells by day 7. At higher cell confluency, hES-MP cells appear to align along their major axis whereas MG-63 cells remain homogenously spread.

For quantitative analysis, images were analysed using CellProfiler3, specifically the IdentifyPrimaryObjects module. This was applied to the fluorescent images, to automatically identify the nuclei and measurements for area and eccentricity of the cell nuclei across the 3 time points and 2 cell types were taken using the MeasureObjectSizeShape module. Eccentricity is a measure of how much a conic section is not circular, with values close to zero indicating a more circular shape and close to 1, elliptical and calculated as the ratio of the distance between the foci of the ellipse and its major axis length (figure 4.3). Area was defined as the number of pixels within each enclosed region, converted to μm^2 using the image scale. Results are presented in figures 4.4 and 4.5 and with the full summary table under appendix (8.2). D'Agostino and Pearson's normality test was applied using the python module `scipy.stats.normaltest` and the data sets were found to be significantly non

parametric. As a result, Kruskal-Wallis test, with Mann-Whitney test as post-hoc pairwise comparison, was used.

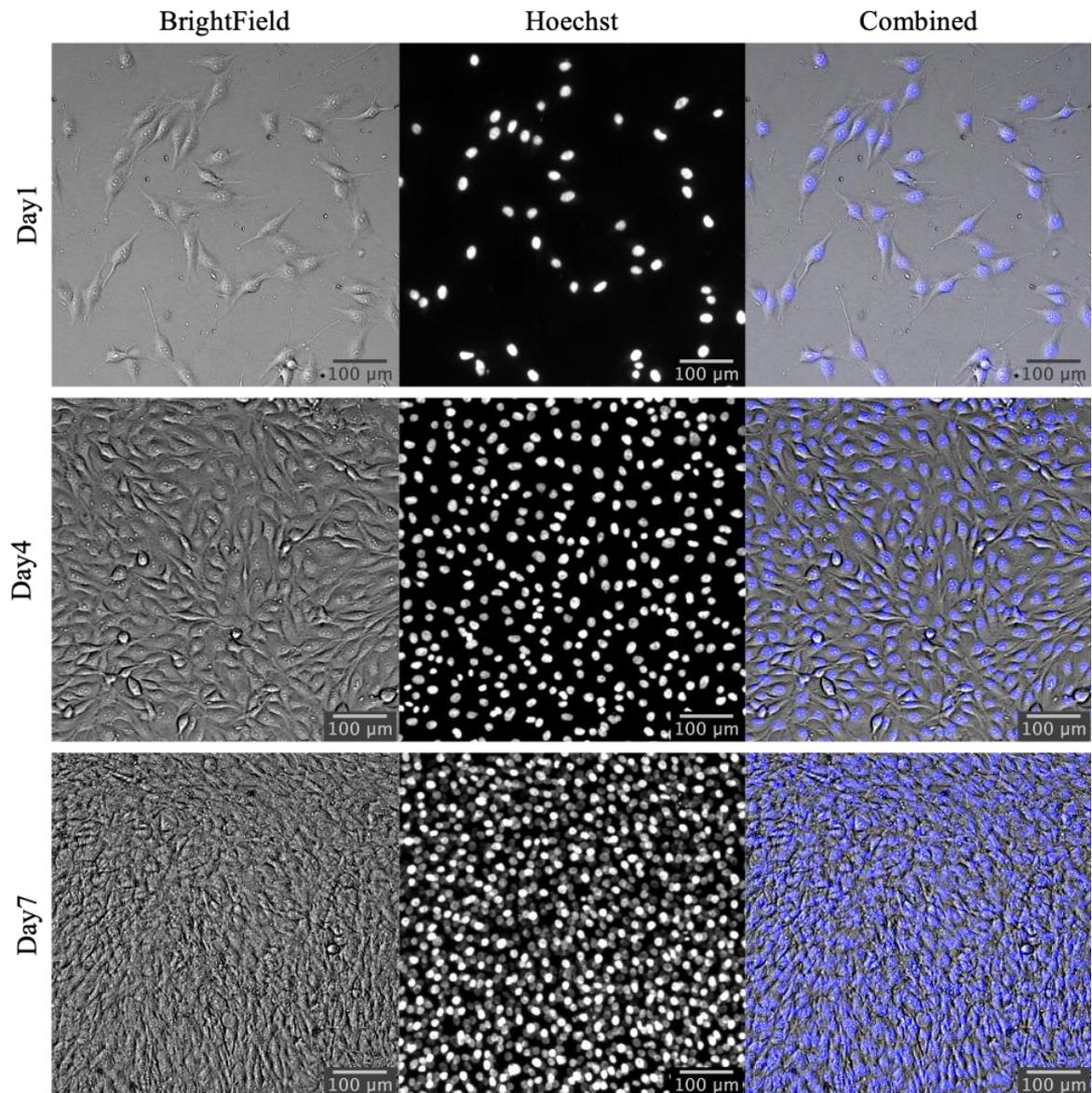


Figure 4.1: MG-63 cells (passage 10) seeded at 10,000 cells per cm^2 and imaged on day 1, 4 and 7. The first column shows images taken under brightfield, the middle column, blue fluorescence to detect the cell nuclei stained with Hoechst and the final column is the merged image combining the first two. At high confluency, the MG-63 cells are no longer in single mono-layers but appear homogeneously spread. The use of a nuclear stain allows for more accurate cell detection, specially at high confluency.

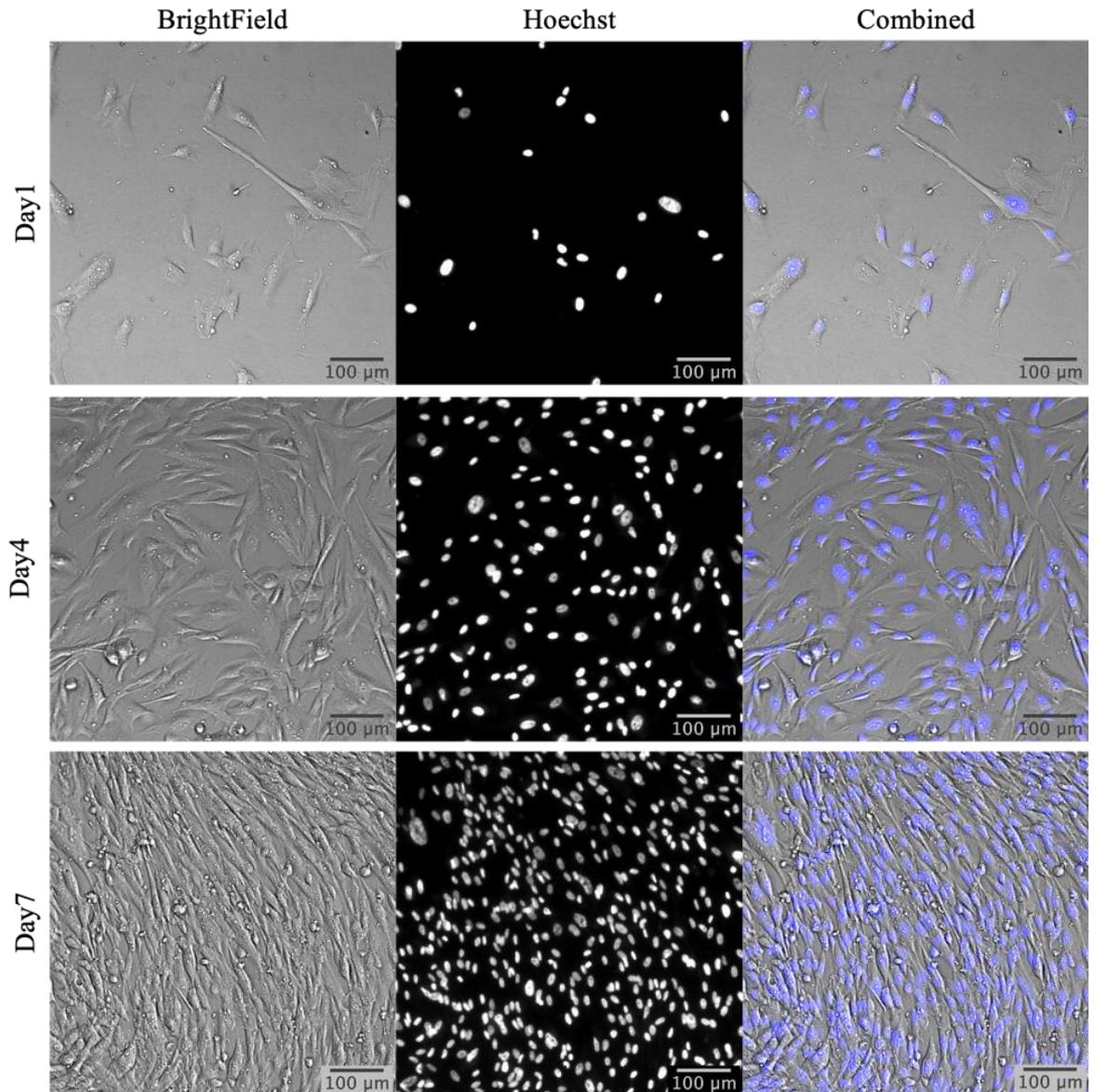


Figure 4.2: hES-MP cells (passage 9) seeded at 10,000 cells per cm² and imaged on day 1, 4 and 7. The first column shows images taken under brightfield, the middle column, blue fluorescence to detect the cell nuclei stained with Hoechst and the final column is the merged image combining the first two. hES-MP cells appear elliptical in shape and at high confluency, align along their major axis.

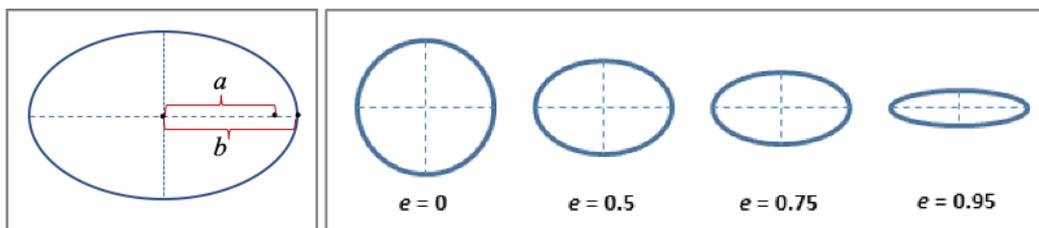


Figure 4.3: Diagrams depicting eccentricity. Left, eccentricity of an elliptical shape is calculated as the ratio of the distance from the centre to the foci (a) and the distance from the centre to the edge along the shape's major axis (b). Therefore eccentricity, $e = a/b$. Right, from CellProfiler3 documentation, shows the relationship between eccentricity value (e) and object shape.

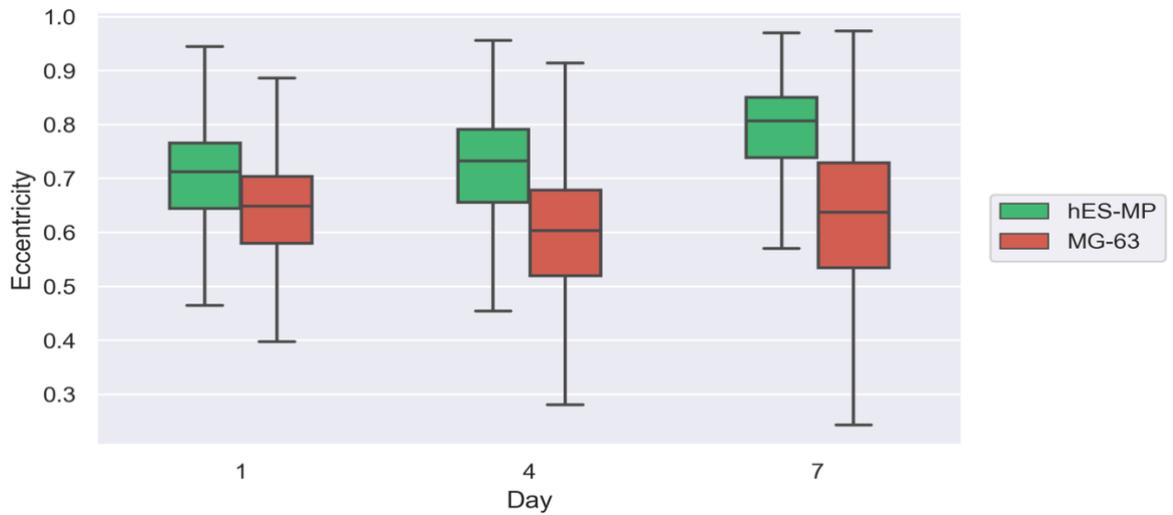


Figure 4.4: Comparison between cell nuclei eccentricity of hES-MP and MG-63 cells at day1, 4 and 7. Boxplot whiskers include fliers and show full range of values. hES-MP cells are skewed towards $e = 1$, indicating a more elliptical shape which increase as the cells grow (increase in confluency). Pair-wise comparison across all pairings showed significant differences ($p < 0.001$ between all pairings except day 1 and day 7 of MG-63 cells at $p < 0.05$) $n=12$.

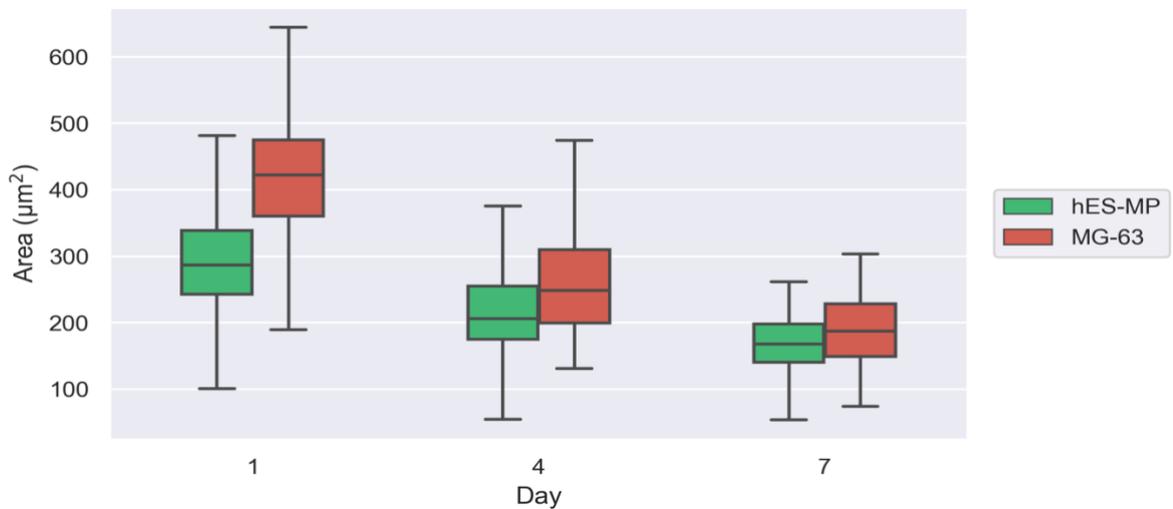


Figure 4.5: Comparison between cell nuclei surface area of hES-MP and MG-63 cells at day1, 4 and 7. hES-MP cells were found to be significantly smaller in area to MG-63 cells ($p < 0.001$). Both cells show a statistically significant ($p < 0.001$) reduce in surface area as they continue to grow. $n=12$.

Both cells were found to be in the elliptical definition of eccentricity with hES-MP cells highly skewed towards $e = 1$ (hES-MP skew: -1.320; MG-63 skew: -0.432) indicating a more elliptical shape. With progression of population growth the skew in the direction of $e = 1$ of the hES-MP cells was found to increase (day1 skew: -1.221; day7 skew: -1.736). hES-MP cells were found to be significantly smaller in area in comparison to MG-63 cells ($p < 0.001$) with both cells showing a statistically significant ($p < 0.001$) reduction in surface area as they continue to grow.

The non-parametric distribution of the cell nuclei area could be due to the limitations of the post-processing methodology, specially at high cell confluency. In order to avoid incorrect detection of clumped cells, a filter was applied, excluding large objects (an example shown in figure 4.6). The conditions for this filter were based on preliminary manual measurements and distribution of data from previous time points. Although, the data was still non-parametric due to the artificially imposed boundary conditions, the observed skew and kurtosis was reduced, results listed in table 4.1

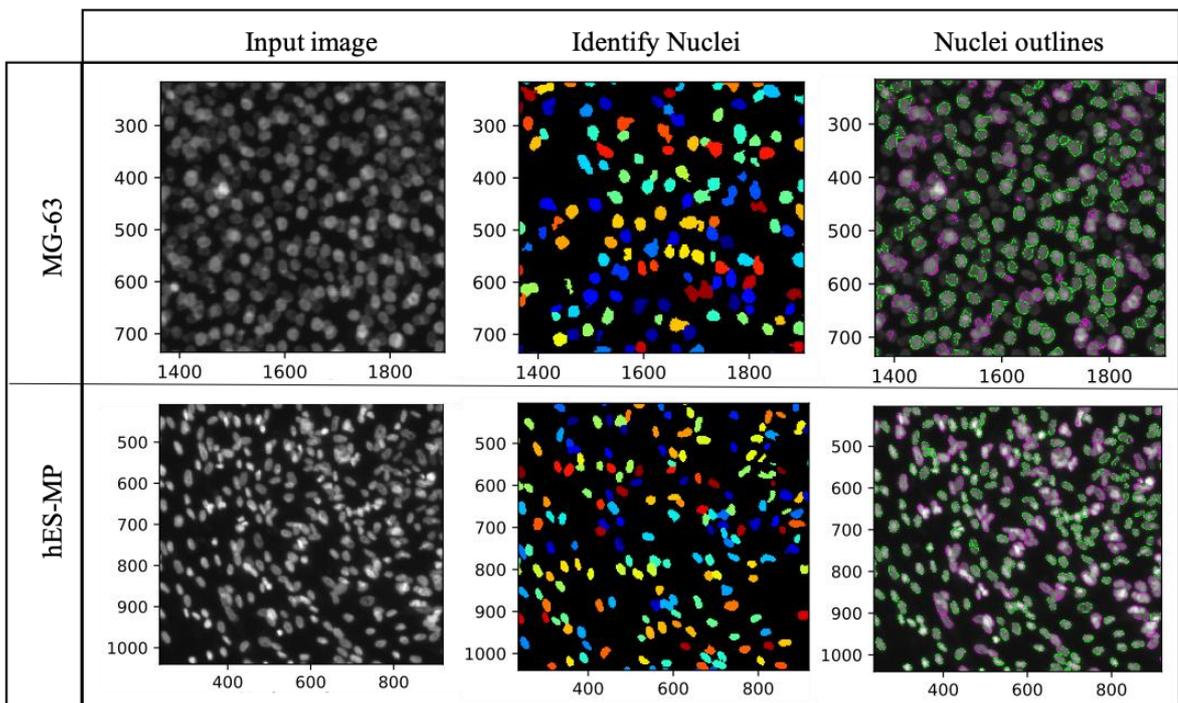


Figure 4.6: Performance of CellProfiler-IdentifyPrimaryObject module used to identify cell nuclei applied to images of MG-63 and hES-MP cells on day 7 at high confluency. In order to avoid cells clumps measured as single cells, boundary conditions were applied which would have contributed to the deviation from normal distribution of the data.

Table 4.1: Reduction in skew and kurtosis of data by applying filter to the CellProfiler pipeline to exclude clumped objects.

| Day | Cell Type | No filter | | Filter | |
|-----|-----------|-----------|----------|--------|----------|
| | | Skew | Kurtosis | Skew | Kurtosis |
| 7 | MG-63 | 1.38 | 3.98 | 0.02 | 0.79 |
| | hES-MP | 2.38 | 8.61 | -0.24 | -0.09 |

4.1.2 Time-lapse live cell imaging

Time-lapse live cell imaging was carried out as per protocol stated in section 3.2.3.3 Cells were seeded in a 48 well plate at three different seeding densities of low (5,000 cells/cm²), medium (10,000 cells/cm²) and high (20,000 cells per cm²) with four separate well replicates per seeding density and left in a forma cell culture incubator 24 hours prior to the start of time-lapse. HEPES buffer was used at a final working concentration of 25 mM to maintain the pH of the culture media and Hoechst stain was used to stain the nucleus of the cells to allow for more accurate cell tracking. Two locations per well were imaged under both brightfield and blue fluorescence channels every 20minutes. In between imaging, the cells were kept in the dark to reduce the effects of photo toxicity on the cells and photo bleaching on the stain. Full media change was done at 48 hours and imaging resumed at the same stage positions. Images were post processed with ImageJ and analysed with Python.

4.1.2.1 Cell tracking

Cell tracking was done using the TrackMate plugin (Tinevez *et al.*, 2017) for ImageJ which enables extraction of cartesian co-ordinates per cell per time point. To ensure accuracy and reliability of data, this process was semi-automated, whereby the initial cell detection and tracking was done by TrackMate and then subsequently checked and corrected manually for each cell. This data was imported into Python for further organisation, analysis and visualisation. Object-orientated code was used to organise the data, linking each cell to itself at timepoint -1 and timepoint +1. This was used to assess single cell motility and track lineage from the start of time-lapse to the end or when the cell was no longer in frame.

Approximately 100 individual hES-MP cells per seeding condition were tracked over 72 hours. As MG-63 cell proliferation was higher, approximately 50 individual cells were tracked over 42 hours. Diagrams depicting the time-line for each individual cell tracked are included in appendix (8.3).

4.1.2.2 Analysis of cell displacement

From initial observations, it was clear the stem cells are more motile than the cancer cells and as a result frequently migrated out of frame. In order to compensate for this, twice as many individual hES-MP cells were tracked as MG-63 cells. Figure 4.7 shows an example of typical cell trajectory for either cell from the medium seeding density experiments.

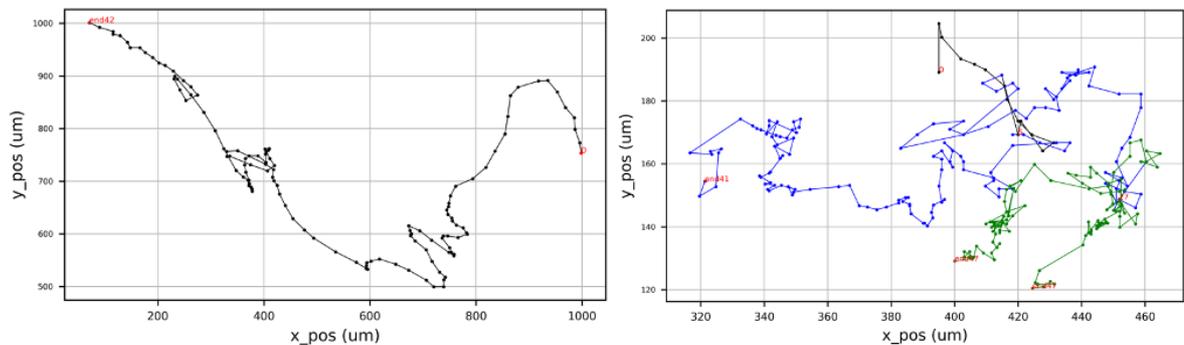


Figure 4.7: Examples of individual cell trajectory for hES-MP (left) and MG-63 (right) cells seeded at 10,000 cells/cm². The start point is marked with a red '0' and endpoint or key events such as mitosis are marked with red numbers at the time they occur. Each black dot represents the cell's position at a single time-point. MG-63 cells were highly proliferative, undergoing mitosis two times on average over the 42 hours. Stem cell mitosis was observed at a very low rate (an average of 1 cell out of 100). The first generation cell is marked in black, 2nd generation in blue and 3rd generation in green. Over the same period of time, the hES-MPs covered a much larger displacement from start to final position in comparison to the MG-63 cells comparable by the range displayed on the axis.

Displacement was defined as the Euclidian distance between the cell's cartesian coordinates from one time frame to the next. The distributions of displacements from each of the three seeding densities and two cell types were compared. The hES-MP cells had a larger range of displacements (0 – 80 μm) across all seeding densities in comparison to the MG-63 cells (0 – 20 μm). In all cases the distribution of displacements measured (shown in figure 4.8) were non-parametric with a strong skew towards the lower displacements. Kruskal-Wallis, followed by Man-Whitney pairwise comparison tests found, a significant difference between all seeding density pairings of each cell (hES-MP medium and high $p < 0.05$ and all other pairings $p < 0.001$). As a result, the relationship between displacements measured

at each time-point and the cell density was explored using Spearman rank-order correlation coefficient. Density was defined as the total number of cells in frame at the time-point the displacement occurred. Figure 4.9 (top) is a visualisation of this relationship for both hES-MPs and MG-63s. A weak negative correlation between displacement and density was found for MG-63 cells: $r = -0.0136$, $p < 0.001$ whereas for hES-MP cells, a weak positive correlation was observed: $r = 0.0521$, $p < 0.001$.

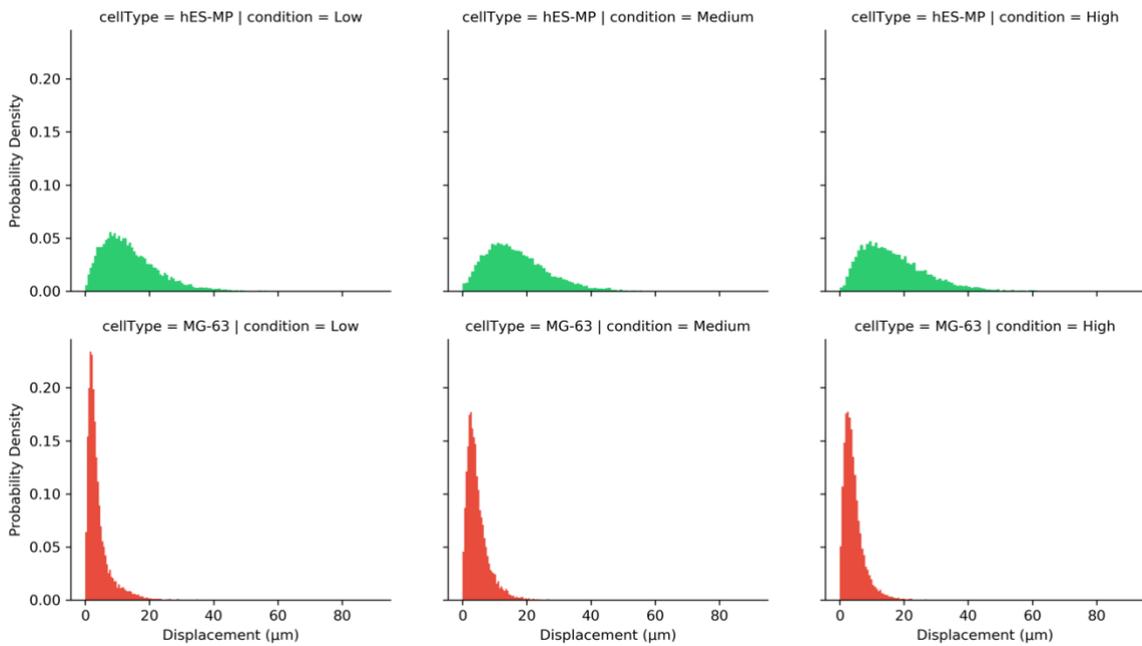


Figure 4.8: Probability distributions of all 6 datasets, top: hES-MP and bottom, MG-63 cells. The hES-MP cells produced a larger range of displacements ($0 - 80 \mu\text{m}$) across all seeding densities in comparison to the MG-63 cells ($0 - 20 \mu\text{m}$). In all cases the distribution of displacements were non-parametric with a strong skew towards the lower displacements. A significant difference between all seeding density pairings of each cell were found (hES-MP medium and high $p < 0.05$ and all other pairings $p < 0.001$).

Also shown in figure 4.9 (bottom) is the relationship between displacement and time. This was looked at to account for the relationship between density and time since, density increases with an increase in time. Therefore, it was considered that the correlation observed might be due to time as opposed to density (for example caused by prolonged exposure to camera flashes during time-lapse acquisition). Again, Spearman rank-order correlation coefficient was calculated for both relationships. A similar weak negative correlation between displacement and time was found for both cells (MG-63 cells: $r = -0.237$, $p < 0.001$ and hES-MP cells: $r = -0.228$, $p < 0.001$). Although, the correlation is classified as weak, in both cases it is much stronger than the correlation established between displacement and density. Additionally for hES-MP cells that were tracked over a media change period (at

time-point = 48hours), it was observed that the cell displacements slowed down considerably directly after media change. This is also clearly visualised in figure 4.9 (bottom, right).

Since MG-63 cells were tracked across multiple generations, the distribution of displacement's measured for each generation was also compared using Man-Whitney and found to be significantly different across all pairings ($p < 0.001$). However, increase in generation is also associated with increase in time. Therefore, it is not clear what the direct cause in change of displacement is without further study.

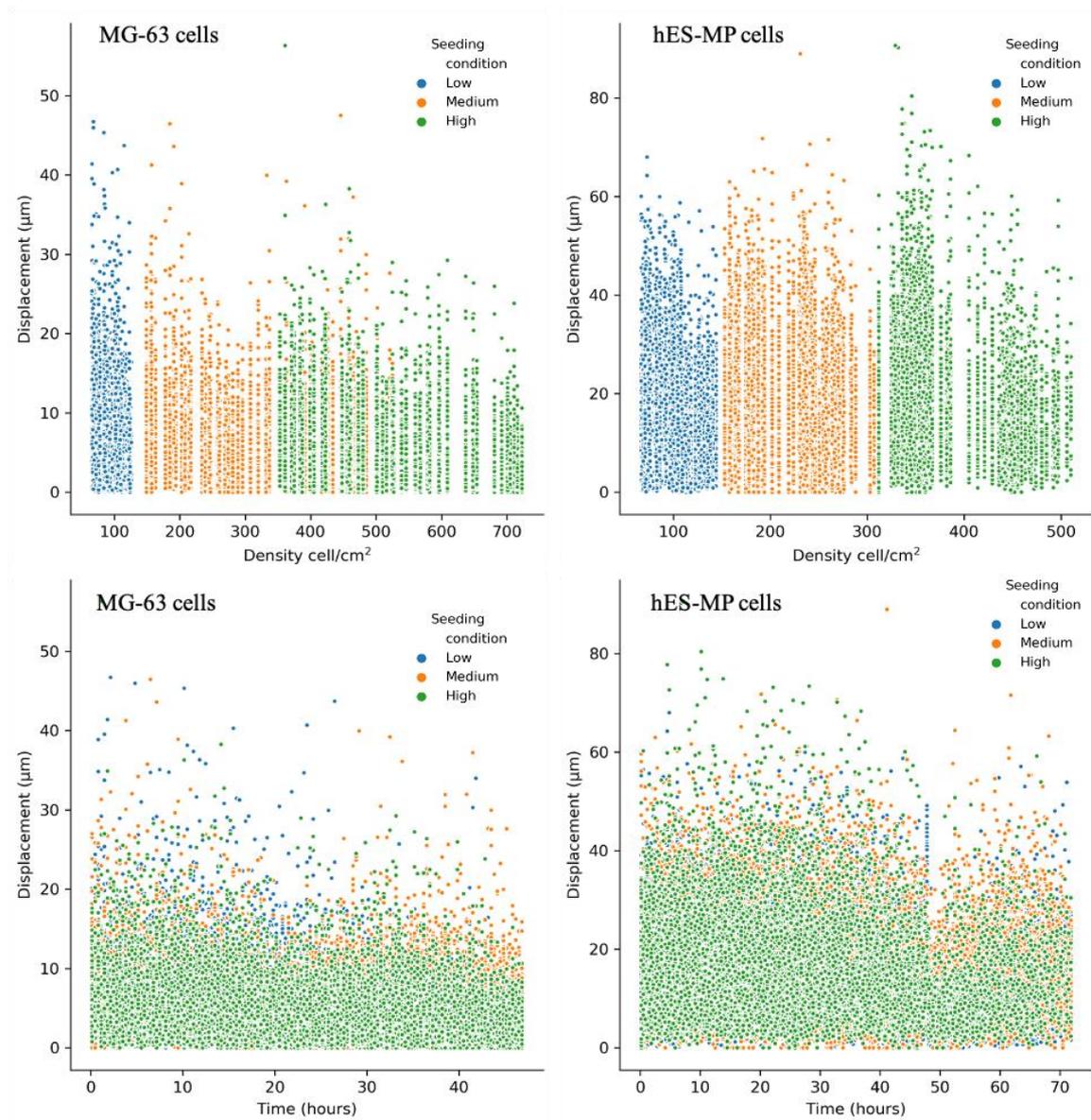


Figure 4.9: Top: shows the relationship between displacement and density. A weak negative correlation was found for MG-63 cells: $r = -0.0136$, $p < 0.001$ whereas for hES-MP cells, a weak positive correlation was observed: $r = 0.0521$, $p < 0.001$. Bottom: the relationship between displacement and time. Both MG-63 and hES-MP cells were found to have a weak negative correlation (MG-63 cells: $r = -0.237$, $p < 0.001$ and hES-MP cells: $r = -0.228$, $p < 0.001$). The correlation between displacement and time is stronger than the correlation between displacement and density. The sudden drop in displacement at time = 48hours in hES-MP cells seen, coincides with the exact time at which the media was changed.

4.1.2.3 Probability distribution model of displacement

The displacements of each cell was fitted to the following three continuous probability distributions, Weibull, Gamma and Lognormal using the `scipy.stats` module in python. These distributions were selected as they are examples of mostly positively skewed distributions. The fitted models were compared against the original displacement data using both Man-Whitney and Wilcoxon rank sum tests. Figure 4.10 shows the plots of these models and table 4.2, the results of the statistical analysis. It was concluded that the lognormal probability distribution model is the best fit for the displacements measured for both cells.

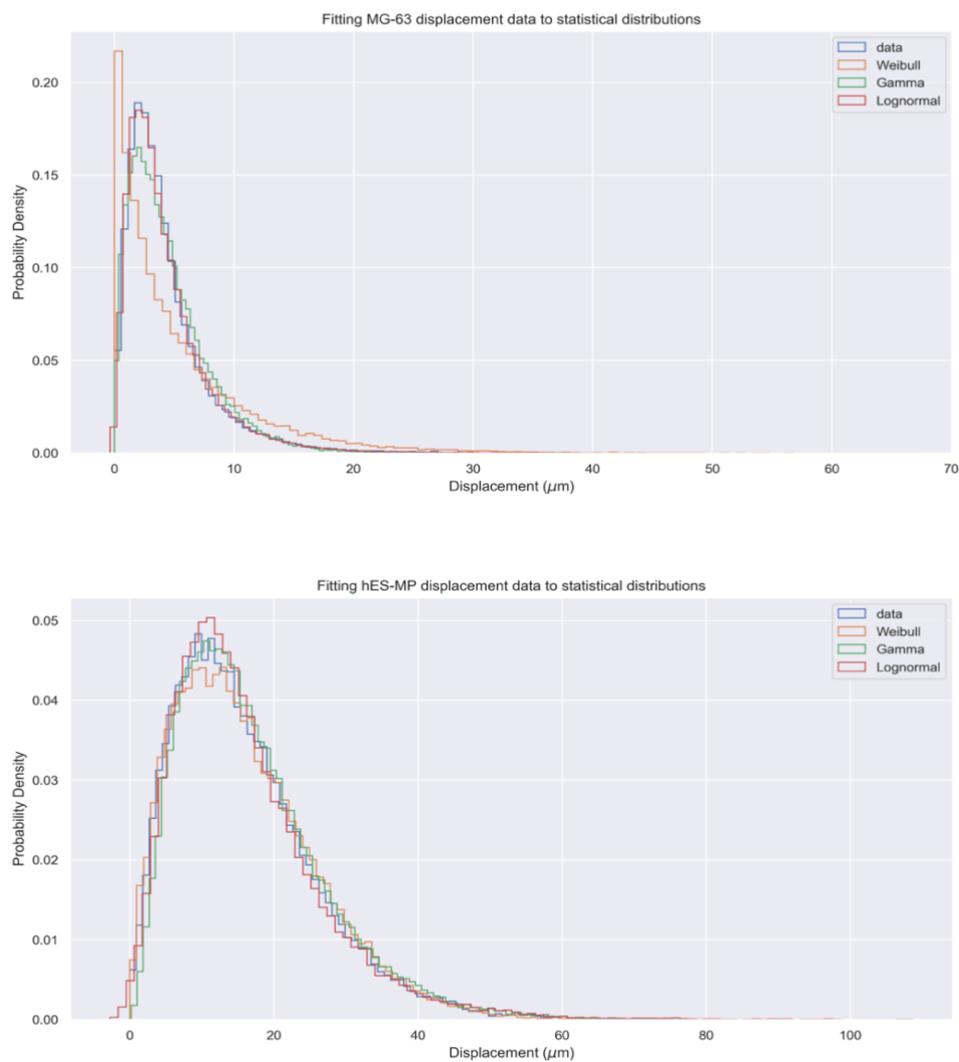


Figure 4.10: The probability distribution of displacing for MG-63 (top) and hES-MP (bottom) cells was fitted to existing continuous probability functions, Weibull, Gamma and Lognormal. Both Man-Whitney U test and Wilcoxon rank sum tests found the lognormal fitted model statistically similar to the raw data for both cell types. The fitted lognormal model was used to generate random cell displacements in the final ABM.

Table 4.2: Table listing the statistical tests results from comparison of each of the fitted models to the raw experimental data. p values in bold indicate non-significant results and hence statistically similarity between model and data. The Lognormal model was found to be the best fit by both tests for both cells.

| Cell Type | Model | Wilcoxon rank sum | | Mann-Whitney test | |
|---------------|-----------|-------------------|------------|-------------------|------------|
| | | p value | Test score | p value | Test score |
| hES-MP | Weibull | 0.0004 | 295666030 | 0.00026 | 595365864 |
| | Gamma | 0.0 | 278212221 | 0.0 | 568475844 |
| | Lognormal | 0.70 | 301572217 | 0.34 | 603457316 |
| MG-63 | Weibull | 0.0 | 466181067 | 0.04 | 1038053871 |
| | Gamma | 0.0 | 496715114 | 0.0 | 1007700970 |
| | Lognormal | 0.06 | 517312059 | 0.09 | 1039738758 |

4.1.2.4 Analysis of cell directionality

Two groups of angle measurements were calculated, the direction of motion (referred to as direction) and the turning angle. Direction is the angle made by the cell velocity vector with respect to the horizontal and turning angle is the angle between the cell's velocity vectors from one time frame to the next. First the displacement in the x and y directions were calculated (as discussed above), this was followed by calculating the direction by applying the arctan function to the ratio between y displacement and x displacement as per trigonometry rules. Boundary conditions for each of the four quadrants were applied in order to ensure the cell direction is in the range $0 - 2\pi$ and in the anti-clockwise direction.

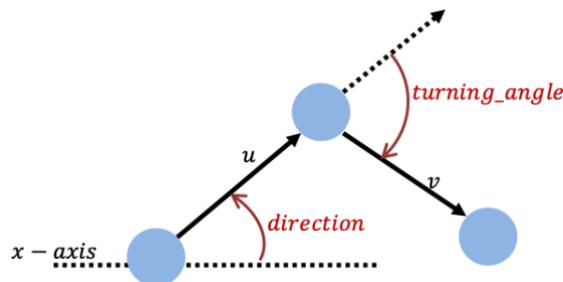


Figure 4.11: A depiction of the definition of cell direction and turning angle: direction is the angle made by the cell velocity vector with respect to the x-axis whereas turning_angle is the angle between the cell's velocity vectors from one time frame to the next

Turning angle is calculated using the atan2 function within the math module of python. This function returns a value for the angle between $-\pi$ and π . The angle is positive in the anti-clockwise direction and negative in the clockwise direction. atan2 is applied to the x, y positions of a vector and is computed with respect to the positive x-axis. To apply it to the angle between two vectors, one vector has to be rotated to the positive x-axis. Equation 4.1 shows the (a) rotation matrix and (b) final equation used to compute the turning angle between vector u and v, given than $u = (x_1 y_1)$ and $v = (x_2 y_2)$.

$$\begin{pmatrix} x_1 & y_1 \\ -y_1 & x_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1x_2 + y_1y_2 \\ x_1y_2 - x_2y_1 \end{pmatrix} \quad 4.1a$$

$$Turning_{angle} = \mathit{math.atan2}(x_1y_2 - x_2y_1, x_1x_2 + y_1y_2) \quad 4.1b$$

Figure 4.12 is the polar histograms of the directions and turning angles calculated for both cells in each of the three seeding conditions. On initial observation, the cells appear to favour a forward direction for turning angle whereas their overall direction with respect to the horizontal appears more uniformly spread. hES-MP cells have a significantly stronger bias to forward motion than MG-63 cells.

As the distributions were found to be non-parametric, Kruskal-Wallis test was used to compare datasets. No significant difference (Chi square = 1.04, $p = 0.60$) was found among the three seeding conditions for MG-63 direction angle and hES-MP turning angle. However for MG-63 turning angle and hES-MP direction angles, a significant difference was found, $p < 0.001$ for both. Man-Whitney test was used to compare within the seeding densities. For MG-63 turning angles, It was found that the high seeding density was significantly different to the low and medium densities ($p < 0.001$) and for hES-MP direction angles, significant difference was found across all pairings ($p < 0.001$).

As a result of these findings, the relationship between turning angle and displacement was explored.

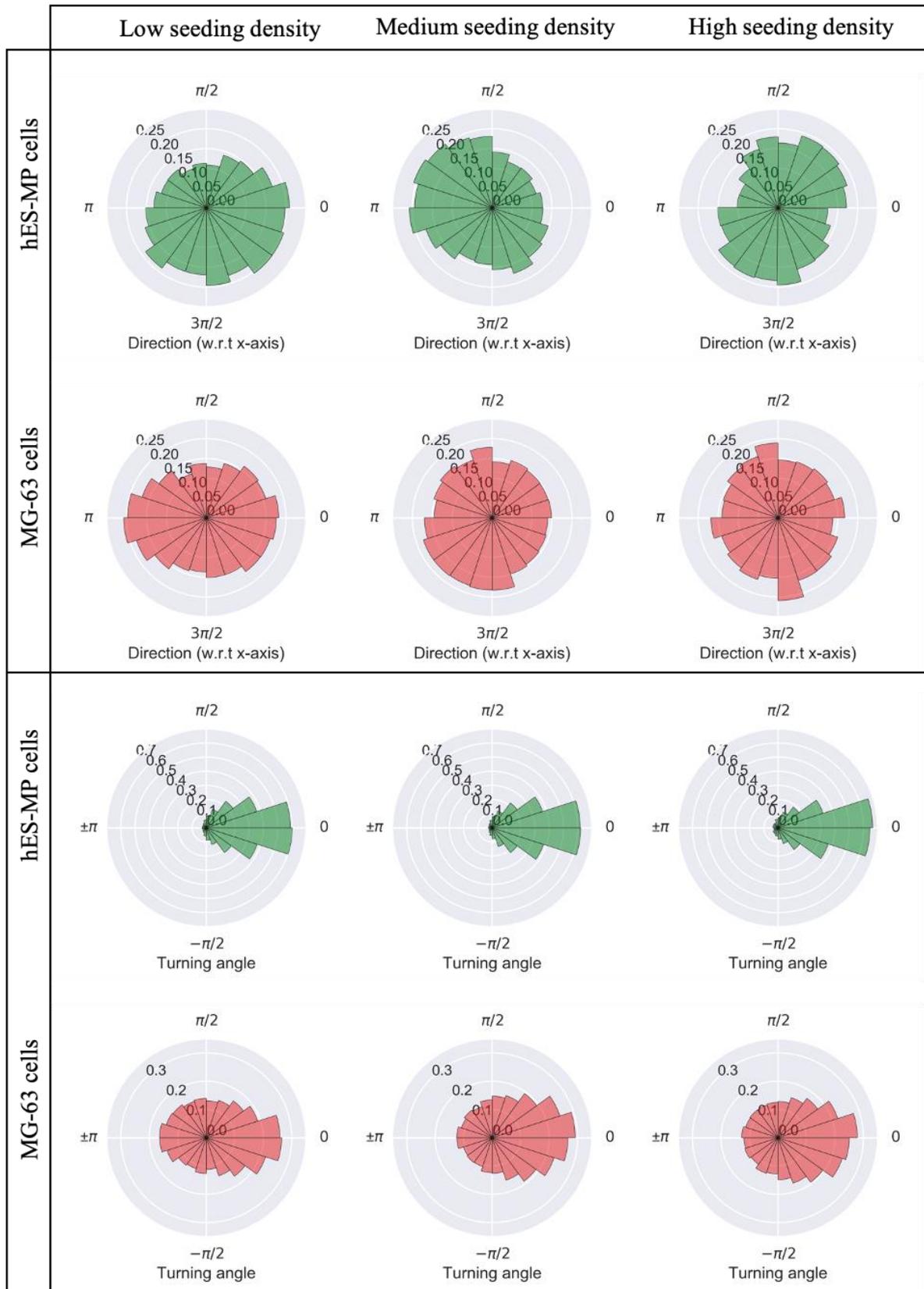


Figure 4.12: Polar histogram of the 'direction' and 'turning angles' with 20 angular bins. No significant difference (Chi square = 1.04, $p = 0.60$) was found among the three seeding conditions for MG-63 direction angle and hES-MP turning angle. However for MG-63 turning angle and hES-MP direction angles, a significant difference was found, $p < 0.001$ for both. Man-Whitney test for MG-63 turning angles found that the high seeding density was significantly different to the low and medium densities ($p < 0.001$) and for hES-MP direction angles, significant difference was found across all pairings ($p < 0.001$).

4.1.2.5 Relationship between displacement and turning angle

To visualise the relationship between the cell displacement and turning angle from one time frame to the next, scatter and joint probability distributions were used. At low displacements, the distribution of turning angle is approximately uniform, whereas at large displacement, the turning angle distribution becomes increasingly concentrated around 0, indicating little to no change in direction and a persistence of forward motion. This can be described using a von Mises distribution, with the displacement used as concentration parameter, κ . Figure 4.13 shows the probability density function of a von Mises distribution for $-\pi < \theta < \pi$ centred on $\theta = 0$. The variance of the distribution is controlled by the concentration parameter, κ , which is inversely proportional to variance in a normal distribution. This means, for the smallest κ values (approximately $\kappa = 0$) the distribution is very close to uniform and as κ is increased, the distribution becomes increasingly concentrated around the centre (0 here).

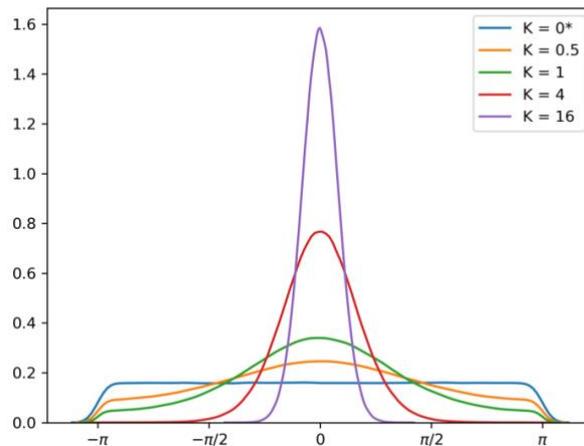


Figure 4.13: The probability density function of a von Mises distribution for $-\pi < \theta < \pi$ centred on $\theta = 0$. For the smallest κ values (*approximately $\kappa = 0$) the distribution is very close to uniform and as κ is increased, the distribution becomes increasingly concentrated around the centre (0 here).

The distribution of turning angle (with respect to displacement), was modelled using a von Mises distribution, by equating cell displacement to κ , therefore associating a unique probability distribution for the turning angle based on the cell's displacement. The model was calibrated to fit the experimental data by implementing a displacement shift parameter to allow for a range of low displacement's to fall into the $\kappa \approx 0$ region (i.e. turning_angle =

vonMises($\kappa = \text{displacement} - \text{shift}$). For hES-MP cells the displacement shift was found to be $5 \mu\text{M}$ and for MG-63 cells $7 \mu\text{M}$. The modelling process is shown in figure 4.14.

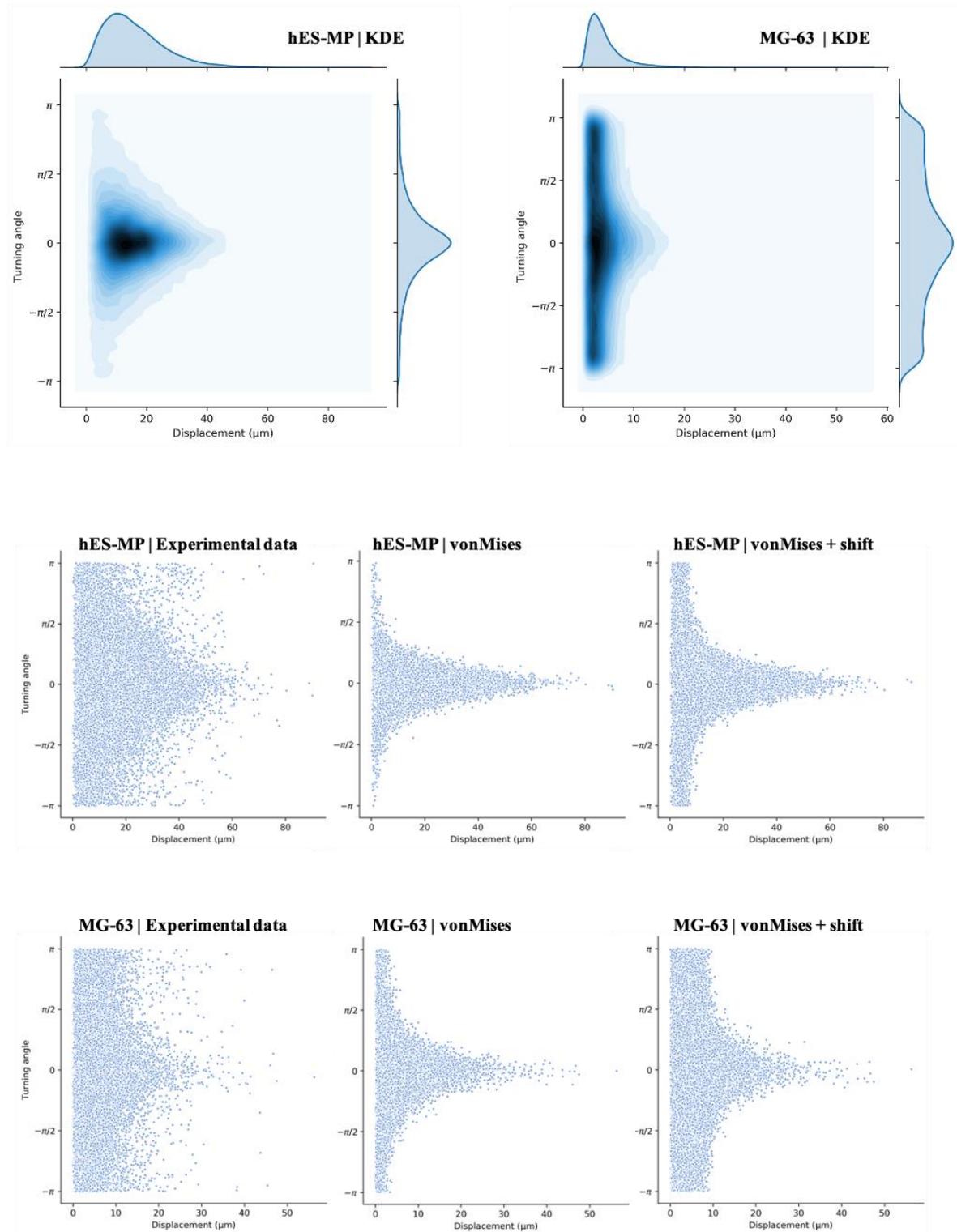


Figure 4.14: Top: the kernel density estimation (KDE) for the joint probability distribution of cell displacement and turning angle. Bottom two rows show the scatter plots or the (from left to right) raw experimental data, the data generated by vonMises($\kappa = \text{displacement}$) and vonMises($\kappa = \text{displacement} - \text{shift}$) respectively. For hES-MP cells the displacement shift was found to be $5 \mu\text{M}$ and for MG-63 cells $7 \mu\text{M}$. The final model using vonMises + displacement shift, is used to draw random turning angles for the cells in the final ABM.

4.1.2.6 Lineage tracking

During the cell tracking data acquisition stage, mitosis events were labelled and links were placed between a cell and its resulting daughter cells. This data (visualised in figure 4.15) was used to obtain an estimate for cell cycle duration and was limited by the number of mitosis events captured per individual cell track. A minimum of two mitosis events per cell was required as cell cycle duration was defined as the time elapsed from one mitosis event to the next. This was only successfully collected for MG-63 cells as for hES-MP cells, mitosis events were not reliably detected/identified.

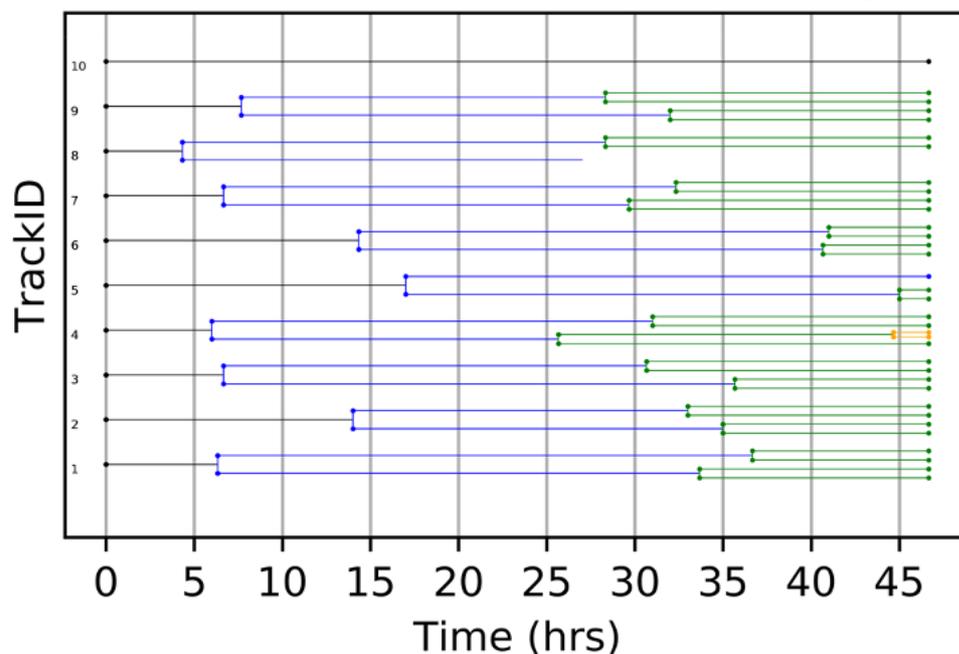


Figure 4.15: An example of cell lineage tracking shown for 10 individual MG-63 cells over 2 days. Full set of track diagrams are included in appendix (8.3). Changes in colour represent change in cell generation. Circular markers are used to label start point, end point and key events. Open ended tracks indicate the cell migrated out of frame. Measuring the time elapsed between one mitosis event to the next can be used as an estimate of cell cycle duration.

147 data points were collected ranging from 18 to 36 h. The data was found to be normally distributed using D'Agostino and Pearson's normality test (Chi statistic = 5.87, $p = 0.53$) with a mean of 24.937 hours and standard deviation of 3.861. There was a slight skew (0.342) in the positive direction (shorter cell cycle duration).

4.1.2.7 Death rate

Cell death was identified visually and recorded during cell tracking. The cell death mechanism observed suggested programmed cell death (apoptosis) as cell membrane blebbing, cell shrinkage and bursting to form apoptotic bodies could be seen in the sequence of time-lapse images. Death rate was calculated as the fraction of cell deaths recorded out of the total number of cells visible per 20 minute time frame. Death rate was found to be 3.283 cells per 100,000 for MG-63 cells and 6.132 cells per 100,000 for hES-MP cells per 20 minute time frame.

4.1.2.8 Multi-nucleated cancer cells

During frame by frame analysis of cancer cell time-lapse images, it was noticed that occasionally MG-63 cells would form what appeared to be clusters that move in unison an example of which is shown in figure 4.16. However on closer analysis, and cross-comparison of brightfield and fluorescent images of the cell nuclei, due to the lack of cell membrane between the nuclei of the cells in the cluster, they are perhaps better described as multi-nucleated cells. This is not an expected cell property documented for this cell line and could be caused by a dysfunction during cytokinesis, the final stage of mitosis, whereby the cell membrane should split into two. Although, this does not explain the instances of cells with odd numbers of nuclei. This is an interesting area for future work.

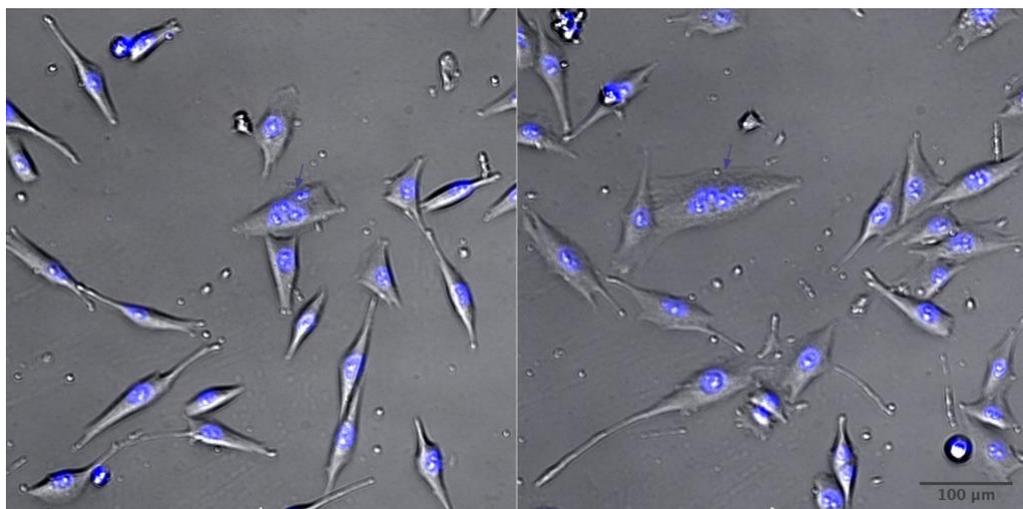


Figure 4.16: Images from time-lapse analysis taken 10 hours apart, showing the same cell (marked with arrow) that appears to have 4 nuclei and could be considered a cluster.

4.2 Multi-cell analysis (population level analysis)

Cell growth obtained from manual cell counting was used as a form of viability study in order to gauge population level behaviour of cells. Results gathered here were used as a success criterion employed in the calibration and validation of the ABM and formed the basis for design of the co-culture experiments.

Both hES-MP and MG-63 cells were seeded (in separate mono-culture experiments) in 24 well plates at four different seeding densities of 2, 500 cells/cm², 5, 000 cells/cm², 10,000 cells/cm² and 20,000 cells per cm² with four separate well replicates per seeding density. This was multiplied by seven, the total number of time-points, as manual cell counting is an endpoint analysis after which cells are discarded. The plates were kept under general cell culture conditions (listed in section 3.2.1.1) with full media change carried out on days 3 and 5 post seeding. Cells were detached (per protocol listed under 3.2.4.2) and counted every 24hours over 7 days. Prior to detachment cell nuclei were stained with Hoechst 33343. Hoechst was used to allow for automated cell counting using a cell detection pipeline designed in CellProfiler (see section 3.2.4.2). The experiment was repeated 3 times.

4.2.1 Rate of growth and seeding density

The results for both cells across all four seeding densities are presented in figures 4.17-4.19. All growth curves obtained are in the growth phase and have not reached the plateau phase. On initial observation, the MG-63 cells were found to grow at a significantly faster rate across all seeding densities in comparison to the hES-MP cells. Day 7 MG-63 cell numbers were a minimum of 3.5 times higher than the respective hES-MP cells seeded at the same initial seeding density (with an increase to 5 fold for the 20,000 cells/cm² condition). On day 1, MG-63 cell numbers are closer to the number of cells seeded, whereas the hES-MP cell numbers fall below, suggesting a lower attachment rate for hES-MP cells or a longer lag phase. The MG-63 cells appear to grow at a similar rate across all four seeding conditions, whereas hES-MP cells seeded at 20, 000 cells/cm² appear to grow at a slower rate than those in lower seeding densities. For further analysis of growth rates, the results were fitted to linear regression models.

Directly fitting the results to linear regression models were found to not meet the conditions for linear regression such as homoscedasticity and normality of residuals. In order to meet these requirements the cell numbers were transformed using Box-Cox

transformation with lambda, $\lambda = 0.3$. The same transformation was applied to all conditions in order to allow for cross comparison. Values of the full statistical analysis are presented in appendix (8.4) and a visual representation of the models, in figure 4.20.

The slopes of the regression models, representing the increase in cell number per timepoint, are very similar across all conditions for MG-63 cells with a slope coefficient, $\beta \approx 18.2$. For the hES-MP cells the slope coefficients are similar ($\beta \approx 10.73$) only for the three smallest seeding densities. This similarity in β coefficients, suggests there might be a relationship between cell density and percentage of mitotic cells at any given time point (further discussed in the next section).

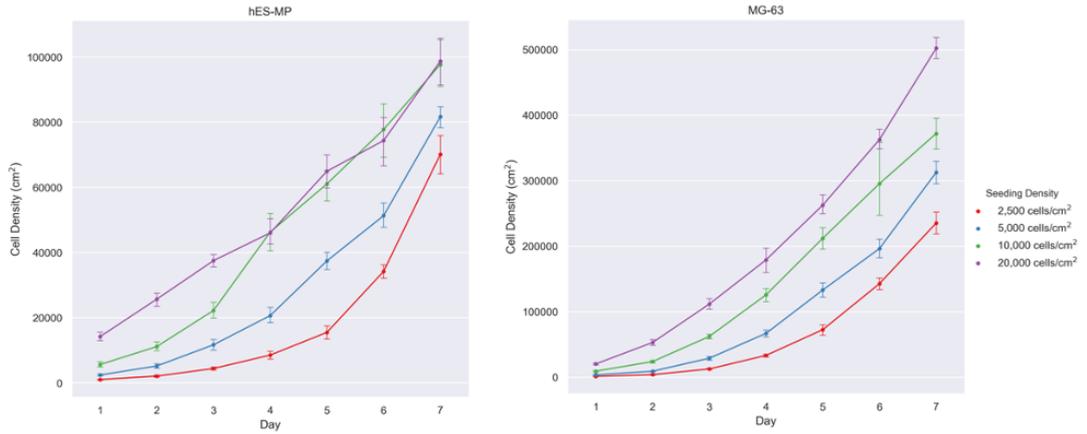


Figure 4.17: Growth curves for hES-MP cells (left) and MG-63 cells (right) seeding at four different initial densities. Cell numbers were counted every 24hours over 7 days and averaged from 4 well repeats and 3 experimental replicates. As evident by the range of cell density shown on each of the y-axis, the cancer cells were found to grow at a significantly higher rate than the stem cells.

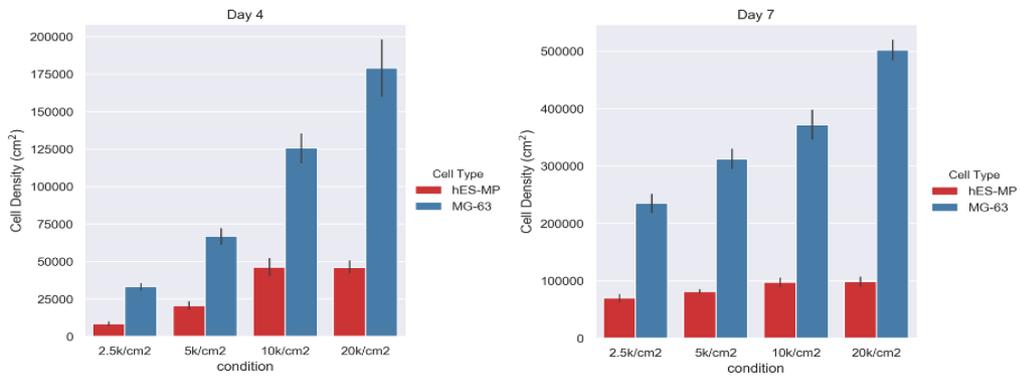


Figure 4.18: Day 4 and day 7 cell density plotted as bar charts to contrast the final cell numbers for the same initial seeding number per cell type. MG-63 cell numbers (in blue) were found to be a minimum of 3.5 times larger than the hES-MP cells for the same seeding density (5 times larger in the 20,000 cells/cm² experiment).

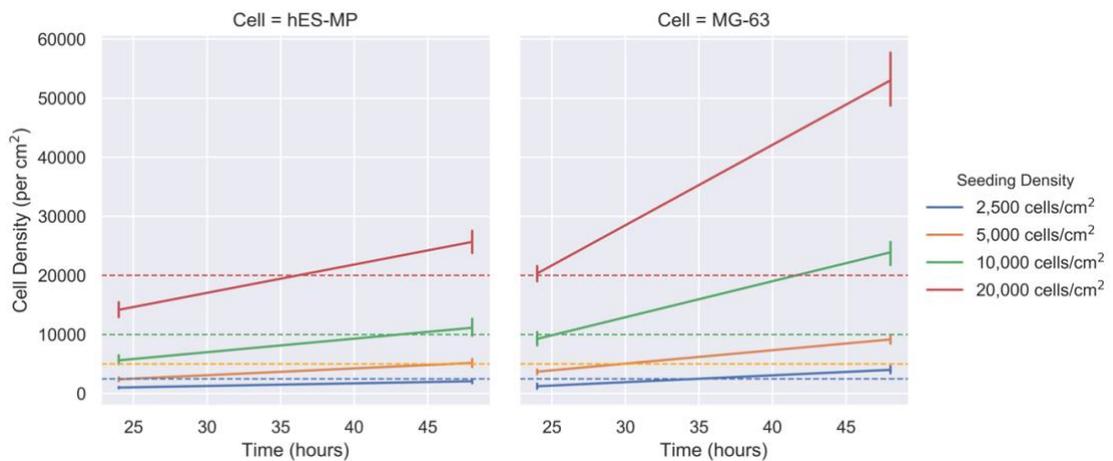


Figure 4.19: Cell density on day 1-2 for both cells. MG-63 cell numbers are closer to the numbers of cells seeded, whereas the hES-MP cell numbers fall below, suggesting a lower attachment rate for hES-MP cells or a longer lag phase.

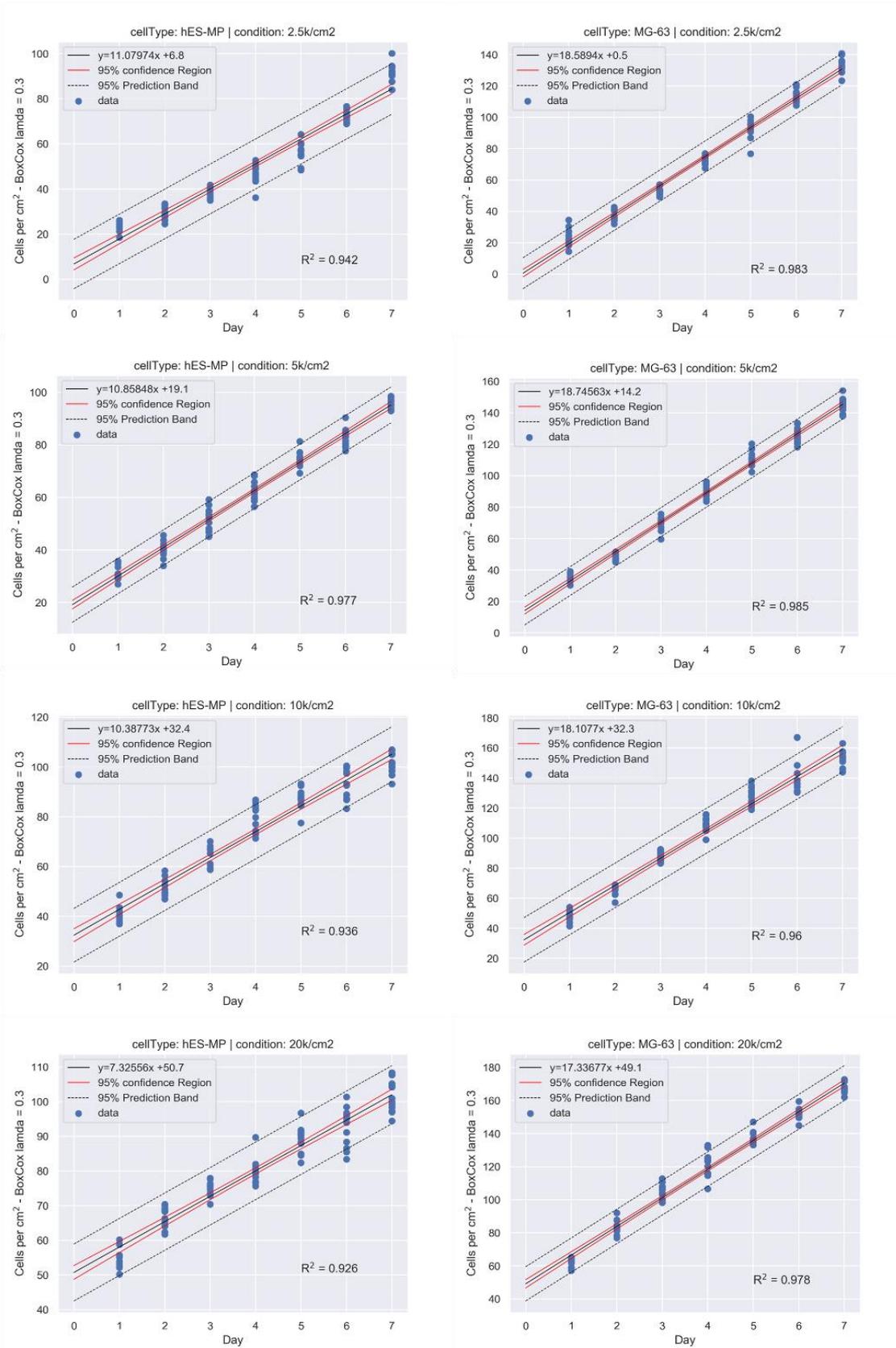


Figure 4.20: Cell growth modelled using linear regression with Box-Cox transformation ($l = 0.3$). The slopes of the regression models, representing the increase in cell number per timepoint, are very similar across all conditions for MG-63 cells and only the smallest three seeding densities for the hES-MP cells.

4.2.2 Mitotic rate model

The relationship between the mitotic rate and cell density was fitted to a linear regression model for each cell. The mitotic rate was defined as the difference in cell number from time = t to time = $t + 24$ h as a percentage of the total number of cells at time = t and the cell density was the total number of cells per mm^2 at time = t . Both models are presented in figure 4.21 with the results of the assumption tests for linear regression listed in table 4.3.

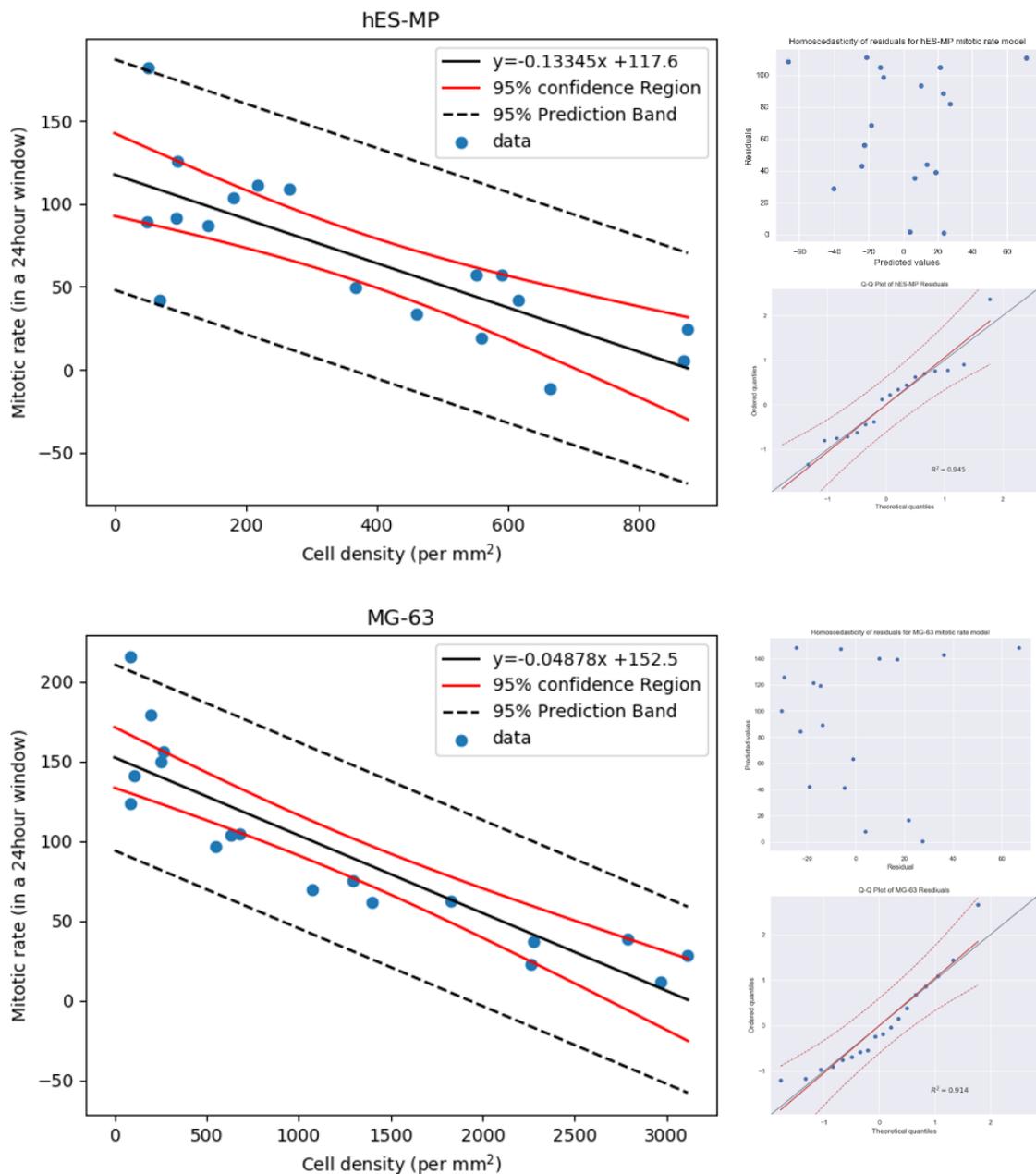


Figure 4.21: The mitotic rate regression models for the relationship between cell density and percentage of mitotic cells in a 24 hour window. To the left of each model the residual plots for homoscedasticity and normality are included. These models were implemented in the ABM to simulate cell growth.

Table 4.3: Table of linear regression assumptions for the mitotic rate models for hES-MP and MG-63 cells.

| Test | hES-MP mitotic rate model | MG-63 mitotic rate model |
|---|---------------------------------|---------------------------------|
| Model: $y = \alpha + \beta x$ | $y = 117.6 - 0.1334 x$ | $y = 152.5 - 0.0488 x$ |
| Shapiro-Wilk normality test for residuals | Test statistic 0.963 | Test statistic 0.914 |
| | p-value 0.651 | p-value 0.103 |
| Harvey Collier test for linearity of residuals | Test statistic 0.038 | Test statistic 1.147 |
| | p-value 0.970 | p-value 0.271 |
| Gold-Quandt test for homoscedasticity | Test statistic 0.642 | Test statistic 0.952 |
| | p-value 0.125 | p-value 0.193 |
| Mean of residuals | $3.4 \times 10^{-14} \approx 0$ | $2.6 \times 10^{-15} \approx 0$ |

4.3 Summary

In this chapter the qualitative and quantitative analysis of mono-culture *in vitro* experiments was presented. These analysis were carried out in order to gain a better understanding of the biological domain leading to the design and parametrisation of the agent based model discussed in the next chapter.

Analysis of the shape and size of the nuclei, found the MG-63 cells on average to be larger than the hES-MP cells. Both cell lines demonstrated a reduction in size (shrinkage) with an increase in density. The hES-MP cells were found to be more elongated than the MG-63 cells and aligned along their major axis at higher cell confluences.

Time-lapse live cell imaging coupled with software assisted object tracking was used to continuously image and track individual cells of both cell lines across 2/3 days resulting in more than 450 cells being tracked over multiple generations. This large amount of data was analysed using object orientated code designed in python and fit to existing statistical models. Cell displacement was successfully modelled using a lognormal distribution which

was found to be statistically significant using both using Mann-Whitney and Wilcoxon rank sum test.

The relationship between the distribution of cell displacement and turning angle was found to fit that of a von Mises distribution with concentration parameter, κ , defined as the cell displacement for both cell lines. This suggests that cells with larger displacements are less likely to change their direction. The cell direction with respect to the horizontal was found to be randomly distributed for MG-63 cells, however for hES-MP cells, there appeared to be a synchronisation of directionality within the cells seeded in the same plate.

For MG-63 cells, lineage tracking was used to find the average cell cycle duration and visual tracking of cell death was used to determine the death rate. Also, the percentage of cells actively in the cell cycle per 24 h time frame (referred to as mitotic rate) was modelled using linear regression.

Multi-cell analysis was used to compare the growth of both cell lines. MG-63 cells were found to grow at a significantly faster rate in comparison to hES-MP cells. Cell numbers 24 h after seeding were compared with the initial numbers seeded and it was found that for an accurate estimation of the number of viable cells seeded, cell count of MG-63s on day 1 is sufficient whereas for hES-MPs day 2 is better. Boxcox transformation coupled with linear regression models were used as a quantitative means of comparing growth rates across experimental conditions.

The next chapter uses the results of both the qualitative and quantitative analysis to design the baseline (null) ABM for each of the cell lines that reflects the experimental lab conditions and specific cell line behaviours.

Chapter 5

Agent-based ‘null’ model

In order to develop a model of interacting cells, where hypothesis can be made regarding their rules of interaction, it was important to first establish a model of cell growth and motility for each individual cell with the assumption of no cell to cell interaction between cell types. This chapter reviews the development and validation of this model, referenced as the ‘null’ model.

5.1 Model development

The model developed is an agent-based model (ABM) with cells modelled as individual agents. ABM is a bottom-up discrete modelling approach which makes it an ideal tool for direct mapping of desired cell attributes and behaviours onto autonomous model

agents in order to explore their subsequent emergent behaviour. The advantages and limitations of ABM, are further discussed in section 2.2.2.1.

Following the CoSMoS process (described in section 3.3.1), during model development, the biological domain is translated into the platform and simulation model (Andrews *et al.*, 2010). Here the biological domain consists of individual cell attributes and rules regarding their motility and growth in an *in vitro* environment and as a result, these are the main components of the developed ABM, referred to as CellABM.

5.1.1 CellABM model outline

Python (ver. 3.7.4) accessed via the Anaconda package manager, was used to encode the CellABM using Spyder IDE. A full list of libraries utilised can be found in section 2.1. Figure 5.1 is the CellABM class diagram. Environment and cell class' are used to simulates a 2D *in silico* cell culture environment of cells growing in mono-layer.

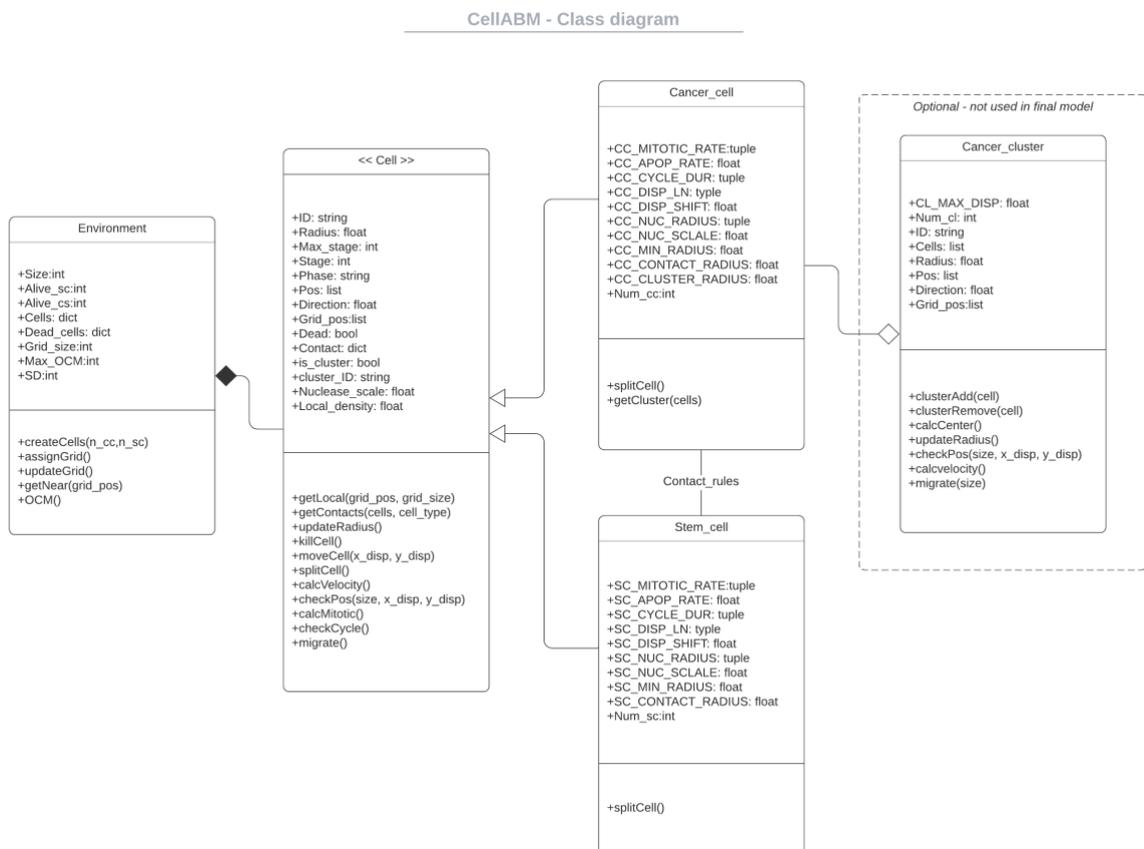


Figure 5.1: CellABM class diagram. There are two main class', Environment and Cell which have a composition relationship. Each cell type, Cancer_cell (cc) and Stem_cell (sc) are inherited from the parent Cell class. Cancer_clusters have an aggregation relationship with Cancer_cells, where Cancer_cell instances can join or leave Cancer_clusters.

The environment class creates and stores instances of the cell class. Only one instance of the environment class is created per simulation. There are two types of cells, cancer cells (cc) and stem cells (sc) both inherited from the parent cell class. An optional cancer cluster aggregate class exists to account for cancer cell clustering. This is not used in the final model simulations, however further explanation can be found in section (5.1.2.6).

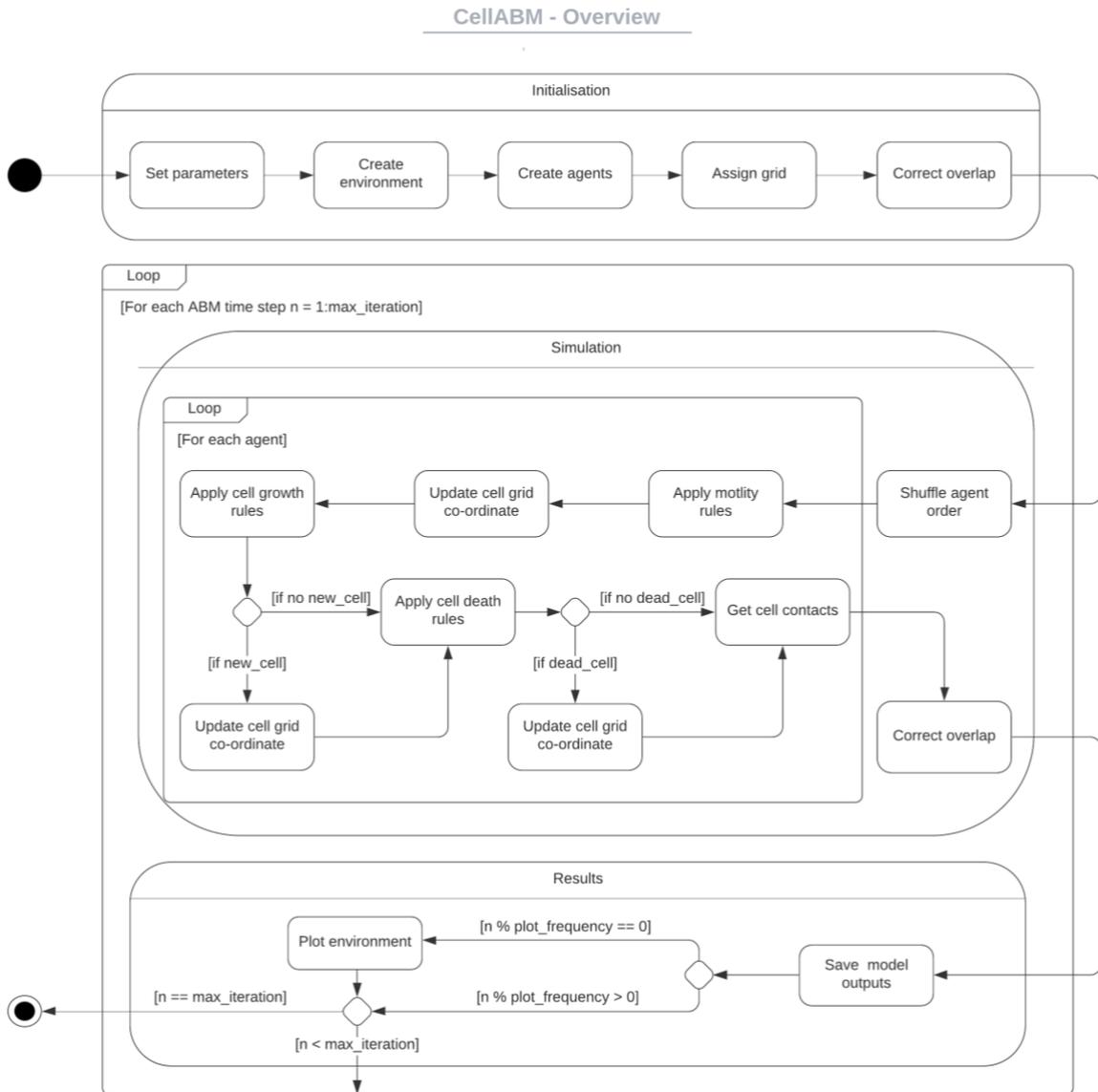


Figure 5.2: State and activity diagram depicting an overview of the CellABM. After model initialisation, the model is executed for each time step. At each time step the agent rules are applied asynchronously for each agent in the model. Overlap correction model is used to physically correct any unwanted overlapping of cells.

Figure 5.2 depicts an overview of the CellABM showing the three main stages, initialisation, simulation and results. Initialisation occurs only once per simulation and it is used to set local parameters, create the environment and populate the environment with agents based on the initial cell numbers (user input external parameter).

As CellABM is a temporal model, simulation and results are processed iteratively per model time step. Each time step simulates 20 minutes of real time. At each time step the model is 'solved' for each cell agent. This involves applying agent rules and is done asynchronously, which means the execution of agent rules per agent is done in series and the model environment is updated after each change. The agent order is randomly shuffled at each model time step in order to remove any bias from using a fixed execution order.

Model results are stored at each time step with an option to produce visual plots of the agents and environment, however as this is one of the more computationally taxing elements of the model and not always necessary to the analysis, there is an option to disable or limit to certain time steps only.

The overlap correction model is a physical model used to artificially correct any unwanted overlapping of cells. Although the mechanism is unrelated to the biological domain, the purpose of its implementation is to ensure that the cells grow in a mono-layer as seen in the biological domain. It is utilised at the end of cell creation in the initialisation stage and also during each time step of the simulation stage after all agent rules have been applied to each agent in the model.

5.1.2 The Agents

As mentioned above and shown in figure 5.1, cells are modelled as agents using the Cell class. Each of the two cell type class' (Cancer_cell and Stem_cell) inherit from the main Cell parent class and as a result have a very similar list of attributes and methods. The main differences are the cell type specific constants and the splitCell method that is used to generate new instances. Cancer and stem cell attributes and methods are based on the MG-63 and hES-MP cells presented in chapter 4.

A unique ID is generated for each new instance of the cell class, made up of a letter denoting the cell type and a unique number increasing in sequence with the number of instances created. The cell ID does not change throughout the simulation as it is the main key used to identify each unique cell at any point in the simulation.

The next sub-sections outlines all other agent attributes, methods and model parameters that are used in CellABM.

5.1.2.1 Cell shape

CellABM simulates cells grown in single layer mono-culture, therefore implementing the model in 2D space was considered a sensible abstraction. Results presented in section 4.1.1 were used as justification for modelling the cell agents as 2D circular objects. Each agent is created with a unique cell radius. The length of this radius is calculated by first generating the radius of the cell nucleus from a gaussian normal distribution using the `numpy.random.normal(μ , σ)` module, with μ and σ defined as the mean and standard deviation of the nuclear radius of cells measured experimentally (work presented in 4.1.1) and unique to each of the two cell types modelled.

In order to convert the nuclear radius into the cell radius, a scaling factor, `nuc_scale`, is applied. This is calculated from the model parameter `NUC_PERCENT` which is drawn from literature and is the percentage of the cell cytoplasm occupied by the nucleus. This calculation is shown in equation 5.1 and is based on the relationship between a sphere's volume and cross-sectional surface area.

$$nuc_scale = \sqrt[3]{100/NUC_PERCENT} \quad 5.1$$

5.1.2.2 Cell Motility

Unless stated otherwise, justification for all cell motility rules, parameters and distributions presented here can be found in the experimental work presented in section 4.1.2.

On model initiation, for each new instance of the cell agent, x and y positions are randomly assigned (enclosed within the environment space). This is done using the `numpy.random.rand` number generator and represents *in vitro* cell seeding. Cell direction is also randomly assigned here by drawing from a uniform distribution ranging from 0 to 2π using the `numpy.random.uniform` module.

Cell direction is defined as the angle of movement (in the anti-clockwise direction) with respect to the horizontal (x-axis). In addition to this angle, the model also uses

turning_angle which is the angle between the cell's velocity vectors from one time step (t) to the next ($t+1$). Both definitions are shown in figure 4.11 (section 4.1.2.4).

Figure 5.4 is the state and activity diagram for the cell motility rules applied for each agent at each model time step. The start point is the generation of a random displacement value from a lognormal distribution using the `scipy.stats.lognorm` module. As this is defined as the magnitude of the velocity vector, it cannot be less than zero, therefore a boundary condition to meet this requirement is applied. `scipy.stats.vonmises` is used to randomly generate the turning angle using the parameter `kappa`, which is calculated using the cell displacement as per findings presented in section 4.1.2.5. The new cell direction is simply the sum of the previous direction and the turning angle kept within the $0-2\pi$ range. Using basic trigonometry with the cell direction as θ , the `x_displacement` and `y_displacement` is calculated (equations 5.2a and b). If applying the new displacements does not force the cell out of the model environment, the cell position is updated with the new co-ordinates. Otherwise the model will loop over a finite number of attempts to try and find a displacement within the boundaries of the environment otherwise, the cell position does not change.

$$x_displacement = cell_displacement.\cos(\theta) \quad 5.2a$$

$$y_displacement = cell_displacement.\sin(\theta) \quad 5.2b$$

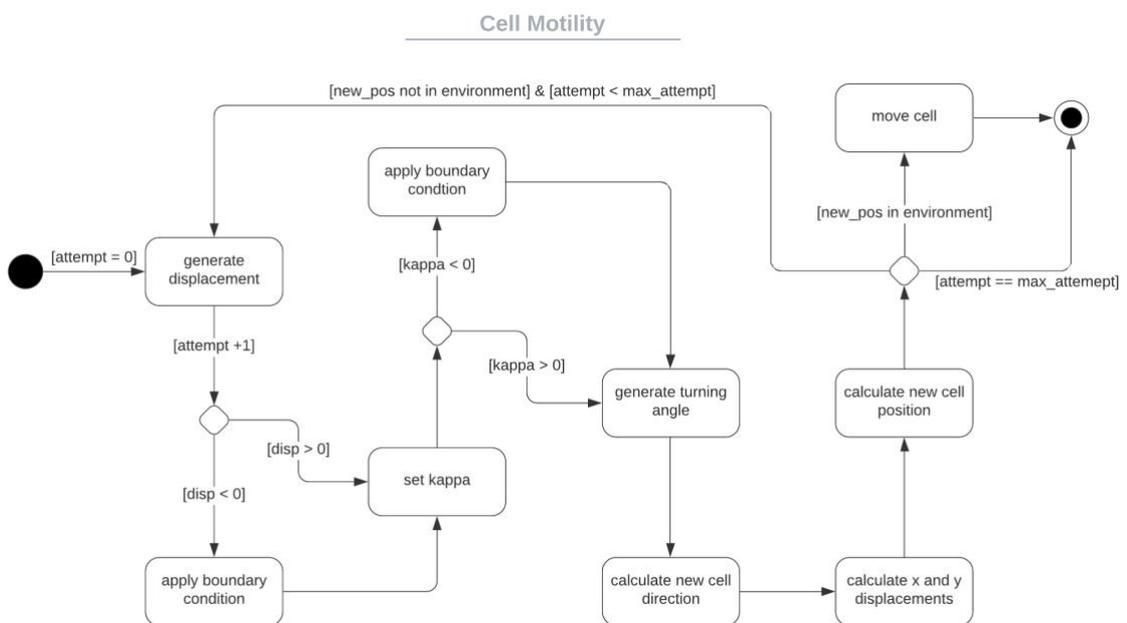


Figure 5.3: Diagram showing the cell motility rules implemented in CellABM. The displacement is randomly generated from a lognormal distribution and after applying boundary conditions it is used to randomly generate the turning angle from a von Mises distribution.

5.1.2.3 Cell growth

The cell growth rules are based on two aspects of the biological domain, the cell cycle and the mitotic rate. The cell cycle (figure 2.4 section 2.1.5) is a sequence of stages that the cell goes through from one cell division to the next. Hence the collective progression of cells through the cell-cycle leads to cell proliferation. This progression is regulated by a series of checkpoints. At the G1 checkpoint, also referred to as the restriction point, the cell size and environmental conditions are checked. If the cell meets these conditions, it can progress onto S phase irreversibly (Phillips *et al.*, 2013). During G1, if the cell cannot progress to S phase, it will enter G0, also referred to as resting or quiescent stage where a semi-permanent halt is placed on cell division. In CellABM, progression through the cell cycle occurs incrementally at each model time step for cells that are not in G0 phase.

To decide if a cell is actively progressing in the cell cycle or is in G0, the mitotic rate regression models presented in 4.2.3 are used. These are regression models of the relationship between the percentage of cells that split in a 24 h window and local cell density. The models, including both the linear regression relationship and confidence intervals, are saved directly as pickle objects and loaded onto CellABM at the initialisation stage (set parameters). The uncertainties module is used to generate a mean and standard deviation at any given local cell density using the linear regression models. This is then used to generate the mitotic rate (*m_rate*) from a gaussian normal distribution. For each agent at each time step, a mitotic rate unique to the cell is generated. In the event of cell splitting, the mitotic rate is used to determine if either of the new daughter cells are in G1 or G0.

The mitotic rate is also used to determine if cells in G0 should enter the cell cycle. In order to do this the mitotic rate per time step (*mRate_it*) is calculated. As the mitotic rate is over a 24hour window and each model time step is equivalent to 20minutes of real time, the mitotic rate here is divided by 72 to obtain the mitotic rate per time step.

During model initialisation, mitotic rate is also used to determine if each cell agent created is in G1 or G0. However for this purpose, a new parameter (*MRATE_S*) is introduced. This is used to calibrate the model to account for the lag phase in cell growth observed experimentally (further discussed in section 5.2.3)

With each new instance of the cell agent, a *max_cycle*, representing the maximum duration of the cell cycle, is randomly generated from a normal gaussian distribution. The mean and standard deviation used here are justified for the cancer cells, MG-63, in the experimental work presented in section 4.1.2.6. However these parameters could not be

experimentally measured for specific stem cells, hES-MPs and were based on literature derived values for general stem cells coupled with calibration.

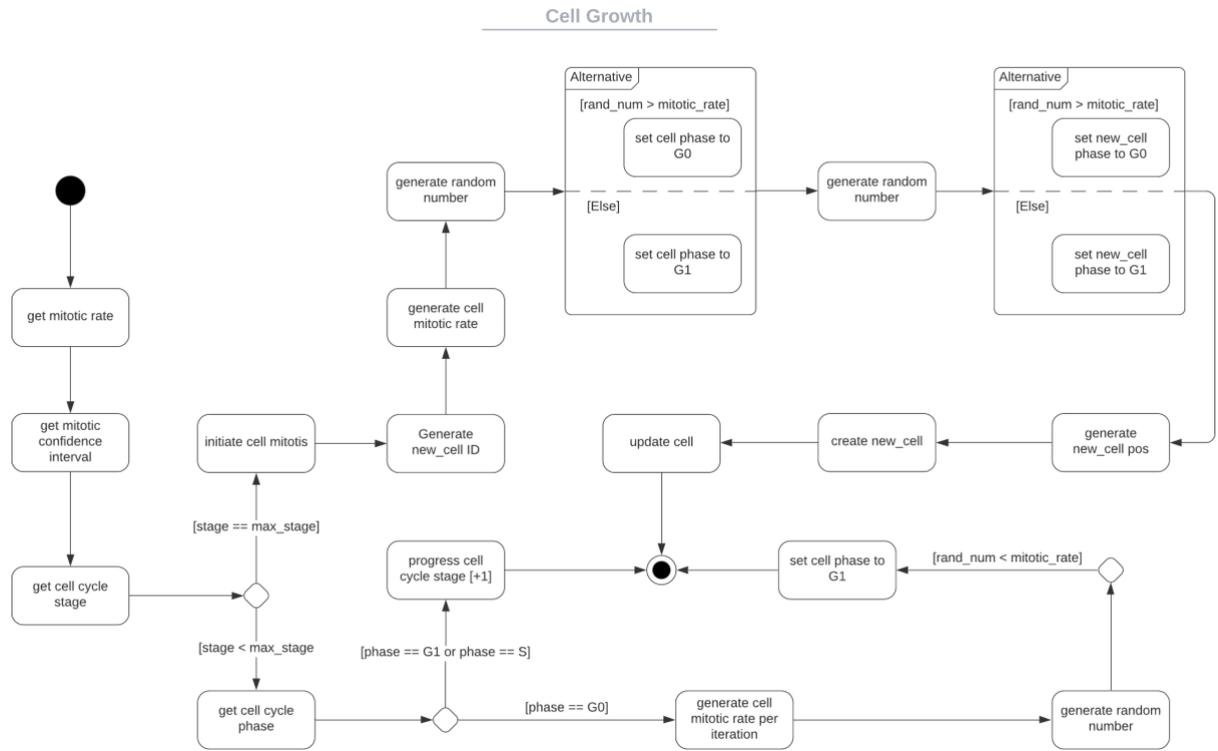


Figure 5.4: Diagram showing the cell growth rules implemented in CellABM.

5.1.2.4 Cell death

Based on the observed apoptosis rate presented in section 4.1.2.7, at each time step, each agent has a chance of undergoing apoptosis. This is done by generating a random number using the `numpy.random.rand` module. If this number is less than the `APOP_RATE` parameter specified for the cell type, then the cell is killed and removed from the environment list of alive cells and stored in a separate dictionary.

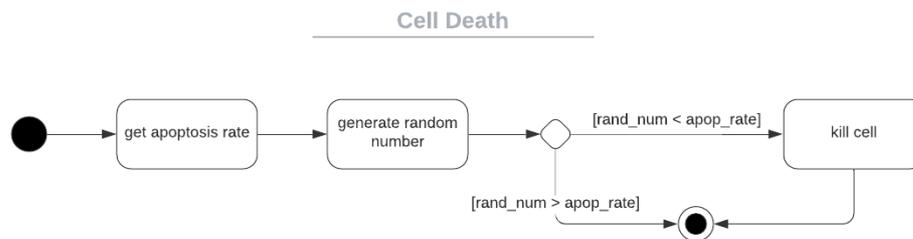


Figure 5.5: Diagram showing the implementation of cell death in CellABM.

5.1.2.5 Cell to cell contact

Cell agents in the ‘null’ CellABM model do not interact with each other. However, as this model is used as a baseline model for the development of the contact rules,, methods for checking contact were implemented. These methods check for direct cell to cell contact between agents of the cancer cell class and stem cell class. Three groups of contact data is gathered, the frequency of cancer to stem cell contact, the number of stem cells in simultaneous contact with a cancer cell and the length of non-transient contact. The effect of changes in input parameters on these three groups is explored using sensitivity analysis (see sections 5.2.5 and 5.2.6).

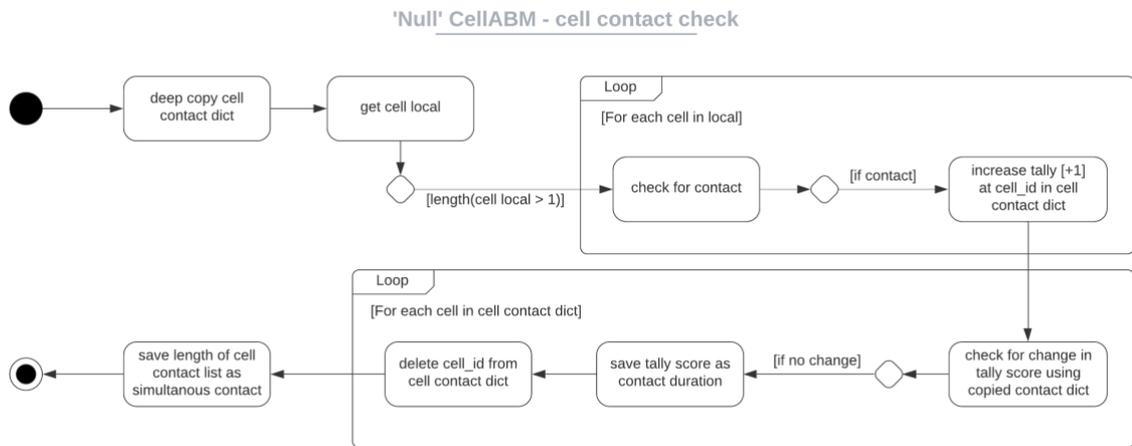


Figure 5.6: Diagram showing the methods implemented for 'null' model cell to cell contact check.

5.1.2.6 Cancer clusters

Cancer cell clustering is implemented in CellABM by employing the use of a cancer cluster aggregate class. This allows for a hierarchical relationship between individual instances of the cancer cell class and any cancer cell clusters that they might be a part of. Cells in a cluster refer to the cancer cell class for rules regarding growth, death and cell to cell contact, however cell motility rules are controlled by the cluster class.

Modelling cancer cell clustering was based on preliminary experimental observations (discussed and presented in section 4.1.2.8), however as further experimental analysis indicated the prevalence of clusters forming to be low, it was not used in final model simulations.

5.1.3 The environment

5.1.3.1 Model initialisation

The culture environment is modelled using the Environment class. On model initialisation, This is the first object created with only one instance. From here, cells are ‘seeded’ using the create_agents method and the overlap correction model (OCM) is initiated and controlled. The environment also holds all the key model elements, shown in the entity relationship diagram below (figure 5.7).

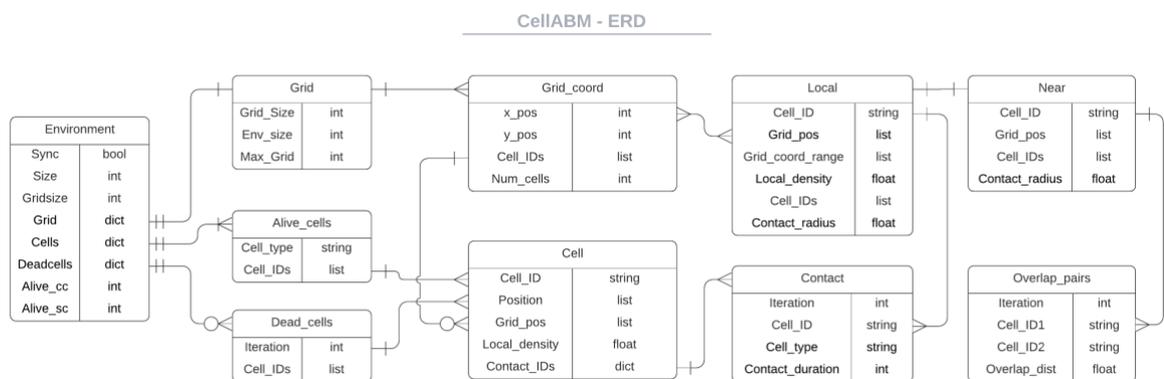


Figure 5.7: ERD showing the relationship between various elements of the ABM and the flow and storage of data.

The environment holds two python dictionaries which are used as a hash tables to store agents, alive_cells and dead_cells. Alive cells are stored using their unique agent ID as the key, whereas dead cells, are stored using the model iteration that they died at, as key. Each cell also uses a dictionary to store all cell to cell interactions. Here, IDs of all other cells interacting with the cell are used as key.

5.1.3.2 The grid

The environment also stores a dictionary for the grid. The grid is used to spatially discretise the environment in order to optimize execution of cell to cell interaction rules (e.g. contact rules and OCM). For these rules, each cell agent is required to have an awareness of its surrounding environment. Implementing a grid system means this is done more efficiently (results discussed in section 5.2.1.1).

The grid is created after the initial cell agents are created. The constant `GRIDSIZE`, is used to divide the environment. Then, for each agent, using the agent's position, the co-ordinates of the grid square it belongs to is calculated and saved. The grid dictionary, using the grid co-ordinates as key, is updated with each cell agent ID. Each grid co-ordinate is associated with a 'local' group which holds the eight squares surrounding each grid square and also a 'near' group, holding only four squares, show in figure 5.10. Local is used for cell to cell contact rules and in the OCM, the correction of overlap, whereas near is used for detection of overlapping pairs.

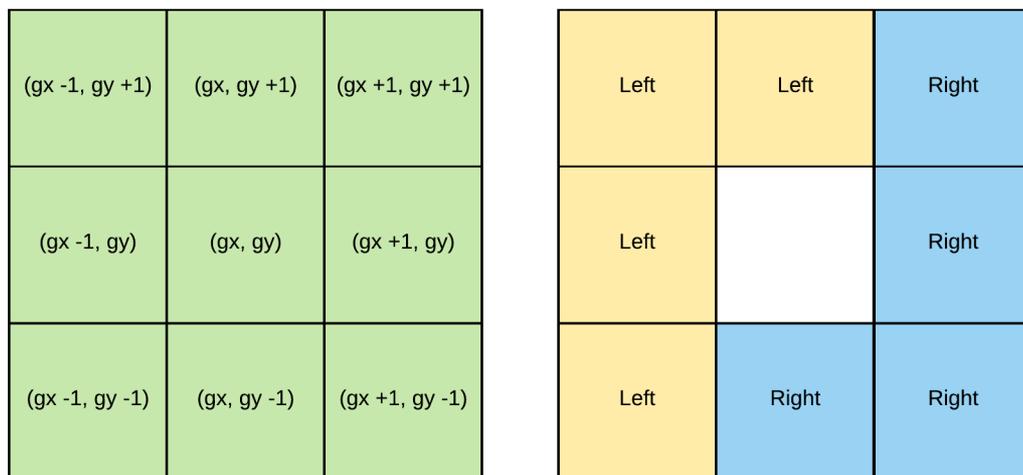


Figure 5.8: The 'local' (left) and 'near'(right) grid formations. Local is used for cell to cell contact and correction of overlap, near is used for detection of overlapping pairs. Both allow for a significantly faster execution of rules involving cell to cell interactions.

5.1.3.3 Overlap-correction model

As previously stated, the overlap-correction model (OCM) is implemented in order to artificially correct any unwanted overlapping of cells in order to ensure cell growth is done in mono-layers. It is executed after cell seeding and at the end of each ABM time step.

This model calculates the individual displacement for each cell using an adaptation of Newton's second law, with damping and no inertia to calculate the required displacement of each cell using the sum of the forces exerted on it by each of its neighbours. The equation implemented has been developed by Meineke *et al* (Meineke, Potten and Loeffler, 2001; Byrne, 2008) and is shown in equations 5.3.

$$x_i(t+\Delta t) = x_i(t) + \frac{\mu}{\eta} \sum u_{ij(t)} (x_{ij(t)} - L_{ij(t)}) \Delta t \quad 5.3a$$

$$y_i(t+\Delta t) = y_i(t) + \frac{\mu}{\eta} \sum u_{ij(t)} (y_{ij(t)} - L_{ij(t)}) \Delta t \quad 5.3b$$

To summarize the equation, the position of cell i at time $t + \Delta t$ is calculated using the positions of cell i and its neighbours, j , at time t , where u_{ij} is the unit vector in the direction of cell i to j , x_{ij} and y_{ij} are the distances between cells i and j (in the x and y directions respectively) and L_{ij} is the expected distance between both cells at zero overlap (equilibrium spring length). μ is the spring constant and η the damping constant. In the OCM, the ratio between these two constants is implemented using the `SD_CONSTANT`, and its optimum value is obtained via calibration (section 5.2.1).

Solving equation 5.3 is an iterative process which is shown in figure 5.11. A cap on the number of OCM iterations is placed using `MAX_OCM` that is also obtained via calibration. At increased cell confluency, where `MAX_OCM` has been reached without full correction of overlap, `reduce_radius` is applied to overlapping cell pairs. This is based on experimental work presented in section 4.1.1, where it was found that cell size decreases as confluency increases.

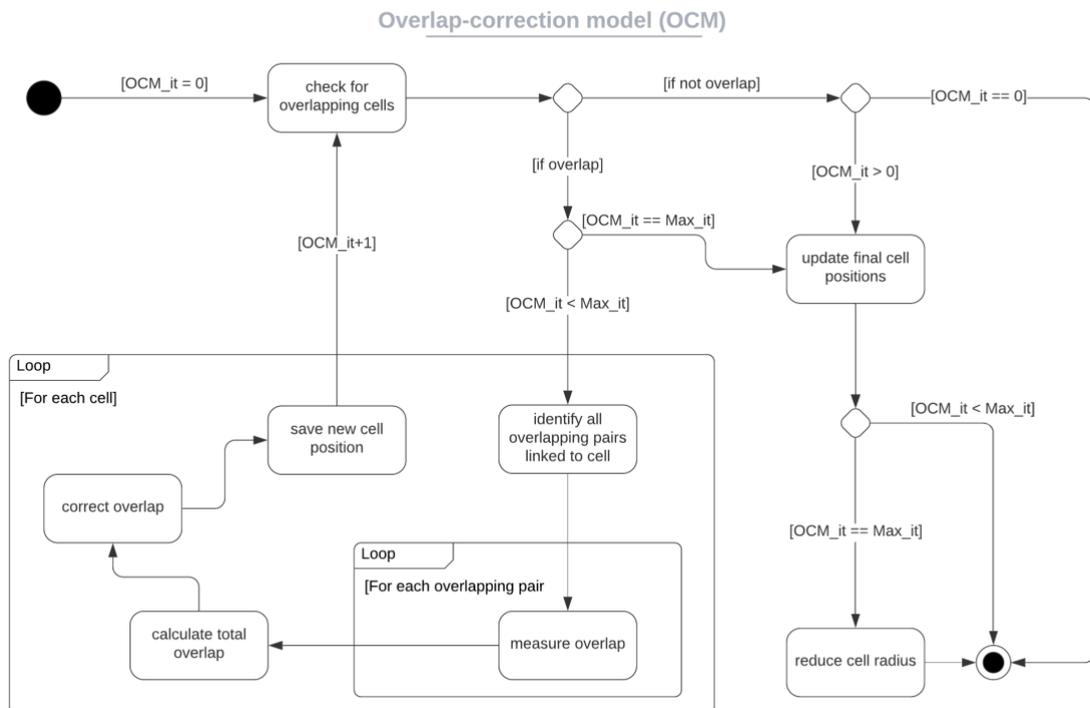


Figure 5.9 Diagram showing the OCM process.

5.1.4 Model parameters

5.1.4.1 Local parameters

Table 5.1: Parameter values. Parameters prefixed by *cc* or *sc* refer to the cancer and stem class respectively. The symbols used in the parameter column are as follows: μ = mean, σ = standard deviation, β = linear regression coefficient, α = linear regression intercept, *CI* = confidence interval, *s* = lognormal shape, *loc* = lognormal local, and *scale* = lognormal scale

| | Parameter | | Value | Unit | Justification/Notes | |
|---------------------------|------------------|--------------|---------------|---|---|---------------------------------------|
| Enviro- -nment | GRIDSIZE | | 50 | μm | Abstracted value (estimated using cell size) | |
| | MAX_OCM | | 100 | N/A | Calibrated (section 5.2.1) | |
| | SD_CONSTANT | | 0.25 | N/A | Calibrated (section 5.2.1) | |
| | PERMITTED_OL | | -5.0 | μm | Calibrated (section 5.2.1) | |
| Cell shape | NUC_PERCENT | | 10 | N/A (%) | Existing literature | |
| | MIN_NUC_RADIUS | | 5.0 | μm | Results presented in section 4.1.1 | |
| | cc.NUC_RADIUS | μ | | 11.457 | μm | Results presented in section 4.1.1 |
| | | σ | | 1.299 | | |
| | sc.NUC_RADIUS | μ | | 9.653 | μm | Results presented in section 4.1.1 |
| | | σ | | 1.527 | | |
| | PERCENT_R_REDUCE | | 10 | N/A (%) | Abstracted value (estimated using min cell size) | |
| Cell motility | | <i>s</i> | 0.661 | | | |
| | cc.DISP_LN | <i>loc</i> | -0.617 | μm | Results presented in section 4.1.2.3 | |
| | | <i>scale</i> | 4.036 | | | |
| | sc.DISP_LN | <i>s</i> | 0.450 | μm | Results presented in section 4.1.2.3 | |
| | | <i>loc</i> | -5.830 | | | |
| | | <i>scale</i> | 19.769 | | | |
| | cc.DISP_SHIFT | | 7.0 | μm | Results presented in section 4.1.2.5 | |
| sc.DISP_SHIFT | | 5.0 | μm | Results presented in section 4.1.2.5 | | |

(contiued)

| | Parameter | Value | Unit | Justification/Notes | |
|--------------------|----------------|---------------------|------------------------|--|--------------------------------------|
| Cell growth | MITOTIC_WINDOW | 24 | hour | Results presented in section 4.2.3 | |
| | | β | -0.0488 | | |
| | cc.MRATE | β CI | $\pm 6.06 \times 10^3$ | N/A | Results presented in section 4.2.3 |
| | | α | 152.5 | | |
| | | α CI | ± 9.697 | | |
| | sc.MRATE | β | -0.1334 | N/A | Results presented in section 4.2.3 |
| | | β CI | $\pm 2.75 \times 10^2$ | | |
| | | α | 117.6 | | |
| | | α CI | ± 12.728 | | |
| | | cc.MRATE_S | 0.8 | N/A | Calibrated (section 5.2.4) |
| | | sc.MRATE_S | 0.5 | N/A | Calibrated (section 5.2.4) |
| | | cc.STAGE_S | 18 | hour | Calibrated (section 5.2.4) |
| | | sc.STAGE_S | 18 | hour | Calibrated (section 5.2.4) |
| | cc.MAX_STAGE | μ | 24.937 | hour | Results presented in section 4.1.2.6 |
| σ | | 3.861 | | | |
| sc.MAX_STAGE | μ | 28 | hour | Existing literature and Calibrated (section 5.2.4) | |
| | σ | 3 | | | |
| Cell death | cc.APOP_RATE | 3.283×10^5 | N/A | Results presented in section 4.1.2.7 | |
| | sc.APOP_RATE | 6.132×10^5 | N/A | Results presented in section 4.1.2.7 | |

5.1.4.2 External parameters

CellABM takes four main input parameters that must be provided for the model to run and five optional parameters.

size (int): Size of one length of the square model environment with unites in μm . For example size = 1000 creates an environment equivalent to 1mm^2 of growth surface area.

ncc (int): Number of cancer cells at model initialisation or ‘cell seeding’.

nsc (int): Number of stem cells at model initialisation or ‘cell seeding’.

iterations (int): Number of iterations to be executed. Each iteration is equivalent to 20minutes of real time. For example, iterations = 288, simulates 4 days of cell growth.

directory (string): Optional to specificity a specific location to save model outputs. If none is provided, model outputs will be saved in the same directory in a folder named ‘CellABM_Results’.

frequency (int): Optional to specify if plots are produced and at what frequency. For example frequency = 5, will produce and save a plot every 5th iteration of the model and frequency = 0, will not produce any plots. If no value is provided, frequency is set to zero, no plots produced.

labels (bool): Optional to produce plots with cell ID labels included. Default set to False, no labels.

sync (bool): Optional to solve the model either synchronously or asynchronously. The default is set to False and all simulations and results in this PhD are done with an asynchronous model.

fixseed (int/None): Optional to allow for fixing of random number generator seed for testing and model reproducibility. Default is set to None with no specific seed set.

5.2 Model calibration and validation

5.2.1 Verification and calibration of OCM

The OCM is implemented to ensure cells maintain a single-layer growth configuration. Due to its iterative nature, it is one of the more computationally taxing parts of the overall CellABM. Model optimization and calibration is used to lower the execution time. The results presented for the OCM are based on the model's implementation in the initialisation stage after cells have been seeded at high density. At this stage, the highest amount of cell overlap is shown as cell positions are randomly assigned, making it ideal for OCM exploration.

5.2.1.1 Verification of optimization

With the implementation of the grid discussed in section 5.1.3.2, The execution time of the OCM decreased by 98.5% shown in figure 5.12 (left). The results are the average of 100 simulations. This was verified by running both models with the same random number generator seeds in order to ensure the model optimisation has not affected the OCM output, this is shown in figure 5.12 (right).

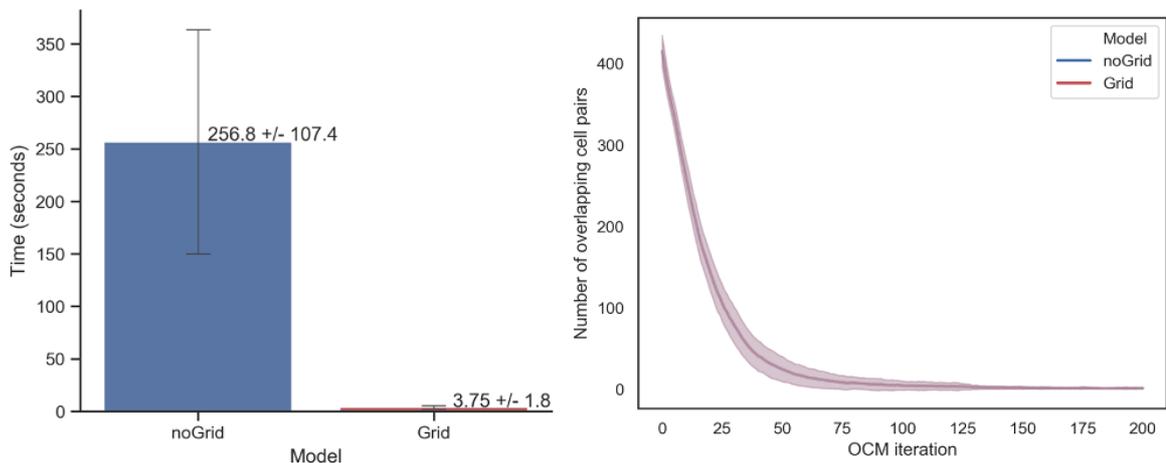


Figure 5.10: Left - bar chart showing the OCM execution time with and without the implementation of the grid. With the implementation of the grid, the execution time of the OCM decreased by 98.5%. Right – The number of overlapping cell pairs at each OCM iteration for both models used as verification that the model optimization did not alter the model output. The results are the average of 100 simulations with the random number generator seed fixed across both models. Errors shown are the standard deviation.

5.2.1.2 Calibration

The following three parameters were calibrated for the OCM:

SD_CONSTANT: this is the ratio between the spring and damping constant of the equation implemented in the model (5.3). This enhances the repulsive force applied to the cell by its neighbours. Therefore, larger values, result in a larger dispersive effect which can reduce the number of iterations needed to complete the model. However, values that are too large will result in the cells being pushed too far apart and can directly interfere with the ABM emergent behaviour.

PERMITTED_OL: this is the permitted overlap distance which is ignored by the model when checking for overlap only. It is a boundary condition applied in order to prevent the model from oscillating unnecessarily over small overlap values.

MAX_OCM: this is the maximum iterations the OCM will run for each time it is called. This limit is set to prevent the OCM getting stuck in situations where the environment is confluent with agents to the extent that overlap cannot be fully removed. In such cases, the cell radius of overlapping pairs of cells is reduced. This is based on the results presented in section 4.1.3 and is capped at the minimum cell radius found experimentally.

Initially various values for SD_CONSTANT (SD) and PERMITTED_OL (OL) were explored visually. This was used to select a sensible subset of values for further calibration, which involved running the model for each of the SD-OL combinations. The random number generator seed was kept the same across each combination and an average of 100 simulations was taken. From the various model outputs analysed, the bar chart in figure 5.13 shows the MAX_OCM.

The SD-OL combinations that produced the lowest MAX_OCM were compared statistically. SD-OL combination of 0.25 and $-5\mu\text{m}$ respectively was found to finish on a significantly smaller MAX_OCM than any other combination ($p < 0.0001$ for the nearest increase in SD at 0.2).

The OCM behaviour at the selected parameters was further analysed in order to determine the final value for MAX_OCM. It was found that the MAX_OCM can be reliably set to 100 with room to further reduce to as low as 40 if needed. OFAT analysis with the MAX_OCM is also carried out in order to assess the sensitivity of the model outputs to changes in this parameter (discussed in 5.2.5).

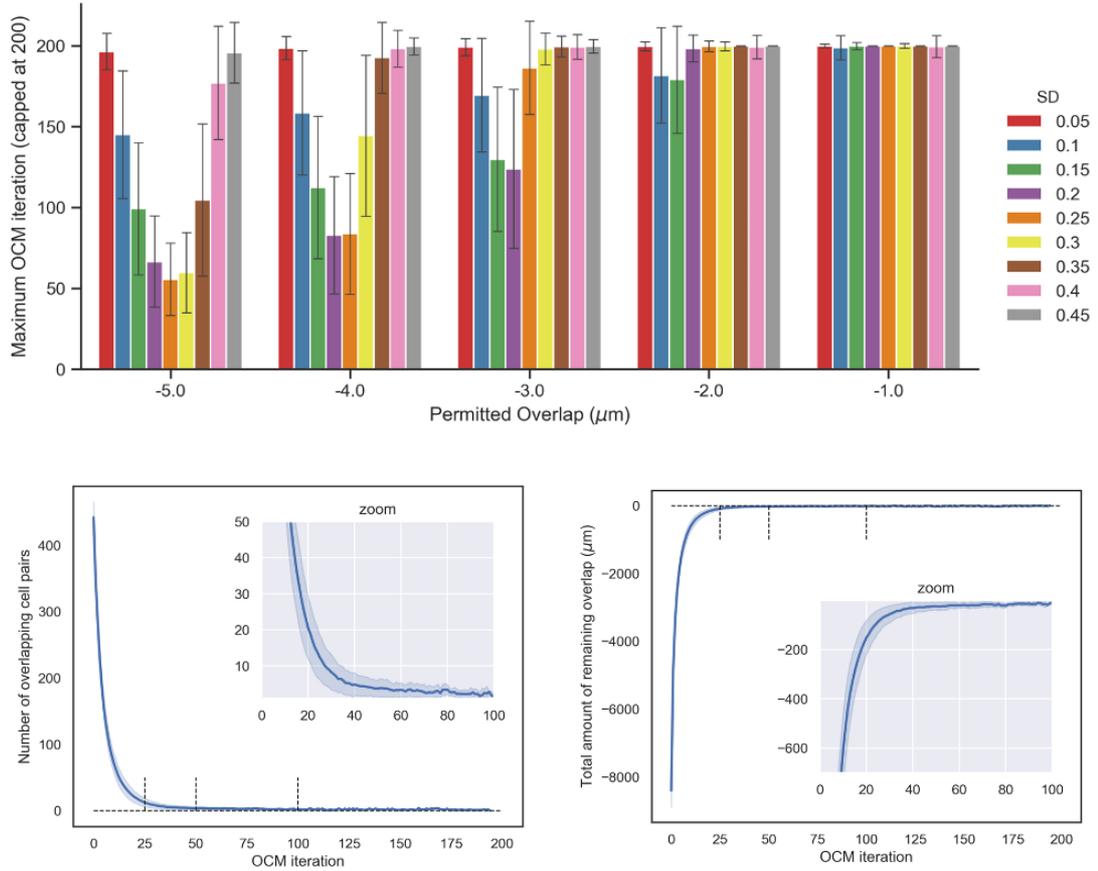


Figure 5.11: Top (bar chart) – The effect of alterations in `SD_CONSTANT` and `PERMITTED_OL` values, on the `MAX_OCM` required to solve the OCM. `SD-OL` pairing at 0.25 `SD` and -5`OL` resulted in the smallest `MAX_OCM` and was subsequently selected. Bottom – the behaviour of the OCM at the selected parameters, showing the changes in the number of overlapping pairs and amount of overlap with each iteration of the model. This was used to determine the `MAX_OCM`.

5.2.2 Aleatory uncertainty analysis

The use of random number generators implemented at various stages of the CellABM introduces stochasticity and as a result aleatory uncertainty. In order to mitigate this, replicate simulations have to be run per analysis. To determine the number of replicates needed (referred to as sample size) the following analysis were carried out.

5.2.2.1 Preliminary uncertainty analysis

Initial analysis was done by calculating the 95% confidence interval and coefficient of variation, defined as ratio of standard deviation to mean, for the model outputs for increase sample sizes. A pool of samples was created by running the model 500 times with internal

parameters fixed. Samples were randomly drawn from and the data was visualised in order to obtain an approximate for the sample size at which a steady state is achieved. An example of this is shown in figure 5.14. The approximate sample size was used for analysis using Vargha-Delaney A-Test.

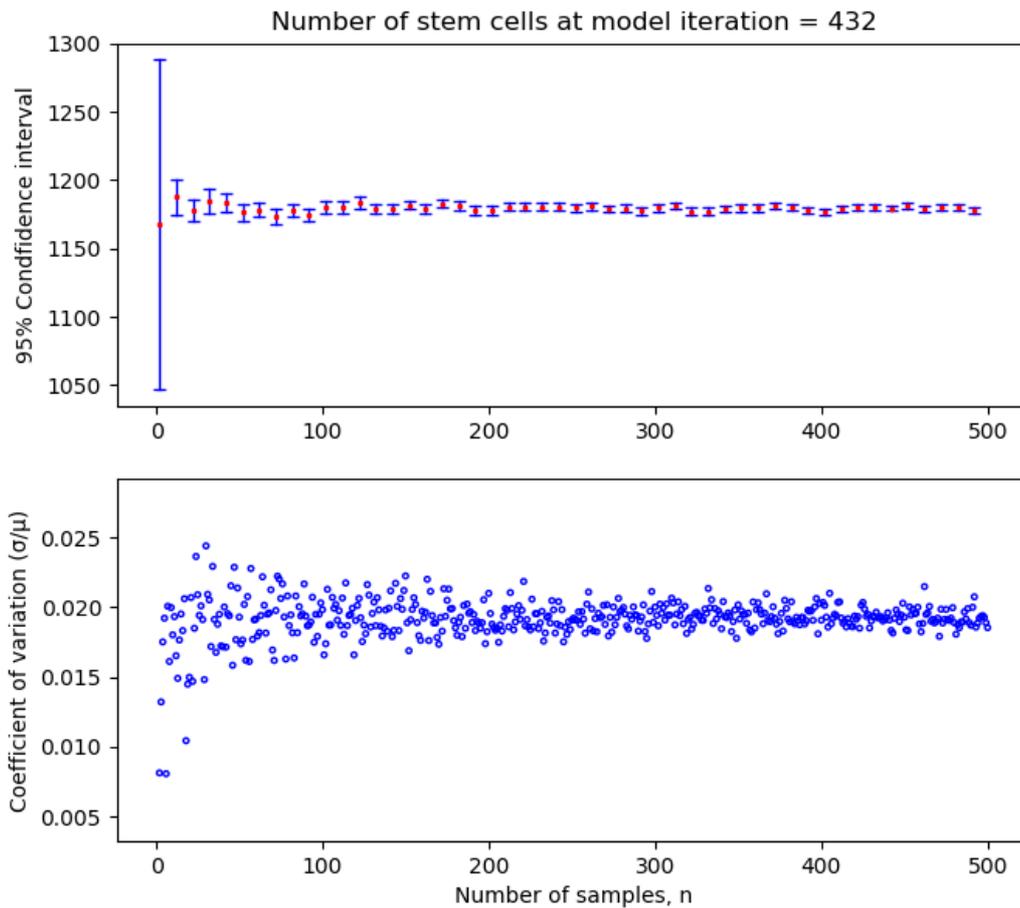


Figure 5.12: A visualisation of the effect of sample size on coefficient of variation and 95% confidence interval for cell number output at a fixed time step of the model. Each sample size plotted was selected randomly from 500 model simulations (at $n = 500$, the entire sample pool was used).

5.2.2.2 Uncertainty analysis with Vargha-Delaney A-Test

Using the initial assessment as a guide, a range of sample sizes were selected for further analysis. These were 1, 5, 10, 25, 50, 100, 150, 200 and 250. For each of the selected sample sizes, 20 sets of simulations were run, resulting in a total simulation number of 15,820. For each sample size, the median response per sets 2-20 was contrasted with the first set's median response using the Vargha-Delaney A-Test (Vargha and Delaney, 2000). This is a non-parametric measure of effect size that allows for the comparison of two populations with

results ranging from 0-1. A result of 0.5 indicates no difference with values above or below this indicating an increase in difference (table 3.3 in section 3.3.4.2). Here as we are only comparing the magnitude of effect (and not its directionality) all results below 0.5, were assigned corresponding values above 0.5.

Out of the six model outputs explored, five produce statistically consistent distributions at relatively low sample size (<10) with the output showing the largest variation being the ratio of cancer cells to stem cells. This is very likely due to the fact that ratio is calculated based on two other model outputs, therefore, creating a compound magnitude effect in instances where both of those outputs vary in opposite directions.

Through this technique, developed by (Read *et al.*, 2011), the number of samples required in order to obtain a statistically consistent output is determined, visualised in figure 5.15. A sample size of 150 for co-culture models and 50 for mono-culture models is concluded sufficient for mitigating aleatory uncertainty in CellABM. Individual results per sample size is included in appendix (8.5).1

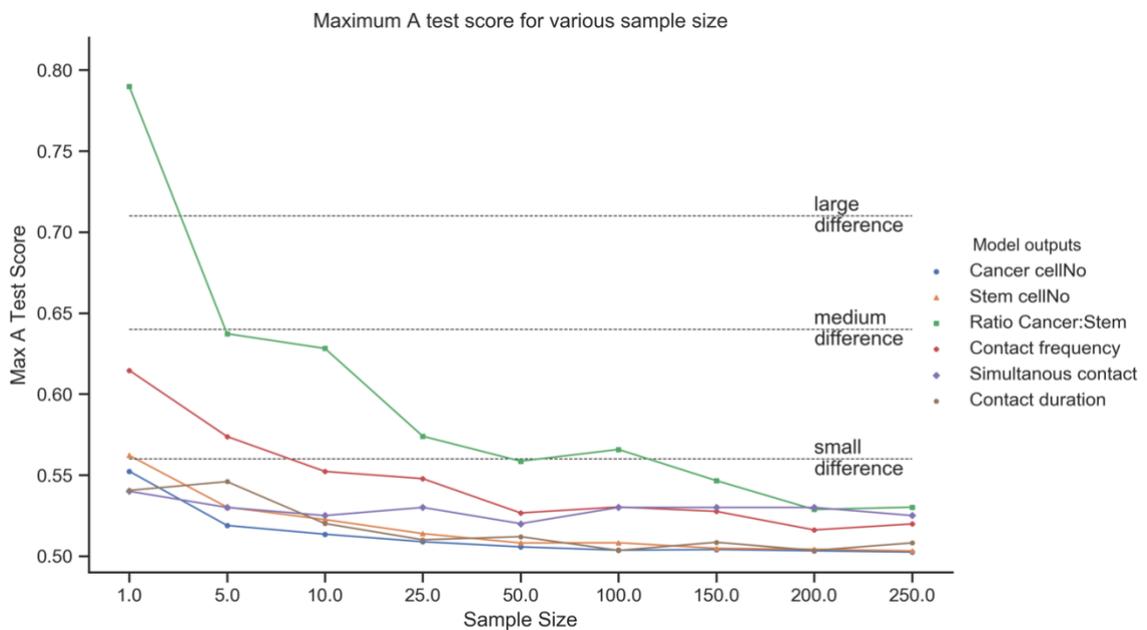


Figure 5.13: The maximum A-test score out of all 20 sets per sample size. Each sample size represents the number of repeat runs of the simulation from which the median responses are used in statistical comparison using the Vargha-Delaney A-Test. Horizontal lines labelled mark the magnitude of effect with a score of 0.5 exactly indicating no effect. Scores below 0.5 were assigned corresponding values above 0.5 as the focus of this analysis was magnitude of effect only and not its direction. It was concluded that a sample size of 50 for mono-culture models, is sufficient whereas a sample size of 150 is required for co-culture models.

5.2.2.3 Effect of model scale on aleatory uncertainty

CellABM is modelled in 2D space, within a confined environment. At the start of the simulation, the number of agents and size of the environment is declared. Keeping the ratio of number of agents to environment surface area fixed, the effect of up scaling the number of agents the model is executed for was explored. In the previous section, the uncertainty analysis done was using 100 agents in an environment size of 1 mm². Here, two additional model scales are explored, 400 and 900 agents with the environment size at 4mm² and 9mm² respectively. This ensures the ratio of agents to surface area is fixed at 10,000 across all three scales, representing 10,000 cells per cm² seeding density (recommend for *in vitro* experiments at medium confluency). For each of the model scales, 500 model simulations were run; 20 sets of simulations of sample size of 25. As before, the median responses of each of the sets 2-20 were compared with that of set1. The maximum A-test scores for each output at each model scale was compared.

The results indicate larger model scales produce more statistically consistent distributions and is observed most clearly with outputs that have a larger variance, such as ratio of cancer to stem cells. This is consistent with the fact that in an ABM, rules and behaviours are executed per agent and so stochasticity is introduced at the agent level. Therefore as larger model scales have larger number of agents, they will produce more statistically consistent distributions.

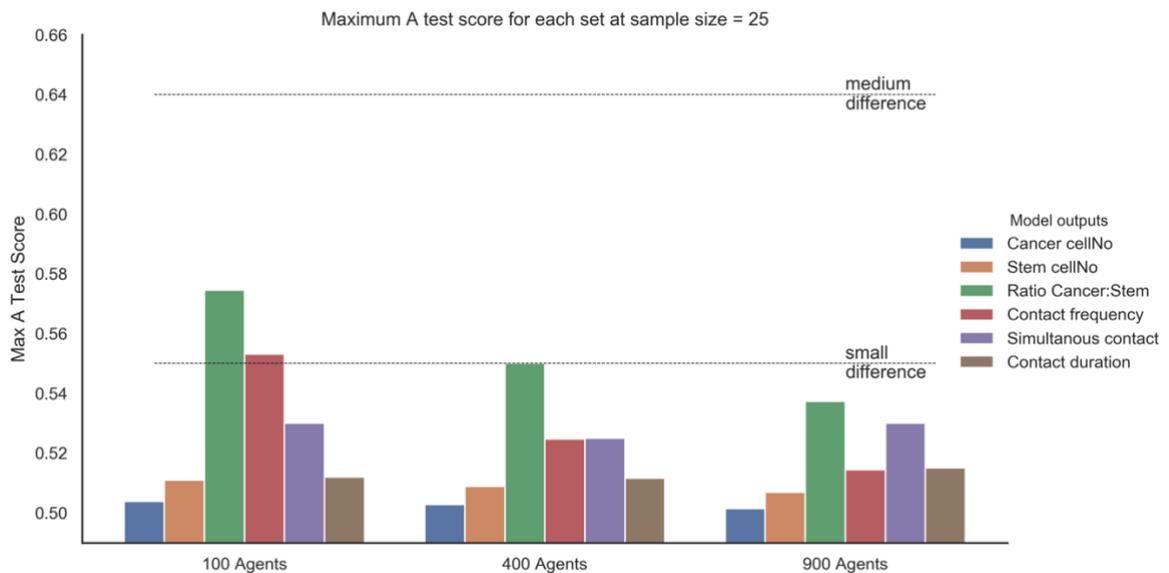


Figure 5.14: Maximum A-test scores for various model scales increasing in agent number while keeping the ratio of agents to environment surface area fixed. Increasing the number of agents, produces more statistically consistent outputs.

5.2.3 Calibration against experimental data

The majority of the parameters used in CellABM were based on lab derived experimental measurements presented in section 4. However, some parameters could not be verified experimentally but were key in the model design. For these, calibration was used.

For each cell type, the model was calibrated to account for the cell specific lag phase observed experimentally. The two parameters key to this were MRATE_S and STAGE_S, both used only in the model initialisation stage. The first refers to the mitotic rate used to determine if a cell is in G1 or G0 and the second, the cell cycle stage the cell starts at. For stem cells a further level of calibration was done for the cell cycle duration (MAX_STAGE). All calibration was done against growth curves gathered experimentally for each cell (presented in section 4).

The model was run 25 times (to mitigate aleatory uncertainty) for various combinations of parameters (referenced to as scenarios) reasonable for the biological domain. The root mean squared error (RMSE) between each model output and its respective experimental time point was measured and averaged for the 25 simulation replicates. This was analysed both for each time point and as a sum across all time points. The parameter combinations with the lowest RMSE (falling in the bottom quantile) were further analysed using ANCOVA (analysis of covariance), with time as the co-variate.

5.2.3.1 Cancer cells

Overall 45 unique scenarios were considered for cancer cells, with MRATE_S ranging over [0.1 : 1.0] and STAGE_S from [4 : 18]hours. The simulation results were compared against growth curves of MG-63 cells seeded at a density of 10, 000 cells/cm². In Figure 5.17, an example output from the first stage of the calibration process is shown. This is the total RMSE summed across all time points before and after a mask was applied to identify the scenarios that fall into the lower quantile which were then analysed using one-way ANCOVA. Using this test allowed for the comparison of the cell numbers produced by the model and experimental data while controlling for time (iteration number). Levene's test was used to ensure homogeneity of variance. Non significance in both tests ($p > 0.05$) indicated similarity in the experimental (*in vitro*) and computational (*in silico*) models. The combination of STAGE_S = 18hours and MRATE_S = 0.8 were found the best fit out of all the scenarios tested.

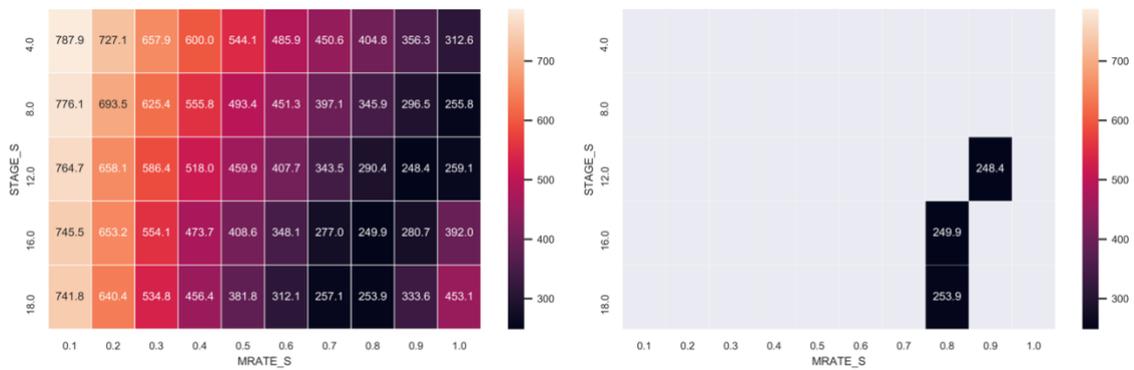


Figure 5.15: An example of the visualisation technique used in model calibration. Here the sum of all RMSE per scenario is displayed in a heatmap (left) and a filter is applied to identify the scenarios falling in the lower quartile (right). These scenarios were further analysed using ANCOVA (with time as the covariate).

5.2.3.2 Stem cells

For stem cells, calibration was more challenging. First, the stem cells used experimentally (hES-MPs) were observed to have a much longer and variable lag phase (up to 48 hours). This is the elapsed time from when the cells are seeded to when they begin growing during which the cells adapt to the culture environment. In order to save on computational time and because this part of cell growth was not relevant to the model, the modelled lag phase for stem cells was limited to the first 24hours. In other words, on model initialisation, the stem cells are assumed to have been in culture for the previous 24hours. To do this, the model was calibrated against hES-MP growth curve measurements with a 24hour shift (e.g. day 1 of the *in silico* model was calibrated with day 2 of the *in vitro* culture measurements).

The other added complexity, was the extra level of calibration needed to account for the uncertainty of the MAX_STAGE value used for the cells. As discussed in the work presented in section 4, the methodology used to extract this parameter value for the cancer cells could not successfully be applied to the stem cells. As a result, calibration was done with 270 unique scenarios with MAX_STAGE ranging over [23 : 28] hours and MRATE_S and STAGE_S the same as before [0.1 : 1.0] and [4 : 18] hours respectively. The full results are presented in appendix (8.6). As before, low RMSE coupled with non-significant ($p > 0.05$) one way ANCOVA with Levene's test for homogeneity was used as secondary criteria to assess model fit. The combination of STAGE_S = 18hours, MRATE_S = 0.5 and MAX_STAGE = 28 hours were established as the best fit out of all the scenarios tested.

5.2.4 Validation against experimental data

The model simulation was validated both qualitatively, through face validation against microscopy images of cell growth, and quantitatively using effect size measures such as two sample Kolmogorov-Smirnov test and Cohen's 'δ' statistic against four separate sets of experimental data collected under different initial seeding conditions (experimental data presented in section 4.2).

5.2.4.1 Face validity

Face validation involved comparison of the simulation output with respective biological domain model representation; microscopic images of each cell at specific time-points. Figures 5.18 and 5.19 show an example of this for model time step 288 equivalent to 4 days of real time growth of cancer (MG-63) cells and 5 days for stem (hES-MP) cells. On initial comparison, the confluency of the cells in the model output appear very similar to the experimental observations. For cancer cells the homogeneity and spread also seems similar. For stem cells this appears not as well matched which is very likely to be due to the abstraction of the cell shape as circular.

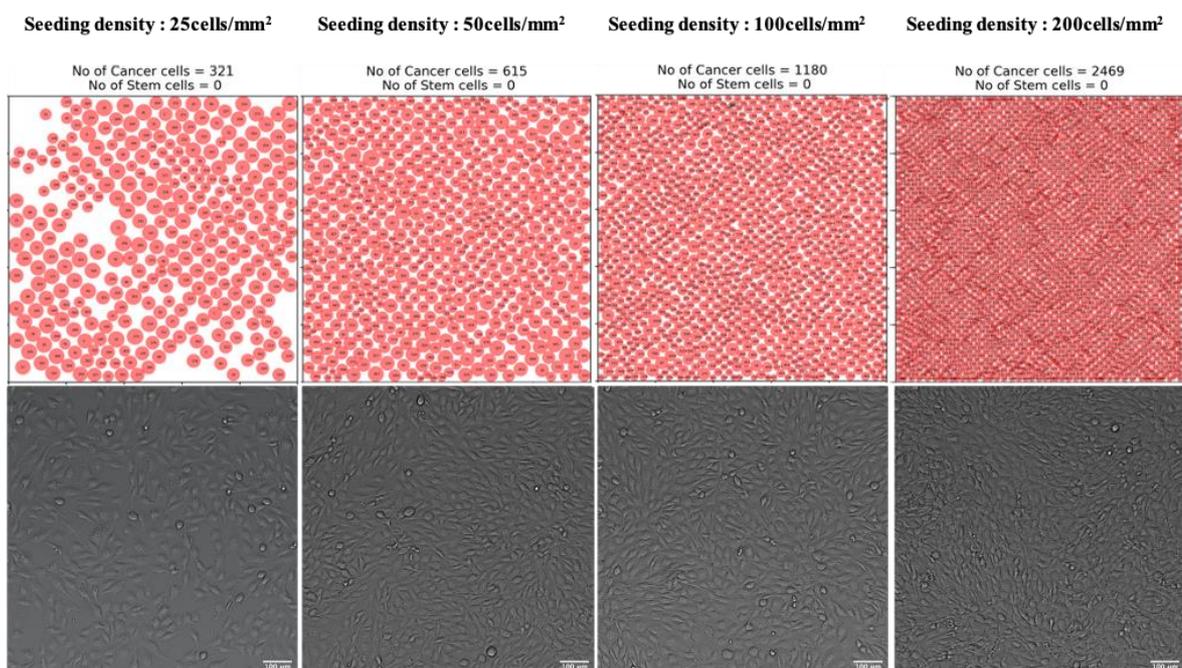


Figure 5.16: Face validation of the ABM simulation outputs of MG-63 growth on day 4 (equivalent to 288 model time-steps) shown for each four different seeding densities.

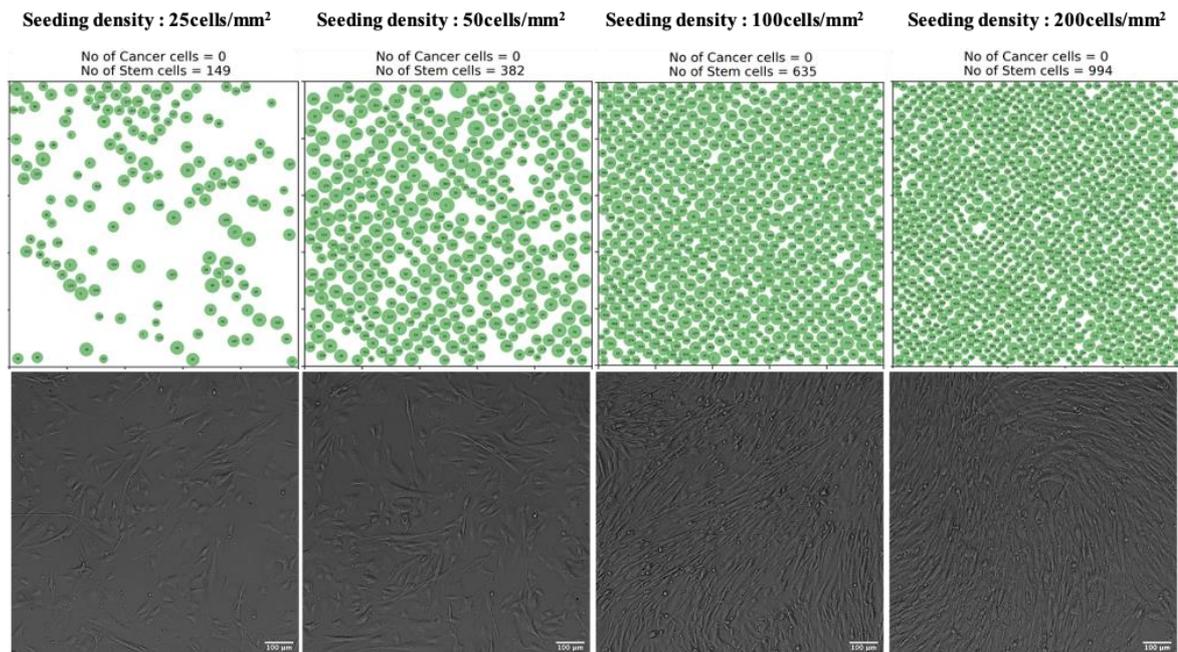


Figure 5.17: Face validation of the ABM simulation outputs of hES-MP growth on day 5 (equivalent to 288 model time-steps) shown for four different seeding densities.

5.2.4.2 Statistical analysis

For a quantitative approach to model validation, the cell numbers produced by the ABM and those obtained experimentally, were compared at key time-points. Two sample Kolmogorov-Smirnov (K-S) test and Cohen's 'δ' statistic were used as effect size measures. For K-S, $p > 0.05$ was considered statistically similar and for Cohen's 'δ' statistic, $\delta < 0.8$ were considered acceptably similar (in the small and medium range of the effect size). A full table of results is presented in appendix (8.7) with visual representations in figures 5.20 and 5.21. The model was found statistically valid for medium to low seeding densities up to 288 model iterations. Both models could not be validated for high seeding densities.

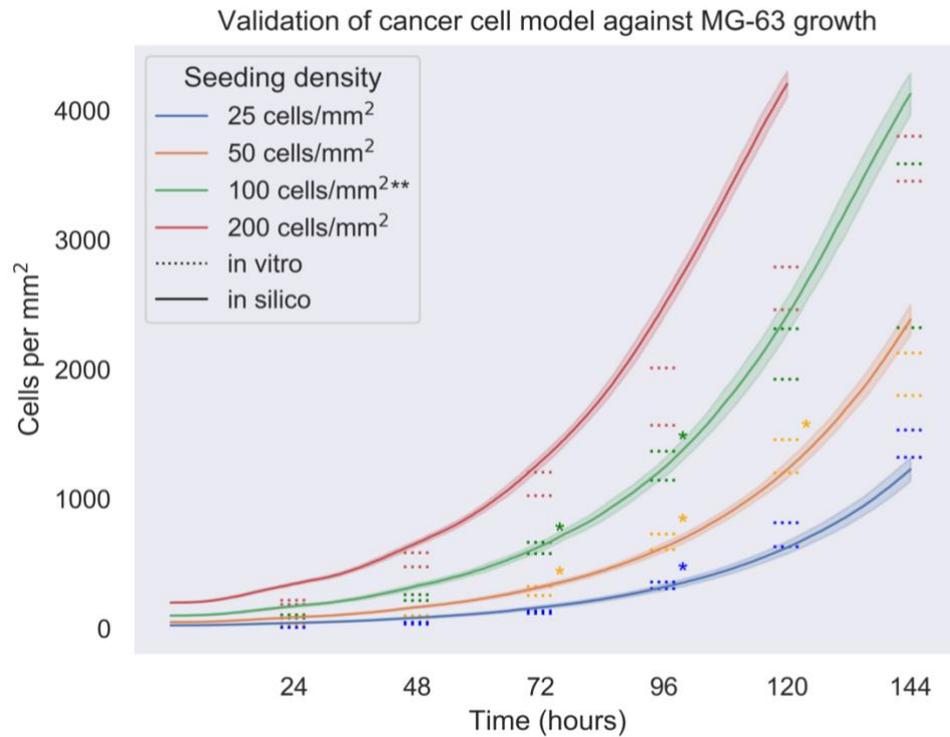


Figure 5.18: Validation of the ABM for cancer cell mono-culture growth against four separate experimental data sets. The model was found statistically similar ($K-S p > 0.05$ and $\delta < 0.8$) for time points marked with * and hence statistically valid for medium to low seeding densities up to the 4 day time point equivalent to 288 model iterations. ** marks the dataset the model was calibrated against.

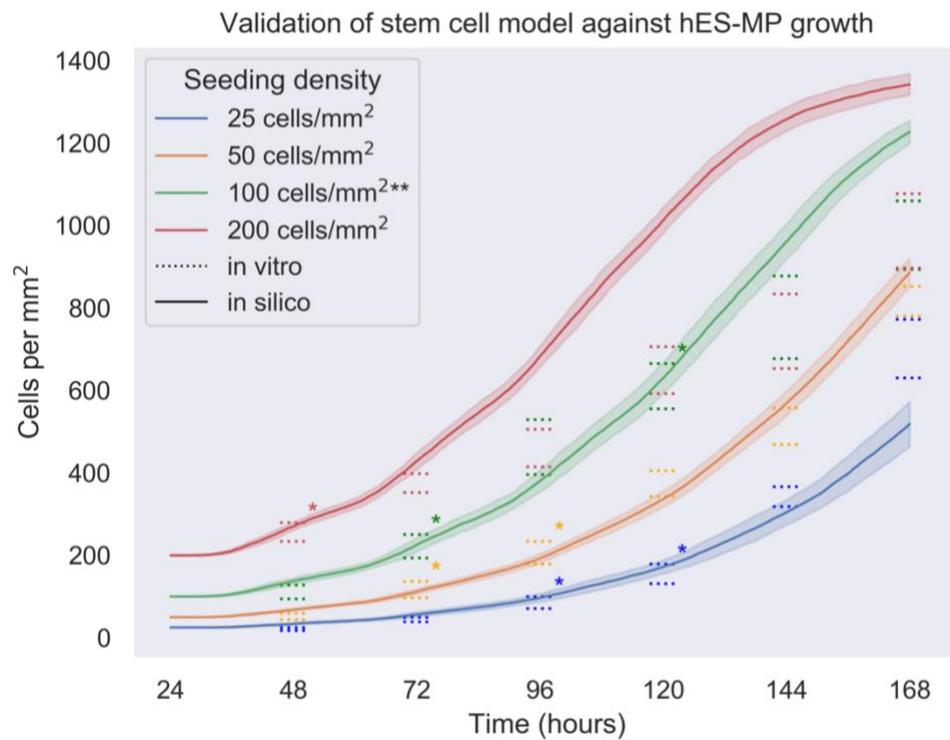


Figure 5.19: Validation of the ABM for stem cell mono-culture growth against four separate experimental data sets. The model was found statistically similar ($K-S p > 0.05$ and $\delta < 0.8$) for time points marked with *

*and hence statistically valid for medium to low seeding densities up to the 5 day time point equivalent to 288 model iterations. ** marks the dataset the model was calibrated against.*

5.2.5 Local sensitivity analysis

Sensitivity analysis is useful for understanding the relative sensitivity of the simulation's output to its inputs. From all the input parameters of the model, those with the most epistemic uncertainty were explored using one-factor-at-a-time (OFAT) sensitivity analysis. These include parameters that were added as boundary conditions, model abstractions or could not be verified experimentally. Each parameter was altered by $\pm 30\%$, $\pm 20\%$, $\pm 10\%$ and $\pm 5\%$ of the final (calibrated or otherwise) value listed in section 5.1.4.1. The model was simulated for 150 repeats for each parameter change and the output simulation results were contrasted with that of the baseline model (with no parameter change) using Vargha-Delaney A-Test (Vargha and Delaney, 2000). A-test scores around the 0.5 mark indicated little to no difference caused by the change in parameter. Scores deviating in both the positive and negative direction from 0.5 indicated increasing differences in the two distributions and higher model sensitivity. The individual output distributions per parameter change was also visualised. Full sensitivity analysis result per parameter is included in appendix 8.8.

From the Vargha-Delaney A-Test scores, the model sensitivity to changes within the -30:30% range was found to be in the small effect region for most parameters, with `nuc_percent` and `gridsize` slightly over, but still well below the medium effect mark.

The parameter `gridsize` was used in the discretisation of the model and defines the area over which the agent considers 'local'. Figure 5.22 shows the distribution of the model response to changes in `gridsize`. The biggest effect is seen in contact outputs. The A-test score for all `gridsize` parameter changes was in the small to no difference range, with the highest difference observed at the -30% parameter value (`gridsize` = 35). This is most likely due to the `gridsize` falling below the cell diameter of some cells. `Gridsize` is limited by cell size and has to be kept larger in order to ensure reliable model output. Interestingly, changes in `gridsize` do not appear to affect cell number despite the relationship between local density and mitotic rate of the cell.

`Nuc_percent` is used to scale up the nucleus size (measured experimentally) to an estimate of cell size. Changes in this parameter affect average contact per cell and average simultaneous contact (figure 5.23). They do not affect average duration of contact or cell number. This is reasonable, as the larger the cell size, the higher the number of contact points, therefore an increase in frequency of contact.

Other notable results were the effect of changes in both cancer cell and stem cell disp_shift parameter. Displacement shift was implemented in order to calibrate the directional change model to the observed experimental data. Increases in displacement shift result in a direction change that is more uniformly distributed across all possible angles. Smaller displacement shift, result in a higher chance of small changes in cell direction. The results of the sensitivity analysis were found to also reflect this, with changes in displacement shift affecting the average duration of contact measure.

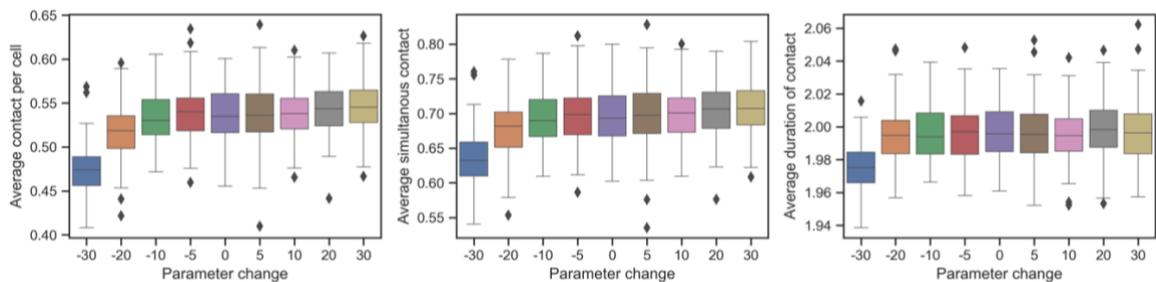


Figure 5.20: Changes in the distribution of contact related outputs with changes in input parameter, gridsize. It was found that reducing the gridsize by more than 20% causes a relatively large change in all contact related output measures. However, the A-test score was found to be less than the medium effect mark.

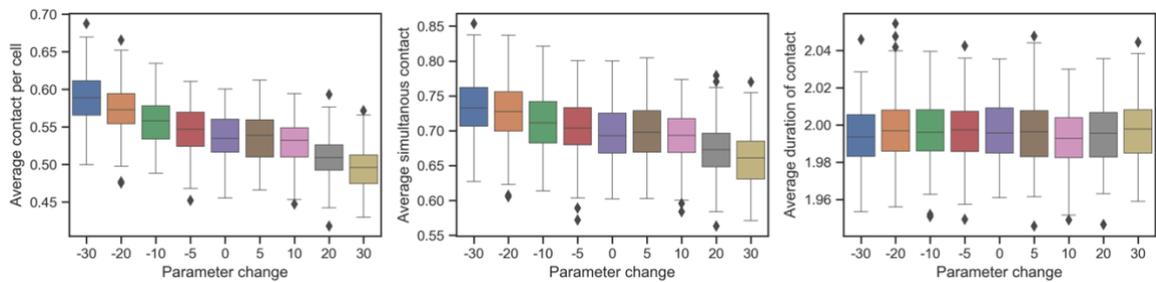


Figure 5.21: Changes in the distribution of contact related outputs with changes in input parameter, nuc_percent. It was found that increasing the percentage of cell volume occupied by the nucleus, reduces the frequency of cell to cell contact and number of simultaneous contacts per cell, however it has very little effect on duration of contact. The effect size of the largest changes in parameter measured by A-test score are still less than the medium effect mark.

5.2.6 Global sensitivity analysis

OFAT investigates the effect of changes in each parameter individually and does not account for relationships between parameters and the compound effects that might occur. Sobol's method (Sobol, 2001) was used as a variance based global sensitivity analysis. 20% uncertainty was introduced into 31 parameters of interest and an input file listing the upper and lower bounds of each parameter was created. Using the SALib python library this input

file was used to generate a list of parameter value combinations using the Saltelli sampling algorithm (Saltelli *et al.*, 2017). As 31 parameters were used and the sample size established in section 5.2.2 was 150, the model was run for a total of 9600 individual input parameter combinations (using the formula: $n(2d + 2)$ with n as the sample size and d as the number of input parameters). The model outputs of interest for each of the input combinations were then used to generate the sensitivity indices using the analyse tool of SALib and visualised appropriately, full results, including the upper and lower boundaries of the input parameters, are included in appendix 8.9.

The three sensitivity indices considered are defined as; first order index (S_1), a measure of a single input parameter's contribution to the model output variance, second order index (S_2), a measure of the interaction of the input paired with other inputs contributing to the model output variance and total order index (S_T), measure of a model parameter's first order and all higher order interactions contributing to the model output variance.

Shown in figure 5.24 is the S_1 and S_T indices across all six output measures and the difference between them. On initial observation, the sc/cc MaxStageMu parameters were found to largely contribute to model uncertainty. These two parameters are the mean values of the normal distribution the maximum cell cycle duration of either cell is randomly drawn from on model initialisation. These results suggests the need for further detailed experimental analysis of cell cycle duration. For each parameter, the difference between S_T and S_1 reflects that parameter's total contribution to all interactions that in turn contribute to the output variance. Both contact output measures related to frequency of contact (average contact per cell and average number of simultaneous contacts) are impacted the most by interactions between the parameters. For second order interactions visual representation of each pairwise combination of parameters was analysed, the results are included in appendix 8.9.

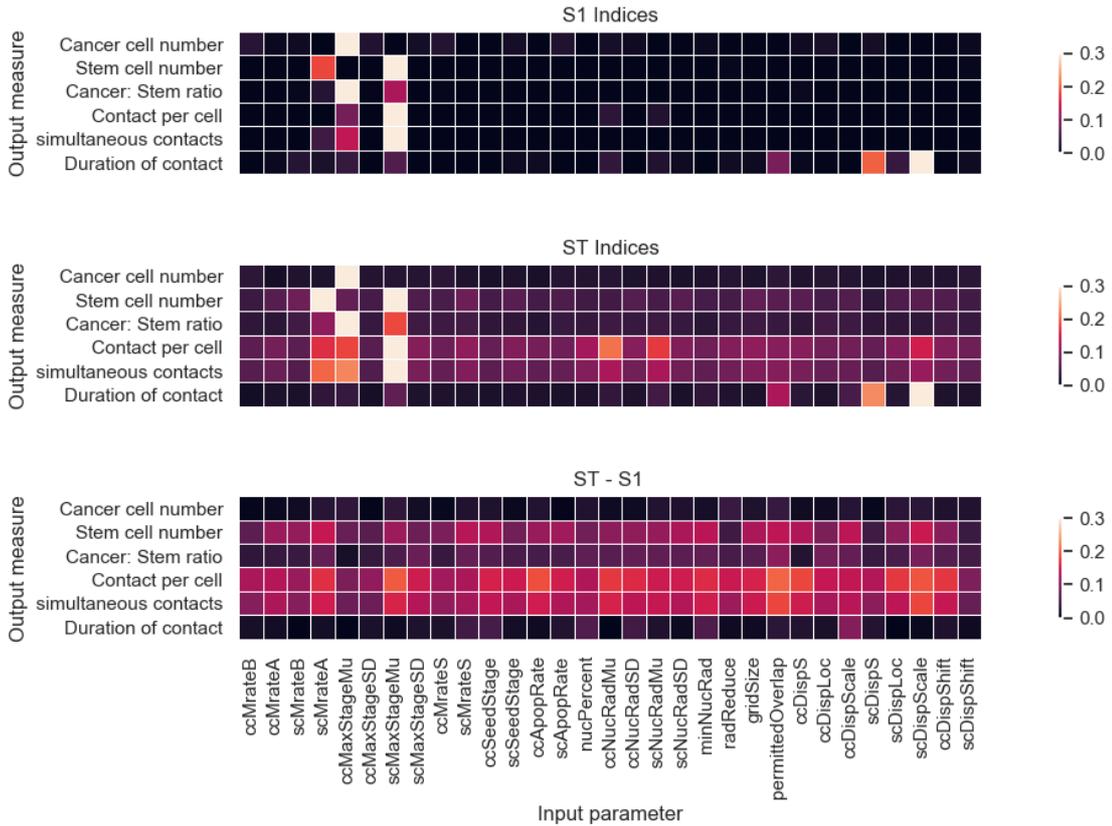


Figure 5.22: The first order (S_1) and total order (S_T) indices for all 31 model parameters for each of the six key model output measures. Also included is the difference between S_T and S_1 which indicates the amount of total interaction each parameter contributes to.

5.3 Summary

In this chapter the agent-based ‘null’ model was presented. The purpose of this model was to reliability capture the behaviour of the cells observed and presented in chapter 4. Object oriented code was used to design and implement the model in Python with each cell type assigned to its own class. An environment class was used to model the physical cell culture environment in addition to acting as virtual storage of agent information.

ABM agent to agent communication was successfully optimised by implementing a grid based search mechanism and the use of hash tables in the form of python dictionaries, to store and locate individual agents. Model optimisation improved the performance of the model by 98.5%. The uncertainty of the model output due to sample size (simulation repeats and/or agent numbers) was explored. A minimum of 25 simulation repeats were found to be sufficient to mitigate the effect of stochastic uncertainty for most of the model outputs.

Local sensitivity analysis was carried out for key parameters that could not be verified experimentally (or were implemented as model abstractions). The model was found to have low to medium sensitivity to changes in these parameters within a $\pm 30\%$ range. All 31 model parameters were subjected to 20% uncertainty the resulting effects of which were explored through global sensitivity analysis for 1st, 2nd and total order indices. On a whole the model was found robust to changes in input parameter with the largest indices falling below 0.3 (normalised). The model outputs related to frequency of contact were found to be effected the most by most parameter interactions. The model was found to be the most sensitive to changes in the mean and standard deviation parameters for cell cycle duration.

The null ABM was calibrated against experimental data with RMSE and ANCOVA analysis used as goodness of fit measures. Further validation was done using three additional experimental data sets collected independently of the calibration set. Kolmogorov-Smirnov test and Cohen “ δ ” scores were used as effect size measures so assess the validity of the model against each unique data set. The model was found valid for both stem and cancer cell growth at seeding densities of 10,000 cells /cm² or lower, up to the four day post seeding timepoint. In the next chapter this validated ABM, coupled with *in vitro* co-culture experimental data, is used to preliminarily explore cell to cell contact

Chapter 6

Co-culture analysis

Co-culture analysis, conducted both *in vitro* and *in silico*, was used to gain an insight into possible cell to cell interactions between stem and cancer cells affecting cancer cell growth. Experimentally, the effect of direct and indirect cell to cell signalling was explored and computationally, the effect of passive contact inhibition was explored.

6.1 Co-culture *in vitro* experiments

In order to investigate if stem cells have an effect on cancer cell proliferation, cell culture experiments of hES-MP and MG-63 grown in a shared environment was carried out.

6.1.1 Methods

An important aspect of co-culture design is selecting an environment suitable for the growth of both cells. As the hES-MP cells require more specialised conditions for growth, the selected shared growth environment used for both cells was that of the hES-MPs (listed in section 3.2.1.3). To assess if the mechanism of the interaction between cells is controlled via direct cell signalling (juxtacrine) or indirect (paracrine), the cells were seeded in co-culture environments designed with and without direct contact permitted through the instalment of a hydrophobic barrier using polydimethylsiloxane (PDMS), a hydrophobic silicone elastomer that is bio-compatible (detailed methodology shared in section 3.2.2.3).

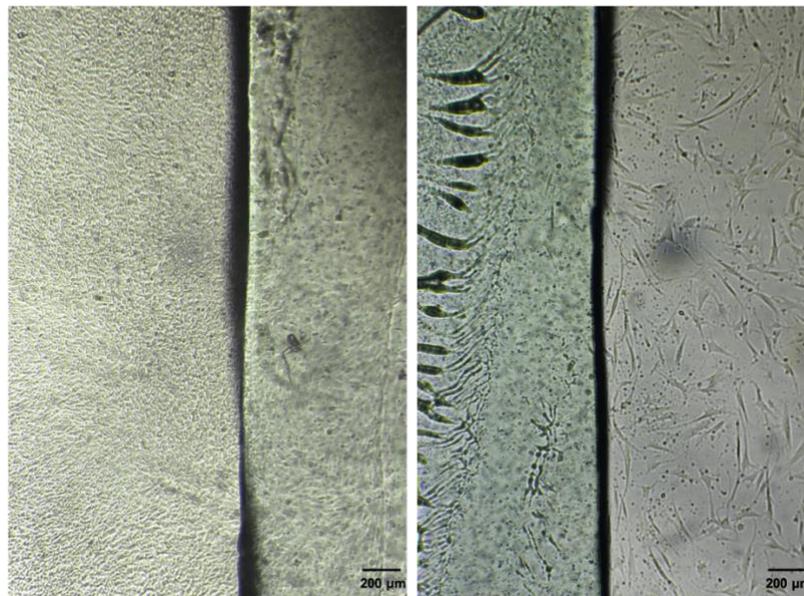


Figure 6.1: light microscope images (at 4x magnification) of the PDMS barrier placed at the centre of the plate with MG-63 cells seeded to the left and hES-MP cells on the right

Four experimental environments were considered, direct co-culture, indirect co-culture, hES-MP and MG-63 mono-cultures. Seeding density was kept consistent across all experimental conditions to account for inevitable differences in culture plate surface area, which was also taken into consideration when applying media or reagents such as staining solution and trypsin. All quantities used were calculated per cm² surface area (Table 6.1).

Table 6.1: Table listing the reagents and volume per culture environment. All measurements are calculated per surface area.. Also shown are the final cell suspension volumes per timepoint post cell detachment. *values stated are per barrier side

| | Surface Area (cm ²) | Hoechst staining solution (uL) | Trypsin (uL) | Neutralising solution (uL) | Final cell suspension including dilution (μL) | | |
|-----------------------------|---------------------------------|--------------------------------|--------------|----------------------------|---|-------|-------|
| | | | | | Days 1-3 | Day 4 | Day 5 |
| 48 well plate | 0.64 | 80 | 32 | 32 | 64 | 128 | 192 |
| 24 well plate with barrier* | 0.80 | 100 | 40 | 40 | 80 | 160 | 240 |
| 24 well plate | 1.91 | 250 | 95 | 95 | 190 | 380 | 570 |
| 12 well plate with barrier* | 1.68 | 200 | 85 | 85 | 170 | 340 | 510 |
| 12 well plate | 3.80 | 500 | 190 | 190 | 380 | 760 | 1140 |

Trypsin-EDTA solution was used for cell detachment and neutralised with HBSS containing 20% FBS. This was used in replacement of media containing phenol red to improve fluorescent imaging and reduce background noise. The higher percentage of FBS allowed for the use of smaller volumes of neutralising solution and in turn a smaller final cell suspension volume, increasing the accuracy of the cell counts at low cell densities.

Prior to seeding, MG-63 cells were labelled with CellTracker as per protocol listed in section 3.2.3.2. During preliminary experiments, hES-MP cells were also labelled separately with a contrasting CellTracker, however, this step was eliminated in the final experiment design. This both increased the simplicity of the experiment and viability of hES-MP cells which were observed to be highly sensitive to changes in environment.

During mono-culture analysis of cell growth (results presented in section 4.2), hES-MP cells were found to have a longer lag phase. Based on this, in co-culture, a 24 h delay was placed between seeding hES-MP and adding MG-63 cells. The timepoint at which MG-63 cells were added is referenced to at day 0 of the experiment. For direct co-culture, homogenous cell suspensions at the desired concentration were added to each well and the well plate gently moved in a figure of 8 motion whilst kept horizontal before placing in the cell culture incubator. For indirect co-culture, cells were gently pipetted in a drop wise manner across the entire half well surface. Care was taken to not move the plate as it was

kept under the cell culture hood to attach for 15 minutes before transfer to the incubator. This was done to ensure homogeneous seeding which originally proved a challenge due to the shape of the well after barrier instalment.

Six hours after the addition of MG-63 cells, additional media was added to all conditions. This was done to bridge over the barrier and allow mixing of the media across both sides of the well plate. Full media change was carried out on days 2 and 4 and manual cell counts were obtained every 24hours. Cells were loaded with Hoechst nuclear stain prior to detachment at each timepoint. This was done to allow for automated cell counting using a pipeline designed in CellProfiler. For verification of results, cells were imaged before detachment. Full methodology for manual cell counting presented in section 3.2.4.2.

The experiment was carried out for four different initial seeding configurations of MG-63 to hES-MPs, starting at an initial ratio of 1:1 (50% MG-63 cells) and gradually increasing the number of hES-MP cells added per MG-63 cell in ratios 1:2, 1:3 and 1:4 (33%, 25% and 20% MG-63 cells respectively). The overall cell density in all seeding ratio configurations was kept constant (table 6.2). The full experiment was replicated a total of 5 times, with 4 well repeats per time-point (n=20).

Table 6.2: Seeding density of MG-63 and hES-MP cells at each ratio configuration and environment condition

| Ratio MG-63:hES-MP (Percentage of MG-63 cells) | Mono-culture (x1000 cells/cm ²) | | Direct co-culture (x1000 cells/cm ²) | | Indirect co-culture (x1000 cells/cm ²) | |
|--|---|--------|--|--------|--|--------|
| | MG-63 | hES-MP | MG-63 | hES-MP | MG-63 | hES-MP |
| | 1:1 (50%) | 12 | 12 | 6 | 6 | 12 |
| 1:2 (33%) | 8 | 16 | 4 | 8 | 8 | 16 |
| 1:3 (25%) | 6 | 18 | 3 | 9 | 6 | 18 |
| 1:4 (20%) | 4.8 | 19.2 | 2.4 | 9.6 | 4.8 | 19.2 |

6.1.2 Results

6.1.2.1 Microscopic co-culture images

Included below is a series of examples of fluorescent images used predominately in preliminary co-culture experiments.

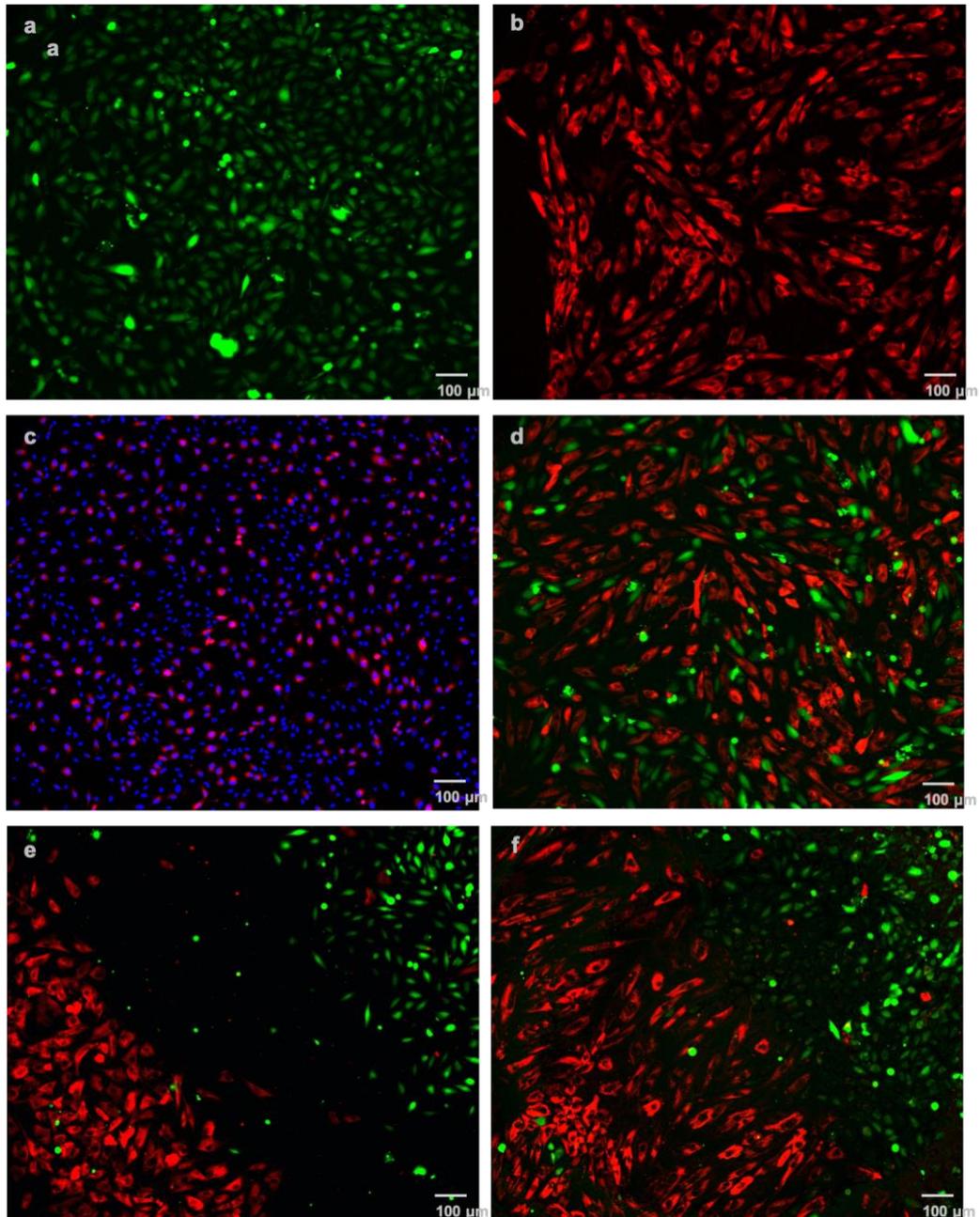


Figure 6.2: Examples of fluorescent images of hES-MP and MG-63 cells in co-culture during preliminary experimental work. (a) MG-63 cells in indirect co-culture stained with CellTracker green. (b) hES-MP cells in indirect co-culture stained with CellTracker red. (c) Direct co-culture of MG-63 cells labelled with CellTracker red and hES-MP cells unlabelled. Both cells nuclei are stained with Hoechst. This was the final staining methodology selected. (d) Direct co-culture of hES-MP (red) and MG-63 (green). (e and f) Direct co-culture of Mg-63 (green) seeded in an island formation before addition of hES-MP (red) 4 and 48 hours after seeding respectively.

6.1.2.2 Verification of day 1 ratios

To verify the actual percentage of MG-63 cells seeded, at time 0, 4 sample wells of hES-MP cells per condition were counted before adding the MG-63 cells. Cell numbers were then subsequently collected for both cells every 24 h. Figure 6.3 shows the results of time 0 to 48 h, 2 days post seeding the MG-63 cells. It was found that the MG-63 cell numbers on day 1 of the experiment were very close to the cell numbers seeded, whereas for the hES-MP cells at high seeding densities or seeded in direct co-culture environment they were higher than the numbers seeded. This resulted in discrepancies in the actual seeding density and initial percentage of MG-63 cells seeded per experiment which had to be taken into consideration when interpreting the results. In order to do this, all data was normalised to day 1 results.

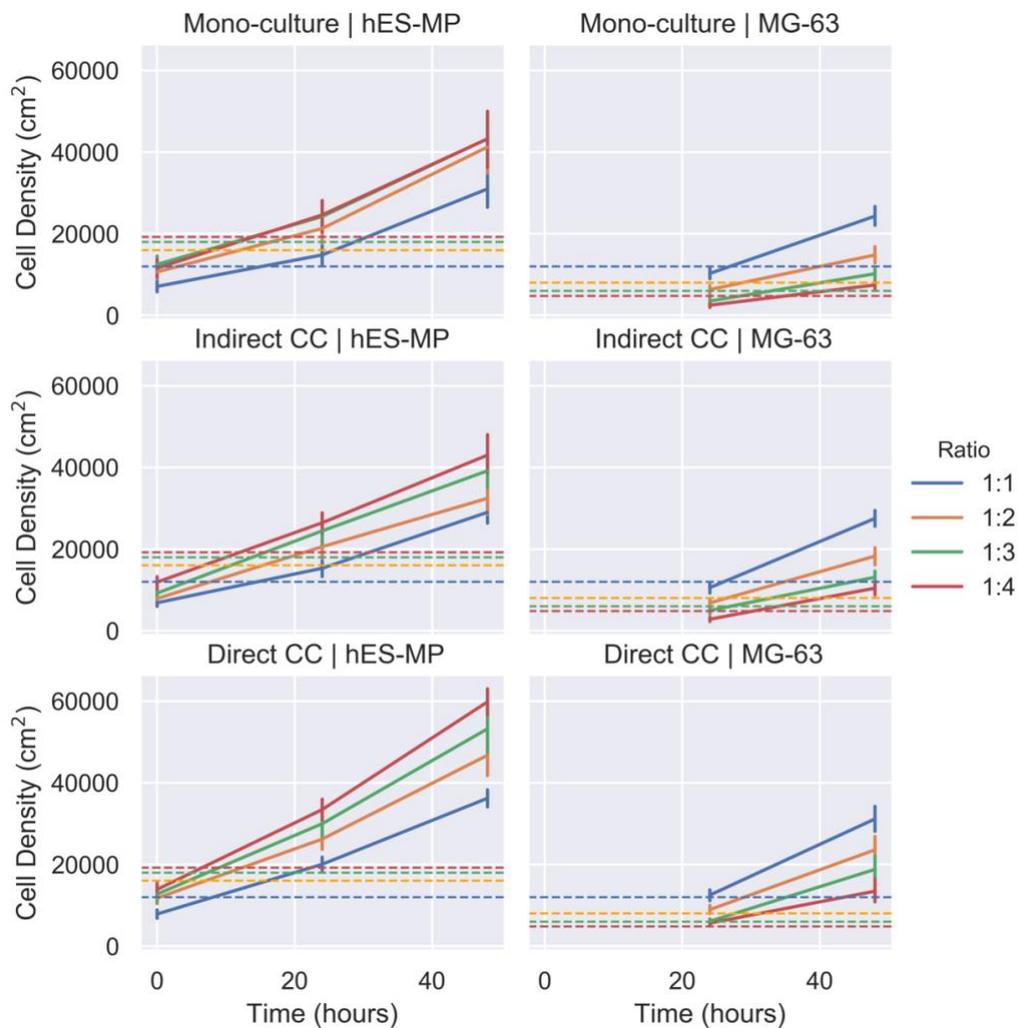


Figure 6.3: Results showing initial cell numbers at time 0 (hES-MP only, 24hours and 48hours post addition of MG-63 cells). Expected cell numbers are drawn with dotted lines per seeding ratio and solid lines represent actual cell numbers. MG-63 cell numbers on day 1 are close to the expected value whereas hES-MP cells were found to be much higher at high initial seeding densities (e.g. ratio 1:4 with MG-63 cells taking up only 20% of the total cells seeded) and/or when seeded in direct contact environment.

6.1.2.3 Comparison of growth

Cell numbers obtained were first normalised to well plate surface area and then normalised to day 1 results, therefore plots presented in figure 6.4-6.6 all start at the 100 mark and trends represent changes in initial seeding ratio (figure 6.4) or cell number (figures 6.5-6.6) as a percentage of day 1 data. It was found that the initial seeding ratio was not maintained in any of the conditions and instead the ratio of MG-63 to hES-MP increased. At low seed ratio conditions (e.g. 1 MG-63 cell to 4 hES-MP cells), this increase was much larger. This is very likely to be due to the differences in rate of growth at different initial seeding densities (as found in mono-culture analysis presented in section 4.2). It is also verified by the fact that this trend in change in ratio in indirect co-culture and mono-culture is very similar.

In direct co-culture however, the increase in ratio of MG-63 to hES-MP is slower than its counterpart experiments at initial seeding conditions with lower percentages of MG-63 cells (e.g. ratios 1:3 (25%) and 1:4 (20%)). Cross comparing with the cell number results, this slower increase in ratio (or total percentage of MG-63 cells) seems to be caused predominately by a slower rate of growth in MG-63 cells as the change in cell numbers for hES-MP was found to be significantly the same across all environmental conditions.

Statistical analysis was carried out on day 4 results using 2-way ANOVA followed by pairwise Tukey-HSD post-hoc tests. The results are visualised as bar charts presented in figures 6.7-6.9. At seeding densities of 1:1 and 1:2, where the number of initial MG-63 cells seeded is the highest (50% and 33% respectively), the rate of growth of the hES-MP cells were found to show the largest differences across environmental conditions. This suggests that MG-63 cells negatively affect hES-MP growth. As hES-MP numbers in both indirect and direct co-culture were found to be lower than in mono-culture, the mechanism for this could be a combination of paracrine and juxtacrine signalling.

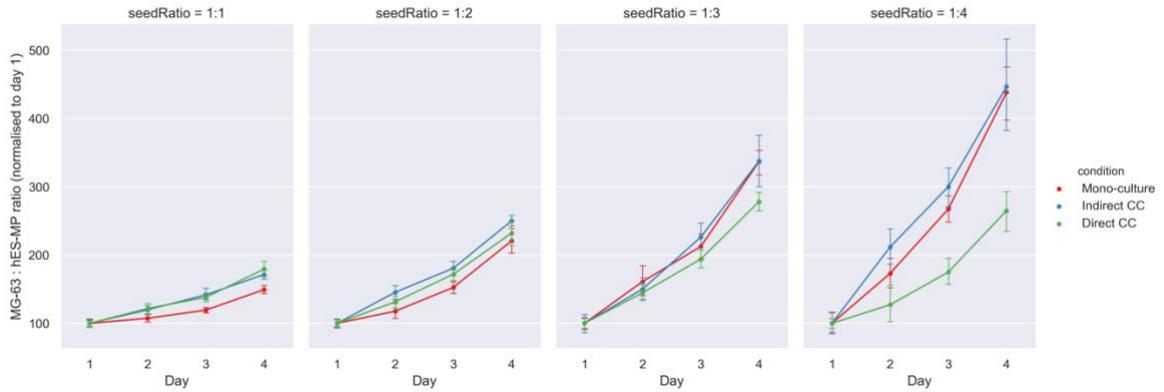


Figure 6.4: The change in ratio of MG-63 cells to hES-MPs across 4 different initial seeding ratios. Seeding ratio here means the percentage of MG-63 cells in the total number of cells seeded (therefore, seeding ratio of 1:1, 1:2, 1:3 and 1:4 or 50%, 33% 25% and 20% MG-63 cells). The data represented on the y-axis is normalised to day 1 data. The ratio of MG-63 to hES-MP cells increases across all seeding conditions, with the highest increase observed at low initial MG-63 cell numbers. Within the low MG-63 to hES-MP seeding conditions it was also found that the increase in ratio occurs at a lower rate in direct co-culture in comparison to indirect or mono-culture.

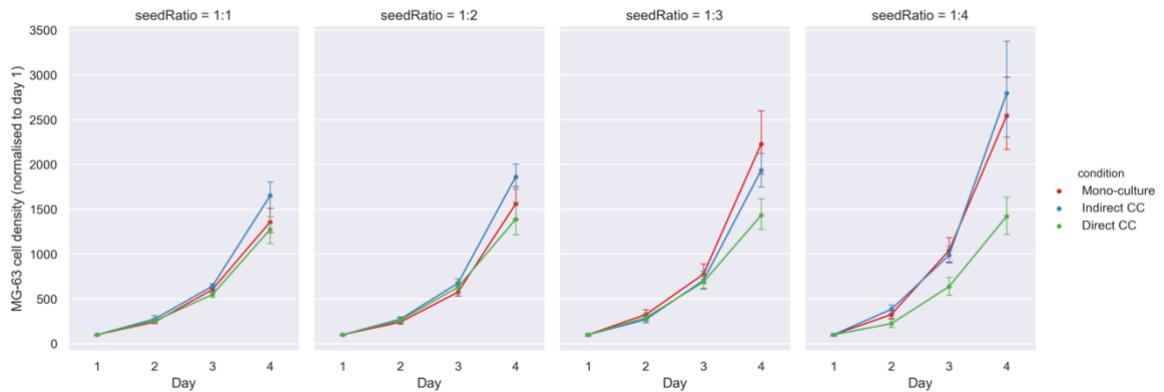


Figure 6.5: Cell numbers of MG-63 normalised to day 1. The behaviour of MG-63 cells was found to be similar in indirect co-culture and mono-culture. This was also matched with direct co-culture at high MG-63 to hES-MP seeding ratios (1:1 and 1:2 or 50% and 33% MG-63s). At lower ratios, the MG-63 cells grew at a slow rate than their counterpart cells restricted to indirect co-culture or mono-culture. This suggests direct co-culture with hES-MP cells has a negative effect at high ratios of hES-MP to MG-63 cells.

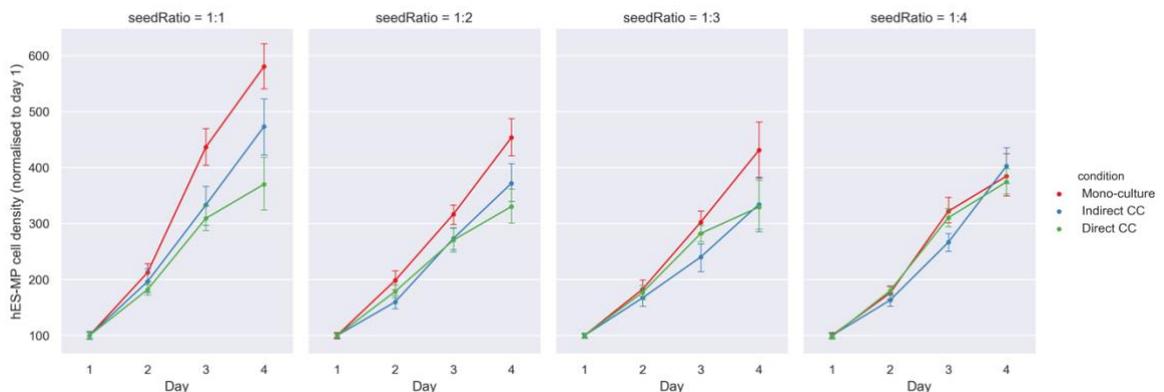


Figure 6.6: Cell numbers of hES-MP cells normalised to day 1. The change in cell number of hES-MP in co-culture was found to be slower than in mono-culture, specially at experiments with higher initial number of MG-63 cells (lower hES-MP to MG-63 ratios). At high ratios of hES-MP to MG-63 cells, co-culture effects appeared to be minimised.

6.1.2.4 Statistical analysis of day 4 results

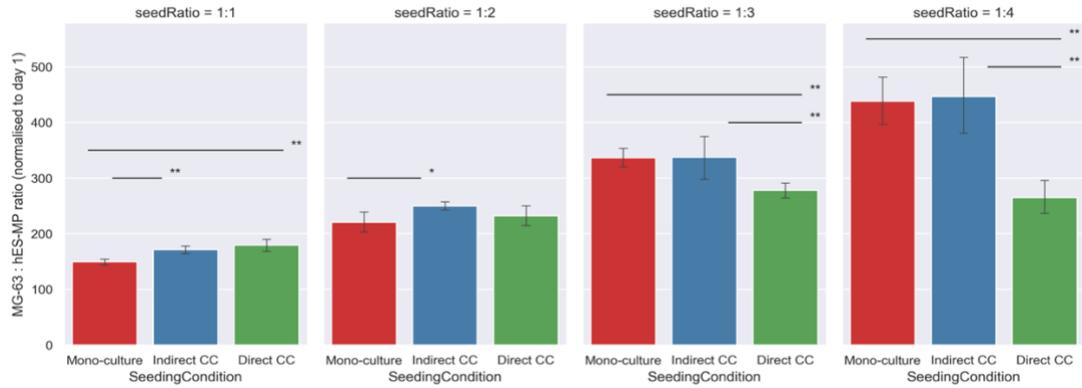


Figure 6.7: Bar charts showing the difference in MG-63 to hES-MP ratio on day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. The relationship between indirect co-culture and mono-culture found at seed ratios of 1:1 and 1:2 were found to be linked to hES-MP behaviour (figure 6.9) and the significant difference observed in direct co-culture at initial seed ratio of 1:3 and 1:4 was found to be a result of changes in MG-63 growth (figure 6.8).

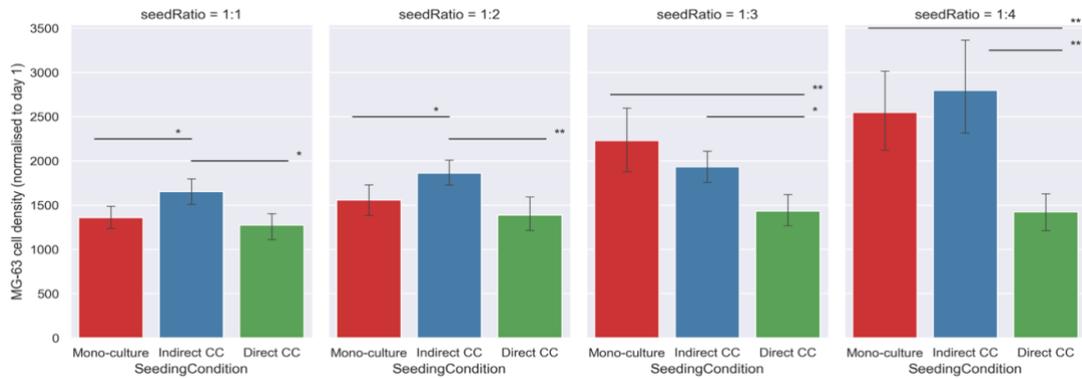


Figure 6.8: Bar charts showing the difference in MG-63 growth by day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. At low MG-63 to hES-MP ratios (1:3 and 1:4) cells in direct co-culture were found to grow significantly slower than their counterparts in mono-culture and indirect co-culture. No significant difference was found between mono-culture and indirect co-culture at these ratios, however at higher ratios of MG-63 to hES-MP cells (1:1 and 1:2), significant differences were found.

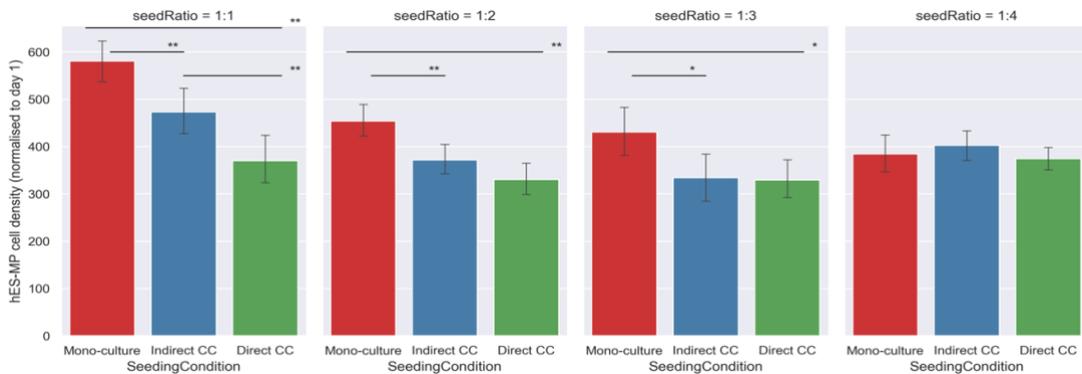


Figure 6.9: Bar charts showing the difference in hES-MP growth by day 4. Results of statistical analysis are presented with * indicating $p < 0.05$ and ** $p < 0.001$. The rate of growth of hES-MPs was found to be significantly slower in co-culture than mono-culture under initial seeding ratios of 1:1, 1:2 and 1:3. No difference was observed at ratio 1:4. Significant difference between direct and indirect co-culture was only found at ratio 1:1.

6.2 Co-culture *in silico*

The experimental seeding conditions used for the *in vitro* co-culture presented above (section 6.1) were applied to the ‘null’ ABM presented in chapter 5. As no contact rules (paracrine or juxtacrine) were implemented at this stage, the indirect co-culture experiments were not included here (i.e. without paracrine rules, these simulation conditions would be exactly the same as the mono-culture simulations). The aim here was to gain an insight into how much if any, the passive contact inhibition or competition between the cells due to sharing an environment, could be responsible for the trends observed *in vitro* (presented in the previous section 6.1.2.4).

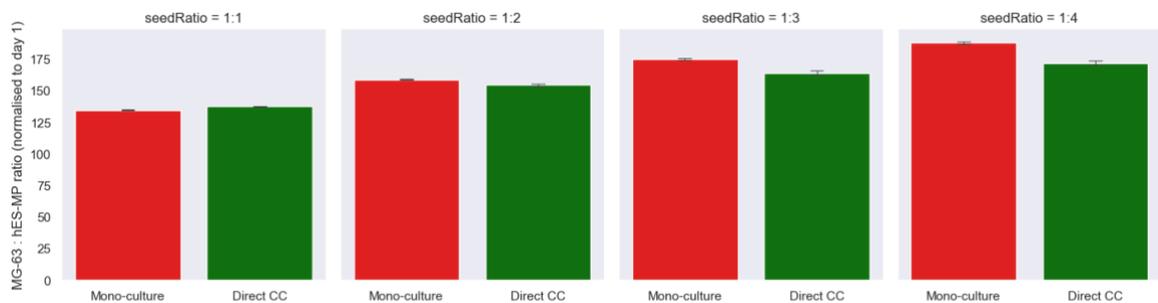


Figure 6.10: The simulation results of no-contact ABM (presented in chapter 5), run under the same seeding conditions as *in vitro* co-culture experiments (results presented in figure 6.7). The results presented here are the day4 simulation results normalised to day 1 data in order to allow for direct comparison with the *in vitro* results.

The *in silico* simulations (figure 6.10) and the *in vitro* (figure 6.7) results share an interesting similarity in the overall trends shown. In both the change in percentage of MG-63 cells (or ratio of MG-63 to hES-MP cells) seeded with respect to day 1 increases slower in mono-culture in the experimental conditions where the initial percentage of MG-63 cells seeded was higher (seed ratio of 1:1 and 1:2 at 50% and 33% MG-63 cells). Contrastingly in both the *in silico* and *in vitro* results, the opposite is seen when the number of hES-MP cells to MG-63 cells seeded is increased (seed ratio of 1:3 and 1:4 at 25% and 20% MG-63 cells). In these experiments the change in percentage of MG-63 cells increases slower in direct co-culture. This suggests that there could be some form of passive contact inhibition resulting from the motile hES-MP cells coming into contact with the MG-63 cells more frequently as a result of these seeding conditions.

6.3 Summary

In this chapter a protocol for co-culture analysis using fluorescent staining and imaging coupled with a semi-automated counting pipeline was presented for both direct and indirect co-culture. hES-MP and MG-63 cells were grown in co-culture over 5 days under 4 different initial seeding conditions of varying ratio of MG-63 to hES-MP cells. These ratios were 1:1 (50% MG-63 cells), 1:2 (33% MG-63 cells), 1:3 (25% MG-63 cells) and 1:4 (20% MG-63 cells). It was observed that the ratio of MG-63 to hES-MP cells increased across all seeding conditions, with the highest increase observed at low initial MG-63 cell numbers (25% and 20%). Within the low MG-63 to hES-MP seeding conditions it was also found that the increase in ratio occurs at a lower rate in direct co-culture in comparison to indirect or mono-culture.

Comparison of raw cell numbers showed that at high initial percentages of MG-63 cells, hES-MP cell growth is slower in co-culture (most significantly in direct co-culture) in comparison to mono-culture. Increasing the seeding ratio of hES-MP to MG-63 cells lowers the difference between hES-MP growth in mono-culture and co-culture. At the highest seeding ratio of hES-MP to MG-63s no significant difference was seen in hES-MP cell numbers between mono-culture and both co-culture (direct and indirect) experiments. Contrastingly in the same experiment, the number of MG-63s increased slower in direct co-culture in comparison to indirect and mono-culture conditions.

Co-culture analysis was also done using the ABM presented in chapter 5. Simulations were done using the same initial seeding configurations as the *in vitro* experiments for mono-culture and direct co-culture conditions. Comparison of the no contact ABM simulation results with the lab based co-culture results showed similarities in trend. The change in percentage of MG-63 cells in direct co-culture increased at a slower rate in comparison to mono-culture in conditions with higher initial seeding ratios of hES-MP to MG-63 cells.

The ABM can be further expanded to test various hypothesised contact rules using *in vitro* co-culture results to validate and compare model emergent behaviour and the lab based observations.

Chapter 7

Discussion and conclusions

The work presented in this thesis was designed with the main aim of testing whether stem cells alone can affect the growth of cancer cells in co-culture, with the added research question of whether stem cell and cancer cell interactions are governed by juxtacrine or paracrine signalling. In order to achieve this, a combination of *in vitro* and *in silico* modelling approaches were used. Below, the six main objectives listed at the start of this thesis are reviewed.

7.1 Single cell analysis

Objective 1 - Qualitative and quantitative exploration of the biological domain via in vitro single cell culture analysis of cancer and stem cell behaviour.

Qualitative exploration of the biological domain was found to be a good starting point and was used as a guide in the modelling process. However, the main focus was to obtain quantitative data that was directly translated into model parameters. This proved a challenge due to the large amounts of data that was required to draw statistically valid conclusions which was both time consuming to collect and to post process. To address this problem, there are many automated tracking tools available (Meijering, Dzyubachyk and Smal, 2012), however with automation some level of accuracy can be lost especially when decisions have to be made regarding unique scenarios. These scenarios can sometimes be key descriptors of the model domain. For example, in collection of cell shape and size data, the automated cell detection software, was found to be less accurate at high cell density. The data regarding the changes in cell morphology at high confluency was an area of interest for the ABM as it directly affects the maximum cell density whilst maintaining a mono-layer formation. Similarly, for cell motility study, at unique scenarios where the cell migration per time-step was large, the software was found to be less accurate at correctly connecting cell trajectories. This was especially problematic for stem cell tracking and the solution employed was to use semi-automated pipelines (as opposed to reducing the time-step which would increase the data size). Therefore, a balance has to be established between accuracy, data size, resolution of data detail and limitations of automation software. The cell tracking data was collected from more than 450 cells tracked across multiple generations over 2-3days. This was considerably higher than other similar research, for example (Wadkin *et al.*, 2017, 2018).

Other limitations of single cell analysis, is the technology available to take measurements at the cell (or smaller) scale. It is important that the methodology used to obtain data of single cell behaviour does not interfere with the cell's true behaviour which can be unavoidable with methods of isolation, culture and imaging. For example, on analysis of time-lapse imaging data, it was found that the cells show a gradual decline in motility as time increases. Although there was not enough data to draw a conclusion, this might have been due to phototoxicity as a result of prolonged exposure to camera flashes (Magidson and Khodjakov, 2013) or repeated excitation of the Hoechst33342 nuclear stain (Purschke *et al.*, 2010).

Included in the challenges of collecting quantitative data, is the establishment of a relationship that can be quantitatively described. During the process of collecting single cell data to inform the ABM, in most situations, it was found that single parameter measurements were not sufficient to reliably describe biological behaviour. For example, in modelling mitotic rate, a linear regression model was used and for cell velocity, lognormal and von Mises distributions.

The lognormal models presented for the distributions of cell motility displacements for both hES-MP cells and MG-63 were found to be significantly valid through both Mann-Whitney and Wilcoxon rank sum tests. Lognormal has also been used to successfully model T-cell movement by Fricke *et al.*, (Fricke *et al.*, 2016).

The relationship between cell movement and turning angle, the directional change made per time-step, was found to resemble that of a von Mises distribution with concentration parameter, κ , defined as the cell displacement. This model was not statistically validated, however at face validity, it resembled both the MG-63 distribution of turning angles and hES-MPs. As both cells have distinctly different motility patterns, this can be taken as a twofold validation of the suitability of this model. Von Mises distribution has been referenced for use in modelling angular distributions (Cantarella *et al.*, 2009), for example in random walk models of reef fish larvae (Codling *et al.*, 2004), however a reference to the adaption of von Mises to describe relationship between displacement and turning angle has not been found in the literature.

The final relationship between turning and displacement was defined by applying a shift factor to the cell displacement, before randomly drawing a turning angle from the resulting distribution. One reason for the need for the displacement shift could be due to experimental error in detecting cell centre during the data acquisition cell tracking stage. With 1 image pixel = 0.645 μM at the magnification the images were taken at, small errors in manual or automated cell detection would affect the resulting displacement. In addition to experimental error, the displacement shift can also be interpreted as the cell's speed without directional persistence.

The average cell nucleus size was found to get smaller as cell density increased for both cells (Li *et al.*, 2010; Phadnis *et al.*, 2015). hES-MP cells were found to be more elliptical in shape than the MG-63 cells with the cells aligned along their major axis (Wadkin *et al.*, 2018). The cell direction with respect to the horizontal was found to be randomly distributed for MG-63 cells, however for hES-MP cells, there appeared to be a

synchronisation of directionality within the cells seeded in the same plate. For future work, it would be interesting to measure the angles for cells seeded in the same seeding condition across a larger number of separate plating samples. Also, in order to assess the interference of the time-lapse protocol with cell behaviour over extended periods of time, the results of later stages of time-lapse, should be contrasted with cells seeded at the same time, but kept in a cell culture incubator as opposed to under time-lapse.

Death rate was calculated by visually identifying and recording instances of cell death captured on time-lapse. The death mechanism observed suggested death due to apoptosis, however this was not verified using apoptotic markers. For future work, a more in depth exploration of death rate should be carried out using appropriate markers and normal cell culture conditions (not subjected to continuous time-lapse).

Also for future work, cell synchronisation assays can be used to gain more detailed picture of the number of cells actively in the cell cycle, the length of various stages of the cycle and overall cycle duration. This data could be compared with the mitotic rate model and cell cycle duration values established via cell tracking and calibration in order to validate their accuracy.

7.2 Null model

Objective 2: Construction of a robust agent-based model (ABM) of the in vitro cell culture environment explored in objective 1. – the ‘null’ model

Objective 3 - Verification of the ‘null’ ABM simulations response using in vitro mono-culture experimental data.

Object oriented code was used to design and implement an ABM of individual stem and cancer cell growth and movement. Agent-based modelling was selected due to its bottom-up discrete nature, allowing for one to one mapping of cells onto agents. Model design was based on existing literature, direct qualitative analysis of hES-MP (stem) and MG-63 (cancer) cell behaviour *in vitro*. The aim here was to create a computational model that represented the experimental conditions, and specific behaviours of the two cell lines when grown in mono-culture. Therefore individual agent behaviour and characteristics were based on statistical models of single cell behaviour collected *in vitro* via time-lapse imaging and cell tracking.

This implementation of an ABM is not typical. Commonly with agent-based modelling, mechanistic theories are applied at the agent level and their subsequent effect on the emergent behaviour is used to gain an understanding of the system. Using statistical models to inform the ABM does not give the same insight into the system. However, as the aim of this ‘null’ ABM was to act as a baseline model for further expansion into contact modelling, this was not an issue.

For future work however, it would be interesting to implement some mechanistic thinking and compare the emergent behaviour with that of the statistically fit ABM. For example, the relationship between cell displacement and turning angle presented in the single cell study was fit to a von Mises distribution, suggesting that the cells with larger displacements showed a persistence in forward motion. This theory could be tested mechanistically using an ABM.

With all ABM, the agent rules must be at the single cell level. Therefore statistical models (which by nature are population reflecting) must be used discretely. To do this, the models were used as probability density functions with agents able to draw dynamic (or static) parameter values using conditional statements (if – else). One example of this is the implementation of the mitotic rate linear regression model. Here the agents use local information to reverse the linear regression model in order to obtain a mean and standard deviation value for mitotic rate. This is used as a normal distribution to draw a random value which is taken as the cells mitotic rate. Applying the statistical models in this way, allowed for each agent to make independent decisions and captured the stochastic uncertainty of the experimental data used to inform the model.

One of the disadvantages of ABM is that it is limited by computational power. Increases in agent numbers also increase the simulation time. One reason for this is that in this type of modelling agents are able to communicate with each other and this can be computationally taxing to implement. Here the ABM agent to agent communication was optimised by implementing a grid based search mechanism and the use of hash tables to store and locate individual agents. Model optimisation improved the performance of the model by 98.5%.

Sensitivity and uncertainty analysis is a very important aspect of developing a reliable model. There are many ways to interpret and carry out sensitivity and uncertainty analysis. First it is important to differentiate between the different types of uncertainty: aleatory and epistemic. Inherent stochasticity in biological and computational systems introduces aleatory uncertainty. This can be mitigated by increasing the number of

replications. On the other hand epistemic uncertainty arises from abstraction or lack of knowledge regarding the biological domain of the model. In order to assess the impact of this uncertainty, the sensitivity of the model to changes in input parameter should be assessed. Generally, this is done by introducing a fixed percentage of uncertainty to each parameter and assessing the effect on the model outputs. By using a fixed uncertainty, the effect on the model output can be compared between input parameters. However, this approach does not factor in the different ranges of uncertainty each parameter might have.

The uncertainty of the model output due to sample size (simulation repeats and/or agent numbers) was explored. A minimum of 25 simulation repeats were found to be sufficient to mitigate the effect of stochastic uncertainty for most of the model outputs. Local sensitivity analysis was carried out for key parameters that could not be verified experimentally (or were implemented as model abstractions). The model was found to have low to medium sensitivity to changes in these parameters within a $\pm 30\%$ range. All 31 model parameters were subjected to 20% uncertainty the resulting effects of which were explored through global sensitivity analysis for 1st, 2nd and total order indices. On a whole the model was found robust to changes in input parameter with the largest indices falling below 0.3 (normalised). The model outputs related to frequency of contact were found to be affected the most by most parameter interactions. The model was found to be the most sensitive to changes in the mean and standard deviation parameters for cell cycle duration.

The null ABM was calibrated against experimental data with RMSE and ANCOVA analysis used as goodness of fit measures. This was perhaps one of the most time consuming and computationally taxing stages of model development. In future, statistical sampling methods such as Latin hypercube sampling should be used in order to reduce the number of parameter combinations.

The model was then further validated using three additional experimental data sets collected independently of the calibration set. Kolmogorov-Smirnov test and Cohen “ δ ” scores were used as effect size measures so assess the validity of the model against each unique data set. The model was found valid for both stem and cancer cell growth at seeding densities of 10,000 cells/cm² or lower up to the four day post seeding timepoint. The model failed validation tests for higher seeding densities for both cells.

Taking into consideration the results of the global sensitivity analysis, the poor performance against experimental data of the model at high densities, could be a result of inaccurate implementation of cell cycle duration in the ABM. For hES-MP cells during experimental cell tracking, data regarding cell cycle duration could not be gathered, as the

cell's did not appear to visually split. For MG-63 cells, this data was relatively small (147 datapoints). Also the protocol used to assess cell cycle duration, did not account for cell's in the G0 phase of the cycle, it assumes that the cell remains in the active phase of cell cycle from one mitotic event to the next. In future work, in depth analysis of cell cycle behaviour is suggested, using cell synchronisation assays (Uzbekov, 2004).

7.3 Co-culture analysis

Objective 4 - Design of in vitro co-culture experiments to differentiate between juxtacrine and paracrine signalling.

Objective 5 - in silico exploration of cell to cell interaction in co-culture between cancer and stem cells within the agent-based modelling environment. – the 'contact' model

Objective 6 - Validation of the 'contact' ABM simulation responses using in vitro co-culture experimental data

Through preliminary mono-culture experiments, *in vitro* co-culture experiments were designed. Adapted protocols for manual cell counting, staining and cell seeding were developed through trial and error.

From the many challenges of co-culture design, establishing the correct time-line for cell seeding proved to be the most difficult. In both mono-culture and co-cultures hES-MP cells were found to have a higher variability in behaviour. This resulted in inaccuracies in expected cell numbers from seeding and actual cell's attached. A solution to this was to normalise all data to results obtained on day 1. The experience gained from *in vitro* experimental design, was that simplified protocols should be favoured as much as possible, introducing extra steps in order to 'perfect' protocols can introduce unforeseen errors. For example in co-culture, initially two stains were used to label each cell respectively. However this was found unnecessary and the added steps involved in staining the highly sensitive hES-MP cells was affecting their viability and attachment and also prolonging the cell seeding process which ideally should be kept short.

The co-culture protocol developed is an alternative to more expensive methods such as using FACS and can be carried out on smaller sample sizes. The use of a PDMS barrier to create an indirect co-culture environment also allowed for easy imaging of the cells using

an inverted microscope which would otherwise not be possible (e.g. if typical methods such as transwell inserts were used). However both manual cell counting and plate preparation made this protocol very time consuming and more susceptible to human errors and inconsistencies. It is important to reduce the time taken to measure each sample as the cumulative effect of long sample processing times can induce variability and inconsistency between samples measured at the start of the experiment and the end. For example, if all samples were seeded at time = t and the cell counting began at time = $t + 24$ h but took 4 hours to complete, then the samples detached towards the end of the session would have been under culture conditions for 4 hours longer than those detached at the start. Alternatively detaching all samples prior to counting could result in samples counted towards the end of the session being in a worse condition (e.g. clumping or losing viability) than those counted at the start.

Software aided cell detection was used in an attempt to lower sample processing time. In addition, a high throughput counting chamber, inspired by the design of a haemocytometer in the form of a microfluidic chamber was built. The current prototype of this counting chamber can be loaded successfully with multiple samples at a time, making it faster than using a haemocytometer, however loading and cleaning between sample sets proved inefficient and as a result failed to reduce the overall experimental time. More work is required to improve the design of this device. Current similar devices available are the counting chambers provided with automated cell counting machines such as precision cell counters. These counting chambers can only take two samples at a time and are disposable, meaning that for a protocol such as the one presented in this thesis, they would not be cost effective.

Another disadvantage of this protocol was the staining method, which limited the longevity of the experiments due to fading over time and exposure to light. Since the stains used were not cell line specific, they could only be loaded prior to seeding in co-culture. For future work, the use of cell specific labelling and staining is recommended.

The results of co-culture analysis were found to support the main hypothesis of the PhD; It was found that hES-MP cells lower MG-63 rate of growth in direct co-culture. This effect is increased when the percentage of hES-MP cells seeded is increased to 75% (ratio of 3:1) and 80% (ratio of 4:1). This relationship between higher doses of stem to cancer cells was also reported by (Ohlsson *et al.*, 2003), however the majority of papers that take into account stem to cancer ratios found increases in ratio of stem cells to cancer cells to have

tumour promoting properties. (Zhu *et al.*, 2006; Karnoub *et al.*, 2007; Fabian L. Muehlberg *et al.*, 2009; Kucerova *et al.*, 2010; Prantl *et al.*, 2010).

Also worth noting is the inhibiting effect of MG-63 cells on the rate of growth of hES-MP cells observed. This effect is seen more clearly in experiments that seeded higher percentages of MG-63 cells (50%, 33% and 25%).

The co-culture conditions were simulated virtually using the ‘null’ ABM (with no contact rules implemented) in order to explore the possibility of passive contact inhibition resulting in the trends observed experimentally. Interestingly, the model emergent output shared some similarity with the experimental data. The change in percentage of MG-63 cells in direct co-culture increased at a slower rate in comparison to mono-culture in conditions with higher initial seeding ratios of hES-MP to MG-63 cells, suggesting that there could be some form of passive contact inhibition resulting from the motile hES-MP cells coming into contact with the MG-63 cells more frequently.

The prerequisites for objectives 5 and 6 have been completed. For future work, the validated null ABM can be used to test hypothesised rules regarding the mechanism of cell to cell contact. These contact versions of the ABM can then in turn be validated against the four independent co-culture experimental data sets. The role of passive and active contact inhibition is one area that could be explored this way. Also given the experimental findings regarding direct co-culture, the ABM can be used to explore various mechanistic juxtacrine rules.

Chapter 8

Appendix

8.1 Statistical tests and sources

Table 8.1: Table listing all python libraries (or otherwise) utilised in statistical analysis.

| Statistical test | Python library | Specific module & function |
|--------------------------|----------------|---------------------------------|
| Shapiro-Wilk | scipy | stats.shapiro |
| D'Agostino and Pearson's | scipy | stats.normaltest |
| Leven's test | scipy | stats.levene |
| ANOVA | pingouin | anova |
| Tukey-HSD | pingouin | pairwise_tukey |
| Kruskal-Wallis | scipy | stats.kruskal |
| Mann-Whitney | scipy | stats.mannwhitneyu |
| Wilcoxon rank sum | scipy | stast.wilcoxon |
| Linear regression | statsmodels | formula.api.ols |
| Box-cox transformation | scipy | stats.boxcox |
| Anderson-Darling | scipy | stats.anderson |
| Goldfeld-Quandt | statsmodels | stats.api.het_goldfeldquandt |
| Harvey-Collier | statsmodels | stats.api.linear_harvey_collier |
| ANCOVA | pingouin | ancova |
| Q-Q plot | pingouin | qqplot |
| Histogram plot | seaborn | hist |
| Kolmogorov-Smirnov test | scipy | stats.ks_2samp |
| Vargha Delaney A -test | - | |
| Cohen-d | - | |
| Grubbs test | - | |

8.2 Results of cell nuclei size and shape study

Table 8.2: Cell nuclei size and shape results

| Day | Cell Type | Average Eccentricity value | Average Area | Average radius | Minimum radius | Maximum radius |
|------------|------------------|-----------------------------------|---------------------|-----------------------|-----------------------|-----------------------|
| 1 | MG-63 | 0.636 ± 0.105 | 417.7 ± 92.53 | 11.5 ± 1.3 | 6.51 | 17.95 |
| | hES-MP | 0.694 ± 0.105 | 300.0 ± 99.8 | 9.65 ± 1.53 | 3.234 | 18.856 |
| 4 | MG-63 | 0.591 ± 0.591 | 258.9 ± 77.3 | 9.0 ± 1.3 | 6.459 | 19.684 |
| | hES-MP | 0.710 ± 0.118 | 231.1 ± 97.21 | 8.43 ± 1.59 | 3.255 | 19.338 |
| 7 | MG-63 | 0.627 ± 0.143 | 187.8 ± 54.5 | 7.64 ± 1.16 | 4.841 | 9.825 |
| | hES-MP | 0.777 ± 0.113 | 167.2 ± 42.6 | 7.23 ± 0.98 | 3.234 | 9.127 |

8.3 Cell track diagrams

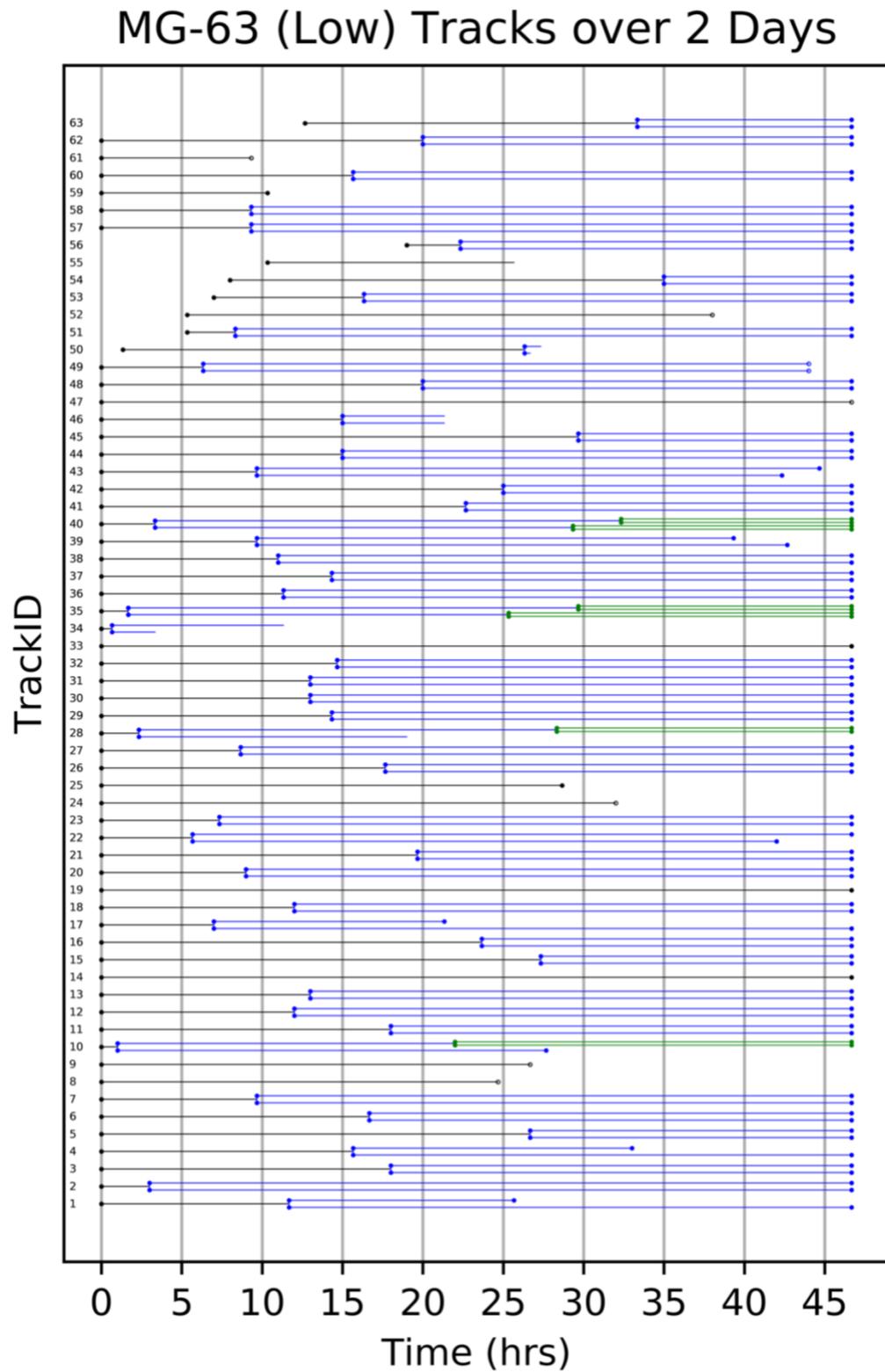


Figure 8.1: MG-63 cells in low seeding density (5, 000 cells/cm²) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by color change. \

MG-63 (Med) Tracks over 2 Days

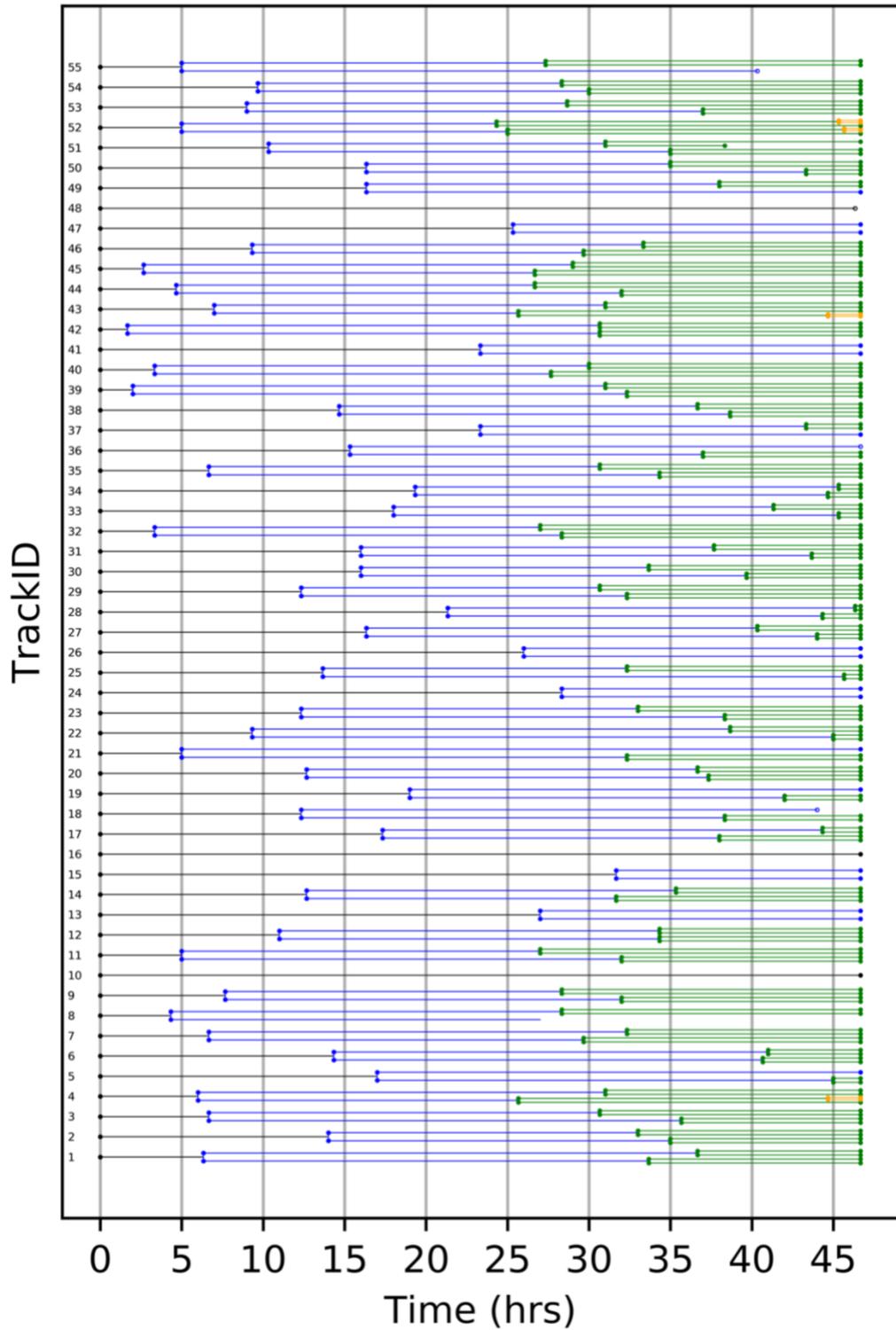


Figure 8.2: MG-63 cells in medium seeding density (10,000 cells/cm²) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by colour change.

MG-63 (High) Tracks over 2 Days

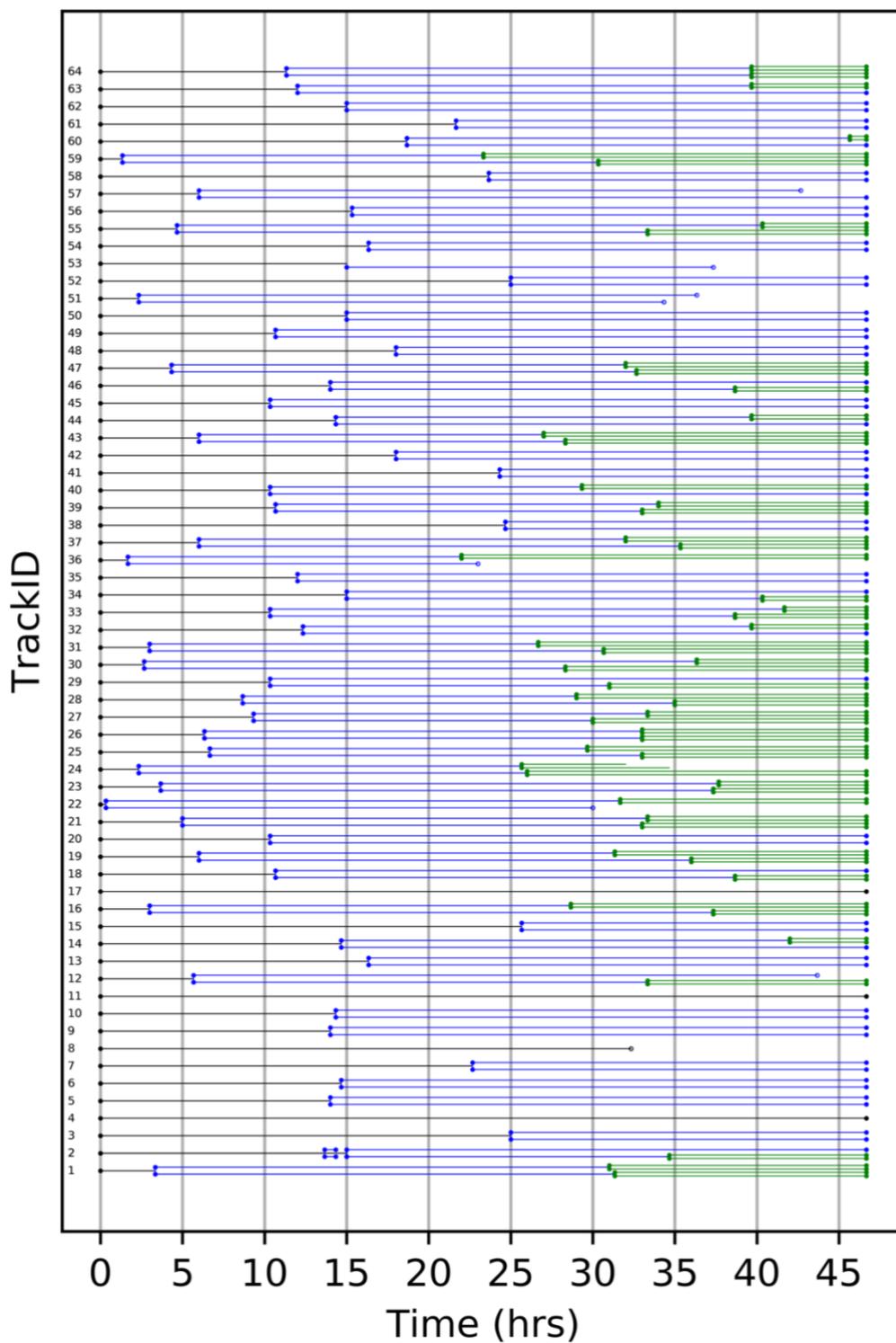


Figure 8.3: MG-63 cells in high seeding density ($20,000 \text{ cells/cm}^2$) tracked over 2 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame. Changes in generation are marked by colour change.

hES-MP (Low) Tracks over 3 Days

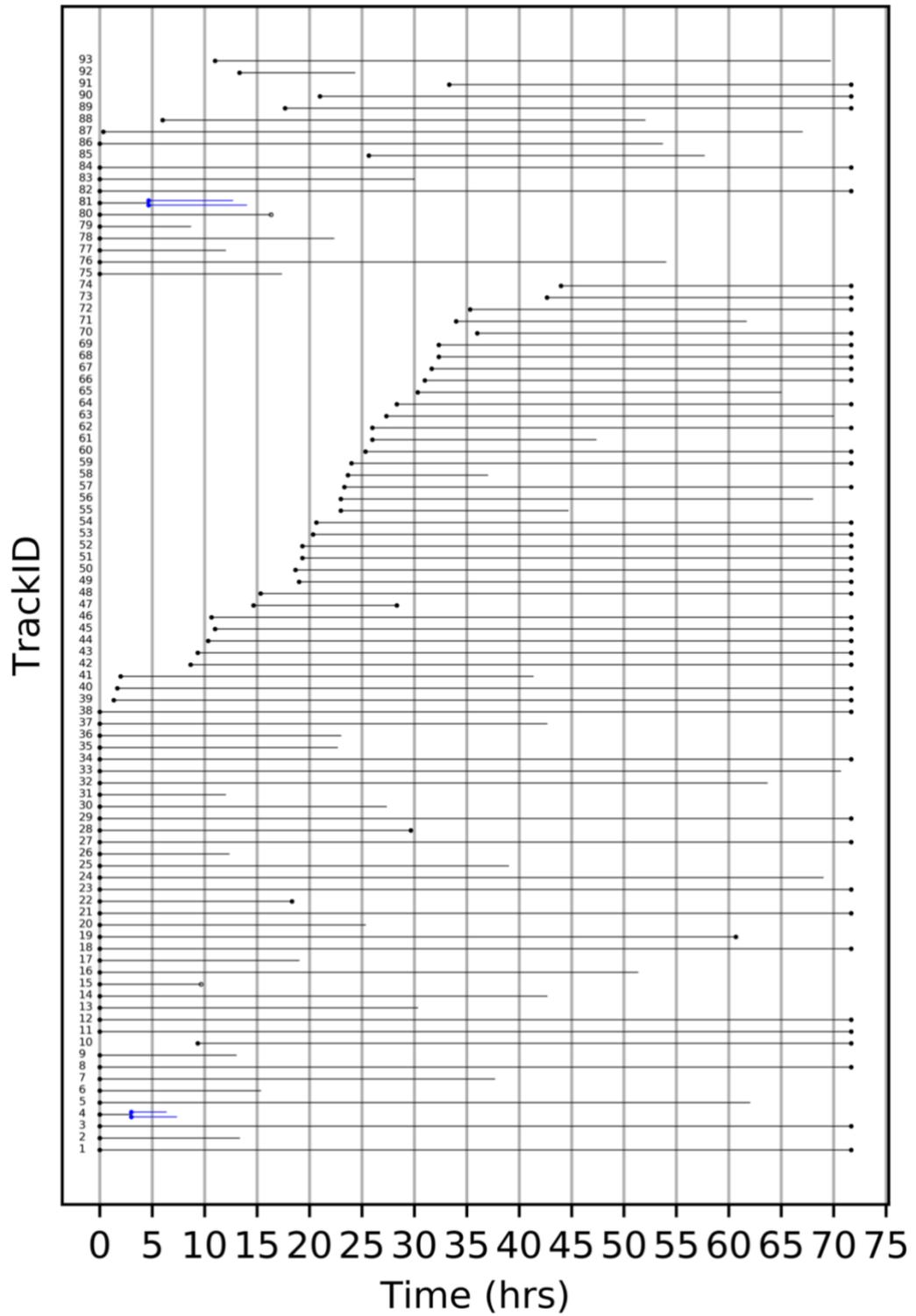


Figure 8.4: hES-MP cells in low seeding density (5,000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

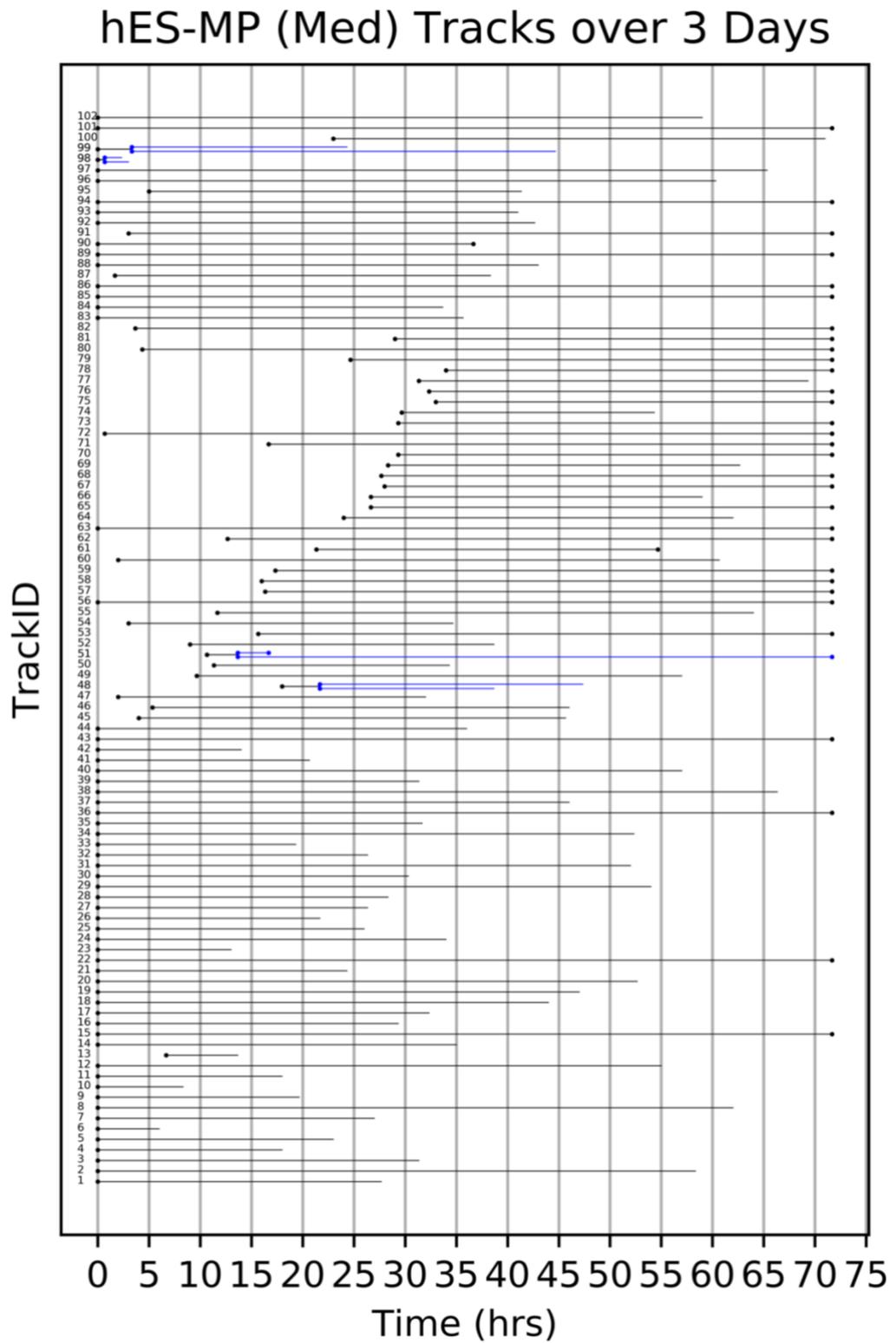


Figure 8.5: hES-MP cells in medium seeding density (10,000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

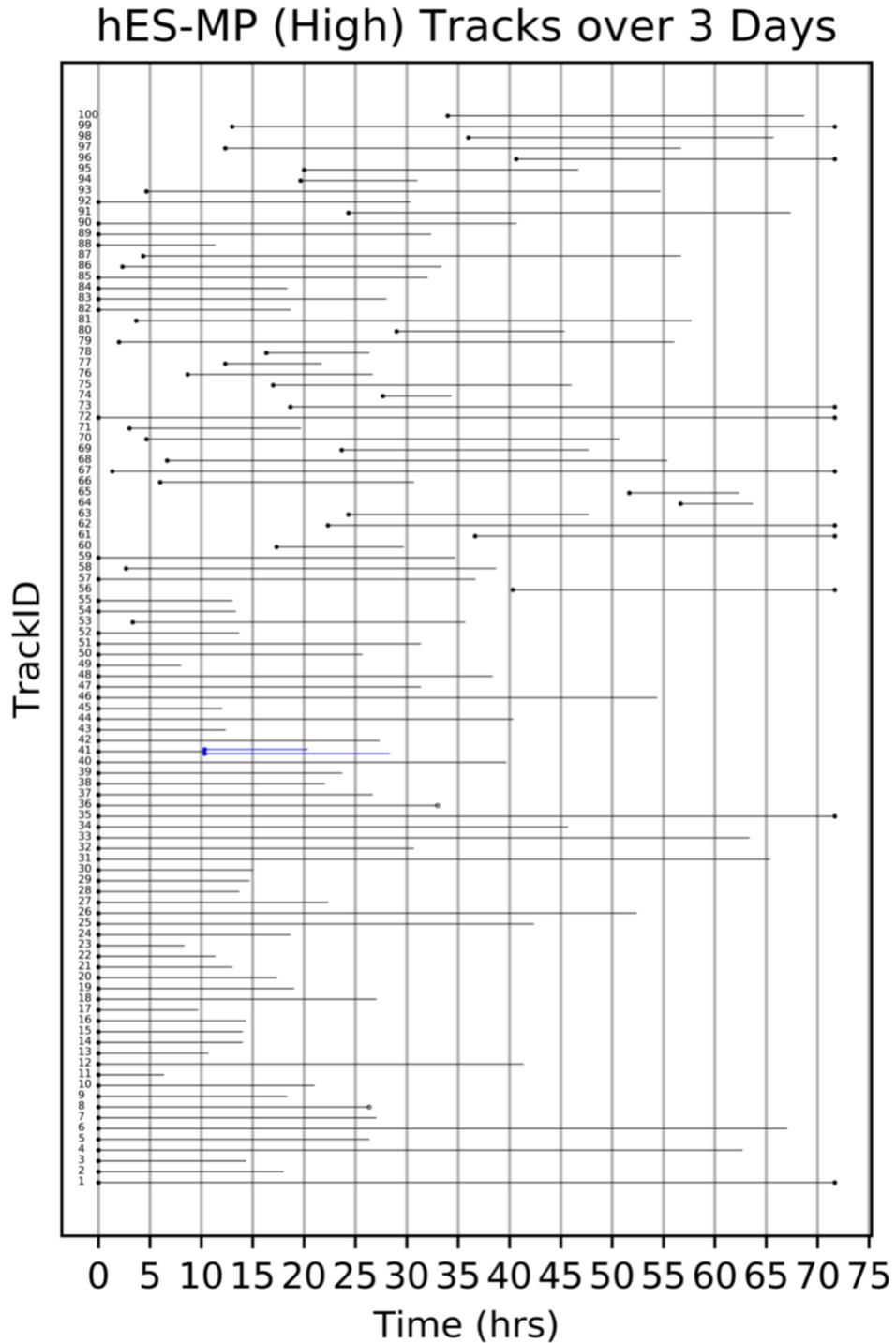


Figure 8.6: hES-MP cells in high seeding density (20,000 cells/cm²) tracked over 3 days. Markers indicate key events, start, mitosis, death, end. Dead cells are marked by an unfilled marker. An open ended track indicates the cell left the frame.

8.4 Linear regression models of cell growth

Table 8.3: Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 2, 500 cells/cm²

Seeding condition = 2, 500 cells/cm²

| Test | hES-MP mitotic rate model | MG-63 mitotic rate model |
|--|--|---|
| Model: $y = \alpha + \beta x$ | $y = 11.08 + 6.77 x$ | $y = 18.59 + 0.54 x$ |
| Shapiro-Wilk normality test for residuals | Test statistic 0.986 p-value 0.483 | Test statistic 0.968 p-value 0.035 |
| Harvey Collier test for linearity of residuals | Test statistic 0.195 p-value 0.846 | Test statistic -0.328 p-value 0.744 |
| Gold-Quandt test for homoscedasticity | Test statistic 0.741 p-value 0.048 | Test statistic 0.828 p-value 0.690 |
| Mean of residuals | $-3.60 \times 10^{-15} \approx 0$ | $-9.56 \times 10^{-15} \approx 0$ |

Table 8.4: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 5, 000 cells/cm²

Seeding condition = 5, 000 cells/cm²

| Test | hES-MP mitotic rate model | MG-63 mitotic rate model |
|--|--|--|
| Model: $y = \alpha + \beta x$ | $y = 10.86 + 19.14 x$ | $y = 18.75 + 14.22 x$ |
| Shapiro-Wilk normality test for residuals | Test statistic 0.988 p-value 0.661 | Test statistic 0.990 p-value 0.800 |
| Harvey Collier test for linearity of residuals | Test statistic -3.139 p-value 0.002 | Test statistic 0.339 p-value 0.735 |
| Gold-Quandt test for homoscedasticity | Test statistic 0.782 p-value 0.724 | Test statistic 0.382 p-value 0.322 |
| Mean of residuals | $-9.35 \times 10^{-15} \approx 0$ | $2.66 \times 10^{-15} \approx 0$ |

Table 8.5: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 10, 000 cells/cm²

Seeding condition = 10, 000 cells/cm²

| Test | hES-MP mitotic rate model | MG-63 mitotic rate model |
|--|---|---|
| Model: $y = \alpha + \beta x$ | $y = 10.39 + 32.41x$ | $y = 18.10 + 32.23 x$ |
| Shapiro-Wilk normality test for residuals | Test statistic 0.983 p-value 0.312 | Test statistic 0.947 p-value 0.002 |
| Harvey Collier test for linearity of residuals | Test statistic -1.043 p-value 0.300 | Test statistic -1.825 p-value 0.072 |
| Gold-Quandt test for homoscedasticity | Test statistic 0.113 p-value 0.071 | Test statistic 0.994 p-value 0.005 |
| Mean of residuals | $-2.08 \times 10^{-14} \approx 0$ | $-3.38 \times 10^{-16} \approx 0$ |

Table 8.6: : Statistical analysis of assumptions for linear regression models of cell growth with initial seeding condition 20, 000 cells/cm²

Seeding condition = 20, 000 cells/cm²

| Test | hES-MP mitotic rate model | MG-63 mitotic rate model |
|--|---|---|
| Model: $y = \alpha + \beta x$ | $y = 7.33 + 50.73 x$ | $y = 17.34 + 49.15 x$ |
| Shapiro-Wilk normality test for residuals | Test statistic 0.989 p-value 0.692 | Test statistic 0.972 p-value 0.060 |
| Harvey Collier test for linearity of residuals | Test statistic -1.230 p-value 0.222 | Test statistic -0.433 p-value 0.666 |
| Gold-Quandt test for homoscedasticity | Test statistic 0.435 p-value 0.094 | Test statistic 0.828 p-value 0.429 |
| Mean of residuals | $-1.13 \times 10^{-14} \approx 0$ | $-3.32 \times 10^{-14} \approx 0$ |

8.5 Aleatory uncertainty analysis results

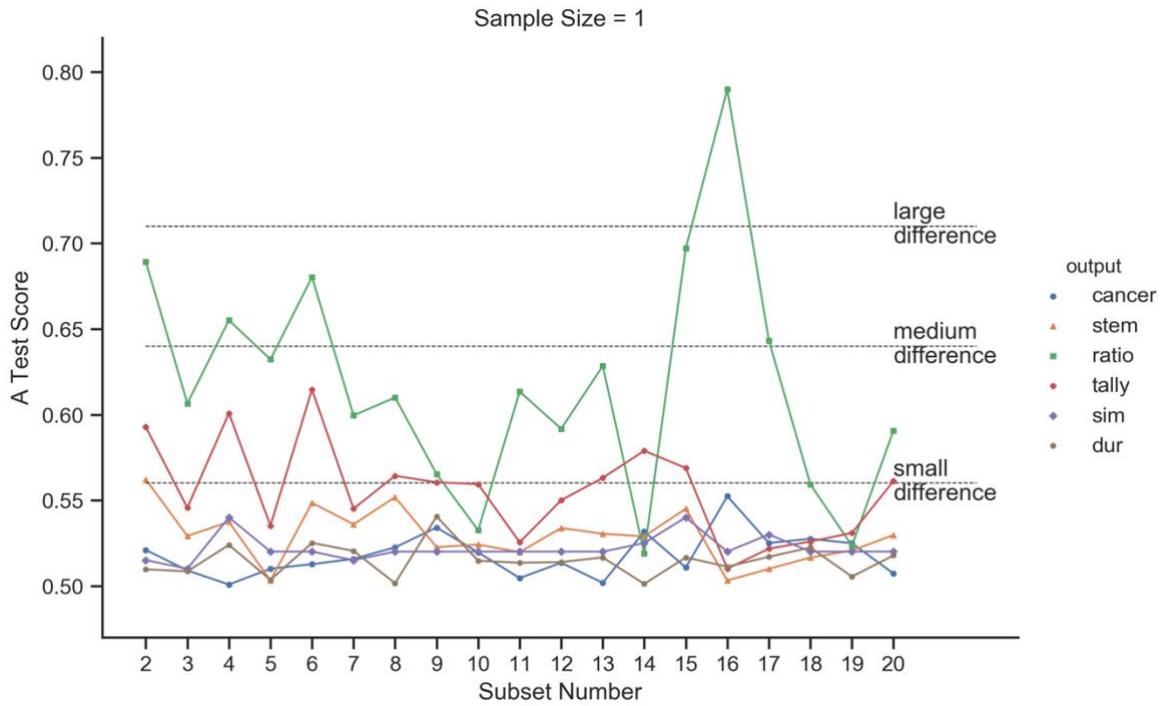


Figure 8.7: A-test score of all sets per sample size of 1 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

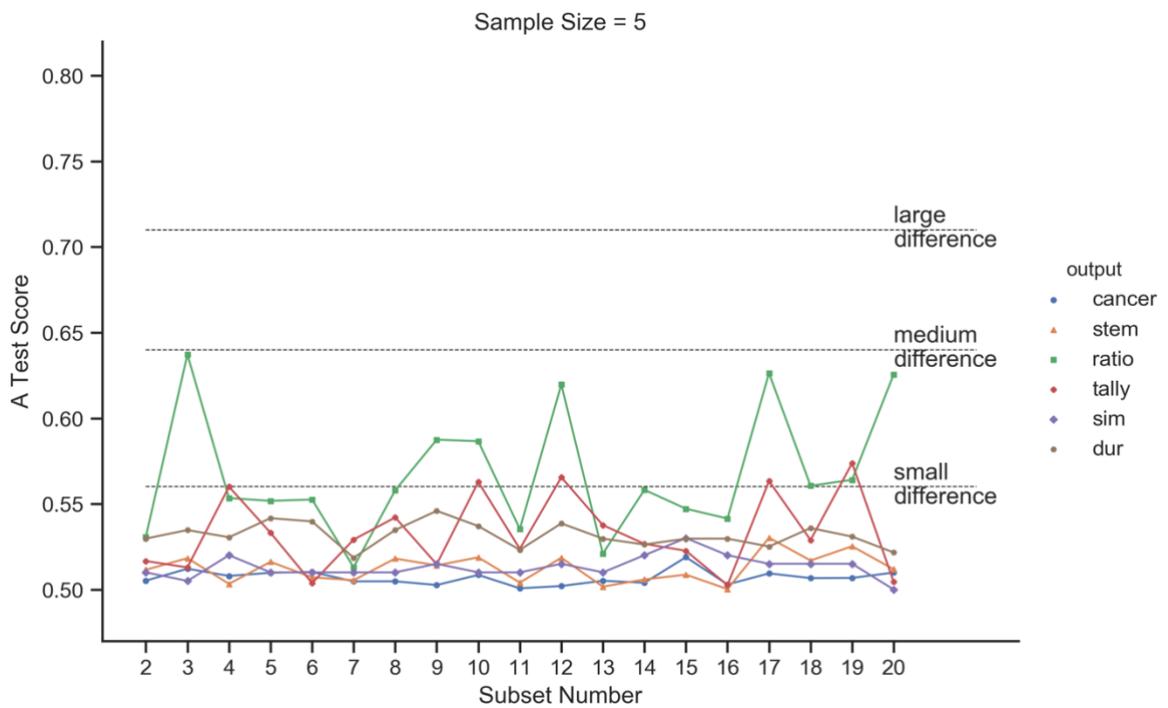


Figure 8.8: A-test score of all sets per sample size of 5 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines

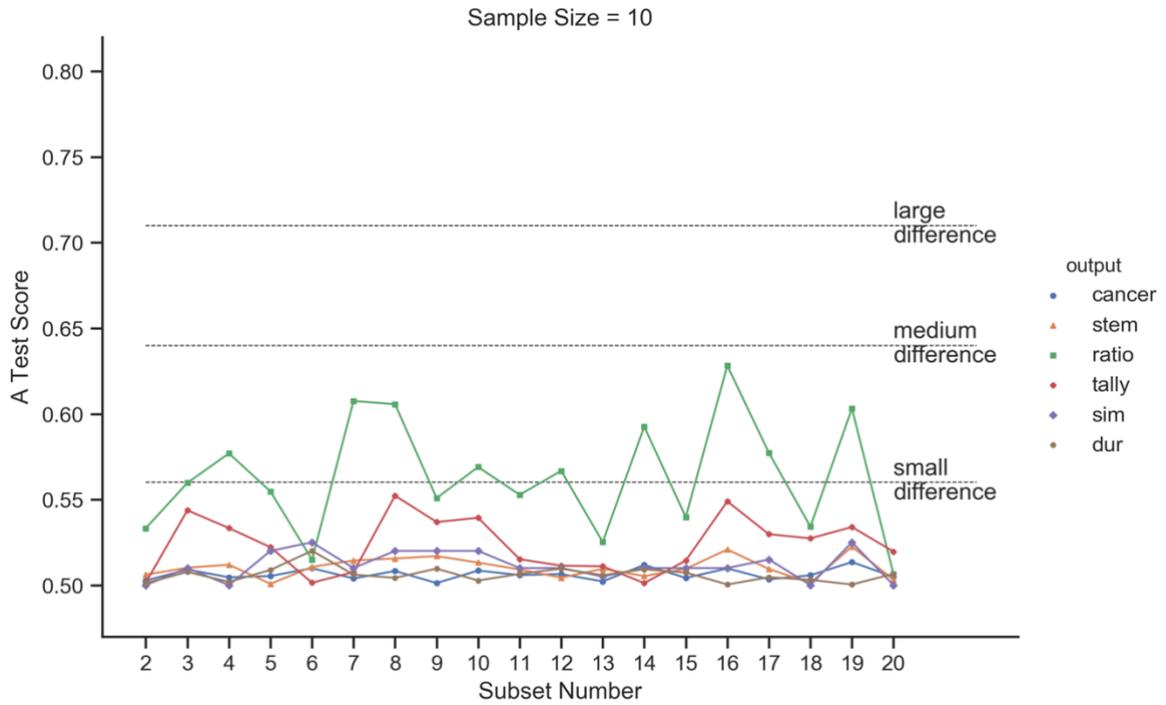


Figure 8.9: A-test score of all sets per sample size of 10 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

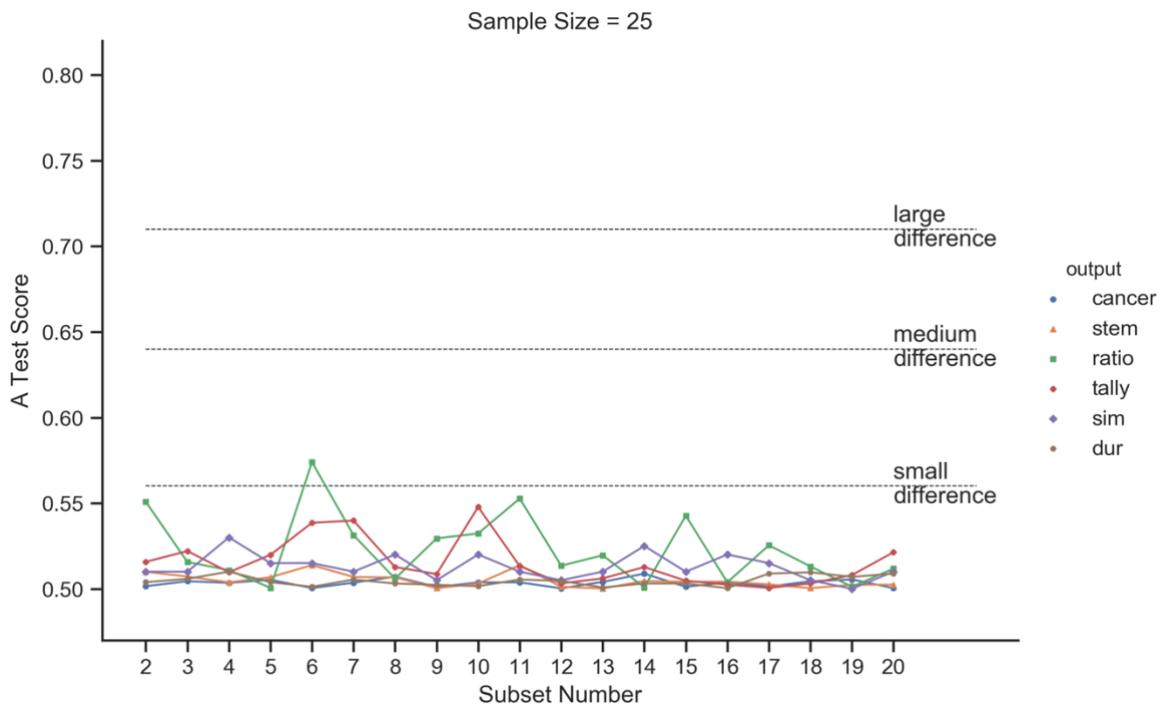


Figure 8.10: A-test score of all sets per sample size of 25 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

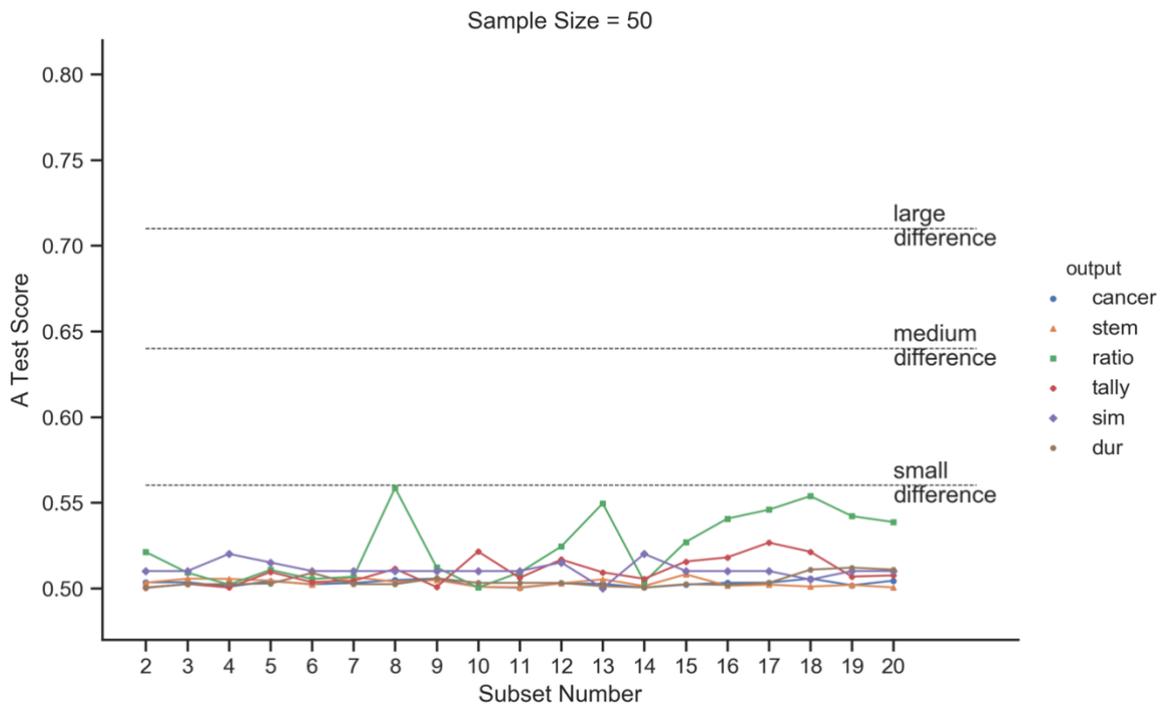


Figure 8.11: A-test score of all sets per sample size of 50 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

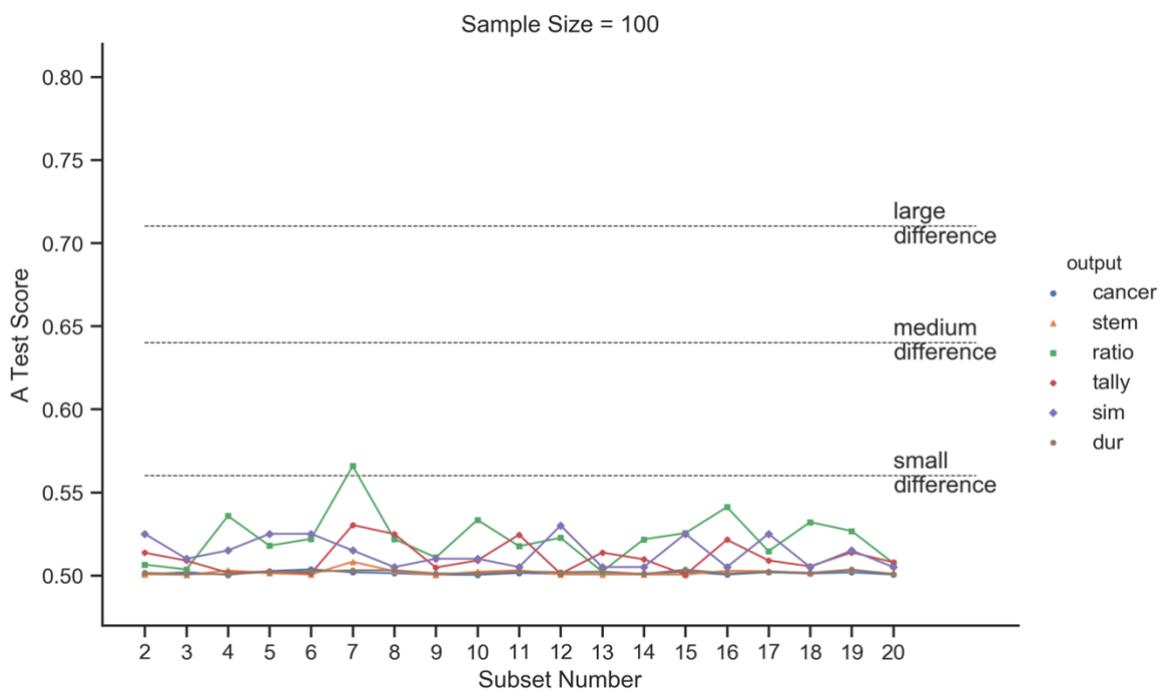


Figure 8.12: A-test score of all sets per sample size of 100 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

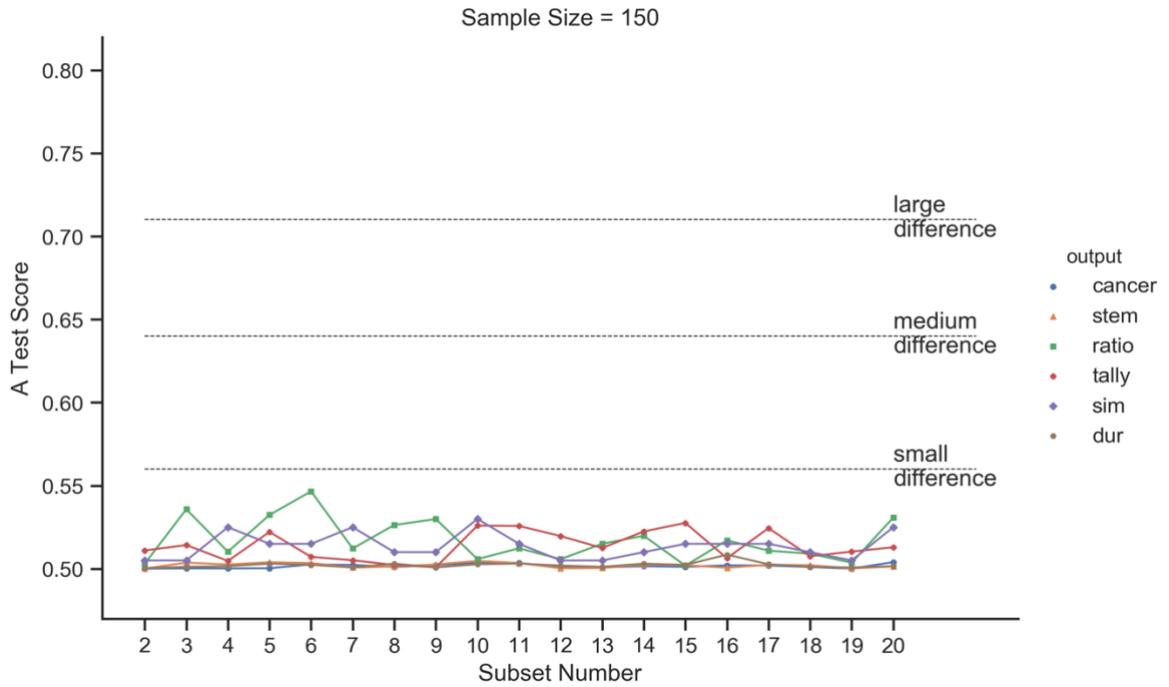


Figure 8.13: A-test score of all sets per sample size of 150 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

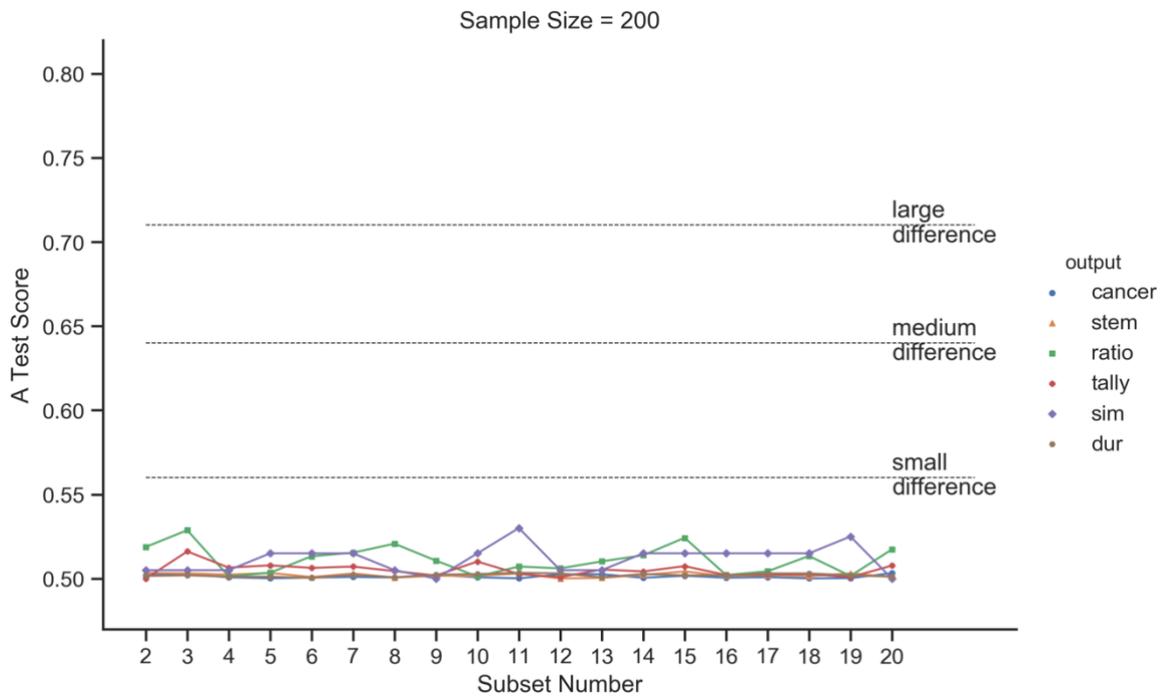


Figure 8.14: A-test score of all sets per sample size of 200 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

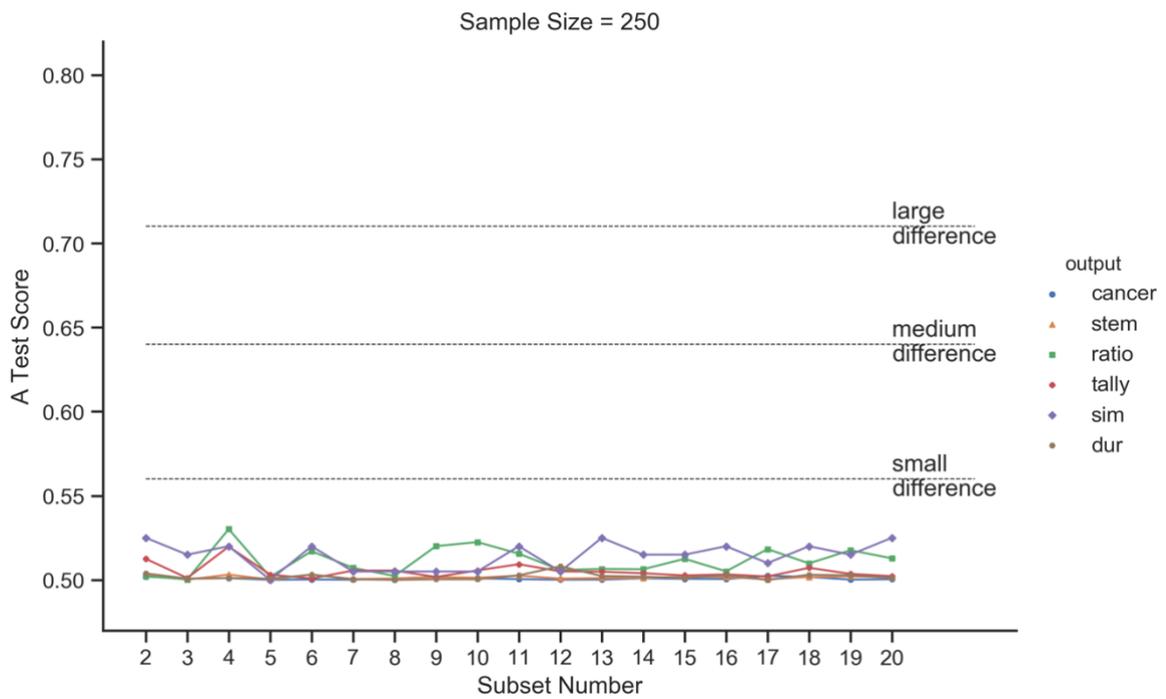


Figure 8.15: A-test score of all sets per sample size of 250 simulation repeat. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.

8.6 Model calibration process heatmaps

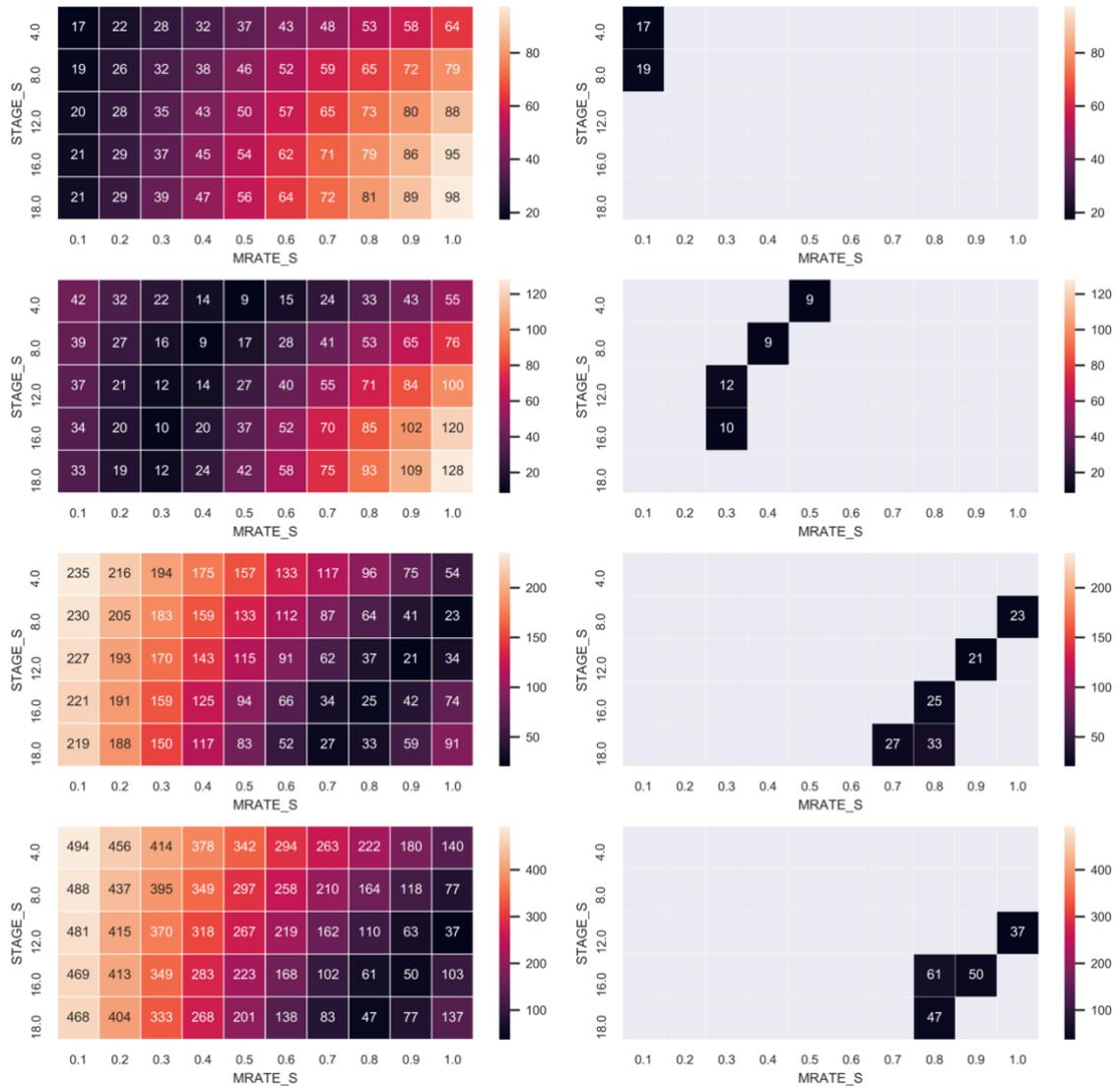


Figure 8.16: Model calibration for cancer growth showing the RMSE between model output and experimental data (from top to bottom) at 24, 48, 72 and 96 hours post seeding. To the right is the lower quartile used to determine the closest fit scenarios which were further analysed using ANCOVA.

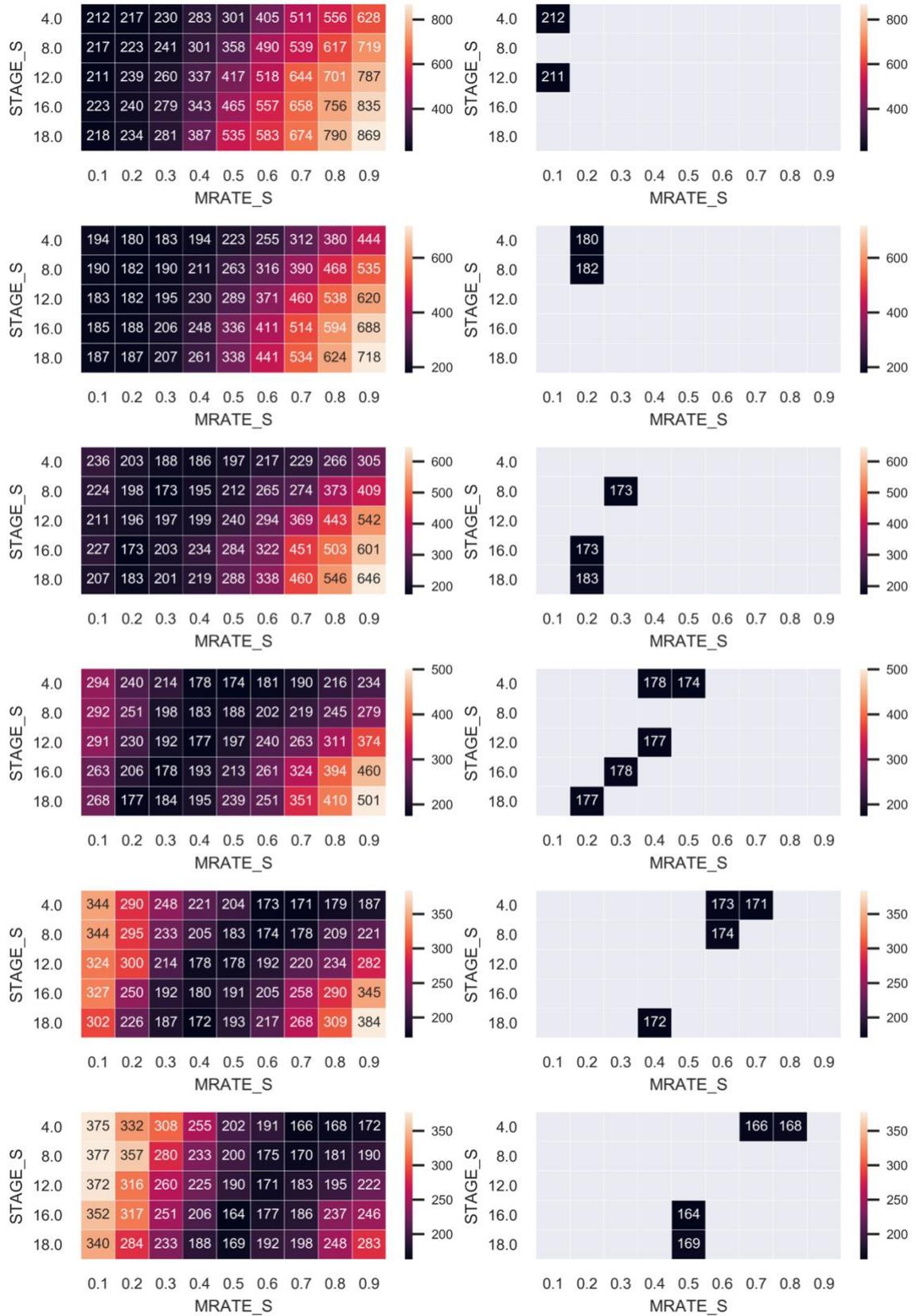


Figure 8.17: Model calibration for stem cell growth showing the total RMSE across all timepoints between model output and experimental data at $cc.MAXSTAGE$ equal to (from top to bottom) 23, 24, 25, 26 and 27 hours, To the right is the lower quartile used to determine the closest fit scenarios which were further analysed using ANCOVA.

Full statistical results of 'null' ABM validation

| Model (Cell type - Seeding density) | Timepoint (real time/model iteration) | Kolmogorov-Smirnov test | | Cohen's 'δ' statistic | |
|--|--|-------------------------|------------|-----------------------|------------|
| | | p value | Test score | Effect size | Test score |
| MG-63 - 25 cells/mm ² | 48 h / 144 | 0.000 | 1.000 | large | 5.706 |
| | 72 h / 216 | 0.000 | 0.833 | large | 2.761 |
| | 96 h / 288 | 0.339 | 0.320 | medium | 0.463 |
| | 120 h / 360 | 0.002 | 0.634 | large | 1.164 |
| | 144 h / 432 | 0.002 | 0.626 | large | 1.734 |
| MG-63 - 50 cells/mm ² | 48 h / 144 | 0.000 | 1.000 | large | 7.679 |
| | 72 h / 216 | 0.051 | 0.458 | large | 0.914 |
| | 96 h / 288 | 0.092 | 0.422 | medium | 0.731 |
| | 120 h / 360 | 0.092 | 0.422 | large | 0.882 |
| | 144 h / 432 | 0.000 | 0.801 | large | 2.536 |
| MG-63 -100 cells/mm ² | 48 h / 144 | 0.000 | 0.100 | large | 3.432 |
| | 72 h / 216 | 0.561 | 0.286 | medium | 0.268 |
| | 96 h / 288 | 0.249 | 0.369 | small | 0.144 |
| | 120 h / 360 | 0.014 | 0.571 | large | 1.400 |
| | 144 h / 432 | 0.000 | 0.833 | large | 1.933 |
| MG-63 -200 cells/mm ² | 48 h / 144 | 0.000 | 0.917 | large | 2.710 |
| | 72 h / 216 | 0.000 | 0.833 | large | 2.054 |
| | 96 h / 288 | 0.000 | 0.967 | large | 3.573 |
| | 120 h / 360 | 0.000 | 1.000 | large | 9.915 |

Table 8.7: Statistical results of the 'null' ABM of cancer growth validated against 4 seeding densities of experimental growth curves of MG-63 cells

| Model (Cell type - Seeding density) | Timepoint (real time/model iteration) | Kolmogorov-Smirnov test | | Cohen's 'δ' statistic | |
|---|---|-------------------------|------------|-----------------------|------------|
| | | p value | Test score | Effect size | Test score |
| hES-MP - 25 cells/mm² | 72 h / 144 | 0.001 | 0.667 | large | 1.909 |
| | 96 h / 216 | 0.033 | 0.487 | large | 0.823 |
| | 120 h / 288 | 0.056 | 0.454 | medium | 0.715 |
| | 144 h / 360 | 0.005 | 0.589 | large | 1.153 |
| | 168 h / 432 | 0.000 | 0.769 | large | 2.476 |
| hES-MP - 50 cells/mm² | 72 h / 144 | 0.155 | 0.384 | small | 0.279 |
| | 96 h / 216 | 0.185 | 0.371 | medium | 0.550 |
| | 120 h / 288 | 0.002 | 0.634 | large | 1.248 |
| | 144 h / 360 | 0.003 | 0.621 | large | 1.265 |
| | 168 h / 432 | 0.004 | 0.602 | large | 1.696 |
| hES-MP -100 cells/mm² | 72 h / 144 | 0.026 | 0.500 | small | 0.132 |
| | 96 h / 216 | 0.006 | 0.583 | large | 1.384 |
| | 120 h / 288 | 0.534 | 0.274 | medium | 0.356 |
| | 144 h / 360 | 0.001 | 0.667 | large | 2.061 |
| | 168 h / 432 | 0.000 | 1.000 | large | 3.480 |
| hES-MP -200 cells/mm² | 72 h / 144 | 0.000 | 0.718 | large | 2.145 |
| | 96 h / 216 | 0.000 | 0.917 | large | 5.178 |
| | 120 h / 288 | 0.000 | 1.000 | large | 6.614 |
| | 144 h / 360 | 0.000 | 1.000 | large | 4.625 |

Table 8.8: Statistical results of the 'null' ABM of stem growth validated against 4 seeding densities of experientnal hES-MP cells

8.7 Local sensitivity analysis results

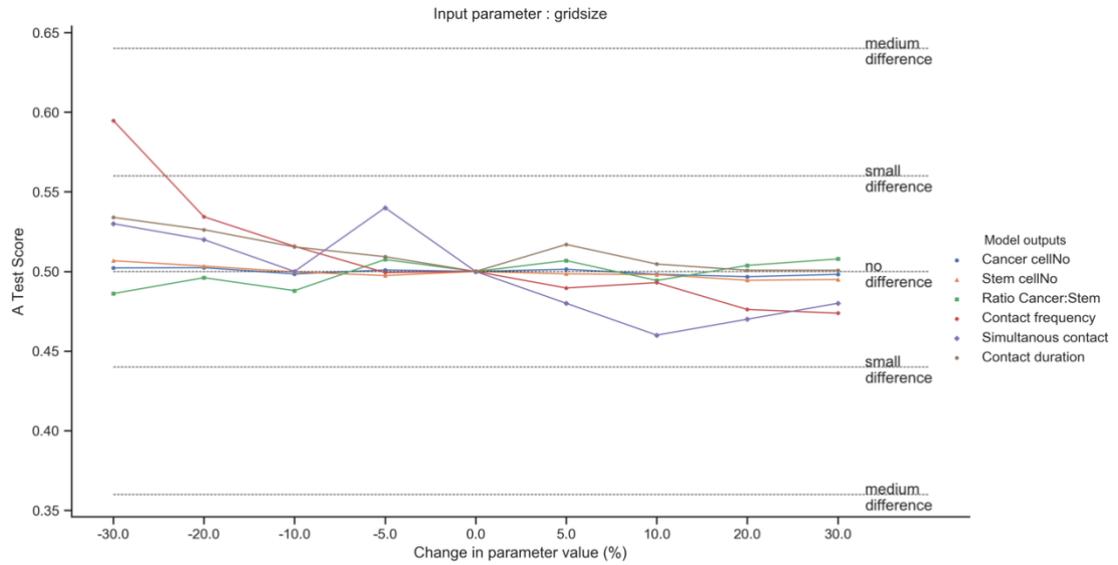
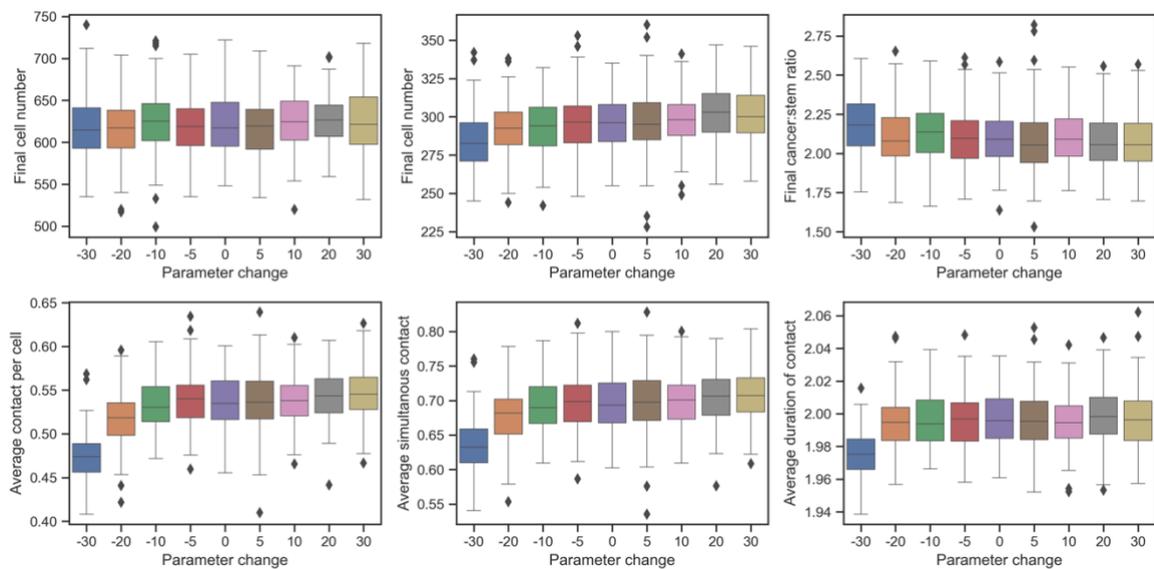


Figure 8.18: A-test score comparing the baseline model output against model outputs resulting from changes in gridsize parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



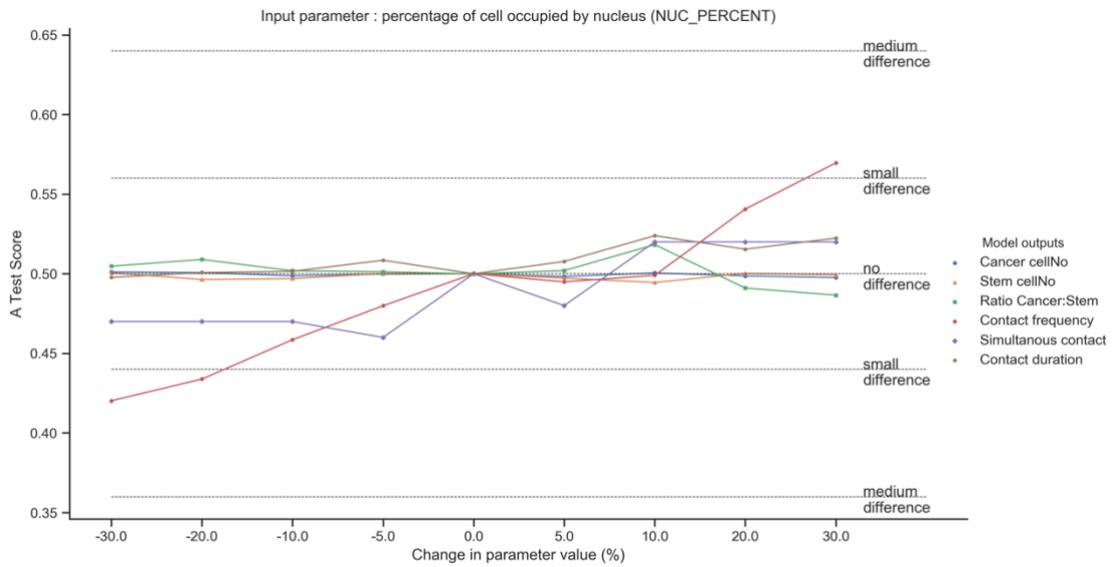
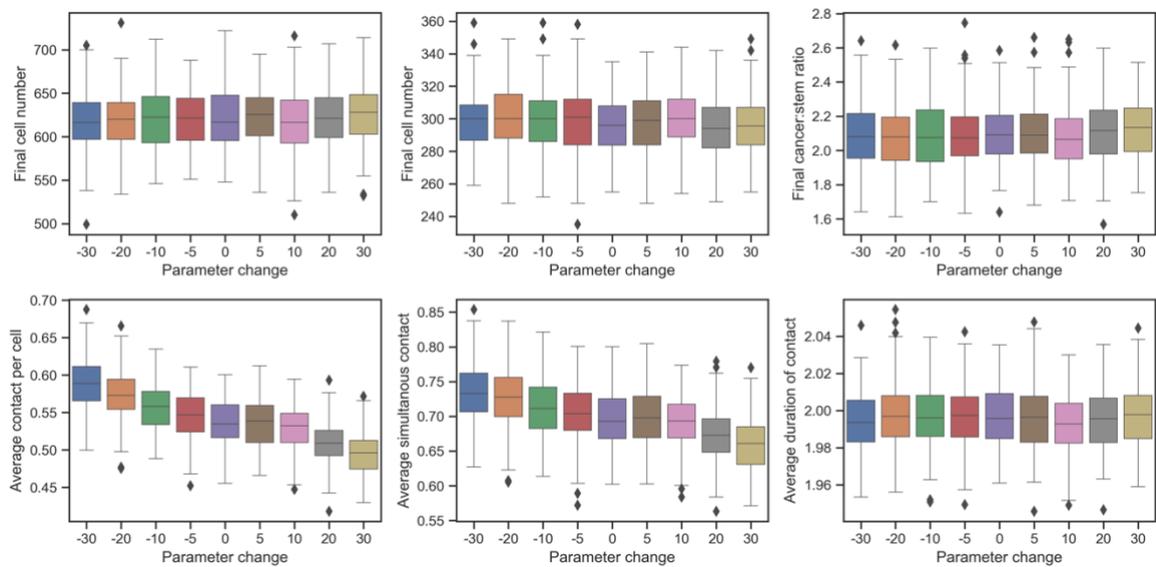


Figure 8.19: A-test score comparing the baseline model output against model outputs resulting from changes in nucl_percent parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



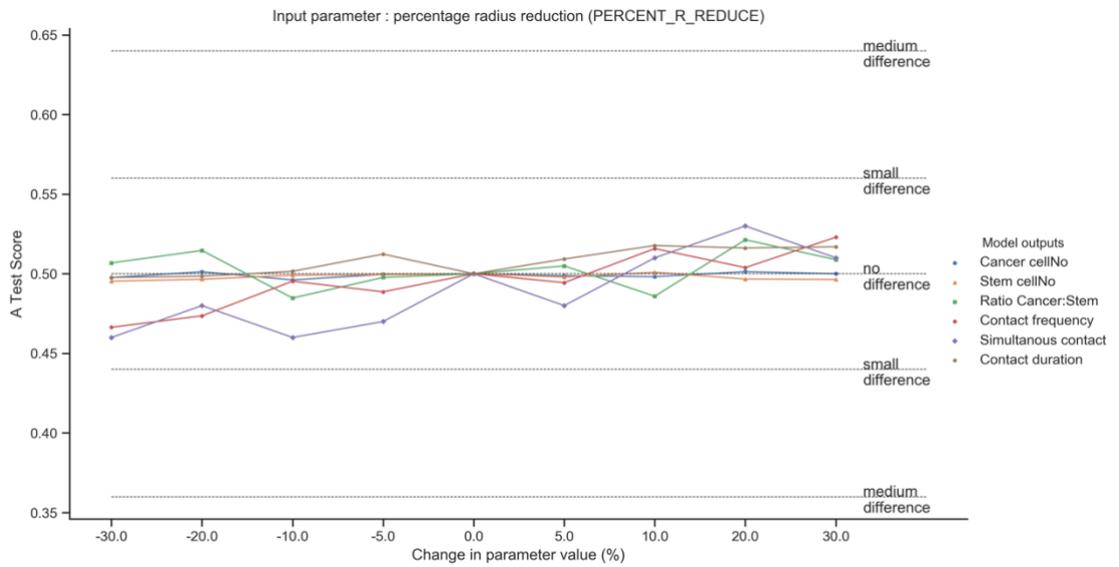
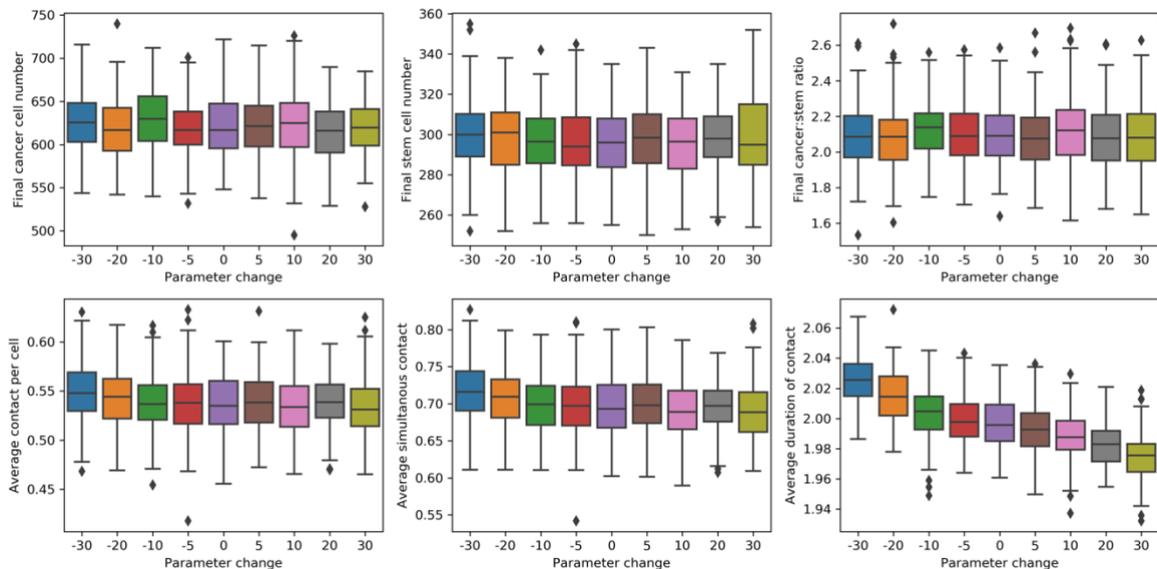


Figure 8.20: A-test score comparing the baseline model output against model outputs resulting from changes in percent_r_reduce parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



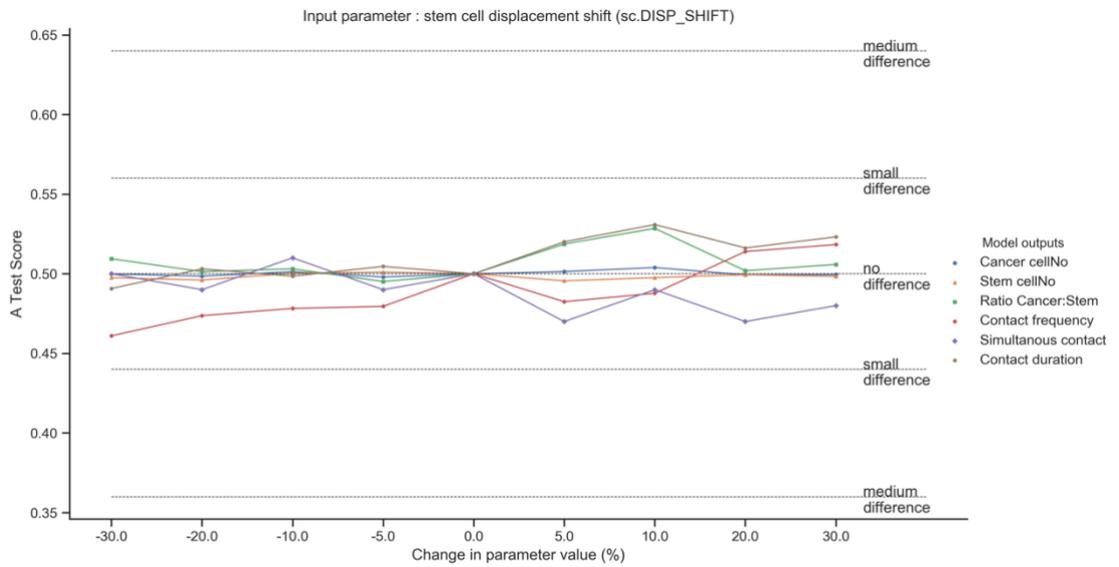
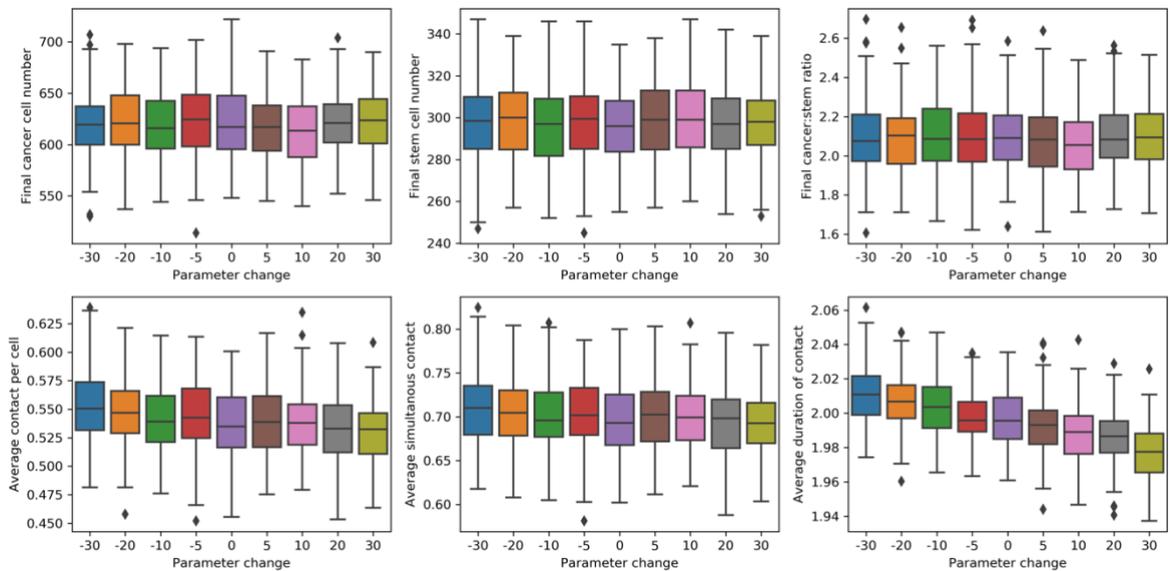


Figure 8.21: A-test score comparing the baseline model output against model outputs resulting from changes in stem cell disp_shift parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



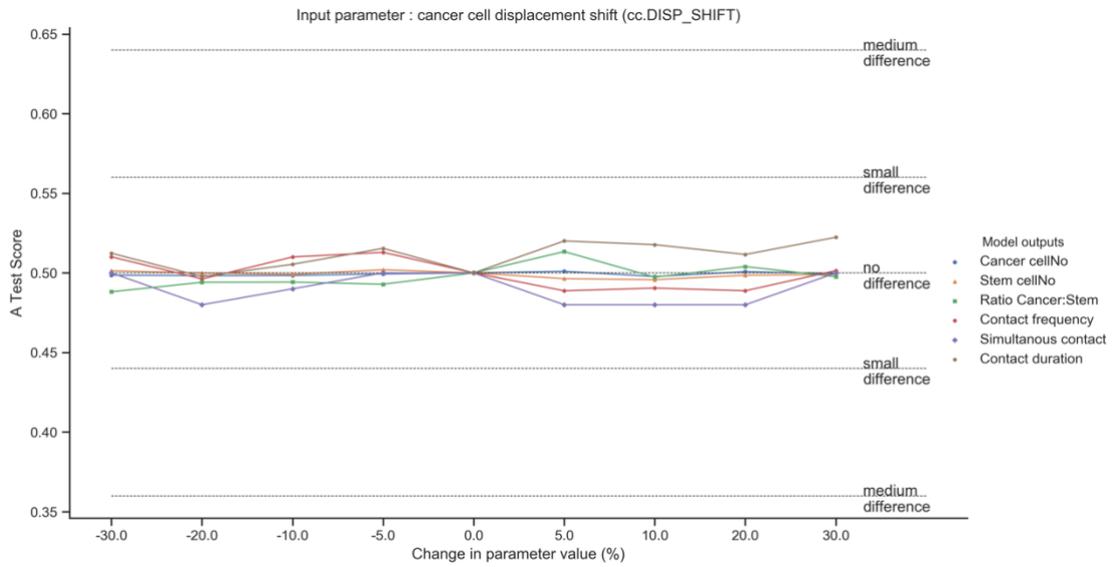
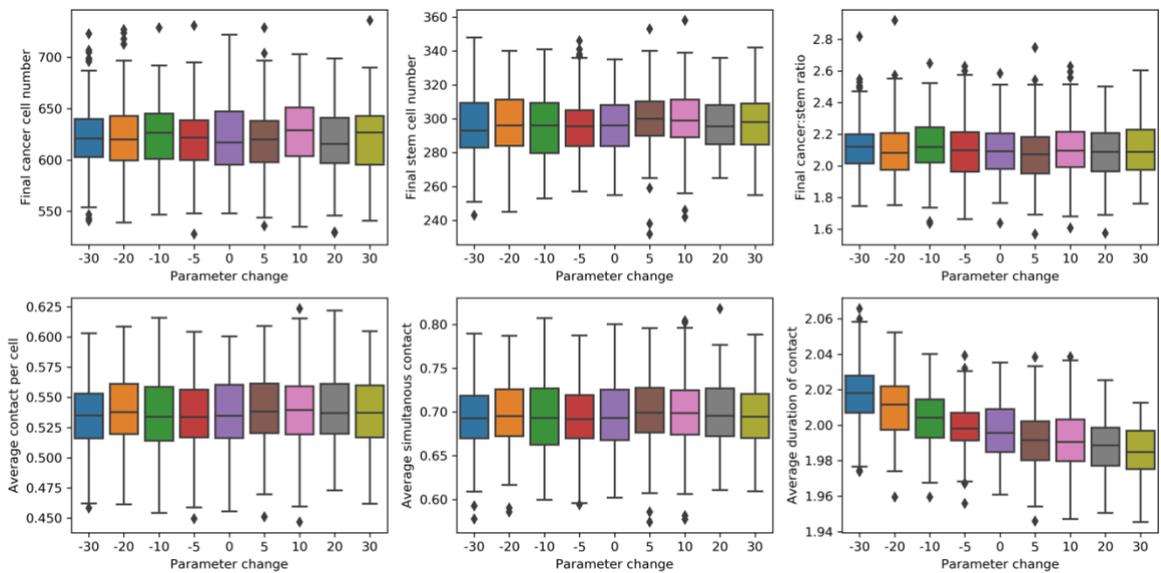


Figure 8.22: A-test score comparing the baseline model output against model outputs resulting from changes in cancer cell disp_shift parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



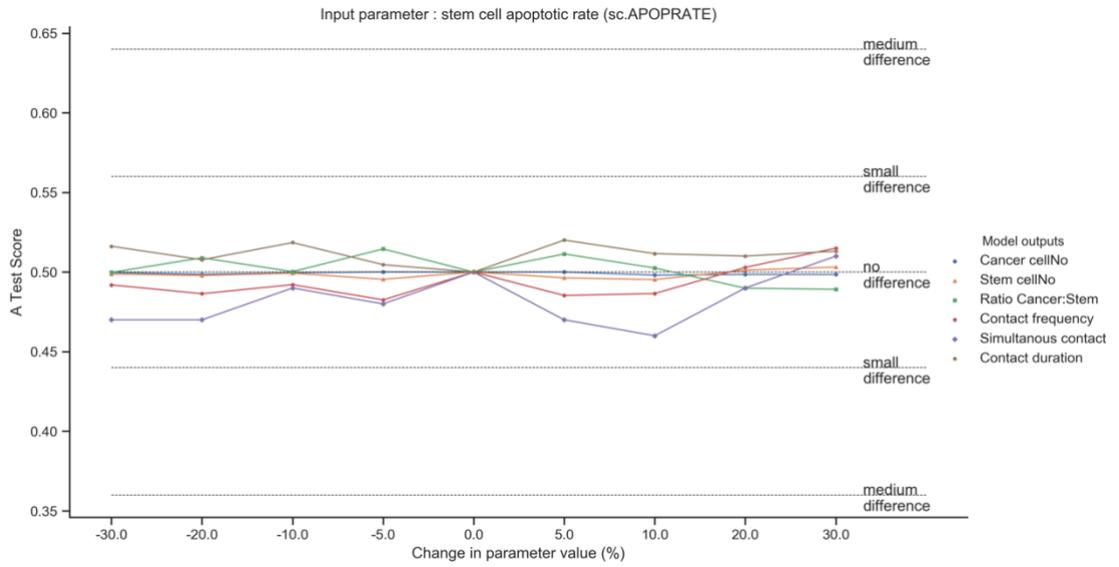
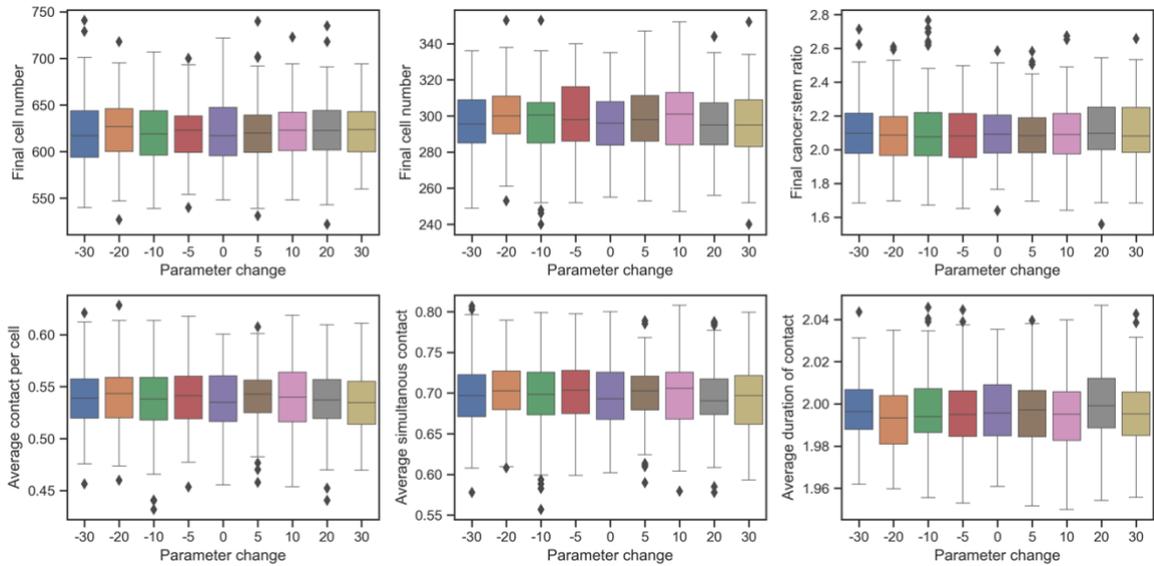


Figure 8.23: A-test score comparing the baseline model output against model outputs resulting from changes in stem cell death rate parameter. Statistical comparison done using the Vargha-Delaney A-Test with magnitude of effect marked as horizontal lines.



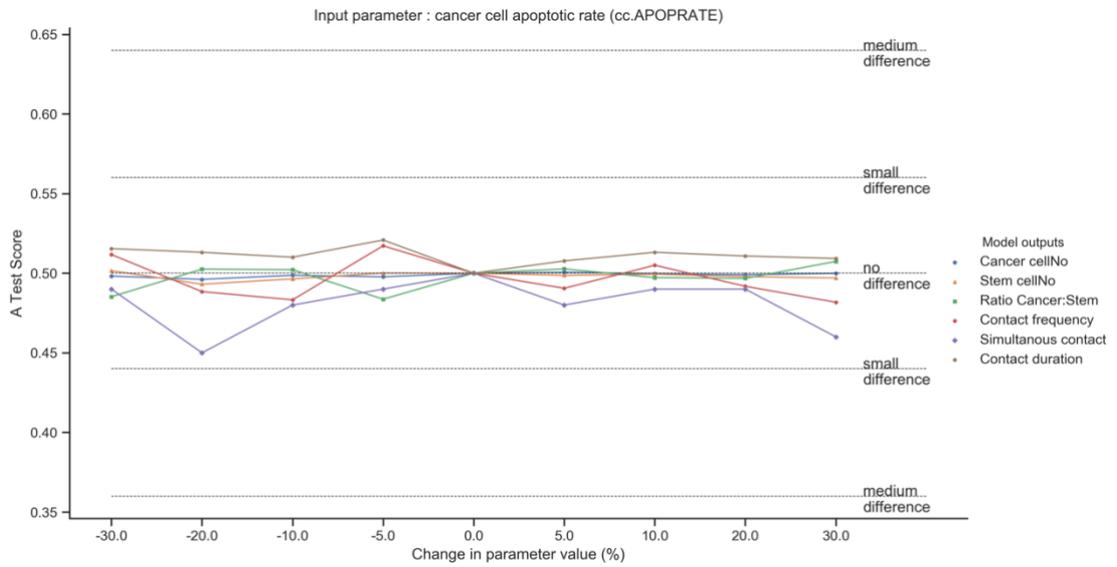
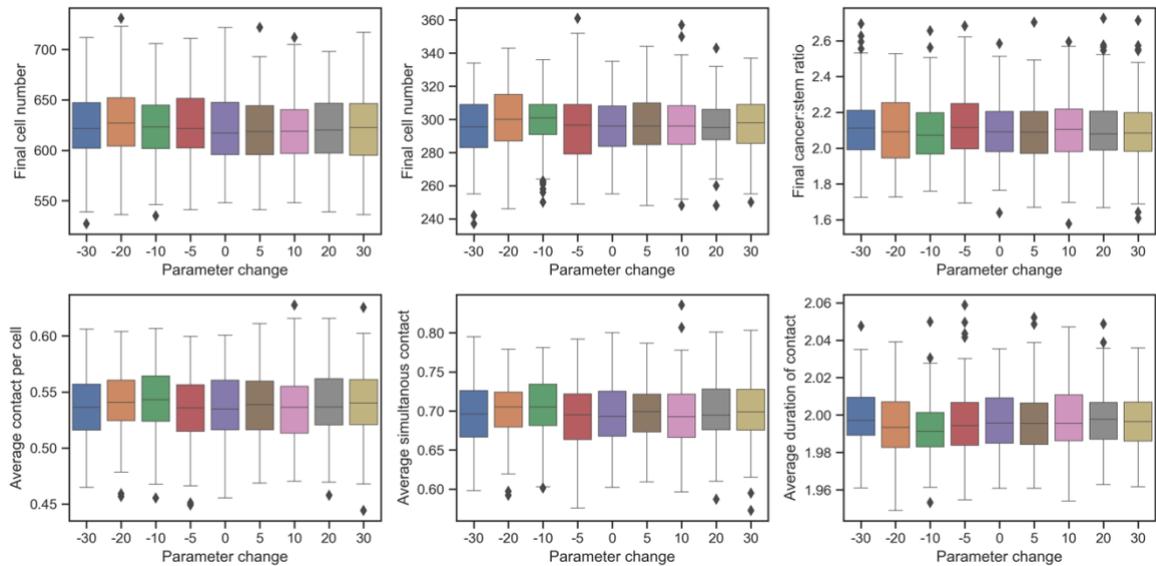
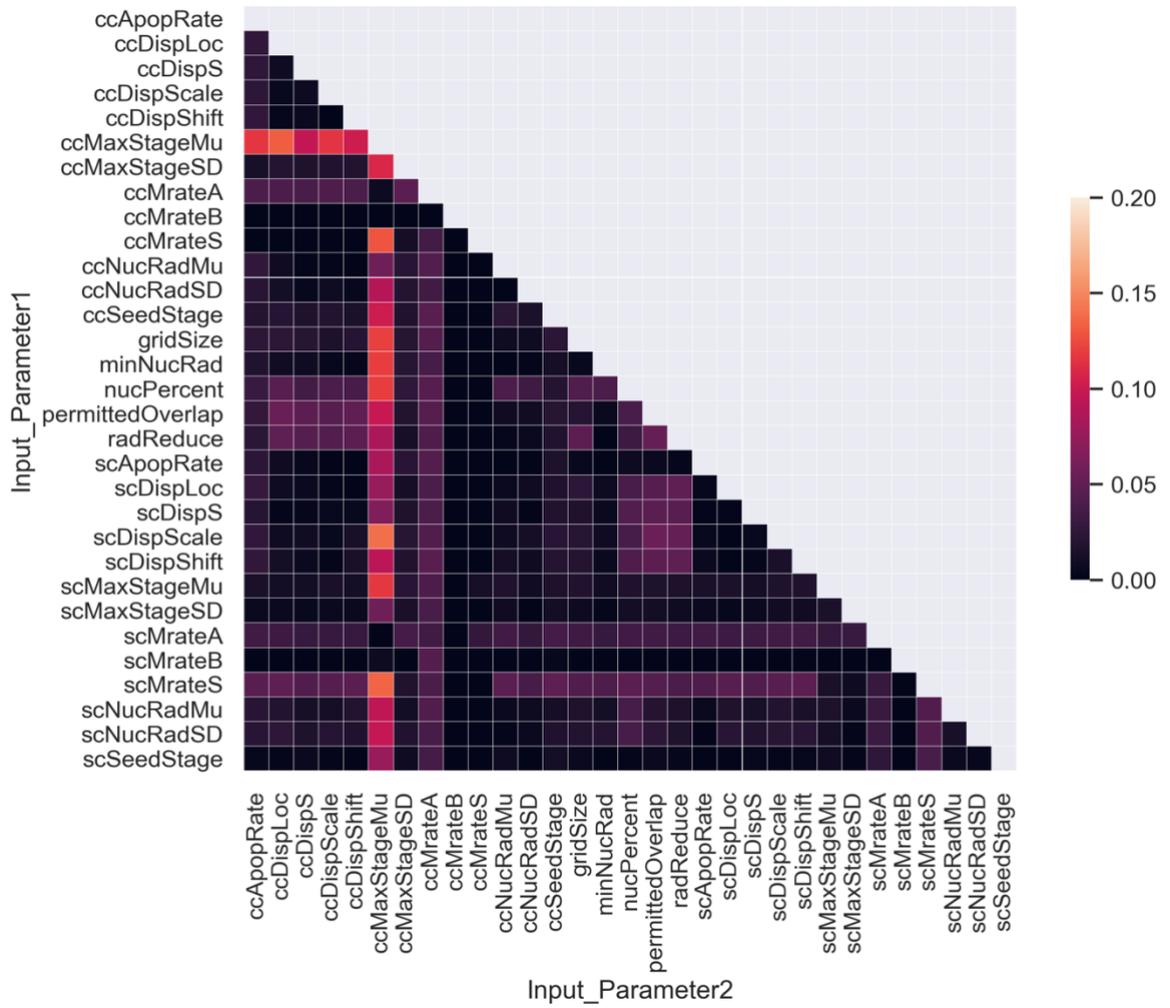


Figure 8.24: A-test score comparing the baseline model output against model outputs resulting from changes in cancer cell death rate parameter. Statistical comparison done using the Vargha-Delany A-Test with magnitude of effect marked as horizontal lines.

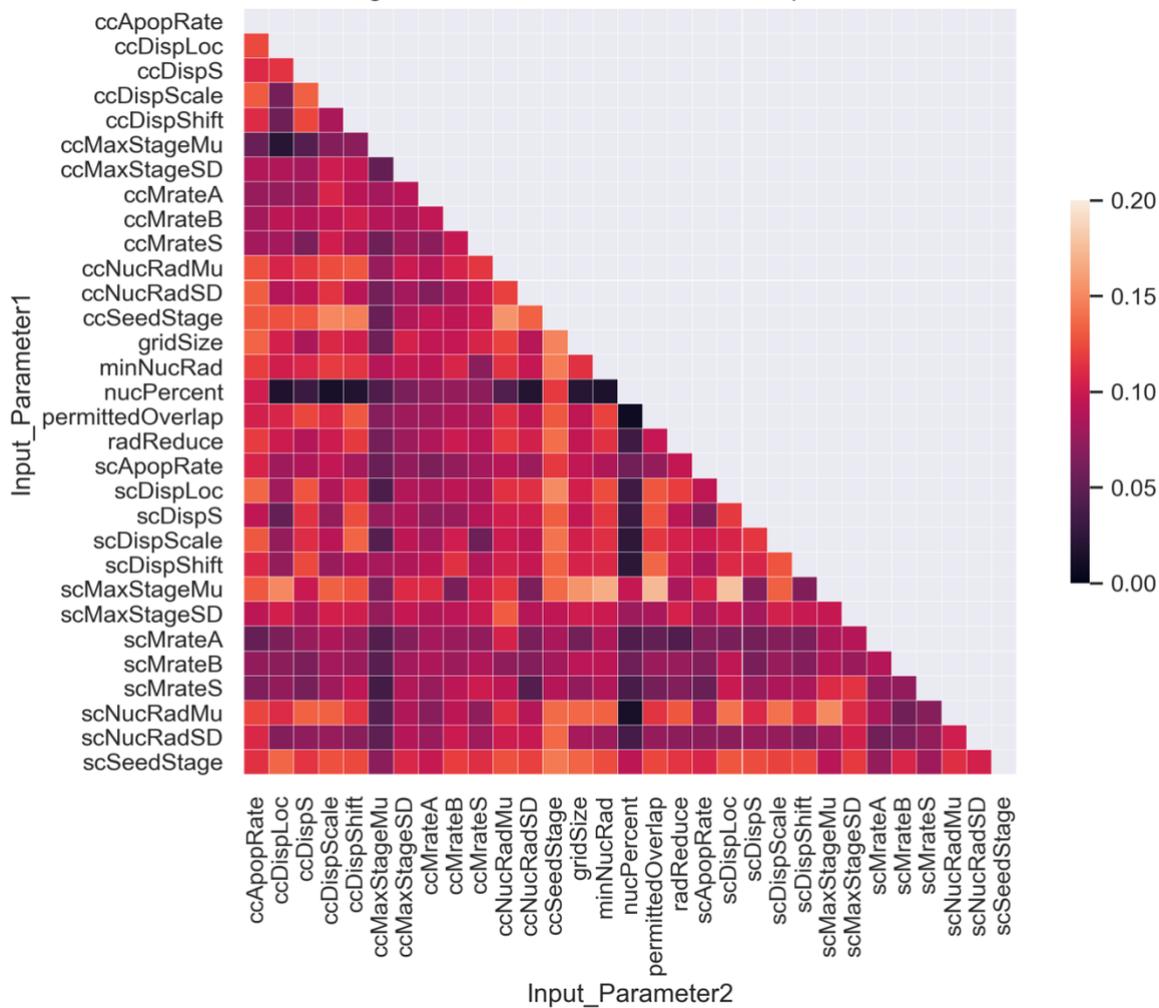


8.8 Global sensitivity analysis results

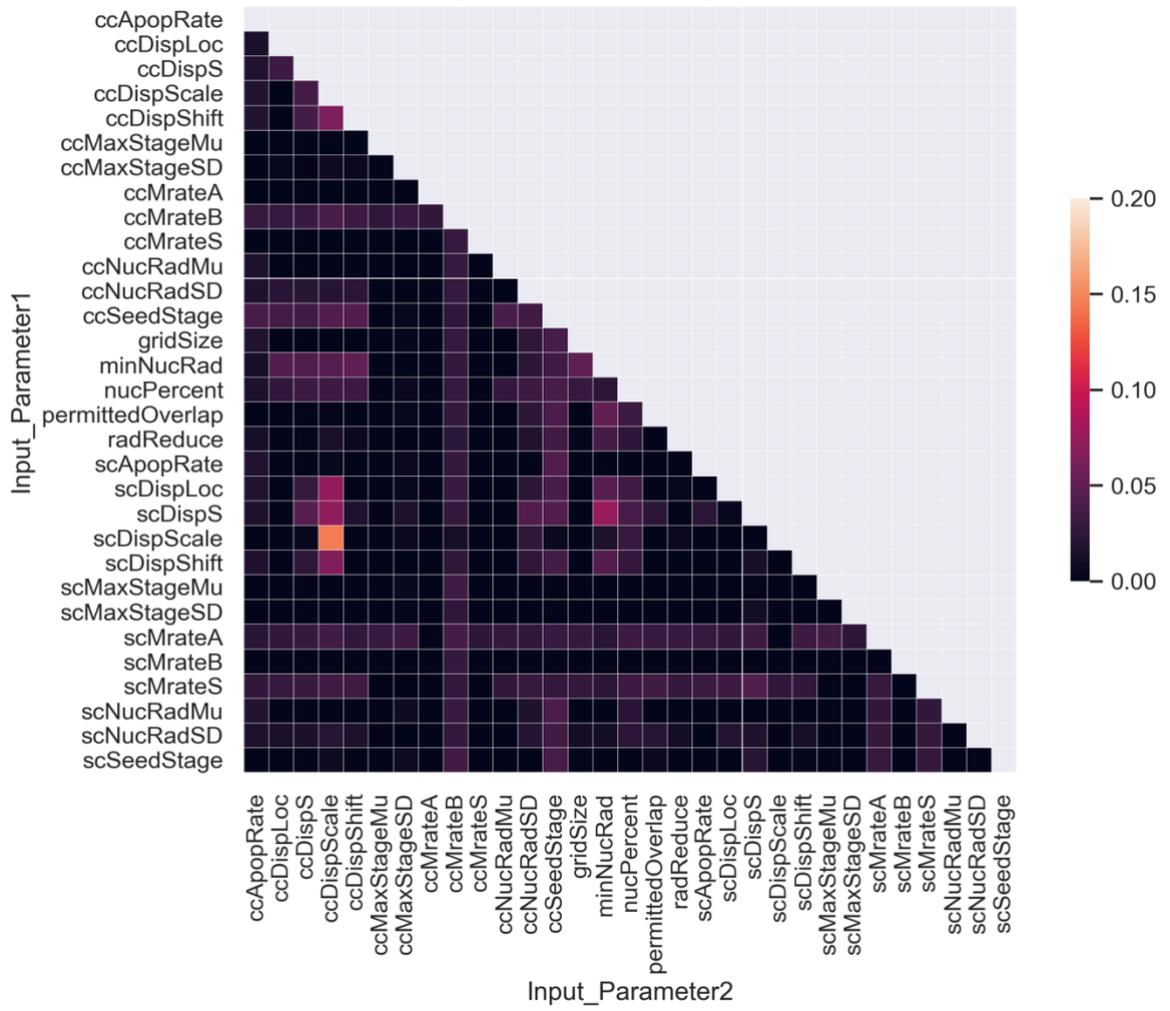
Cancer cell number



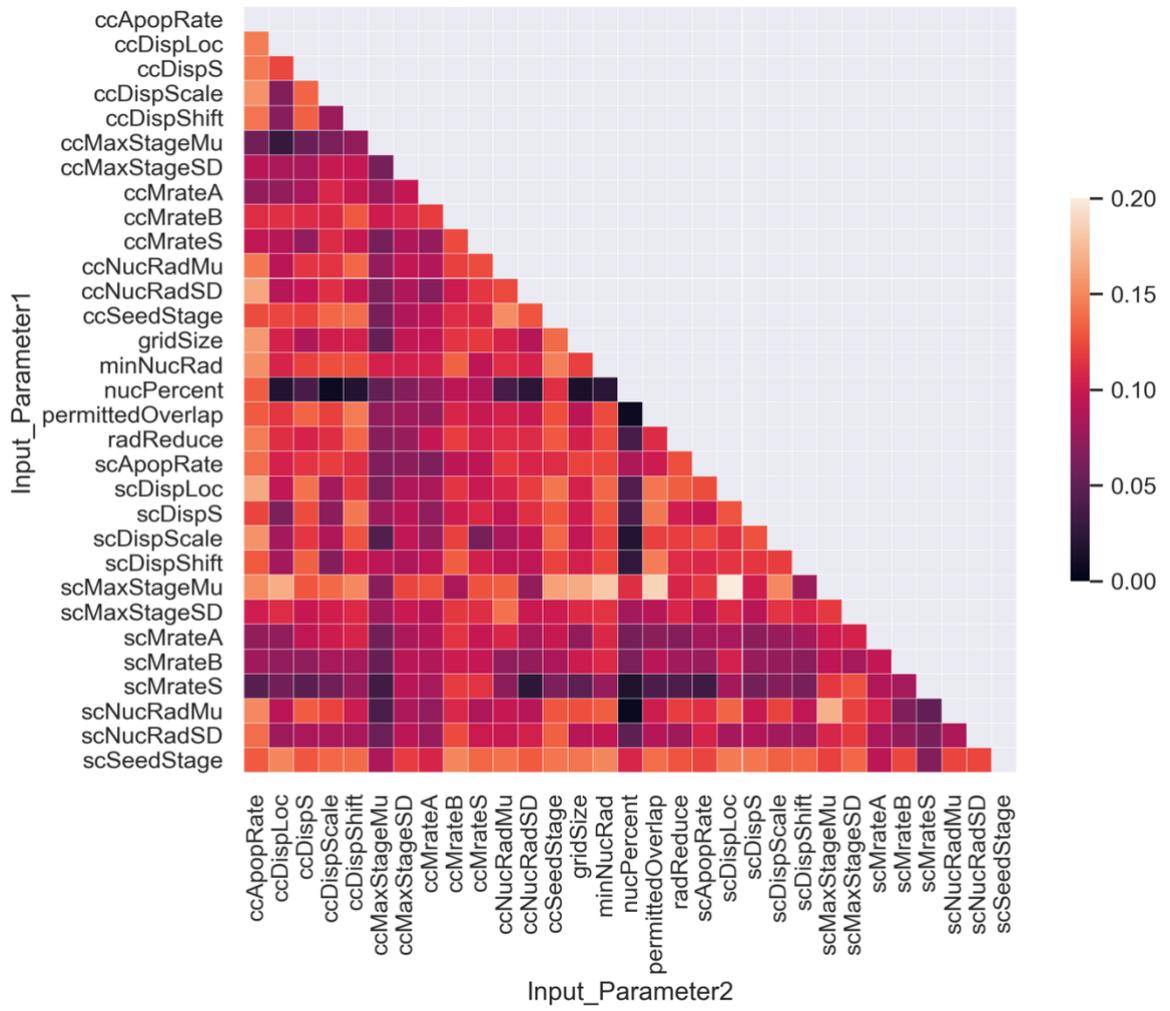
Average number of simultaneous contacts per cancer cell



Average duration of contact per cancer cell



Average number of contacts per cancer cell



8.9 Global sensitivity analysis parameters

Table 8.9: Table of parameters used in global sensitivity analysis and their respective upper and lower bounds.

| Parameter | Upper bound | Lower bound |
|------------------|------------------------|------------------------|
| ccMrateB | -5.6×10^{-2} | -3.9×10^{-2} |
| ccMrateA | 122 | 183 |
| scMrateB | -0.16 | -0.17 |
| scMrateA | 94.06 | 141.1 |
| ccMaxStageMu | 19.95 | 29.92 |
| ccMaxStageSD | 3.089 | 4.634 |
| scMaxStageMu | 22.4 | 33.6 |
| scMaxStageSD | 2.4 | 3.6 |
| ccMrateS | 0.64 | 0.96 |
| scMrateS | 0.4 | 0.6 |
| ccSeedStage | 43.2 | 64.8 |
| scSeedStage | 43.2 | 64.8 |
| ccApopRate | 2.627×10^{-5} | 3.940×10^{-5} |
| scApopRate | 4.906×10^{-5} | 7.358×10^{-5} |
| nucPercent | 8 | 12 |
| ccNucRadMu | 9.166 | 13.75 |
| ccNucRadSD | 1.039 | 1.558 |
| scNucRadMu | 7.722 | 11.59 |
| scNucRadSD | 1.221 | 1.832 |
| minNucRad | 4 | 6 |
| radReduce | 8 | 12 |
| gridSize | 40 | 60 |
| permittedOverlap | -6 | -4 |
| ccDispS | 0.529 | 0.793 |
| ccDispLoc | -0.741 | -0.494 |
| ccDispScale | 3.229 | 4.843 |
| scDispS | 0.360 | 0.540 |
| scDispLoc | -6.997 | -4.664 |
| scDispScale | 15.82 | 23.73 |
| ccDispShift | 5.600 | 8.400 |
| scDispShift | 4.000 | 6.000 |

References

Aboody, K. S. *et al.* (2000) ‘Neural stem cells display extensive tropism for pathology in adult brain: evidence from intracranial gliomas.’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(23), pp. 12846–51. doi: 10.1073/pnas.97.23.12846.

Adra, S. *et al.* (2010) ‘Development of a three dimensional multiscale computational model of the human epidermis.’, *PloS one*, 5(1), p. e8511. doi: 10.1371/journal.pone.0008511.

An, G. *et al.* (2009) ‘Agent-based models in translational systems biology’, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(2), pp. 159–171. doi: 10.1002/wsbm.45.

Andrews, P. S. *et al.* (2010) ‘The CoSMoS Process, Version 0.1: A Process for the Modelling and Simulation of Complex Systems’, *Iceccs 2010*, (March), pp. 276–285. Available at: <http://www-users.cs.york.ac.uk/susan/bib/ss/nonstd/tr453.pdf>.

Balci, O. (1998) *Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice. Chapter Verification, Validation and Testing*. New York: John Wiley & Sons.

Baskar, R. *et al.* (2012) ‘Cancer and radiation therapy: current advances and future directions.’, *International journal of medical sciences*, 9(3), pp. 193–9. doi: 10.7150/ijms.3635.

Bauer, A. L., Beauchemin, C. A. . and Perelson, A. S. (2008) ‘Agent-based modeling of host-pathogen systems: The successes and challenges’, *NIH Public Access*, 179(10), pp. 1379–1389.

Bigdeli, N. *et al.* (2010) ‘Superior osteogenic capacity of human embryonic stem cells adapted to matrix-free growth compared to human mesenchymal stem cells’, *Tissue Engineering - Part A*, 16(11), pp. 3427–3440. doi: 10.1089/ten.tea.2010.0112.

Billiau, A. *et al.* (1977) ‘Human interferon: mass production in a newly established cell line, MG 63’, *Antimicrobial Agents and Chemotherapy*, 12(1), pp. 11–15. doi: 10.1128/AAC.12.1.11.

Boyle, P. and Levine, B. (2008) *World Cancer Reports 2008*. IARC Press, International Agency for Research on Cancer.

Bupa (2014) *Lymphadenectomy*. Available at: <http://www.bupa.co.uk/individuals/health-information/directory/l/lymphadenectomy>.

Byrne, H. (2008) ‘Multiscale Modelling of Solid Tumour Growth’, in *Selected Topics in Cancer Modelling*, pp. 449–470.

Byrne, H. and Drasdo, D. (2009) ‘Individual-based and continuum models of growing cell populations: a comparison.’, *Journal of mathematical biology*, 58(4–5), pp. 657–687. doi: 10.1007/s00285-008-0212-0.

Cai, J., Weiss, M. L. and Rao, M. S. (2004) ‘In search of “stemness”’, *Experimental Hematology*, 32(7), pp. 585–598. doi: 10.1016/j.exphem.2004.03.013.

Cancer Research UK (2014) *Stem cell transplants*. Available at: <http://www.cancerresearchuk.org/cancer-help/about-cancer/treatment/transplant/stem-cell-transplants>.

Cancer Research UK (2020a) *Cancer incidence for common cancers*. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence#heading-One>.

Cancer Research UK (2020b) *Cancer Research UK*. Available at: <https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer>.

Cantarella, C. *et al.* (2009) ‘Analysis and modelling of motility of cell populations with MotoCell’, *BMC Bioinformatics*, 10(Suppl 12), p. S12. doi: 10.1186/1471-2105-10-S12-S12.

Carey, P. J. (2003) ‘Drug-Induced Myelosuppression’, *Drug Safety*, 26(10), pp. 691–706.

Cellosaurus (2012) *MG-63 CVCL_0426*. Available at: https://web.expasy.org/cellosaurus/CVCL_0426.

Coakley, S., Smallwood, R. and Holcombe, M. (2006) ‘Using X-machines as a Formal Basis for Describing Agents in Agent-Based Modelling’.

Codling, E. A. *et al.* (2004) ‘Random walk models for the movement and recruitment of reef fish larvae’, *Marine Ecology Progress Series*, 279, pp. 215–224. doi: 10.3354/meps279215.

Coleman, M. P. *et al.* (2008) ‘Cancer survival in five continents: a worldwide population-based study (CONCORD).’, *The lancet oncology*, 9(8), pp. 730–56. doi: 10.1016/S1470-2045(08)70179-7.

Concise Oxford English Dictionary. Eleventh e (2009). New York: Oxford University Press.

Corsten, M. F. and Shah, K. (2008) ‘Therapeutic stem-cells for cancer treatment: hopes and hurdles in tactical warfare’, *The Lancet Oncology*, 9(4), pp. 376–384. doi: 10.1016/S1470-2045(08)70099-8.

Cousin, B. *et al.* (2009) ‘Adult stromal cells derived from human adipose tissue provoke pancreatic cancer cell death both in vitro and in vivo.’, *PloS one*, 4(7), p. e6278. doi: 10.1371/journal.pone.0006278.

Crooks, A. T. and Heppenstall, A. J. (2012) ‘Chapter 5: Introduction to Agent-Based Modelling’, in Heppenstall, A. J. *et al.* (eds) *Agent-Based Models of Geographical Systems*. Dordrecht: Springer Netherlands, pp. 347–360. doi: 10.1007/978-90-481-8927-4.

Dasari, V. R., Velpula, K. K., *et al.* (2010) ‘Cord blood stem cell-mediated induction of apoptosis in glioma downregulates X-linked inhibitor of apoptosis protein (XIAP).’, *PloS one*, 5(7), p. e11813. doi: 10.1371/journal.pone.0011813.

Dasari, V. R., Kaur, K., *et al.* (2010) ‘Upregulation of PTEN in glioma cells by cord blood mesenchymal stem cells inhibits migration via downregulation of the PI3K/Akt pathway.’, *PloS one*, 5(4), p. e10350. doi: 10.1371/journal.pone.0010350.

Deisboeck, T. S. and Stamatakis, G. S. (eds) (2011) *Multiscale Cancer Modelling*. First Edit. London: CRC Press Taylor & Francis Group.

Deisboeck, T. S. *et al.* (2011) ‘Multiscale Cancer Modeling’, *National Institutes of Health*, 15(13).

Djouad, F. *et al.* (2003) ‘Immunosuppressive effect of mesenchymal stem cells favors tumor growth in allogeneic animals.’, *Blood*, 102(10), pp. 3837–44. doi: 10.1182/blood-2003-04-1193.

Dougan, M., Dranoff, G. and Dougan, S. K. (2019) ‘Cancer immunotherapy: Beyond checkpoint blockade’, *Annual Review of Cancer Biology*, 3(1), pp. 55–75. doi: 10.1146/annurev-cancerbio-030518-055552.

Fairley, R. E. (1976) ‘Dynamic testing of simulation software’, in *1976 Summer Computer Simulation Conference*. Washington, D.C., pp. 40–46. doi: 10.1109/C-M.1978.218132.

Ferraro, F., Lo Celso, C. and Scadden, D. (2010) ‘Adult stem cells and their niches’,

Advances in Experimental Medicine and Biology, 695, pp. 155–168. doi: 10.1007/978-1-4419-7037-4_11.

Fricke, G. M. *et al.* (2016) ‘Persistence and Adaptation in Immunity: T Cells Balance the Extent and Thoroughness of Search’, *PLOS Computational Biology*. Edited by M. Meier-Schellersheim, 12(3), p. e1004818. doi: 10.1371/journal.pcbi.1004818.

Friesen, R. D. M. and M. R. (2018) ‘A Simple Example of Emergence’, (June), pp. 1–6.

Galiè, M. *et al.* (2008) ‘Mesenchymal stem cells share molecular signature with mesenchymal tumor cells and favor early tumor growth in syngeneic mice.’, *Oncogene*, 27(18), pp. 2542–51. doi: 10.1038/sj.onc.1210920.

Genot, A. J., Fujii, T. and Rondelez, Y. (2013) ‘In vitro regulatory models for systems biology.’, *Biotechnology advances*. Elsevier Inc., 31(6), pp. 789–96. doi: 10.1016/j.biotechadv.2013.04.008.

Goroehowski, T. E. (2016) ‘Agent-based modelling in synthetic biology’, *Essays in Biochemistry*, 60(4), pp. 325–336. doi: 10.1042/EBC20160037.

Grimm, V. *et al.* (2006) ‘A standard protocol for describing individual-based and agent-based models’, *Ecological Modelling*, 198(1–2), pp. 115–126. doi: 10.1016/j.ecolmodel.2006.04.023.

Hanahan, D., Weinberg, R. A. and Francisco, S. (2000) ‘The Hallmarks of Cancer Review University of California at San Francisco’, *Cell*, 100, pp. 57–70.

Hanahan, D. and Weinberg, R. a (2011) ‘Hallmarks of cancer: the next generation.’, *Cell*. Elsevier Inc., 144(5), pp. 646–74. doi: 10.1016/j.cell.2011.02.013.

Hesketh, R. (2013) *Introduction to Cancer Biology*. First edit. Cambridge: Cambridge University Press.

Huang, G. *et al.* (2012) ‘Engineering three-dimensional cell mechanical microenvironment with hydrogels.’, *Biofabrication*, 4(4), p. 042001. doi: 10.1088/1758-5082/4/4/042001.

Hyland, K. M. (2007) ‘Cell proliferation and its regulation’, *University of California*, pp. 1–57. doi: 10.1109/NGMAST.2010.31.

Ivanova, N. B. *et al.* (2002) ‘A Stem Cell Molecular Signature’, *Science*, 298(5593), pp. 1–45.

Jennings, N. R. (2000) ‘On agent-based software engineering’, *Artificial Intelligence*, 117(September 1999), pp. 277–296.

Karlsson, C. *et al.* (2009) ‘Human embryonic stem cell-derived mesenchymal

progenitors-Potential in regenerative medicine’, *Stem Cell Research*, 3(1), pp. 39–50. doi: 10.1016/j.scr.2009.05.002.

Karnoub, A. E. *et al.* (2007) ‘Mesenchymal stem cells within tumour stroma promote breast cancer metastasis.’, *Nature*, 449(7162), pp. 557–63. doi: 10.1038/nature06188.

Kent, M. (2000) *Advanced Biology*. Oxford University Press.

Khakoo, A. Y. *et al.* (2006) ‘Human mesenchymal stem cells exert potent antitumorigenic effects in a model of Kaposi’s sarcoma.’, *The Journal of experimental medicine*, 203(5), pp. 1235–47. doi: 10.1084/jem.20051921.

Klopp, A. H. *et al.* (2011) ‘Concise review: Dissecting a discrepancy in the literature: do mesenchymal stem cells support or suppress tumor growth?’, *Stem cells*, 29(1), pp. 11–19. doi: 10.1002/stem.559.

Knowles, M. A. and Selby, P. J. (2005) *Introduction to the Cellular and Molecular Biology of Cancer*. Fourth Edi. Edited by M. A. Knowles and P. J. Selby. New York: Oxford University Press.

Kohl, P. *et al.* (2000) ‘Computational modelling of biological systems: tools and visions’, *The Royal Society*, 358, pp. 579–610.

Kucerova, L. *et al.* (2010) ‘Tumor cell behaviour modulation by mesenchymal stromal cells.’, *Molecular cancer*, 9, p. 129. doi: 10.1186/1476-4598-9-129.

Kumar, P. and Clark, M. (2005) *Clinical Medicine*. Sixth Edit. doi: 10.1017/CBO9781107415324.004.

Lee Ventola, C. (2017) ‘Cancer immunotherapy, part 3: Challenges and future trends’, *P and T*, 42(8), pp. 514–521. doi: 10.1007/978-3-319-17900-1_176.

Li, L. *et al.* (2010) ‘Individual Cell Movement, Asymmetric Colony Expansion, Rho-Associated Kinase, and E-Cadherin Impact the Clonogenicity of Human Embryonic Stem Cells’, *Biophysical Journal*, 98(11), pp. 2442–2451. doi: 10.1016/j.bpj.2010.02.029.

Li, O. *et al.* (2013) ‘Human Embryonic Stem Cell-Derived Mesenchymal Stroma Cells (hES-MSCs) Engraft In Vivo and Support Hematopoiesis without Suppressing Immune Function: Implications for Off-The Shelf ES-MSC Therapies’, *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0055319.

Li, X. *et al.* (2013) ‘Skin stem cell hypotheses and long term clone survival--explored using agent-based modelling.’, *Scientific reports*, 3, p. 1904. doi: 10.1038/srep01904.

Lin, G. *et al.* (2010) ‘Effects of Transplantation of Adipose Tissue-Derived Stem Cells on Prostate Tumor’, 70(10), pp. 1066–1073. doi: 10.1002/pros.21140.Effects.

Lostumbo, L. *et al.* (2004) ‘Review: Prophylactic mastectomy for prevention of

breast cancer.’, *Cochrane Database of Systematic Reviews*, 48(4), pp. 252–253. doi: 10.1002/14651858.cd002748.pub2.

Lu, Y. *et al.* (2008) ‘The growth inhibitory effect of mesenchymal stem cells on tumor cells in vitro and in vivo’, pp. 245–251.

Luengo-Fernandez, R. *et al.* (2013) ‘Economic burden of cancer across the European Union: a population-based cost analysis.’, *The Lancet. Oncology*. Elsevier Ltd, 14(12), pp. 1165–74. doi: 10.1016/S1470-2045(13)70442-X.

Macal, C. M. and North, M. J. (2005) ‘Tutorial on Agent-Based Modeling and Simulation’, in *Proceedings of the 2005 Winter Simulation Conference*, pp. 2–15.

Macal, C. M. and North, M. J. (2010) ‘Tutorial on Agent-Based Modeling and Simulation’, *Journal of Simulation*. Palgrave Macmillan, 4(3), pp. 151–162. doi: 10.1057/jos.2010.3.

Magidson, V. and Khodjakov, A. (2013) ‘Circumventing Photodamage in Live-Cell Microscopy’, in, pp. 545–560. doi: 10.1016/B978-0-12-407761-4.00023-3.

Mansury, Y. *et al.* (2002) ‘Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model.’, *Journal of theoretical biology*, 219(3), pp. 343–70. doi: 10.1006/jtbi.2002.3131.

Meijering, E., Dzyubachyk, O. and Smal, I. (2012) ‘Methods for cell and particle tracking’, *Methods in Enzymology*, 504(February), pp. 183–200. doi: 10.1016/B978-0-12-391857-4.00009-4.

Meineke, F. A., Potten, C. S. and Loeffler, M. (2001) ‘Cell migration and organization in the intestinal crypt using a lattice-free model’, (1993), pp. 253–266.

Muehlberg, Fabian L *et al.* (2009) ‘Tissue-resident stem cells promote breast cancer growth and metastasis.’, *Carcinogenesis*, 30(4), pp. 589–97. doi: 10.1093/carcin/bgp036.

Muehlberg, Fabian L. *et al.* (2009) ‘Tissue-resident stem cells promote breast cancer growth and metastasis’, *Carcinogenesis*, 30(4), pp. 589–597. doi: 10.1093/carcin/bgp036.

Noble, D. (2002) ‘The rise of computational biology.’, *Nature reviews. Molecular cell biology*, 3(6), pp. 459–63. doi: 10.1038/nrm810.

Ohlsson, L. B. *et al.* (2003) ‘Mesenchymal progenitor cell-mediated inhibition of tumor growth in vivo and in vitro in gelatin matrix’, *Experimental and Molecular Pathology*, 75(3), pp. 248–255. doi: 10.1016/j.yexmp.2003.06.001.

Otsu, K. *et al.* (2009) ‘Concentration-dependent inhibition of angiogenesis by mesenchymal stem cells’, *Blood*, 113(18), pp. 4197–4205. doi: 10.1182/blood-2008-09-176198.

Pautke, C. *et al.* (2004) ‘Characterization of osteosarcoma cell lines MG-63, Saos-2 and U-2 OS in comparison to human osteoblasts’, *Anticancer Research*, pp. 3743–3748.

Peirce, C. S. (1903) ‘A Syllabus of Certain Topics of Logic’. Boston: Alfred Mudge & Son.

De Peppo, G. M. *et al.* (2010) ‘Human embryonic mesodermal progenitors highly resemble human mesenchymal stem cells and display high potential for tissue engineering applications’, *Tissue Engineering - Part A*, 16(7), pp. 2161–2182. doi: 10.1089/ten.tea.2009.0629.

Phadnis, S. M. *et al.* (2015) ‘Dynamic and social behaviors of human pluripotent stem cells’, *Scientific Reports*, 5(1), p. 14209. doi: 10.1038/srep14209.

Phillips, R. *et al.* (2013) *Physical Biology of the Cell*. Second Edi. Garland Science, Taylor & Francis Group.

Pogson, M. *et al.* (2006) ‘Formal agent-based modelling of intracellular chemical interactions.’, *Bio Systems*, 85(1), pp. 37–45. doi: 10.1016/j.biosystems.2006.02.004.

Prantl, L. *et al.* (2010) ‘Adipose tissue-derived stem cells promote prostate tumor growth.’, *The Prostate*, 70(15), pp. 1709–15. doi: 10.1002/pros.21206.

Purschke, M. *et al.* (2010) ‘Phototoxicity of Hoechst 33342 in time-lapse fluorescence microscopy’, *Photochemical & Photobiological Sciences*, 9(12), p. 1634. doi: 10.1039/c0pp00234h.

Qiao, L., Xu, Z.-L., *et al.* (2008) ‘Dkk-1 secreted by mesenchymal stem cells inhibits growth of breast cancer cells via depression of Wnt signalling.’, *Cancer letters*, 269(1), pp. 67–77. doi: 10.1016/j.canlet.2008.04.032.

Qiao, L., Xu, Z., *et al.* (2008) ‘Suppression of tumorigenesis by human mesenchymal stem cells in a hepatoma model.’, *Cell research*, 18(4), pp. 500–7. doi: 10.1038/cr.2008.40.

Rand, W. (2006) *Machine learning meets agent-based modeling: When not to go to a bar*. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0031938416312148>.

Read, M. *et al.* (2011) ‘Techniques for Grounding Agent-Based Simulations in the Real Domain : a case study in Experimental Autoimmune Encephalomyelitis’, *Mathematical and Computer Modelling of Dynamical Systems*, 00(00), pp. 1–19. doi: 10.1080/1387395YY.

Richmond, P. *et al.* (2010) ‘High performance cellular level agent-based simulation with FLAME for the GPU.’, *Briefings in bioinformatics*, 11(3), pp. 334–47. doi: 10.1093/bib/bbp073.

Robert A. Weinberg (2013) *The Biology of Cancer*. Second Edi. New York: Garland

Science, Taylor & Francis Group.

Rosenberg, S. a, Yang, J. C. and Restifo, N. P. (2004) ‘Cancer immunotherapy: moving beyond current vaccines.’, *Nature medicine*, 10(9), pp. 909–15. doi: 10.1038/nm1100.

Saltelli, A. *et al.* (2017) ‘A new sample-based algorithms to compute the total sensitivity index’, (March). Available at: <http://arxiv.org/abs/1703.05799>.

Sargent, R. G. (2004) ‘Validation and verification of simulation models’, *Proceedings of the 2004 Winter Simulation Conference*.

Sarpe, V. and Jacob, C. (2013) ‘Simulating the decentralized processes of the human immune system in a virtual anatomy model’, *BMC Bioinformatics*, 14(6).

Schindelin, J. *et al.* (2012) ‘Fiji: An open-source platform for biological-image analysis’, *Nature Methods*, 9(7), pp. 676–682. doi: 10.1038/nmeth.2019.

Secchiero, P. *et al.* (2010) ‘Human bone marrow mesenchymal stem cells display anti-cancer activity in SCID mice bearing disseminated non-Hodgkin’s lymphoma xenografts.’, *PloS one*, 5(6), p. e11140. doi: 10.1371/journal.pone.0011140.

Shah, K. (2013) ‘Encapsulated stem cells for cancer therapy.’, *Biomatter*, 3(1), pp. 1–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3732324&tool=pmcentrez&rendertype=abstract>.

Shinagawa, K. *et al.* (2010) ‘Mesenchymal stem cells enhance growth and metastasis of colon cancer.’, *International journal of cancer. Journal international du cancer*, 127(10), pp. 2323–33. doi: 10.1002/ijc.25440.

Shirinifard, A. *et al.* (2009) ‘3D multi-cell simulation of tumor growth and angiogenesis.’, *PloS one*, 4(10), p. e7190. doi: 10.1371/journal.pone.0007190.

Siclari, V. A. and Qin, L. (2010) ‘Targeting the osteosarcoma cancer stem cell’, *Journal of Orthopaedic Surgery and Research*, 5(1), pp. 1–10. doi: 10.1186/1749-799X-5-78.

Sobol, I. M. (2001) ‘Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates’, *Mathematics and Computers in Simulation*, 55(1–3), pp. 271–280. doi: 10.1016/S0378-4754(00)00270-6.

Srblijinovi, A. and Skunca, O. (2003) ‘AN INTRODUCTION TO AGENT BASED MODELLING AND’, *Interdisciplinary Description of Complex Systems*, 1, pp. 1–8.

Studený, M. *et al.* (2004) ‘Mesenchymal stem cells: potential precursors for tumor stroma and targeted-delivery vehicles for anticancer agents.’, *Journal of the National Cancer*

Institute, 96(21), pp. 1593–603. doi: 10.1093/jnci/djh299.

Sun, T. *et al.* (2007) ‘An integrated systems biology approach to understanding the rules of keratinocyte colony formation.’, *Journal of the Royal Society, Interface*, 4(17), pp. 1077–92. doi: 10.1098/rsif.2007.0227.

Sun, T. *et al.* (2008) ‘Agent based modelling helps in understanding the rules by which fibroblasts support keratinocyte colony formation.’, *PloS one*, 3(5), p. e2129. doi: 10.1371/journal.pone.0002129.

Swat, M. *et al.* (2013) ‘Multi-Scale Modeling of Tissues Using CompuCell3D’, *National*, (1), pp. 325–366. doi: 10.1016/B978-0-12-388403-9.00013-8.Multi-Scale.

Takahashi, K. *et al.* (2007) ‘Induction of pluripotent stem cells from adult human fibroblasts by defined factors.’, *Cell*, 131(5), pp. 861–72. doi: 10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) ‘Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.’, *Cell*, 126(4), pp. 663–76. doi: 10.1016/j.cell.2006.07.024.

Thomson, J. a. (1998) ‘Embryonic Stem Cell Lines Derived from Human Blastocysts’, *Science*, 282(5391), pp. 1145–1147. doi: 10.1126/science.282.5391.1145.

Thomson, J. A. (1998) ‘Embryonic stem cell lines derived from human blastocysts’, *Science*, 282(5391), pp. 1145–1147. doi: 10.1126/science.282.5391.1145.

Thorne, B. C., Bailey, A. M. and Peirce, S. M. (2007) ‘Combining experiments with multi-cell agent-based modeling to study biological tissue patterning.’, *Briefings in bioinformatics*, 8(4), pp. 245–57. doi: 10.1093/bib/bbm024.

Tinevez, J. Y. *et al.* (2017) ‘TrackMate: An open and extensible platform for single-particle tracking’, *Methods*. The Author(s), 115, pp. 80–90. doi: 10.1016/j.ymeth.2016.09.016.

Towns, C. R. and Jones, D. G. (2004) ‘Stem cells, embryos, and the environment: A context for both science and ethics’, *Journal of Medical Ethics*, 30(4), pp. 410–413. doi: 10.1136/jme.2003.002386.

Underwood, J. C. E. (2004) *General and Systematic Pathology*. Fourth Edi. London: Churchill Livingstone.

Uzbekov, R. E. (2004) ‘Review: Analysis of the cell cycle and a method employing synchronized cells for study of protein expression at various stages of the cell cycle’, *Biokhimiya*, 69(5), pp. 597–611. doi: 10.1023/b:biry.0000029845.11184.30.

Vargha, A. and Delaney, H. D. (2000) ‘A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong’, 25(2), pp. 101–132.

Viceconti, M. (2011) ‘A tentative taxonomy for predictive models in relation to their falsifiability.’, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 369(1954), pp. 4149–61. doi: 10.1098/rsta.2011.0227.

Wadkin, L. E. *et al.* (2017) ‘Dynamics of single human embryonic stem cells and their pairs: A quantitative analysis’, *Scientific Reports*. Springer US, 7(1), pp. 1–12. doi: 10.1038/s41598-017-00648-0.

Wadkin, L. E. *et al.* (2018) ‘Correlated random walks of human embryonic stem cells in vitro’, *Physical Biology*, 15(5), pp. 1–19. doi: 10.1088/1478-3975/aac008.

Waldman, A. D., Fritz, J. M. and Lenardo, M. J. (2020) ‘A guide to cancer immunotherapy: from T cell basic science to clinical practice’, *Nature Reviews Immunology*. Springer US, 20(11), pp. 651–668. doi: 10.1038/s41577-020-0306-5.

Walker, D. *et al.* (2006) ‘An integrated agent-mathematical model of the effect of intercellular signalling via the epidermal growth factor receptor on cell proliferation’, *Journal of Theoretical Biology*, 242(3), pp. 774–789. doi: 10.1016/j.jtbi.2006.04.020.

Walker, D. C. *et al.* (2004) ‘The epitheliome: agent-based modelling of the social behaviour of cells.’, *Bio Systems*, 76(1–3), pp. 89–100. doi: 10.1016/j.biosystems.2004.05.025.

Walker, D. C. and Southgate, J. (2009) ‘The virtual cell--a candidate co-ordinator for “middle-out” modelling of biological systems.’, *Briefings in bioinformatics*, 10(4), pp. 450–61. doi: 10.1093/bib/bbp010.

Wei, J. *et al.* (2004) ‘Embryonic endothelial progenitor cells armed with a suicide gene target hypoxic lung metastases after intravenous delivery.’, *Cancer cell*, 5(5), pp. 477–88. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15144955>.

WHO (World Health Organisation) (2020) *Cancer Prevention and Control*. Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer>.

Xiang, X. *et al.* (2005) ‘Verification and Validation of Agent-based Scientific Simulation Models’, in *Agent-Directed Simulation Symposium*, pp. 47–55.

Yu, J. *et al.* (2008) ‘Mesenchymal Stem Cells Derived from Human Adipose Tissues Favor Tumor Cell Growth in vivo.’, *Stem Cells and Development*, 17(3), pp. 463–474. doi: 10.1089/scd.2007.0181.

Yu, J. M. *et al.* (2008) ‘Mesenchymal stem cells derived from human adipose tissue favor tumour cell growth in vivo’, *Stem Cells and Development*, 17(3), pp. 4463–474. doi: 10.1089/scd.2007.0181.

Zahedmanesh, H. and Lally, C. (2012) ‘A multiscale mechanobiological modelling

framework using agent-based models and finite element analysis: Application to vascular tissue engineering', *Biomechanics and Modeling in Mechanobiology*, 11(3–4), pp. 363–377. doi: 10.1007/s10237-011-0316-0.

Zhang, L., Wang, Z., *et al.* (2009) 'Multiscale agent-based cancer modeling.', *Journal of mathematical biology*, 58(4–5), pp. 545–59. doi: 10.1007/s00285-008-0211-1.

Zhang, L., Strouthos, C. G., *et al.* (2009) 'Simulating Brain Tumor Heterogeneity with a Multiscale Agent-Based Model: Linking Molecular Signatures, Phenotypes and Expansion Rate.', *Mathematical and computer modelling*. Elsevier Ltd, 49(1–2), pp. 307–319. doi: 10.1016/j.mcm.2008.05.011.

Zhang, L., Athale, C. A. and Deisboeck, T. S. (2007) 'Development of a three-dimensional multiscale agent-based tumor model: Simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer', *Journal of Theoretical Biology*, 244, pp. 96–107. doi: 10.1016/j.jtbi.2006.06.034.

Zhang, Y. *et al.* (2011) 'Computer Simulations of Cell Sorting Due to Differential Adhesion', *PLoS One*, 6(10). doi: 10.1371/journal.pone.0024999.

Zhu, W. *et al.* (2006) 'Mesenchymal stem cells derived from bone marrow favor tumor cell growth in vivo.', *Experimental and molecular pathology*, 80(3), pp. 267–74. doi: 10.1016/j.yexmp.2005.07.004.

Zhu, Y. *et al.* (2009) 'Human mesenchymal stem cells inhibit cancer cell proliferation by secreting DKK-1.', *Leukemia*, 23(5), pp. 925–33. doi: 10.1038/leu.2008.384.