Georgia State University
ScholarWorks @ Georgia State University

Mathematics Dissertations | Department of Mathematics and Statistics

12-15-2016

Statistical Models and Analysis of Growth Processes in Biological Tissue

Jun Xia

Follow this and additional works at: https://scholarworks.gsu.edu/math_diss

Recommended Citation
Xia, Jun, "Statistical Models and Analysis of Growth Processes in Biological Tissue." Dissertation, Georgia State University, 2016.
https://scholarworks.gsu.edu/math_diss/39

This Dissertation is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

STATISTICAL MODELS AND ANALYSIS OF GROWTH PROCESSES IN BIOLOGICAL TISSUE

by

JUN XIA

Under the Direction of Remus Osan, PhD

ABSTRACT

The mechanisms that control growth processes in biology tissues have attracted continuous research interest despite their complexity. With the emergence of big data experimental approaches there is an urgent need to develop statistical and computational models to fit the experimental data and that can be used to make predictions to guide future research. In this work we apply statistical methods on growth process of different biological tissues, focusing on development of neuron dendrites and tumor cells.

We first examine the neuron cell growth process, which has implications in neural tissue regenerations, by using a computational model with uniform branching probability and a maximum overall length constraint. One crucial outcome is that we can relate the parameter fits from our model to real data from our experimental collaborators, in order to examine the usefulness of our model under different biological conditions. Our methods can now directly compare branching probabilities of different experimental conditions and provide confidence intervals for these population-level measures. In addition, we have obtained analytical results that show that

the underlying probability distribution for this process follows a geometrical progression increase at nearby distances and an approximately geometrical series decrease for far away regions, which can be used to estimate the spatial location of the maximum of the probability distribution. This result is important, since we would expect maximum number of dendrites in this region; this estimate is related to the probability of success for finding a neural target at that distance during a blind search.

We then examined tumor growth processes which have similar evolutional evolution in the sense that they have an initial rapid growth that eventually becomes limited by the resource constraint. For the tumor cells evolution, we found an exponential growth model best describes the experimental data, based on the accuracy and robustness of models. Furthermore, we incorporated this growth rate model into logistic regression models that predict the growth rate of each patient with biomarkers; this formulation can be very useful for clinical trials. Overall, this study aimed to assess the molecular and clinic pathological determinants of breast cancer (BC) growth rate in vivo.

INDEX WORDS: Neuronal Tree, Stochastic Models, Maximum Length Constraint, Breast Cancer, Tumor Size Growth Rate, Survival Analysis, Kaplan-Meier Estimate, Cox Proportional Hazard Ratio, Model Selection

STATISTICAL MODELS AND ANALYSIS OF GROWTH PROCESSES IN BIOLOGICAL TISSUE

by

JUN XIA

Under the Direction of Remus Osan, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016

Copyright by

Jun Xia

2016

STATISTICAL MODELS AND ANALYSIS OF GROWTH PROCESSES IN BIOLOGICAL TISSUE

by

JUN XIA

Committee Chair: Remus Osan

Committee: Yichuan Zhao

Xin Qi

Gennady Cymbalyuk

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2016

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my mentor Professor Remus Osan, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his guidance and persistent help, this dissertation would not have been possible.

I would like to thank my committee members, Professor Yichuan Zhao, Professor Xin Qi and Professor Gennady Cymbalyuk. Their insights and comments guided me throughout the whole preparation for the defense. What is more, I took many classes with them and learnt a lot from them, which is so valuable for my future career. In addition, I extend a special thank you to Professor Gengsheng Qin, who brought me to Georgia State University. Also, I want to thank Professor Yi Jiang, Professor Derby Charles, Professor Guantao Chen, Dr. Manfred Schmidt, Professor Igor Belykh, and all other professors, who helped me during the past five years.

I also want to thank all my fellows in the Mathematics and Statistics department, especially Jie Zhang with whom I collaborated on some aspects of this dissertation. What is more, I would like to thank all my friends in the US and also back in China, for the memories we had during the past five years.

I dedicate my dissertation to my parent and Xiaolei (Lily) Zhang, who have been supported me during the past five years and will keep supporting me in the future!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **iv**

TABLE OF CONTENTS ..... **v**

LIST OF TABLES ..... **ix**

LIST OF FIGURES ..... **xi**

1 INTRODUCTION **1**

1.1 Motivation for Studying Neural Growth **1**

1.1.1 Previous studies ***1***

1.1.2 Modeling neural dendritic tree growth using stochastic processes ***5***

1.1.3 Our study on the discrete neural growth model under limited resource condition 8

1.2 Motivation for Tumor Growth Project **9**

1.2.1 Previous study ***9***

1.2.2 Models for tumor growth process ***11***

1.2.3 Application of growth rate models in the survival analysis ***12***

1.2.4 Using survival analysis to analyze breast cancer data ***13***

1.2.5 Fit Logistic regression model to predict growth rate with biomarkers ***14***

1.2.6 Extension of the survival analysis to recurrence of breast cancer ***15***

2 METHODS **17**

2.1 Neuronal Culture and Analysis **17**

2.2 Discrete statistical model for neural growth without restriction 18

2.2.1 Recurrence formulas for the probability distribution of all possible neural trees 18

2.2.2 Derivation of the recurrence formula for all possible neuronal trees 20

2.2.3 Derivation for the expected number at distance k: $Pn = 1 + pn$ 21

2.3 Neural Growth Model with Restrictions 23

2.3.1 Optimal targeting region for trees with size restrictions 23

2.3.2 Right tail series expansion up to order $p3$ 23

2.3.3 Determining the approximation for the peak location 25

2.4 Tumor Growth Models 27

2.4.1 Materials and methods for predicting tumor growth rate 27

2.4.2 Model selection for growth rate 28

2.4.3 Biomarkers in the model 30

2.4.4 Statistical analysis on the growth rate 31

2.4.5 Choose the cut point for growth rate 32

2.4.6 Fitting the logistic regression model to predict growth rate 33

2.5 Extending survival analysis on the recurrence of breast cancer 34

2.5.1 Cohort study 34

2.5.2 Follow up 35

2.5.3 Statistical Analysis 35

3 RESULTS 37

3.1 Discrete Model 37

3.1.1 Unrestricted neural growth model 37

3.1.2 Discrete probability distribution for the evolving family of trees 39

3.2 Model with conditional constraint 41

3.2.1 Neuronal growth model with a length constraint 41

3.2.2 Long-range expected performances are also characterized by geometrical series 44

3.2.3 Approximating the location of optimal targeting performances 47

3.3 Comparison with lab data 49

3.3.1 Comparison of the theoretical and experimental results for neural growth 49

3.4 Results for Tumor growth rate 55

3.4.1 Tumor volume and growth rate: 55

3.4.2 Results from neural growth model selection 58

3.4.3 Correlation study on growth rate and biomarkers 66

3.4.4 Results from survival analysis on the tumor growth data 74

3.4.5 Results from optimize the cutpoint 75

3.4.6 Results from logistic regression modeling on the growth rate 75

3.5 Results from extending the survival analysis on the breast cancer recurrence 77

3.5.1 Clinico-pathological characteristics of patients 77

3.5.2 Recurrence pattern among racially distinct patients 83

3.5.3 Recurrence patterns among racially distinct patients following each form of treatment 85

3.5.4 Recurrence rates among racially distinct breast cancer patients in different stages 88

3.5.5 Survival outcomes among racially distinct patients displaying recurrence 91

4 CONCLUSIONS 93

4.1 Conclusion part for neural growth model 93

4.2 Conclusion part for tumor growth model 95

4.3 Conclusion from extending survival analysis on breast cancer recurrence 98

REFERENCES 101

LIST OF TABLES

Table 1 List of the trees that can be generated when maximum length of the tree is equal to 7. . 23

Table 2 Coefficients of first 3 orders from the proportions between steps of an example with 26 steps 24

Table 3 Average number of active branches and the associated variances for the first two time steps 39

Table 4 Expected mean and variance for the active (terminal) branches after N timesteps 40

Table 5 Approximate geometrical series for the right tail of the probability distribution. 46

Table 6 ANOVA results from estimating branching probability 52

Table 7 Comparison between branching probabilities of two processes 54

Table 8 Comparison between branching probabilities of two processes 54

Table 9 Clinicopathological features of cases 57

Table 10 ANOVA results from linear model 58

Table 11 ANOVA results from exponential neural growth model 61

Table 12 ANOVA results from Gompertz neural growth model 64

Table 13 Relationship between growth rate and biomarkers 69

Table 14 Relation between Ki67 and Caspase3 expression 73

Table 15 Survival analysis for monthly growth rate 74

Table 16 Survival analysis with adjusting age, Ki67, and grade 74

Table 17 NH demographics and breast cancer clinico-pathological compared between patients with or without tumor recurrence 79

Table 18 Broad spectrum recurrence patterns among racially distinct populations 84

Table 19 Recurrence rates and patterns after receiving any form of treatment among racially distinct populations 85

Table 20 Overall recurrence rates among racially distinct staged breast cancer patients..... 89

LIST OF FIGURES

Figure 1 Overexpression of vypin from DIV 6-10 increases proximal branching and total dendrite number but decreases average dendrite length 3

Figure 2 Three different scenarios during the neural tree regeneration [14]. 4

Figure 3 Summary figure for the evolution of the trees for the first three time steps. 19

Figure 4 Ellipsoid Tumor 28

Figure 5 Histogram of individual growth rate from three models 29

Figure 6 Diagram of all possible tree instantiations after 3 time steps. 38

Figure 7 All possible trees of lengths below a maximum value of 7. 41

Figure 8 The expected number of branches as a function of distance for families of trees with maximum length 7 and branching probability p = 0.5. 42

Figure 9 Sholl plots for different branching probabilities p. 44

Figure 10 Illustration of trees generated at very long distances. 45

Figure 11 Comparison of naïve vs exact results for a tree of maximum length L = 31. 48

Figure 12 Comparison of experimental and theoretical results. 50

Figure 13 Results from estimating the branching probability 52

Figure 14 Comparison between NOS vs GFP 53

Figure 15 Comparison between GFP and Cypin 54

Figure 16 Histogram of tumor volume at diagnosis and screening and also histogram of the change in tumor size 56

Figure 17 Fit plot of linear neural growth rate model 59

Figure 18 Diagnostic results of linear neural growth model 60

Figure 19 Fit plot for exponential neural growth rate model 62

Figure 20 Diagnostic results of exponential neural growth model 63

Figure 21 Fit plot of Gompertz neural growth model 65
Figure 22 Diagnostic results of Gompertz neural growth model 66
Figure 23 Significant correlation between tumor growth rate with (A) Ki67; (B) Caspase3; (C) Grade; (D) NPI and (E) Mitosis 68
Figure 24 Relationship between Ki67 and Caspase-3 expression 69
Figure 25 Comparison of growth rate among different type of Breast Cancer 73
Figure 26 Survival Curve for different groups 75
Figure 27 NH demographics and breast cancer clinico-pathological characteristics compared between patients with or without tumor recurrence 78
Figure 28 AA exhibit lower survival duration than EA among recurrent breast cancer patients . 92

1 INTRODUCTION

1.1 Motivation for Studying Neural Growth

Injuries to the central nervous system often result in extreme functional deficits because of physical damage to neurons in the brain and/or spinal cord. Currently, there is no neuroengineering solution to either traumatic brain injury (TBI) or spinal cord injury (SCI) despite intense research in both of these areas, likely because the central nervous system has a limited capacity to regenerate itself [1]. Especially in the mammalian central nervous system, the regeneration ability of axons is quite limited after injury. Thus, the axonal damage will lead to functional deficits, such as paralysis, Alzheimer and so on. Therefore, in the recovery process after central nervous system injury, axonal regeneration plays a key role and influences the speed of recovery. In this study, we model neuronal growth with the aim that the insight gained would help develop therapies aimed at fostering connection regrowth.

With the subject of regeneration in mind, we assumed that the neural tree is formed as a result of a branching process, which is subjected to limited resources, such as maximum total neural tree length [2]. Since traditional branching processes do not have such limitations, making it easy to obtain the mean and variance along with each step, we aim to extend that formula to the practical process with the limitations mentioned above. We believe that the regeneration process with limited resources can be applied to so many other fields, which have the similar scenarios; such examples can be seen in the marketplace, nuclear, chemistry and so on. Usually, for these regeneration processes, we are interested in the size of the whole tree for each step, under limited resources.

1.1.1 Previous studies

Much effort has been devoted to understanding the mechanisms that govern dendritic outgrowth in neurons because proper dendrite morphology is essential to neuronal communication. The shape of the dendritic arbor determines which neurons receive and process information, thereby affecting postsynaptic activity and the way neurons are integrated into the network [2]. During disease or after trauma [3], dendritic morphology is often negatively affected, resulting in either too few or too many dendrites. The structure of the network is then disrupted,

which can result in network dysfunction and cognitive deficits. Moreover, it has been shown that dendrites are not merely passive structures that deliver information to the soma. Rather, dendrites have the ability to perform computations, influence synaptic output, and regulate protein translation [4-6]. These studies illustrate the importance of a properly formed dendritic arbor, which not only receives and conducts information but also influences the nature of the information.

During neuronal development, dendrite outgrowth is influenced by the simultaneous ingrowth of axons as well as the formation of synapses. Similar to axonal growth, dendritic growth is characterized by the appearance of filopodia that may develop into growth cones and lead to mature branches [5]. Importantly, dendrites are highly dynamic and undergo both growth and retraction during development, and most filopodia do not develop into dendrites. Early in development, dendrite outgrowth is encouraged by the presence of synaptic activity, and then later in development, synaptic activity stabilizes dendrites [5, 7]. While fewer cues are present during two-dimensional *in vitro* development, neurons proceed through similar steps, in which filopodia-like structures develop into immature neurites. One of these neurites becomes the axon, and the rest may become dendrites if they continue to grow [8,9]. Similar to *in vivo* development, neurons that develop *in vitro* are also highly dynamic. This similarity of maturation processes allows us to study the effects of extrinsic and intrinsic regulators of branching, and researchers have identified many of these factors that modulate dendrite number *in-vitro* and *in vivo* [10]. Being able to study neuronal developments *in vitro* has allowed us and others to identify the protein-protein interactions and cellular processes necessary for neuronal development. For example, the Firestein laboratory has identified cytosolic PSD-95 interactor (cypin) as an intrinsic regulator of dendrite branching [6]. They have found that cypin significantly increases primary and secondary branching by promoting microtubule polymerization (figure 1). Additionally, they recently found that brain-derived neurotrophic factor (BDNF), a very well-known extrinsic growth factor, increases proximal dendrite branching through transcriptional regulation of cypin [6]. Many other factors have also been shown to modulate dendrite branching in specific ways; studying neuronal development *in vitro* is allowing scientists to determine which factors may be useful as therapeutic strategies for neurodegenerative diseases and CNS injury.

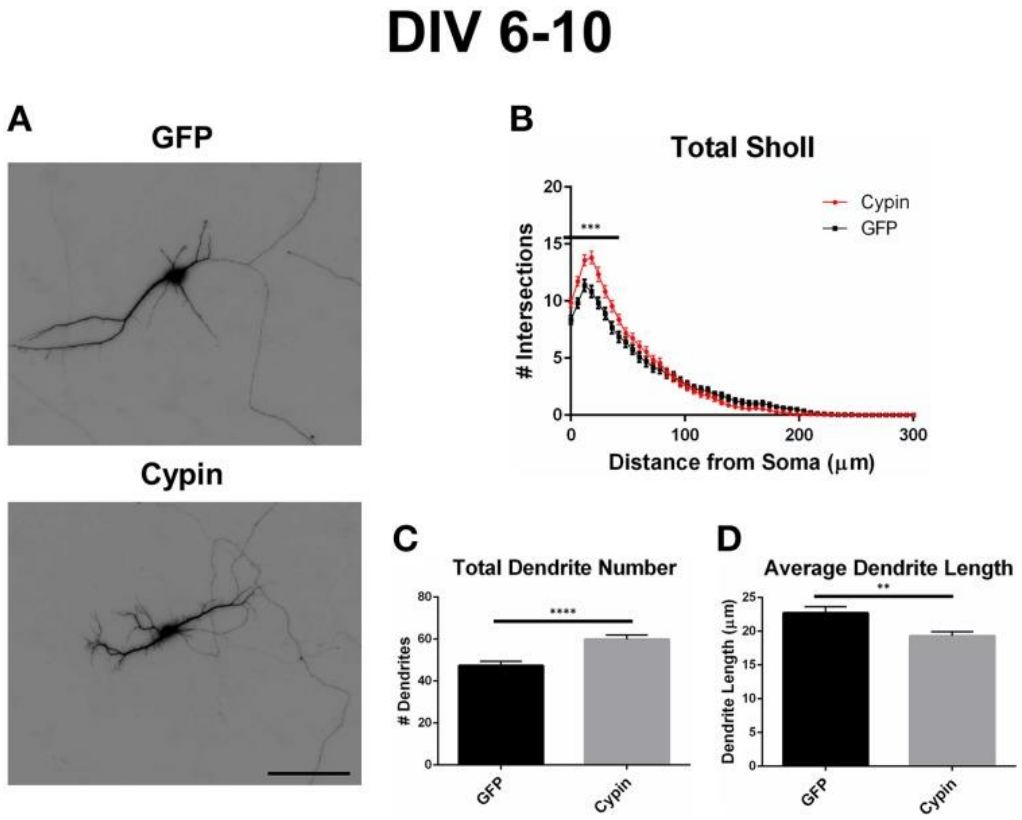


Figure 1 *Overexpression of vypin from DIV 6-10 increases proximal branching and total dendrite number but decreases average dendrite length*

(A) Representative images of hippocampal neurons overexpressing GFP or GFP-cypin (cypin) from DIV 6–10. Scale bar = 100 µm. (B) Sholl analysis of all orders of branches (Total Sholl) shows that overexpression of cypin significantly increases dendrite branching at 0–42 µm from the cell body (***p < 0.001). Statistics were calculated using Two-Way ANOVA followed by Bonferroni multiple comparisons test.(C) Overexpression of cypin results in a significant increase in the total number of dendrites (****p < 0.0001). A total number of dendrites represents the sum of all dendrites, regardless to what category they belong. Statistics were calculated by unpaired, two-tailed Student's t-test. (D) Overexpression of cypin results in a significant decrease in the average length of dendrites (**p < 0.001). The average length is the mean length of all dendrites, regardless to what category they belong. Statistics were calculated by unpaired, two-tailed Student's t-test with Welch's correction. Error bars indicate SEM. n = 50 neurons for GFP, and n = 55 neurons for cypin. Figure 1 was reproduced from [1] with the permission from the authors.

To assess the overall shape of the dendritic arbor, several tools have been developed, the most widely known being Sholl analysis (figure 2) [11]. This method counts dendrites by drawing concentric circles at fixed distances from the cell body and counting the number of intersections

at each distance. When performed manually, the process is time-consuming and prone to errors. To this end, several groups have developed automated or semi-automated Sholl analysis software packages to increase the accuracy of results and decrease analysis time. Sholl is an invaluable tool in the field of neuroscience because it allows for the comparison of branch number at specific distances from the soma, which provides more information than total branch number alone. However, statistical analyses such as two-way analysis of variance only take into account only differences at specific distances, not between the different probability distributions. For example, a Sholl curve shifted to the left or the right from the control curve will show statistical significance at several locations, but no single parameter indicates this specific shift. With our model, we were also able to discriminate between different Sholl curves of control neurons versus treated neurons.

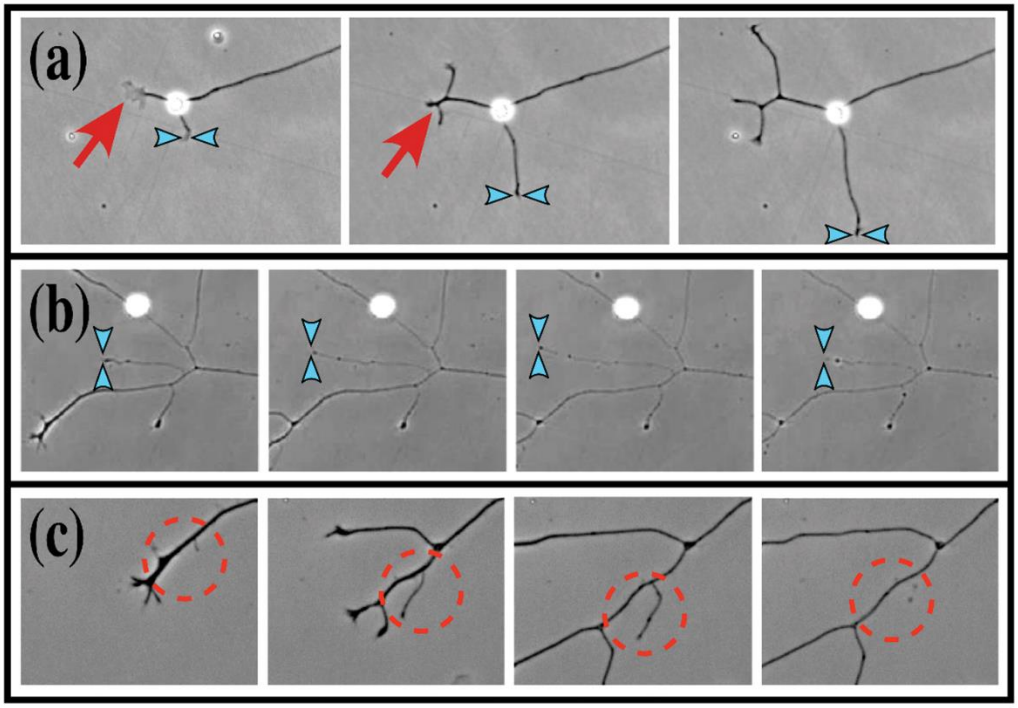


Figure 2 *Three different scenarios during the neural tree regeneration [14].*

Figure 2 Time sequences showing branching and pruning of dissociated E11 chick dorsal root ganglion neurites. (a) Branching (red arrow) and extension (blue arrowheads) of primary axons. (b) Extension and retraction (blue arrowheads) of neurite tip. (c) Tertiary branching and pruning (encircled). Cultures are grown in the presence of glia in 5% CO2/ 37uC on Poly-L-lysine/laminin in N3 complete serum-free media. Phase- contrasts l i v e imaging at 28 hrs post-plating. The time interval between acquisitions for each time series is as follows: (a) 30 mins, (b) 75 mins, (c) 75

mins. Snapshots are contrast enhanced for visual clarity of the neurites. Figure 2 was reproduced with permissions from authors.

While the arborization patterns of neurons have received extensive attention in experiments, it is not clear what the rules are that govern the final shape of neurons despite multiple computational models that were proposed [2, 4, 6,12]. Some of the simplest models assume that outgrowth is a result of a stochastic process, where the neuron starts with an initial branch that generates additional branches uniformly at the periphery. Osan et al. ([2]) examined these growth rules in a computational model [2] in which a blind search for targets was performed. To accomplish this, they cultured neurons under conditions mimicking SCI, then compared the results with our model. In order to derive analytical expectations for these computations, they also introduced the use of a probabilistic realization of all possible trees to compute the average length of no-longer-evolving branches. In this current study, we extend these results to determine the probabilistic performance profiles for all trees of a given maximum length. Importantly, we then show that our model is consistent with experimental data gathered from hippocampal neurons cultured *in vitro*.

1.1.2 Modeling neural dendritic tree growth using stochastic processes

Branching process is a Markov process that models the regeneration of population under the assumption that each individual in the generation k can produce a random number of individuals for next generation k+1. In our case, each branch of the neural tree can generate a random number of next level of branches.

In our study, we first made some naïve assumptions, for example, the probability for each individual to produce next generation is independent of others. The probability mass function for offspring is often called the offspring distribution and is given by

$$p_i = P(number\ of\ offspring = i),$$

$$for\ i = 0, 1, 2, \ldots$$

We first assume that each individual in our model can only generate no more than 2 individuals

for next generation, which means that:

$$p_i = 0 \qquad for\ i > 2$$

This model was introduced by F. Galton, in the late 1800s, to study the disappearance of family names; in this case p_i is the probability that a man has i sons. We will start with a single individual in generation 0, and generate the resulting random family tree. This tree is either finite (when some generation produces no offspring at all) or infinite — in the former case, we say that the branching process dies out, and in the latter that it survives.

Proposition 1: Any recurrent class is a closed subset of states. [13]

This process is equivalent to a Markov chain, where X_n is the number of individuals at generation n. Let's start with the following observations.

- If X_n reaches 0, it stays there, therefore 0 is an absorbing state.
- If $p_0 > 0$, $P(X_{n+1} = 0|X_n = k) > 0$ for all k, which means
- Therefore, by Proposition 1, all states other than 0 are transient if $p_0 > 0$; the population must either die out or increase to infinity. If $p_0 = 0$, then the population cannot decrease, and increases each generation with probability at least $1 - p_1$, therefore must increase to infinity.

It is possible to write down the transition probabilities for this chain, but they have an explicit form, as

$$P(X_{n+1} = i|X_n = k) = P(W_1 + W_2 + \ldots + W_k = i)$$

where $W_1, \ldots, W_k$ are independent random variables, each with the offspring distribution. Recall that we are assuming that $X_0 = 1$.

Let

$$\delta_n = P(X_n = 0)$$

be the probability that the population is extinct by generation (which we also think of as time) n. The probability π0 that the branching process dies out is then the limit of these probabilities:

$$\pi_0 = P(the\ process\ dies\ out) = P(Xn = 0\ for\ some\ n) = \lim_{n\to\infty} P(X_n = 0) = \lim_{n\to\infty} \delta_n$$

Note that $\pi_0 = 0$ if $p_0 = 0$. Our main task will be to compute π_0 for general probabilities p_k. We start, however, with computing expectation and variance of the population at generation n.

Let μ and σ^2 be the expectation and variance of the offspring distribution, that is,

$$\mu = EX_n = \sum_{k=0}^{\infty} kp_k$$

and

$$\sigma^2 = Var(X_n).$$

Theorem 1: Expectation and variance of sums with a random number of terms: Assume that $X_1, X_2, \ldots$ is an $i.i.d$ sequence of random variables with finite $E(X) = \mu$ and $Var(X) = \sigma^2$. Let N be a nonnegative integer random variable, independent of all X_i, and let $S = \sum_{i=1}^{N} X_i$

Then

$$E(S) = \mu E(N)$$

$$Var(S) = \sigma^2 E(N) + \mu^2 Var(N)$$

Let $m_n = E(X_n)$ and $v_n = Var(X_n)$. Now, X_{n+1} is the sum of a random number, which equalsX_n, of independent random variables, each with the offspring distribution. Thus, we have, by Theorem 1 above,

$$m_{n+1} = m_n \mu,$$

and

$$v_{n+1} = m_n \sigma^2 + v_n \mu^2 .$$

Together with initial conditions $m_0 = 1, v_0 = 0$, the two recursive equations determine m_n and v_n. We can very quickly solve the first recursion to get $m_n = \mu^n$, and consequently we obtain:

$$v_{n+1} = \mu^n \sigma^2 + v_n \mu^2 .$$

This recursion has a general solution of the form $v_n = A\mu^n + B\mu^{2n}$. The constant A must satisfy

$$A\mu^{n+1} = \sigma^2\mu^n + A\mu^{n+2},$$

so that, when $\mu \neq 1$,

$$A = \frac{\sigma^2}{\mu(1-\mu)}.$$

From $v_0 = 0$, we get $A + B = 0$ and the solution is given in the next theorem.

So, we have the Theorem1.1.1: Expected m_n and variance v_n of the n_{th} generation count

We have

$$\begin{cases} m_n = \mu^n \\ and \\ v_n = \begin{cases} \dfrac{\sigma^2}{\mu(1-\mu)}\mu^n - \dfrac{\sigma^2}{\mu(1-\mu)}\mu^{2n} & if\ \mu \neq 1 \\ n\sigma^2 & otherwise \end{cases} \end{cases}$$

1.1.3 Our study on the discrete neural growth model under limited resource condition

The structure of neuronal trees is essential for individual cell and network function, yet neuronal synaptic and membrane dynamics have received significantly more attention in the experimental and computational neuroscience community. This is in part due to the fact that our understanding of experimental data has traditionally focused on descriptive approaches of the network structure, and conversely, the corresponding mathematical models are much less developed.

In the current study, we extend our work to hippocampal neurons that develop normally. We study these neurons as they are often damaged as a result of traumatic brain injury and are also subject to alteration due to neurodegeneration, such as Alzheimer's Disease, or neurocognitive disorders, such as schizophrenia. Our results can be used to determine the probabilistic

performance profiles for all trees of a given maximum length. Importantly, we then show that our model is consistent with experimental data gathered from primary cultures of rat hippocampal neurons. More precisely, the Sholl curves predicted by our model exhibit the same trends seen in the experiments: a region of geometric growth in the nearby regions, which soon slows down to produce a peak in expected number of branches that is followed by a steady approximately geometric decrease at long distances. Furthermore, our model allows us to generate accurate fits of the experimental data using a small number of parameters, with the uniform branching probability playing a crucial role. Based on our results, we now have a baseline for dendritic growth and can compare this model to ones in which neurons are damaged due to trauma or disease, in order to work toward predicting which intrinsic or extrinsic factors can be targeted therapeutically to treat patients. Our goal is to compare features such as the branching probabilities under different conditions and generate statistical tests for these comparisons in order to determine statistical significance. Our future work will examine the effects induced by uniform rate pruning as well as parameter-based branching and pruning.

1.2 Motivation for Tumor Growth Project

1.2.1 Previous study

Breast cancer is the heterogeneous disease with different clinicopathological features, recurrence patterns, and survival. [14] The major molecular subtypes: estrogen receptor (ER) positive, human epidermal growth factor receptor (HER2) positive and Triple Negative are widely used to predict prognosis and response to treatment in breast cancer patients [15]. Risk factors for breast cancer range from excessive use of exogenous hormones to age at menarche and age at menopause. [16] Family history of breast cancer is also considered a highly important risk factor while the non-reproductive factors include age, high BMI, the excessive alcohol intake, sedentary lifestyle, poor diet and exposure to medical radiation. [16-18]

Breast cancer has highly variable rates of growth. Contemporary thoughts assume that the gross tumor growth rate decelerates with increasing tumor mass. The study of tumor growth rates is necessary to understand the biology and natural history of the malignant diseases[19]. In developed countries, mammography screening is becoming an established part of health services.

However, there is an ongoing discussion related to optimization of the mammography screening, including the determination of optimal time intervals between screening and which age group to initiate. Therefore, the knowledge of tumor growth is important in planning and evaluation of screening programs. [20] Despite these facts, Mammography has a number of limitations wherein it yields false-positive results, leading to an increased number of breast biopsies especially among young patient. [21] Breast cancers that are detected in the interval after a negative mammographic result are the interval cancer. [22] It is now a known fact that breast cancer screening has been proven to save lives and helps in better treatment regimens wherein Mammography remains the mainstay of screening[23]. But In US, the overall progress in reducing the breast cancer-related mortality following screening has not been reduced instead the probability of death is snoring. [24]

The tumor growth rate has not been implicated as a prognostic variable in clinical practice because of its difficulty in evaluating it in the short interval of diagnosis and treatment. However, Yoo demonstrated that there is no any association of tumor growth rate with patient's survival. [25] Also, in the study carried out by Tubiana [26], there was no any survival difference between the subgroups of patients with rapid or intermediate growth rate after the follow-up exceeding 8 years. [26]

In vivo, breast cancer growth rate is carefully regulated. There is a precise balance between growth fraction – the proportion of tumor cells that are proliferating – rate of tumor cell loss by apoptosis, (and necrosis) and cell doubling time. [27, 28] Heterogeneities in growth rate between individual tumor cells, the degree of angiogenesis, [29] and interactions between the tumor and surrounding stromal cells also play a part. [30] The growth rate holds vital importance in prognostication, as it been shown that tumors with a faster growth rate in vivo, are more aggressive and thus offer a worse prognosis. [31, 32]

Essentially, tumor growth progression is characterized by a net increase in the number of tumor cells, which could be due to increased proliferation and/or decreased apoptosis or both. Although a lot of such correlative studies are available in cultured cells, no study has yet systematically evaluated proliferation and apoptosis during in vivo progression of tumor growth. From a clinical

standpoint, several proliferation markers have demonstrated utility in the clinic for prognostication. The study aims to identify correlations between clinic-pathological parameters, biomarker expression and in vivo tumor growth rate, determined by the change in tumor size between sequential mammograms. Tumor growth rate may be a significant consideration when subtyping breast carcinomas, and consequently could be important in determining patient outcome. The study is expected to improve our understanding of breast cancer biology and growth determinates but also will provide prognostic information and evidence-based data that can be used in the medicolegal practice.

1.2.2 Models for tumor growth process

There are so many continuous models for the growth rate that can be used to model tumor growth [33-36]. The simple list one is linear growth model:

$$V(t = t_1) = V_0 + \alpha(t_1 - t_0)$$

Here V_0 is the initial tumor size, t_0 and t_1 are the time of diagnostic and time of revisit, α is the tumor growth rate. And we have the result for growth rate is:

$$\alpha = \frac{V_{t_1} - V_{t_0}}{t_1 - t_0}$$

We also tried the second model, which is Exponential Linear Model [38]:

$$\begin{cases} \dfrac{dV}{dt} = a_0 V, & t \le \tau \\ \dfrac{dV}{dt} = a_1, & t > \tau \\ V(t = 0) = V_0 \end{cases}$$

Here, coefficient a_0 is the coefficient for the exponential growth part, which is the parameter that we are interested in predicting. The coefficient a_1 is the coefficient for the linear growth part.

The theory behind this model is that all cells proliferate with constant cell cycle durationT_c. Thus, this leads to an exponential growth model. This model also valid when the fraction of proliferative of the tumor size is constant or when the cell cycle period is a random variable with exponential

distribution. Here for our model, we assume the exponential phase to be followed by a linear growth phase.

In our study, since there were only have two screening results and only a limited number of samples (less than 100 trials), we simply set T_c to be the maximum of time between two screenings. Then the solution to obtain the growth rate for our model is:

$$a_0 = \frac{\log\left(\frac{V_1}{V_0}\right)}{t_1 - t_0} = \frac{\log(V_1) - \log(V_0)}{t_1 - t_0}$$

Where V_0 and V_1 are the tumor volume at the time of diagnosis and time of the first revisit, at the time t_0 and t_1 respectively.

The last model we tried to fit the growth model is Logistic and Gompertz models [36]:

$$\begin{cases} \frac{dV}{dt} = aV\left(1 - \frac{V}{K}\right) \\ V(t=0) = 1 \end{cases}$$

Here a is a coefficient related to proliferation kinetics, K stands for the carrying capacity, where all tumor volumes converge. This model can be interpreted as growth model under limited carrying capacity. Again, due to lack of enough data and making calculation simple, we set the carrying capacity to be the maximum of tumor size screened. Here an approximation of the solution for the growth rate is:

$$a = \frac{\log\left(V_{t_1}\right) - \log(V_{t_0})}{(1 - e^{-\frac{t_1 - t_0}{K}})}$$

1.2.3 Application of growth rate models in the survival analysis

As described above, the growth rate has been less studied or used in clinical trials to classify the type of breast cancer patients. Part of the reason is it is not linked to biomarkers, which are widely used for prognostication. In order to connect change in tumor size and type of breast cancer, we further apply survival analysis on the data.

We first try to find the best cut point for the growth rate, which would be used later to separate fast growth patients from slow growth patients. The practice of dichotomizing continuous covariates is widely used in medical and epidemiological research for both clinical and statistical reasons. From the clinical perspective, we can use a cutpoint split continuous variables into 'low' and 'high', establishing a threshold for future studies. Using the criteria, patients can be classified into these two groups and their prognosis can be estimated, which will be used for choosing proper treatments. From the statistical perspective, binary covariate provides a simpler interpretation of effect measures, such as odds ratios and relative risks. In our study, by comparing hazard ratios between groups, we can have not only a clearer understanding of the results but also an easier way for future implementation. Another benefit of using binary covariates is that one does not need to check statistical assumptions before using them to build up our model [39].

The way to choose the best cut point is choosing the one that provides the greatest difference between survival functions of two groups. After that, we fit a logistic regression to predict growth rate mode with biomarkers, such as KI67 scores, CASPS3 scores.

1.2.4 Using survival analysis to analyze breast cancer data

Survival analysis is a set of methods for analyzing data with the outcome as 'time to event', which, in our study, is time to death and later was extended to time to recurrence.

There are several common terms defined in our survival analysis:

- Event: In our study, the events we are interested in is Death and Recurrence;
- Time: The time from the diagnosis to the event of interest;
- Censoring: If a subject does not have an event during the observation time, then these cases are defined as censored. There is no information about the occurrence of the event after observation;
- Survival function: The probability that a patient survives longer than time t, in which 'survives' was defined as still being alive at observation time.

In our study, we mainly used Kaplan-Meier curves and Cox proportional hazard regression to analyze the breast cancer data.

Kaplan-Meier curve is a non-parametric statistic used to estimate the survival function, which was used in our study to measure the length of survival time for patients with breast cancer [40]. There are three assumptions that are important in this analysis. First, all the censored patients have the same survival pattern as those patients who remains in the study. Secondly, all patients have the same survival probability no matter when they joined the study. Thirdly, we can only observe the event at the specified time, which means we can only know what happened when carrying out the follow-up. And the survival probability at particular time is calculated by the formula given below [40]:

$$S_t = \frac{number\ of\ subjects\ remained\ alive\ at\ end\ time}{number\ of\ subjects\ alive\ at\ the\ start\ time}$$

Another survival analytic tool we used is Cox proportional hazard model, where the hazard at time t for an individual with covariates x_i is assumed to be

$$\lambda_i(t|x_i) = \lambda_0(t)\exp(x_i'\beta)$$

In this model, $\lambda_0(t)$ is a baseline hazard function that describes the risk for individuals with $x_i = 0$. In our study, we select the lowest level of each biomarker to be the baseline. Here $\exp(x_i'\beta)$is the relative risk, a proportional increase or decrease in risk, which is what we want to estimate from the model.

1.2.5 Fit Logistic regression model to predict growth rate with biomarkers

In our study, we tried to build up model to predict the type of growth rate for each patient with their prognostic characteristics. This model will be used in the future to predict whether the patient has a fast growth rate or a slow growth rate, which indicates whether he or she would face a high or low risk to death.

The model we choose is logistic regression model. There are several reasons why we made this choice:

- The outcome of our data is a categorical variable, so our model should be the probability of belonging to each class.
- Our model would be applied to the clinical practice, which means our model should not be too complicated for implementation.
- When selection the biomarkers in the model, the logistic regression is an easy way to compare the results from different models. We choose the best model based on the predicting accuracy and model diagnostic result.

1.2.6 Extension of the survival analysis to recurrence of breast cancer

After we obtain the connection between growth rate and risk to death, we want to extend the usage on more clinical practice. Since the idea of survival analysis can be applied to some other time to event study, we just need to change the event of interest to the recurrence of breast cancer. Since people from public health are interested in comparing risks to recurrence among difference races, I extended the survival analysis to the recurrence of breast cancer and tried to compare that of patients with different races.

Clinical studies have revealed a higher risk of breast tumor recurrence in African-American (AA) patients compared to European-American (EA) patients, contributing to the alarming inequality in clinical outcomes among the ethnic groups. However, distinctions in recurrence patterns upon receiving hormone, radiation, and/or chemotherapy between the races remain poorly characterized.

The significant divide in breast cancer mortality between African-American (AA) and European-American (EA) patients remains a challenge for clinicians. Despite a similar number of reported incidences of breast cancer among AA and EA women, AAs experience notably higher severity in clinical outcomes and exhibit a 40% higher death rate than EAs among premenopausal and menopausal breast cancer patients [41-43]. Recurrent breast cancer has impeded successful

management of the disease for decades and is one of the primary factors for this racial division in prognosis [44]. Statistics demonstrate that approximately 40% of all breast cancer survivors will experience a recurrence episode during their lifetime, which has been suggested to play a principal role in breast cancer mortality [45]. Clinical studies have revealed a higher risk of recurrence in AA compared to EA, presumably contributing to the inequality in clinical outcomes among the ethnic groups [41]. This statistic has provided an impetus for clinicians to devise and implement robust prognosticative measures to preclude recurrence in AA breast cancer patients. However, distinctions in recurrence rates and patterns following various forms of treatment between the races have not been thoroughly evaluated. This warrants more investigation to potentially attenuate the observed racial disparity in recurrence in the clinic. Hence, we conducted a large institutional study based in Atlanta, Georgia, in which we analyzed rates and patterns of tumor recurrence post hormone, radiation, and chemotherapy among AA and EA breast cancer patients. This retrospective clinical study uncovered previously unrecognized distinctions in recurrence patterns following each conventional form of treatment among racially distinct breast populations and may impart valuable clinical insight into preclusive measures for mitigating the ethnic disparity in breast tumor recurrence.

We compared patterns and rates (per 1000 cancer patients per 1 year) of recurrence following each form of treatment between AA (n=1850) and EA breast cancer patients (n=7931) from a cohort of patients (n=10504) treated between 2005-2015 at Northside Hospital in Atlanta, GA. Multivariate models were used to examine the effect of age, grade, and stage on our results and 95% confidence intervals were used to determine if there are significant difference between the two groups.

2 METHODS

2.1 Neuronal Culture and Analysis

We used data from the Firestein laboratory at Rutgers University to verify the validity of our models against experimental results, all the data were used with the permission from them. The neuronal images used for this study were previously analyzed [46]. But were reanalyzed using the Firestein laboratory's method of semi-automated Sholl analysis, called Bonfire [47, 48]. The Sholl curves generated for this study were not included in previous study [46]. The experimental procedures used are as follows *Cell culture, transfection, and immunostaining:*

Hippocampal neurons were isolated from embryonic rats at 18 days of gestation (E18) as previously described [47, 48]. Briefly, the hippocampi from Sprague-Dawley rats were isolated and mechanically dissociated. Hippocampal neurons were then plated on poly-D-lysine (PDL)-coated glass coverslips (12 mm diameter) at a density of approximately 1800 cells/mm2. Cells were cultured in Neurobasal medium supplemented with B27, Glutamax, and 1% penicillin/streptomycin (Life Technologies). At 5 days in vitro (DIV), neurons were transfected with cDNA encoding green fluorescent protein (GFP) in the pEGFP-C1 vector (Clontech) using Lipofectamine 2000 (Invitrogen). GFP is expressed throughout the entire neuron and ensures accurate assessment of dendrite number. At 7 DIV, immunostaining was performed to enhance the natural fluorescence of GFP. Neurons were fixed in 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS), after which they were incubated in blocking buffer. Primary antibody incubation (1:1000 dilution of rat anti-GFP from Dr. Shu-Chan Hsu of Rutgers University) occurred at 4˚C overnight. Coverslips were washed 3 times with PBS and then incubated with secondary antibody (1:250 dilution of Cy2-anti rat IgG from Jackson Immunoresearch) for 1 hour at room temperature. Coverslips were washed twice more with PBS and then incubated with Hoechst for 5 min at room temperature to stain the nuclei. Coverslips were washed a final time with PBS and were then mounted onto glass microscope slides using Fluoromount (Southern Biotechnology).

Transfected cells were imaged using an Olympus Optical IX50 microscope with a Cooke SensiCam charge-coupled device (CCD) cooled camera fluorescence imaging system and ImagePro software (Media Cybernetics). All images were taken at 200x magnification.

Images were processed as previously described [47, 48], using customs scripts written in Matlab (MathWorks). Briefly, cell bodies and dendrites were traced in ImageJ using the NeuronJ plugin (NIH, Bethesda, MD). The data were exported to NeuronStudio and checked to ensure proper connectivity of dendrites. Sholl analysis was then performed at 6 μm intervals starting at 0 μm from the soma using the Bonfire program [47, 48].

2.2 Discrete statistical model for neural growth without restriction

2.2.1 Recurrence formulas for the probability distribution of all possible neural trees

The results from the previous section can be proven using the discrete probability distribution function at any time step N, with p as the branching probability and q is the elongation probability. The evolution of all possible trees may be described using a recursive function $f_n(p, q)$ as follows: For the first step (step 0), use $f_0(p, q) = 1$ to denote a single branch (see Table 1, step 0). At step 1, $f_1(p, q) = p + q$, (see Table 1, step 1), indicates the existence of two trees: the first with two active branches and probability of instantiation p, and the second containing a single branch and instantiated with probability q. Note that since $p + q = 1$, the sum of all probabilities adds up to 1, as needed. At step 2, we can describe the existing trees using

$$f_2(p, q) = q * p + \mathrm{q} * \mathrm{q} + \mathrm{p} * (\mathrm{p}^2 + 2\,\mathrm{p}\,\mathrm{q} + \mathrm{q}^2) = \big(\mathrm{q} + \mathrm{p}(\mathrm{p} + \mathrm{q})\big) * (\mathrm{p} + \mathrm{q}).$$

These trees can be identified in the tree evolution (Figure 1) for steps 1-2, and are listed in Table 1 for step 2. A summary figure is listed in Figure 3.

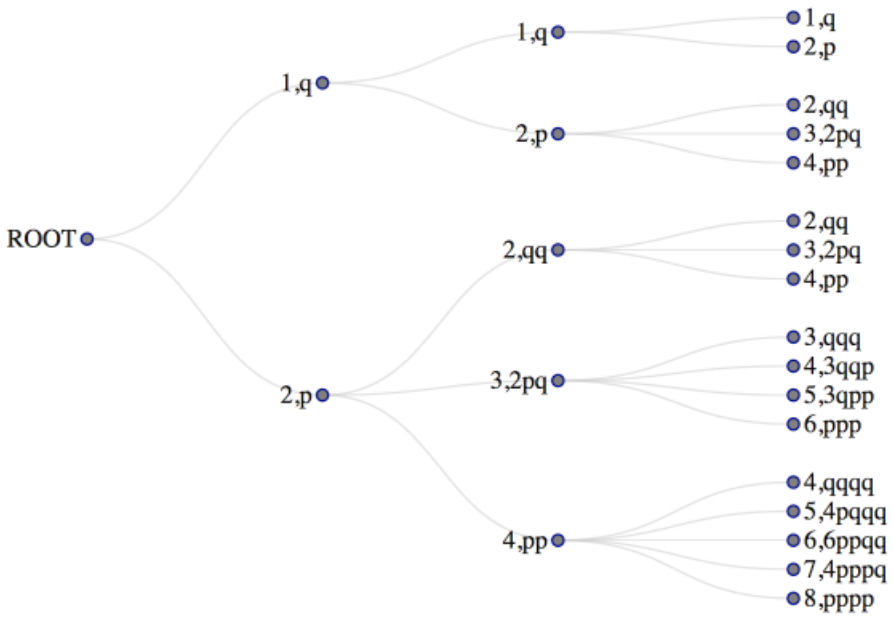


Figure 3 *Summary figure for the evolution of the trees for the first three time steps.*

The number of terminal branches is listed for each tree, followed by the probability of obtaining that particular outcome starting from the previous step. The overall probability of a certain outcome can be obtained by multiplying the probabilities along the path.

Intuitively, the single-branch tree can evolve on two possible paths (extend or branch) as indicated by the first two terms. The tree with two terminal branches can generate trees with 4, 3 and 2 terminal branches, respectively. The formula for step 3, $f_2(p,q)$, can now be obtained by substituting p by $p*(p+q)$ in formula $f_1(p,q)$ and multiplying the result with $(p+q)$. We thus hypothesize the following recurrence formula for generating all trees at step n:

$$f_{n+1}(p,q) = f_n(p*(p+q),q)*(p+q) \qquad (5)$$

A proof of this formula is provided in the next section. This explicit recursive description of the discrete probability distribution now allows us to compute the expected number of mean

branches and the associated variance in the general case, extending the enumerative example shown in Table 1. As a result, when the number of steps is small enough such that all trees have total lengths below the maximum allowed value (i.e. before they run out of resources and cannot extend anymore), we can prove that the expected mean indeed follows the general formula $E(n) = (1+p)^n$. Furthermore, we were able to derive a recurrence formula that described the associated variance:

$$Var(S_n) = p * (1-p) * (1-p)^{n-1} * g(n) \quad (6)$$

where $g(n+1) = p * g(n) + g(n) + 1$, with $g(1) = 1$

This has the following solution:

$$\mathrm{Var_n(X)} = (1+\mathrm{p})^{2*(\mathrm{n}-1)} * (1-\mathrm{p}^2) + (1+\mathrm{p})^{\mathrm{n}-1} * (\mathrm{p}-1) \;\; \mathrm{n} = 1,2,\ldots\ldots \quad (7)$$

2.2.2 Derivation of the recurrence formula for all possible neuronal trees

We now prove the formula for generating all trees: $f_{n+1}(p,q) = f_n(p.(p+q), q).(p+q)$. We write the formula for step n as $f_n(p,q) = \sum_{k=1}^{2^n-1} p^k g_k(q)$ where $g_k(q)$ are polynomials in q. Obviously, this is true for the first 3 time steps considered in Table 1. Each individual term in the $f_n(p,q)$, of the form $p_k \; g_k(q)$ represents the probability of generating a tree with $k+1$ active branches. Some of the trees are isomorphs, for simplicity that will be absorbed into the coefficients from $g_k(q)$. In order to derive the general formula for $n+1$ step, we note that the trees that contain a p^k term have k + 1 terminal or active branches (see Table 1). Then, at the next step, these active branches can generate between 0 and $(k+1)$ new active branches. Taking into account the degeneration (isomorph trees), the probabilities for these new branches are described by the combinations from the $(p+q)^{k+1}$ formula. We then obtain the following:

$$f_{n+1}(p,q) = \sum_{k=1}^{2^n-1} p^k (p+q)^{k+1} g_k(q) = (p+q) \sum_{k=1}^{2^n-1} p^k (p+q)^k g_k(q)$$

$$= (p+q) \sum_{k=1}^{2^n-1} (p(p+q))^k g_k(q) = (p+q) f_n(p,q)$$

This proves the validity of formula of mean and variance.

2.2.3 Derivation for the expected number at distance k: $\boldsymbol{P(n) = (1 + p)^n}$

We want now to prove that the expected number of branches at step n is $(1 + p)^n$. In order to achieve this, we need to prove the following intermediary steps. First, we can determine the expected number of active branches at step n as the derivative of the $f_n * p$ function: $(f_n(p, q) * p)$', where the symbol ' stands for derivative with respect to p. Since the terms in the sum of $f_n(p, q)$ are all the entries in the global probability table, we now have the expected number of active branches at time step n (and distance $n + 1$ away from the origin):

$$E(number\ active\ branches\ at\ step\ n) = E(b_n)$$
$$= \sum_{k=all\ trees} number\ active\ branches\ on\ tree * tree\ probability$$

We rewrite this to be:

$$f_n(p, q) = \sum_{k=1}^{2^n-1} p^k g_k(q)$$

We now obtain the following:

$$E(b_n) = \sum_{k=all\ trees} number\ active\ branches\ on\ tree * probability\ y\ tree$$
$$= \sum_{k=0}^{2^n-1} (k+1) p^k g_k(q) = \left(\sum_{k=1}^{n} p^{k+1} g_k(q) \right)' = \left(p. \sum_{k=1}^{n} p^k g_k(q) \right)'$$
$$= (p. f_{n+1}(p, q))'$$

Using these intermediary results, the expected number of branches evaluates as $(1 + p)^n$, as proved by induction below.

$$E(b_n) = (1 + p)^n$$

$$E(b_n) = (p.f_{n+1}(p,q))' = \left(p.\sum_{k=0}^{2^n-1} p^k(p+q)^{k+1}g_k(q)\right)' = \left(\sum_{k=0}^{2^n-1} p^{k+1}(p+q)^{k+1}g_k(q)\right)'$$

$$= \left(\sum_{k=0}^{2^n-1} (p.(p+q))^{k+1}g_k(q)\right)' = \sum_{k=0}^{2^n-1} (k+1)(2p+q)(p.(p+q))^k g_k(q)$$

$$= \sum_{k=0}^{2^n-1} (k+1)(1+p)(p.(p+q))^k g_k(q)$$

$$= (1+p)\sum_{k=0}^{2^n-1} (k+1)\big(p.(p+q)\big)^k g_k(q)$$

$$= (1+p)\sum_{k=0}^{2^n-1} (k+1)(2p+q)p^k g_k(q) = (1+p)\big(p.f_n(p,q)\big)' = (1+p)^{n+1}$$

Furthermore, we can extend this approach and compute the variance for this expected number, using $var(X) = E(X^2) - E(x)^2 = ((f_n(p,q) * p)' * p)' - ((1+p)^n)^2$

After simplifications, this can also be written as a recursion formula:

$$Var(S_n) = p.(1-p).(1-p)^{n-1}.g(n)$$

$$g(n+1) = p.g(n) + g(n) + 1$$

$$g(1) = 1$$

After solving this difference equation, we obtain the solution for the general formula of the variance:

$$\mathrm{Var_n(X)} = (1+\mathrm{p})^{2(\mathrm{n}-1)}.(1-\mathrm{p}^2) + (1+\mathrm{p})^{\mathrm{n}-1}.(\mathrm{p}-1) \;\; \mathrm{n} = 1,2,$$

2.3 Neural Growth Model with Restrictions

2.3.1 Optimal targeting region for trees with size restrictions

At larger distances from the origin, some trees should be eliminated from the analysis. For example, if the maximum length of the tree is equal to 7, the larger trees will not generate any additional subtrees after time step 3. However, the smaller ones do, and the formulas for the number of searching sites at distance n that allow for a tree of maximum size 7 are the following:

Table 1 List of the trees that can be generated when maximum length of the tree is equal to 7.

Step 1	$p + 1$	enough resources
Step 2	$(p + 1)^2$	enough resources
Step 3	$(p - 1)^2 * (2 * p^3 + 5 * p + 1)$	some trees are eliminated
Step 4	$-(p - 1)^3 * (-2 * p^2 + 3 * p + 1)$	some trees are eliminated
Step 5	$(p - 1)^4 * (p + 1)$	the one that does not branch at all and the one that branches after 4 steps
Step 6	$(p - 1)^6$	only the tree that does not branch at all reaches this far

2.3.2 Right tail series expansion up to order p^3

The series expansion for the right tails follows the formula $\frac{f(N-k)}{f(n-k+1)} = f_1(k) \cdot p + f_2(k) \cdot p^2 + f_3(k) \cdot p^3 + \cdots = \sum_{j=1}^{k} f_j(k) \cdot p^j$, where the f_j functions are defined below:

$$f_1(0) = q^N, f_2(0) = 0, f_3(0) = 0, f_4(0) = 0$$

$$f_1(1) = q^{N-1}, f_2(1) = 2pq^{N-2}, f_3(1) = 0, f_4(1) = 0, f_1(k) = q^{N-k}$$

$$f_2(2) = 2pq^{N-2} + 2pq^{N-3}, f_3(2) = 0, f_4(2) = 0$$

$$f_2(k) = 2pq^{N-2}\sum_{i=1}^{k} p^{i-k}, f_3(2) = 0, f_4(2) = 0$$

$$f_3(3) = 6p^2q^{N-4}, f_4(3) = 0$$

$$f_3(4) = 6p^2q^{N-4} + 6p^2q^{N-4}, f_4(4) = 4p^3q^{N-6}$$

$$f_3(k) = 6p^2q^{N-4}\sum_{i=3}^{k} g(i)p^{i-k}$$

$$g(0) = g(1) = g(2) = 0, g(3) = 1, g(4) = 1, g(5) = 2, g(k) = g(k-2) + 1$$

$$f_4(k) = 4p^3q^{N-6}\sum_{i=5}^{k} h(i)p^{i-k}$$

$$h(0) = h(1) = h(2) = h(3) = h(4) = 0, h(5) = 1, h(6) = 1, h(7) = h(5) + 6 = 7$$
$$h(8) = h(6) + 7 = 8, h(9) = h(7) + 7 = 14$$

$$h(3j+4) = h(3j+2) + 6j$$
$$h(3j+5) = h(3j+3) + 6j + 1$$
$$h(3j+6) = h(3j+4) + 6j + 1$$

Therefore, up to the 4th order approximation in p we have:

$$R(N-k) = \frac{f(k+1)}{f(k)} = \frac{f_1(k+1) + f_2(k+1) + f_3(k+1) + f_4(k+1)}{f_1(k) + f_2(k) + f_3(k) + f_4(k)}$$

$$R(N-k) = \frac{q^{N-(k+1)} + 2pq^{N-2}\sum_{i=1}^{k+1} p^{i-k} + 3p^2q^{N-4}\sum_{i=3}^{k+1} 2g(i)p^{i-k} + 4p^3q^{N-6}\sum_{i=5}^{k+1} h(i)p^{i-k}}{q^{N-k} + 2pq^{N-2}\sum_{i=1}^{k} p^{i-k} + 3p^2q^{N-4}\sum_{i=3}^{k} 2g(i)p^{i-k} + 4p^3q^{N-6}\sum_{i=5}^{k} h(i)p^{i-k}}$$

Table 2 Coefficients of first 3 orders from the proportions between steps of an example with 26 steps.

$\frac{Step\ (N-1)^{th}}{Step\ N^{th}}$	Coefficient of 1st order	Coefficient of 2nd order	Coefficient of 3rd order
N = 26	3	5	7
N = 25	3	-1	3
N = 24	3	-1	17
N = 23	3	-7	47
N = 22	3	-7	55
N = 21	3	-13	135
N = 20	3	-13	141

2.3.3 Determining the approximation for the peak location

We can use this table 2 to derive a better, second order approximation that improves the first order approximation derived in formula (3), namely $k = \frac{3+5p}{6p}$. Similar to the derivation of this equation, the exact location where the ratio becomes 1 can be used to approximate where the spatial location of the peak of the distribution occurs. This occurs at step N – k, when $\frac{P(N-k-1)}{P(N-k)}$, which starts with an initial value of $(1 + 3p + 5p^2 + f_3(k)p^3)$, reaches a value of 1 after experiencing continuous decreases in value. This is determined by:

$$1 + 3p + (5 - 6k)p^2 + f_3(k)p^3 = 1$$

which can be easily rewritten as

$$3 + (5 - 6k)p + f_3(k)p^2 = 0$$

which can only be solved numerically in k due to the complicated expression for the function $f_3(k)$.

The condition that all neurite tips reach a target distance D translates to:

$$D = L_0 N_{max} = L_0 \log_2(1 + \frac{L_{max}}{L_0})$$

Now we can set up the comparison: using the geometrical series result, we determine that the average number of branches at time step k is $(1+p)^k$, and therefore we obtain the total length of the tree after completion of n steps:

$$1 + (1+p) + (1+p)^2 + \cdots + (1+p)^n = \sum_{k=0}^{n} (1+p)^k = \frac{(1+p)^{n+1} - 1}{1 + p - 1} = \frac{(1+p)^{n+1} - 1}{p}$$

Setting up the condition that the tree stops expanding when it reaches maximal allowed length L, we get:

$$L = \frac{(1+p)^{n+1} - 1}{p}$$

We can now compute p from this equation: $p \cdot L = (1+p)^{n+1} - 1$, or equivalently $1 + p \cdot L = (1+p)^{n+1}$ Taking log on both sides: $\ln(1 + \mathrm{pL}) = (n+1)\ln(1+p)$, or equivalently $n + 1 = \frac{\ln(1+pL)}{\ln(1+p)}$. We finally obtain:

$$n = \frac{\ln(1+pL)}{\ln(1+p)} - 1$$

Note that when p → 0, the single-branch tree expands all the way to x = N (in N - 1 time steps) since

$$\lim_{p\to 0} n = \lim_{p\to 0}(\frac{\ln(1+pL)}{\ln(1+p)} - 1) = L'Hospital's\ Rule = \lim_{p\to 0}\left(\frac{ln'(1+pL)}{ln'(1+p)}\right) - 1$$

$$= \lim_{p\to 0}\left(\frac{\frac{L}{1+pL}}{\frac{1}{1+p}}\right) - 1 = L - 1$$

Because the probability of this tree is $f(p) = (1-p)^{N-1}$, the function f is obviously maximized at $p = 0$. Not surprisingly, the symbolic/numeric evaluation yields the same result. At the other

end of the spectrum, when p → 1, the tree expands as $1 + 2 + 4 + 8 + 16 + \ldots = N$ and runs out of resources. As an example, when $L = 32$, we obtain:

$$n = \frac{\ln(1+1L)}{\ln(1+1)} - 1 = \frac{\ln(1+L)}{\ln(2)} - 1 = ln_2(1+L) - 1 = ln_2(1+31) - 1 = 5 - 1 = 4$$

These results assume that the number of tree branches doubles without fail at each time step until the tree will finally run out of resources. At that point, the number of terminal branches will be maximal, meaning that the tree peaked at the right location and it will have the highest chance to find a target in this blind search there. Note that in contrast to previous results, what we seek here is the optimal branching probability that would allow this to happen. The optimization problem here first selects the distance at which maximal performances are sought, then the uniform branching probability is determined.

2.4 Tumor Growth Models

2.4.1 Materials and methods for predicting tumor growth rate

Patients: A total of 114 patients were derived from the database of screen-detected primary breast carcinoma patients presented to Nottingham City Hospital (UK) from 1988 to 2008 with interval cancer, and for whom revision of the previous screening mammography showed a previously undetected cancer at the same affected site, all the data was used with permission from our collaborators. This misdiagnosis might be due to a minimal visible diagnosis in the previous mammogram. The maximum tumor dimension identified by the mammography was available for each case. We excluded cases presented with predominant calcification. All the clinicopathological information was available for 92 patients. For each case, information relating to changes in tumor size between the time of screening and of diagnosis as well as the dates of screening and diagnosis was used to estimate tumor growth rate. Clinical and pathological data of the patients including age, histological tumor type, primary tumor size, lymph node status, histological status, NPI and vascular invasion were obtained in a standard manner at the time of diagnosis was compiled.

Calculated Tumor volume: The two measurements in the mammogram screening were assumed as tumor diameter and tumor height. The highest mammogram reading was assumed as height and the other as diameter to calculate the volume using the formula $\frac{4}{3}\pi ab^2$ (assuming the tumor to be ellipsoidal). The measured tumor height was introduced into the formula in the position b after dividing by 2 whereas the tumor diameter was introduced in the position of a after dividing by 2. We simply used the formula

$$Tumor\ Volume = \frac{4}{3}\pi ab^2$$

to calculate the tumor volume at the time of diagnosis and tumor volume at the time of screening to determine the change in volume.

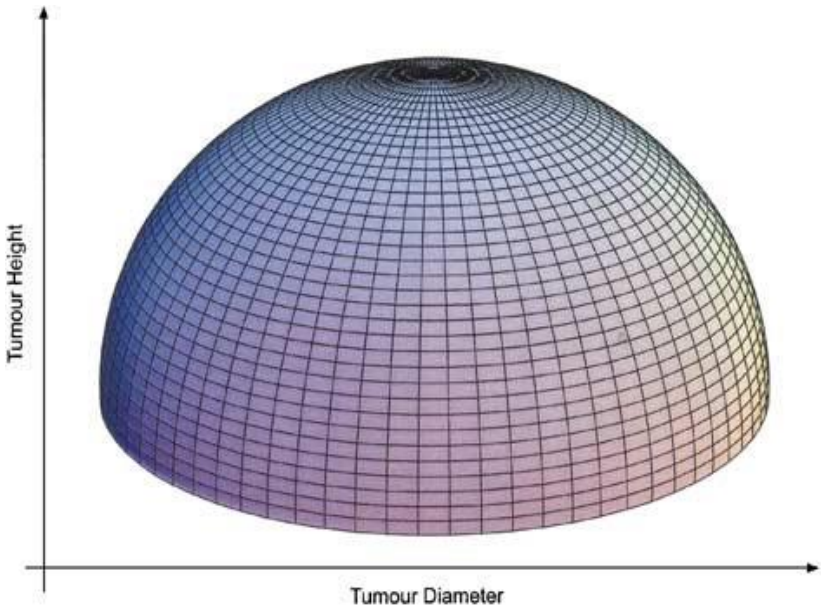


Figure 4 *Ellipsoid Tumor*

Ellipsoid tumor calculated by tumor diameter and tumor height with the formula 4/3 πab^2. Graph reproduced from Richtig study [80].

2.4.2 Model selection for growth rate

As described in the introduction part, we have three models to calculate the growth rate. And based on the results, we choose the Exponential model, which provides the best result from

survival analysis, more results will be shown in the Result section 3.4.2.

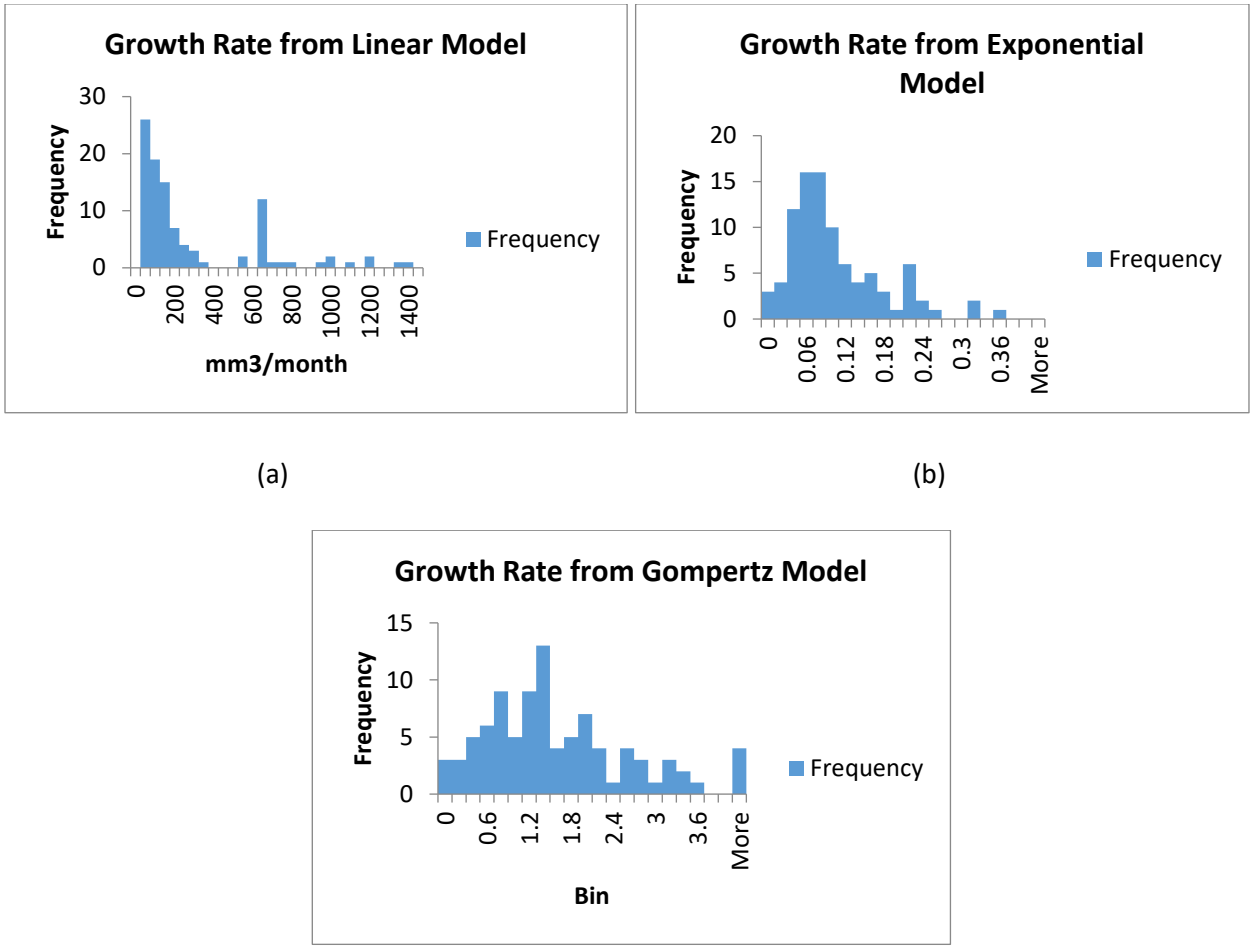


Figure 5 *Histogram of individual growth rate from three models*

Figure 5 shows the histogram of growth rate from three models. (a) is from the linear model which gives most skewed results, which is not useful for further study. (b) and (c) are from the exponential model and the logistic model, both of which are less skewed. The plot for (c) however shows that there are more outliers from the third model. Since (b) is closer to a bell shaped distribution, we choose it for further studies.

Calculated Tumor Growth Rate:

$$\alpha = \frac{\log(Volume_t) - \log(Volume_{t0})}{t - t0}$$

Where,

$Volume_t$ is volume at time of diagnosis calculated as:

$$Volume_t = Volume_{t0} * e^{\alpha(t-t0)}$$

$Volume_{t0}$ is volume at time of screening. $e^{t-t_0}$is an exponential function with α being growth rate and $t - t_0$is the time difference between screening and diagnosis.

2.4.3 Biomarkers in the model

Immunohistochemistry: For each patient, a representative formalin-fixed paraffin wax-embedded tumor block was obtained from the Nottingham breast tumor bank. 4 µm thick full-face sections were prepared and placed onto glass slides (Xtra®, Surgipath).

ER, PR, HER2, CK5/6, Ki67, BCl2, and MCM2 staining: We applied IHC to FFPE tissue sections using Novolink™ Max Polymer Detection System from Leica Biosystems (Leica, Newcastle, UK). Heat-induced retrieval of antigen epitopes was performed in citrate buffer (pH 6) using a microwave for 20 minutes, followed by immediate cooling. The slides were rinsed with Tris Buffered Saline (TBS, pH 7.6). The primary antibody was applied for 30 minutes at room temperature. Optimal antibody dilutions have based on a study by El Rehim et al, (2005). Dako® Antibody Diluent (ref # 50809, let # 10032882) was used. To visualize antibody binding, 3-3' Diaminobenzidine tetrahydrochloride (Novolink DAB substrate buffer plus) was freshly prepared and used as a chromogen. The tissue sections were counter stained with Mayer's hematoxylin for 6 minutes. Slides were dehydrated in alcohol, cleared in Xylene then mounted with DPX.

Cleaved Caspase-3 Staining: This marker was stained for using a pre-fabricated detection kit (SignalStain® Cleaved Caspase-3 (Asp175) IHC Detection Kit #8120, Cell Signaling Technology) following manufacturer's instructions. As before, slides were de-waxed and rehydrated before antigen retrieval was carried out. Pre-diluted cleaved caspase-3 Primary Antibody (# 9661, clone D175, polyclonal rabbit) was applied and incubated overnight at 4°C ensuring they did not dry out. Following incubation, prediluted Biotinylated Secondary Antibody was applied. Slides were closely monitored after the application of NovaRedTM Substrate Chromogen solution and were immersed in dH20 when red-brown staining was observed (typically after approximately 7-8

minutes). They were counterstained with Haematoxylin for 6 minutes, dehydrated in denatured ethanol and xylene as before, before mounting with cover slips.

Assessment of IHC staining:

For each slide, only the invasive tumor component was scored. Cases were scored twice using light microscopy; the second time blind to the first results. Scores were then averaged. The cases were scored without the prior knowledge of patients' clinicopathological data. For ER and PR nuclear immunoreactivity, H-Score was employed. An H-score of 10 was used as the cut-off, in accordance with normal clinical practice at Nottingham.

For HER-2, overall membranous staining intensity was scored on a 0-3 basis using guidelines by the American Society of Clinical Oncology and College of American Pathologists, with 0 representing no visible staining, and 3 representing extremely intense membranous staining.

For Ki67, only nuclear immunoreactivity of proliferating cells was recorded, whilst for cleaved Caspase-3, only cells undergoing apoptosis were noted. For both markers, the invasive tumor component in each case was scanned until the staining "hot-spot" was identified. At 40x magnification, the number of cell nuclei stained regardless of staining pattern or intensity per 1000 invasive tumor cells was then counted. The % staining in the "hot-spot" field was then calculated.

CK5/6 was scored as the overall percentage of positively staining cells within the whole invasive tumor component. For the purposes of this study, and based previous studies a cut-off of 10% staining was used to distinguish between positive and negative expression groups. Breast cancer molecular subtypes were defined based on their IHC expression profile for ER, PgR, HER2, CK5/6 into1) luminal (ER+ and/or PR+ /HER2-), 2) HER2+ (HER2 positive), 3) Triple negative (TN; ER-, PR-, HER2-) and Basal-like Breast cancer (BLBC: TN+ CK5/6 +). [50]

2.4.4 Statistical analysis on the growth rate

All the statistical analysis in this paper was performed using Statistical Analysis System (SAS 9.4, Cary, North Carolina, US). For the following analysis, P-value less than 0.05 were taken as significant, for both one-tailed tests and two-tail tests. Spearman's Rank-Order Correlations were

used to test the correlation between the growth rate and each biomarker. We also apply Fisher's exact test on the contingency table to examine the association between type of breast cancer and each biomarker. To compare the growth rate among patients with different levels of biomarkers, we did the analysis of variance (ANOVA) on the biomarkers, which are associated with growth rate determined by Spearman's Rand-Order correlation test. At last, Multivariate Cox Proportional Hazard Model was fitted with growth rate and potential confounders selected from the previous analysis, in order to determine the effect of a unit increment in the growth rate on the change of hazard rate.

2.4.5 Choose the cut point for growth rate

In order to find the optimal cut point, we first have a set of possible of candidates and then choose the best one based on the results of dichotomizing the patients with different performances. The maximum number of candidate cut points is k- 1, where k is the number of unique values in the data of that continuous covariate. Some researchers have suggested excluding the outer 10-20% of the continuous covariate distribution to avoid having small numbers in one of the groups the following dichotomization, thereby preventing substantial losses in statistical power [39]. The inner 80-90% of the distribution from which a cutpoint is chosen is referred to as the selection interval. For each candidate cutpoint within a specified selection interval, an appropriate two-sample test with concomitant test statistic and p-value (Pc) is determined. A cutpoint model may be appropriate if any P_c is less than or equal to some allowable level of Type I error. The optimal cutpoint is often defined as that candidate cutpoint with the smallest Pc. This method for estimating a cutpoint is referred to as the minimum p-value approach, or alternatively the maximum statistic approach. Other criteria for choosing an optimal cutpoint have been suggested, including maximum effect size and the maximum precision of estimates, but have received less support.

We first choose all possible candidates of cutpoints, then we calculated the p-values from log-rank test [51]. Each log-rank test is trying to test the hypothesis that the survival curves of two groups are the same. Then the cutpoint with the lowest p-value is the one that split the covariates best.

After I find the optimal cutpoint, I used SAS proc lifetest to test the efficiency of separating the patients into two distinct groups with different level of risk to death. First, the survival curve is estimated by Kaplan-Meier method.

Then, in order to test the difference between the risks to death of two groups, I used log-rank test which is well known and widely used [52]. It is used to test the null hypothesis that survival curves of two populations have no significant difference. The test statistic is calculated as follows:

$$\chi^2(\log rank) = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Where the O_1 and O_2 are the total numbers of observed events in group 1 and 2, respectively, and E_1 and E_2 the total number of expected events. The total expected number of events for a group is the sum of the expected number of events at the time of each event. The expected number of events at the time of an event can be calculated as the risk to death at that time multiplied by the number alive in the group. Under the null hypothesis, the risk to death can be calculated from the combined data for both groups.

2.4.6 Fitting the logistic regression model to predict growth rate

In order to build up model to predict the growth rate, we fitted a logistic regression model with patient's biomarker. Here we used backward model selection method to choose the best model for the future use, which has been done with following steps:

Variable transformation: Our prognostic data include different types of variables, such as strings, numeric and symbolic. Before we use them in the model, we transformed them into proper form based on both statistical and clinical reasons. From the statistical perspective, all variables that have potential problem of multicollinearity and non-normality must be transformed, by either factorized or scaled. From the clinical perspective, we need to transform variables in order to make them easy to be used in the future implementation.

Step-wise model selection: We put all variables in the model in the beginning and calculate all the statistics of the model, such as R-squared, AIC, ANOVA, and so on. Then, based on the significance of the coefficients, we eliminated the least influencing variable from the model and

calculated the statistics again. We kept doing this until obtaining a model with all significant variables.

Cross validation: In order to make our model robust for new data, we used cross-validation method to assess our model. We choose 80% of the data to be the training data, with 20% of the data to be the validation set. The best model is the one not only works well for training data, but also works well for the validation set.

2.5 Extending survival analysis on the recurrence of breast cancer

2.5.1 Cohort study

In this study, a large cohort of breast cancer patients treated at Northside Hospital (NH) in Atlanta, Georgia from 2005 to 2015, were examined, all data was used with permission from our collaborator Ritu. We received approval and permission by the institutional review board at Northside Hospital to access patient clinico-pathological information used in this study and have a written human subject's assurance on file. The demographics and clinico-pathological characteristics of each patient were recorded to generate a database of 10,504 patients. Patient demographic information recorded in the database included age at the time of diagnosis and ethnicity. Age at diagnosis among patients was divided into three subgroups, comprised of patients below the age of 48 (premenopausal), over the age of 55 (postmenopausal), and in between (premenopausal), to precisely describe menopausal status. The races of patients in the database were primarily comprised of African-Americans (AA) and European-Americans (EA). The "unknown/others" subcategory denote patients of all other ethnicities (excluding AA and EA) and patients lacking race information. Ethnicity was reported according to the patient's claim. Breast tumor characteristics that were recorded for each patient consisted of nuclear grade, Nottingham (NGH) grade, stage, nodal status, T (primary tumor), N (lymph node metastasis) and M (distant metastasis) classifications. The 7th edition of the American Joint Committee on Cancer (AJCC)/Union for International Cancer Control (UICC) TNM Classification and Stage groupings for breast carcinoma was used in this article. All patient treatments were recorded, including chemotherapy, hormone, and radiation therapy. Patients that underwent chemotherapy were

subcategorized into neoadjuvant and adjuvant depending on the timing of treatment. Additionally, any combination of the hormone, radiation, and chemotherapy that patients received was labeled as a combination of adjuvant therapies. Follow-up data was collected to determine breast cancer recurrence episodes, as well as the site of recurrence, such as local, regional or distant sites. Local recurrences comprise recurrence of the tumor in the primary site. Regional recurrence encompasses recurrence of the breast cancer in adjacent lymph nodes. Distant recurrences involve metastatic breast cancer in remote organs such as distant lymph nodes, bone, liver or others.

2.5.2 Follow up

Both follow-up of patients and initial diagnosis occurred between the years of 2005 and 2015. Initial diagnosis dates, as well as treatment start and completion dates for any therapies, were documented. Dates of the last contact for all patients were recorded. Survival status (alive/dead) was reported for each patient along with survival time. Dates of the first recurrence were noted. February 19, 2015, was the final follow-up for the last patient seen.

2.5.3 Statistical Analysis

A significance level of 0.05 and 95% confidence intervals were selected for all analyses. Sample sizes were based on the available patients that comprised each category in the NH database and not power analysis. Chi-square tests were performed to examine significant differences in clinico-pathological characteristics, therapy administration, and recurrence characteristics between recurrence and non-recurrence patients as well as between AA and EA breast cancer patients. Recurrence rates were calculated as per 1000 person-years (incidence rate) from the date of diagnosis until the first incidence of recurrence over a 10-year period irrespective of specific treatment and for each form of treatment administered. Recurrence was identified as tumors that either reformed from the primary tumor or cells that metastasized from the primary site and colonized at a distant site. Test statistics were computed using MATLAB (MATLAB and Statistics Toolbox Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States) program and 1-tailed univariate p-values were reported. The one-tailed analysis was preferred over two-tailed for this particular study to adequately reflect the presumption that treatment is expected to

improve patient outcome. Multivariate Cox proportional hazard models were computed to determine significant differences in recurrence rates and patterns between the racial groups (6, 7). These statistical models were additionally modified to control for variables of age, grade, and stage. The Kaplan-Meier analysis was conducted in SAS 9.4 program to estimate survival function for AA and EA with the recurrent disease over a 10-year period from the time of first tumor reappearance until death or end of follow-up. A log-rank test was conducted to evaluate significance level for between-race differences in survival. Finally, a t-test was used to compare mean time from first recorded recurrence event until death among patients with distant recurrence.

In order to extend the survival analysis on the recurrence of breast cancer, we borrowed the concept of time to death and applied on the time to recurrence. Then instead of risk to death, now we can compare the risk to recurrence among different groups.

3 RESULTS

3.1 Discrete Model

3.1.1 Unrestricted neural growth model

In this study, we used a simplified stochastic model to investigate how changes in branching probability influence a neuron's success in targeting locations that differ in distance from the initiation point. Since our model aims to mimic experimental results that occur in laboratory tissue growth, we assume that each neuron performs a blind search in their quest to connect to the other neurons, and as such, it cannot benefit from the chemical cues that exist during normal development. Therefore, success in finding a target within a certain region depends on the number of active branches that reach that portion of space. The rules for generating a simplified neuronal tree, described in our previous work [2], are as follows. Initially, there is one active branch that at each discrete time step can branch into two neurites with probability p or further extend with probability$q = 1 - p$. Furthermore, each active branch then acts independently of the others and can further extend or branch out to create one additional new neurite. We only account for an extra neurite because we can make sure that the probability to create more than two branches is too small to be included in our model, by choosing small enough time intervals. As a result, multiple branching processes shape the structure of the resulting neuronal tree in a random fashion. While in our previous work each branch is allowed to evolve stochastically in space, here we ignore the spatial structure of these branches and assume that they evolve linearly after branching at small angles to facilitate the statistical estimates for the resulting probability distribution. This simplification allows us to obtain analytical results while taking into account the most important feature of the stochastic neuronal tree, namely the random generation of extra neurites. The advantage of this approach is that for each step, we can generate the full probability distribution for each possible outcome. An illustration for all possible outcomes is shown in Figure 6 for the first three time steps, reproduced from [2].

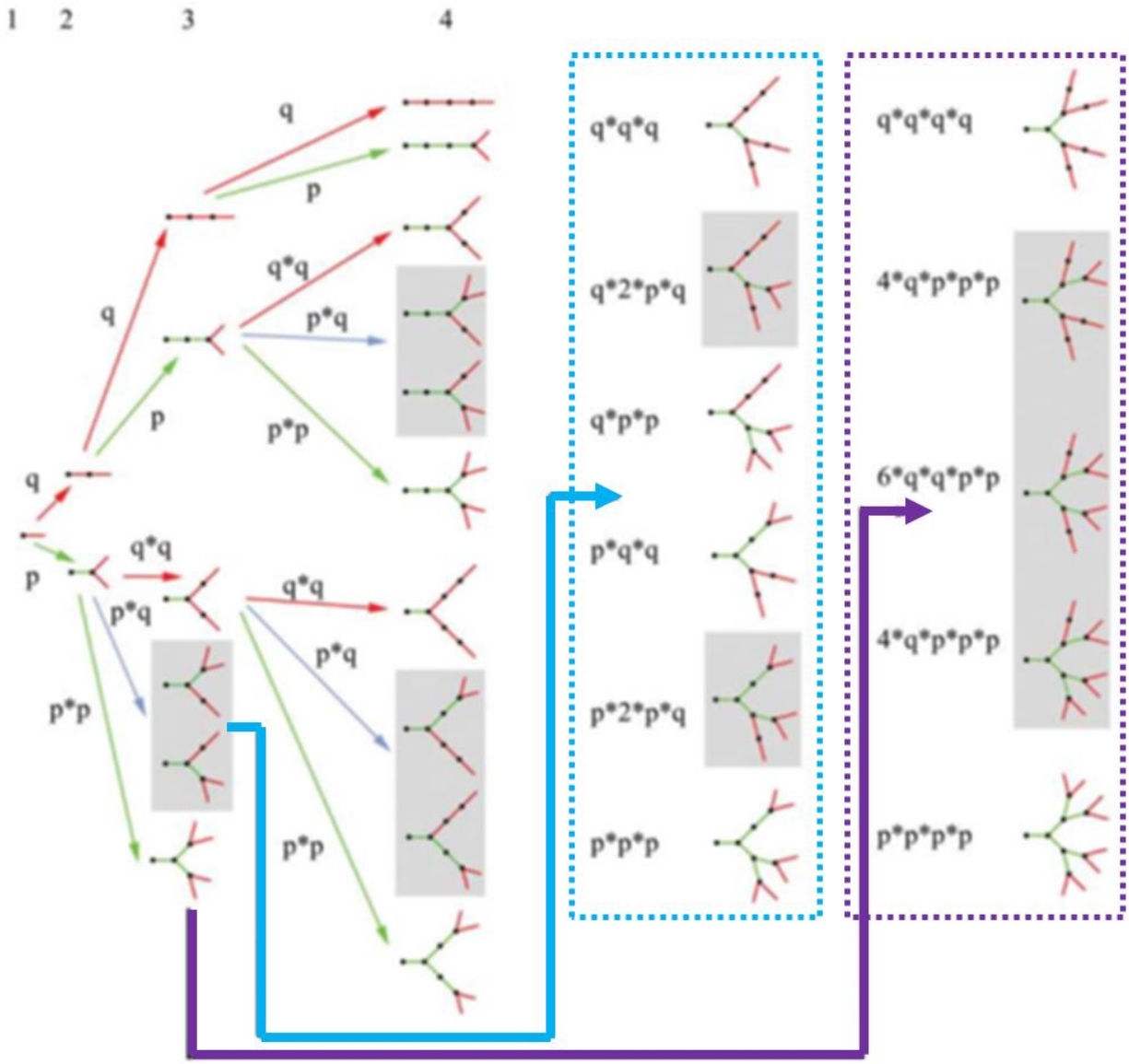


Figure 6 *Diagram of all possible tree instantiations after 3 time steps.*

The full population distribution can be generated by examining all possible tree configurations that can be achieved after each time step. For example, at $t = 3$, the simplest tree is a single evolving branch of length 4 that is obtained with a probability of q^3. At the opposite end of the spectrum, the most complex tree contains 8 active branches of length 1, obtained with probability p^7. The associated probabilities can be determined by computing the products of individual probabilities along the arrows. Each tree has an associated probability and some of the trees listed have multiple replicates (shaded boxes)

3.1.2 Discrete probability distribution for the evolving family of trees

An example of a probability table thus generated is shown in Table 3 for the first three time steps depicted in Figure 6. We assume that the parameter most reflective of the probability of finding a target at a given distance from the starting point is the number of active branches at that distance. This is due to the fact that since in the first order approximation we do not care about the detailed neural structure at a distance from the origin, the probability of success in a blind search is proportional to the number of local branches. The stepwise mean and variance of the expected number of branches at this distance from the origin can then be determined as a function of the branching probability p. We use the following formulas for expected mean and variance: $\overline{x} = \sum_{i=1}^{n} x_i \cdot p_i$ and $Variance = \sigma^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot p_i$ for the discrete probability distributions listed in Table 3 for the first few time steps.

Table 3 Average number of active branches and the associated variances for the first two time steps.

Tree	Probability	# active branches	Total length	Expected value	Variance σ^2
Step 0				**1**	**0**
1	**1**	**1**	**1**		
Step 1				**(1+p)**	**$-p^2 + p$**
1	**q**	**1**	**2**		
2	**p**	**2**	**3**		
Step 2				**$(1 + p)^2$**	**$-p^4 - 2p^3 + p^2 + 2p$**
1	**q·q**	**1**	**3**		
2	**q·p**	**2**	**4**		

3	**$p \cdot q^2$**	**2**	**5**		
4	**$p \cdot (2 \cdot p \cdot q)$** Two possible trees	**3**	**6**		
5	**$p \cdot p^2 = p^3$**	**4**	**7**		

These results suggest a clear formula for the expected mean:

$$E(n) = (1 + p)^n$$

This result is intuitive; for example, if there are 100 active branches and a 10% probability of branching, at the next time step we expect 110 active branches, followed by 121 at the next time step and so on. We prove this formula analytically using the complete discrete probability distribution of possible neuronal trees in the Methods Section, where we also provide recurrence formulas and analytical solutions for the variances. The expected number of active branches, as well as the associated variances, for the first three time steps, are listed in Table 4.

Table 4 Expected mean and variance for the active (terminal) branches after N timesteps

N	Expected mean	Expected variance
1	$1 + p$	$-p^2 + p$
2	$(1 + p)^2$	$-p^4 - 2p^3 + p^2 + 2p$
3	$(1 + p)^3$	$-p^6 - 4p^5 - 5p^4 + p^3 + 6p^2 + 3p$

3.2 Model with conditional constraint

3.2.1 Neuronal growth model with a length constraint

Similar to our previous results, we assume that the family of neuronal trees is subject to a maximal length constraint since otherwise, the only optimal strategy would be to branch as much as possible, which is in contrast with experimental results. Obviously when the total length of the trees cannot exceed a set maximum value, some of the trees generated using formula cannot be instantiated. This is illustrated in Figure 7 for trees with a maximum length of 7.

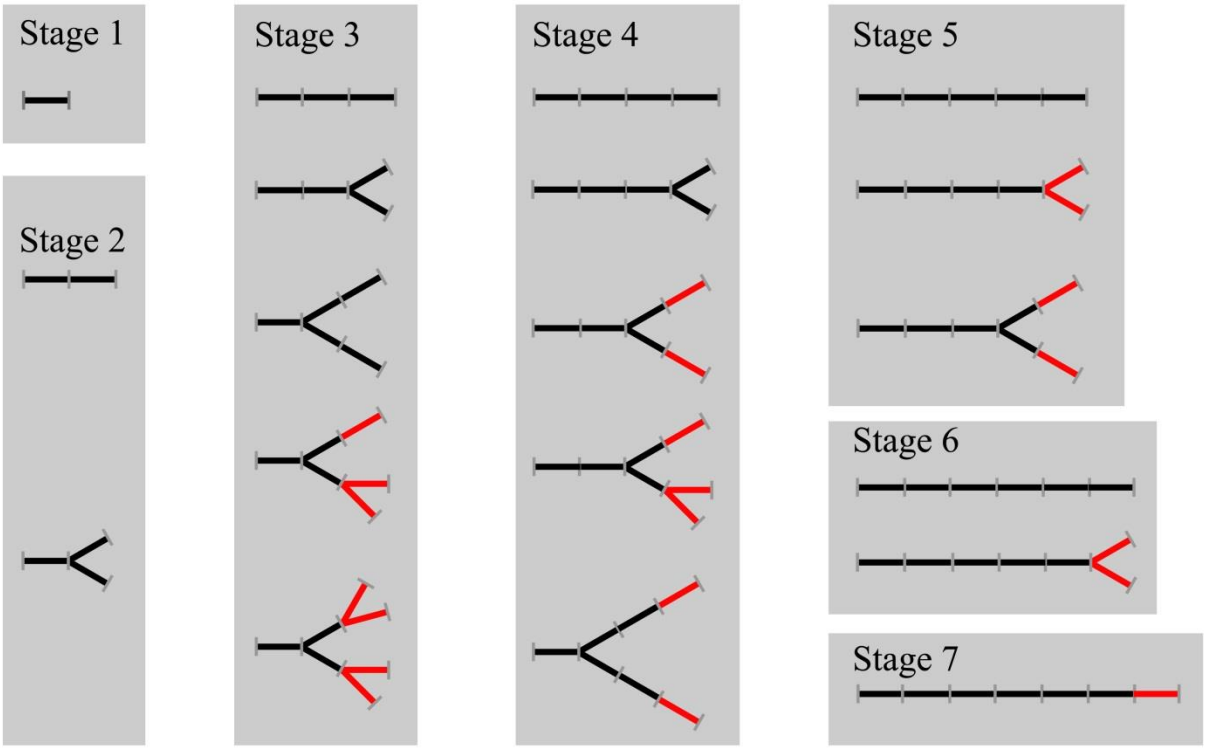


Figure 7 *All possible trees of lengths below a maximum value of 7.*

Trees that can no longer generate offspring have terminal branches shown in red. For example, the 3-terminal branch tree listed at stage 3 already has a total length of 6. Therefore, even if all its terminal branches merely extend, the resulting tree will have a length of 9, exceeding the maximal possible value of 7. All other trees will have even larger total lengths.

As a result, the expected geometrical growth is possible only near the origin, as the densest possible trees start exceeding length constraints at larger distances. In fact, growth slows down further away from the origin, reaches a peak value, and achieves a longer tail of decaying values corresponding to neural trees that seldom branch. This is shown in Figure 8, again for the family of trees that cannot exceed a total length of 7.

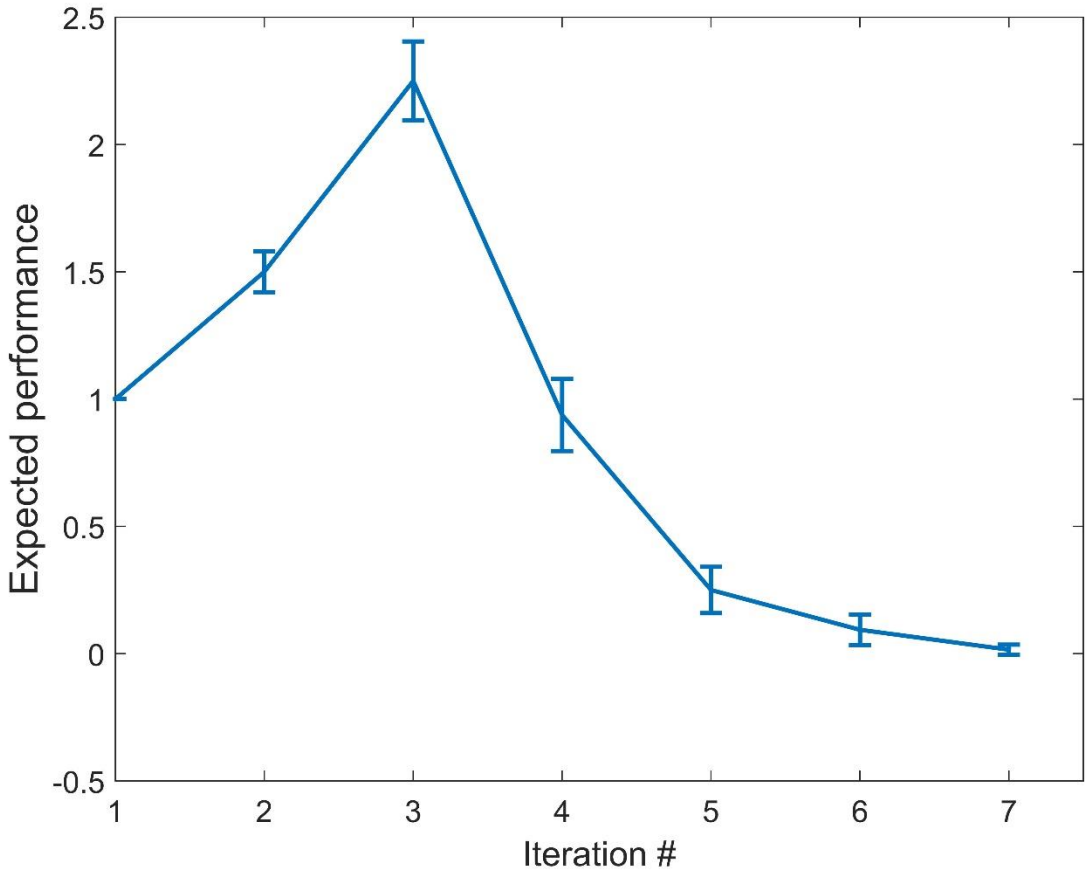


Figure 8 *The expected number of branches as a function of distance for families of trees with maximum length 7 and branching probability p = 0.5.*

Close to the origin, the expected value of searching sites increases as a geometrical series (see performances at step 2). However, this picture changes at large distances from the origin. There is an optimal targeting region, indicated by the peak at step 3, and a long tail of decaying values for steps 4 to 7. There are no possible trees for steps 8 or larger since a tree of the maximal size of 7 cannot reach beyond a total distance of 7. Only one tree, the one that does not branch at all and has a probability q^6, will reach a distance d = 7 and will have only one active branch. Error bars indicated the standard error at each iteration point and are computed for a sample size of 39 (this is the sample size of one of the experimental data sets we later used for comparison).

As illustrated in Figure 9, we can use the probability distribution for all possible instantiation of neuronal trees to determine the expected search performances of trees of maximum length L. The coverage area extends from x = 1 to L. The initial rise in performance follows a geometrical series trend that is determined by the exact value of the branching parameter p. The performance at large distances away from the origin has a long tail, corresponding to the increasingly smaller number of trees that can make it further and further. In terms of the exact values of branching probability, trees that branch often will tend to cover the nearby area well and will have very small chances of extending far away. In contrast, trees that seldom branch will have a much-improved chance to explore farther in space, albeit while sending out fewer branches. It is then intuitive that the optimal targeting region is determined by the exact value of the branching probability parameter p, as illustrated in Figure 9. Here the average number of dendrites at different time steps is shown for different values of branching probability p. This is essentially the equivalent of a Sholl plot for neuronal trees. [53,54].

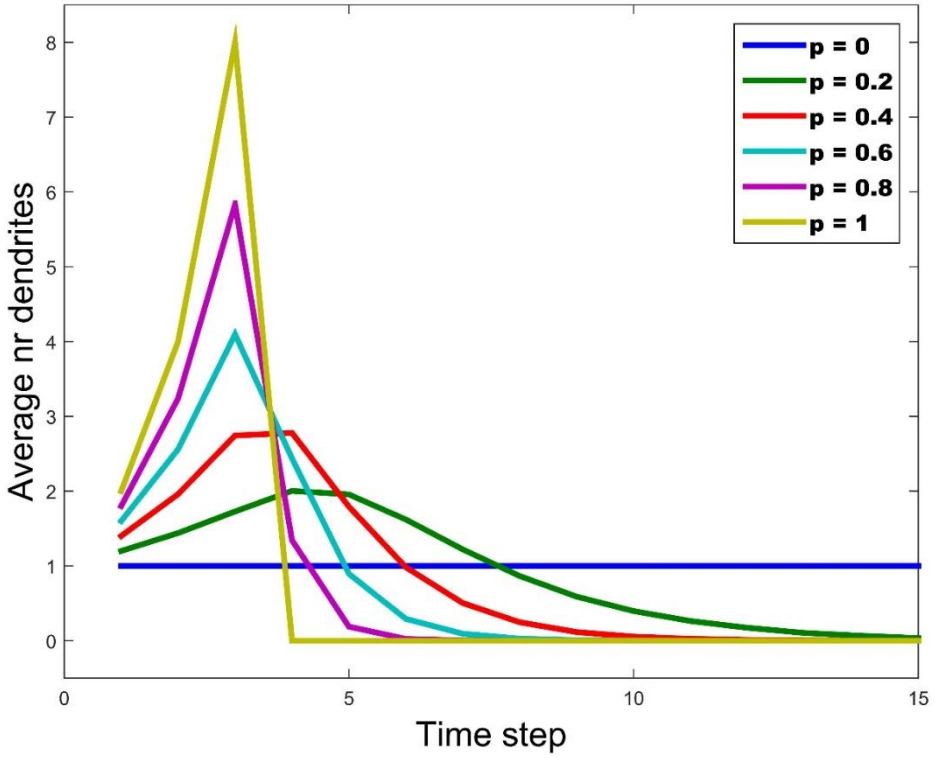


Figure 9 *Sholl plots for different branching probabilities p.*

Expected number of dendrites at each time step is plotted for a tree of maximum possible length of 16, for the branching probabilities in the set {0, 0.2, 0.4, 0.6, 0.8, 1}. Since the tree starts with one branch, it can only extend for a maximum of 15 extra number of steps. The tree with $p = 0$, shown in blue, simply extends until it runs out of resources and generates a flat line of 1 expected branch at each time step. In contrast, the tree with $p = 1$ doubles the number of branches at each time step and runs out of resources at step 3, achieving a maximum at the time it stops. In between these two extremes, families of trees shift their optimal targeting regions farther away as the branching probability decreases, at the cost of reducing the overall amplitude of the corresponding peak values (success rates).

3.2.2 Long-range expected performances are also characterized by geometrical series

Surprisingly, the right tail of the probability distribution also follows a geometrical distribution with a factor of $(1 + 3p)$, although higher order corrections are needed when p takes on larger values. This trend can be determined by examining the trees that have the furthest reach, which

is listed in Figure 10.

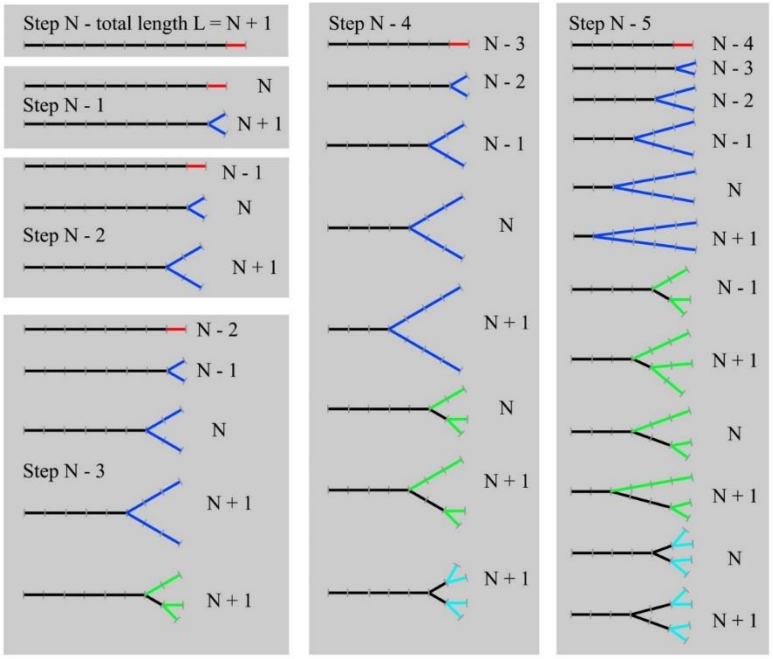


Figure 10 *Illustration of trees generated at very long distances.*

Figure 10: We use the longest possible tree to generate all the trees that can be instantiated at the previous time steps. For the cases listed here, only trees with a small number of terminal branches, 1 to 4, can be instantiated, as the trees are back-generated for 5 time steps.

Using Figure 10, it is easy to see the expected number of branches at large distances, computed under the assumption that we ignore trees with more than 4 terminal branches, follows the recursive structure listed in column 2 of table 5. This approximation becomes more and more accurate as the branching probability p goes to 0. Consequently, the ratio of performances at successive time steps (going backward) is approximately $1 + 3p$, as shown in column 3 of table 5.

Table 5 Approximate geometrical series for the right tail of the probability distribution.

Step	Expected number	Ratio $\frac{P(k)}{P(k+1)}$
N	q^N	
N – 1	$q^{N-1} + 2pq^{N-2}$	$1 + 3p + 5p^2 + 7p^3 + 9p^4 + 11p^5 + 0(p^6)$
N – 2	$q^{N-2} + 2pq^{N-3} + 2pq^{N-2}$	$1 + 3p - p^2 + 3p^3 - p^4 + 3p^5 + 0(p^6)$
N – 3	$q^{N-3} + 2pq^{N-4} + 2pq^{N-3} + 2pq^{N-2}$	$1 + 3p - 7p^2 + 29p^3 - 99p^4 + 357p^5 + 0(p^6)$
N – 4	$q^{N-4} + 2pq^{N-5} + 2pq^{N-4} + 2pq^{N-3} + 2pq^{N-2}$	$1 + 3p - 13p^2 + 85p^3 + 507p^4 + 3073p^5 + 0(p^6)$
N – 5	$q^{N-5} + 2pq^{N-6} + 2pq^{N-5} + 2pq^{N-4} + 2pq^{N-3} + 2pq^{N-2}$	$1 + 3p - 19p^2 + 171p^3 - 1447p^4 + 12341p^5 + 0(p^6)$

Summarizing the results from table 5, for very small branching probability p, we obtain the following approximation:

$$\frac{P(N-2k-1)}{P(N-2k)} = 1 + 3p + (5 - 6k)p^2 \text{ , } \frac{P(N-2k-2)}{P(N-2k-1)} = 1 + 3p + (5 - 6(k+1))p^2 \qquad (1)$$

This approximation is correct when discounting trees with more than two terminal branches, as these trees will introduce corrections of order p^2 and larger. We note while high values of the branching probability p render this approximation more and more inaccurate, in fact, low values of p are needed for the neural trees to be able to reach large distances; therefore, this is a useful approximation.

3.2.3 Approximating the location of optimal targeting performances

We can now use the formula for long-range expected performances to compute an approximation for the location of optimal targeting region. The ratio of performances that is initially larger than 1 keeps decreasing as one moves away from the longest possible tree of length L toward and closer to the origin. Eventually, this ratio, again based on extensions of formula (1), will become 1 before further decreasing at distances closer and closer to the origin. The exact location where the ratio becomes 1 can be used to approximate where the spatial location of the peak of the distribution occurs. This occurs at step N – k, when $\frac{P(N-k-1)}{P(N-k)}$, which starts with an initial value of $(1 + 3p + 5p^2)$, reaches a value of 1 after experiencing continuous decreases in value. This is determined by the following equation:

$$1 + 3p + (5 - 6k)p^2 = 1$$

which can be easily rewritten as $3 + (5 - 6k).p = 0$, yielding the following value of the peak:

$$k = \frac{3 + 5p}{6p}, 6pk = 3 + 5p, (6k - 5)p = 3, p = \frac{3}{(6k - 5)}$$

A recurrence formula that takes into account terms up to order p^3 is provided in the Methods Section. When allowing for correction terms of orders up to p^3, only numerical solutions for this equation can be obtained since the coefficients have relatively complex recurrence formulas.

We now list another way of approximating the optimal targeting region. In previous study [2], we derived an estimate of the location of optimal targeting performances by examining a tree with neurites that branches periodically until growing a fixed length L_0, which is determined by the branching probability p. After computing how L_0 depends on p, we now can determine the optimal targeting region after n steps:

$$n = \frac{\ln(1 + pL)}{\ln(1 + p)} - 1$$

Derivation of this equation is listed in the Methods Section. This equation determines the optimal targeting region for a tree with branching probability p. Note that here n is a continuous function of p since the branching probability can take continuous values. This 'naïve' approximation is in

very good agreement with the results from the discrete probability distribution obtained by instantiating all possible trees that do not exceed the maximal value L up to time step n, as shown in Figure 11.

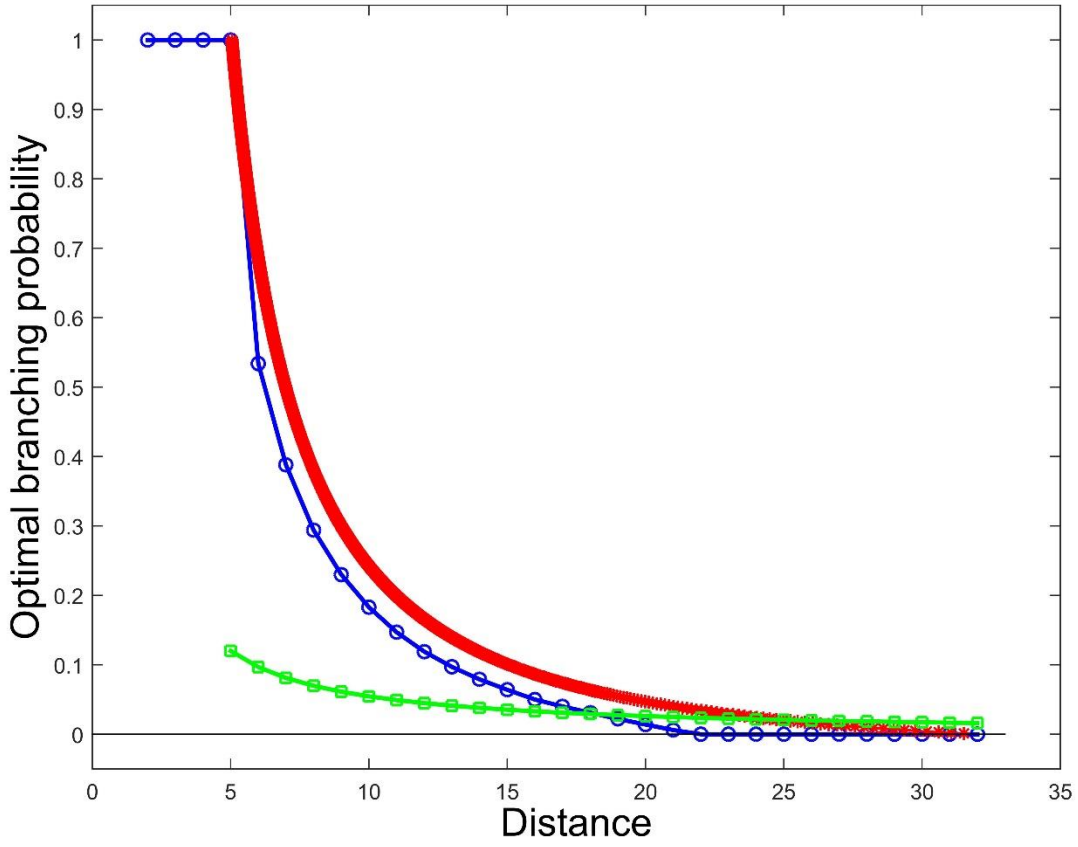


Figure 11 *Comparison of naïve vs exact results for a tree of maximum length L = 31.*

We use the explicit discrete probability distribution to determine the optimal targeting regions, shown as blue circles, while the red line is given by equation$n = \frac{\ln(1+pL)}{\ln(1+p)} - 1$. For the discrete distribution, since the total length of the tree is 31, there are enough resources for all possible trees until step 4. Therefore, the optimal branching probability is$p = 1$, corresponding to creation of the densest trees. As a result, steps 1-4 are excluded from the comparison. The blue points are computed using the explicit symbolic equation (polynomials in$p$) for the expected value s_k at different locations$k < n$. Then the solution for the probability p that maximizes the expected value $s_k(p)$ is computed and plotted at location k as a blue circle. Values for p are restricted to be between 0 and 1 to exclude other potential solutions for the polynomial equations. The graph indicates that while the red curve obtained from the 'naïve' theoretical expectation overestimates the analytical expected results, these curves are in agreement and

exhibit the same trends. Furthermore, as expected, the approximation used in equation $p = \frac{3}{(6k-5)}$, shown here in green, works relatively well at large distance, but it loses accuracy at distances close to the origin. Nevertheless, higher order terms need to be used in order to obtain a more useful approximation.

3.3 Comparison with lab data

3.3.1 Comparison of the theoretical and experimental results for neural growth

In order to assess the validity of our model, we performed a comparison with experimental data from neuronal growth tissue from the Firestein laboratory. An illustration of two sample neurons is provided in Figure 12. We performed Sholl analysis for the family of neurons used here in order to generate the targeting profiles. We fit our models to these profiles by allowing changes in the following parameters: spatial distance covered in between two potential branching events (width of the distribution), the number of initial branches (peak of the distribution) and branching probability/total length. We were able to produce very accurate fits for the Sholl analysis curves (Fig 12), indicating that our model is in very good agreement with the experimental data.

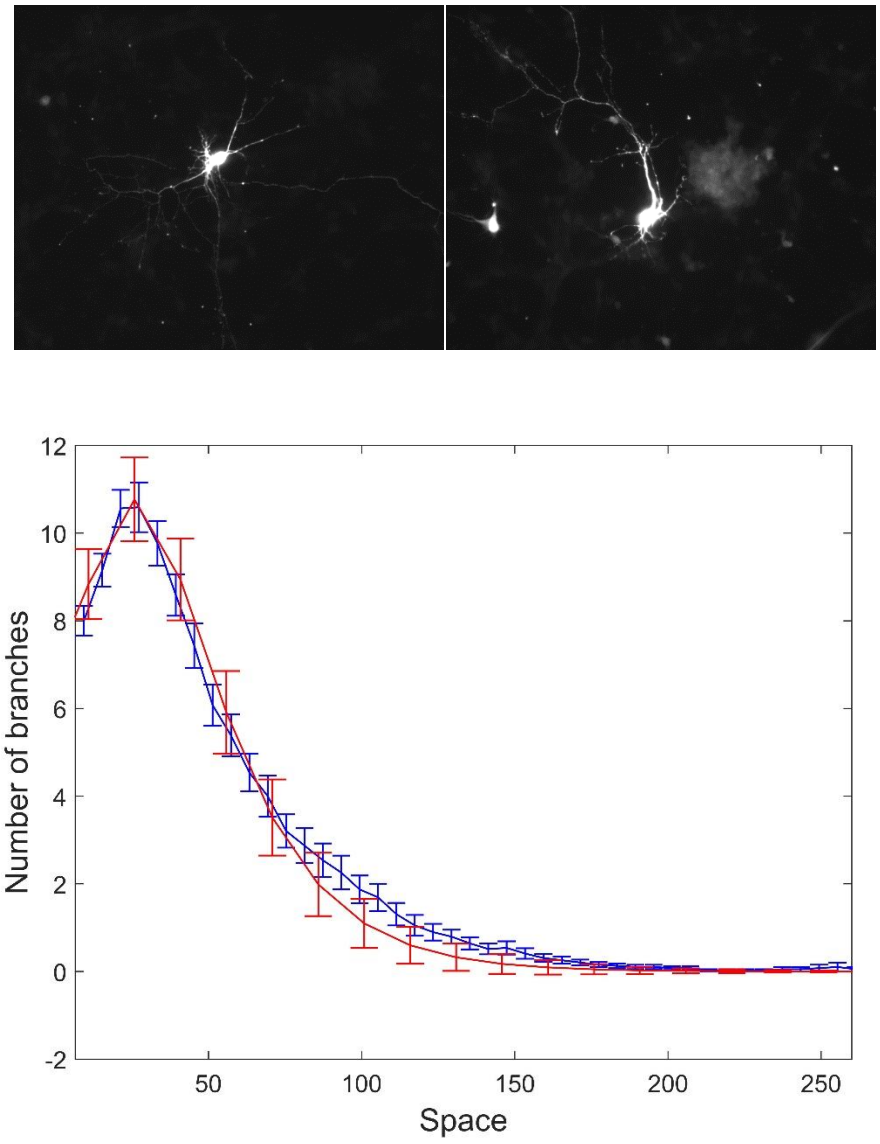


Figure 12 *Comparison of experimental and theoretical results.*

Arborization profiles for two example neurons (A and B) result in typical profile curves (blue curve in panel C) for this family of neurons (n = 30). Fits of the probabilistic neuronal growth model are in excellent agreement (red curve in panel C).

Our model suggests that the logarithm of the number of active branches is a more appropriate way of constructing Sholl plots, as the average function of the neuronal tree population is predicted to have a linear component near the origin and approximate linear component at large distances. The use of both of these components can provide a better statistical determination of branching values. Also, if the large-distance component is in disagreement

with the near-origin component, this may indicate other effects, such as pruning at larger distances or parameter-dependent branching.

With our model, we are able to connect branching parameters with real branching probabilities and also reflect the effect of chemicals on branching probability, such as Cypin , GFP and NOS..

- Cypin: a guanine deaminase that increases dendrite number by binding to tubulin heterodimers and promoting microtubule assembly [55]. In our study, it is taken as a chemical that increases the branching probability.
- Green fluorescent protein (GFP): a protein composed of 238 amino acid residuals that exhibit bright green fluorescence when exposed to light in the blue to ultraviolet range. Iin our study, it is used as a control group [56].
- Nitric Oxide synthase: a family of enzymes catalyzing the production of nitric oxide from L-arginine [57]. In our study, it is taken as a chemical that decreases the branching probability.

In order to accomplish this task, we first tried to build up our model with data from our collaborator to estimate the branching probability. We set the total length to be 34, due to the limitation of our capacity in generating all possible trees Results show below:

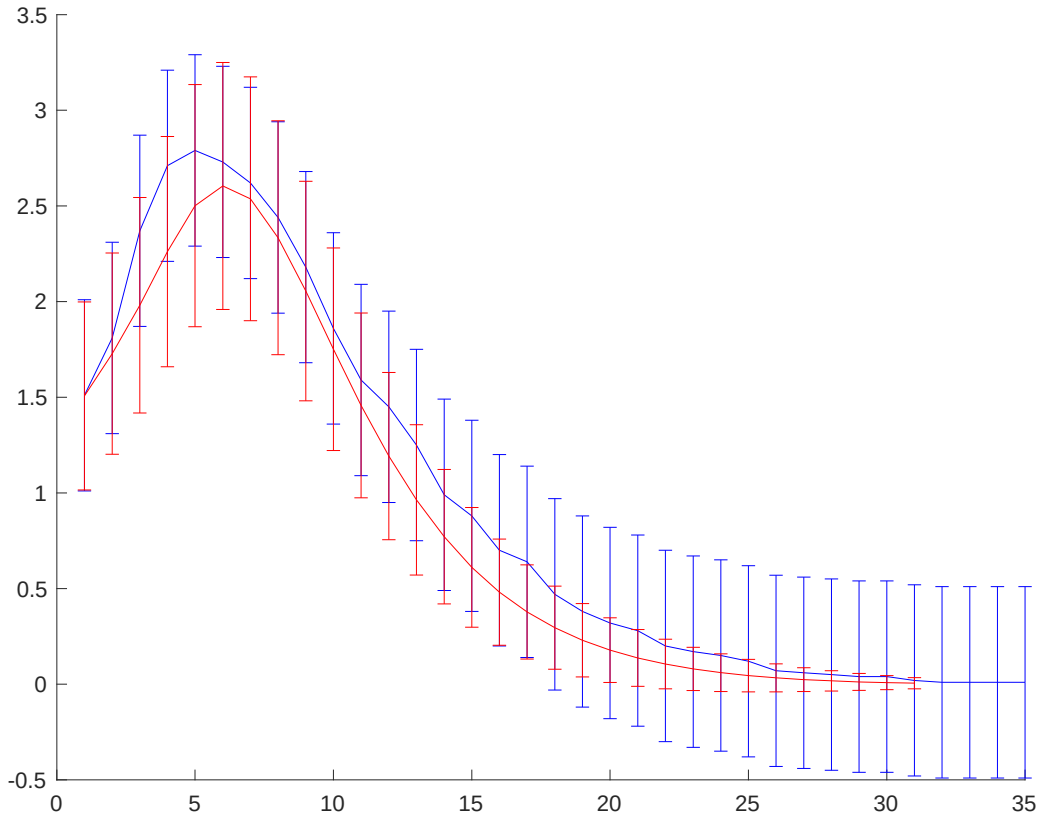


Figure 13 *Results from estimating the branching probability*

Figure13: The figure shows the comparison between the expected value from our model (in blue) and the real data (in red). The estimate of branching probability was obtained when we have the least difference between the model and real data. Meanwhile, horizontal shift was applied to the model, in order to have the proper initial number of branches. We can see that their confidence intervals overlap, which means that our model fits well with the real data.

We also have the ANOVA result:

Table 6 ANOVA results from estimating branching probability

Nonlinear Regression Model:				
$y \sim fit\ (branching\ probability, Step)$				
Estimated Coefficients:				
	Estimate	SE	tStat	P-value

Branching probability	0.1465	0.0059766	24.512	2.127e-21
Number of observations: 31, Error degrees of freedom: 30				
Root Mean Squared Error: 0.186				
R-Squared: 0.965, Adjusted R-Squared: 0.965				
F-statistic vs. zero model: 1.83e+03, P-value: 1.9e-28				

From the results, we can see that our model fits well and R-square looks promising for applying the model for new data. Then we tried to compare the branching probability between neural growth processes controlled by different chemicals, which influence the branching probability.

We first compared NOS vs GFP, expecting a reduction of branching probability between NOS condition and control (GFP), [55-57]:

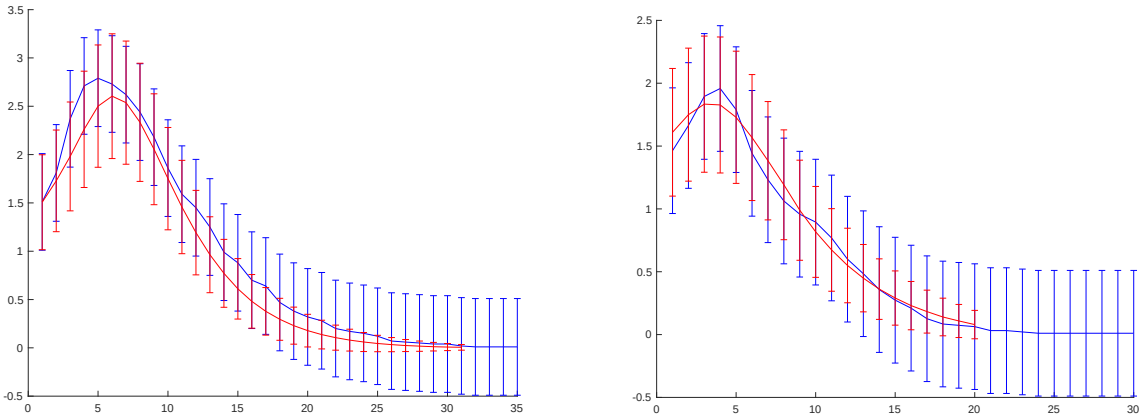


Figure 14 Comparison between NOS vs GFP

Figure 14: comparison between two regeneration controlled by NOS (left) and GFP(right). With our model, we estimate the branching probability to obtain the best fit. We also move our model horizontally with different units, according to the different initial number of branches. We also have the comparison between the estimate two branching probabilities listed below

Table 7 Comparison between branching probabilities of two processes

	NOS			GFP		
	Estimate	95% Confidence Interval			95% Confidence Interval	
Branching probability	0.1	0.0962	0.1038	0.1465	0.1405	0.1525

Then, we compared Cypin vs GFP:

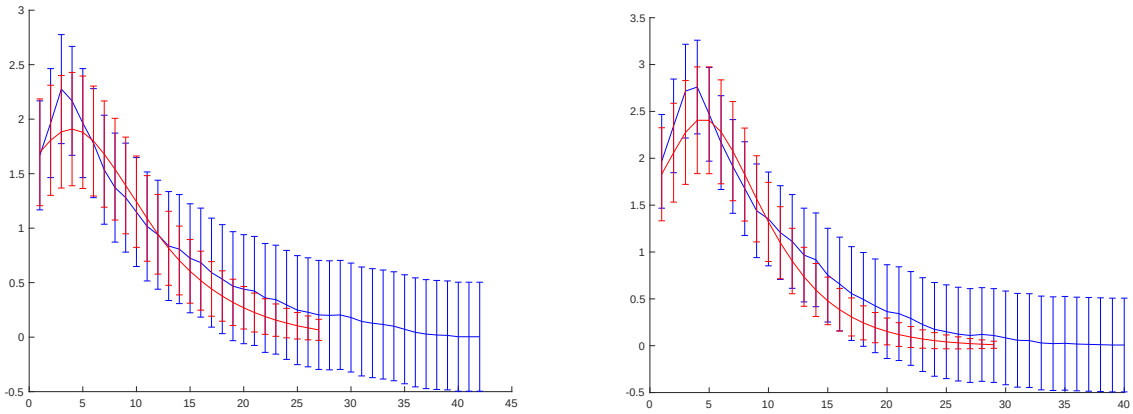


Figure 15 *Comparison between GFP and Cypin*

Figure 15: comparison between two regeneration controlled by GFP (left) and Cypin(right). Same as first pair, we obtain the estimate the branching probability based on the goodness of fit. Horizontal shift was also applied here to have the proper initial number of branches. We also have the comparison between two estimate of branching probabilities listed below

Table 8 Comparison between branching probabilities of two processes

	GFP		Cypin	
	Estimate	95% Confidence Interval		95% Confidence Interval

Branching probability	0.0788	0.0708	0.0868	0.1284	0.1216	0.1352

So from the results above, we can see that the differences between branching probabilities are significant, none of the confidence intervals overlap. This tells us that our model is capable reflecting the effect from these medicals on the branching probability. For example, Cypin was expressed early in culture and continued to be expressed as neurons matured [58 in invivo]. So it is expressed in developing dendrites and increases the branching probability, which is in accordance with our results in table 8. The branching probability of process with Cypin is significantly higher than the other.

In the future, we aim to test our model on processes controlled by same chemical substances, but with different concentrations, as experimental data becomes available.

3.4 Results for Tumor growth rate

3.4.1 Tumor volume and growth rate:

Tumor volume at the time of screening ranged from 33-25918mm3 (mean=1382.7mm3 and median 472.29mm3). This contrasted with tumor volume at diagnosis, which was found to range from 9 to 37819mm3 (mean=4591.74mm3 and median =2365.80mm3).

Overall, change in volume between screening and diagnosis in each patient varied significantly. Six cases showed no change in size, while the largest change in tumor volume was 18438.1mm3. The mean time difference between dates of screening and that of diagnosis was 18 months, (median = 17.5 months), and ranged from 4 to 47 months.

The growth rate was found to differ considerably from patient to patient, ranging from 0 to 1390 mm3/month. (Figure 18) Tumors with a growth rate of <92mm3 per month were deemed slow-growing and comprised 47 cases (49.5%). In contrast, those with a growth rate of ≥92mm3 per month (n = 48, 50.5%) were placed in the fast-growing group.

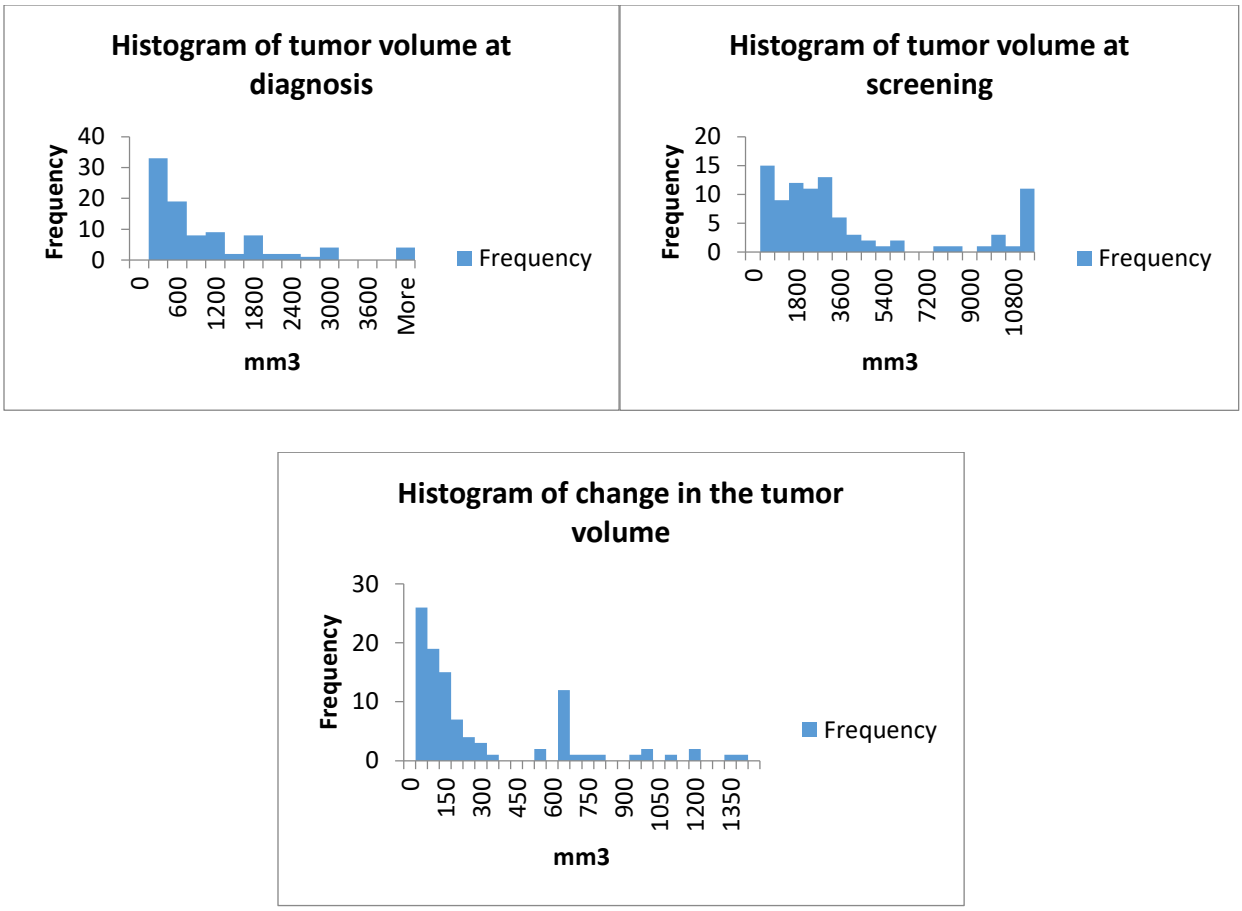


Figure 16 *Histogram of tumor volume at diagnosis and screening and also histogram of the change in tumor size*

Age at the time of diagnosis ranged from 50 to 73 years of age (mean = 60.3 years, median = 61.0 years). Most cases (n=60) were negative for vascular invasion (66.2%), whilst 21 (22.8%) were found to demonstrate definite vascular invasion, and 11 (11.0%) were rated as probable. The majority of tumors (60, 65.2%) were stage 1, 24 were stage 2 (26.1%), and 7 cases were stage 3 (8.7%). 16 tumors were graded as 1 (17.4%), using the Nottingham Grading System, 42 cases (45.7%) were grade 2, and 34 (36.9%) grade 3. DCIS was not observed in 21 cases (22.8%). Of the remaining 73 cases, 20-showed low-grade DCIS (21.7%), 16 were of intermediate grade (17.4%), 5 were intermediate/high grade (5.9%), and the remaining 29 high grade (31.5%). The mean NPI value was 4.00, with a range of 2.08 to 6.56 (median = 4.01). The clinicopathologic features of patients are summarized in the following table 9.

Table 9 Clinicopathological features of cases

Parameters	Number of cases (N)	Percentage (%)
Age		
≤65	74	77.9
>65	18	22.1
Tumor Grade		
1	16	17.4
2	42	45.7
3	34	36.9
Clinical stage		
I	60	65.2
II	24	26.1
III	7	8.7
HORMONE STATUS		
ER Positive	78	84.8
PR Positive	59	64.1
HER2 Expression		
Positive	5	5.85
Negative	81	94.2
Tumor type		
NST	50	54.3
LOB	17	18.5
TUB	11	12.0
MUC	2	2.2
MIX	12	13
Vasicular Invasion		
NEG	60	66.2
DEFINITE	21	22.8
PROBABLE	11	11.0
KI67		
High	44	47.8
Low	48	52.2
CK5/6		
Negative	80	87
Positive	12	13
Caspase3		
0	45	49
≥1	47	51
Subtypes		
Luminal	69	75
Her2	8	8.7
Basal	4	4.3
TN	11	12

Vital Status		
Alive	62	67.4
Dead	30	32.6

Tumor Biomarker Profiles

Of the 92 cases, 78 cases were positive for ER (84.8%) and 59 cases were positive for PR (64.1%) while only 5 cases were positive for HER-2 based on protein expression (5.8%). The degree of Ki67 staining ranged from 0 to 96% of tumor cells stained. In the case of CK5/6, 80 cases (87%) were negative, and just 12 (13%) showing positive staining. Cleaved caspase-3 immunoreactivity ranged from 0 to 6.0% of tumor cells (Table: 10).

3.4.2 Results from neural growth model selection

Here we have the results from three models.

The first model is linear model, with formula:

$$V_t = \alpha * \Delta t + V_0$$

Then, we fit the linear model based on the formula and estimate the growth rate.

The results from linear model:

Table 10 ANOVA results from linear model

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	58924223	58924223	2.57	0.1123
Error	90	2062334567	22914829		
Corrected Total	91	2121258789			

Root MSE	4786.94355	**R-Square**	0.0278
Dependent Mean	3208.99969	**Adj R-Sq**	0.0170
Coeff Var	149.17245		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	1649.30376	1093.20595	1.51	0.1349
input_linear	**1**	86.64977	54.03545	1.60	0.1123

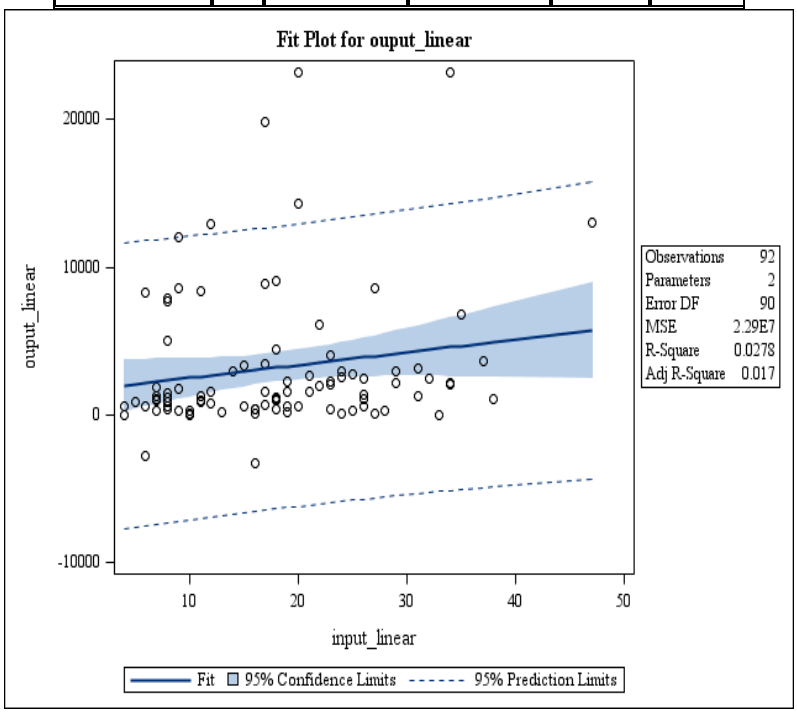


Figure 17 *Fit plot of linear neural growth rate model*

From the results, we can see that the p-value of F-test is greater than 0.05, which means that the model is not good at overall fitting. The r-squared is 0.02, which means that the model only explains 2% of the variation of output. This indicates that the linear model might not be a good choice for the neural growth rate.

We also have the model diagnostic results as shown in figure 18.

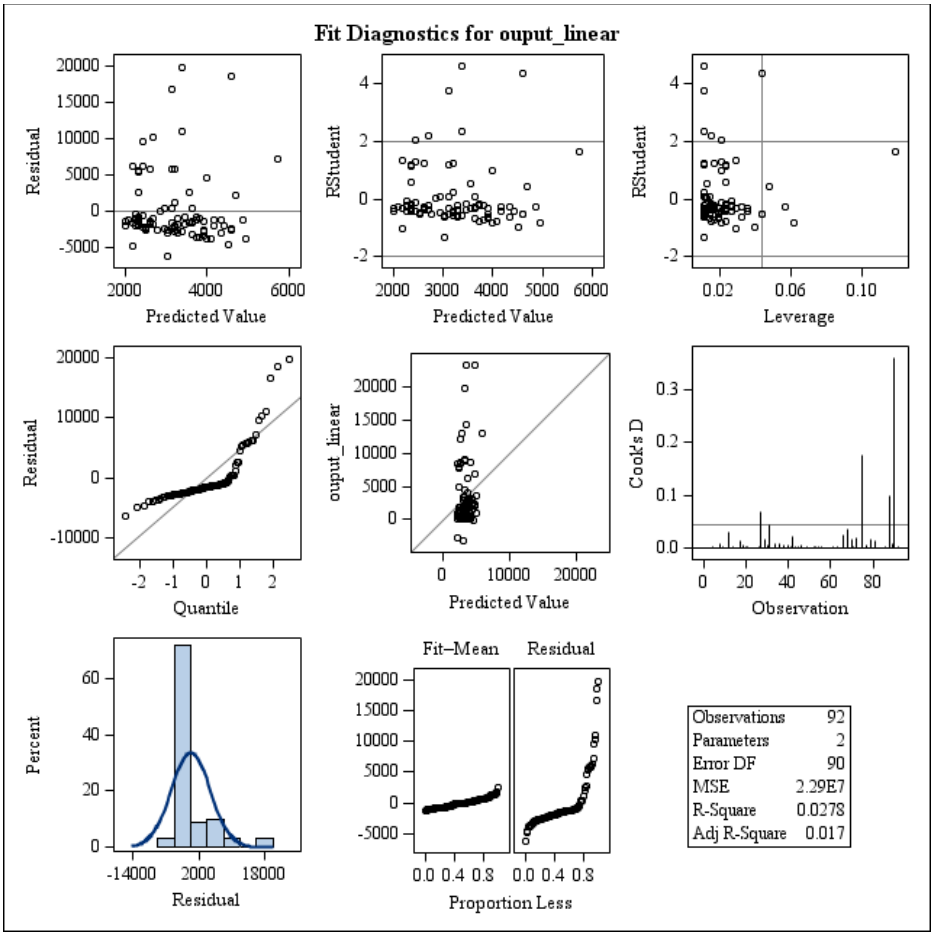


Figure 18 *Diagnostic results of linear neural growth model*

Figure 18 is the results from model diagnosis. From the top residual plots, we can see that the residuals are not normally distributed and also have clear pattern as the predicted value increases, which mean the model can be improved. Also from the qq-plot and residual histogram, we can see that the model violates several modelling assumptions. Therefore, the linear model is not a good choice to predict the growth rate.

The second model we tried is exponential model with the formula:

$$V_t = V_0 * e^{\alpha * \Delta t}$$

Before we fit the model to estimate the growth rate, we did the transformation to have the linear form as:

$$\ln\left(\frac{V_t}{V_0}\right) = \alpha * \Delta t$$

Then we have the results from exponential model:

Table 11 ANOVA results from exponential neural growth model

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	15.08707	15.08707	15.87	0.0001
Error	90	85.55486	0.95061		
Corrected Total	91	100.64193			

Root MSE	0.97499	**R-Square**	0.1499
Dependent Mean	1.47208	**Adj R-Sq**	0.1405
Coeff Var	66.23239		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	0.68286	0.22266	3.07	0.0029
input_exp	**1**	0.04385	0.01101	3.98	0.0001

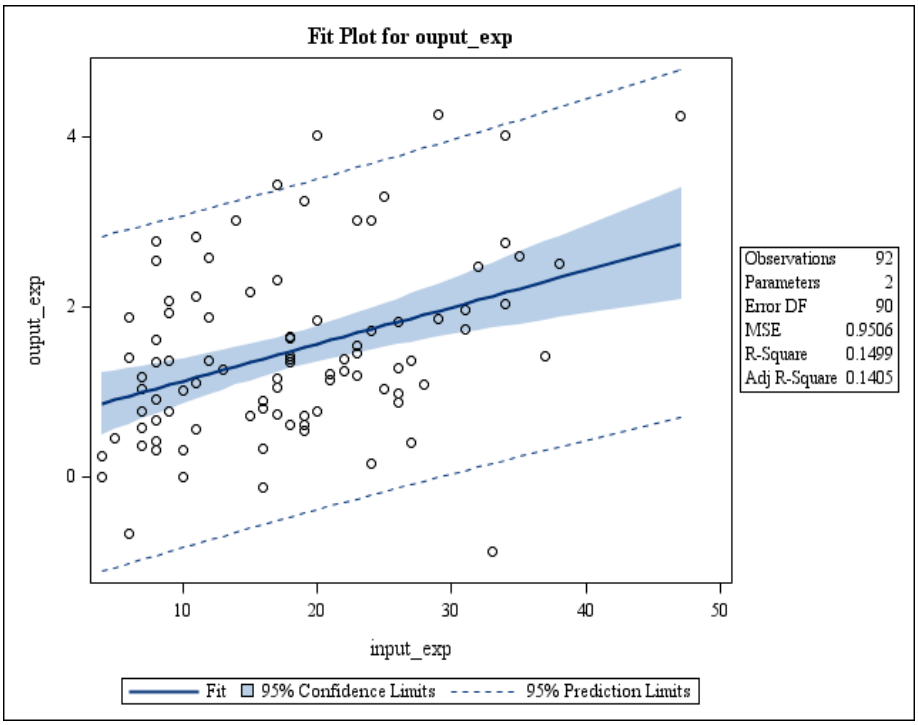


Figure 19 *Fit plot for exponential neural growth rate model*

From the results, we can see that the p-value of F-test is less than 0.0001, which means that the model is good at overall fitting. The r-squared is 0.1499, which means that the model only explains 14% of the variation of output.

We also have the model diagnostic results as shown in figure 20.

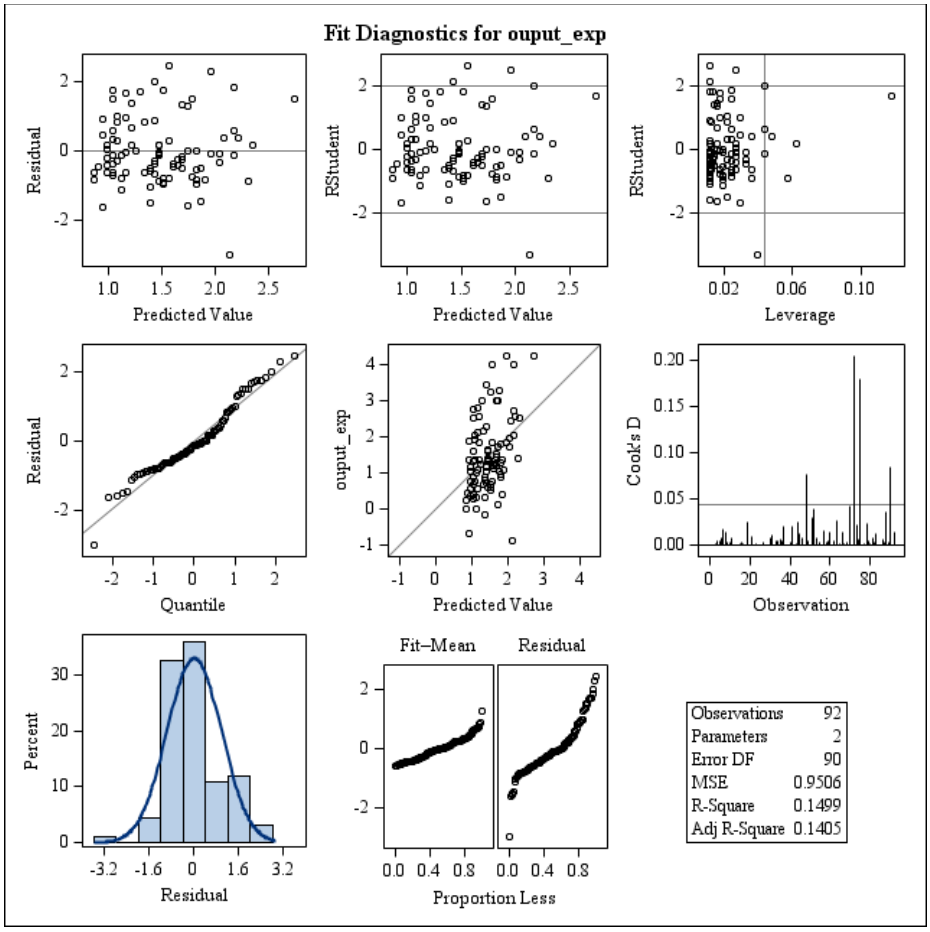


Figure 20 *Diagnostic results of exponential neural growth model*

Figure 20 is the results from model diagnosis. From the top residual plots, we can see that there are not clear patterns in the residual plots, which mean the model has good robustness. Also from the qq-plot and residual histogram, we can see that the model meet most of modelling assumptions. Therefore, the exponential model is a good choice to predict the growth rate.

The last model we tried is Gompertz model:

$$V_t = V_0 * e^{1-\frac{\frac{\alpha}{\Delta T}}{k}}$$

We also transformed it into linear form as:

$$\ln\left(\frac{V_t}{V_0}\right) = \frac{\alpha}{1 - \frac{\Delta T}{k}}$$

The result from fitting Gompertz model:

Table 12 ANOVA results from Gompertz neural growth model

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	12.51102	12.51102	12.78	0.0006
Error	90	88.13091	0.97923		
Corrected Total	91	100.64193			

Root MSE	0.98956	**R-Square**	0.1243
Dependent Mean	1.47208	**Adj R-Sq**	0.1146
Coeff Var	67.22212		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	**1**	1.90810	0.15976	11.94	<.0001
input_third	**1**	0.39765	0.11125	3.57	0.0006

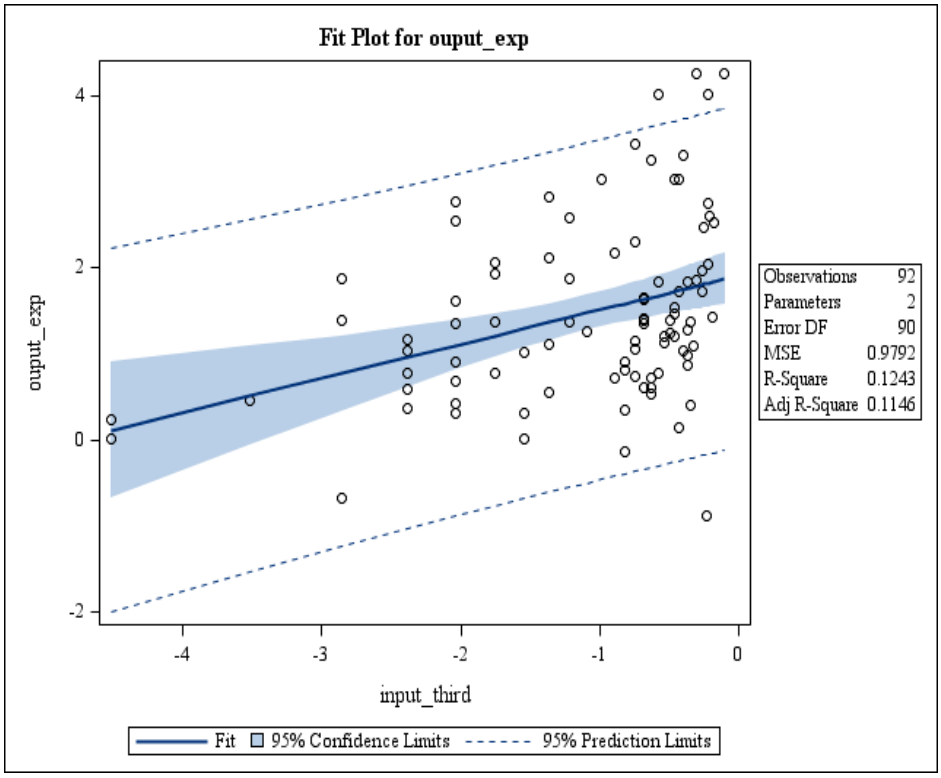


Figure 21 *Fit plot of Gompertz neural growth model*

From the results, we can see that the p-value of F-test is 0.0006, which means that the model is good at fitting. The r-squared is 0.1243, which means that the model only explains 12% of the variation of output.

We also have the model diagnostic results as shown in figure 22.

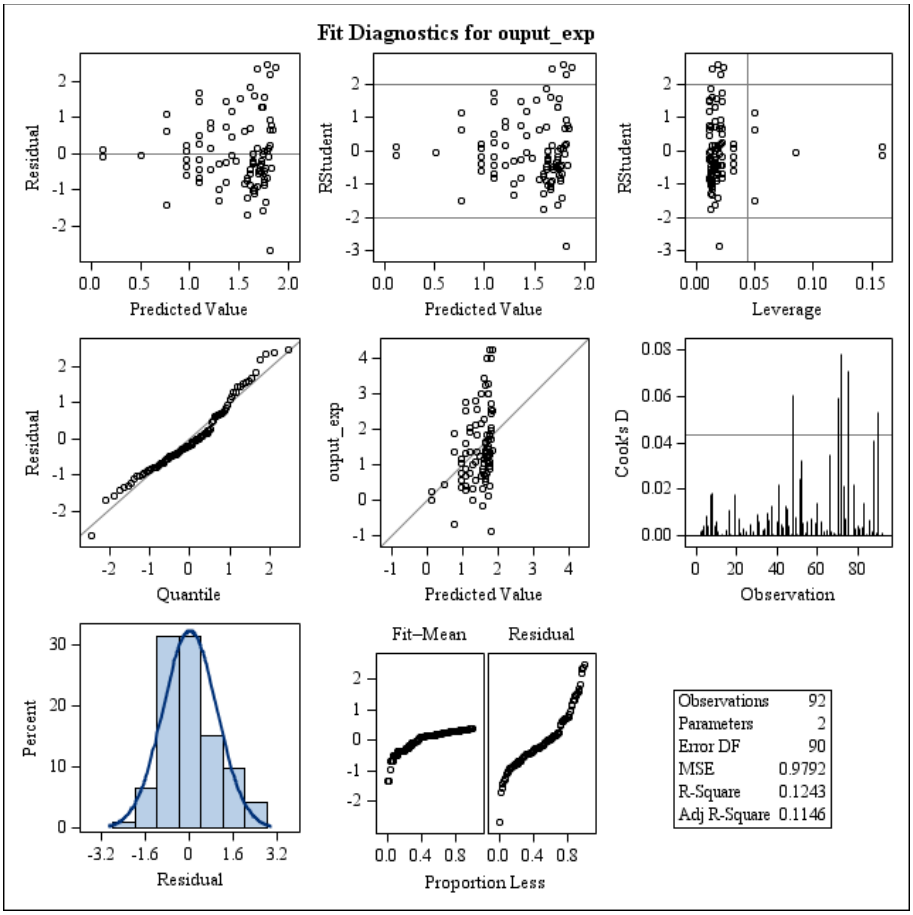


Figure 22 *Diagnostic results of Gompertz neural growth model*

Figure 22 is the results from model diagnosis. From the top residual plots, we can see that there are clear expanding patterns in the residual pattern as the prediction increases, which mean the model is good in predicting. But from the qq-plot and residual histogram, we can see that the model meets the modelling assumptions. In sum, the Gompertz model is the best choice to predict the growth rate.

Based on the results above, we have the conclusion that exponential model is the best one among the three candidates.

3.4.3 Correlation study on growth rate and biomarkers

We also tested the correlation between the growth rate and biomarkers, trying to define the effect of the biomarkers on determining the growth rate. By doing this, we will test the hypothesis that patients at higher cancer stage tend to have faster tumor growth rate. What is

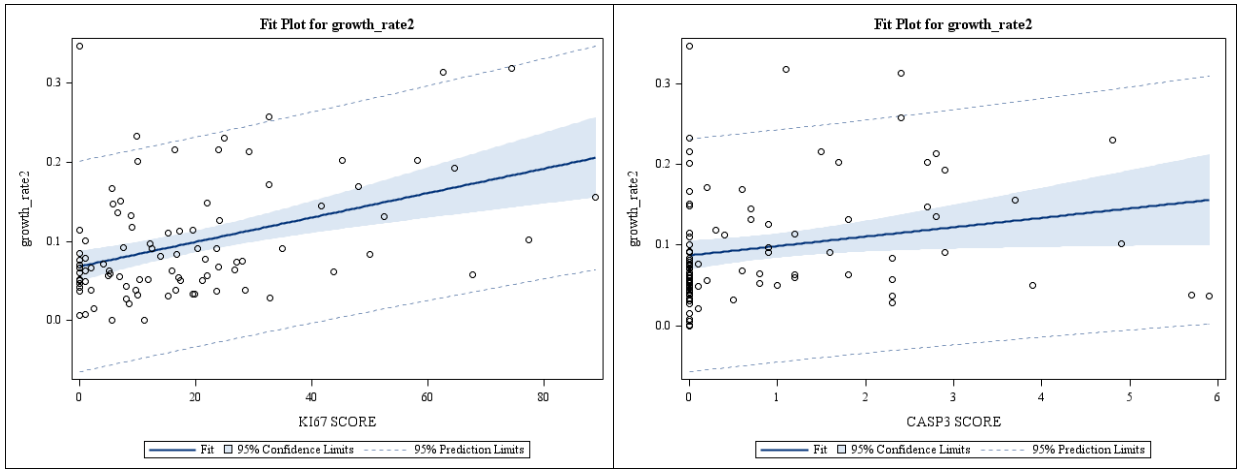
more, this study will help us to choose the biomarkers to build the model predicting the tumor growth rate.

Correlations between Other Clinicopathological Parameters and Growth Rate:

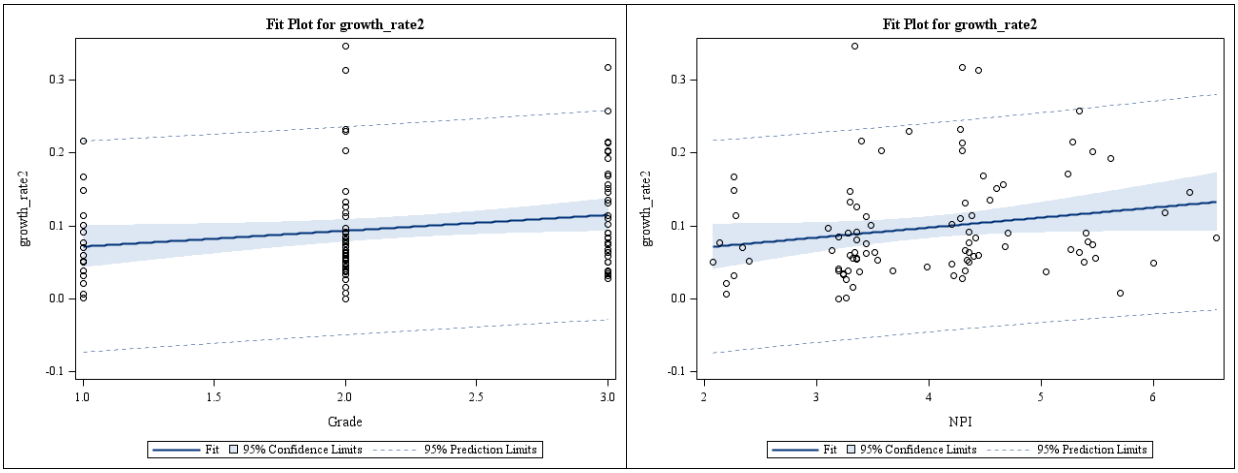
The Growth rate in breast cancer was significantly and positively correlated with tumor grade (p = 0.0247), histological size (p=0.001) and vascular invasion (p = 0.011). When sub-dividing tumor grade into its 3 components, the degree of tumor differentiation was found to have no significant relationship with growth rate (p = 0.486), however, mitotic index was significantly correlated (p = 0.0032). Stage showed no significant relationship with growth rate (Table: 13).

Correlations between Biomarker Expression, clinicopathological parameters, and Growth Rate :

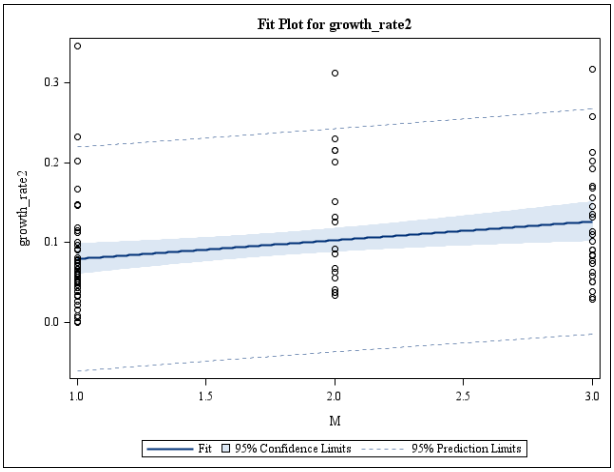
A highly significant correlation was found between the proliferative marker Ki67 and tumor growth rate. (Pearson correlation coefficient=0.36457; p=0.0004). Higher growth rate was also positively correlated with increasing Mitosis (Pearson correlation coefficient=0.31, p=0.0019), higher Grade (Pearson correlation coefficient=0.23411, p=0.0247), NPI (Pearson correlation coefficient=0.27698, p=0.0075) and Caspase3 (Pearson correlation coefficient=0.30380, p=0.0373) as shown in figure 23.



(a) (b)



(c) (d)



(e)

Figure 23 *Significant correlation between tumor growth rate with (A) Ki67; (B) Caspase3; (C) Grade; (D) NPI and (E) Mitosis*

Correlation among tumor biomarkers: A highly significant relationship was observed between cleaved Caspase-3 and Ki67 expression, with the two being positively correlated (p = 0.001, Pearson's correlation coefficient = 0.259, Figure 24).

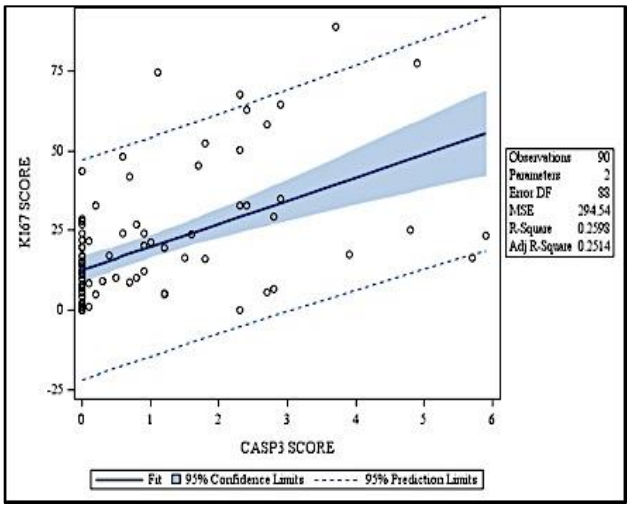


Figure 24 *Relationship between Ki67 and Caspase-3 expression*

We also have

Table 13 Relationship between growth rate and biomarkers

Variables	Growth rate					
	Low (<92 mm^3/month)		High (≥92mm^3/month			
	#	%	#	%	**X2**	P-VALUE
Grade						
1	4	8.16	12	27.91	11.2813	**0.0036**
2	20	40.82	22	51.16		
3	25	51.02	9	20.93		
Total	49		43			

T						
1	2	4.08	2	4.65		
2	14	28.57	19	44.19	2.5772	0.2757
3	33	67.35	22	51.16		
Total	49		43			
P						
1	0	0.00	3	6.98		
2	18	36.73	21	48.84	5.7439	0.0566
3	31	63.27	19	44.19		
Total	49		43			
M						
1	18	36.73	30	69.77		
2	10	20.41	7	16.28	11.5204	**0.0032**
3	21	42.86	6	13.95		
Total	49		43			
TYPE						

LOB	11	22.45	6	13.95		
MED	0	0.00	0	0.00		
MIX	5	10.20	7	16.28	5.5343	0.2367
MUC	0	0.00	2	4.65		
NST	29	59.18	21	48.84		
TUB	4	8.16	7	16.28		
Total	49		43			
STAGE						
1	28	57.14	33	76.74		
2	17	35.42	7	16.28	4.6488	0.0978
3	4	8.33	3	6.98		
Total	49		43			
Vi						
Negative	24	48.98	36	83.72		
Probable	7	14.29	4	9.30	13.599	**0.0011**
Definite	18	36.73	3	6.98		
Total	49		43			

Histological size						
≤15mm	8	16.32	24	55.81	14.7778	**0.0001**
>15mm	41	83.68	19	44.19		
Total	49		43			
NPI						
GPG	10	20.40	26	60.46	16.3576	**0.0003**
MPG	31	63.26	15	34.88		
PPG	8	16.32	2	4.65		
Total	49		43			

Table13: Grade = overall tumor grade, (based on Nottingham grading System), Type = histological tumor type (Lobular, Atypical Medullary, Mixed, Mucinous, Non-specific Type, Tubular), Stage = tumor stage, Vi = degree of vascular invasion, Hist. Size = histological tumor size (15mm being the standard clinical cut-off for dichotomization), NPI = Nottingham prognostic index (using standard clinical cut-offs). Components of the Nottingham grading system; T= degree of tumor differentiation, P = nuclear pleomorphism, M = mitotic index. NPI calculated as (0.2 x Hist Size in cm) / (stage + grade), GPG = Good prognostic group (NPI<3.41), MPG = Moderate prognostic group (NPI >=3.41 and <=5.40), PPG = Poor prognostic group (NPI >5.4)

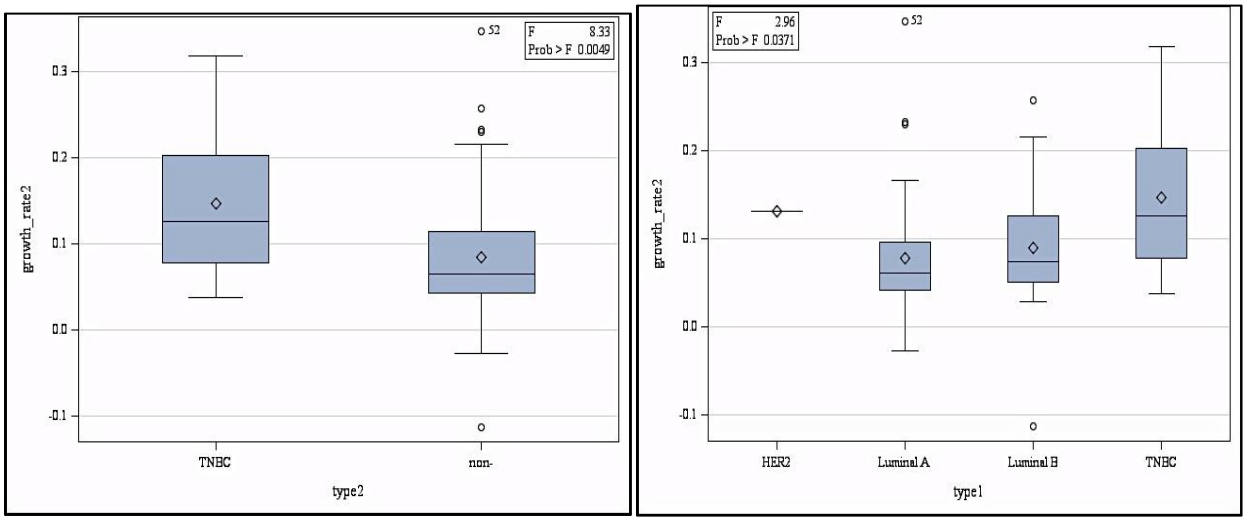
Furthermore, we also tested the correlation between the growth rate and the difference between Ki67 and cleaved Caspase-3. We observed a significant correlation (p = 0.0017) indicating that the tumor growth rate is correlated with the balance between ki67 and caspase3.

Table 14 Relation between Ki67 and Caspase3 expression

	Change in growth rate	
Difference in Ki67 and cleaved Caspase3	Pearson correlation Coefficient	P-value
	0. 41766	<0.001

Correlation between breast cancer subtype and growth rate:

A highly significant positive correlation ($p<0.05$) was observed between subtypes of breast cancer and tumor growth rate. The tumor growth rate varied significantly between the breast cancer subtypes (Figure: 25). Furthermore, when growth rate between TNBC and non-TNBC subtypes were compared we observed a higher tumor growth rate among TNBC cohort compared to other.



(a) ***(b)***

Figure 25 *Comparison of growth rate among different type of Breast Cancer*

3.4.4 Results from survival analysis on the tumor growth data

The increase in tumor growth rate significantly increases the risk to death: When tumor growth rate was considered as the predictive marker for survival, a univariate analysis revealed that the risk to death increases as the growth rate increases. With every unit of positive change in tumor growth change, the risk to death also increases (HR=1.073, P=0.0268) Table: 15. Furthermore, when Age, Grade, and ki67 were adjusted in the multivariate approach, tumor growth rate is still a significant predictor of overall survival (HR= 1.071 and p= 0.0435)

Table 15 Survival analysis for monthly growth rate

Variable	Univariate Analysis		
	Hazard Ratio	95% Confidence interval	P-value
Growth rate	1.073	1.008-1.143	0.0268

Table 16 Survival analysis with adjusting age, Ki67, and grade

Variable		Multivariate analysis		
		Hazard Ratio	95% Confidence interval	P-value
Growth rate		1.071	1.002-1.145	**0.0435**
Age		0.995	0.976-1.014	0.6111
Grade		1.033	0.545-1.96	0.9204
Ki67		1.002	0.977-1.027	0.8894

3.4.5 Results from optimize the cutpoint

After finding the best cutpoint 0.1, we plot the survival curves for both fast growth group and slow growth group. The difference between them is significant. We have the p-value for Chi-Square Log-Rank test as 0.0245, which means that the patients with fast growth rate have a higher risk to death.

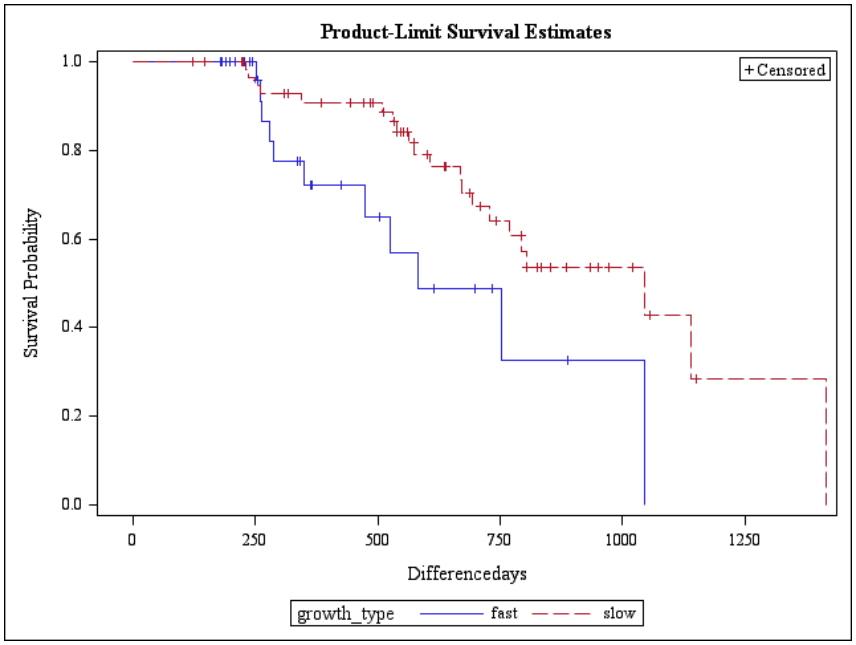


Figure 26 *Survival Curve for different groups*

Figure 26 shows the results for the survival curves of two groups when we choose 0.1 as the cutpoint to separate fast growth group from slow growth group. The p-value for the Chi-Square test Log-Rank test is 0.0245

3.4.6 Results from logistic regression modeling on the growth rate

After applying the method described in the section 2.4.6, we obtained the best model as following:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * Casps3_{Score} + \beta_2 * Grade + \beta_3 * HER + \beta_4 * Ki67_{Score} + \beta_5 * M + \beta_6 * Age + \beta_7 * Age * T + \beta_8 * Casps3_{Score} * Age$$

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	117.059	115.036
SC	119.547	142.411
-2 Log L	115.059	93.036

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	22.0231	10	0.0150
Score	20.2071	10	0.0274
Wald	14.7686	10	0.1407

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
CASP3_SCORE	1	1.3969	0.2372
Grade	2	4.3468	0.1138
HER11	1	1.1243	0.2890
KI67_SCORE	1	3.9386	0.0472
M	2	3.5307	0.1711
AGE	1	2.9366	0.0866
AGE*T	1	2.8339	0.0923
CASP3_SCORE*AGE	1	1.4302	0.2317

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7431	2.3468	0.1003	0.7515
CASP3_SCORE		1	3.4762	2.9412	1.3969	0.2372
Grade	**1**	1	-1.4743	0.7986	3.4085	0.0649
Grade	**2**	1	0.5804	0.5030	1.3317	0.2485
HER11	**0**	1	-0.7044	0.6643	1.1243	0.2890
KI67_SCORE		1	-0.0365	0.0184	3.9386	0.0472
M	**1**	1	1.2064	0.8127	2.2032	0.1377
M	**2**	1	-0.6599	0.4431	2.2178	0.1364
AGE		1	0.0830	0.0484	2.9366	0.0866
AGE*T		1	-0.0183	0.0109	2.8339	0.0923

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
CASP3_SCORE*AGE		1	-0.0577	0.0483	1.4302	0.2317

3.5 Results from extending the survival analysis on the breast cancer recurrence

3.5.1 Clinico-pathological characteristics of patients

The demographics, breast clinico-pathological characteristics, therapies administered and patterns of recurrence among the patients in the cohort are illustrated in the figure 27. From this cohort of 10,504 NH patients, 225 were recorded as having experienced a recurrence episode and 6,009 were determined as displaying no breast tumor recurrence. The remaining patients did not have recorded data indicating the presence of recurrence or lack thereof. Among patients displaying recurrence, higher risk of recurrence was more prevalent among younger patients ($p<0.0001$).

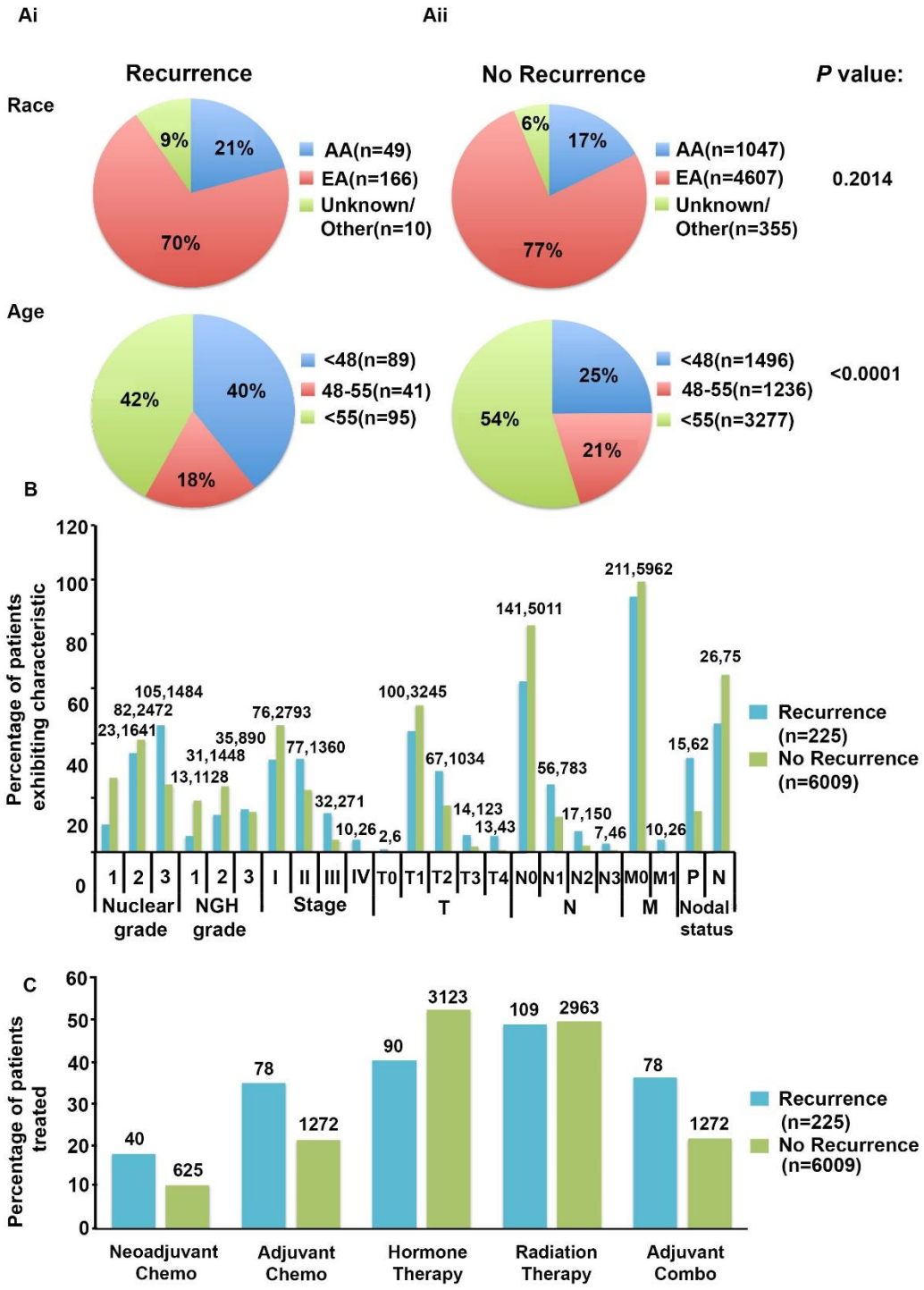


Figure 27 *NH demographics and breast cancer clinico-pathological characteristics compared between patients with or without tumor recurrence*

This result is consistent with previous studies that have observed an association between

younger age and increased risk for recurrence [58, 59]. Among patients with no missing recurrence data, approximately 61% of patients who experienced recurrence were under the age of 48, compared to only 39% who did not experience any recurrence. Among breast clinico-pathological characteristics, recurrence was significantly more associated with higher nuclear grade, NGH grade, stage, as well as T, N, and M classifications ($p<0.0001$). Moreover, recurrence was weakly associated with lymph node metastasis with roughly 35% of patients with recurrence displaying a positive nodal status compared to only 15% of non-recurrence patients ($p=0.121$). These results further confirm previous findings of increased risk of recurrence associated with more aggressive tumor characteristics (48). Regarding treatment, there were significant differences in the distribution of recurrence and non-recurrence patients who were administered neoadjuvant and adjuvant chemotherapy, hormone therapy, and a combination of adjuvant therapies ($p<0.0001$). There was a weak statistical significant difference between the proportion of recurrence and non-recurrence patients that received radiation therapy ($p=0.065$). Please visit Table 17 for details.

Table 17 NH demographics and breast cancer clinico-pathological compared between patients with or without tumor recurrence

	Recurrence		**No recurrence**		
Clinical characteristic	**n=225**	**%**	**n=6009**	**%**	***P***
Ethnicity					
AA	49	21.78	1047	17.42	0.2014
EA	166	73.78	4607	76.67	
Unknown/Other	10	4.44	355	5.91	
Menopausal status					

<48	89	39.56	1496.00	24.90	
48-55	41	18.22	1236.00	20.57	<0.0001
>55	95	42.22	3277	54.53	
Nuclear grade					
1	23	10.22	1641.00	27.31	<0.0001
2	82	36.44	2472.00	41.14	
3	105	46.67	1484.00	24.70	
Missing	15	6.67	412.00	6.86	
Nottingham grade					
1	13	5.78	1128.00	18.77	<0.0001
2	31	13.78	1448.00	24.10	
3	35	15.56	890.00	14.81	
Missing	146	64.89	2543.00	42.32	
Stage					
0	24	10.67	1518.00	25.26	<0.0001
I	76	33.78	2793.00	46.48	
IIa,b	77	34.22	1360.00	22.63	
IIIa,b,c,NOS	32	14.22	271.00	4.51	
IV	10	4.44	26.00	0.43	

Missing/unknown	6	2.67	41.00	0.68	
Nodal status					
Positive	78	34.67	911.00	15.16	0.121
Negative	106	47.11	3920.00	65.24	
Missing/unknown	41	18.22	1170.00	19.47	
TNM Staging					
T					
T0	2	0.89	6.00	0.10	<0.0001
TX	2	0.89	30.00	0.50	
Tis	24	10.67	1518.00	25.26	
T1	100	44.44	3245.00	54.00	
T2	67	29.78	1034.00	17.21	
T3	14	6.22	123.00	2.05	
T4	13	5.78	43.00	0.72	
Unknown	3	1.33	10.00	0.17	
N					
N0	141	62.67	5011.00	83.39	<0.0001
NX	1	0.44	9.00	0.15	
N1	56	24.89	783.00	13.03	

N2	17	7.56	150.00	2.50	
N3	7	3.11	46.00	0.77	
Unknown	3	1.33	10.00	0.17	
M					
M0	211	93.78	5962.00	99.22	<0.0001
MX	1	0.44	11.00	0.18	
M1	10	4.44	26.00	0.43	
Unknown	3	1.33	10.00	0.17	
Lymph node surgery					
Yes	158	70.22	3793	63.12	0.0169
No	67	29.78	2216	36.88	
Unknown	0	0.00	0	0.00	
Chemotherapy					
Neoadjuvant	40	17.78	625	10.40	<0.0001
Adjuvant	78	34.67	1272	21.17	
None	92	40.89	3952	65.77	
Missing	15	6.67	160	2.66	
Hormone therapy					

Yes	90	40.00	3123	51.97	<0.0001
No	119	52.89	2705	45.02	
Unknown	16	7.11	181	3.01	
Radiation therapy					
Yes	109	48.44	2963	49.31	0.0649
No	96	42.67	2732	45.47	
Unknown	20	8.89	314	5.23	
Adjuvant therapy					
Yes	78	35.94	1272	21.42	<0.0001
No	132	60.83	4577	77.07	
Unknown	15	6.91	160	2.69	

3.5.2 Recurrence pattern among racially distinct patients

Recurrence rates and patterns, expressed in terms of incidence rates, were compared broadly between AA and EA patients (Table 18); the analysis indicated that AA exhibited higher overall tumor recurrence rates than EA (p=0.002; HR: 1.676; CI: 1.210-2.323). AA also displayed higher rates of distant recurrence than EA (p=0.023; HR: 1.699; CI: 1.075-2.684); however, these differences did not remain statistically significant after controlling for age, grade, and stage, likely owing to low patient numbers. Additionally, AA experienced higher rates of single tumor recurrence episodes than EA (p=0.003; HR: 1.758; CI: 1.208-2.557) and higher rates of distant recurrence to a single site than EA breast cancer patients (p=0.012; HR: 1.742; CI: 1.130-2.684), although statistical significance was not maintained after adjusting for age, grade, and stage

Table 18 Broad spectrum recurrence patterns among racially distinct populations

	EA		AA		p value; HR (95% CI)	p value; HR (95% CI)
	n	**IR**	**n**	**IR**	**Unadjusted Model**	**Adjusted Model**
Overall	166	13.44	49	21.77	**0.002; 1.676 (1.210, 2.323)**	0.319; 1.192 (0.844, 1.683)
Recurrence site						
Local	48	3.89	12	5.33	0.373; 1.349 (0.698, 2.606)	0.665; 0.857 (0.428, 1.718)
Regional	27	2.19	10	4.44	0.188; 1.701 (0.772, 3.747)	0.151; 1.749 (0.815, 3.752)
Distant	84	6.8	27	12	**0.023; 1.699 (1.075, 2.684)**	0.280; 1.299 (0.809, 2.085)
Number of recurrences						
Single	131	10.6	41	18.21	**0.003; 1.758 (1.208, 2.557)**	0.218; 1.287 (0.861, 1.923)
Multiple	35	2.83	8	3.55	0.754; 1.139 (0.505, 2.573)	0.315; 0.652 (0.283, 1.503)
Distant recurrence						
Single site	73	5.91	24	10.66	**0.012; 1.742 (1.130, 2.684)**	0.451; 1.220 (0.728, 2.043)
Multiple sites	11	0.89	3	1.33	0.492; 1.566 (0.436, 5.625)	0.617; 0.672 (0.142, 3.187)

Abbreviations: AA, African-American; EA, European-American; HR, hazard rate; IR, incidence rate (1000 person-years); CI, confidence interval.

Adjusted Cox hazard model variables: age at diagnosis, grade (1,2,3), and stage (I,II,III,IV).

**P* values were calculated using the student t-test.

3.5.3 Recurrence patterns among racially distinct patients following each form of treatment

Incidence rates and patterns of recurrence were compared between AA and EA after they received hormone, radiation, chemotherapy, and/or any combination of adjuvant therapy to determine distinctions in recurrence patterns between therapies among the racial groups (Table 19). AA exhibited unadjusted higher rates of recurrence (p=0.041; HR: 1.612; CI: 1.021-2.545) and a trend towards higher incidence of distant recurrence than EA post radiation therapy (p=0.065; HR: 1.732; CI: 0.967-3.100). The same trend of higher overall and distant recurrence was observed among recurrent patients who received hormone therapy and any combination of adjuvant therapies. Among patients who underwent hormone therapy, AA displayed stronger overall tendencies than EA to suffer from recurrence (p=0.112; HR: 1.541; CI: 0.906-2.623) and distant recurrence (p=0.123; HR: 1.692; CI: 0.868-3.301). Following any combination of adjuvant therapy, AA displayed higher recurrence rates than EA after adjusting for age, grade, and stage (p=0.015; HR: 1.699; CI: 1.108-2.606). Moreover, unadjusted analyses reveal AA displayed higher rates of distant recurrence than EA (p=0.003; HR: 2.164; 1.290-3.629) as well as stronger tendencies toward regional recurrence (p=0.104; HR: 2.043; CI: 0.863-4.837) after receiving any combination of adjuvant therapy.

Table 19 Recurrence rates and patterns after receiving any form of treatment among racially distinct populations

	EA	AA	p value; HR (95% CI)	p value; HR (95% CI)

Treatment	n	IR	n	IR	Unadjusted model	Adjusted model
Chemotherapy						
Overall	85	22.73	23	26.66	0.466; 1.181 (0.755, 1.846)	0.807; 0.943 (0.587, 1.514)
Local	13	3.48	7	7.46	0.125; 2.053 (0.818, 5.151)	0.394; 1.548 (0.567, 4.226)
Regional	20	5.35	6	6.4	0.594; 1.284 (0.512, 3.219)	0.749; 1.169 (0.450, 3.041)
Distant	50	13.37	12	12.79	0.832; 0.934 (0.498, 1.751)	0.613; 0.840 (0.426,1.653)
Neoadjuvant chemotherapy						
Overall	32	28.95	6	19.77	0.373; 0.673 (0.281, 1.609)	0.409; 0.690 (0.286, 1.664)
Local	2	1.81	4	13.18	0.026; 6.857 (1.256, 37.447)	0.024; 7.134 (1.295, 39.313)
Regional	7	6.33	0	0	N/A	N/A
Distant	23	20.81	2	6.59	0.112; 0.310 (0.073, 1.315)	0.136; 0.332 (0.078, 1.417)
Adjuvant chemotherapy						
Overall	57	20.96	18	26.16	0.405; 1.253 (0.737, 2.130)	0.891; 1.039 (0.603, 1.788)

Local	12	4.41	3	4.36	0.865; 0.897 (0.255, 3.153)	0.500; 0.645 (0.181, 2.303)
Regional	15	5.52	6	8.72	0.333; 1.598 (0.619, 4.125)	0.843; 1.102 (0.421, 2.885)
Distant	30	11.03	9	13.08	0.664; 1.179 (0.561, 2.480)	0.100; 1.000 (0.463, 2.159)
Hormone therapy						
Overall	69	10.45	17	15.94	0.112; 1.541 (0.906, 2.623)	0.949; 1.020 (0.568, 1.830)
Local	15	2.27	2	1.87	0.676; 0.731 (0.169, 3.172)	0.290; 0.332 (0.043, 2.558)
Regional	14	2.12	4	3.75	0.369; 1.654 (0.552, 4.959)	0.580; 1.380 (0.442, 4.305)
Distant	40	6.06	11	10.31	0.123; 1.692 (0.868, 3.301)	0.482; 1.307 (0.619, 2.757)
Radiation therapy						
Overall	79	12.62	23	19.34	0.041; 1.612 (1.021, 2.545)	0.986; 1.004 (0.609, 1.658)
Local	22	3.52	6	5.04	0.450; 1.414 (0.575, 3.475)	0.689; 0.816 (0.302, 2.205)
Regional	10	1.6	3	2.52	0.532; 1.503 (0.419, 5.392)	0.736; 1.264 (0.324, 4.490)

Distant	47	7.51	15	12.61	0.065; 1.732 (0.967, 3.100)	0.810; 1.083 (0.568, 2.063)
Adjuvant radiation, hormone, and chemotherapy						
Overall	101	11.94	30	19.63	0.013; 1.678 (1.115, 2.524)	0.015; 1.699 (1.108, 2.606)
Local	31	3.66	3	1.96	0.279; 0.520 (0.159, 1.698)	0.145; 0.405 (0.121, 1.364)
Regional	19	2.25	7	4.58	0.104; 2.043 (0.863, 4.837)	0.558; 1.310 (0.531, 3.230)
Distant	51	6.03	20	13.09	0.003; 2.164 (1.290, 3.629)	0.101; 1.607 (0.912, 2.833)

Abbreviations: AA, African-American; EA, European-American; HR, hazard rate; IR, incidence rate (1000 person-years); CI, confidence interval.

Adjusted Cox hazard model variables: age at diagnosis, grade (1,2,3), and stage (I,II,III,IV).

**P* values were calculated using the student t-test.

3.5.4 Recurrence rates among racially distinct breast cancer patients in different stages

Overall incidence rates of recurrence were compared between AA and EA in both early (I–II) and late stage (III-IV) breast cancer patients (Table 20). Our data revealed that AA displayed higher recurrence rates than EA among stage I patients (p=0.001; HR: 2.165; CI: 1.348-3.476), even after adjusting for age, grade, and stage (p=0.031; HR: 1.736; CI: 1.052-2.864). Among early stage (I-II) patients, AA also exhibited higher recurrence rates than EA (p=0.002; HR: 1.793; CI: 1.252-2.567), and trending higher in AA after controlling for age, grade, and stage (p=0.131; HR: 1.339; CI:

0.917-1.956). Furthermore, AA displayed higher recurrence rates than EA among T1 classified patients, irrespective of age, grade, and stage (p=0.003; HR: 2.009; CI: 1.263-3.197). Moreover, unadjusted models reveal that AA displayed higher rates of recurrence than EA among N0 (p=0.005; HR: 1.777; CI: 1.186-2.661) and M0 (p=0.002; HR: 1.682; CI: 1.210-2.338) classified patients. However, rates of recurrence were not significantly higher in AA as compared to EA among late stage patients. Thus, these results suggest that AAs are at higher risk than EAs for tumor recurrence among patients with non-invasive or minimally invasive breast cancer.

Table 20 Overall recurrence rates among racially distinct staged breast cancer patients

	EA		**AA**		**p value; HR (95% CI)**	**p value; HR (95% CI)**
	n	**IR**	**n**	**IR**	**Unadjusted Model**	**Adjusted Model**
Grouped stage						
Early (I-II)	130	11.14	39	19.08	**0.002; 1.793 (1.252, 2.567)**	0.131; 1.339 (0.917, 1.956)
Late (III-IV)	31	55.17	9	50.65	0.857; 0.934 (0.445, 1.962)	0.637; 0.823 (0.366, 1.850)
Individual Stage						
I	70	7.76	23	15.84	**0.001; 2.165 (1.348, 3.476)**	**0.031; 1.736 (1.052, 2.864)**
II	60	22.67	16	27.02	0.447; 1.239 (0.713, 2.154)	0.823; 0.936 (0.523, 1.674)

III	25	48.01	7	45.01	0.902; 0.949 (0.410, 2.195)	0.590; 0.774 (0.306, 1.959)
IV	6	145.8	2	90.29	0.822; 0.832 (0.167, 4.152)	0.967; 0.964 (0.168, 5.518)
TNM Staging						
T						
T0	2	130.83	0	N/A	N/A	N/A
T1	67	9.75	28	25.63	**<0.0001; 2.776 (1.781, 4.326)**	**0.003; 2.009 (1.263, 3.197)**
T2	54	28.48	11	22.3	0.504; 0.801 (0.419, 1.534)	0.215; 0.647 (0.325, 1.287)
T3	12	49.67	1	13.31	0.215; 0.275 (0.035, 2.115)	0.161; 0.228 (0.029, 1.796)
T4	9	106.73	4	121.36	0.680; 1.282 (0.394, 4.173)	0.983; 1.015 (0.241, 4.270)
N						
N0	101	9.81	31	16.89	**0.005; 1.777 (1.186, 2.661)**	0.211; 1.319 (0.854, 2.037)
N1	44	27.05	12	39.2	0.201; 1.518 (0.801, 2.877)	0.828; 1.079 (0.545, 2.136)
N2	13	46.45	4	53.23	0.744; 1.207 (0.391, 3.719)	0.965; 0.970 (0.258, 3.646)

N3	5	55.06	1	43.27	0.742; 0.697 (0.081, 5.970)	0.974; 0.962 (0.095, 9.711)
M						
M0	157	12.82	46	20.8	**0.002; 1.682 (1.210, 2.338)**	0.288; 1.210 (0.851, 1.721)
M1	6	145.8	2	90.29	0.822; 0.832 (0.167, 4.152)	0.967; 0.964 (0.168, 5.518)

Abbreviations: AA, African-American; EA, European-American; HR, hazard rate; IR, incidence rate (1000 person-years); CI, confidence interval.

Adjusted Cox hazard model variables: age at diagnosis, grade (1,2,3), and stage (I,II,III,IV).

**P* values were calculated using the student t-test.

3.5.5 Survival outcomes among racially distinct patients displaying recurrence

Survival duration after the initial recorded recurrence was compared between the racial groups (Figure 28). AA exhibited only a very weak trend toward shorter survival time than EA after experiencing their first episode of recurrence (p=0.231). The average time until death was compared between EA and AA patients who experienced distant recurrences. Interestingly, AA and EA patients exhibiting distant recurrence were comprised of similar percentages of alive patients, however, AA (n=26) died considerably sooner than EA (n=80) (p=0.015). More precisely, AA patients who experienced distant recurrences died approximately one year earlier than EA distant recurrent patients. However patient numbers were too low to control for variables of age, grade, and stage.

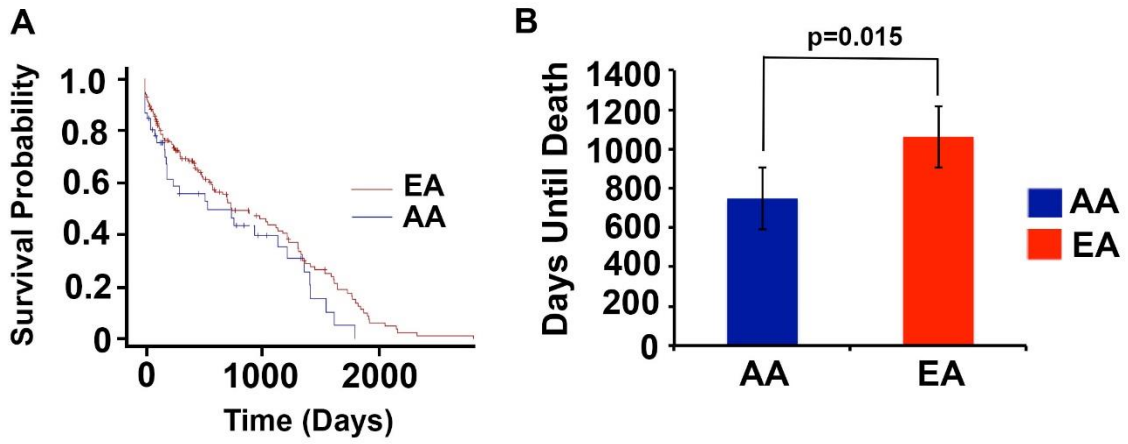


Figure 28 AA exhibit lower survival duration than EA among recurrent breast cancer patients

Figure 28: (A) Survival time from first recurrence episode until death was compared between AA and EA breast cancer patients. The log-rank analysis was conducted to determine statistical differences between the racial groups. AA exhibited a weakly significant lower survival time than EA (p=0.231). (B) The mean time (days) until death was compared between AA and EA breast cancer patients displaying distant recurrence. AA died notably sooner than EA patients (p=0.015). A t-test was performed to determine significant differences between the racial groups.

4 CONCLUSIONS

4.1 Conclusion part for neural growth model

Our analysis of the discrete probability distributions corresponding to stochastic neuronal growth models indicates that the most important component of our model, namely the branching probability p, has a clear impact on targeting performances. More precisely, the expected success near the origin increases as a geometrical series with a factor of $(1 + p)$. Furthermore, the decrease in targeting performance at large distances approximately follows a geometrical series with $(1 + 3p)$ and additional higher order corrections. In addition to computing the values of the expected number of branches as a function of distance, we also determined an expression for the corresponding variance values, as well as approximations for the optimal targeting regions.

These probabilistic models are in excellent agreement with experimental data from the Firestein laboratory and indicate that the parameters used in the theoretical model capture most of the variability seen in the experimental data. One particularly interesting finding in this study is that our model very accurately predicts dendrite growth for relatively young neurons in culture, at days in vitro (DIV) 7 and 8. From our results, we can see that the estimate of growth rate varies when we have different chemicals controlling the evolution process, for example, estimate of branching probability of control is lower than that of Cypin. All the results from our model were in the agreement with previous researches, which means our model can detect the change in the regeneration process when the condition changes.

An interesting outcome of these results is to suggest that the logarithm of the number of active branches might be a more appropriate way of constructing Sholl plots, as this transformation of the expected values for the dendritic tree population is predicted to have a linear component near the origin and approximate linear component at large distances. This is due to the fact that logarithm transformation of geometric series results in linear sequences. The use of both of these short and long-range components can provide a better statistical determination of branching values, for example examining the data at those scales using linear regression models. If the large-distance component is in disagreement with the predicted linear trend, this may indicate

other effects, such as pruning at larger distances or parameter-dependent branching.

When neurons are injured in the brain or spinal cord, they may lose dendritic or axonal branches, depending on the severity of the injury. Losing branching will disrupt the network and lead to a loss of function. To this end, much work in regenerative neuroscience has focused on promoting nerve growth after injury as a means of restoring function. Our goal with this computational model is to better understand the rules that govern neuronal branching under control and injury conditions and to confirm these findings with an in vitro system (cultured hippocampal neurons). Future iterations of this model will allow us to determine whether promoting a particular type of branching pattern – for example, that characteristic of treatment with BDNF – will allow a neuron to best reach the target area after injury and become reintegrated with the network after injury.

Additionally, we were able to use these models to derive statistical tests for determining differences between conditions, such as normal neurons versus neurons treated with growth factors (e.g. BDNF; [60]) or neurons engineered to overexpress a particular protein (e.g. cypin; [61, 62]). Such tests could replace the less informative methods currently used, including variance analysis, which only provide information regarding statistical differences at specific distances as opposed to analysis of the whole curve. In contrast, our analysis will produce a test of hypothesis for relevant parameters, such as branching probability, in order to determine if different experimental conditions produce changes in the fit values of these parameters that are statistically significant. As an illustration, imagine comparing families of neurons with fewer resources and a slower evolution speed to ones that have more resources but a higher growth speed; while the branching rates are similar, the dendritic trees become scaled up or scaled down versions of each other. The resulting Sholl curves would have the same overall shape but are “stretched out” or “compressed” when compared to each other. In this imagined scenario, the ANOVA analysis would indicate statistical differences between these two types of Sholl curves at many locations without providing any intuition about their similarities.

In the future, we will continue to apply our model on different neuron growth process controlled by the same chemical, but with different concentration. By doing this, we will be able to quantitatively connect the change in branching probability with the change in the concentration.

This will definitely be useful in the future in the clinical practice, in order to help to control the development of neuron cells.

4.2 Conclusion part for tumor growth model

The relation between tumor size and biological parameters has been well documented and widely used in prognostication. However, less attention was paid on the effect of clinicopathological factors, such as grade, stage, ki_67 score, on the in vivo breast cancer growth rate [63-67].

This study was aimed not only to reaffirm our current knowledge but also to explore the impact of more novel histological biomarkers on tumor growth rate. Of particular interest was the balance between pro-apoptotic and anti-apoptotic signaling proteins. Those investigated were the proliferative marker Ki67 and the apoptotic marker cleaved Caspase-3, an activated protease enzyme acting in the late stages of the apoptotic pathway. [68] Additionally, the correlation between growth rates – calculated as the change in volume over time across sequential mammograms - and ER, PR, and HER-2 status were investigated, as well as that of the basal marker CK 5/6. However, one problem with studies looking at early screen-detected tumors is that, by their very nature, most tumors being investigated will be small slower-growing Luminal tumors, expressing mostly ER and PR (75% of cases). This made the study somewhat biased, as very few tumors of the HER-2 or basal (CK 5/6 expressing) or TNBC subtypes were studied (8.7%, 4% and 12% of cases respectively). These tumor subtypes are typically more aggressive and much faster growing, making them far less likely to be screening detected [69,70]. It was, therefore, difficult to establish significant correlations between these markers/subtypes and other parameters.

Interestingly, different studies looking at patient outcome have used a wide range of different positive/negative cut-off points for both Ki67 and apoptotic marker staining. This has led to debate as to the most appropriate cut- off point. Some studies have proposed that the cut-off should vary depending on the clinical objective of the study.[71-73]

Few studies have looked at the growth rate in terms of the balance between pro-and anti-apoptotic signaling proteins, and so no consensus on cut-off has been agreed in this context. In this study, in order to keep the data as unbiased as possible, Ki67 and cleaved Caspase-3 were analyzed as continuous variables. It was shown that increasing Ki67 expression correlated significantly with increased tumor growth rate. This was to be expected, with Ki67 being a proliferative marker expressed during all active phases of the cell cycle, peaking during the mitotic phase. [71-74] A significant association was also seen between cleaved Caspase-3 and growth rate. Moreover, Cleaved Caspase-3 was found to positively correlate with Ki67 expression, suggesting that an increase in proliferative markers in breast tumor cells is usually accompanied by an increase in apoptotic proteins as a counter-balancing measure. A previous larger breast cancer study showed similar observations. [75]

Correlations were observed between growth rate and tumor grade and mitotic index. The relationship with grade has previously been observed in many studies and reflects the fact that mitotic index is the most powerful component of tumor grade. [74, 76] In fact, some studies have shown tumor size to be positively correlated with Ki67 expression, whilst have others found no significant relationship. [74] Degree of vascular invasion also significantly correlated with growth rate. This is accounted for by the increased provision of nutrients to the tumor cells from the newly invaded blood vessels, allowing a greater degree of cell proliferation. Other studies have made similar findings, and it has been shown that extensive vascular invasion is also closely related to poorer prognosis, and pronounced Ki67 expression. [76]

A previous study adopted an approach to measure tumor growth rate as the change in tumor size over time in only one dimension – the largest tumor dimension. [77] However this method introduced considerable inaccuracy into the growth rate calculations and was based on the assumption that the proportional increase in tumor size was the same in all three dimensions. In fact, it has been shown that growth rate is influenced heavily by stromal-tumor interactions and the presence of limiting anatomical barriers. [78, 79] A tumor thus is unlikely to grow uniformly in all directions, and will most probably favor the direction of least resistance. A more accurate approach would have been to measure the change in tumor size in two or even three dimensions (tumor volume) if possible. Here we calculate tumor volume from mammogram readings. We

assumed the breast cancer tumor to be ellipsoidal and the highest mammogram reading was assumed as height and the other as diameter. The tumor volumes were between 9 to 37819mm3 with a mean of 4591.74mm3. Thus the tumor volume calculated could serve as a factor for demonstrating the tumor expansion rate rather than tumor size alone, which is a biased parameter. [80]

Interestingly, a previous study has suggested that tumor growth rate, independent of stromal interactions, is non-linear, and grows in a so-called “stepwise” manner, undergoing phases of relative stability intermittently interrupted by sudden surges in growth. It is proposed that these periods of inactivity may be due to: (a) periods of balanced proliferation and apoptosis, or (b) periods when the growth fraction is zero. The latter would presumably be accompanied by a massively decreased expression of proliferative signaling proteins such as Ki67. [81] In this study, the mean time between screening and diagnosis was just 18 months. In future studies growth rate could be calculated over a longer time period, offering a more accurate long-term picture of changes in tumor volume.

In our study, we observed a significant correlation between breast cancer subtypes and tumor growth rate. Tumor growth rate varied significantly between the subtypes with higher growth rate among TNBC patients. Thus, the growth rate may be a significant consideration when we try to separate the breast cancer into different subtypes.

Furthermore, we investigated the relationship between growth rate and patient outcome. Our analysis revealed that increase in growth rate increases the risk to death. However, the tumor growth rate has not been implicated as a prognostic variable in clinical practice because of its difficulty in evaluating it in the short interval of diagnosis and treatment. Many studies have addressed that faster growing tumors have a worse survival. [24, 82-84] However, some studies demonstrated that there is no any association of tumor growth rate with patient’s survival. [25] Also, in the other study carried out by Tubiana et.al, there was no any survival difference between the subgroups of patients with rapid or intermediate growth rate after the follow-up exceeding 8 years [26].

One final avenue that could be explored more is the possible correlations between growth rate

and other aspects of mammographic appearance, and also between biomarker expression and mammographic appearance. Some work has already been done in this area, which shows that luminal tumors showed more speculate lesions on mammographic shadow, whilst HER-2 positive and basal tumors were characterized by more ill-defined masses on mammography. [57, 66] However there is certainly scope for further investigation in this area, looking at density, calcifications, and other features.

In conclusion, the study has demonstrated the relationship between breast cancer growth rate and the expression of the proliferative marker Ki67. Well-documented relationships between various biomarkers and clinicopathological parameters classically used in breast cancer prognostication have also been reaffirmed. The activity of the pro-apoptotic protein cleaved Caspase-3 in counterbalancing proliferative activity – controlled by Ki67 – was clearly observed. Overall the study proposes that growth rate is a significant consideration when separating the type of breast carcinomas, provide the basis for further refinement of the current classification system, as well as the discovery of new molecular subtypes and consequently, it is important in determining patient outcome. This study has revealed prognostic information and evidence-based data that can be used in the medicolegal practice.

4.3 Conclusion from extending survival analysis on breast cancer recurrence

Our analysis revealed higher incidence rates of recurrence in AA compared to EA among patients that received any combination of adjuvant therapy. Moreover, our data demonstrates an increased risk of tumor recurrence in AA than EA among patients diagnosed with minimally invasive disease. Additionally, this is the first clinical study to suggest that neoadjuvant chemotherapy improves breast cancer recurrence rates and patterns in AA.

This clinical study is the first extensive investigation into the rates and patterns of tumor recurrence in breast cancer patients following conventional treatments among racially distinct populations. Our study has revealed notable distinctions in recurrence patterns among EA and AA patients. First, AA displayed considerably higher rates of recurrence than EA. Second,

intriguingly, we observed higher severity in recurrence patterns displayed by AA for whom we discerned stronger trends in AA of tumor recurrence to regional and distant sites. This trend was evident after patients received radiation, hormone, and any combination of adjuvant therapies. Overall, these observed trends were quite significant since local recurrence tends to elicit a more favorable clinical prognosis compared to distant recurrence, while the latter trends type precedes a poorer clinical prognosis. Triple-negative breast cancer (TNBC) patients have been shown to display an increased risk for recurrence and particularly for recurrence to distant sites, while non-TNBC patients exhibit higher trends of recurrence to local sites [58]. These findings parallel our observations of an increased risk of overall and especially distant recurrence in AA, as well as an increased risk of local recurrences in EA. This tendency reflects the well-reported higher incidence of TNBC phenotypes in AA patients and a higher prevalence of non-TNBC subtypes in EA patients. Furthermore, we observed a trend of a higher number of recurrence episodes in AA compared to EA. The more recurrence episodes of the patient experiences, the more fatal the outcome is likely to be, and the more the patient's quality of life is potentially compromised. Additionally, we discerned stronger inclinations of distant recurrence to multiple organs in AA compared to EA. These observed aggressive recurrence patterns reveal that AA patients exhibit an increased prospect of a poor clinical prognosis, theoretically contributing to their higher mortality rates than EA patients. Recurrence rates were also found to be higher in AA than EA among early stage, minimally invasive breast cancer patients, irrespective of age, grade, or stage. This data presents an intriguing paradox as the advanced stage upon diagnosis is typically associated with increased risk for recurrence. Thus, these findings suggest that AA patients of all clinical stages should be closely evaluated for the prospect of tumor recurrence. Neoadjuvant chemotherapy seemed to reverse these observed recurrence trends. Among patients who received neoadjuvant chemotherapy, AA displayed a lower rate of recurrence than EA; however due to a low number of recorded patients that received neoadjuvant chemotherapy, statistical significance was diminished. In addition, higher incidences of aggressive recurrence patterns in AA were notably attenuated after these patients underwent neoadjuvant chemotherapy. This data suggest preoperative chemotherapy may reduce the severity of recurrence rates and patterns in AA patients. This study suggests that neoadjuvant chemotherapy should be

recommended for AA patients who are at higher risk for developing tumor recurrence. A recent clinical study reported that in fact, neoadjuvant chemotherapy is administered more frequently to AA than EA patients likely as a result of their higher prevalence of advanced stage, grade, and triple negative receptor status [59]. Biological research is needed to be conducted to understand the molecular basis underlying inequalities in recurrence patterns among EA and AA patients for potential therapeutic targets. Augmented prognosticative clinical measures need to be exploited in order to foretell the likelihood of disease recurrence. Hence, rigorous and intensive supervision of the prospect of recurrence is conceivably compulsory to palliate the elevated risk of recurrent breast cancer demonstrated by AA patients.

Although prior clinical studies have exposed disparities in recurrence risk among EA and AA, this study is one of the first to uncover distinctions in rates and patterns of tumor recurrences following conventional forms of breast cancer treatments among the racial groups and thus highlights the need for further investigation and surveillance. Our comprehensive analysis has also illuminated previously unrecognized differences in the rates and patterns of recurrence post-chemotherapy among racially distinct populations by suggesting that AAs respond better to neoadjuvant chemotherapy. Additionally, no study has yet elucidated the significantly higher risk for recurrence among early stage AA patients. Overall, this study further advocates that race should be considered among the potential decisive risk factors in the clinic for recurrence. Awareness of the higher rate of recurrence in AA may compel clinicians to consider race as a critical factor in evaluating the prospect of cancer returning after patients enter remission, and allow this factor to play a major role in treatment decisions. Hereinafter, enriched comprehensive screening programs and tailored treatment plans may be imperative to impede augmented the risk of tumor recurrence and aggressive recurrence patterns in AA patients that may be reinforcing their poor clinical outcomes.

REFERENCES

[1] Philip J. Horner & Fred H. Gage (2000) Regenerating the damaged central nervous system. Nature 407, 963-970 (26 October 2000) | doi:10.1038/35039559.

[2] Maskery S, Buettner HM, Shinbrot T (2004) Growth cone pathfinding: a competition between deterministic and stochastic events. BMC Neurosci 5:22:31.

[3] Borisyuk R, Cooke T, Roberts A (2008) Stochasticity and functionality of neural systems: mathematical modelling of axon growth in the spinal cord of tadpole. Biosystems 93: 101–114.

[4] Uylings HBM, and Van Pelt J (2002). Measures for quantifying dendritic arborizations. Network: Comput. Neural Syst. 13: 397-414.

[5] Krottje JK, van Ooyen A (2007) A mathematical framework for modeling axon guidance. Bulletin of Mathematical Biology 69: 3–31.

[6] O'Neill KM, Akum BF, Dhawan ST, Kwon M, Langhammer CG, and Firestein BL (2015) Assessing effects on dendritic arborization using novel Sholl analyses. Frontiers in Cellular Neuroscience, 9:285.

[7] Hentschel HGE, van Ooyen A (1999) Models of axon guidance and bundling during development. Proceedings of the Royal Society B-Biological Sciences 266: 2231–2238.

[8] Borisyuk R, Cooke T, Roberts A (2008) Stochasticity and functionality of neural systems: Mathematical modelling of axon growth in the spinal cord of tadpole. Biosystems 93: 101–114.

[9] van Ooyen A (2011) Using theoretical models to analyse neural development (vol 12, pg 311, 2011). Nature Reviews Neuroscience 12.

[10] Zubler F, Douglas R (2009) A framework for modeling the growth and development of neurons and networks. Frontiers in Computational Neurosci- ence 3.

[11] Langhammer CG, Previtera ML, Sweet ES, Sran SS, Chen M, and Firestein BL (2010) Automated Sholl analysis of digitized neuronal morphology at multiple scales: whole cell Sholl analysis versus Sholl analysis of arbor subregions. CytometryPartA 77A: 1160–1168.

[12] Oşan R, Su E, Shinbrot T. (2011) The interplay between branching and pruning on neuronal target search during developmental growth: functional role and implications. PLoS One. 6(10):e25135.

[13] Lecture Notes for Introductory Probability Janko Gravner

[14] Carey, L.A., et al., Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama, 2006. 295(21): p. 2492-502.

[15] de Ronde, J.J., et al., Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. Breast Cancer Res Treat, 2010. 119(1): p. 119-26.

[16] Beral, V., et al., Hormone replacement therapy and high incidence of breast cancer between mammographic screens. Lancet, 1997. 349(9058): p. 1103-4.

[17] Hin-Peng, L., Diet and breast cancer: an epidemiologist's perspective. Crit Rev Oncol Hematol, 1998. 28(2): p. 115-9.

[18] Pharoah, P.D., et al., Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer, 1997. 71(5): p. 800-9.

[19] Spratt, J.S., J.S. Meyer, and J.A. Spratt, Rates of growth of human solid neoplasms: Part I. J Surg Oncol, 1995. 60(2): p. 137-46.

[20] Berry, D.A., et al., Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med, 2005. 353(17): p. 1784-92.

[21] Baker, L.H., Breast Cancer Detection Demonstration Project: five-year summary report. CA Cancer J Clin, 1982. 32(4): p. 194-225.

[22] Holland, R., et al., So-called interval cancers of the breast. Pathologic and radiologic analysis of sixty-four cases. Cancer, 1982. 49(12): p. 2527-33.

[23] Freimanis, R.I. and M. Yacobozzi, Breast cancer screening. N C Med J, 2014. 75(2): p. 117-20.

[24] Bailar, J.C., 3rd and E.M. Smith, Progress against cancer? N Engl J Med, 1986. 314(19): p. 1226-32.

[25] Yoo, T.K., et al., In Vivo Tumor Growth Rate Measured by US in Preoperative Period and Long Term Disease Outcome in Breast Cancer Patients. PLoS One, 2015. 10(12): p. e0144144.

[26] Tubiana, M., et al., Growth rate, kinetics of tumor cell proliferation and long-term outcome in human breast cancer. Int J Cancer, 1989. 44(1): p. 17-22.

[27] Arnerlov, C., et al., Breast carcinoma growth rate described by mammographic doubling time and S-phase fraction. Correlations to clinical and histopathologic factors in a screened population. Cancer, 1992. 70(7): p. 1928-34.

[28] Boyd, N.F., et al., Clinical estimation of the growth rate of breast cancer. Cancer, 1981. 48(4): p. 1037-42.

[29] Locopo, N., M. Fanelli, and G. Gasparini, Clinical significance of angiogenic factors in breast cancer. Breast Cancer Res Treat, 1998. 52(1-3): p. 159-73.

[30] Casey, T., et al., Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. Breast Cancer Res Treat, 2009. 114(1): p. 47-62.

[31] Galante, E., et al., Growth rate of primary breast cancer and prognosis: observations on a 3- to 7-year follow-up in 180 breast cancers. Br J Cancer, 1986. 54(5): p. 833-6.

[32] Kristek, J., et al., Tumor growth fraction, expression of estrogen and progesterone receptors, p53, bcl-2 and cathepsin D activity in primary ductal invasive breast carcinoma and their axillary lymph node metastases. Coll Antropol, 2007. 31(4): p. 1043-7.

[33] Marušić M, Vuk-Pavlović S (1993) Prediction power of mathematical models for tumor growth. J Biol Syst 1: 69–78

[34] Olea N, Villalobos M, Nuñez MI, Elvira J, Ruiz de Almodóvar JM, et al. (1994) Evaluation of the growth rate of MCF-7 breast cancer multicellular spheroids using three mathematical models. Cell Prolif 27: 213–223

[35] Wallace DI, Guo X (2013) Properties of tumor spheroid growth exhibited by simple mathematical models. Front Oncol 3: 51

[36] Vaidya VG, Alexandro FJ (1982) Evaluation of some mathematical models for tumor growth. Int J Biomed Comput 13: 19–36

[37]Laird AK (1965) Dynamics of tumour growth: comparison of growth rates and extrapolation of growth curve to one cell. Br J Cancer 19: 278–291

[38] Keshgegian, A.A. and A. Cnaan, Proliferation markers in breast carcinoma. Mitotic figure count, S-phase fraction, proliferating cell nuclear antigen, Ki-67 and MIB-1. Am J Clin Pathol, 1995. 104(1): p. 42-9.

[39] Brent A. Williams, MS, Jayawant N. Mandrekar, PhD, Sumithra J. Mandrekar, PhD, Stephen S. Cha, MS, Alfred F. Furth, MS , (June 2006) Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes

[40] Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res. 2010 Oct-Dec; 1(4): 274–278.

[41] Ahmad A. Pathways to breast cancer recurrence. ISRN Oncol. 2013;2013:290568.

[42] Gerber B, Freund M, Reimer T. Recurrent breast cancer: treatment strategies for maintaining and prolonging good quality of life. Dtsch Arztebl Int . 2010;107(6):85-91.

[43] Keenan T, Moy B, Mroz EA, et al. Comparison of the Genomic Landscape Between Primary Breast Cancer in African American Versus White Women and the Association of Racial Differences With Tumor Recurrence. J Clin Oncol, 2015;33(31):3621-7.

[44] Breslow, NE. Analysis of Survival Data under the Proportional Hazards Model. Int Stat Rev. 1975; 43(1): 45–57.

[45] Cox, DR. Regression Models and Life-Tables. J R Stat Soc Series B Stat Methodol. 1972; 34(2):187–220.

[46] Carrel et al. (2009). NOS1AP Regulates Dendrite Patterning of Hippocampal Neurons through a Carboxypeptidase E-Mediated Pathway. J Neurosci. 2009 Jun 24; 29(25): 8248.

[47] Kutzing et al., 2010. Automated Sholl analysis of digitized neuronal morphology at multiple scales. J Vis Exp. 2010 Nov 14;(45). pii: 2354. doi: 10.3791/2354.

[48] Langhammer et al., 2010, Automated Sholl analysis of digitized neuronal morphology at multiple scales: Whole cell Sholl analysis versus Sholl analysis of arbor subregions. Cytometry A. 2010 Dec;77(12):1160-8. doi: 10.1002/cyto.a.20954.

[49] Firestein et al., 1999. Cypin: a cytosolic regulator of PSD-95 postsynaptic targeting. Neuron. 1999 Nov;24(3):659-72.

[50] Rakha, E.A., et al., Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clin Cancer Res, 2009. 15(7): p. 2302-10.

[51] Douglas G Altman , The logrank test J Martin Bland, BMJ, 2004 May 1;328(7447): 1073

[52] RhoA Regulates Dendrite Branching in Hippocampal Neurons by Decreasing Cypin Protein Levels Hongxin Chen and Bonnie L. Firestein

[53] Landgraf M and Evers JF (2005) Control of dendritic diversity. Current Opinion in Cell Biology 17: 690–696.

[54] Kutzing MK, Langhammer CG, Luo V, Lakdawala H, and Firestein BL (2010) Automated Sholl analysis of digitized neuronal morphology at multiple scales. J Visual Exp 45: 2354.

[55] Maxine Chen, Kenyatta G. Lucas, Barbara F. Akum, Gaithri Balasingam, Tamara M. Stawicki, Janine M. Provost, Gary M. Riefler, Rebecka J. Jörnsten, and Bonnie L. Firestein. A Novel Role for Snapin in Dendrite Patterning: Interaction with Cypin. Mol Biol Cell. 2005 Nov; 16(11): 5103–5114.

[56] Roberto Malinow, Yasunori Hayashi, Mirjana Maletic-Savatic, Shahid H. Zaman, Jean-Christophe Poncer, Song-Hai Shi, José A. Esteban, Pavel Osten, and Ken Seidenman, Introduction

of Green Fluorescent Protein (GFP) into Hippocampal Neurons through Viral Infection, Cold Spring Harb Protoc. 2010 Apr;2010(4):pdb.prot5406. doi: 10.1101/pdb.prot5406.

[57] Neuronal nitric oxide synthase: Structure, subcellular localization, regulation, and clinical implications Li Zhou, Dong-Ya Zhu

[58] Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res. 2007;13(15, pt 1):4429-34.

[59] Killelea BK, Yang VQ, Wang SY, et al. Racial Differences in the Use and Outcome of Neoadjuvant Chemotherapy for Breast Cancer: Results From the National Cancer Data Base. J Clin Oncol. 2015;33(36):4267-76.

[60] Kwon, M., et al., BDNF-promoted increases in proximal dendrites occur via CREB-dependent transcriptional regulation of cypin. Journal of Neuroscience, 2011.31(26): p 9735-9745

[61] Akum, B.F., et al., Cypin regulates dendrite patterning in hippocampal neurons by promoting microtubule assembly. Nature Neuroscience, 2004. 7(2): p. 145-152.

[62] O'Neill, K.M., et al., Assessing effects on dendritic arborization using novel Sholl analyses. Frontiers in Cellular Neuroscience, 2015. 9:p.285

[63] Blamey, R.W., et al., Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. Eur J Cancer, 2007. 43(10): p. 1548-55.

[64] Elston, C.W., I.O. Ellis, and S.E. Pinder, Pathological prognostic factors in breast cancer. Crit Rev Oncol Hematol, 1999. 31(3): p. 209-23.

[65] Fitzgibbons, P.L., et al., Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med, 2000. 124(7): p. 966-78.

[66] Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. 406(6797): p. 747-52.

[67] Taneja, S., et al., The mammographic correlations of a new immunohistochemical classification of invasive breast cancer. Clin Radiol, 2008. 63(11): p. 1228-35.

[68] Kobayashi, T., et al., Prognostic significance of the immunohistochemical staining of cleaved caspase-3, an activated form of caspase-3, in gliomas. Clin Cancer Res, 2007. 13(13): p. 3868-74.

[69] Hu, Z., et al., The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics, 2006. 7: p. 96.

[70] Sorlie, T., Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. Eur J Cancer, 2004. 40(18): p. 2667-75.

[71] de Azambuja, E., et al., Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer, 2007. 96(10): p. 1504-13.

[72] MacGrogan, G., et al., Comparison of quantitative and semiquantitative methods of assessing MIB-1 with the S-phase fraction in breast carcinoma. Mod Pathol, 1997. 10(8): p. 769-76.

[73] Spyratos, F., et al., Correlation between MIB-1 and other proliferation markers: clinical implications of the MIB-1 cutoff value. Cancer, 2002. 94(8): p. 2151-9.

[74] Urruticoechea, A., I.E. Smith, and M. Dowsett, Proliferation marker Ki-67 in early breast cancer. J Clin Oncol, 2005. 23(28): p. 7212-20.

[75] Lipponen, P., et al., Apoptosis in breast cancer as related to histopathological characteristics and prognosis. Eur J Cancer, 1994. 30a(14): p. 2068-73.

[76] Pinder, S.E., et al., Assessment of the new proliferation marker MIB1 in breast carcinoma using image analysis: associations with other prognostic factors and survival. Br J Cancer, 1995. 71(1): p. 146-9.

[77] Brekelmans, C.T., et al., Histopathology and growth rate of interval breast carcinoma. Characterization of different subgroups. Cancer, 1996. 78(6): p. 1220-8.

[78] Kim, J.B., R. Stein, and M.J. O'Hare, Tumour-stromal interactions in breast cancer: the role of stroma in tumourigenesis. Tumour Biol, 2005. 26(4): p. 173-85.

[79] Silberstein, G.B., Tumour-stromal interactions. Role of the stroma in mammary development. Breast Cancer Res, 2001. 3(4): p. 218-23.

[80] Richtig, E., et al., Calculated tumour volume as a prognostic parameter for survival in choroidal melanomas. Eye (Lond), 2004. 18(6): p. 619-23.

[81] Speer, J.F., et al., A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Res, 1984. 44(9): p. 4124-30.

[82] Heuser, L.S., et al., The association of pathologic and mammographic characteristics of primary human breast cancers with "slow" and "fast" growth rates and with axillary lymph node metastases. Cancer, 1984. 53(1): p. 96-8.

[83] Olsson, A., et al., Tumour-related factors and prognosis in breast cancer detected by screening. Br J Surg, 2012. 99(1): p. 78-87.

[84] Spratt, J.S. and S.W. Spratt, Medical and legal implications of screening and follow-up procedures for breast cancer. Cancer, 1990. 66(6 Suppl): p. 1351-62.