UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Trabalho de Conclusão de Curso

MODELAGEM DOS EFEITOS DE ALTERAÇÕES GENÉTICAS NA PROGRESSÃO TUMORAL

Ícaro Fernando Fonsêca Braga

Natal/RN
2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

MODELAGEM DOS EFEITOS DE ALTERAÇÕES GENÉTICAS NA PROGRESSÃO TUMORAL

Trabalho de conclusão de curso apresentado ao Departamento de Engenharia Biomédica da Universidade Federal do Rio Grande do Norte para obtenção do título de Graduado em Engenharia Biomédica.

Graduando: Ícaro Fernando Fonsêca Braga

Orientador: Professora Dra. Beatriz Stransky

Co-orientador: Professor Dr. Marcus Nunes

Natal/RN

2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

MODELAGEM DOS EFEITOS DE ALTERAÇÕES GENÉTICAS NA PROGRESSÃO TUMORAL

Banca Examinadora do Trabalho de Conclusão de Curso:

Profa. Dra. Beatriz Stransky Ferreira _____

UFRN – Orientadora

Prof. Dr. Marcus Alexandre Nunes _____

UFRN – Co-orientador

Prof. ---------- _____

UFRN – Avaliador Externo

Prof. ---------- _____

UFRN – Avaliador Externo

Natal/RN

2016

À minha família, que ao longo desses anos nunca deixou de acreditar no meu potencial. Em especial minha mãe, Sônia Fonsêca, que nunca mediu esforços no sentido de me apoiar em minhas decisões e que sempre me ensinou a pensar nos outros antes de mim mesmo. Sem você nada disso seria possível.

AGRADECIMENTOS

À Deus pelo dom da vida, sem a sua presença eu não seria nada.

À minha família, que sempre acreditou em mim em cada luta que travei. Minha mãe, Sônia, por todo o seu amor, do qual nunca serei digno o suficiente. Minha avó, Isaura, meu tio e primo, Sóstenes e Firmo, minha tia e primas, Solange, Karla, Kalyanne e Kyara. Uma família que quando em meio a problemas ainda encontram espaço para sorrir.

À minha namorada, Moema, que mesmo quando eu estive triste nunca desistiu de mim e de me motivar.

À Ordem DeMolay, que desde os meus quatorze anos me ensinou e me ensina como ser uma pessoa melhor e digna. Sem o aprendizado que lá obtive seria metade do homem que sou hoje. Nesta hora, acho por bem fazer uma citação a uma de nossas mais bonitas cerimônias, a das Luzes.

"*No entanto, cada um de vocês traz dentro de seu coração uma chama, um facho para lhe guiar através da escuridão. Se puder fazer esta luz brilhar sobre outra pessoa, se puder penetrar nas profundezas de sua alma e acender a chama que ali está, então aí está sua finalidade de viver".*

À toda a equipe da Engenharia Clínica do Hospital Universitário Onofre Lopes, que nunca pensaram duas vezes antes de me ensinar algo. Em especial, Rafael, que acreditou no meu potencial e deu a oportunidade de participar efetivamente da equipe.

Aos meus amigos e sócios da Airmid, que compartilharam comigo muitos fardos ao longo desse ano. Essa empresa tem muito futuro.

A todos os meus amigos, tenho só que agradecer a Deus por ter me colocado no caminho de todos vocês.

Aos meus orientadores, Beatriz Stransky e Marcus Nunes, que em nenhum momento mediram esforços no sentido de me ajudar a concluir este trabalho. Beatriz, que desde janeiro acreditou em mim e me confiou essa linha de pesquisa e Marcus que desde que se tornou co-orientador não parou de somar ao trabalho. Esta conquista é nossa e muito ainda está por vir.

Junto aos orientadores está toda a equipe do BioMe, Professor Sandro Souza, Vandeclecio, Jorge e André, que contribuíram, cada um com sua área de saber, para o desenvolvimento e conclusão desse trabalho de conclusão de curso.

Por fim, a todos os professores que me acompanharam desde pequeno e que formaram o cidadão que sou hoje.

SUMÁRIO

LISTA DE FIGURAS..... 7

LISTA DE TABELAS..... 8

LISTA DE EQUAÇÕES..... 9

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS..... 10

RESUMO..... 11

ABSTRACT..... 12

1 INTRODUÇÃO..... 13

2 OBJETIVOS..... 14

2.1 GERAL..... 14

2.2 ESPECÍFICOS..... 15

3 REVISÃO BIBLIOGRÁFICA..... 16

3.1 Câncer..... 16

3.2 Câncer como um processo evolutivo..... 17

3.2.1 Autossuficiência em sinais de crescimento..... 19

3.2.2 Insensibilidade à sinais anti-crescimento..... 19

3.2.3 Evasão de apoptose..... 19

3.2.4 Potencial de replicação ilimitado..... 19

3.2.5 Manutenção da angiogênese..... 20

3.2.6 Invasão de tecidos & Metástase..... 20

3.3 Genoma do Câncer..... 20

3.4 Modelos de Dinâmica Evolutiva do Câncer..... 21

3.5 Modelo estatístico estocástico de Galton-Watson..... 22

4 METODOLOGIA..... 24

4.1 Levantamento de dados e parâmetros da dinâmica evolutiva..... 24

4.1.1 Valores de crescimento..... 24

4.1.2 Proporção *Driver-Passenger*..... 25

4.2 Implementação e simulação..... 25

4.2.1 Modelos – dinâmica evolutiva..... 25

4.2.1 Parâmetros..... 27

4.3 Bioinformática..... 27

4.3.1 Banco de dados DB Mutation..... 28

4.3.2 Análise das mutações genéticas através de preditores..... 28

4.4 Análise Estatística – Modelo de regressão..... 30

5 RESULTADOS E DISCUSSÃO..... 32

5.1 Modelo Estocástico e Simulações..... 32

5.2 Bioinformática e Análise de Dados Genômicos..... 37

6 CONCLUSÃO..... 40

7 REFERÊNCIAS..... 41

LISTA DE FIGURAS

Figura 1 - Tempo necessário para uma nova mutação do tipo *driver* (BOZIC et al., 2010). 18

Figura 2 - Capacidades adquiridas do câncer (adaptado de HANAHAN et al., 2000). 18

Figura 3 - Comportamento básico de um processo de ramificação. 23

Figura 4 - Fluxograma das etapas do projeto. 24

Figura 5 - Dinâmica de evolução tumoral e os possíveis resultados para cada célula, em cada iteração (adaptado de ROGOES, 2010). 26

Figura 6 – Quantidade de simulações obtidas com o número máximo de *drivers*..... 33

Figura 7 – Representação das simulações – número de células por tempo de simulação (idade do tumor). 35

Figura 8 – Número de *passengers* obtidos em simulações com número máximo de *drivers*..... 36

Figura 9 – Proporção de *drivers* e *passengers* para a base dados COSMIC. 38

Figura 10 – Proporção de *drivers* e *passengers* para a base dados TCGA. 39

LISTA DE TABELAS

Tabela 1 - Estimativa do número de casos de câncer em 2016 no Brasil (INCA, 2016). 17

Tabela 2 - Distribuição dos dez tipos de câncer mais comuns estimados para 2016 por sexo (INCA, 2016). 17

Tabela 3 – Parâmetros e respectivos valores utilizados no modelo. 27

Tabela 4 – Preditores utilizados no DB Mutation e suas definições (LIU, 2003). 29

LISTA DE EQUAÇÕES

Equação 1 - Equação base do método de Galton-Watson. 23

Equação 2 – Determinação do número de mutações *passenger*..... 25

Equação 3 – Cálculo da probabilidade de estagnação..... 26

Equação 4 – Valor do R^2. 30

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS

BiOME – Núcleo Interdisciplinar de Bioinformática

CHASM – Cancer-specific High-throughput Annotation of Somatic Mutations

COSMIC – Catalog of Somatic Mutations in Cancer

IARC – International Agency for Research on Cancer

IMD – Instituto Metrópole Digital

INCA – Instituto Nacional de Câncer José de Alencar Gomes da Silva

OMS – Organização Mundial da Saúde

TCGA – The Cancer Genome Atlas

BRAGA, Ícaro Fernando. **Modelagem dos efeitos de alterações genéticas na progressão tumoral**. Trabalho de Conclusão de Curso, Graduação em Engenharia Biomédica, Universidade Federal do Rio Grande do Norte, 44p., 2016.

RESUMO

Alterações genéticas que predispõem um indivíduo a um determinado câncer tem sido identificadas em estudos realizados na escala de todo o genoma. O câncer é resultado de alterações genéticas que lhe garantem vantagens adaptativas, como por exemplo a habilidade de desenvolver metástase. Neste contexto, as sucessivas mutações decorrentes do crescimento tumoral proporcionam um aumento na vantagem adaptativa da célula que sofre estas mutações e tem como consequência a expansão do câncer. No entanto, a nossa compreensão dos aspectos que referem-se à dinâmica da replicação, sobrevivência e evolução da população de células tumorais ainda não atingiu o nível dos avanços obtidos na genética molecular. Neste trabalho de conclusão de curso utilizamos um modelo matemático baseado no processo de ramificação estocástica (*branching process*) capaz de relacionar o crescimento tumoral ao acréscimo de mutações *driver*, responsáveis por este crescimento tumoral, e às mutações *passenger*, que não influenciam no crescimento. A partir deste modelo, comparamos os resultados de simulação com dados genômicos de amostras provenientes do banco de dados DB Mutation do Núcleo Interdisciplinar de Bioinformática da Universidade Federal do Rio Grande do Norte. Após realizarmos cem mil simulações no cluster do Instituto Metrópole Digital, obtivemos resultados que são similares a literatura atual de dinâmica evolutiva. Nossos resultados suportam a premissa que após adquirir cerca de quatro mutações *driver* o tumor será hábil o suficiente para se adaptar ao organismo, sendo difícil morrer. Observamos ainda que o número de *passengers* decorrentes das sucessivas mutações *driver* cresce conforme as mutações *drivers* aumentam, mas que após certo tempo estabilizam em 29 *passengers*, em razão do tempo de simulação proposto. Esperamos que a introdução de dados genômicos no modelo matemático gere novas informações e hipóteses acerca da dinâmica evolutiva de tumores e possibilite modelarmos futuramente aplicações de tratamentos experimentais em ambiente simulado.

Palavras-chave: Câncer. Dinâmica evolutiva. Modelo estocástico. Mutação '*Driver*'.

BRAGA, Ícaro Fernando. **Modeling of the effects of genetic alterations in tumor progression.** Conclusion Work Project, Biomedical Engineering Bachelor Degree, Federal University of Rio Grande do Norte, 44p., 2016.

ABSTRACT

Genetic changes that predispose an individual to a particular cancer have been identified in studies conducted at the whole genome scale. Cancer is the result of genetic changes that ensure adaptive advantages, such as an ability to develop metastasis. In this context, the successive mutations resulting from tumor growth provide an increase in the adaptive advantage of the cell that undergoes these mutations and has a consequence in cancer expansion. However, our understanding of the problems that refer to the dynamics of replication, survival and evolution of the tumor cell population has not yet reached the level of advances in molecular genetics. In this work, we use a mathematical model based on the process of stochastic branching (branch process) capable of relating the tumor growth to the addition of driver mutations, which confer growth of the tumor growth, and the passenger mutations, who do not influence any growth. From this model, we compare the simulation results with the genomic data samples from the database DB Mutation of the Interdisciplinary Nucleus of Bioinformatics of the Federal University of Rio Grande do Norte. After performing one hundred thousand simulations in the cluster of the Digital Metropolis we obtained results that are similar to a current evolutionary dynamics literature. Our results support a premise that after obtaining four driver mutations the tumor will be sufficient able to adapt to the organism, being hard to die. We also observed that the number of passengers resulting from the successive driver mutations increases as the number of driver increase, but after a certain time the stabilize in 29 passengers, due to the proposed simulation time. We hope that the introduction of genomic data into the mathematical model will generate new information and hypotheses about the evolutionary dynamics of tumors and will enable future modeling of applications of experimental treatments in a simulated environment.

Keywords: Cancer. Evolutionary Dynamics. Stochastic model, ‘Driver’ mutation.

1 INTRODUÇÃO

Segundo o *World Cancer Report 2014* da *International Agency for Research on Cancer*, braço da Organização Mundial de Saúde, é inquestionável a premissa de que o câncer é um problema de saúde pública. Espera-se que, nas próximas décadas, o câncer na população corresponda a 80% dos mais de 20 milhões de novos casos estimados para 2025 (INCA, 2014).

Após centenas de anos de avanços na ciência, é possível observar a quantidade de informação adquirida na pesquisa do câncer, revelando-o como uma doença baseada em mudanças no genoma. Embora não haja registro científico inaugural concreto que fale sobre a doença, existem menções a egípcios, persas e indianos que identificaram tumores malignos (STRATTON, 2009). A base de toda a pesquisa está na descoberta de alterações nos genes que modificam a função do produto gênico – a proteína, e que geram um ganho ou perda de função da mesma. Este ganho ou perda de função se relaciona com a aquisição de marcos que o tumor precisa para se proliferar. Estes comportamentos foram identificados através de observações de alterações em células cancerígenas humanas e de animais (STRATTON, 2009).

O câncer pode ter origem a partir alterações causadas por de fatores internos, como uma mutação genética herdada de um dos progenitores, e externos, ocasionada por algum agente ambiental mutagênico, sendo que ambos podem levar a um aumento de replicação celular. No início do desenvolvimento tumoral, a célula inicialmente não possui alta vantagem adaptativa, sendo o processo de transformação lento e dependente do surgimento de novas mutações que irão, ao longo do tempo, garantir a sobrevivência do tumor. Entretanto, após acumular uma quantidade suficiente de mutações esta célula possuirá uma vantagem em relação às outras. Além de proliferar mais rápido, estas modificações poderão conferir outras habilidades que lhe proporcionem uma melhor adaptação e crescimento, tais como evadir sinais de apoptose celular ou vascularização do tumor (HANAHAN; WEINBERG, 2000).

Desta forma, um ponto chave da pesquisa atualmente está em distinguir o surgimento de mutações do tipo *driver*, que conferem uma vantagem adaptativa à célula e mutações *passenger* (mutações neutras), que não apresentam este efeito (BOZIC; NOWAK, 2010). Entretanto, métodos atuais têm obtido fracos resultados na tentativa de modelar a dinâmica de uma população tumoral em vista das alterações genéticas, como

por exemplo, o de Wright-Fisher, que utiliza uma perspectiva randômica que não avalia o ganho de replicação adicional das células a cada nova mutação do tipo *driver* (BEERENWINKEL et al, 2007).

A investigação deste ganho sucessivo de vantagem adaptativa e seu efeito na progressão do câncer é o que motiva essa pesquisa. Uma célula com uma mutação *driver* no início de um processo tumorigênico possui pouco valor adaptativo, porém, produz alterações no crescimento do tumor que poderá ocasionar uma nova mutação com mais facilidade. Portanto, a cada nova mutação *driver,* teremos um ganho de replicação, e em consequência, um aumento na possibilidade de novas mutações. Esta irá conferir um novo fator adaptativo, como por exemplo a ineficiência de sinais para apoptose celular. Vê-se claramente que a cada nova mutação do tipo *driver*, haverá uma nova mudança na qualidade adaptação do tumor, corroborando para a sua expansão (STRATTON *et al*, 2009).

Esse trabalho consiste em implementar um modelo capaz de simular a dinâmica evolutiva do câncer a partir de suas mutações *driver* e comparar os dados da simulação com dados reais, obtidos através de sequenciamento e análise do genoma de pacientes. O estudo teve como base a pesquisa feita por Bozic em 2010, onde através de um modelo estocástico foi possível entender melhor a participação das mutações do tipo *driver* na vantagem adaptativa e crescimento dos tumores. Para comparar nossos dados de simulações, também realizamos estudo com dados reais existentes no banco de dados DB Mutation.

2 OBJETIVOS

2.1 GERAL

Implementar um modelo matemático-computacional de dinâmica evolutiva, capaz de simular a progressão de crescimento de tumores, para avaliar a influência das mutações tipo *driver* no crescimento tumoral.

2.2 ESPECÍFICO

Os objetivos específicos consistem em:

- Implementar o modelo matemático em ambiente computacional;
- Simular a dinâmica de crescimento tumoral a partir de parâmetros relativos ao câncer de mama – como u, taxa de aquisição de mutação; v, taxa de aquisição de mutações neutras, s, vantagem adaptativa, T, tempo de dobramento (iteração) e t, número de iterações entre t_0 e o início da última linhagem;
- Analisar os genomas de tumores de mama do banco de dados DB Mutation;
- Comparar os resultados da simulação com informações obtidas da análise bioinformática (dados DB Mutation) em relação à proporção de mutações *passengers* e *drives.*

3 REVISÃO BIBLIOGRÁFICA

Nessa seção será apresentada uma revisão bibliográfica que contempla os tópicos mais relevantes desse trabalho. Os assuntos abordados aqui irão contribuir para um melhor entendimento do conteúdo que será apresentado nos tópicos seguintes. A primeira parte dessa seção apresentará uma revisão do câncer, evidenciando os efeitos de mudanças genéticas no processo de crescimento tumoral. Posteriormente será abordado o método estatístico em que foi baseada esta modelagem, incluindo aspectos de sua teoria e modelagem.

3.1 Câncer

O câncer tem recebido bastante atenção ao longo dos anos devido à sua alta incidência na população. A estimativa mundial realizada em 2012, pelo projeto Globocan/Iarc, apontou que dos 14 milhões de novos casos estimados, mais de 60% ocorreram em países em desenvolvimento. Para a mortalidade, a situação é ainda mais agravante, pois foi constatado que dos 8 milhões de óbitos previstos, 70% ocorreram nesses mesmos países (IARC, 2014). Os tipos de câncer mais incidentes no mundo foram pulmão (1,8 milhão), mama (1,7 milhão), intestino (1,4 milhão) e próstata (1,1 milhão). Nos homens os mais frequentes foram pulmão, próstata, intestino, estômago e fígado. Em mulheres, os cânceres de mama, intestino, colo do útero, pulmão e estômago figuraram entre os principais.

A estimativa no Brasil (Figura 1 e Tabela 1) para o biênio de 2016/2017 aponta a ocorrência de cerca de 600 mil casos novos de câncer. Este perfil se assemelha aos da América Latina e Caribe, onde a maioria dos países se encontra também em desenvolvimento. Com isso, constatamos um problema de saúde pública, onde esta doença se torna cada vez mais comum, com o aumento da expectativa de vida da população (INCA, 2016).

Tabela 2 - Estimativa do número de casos de câncer em 2016 no Brasil (INCA, 2016).

Localização Primária Neoplasia Maligna	Estimativa dos Casos Novos							
	Homens				Mulheres			
	Estados		Capitais		Estados		Capitais	
	Casos	Taxa Bruta	Casos	Taxa Bruta	Casos	Taxa Bruta	Casos	Taxa Bruta
Próstata	61.200	61,82	13.940	64,93	-	-	-	-
Mama Feminina	-	-	-	-	57.960	56,20	18.990	79,37
Colo do Útero	-	-	-	-	16.340	15,85	4.550	19,07
Traqueia, Brônquio e Pulmão	17.330	17,49	4.430	20,59	10.890	10,54	3.230	13,49
Cólon e Reto	16.660	16,84	5.560	25,80	17.620	17,10	6.210	25,95
Estômago	12.920	13,04	3.130	14,54	7.600	7,37	2.180	9,07
Cavidade Oral	11.140	11,27	2.780	12,95	4.350	4,21	1.230	5,04
Laringe	6.360	6,43	1.600	7,50	990	0,94	320	0,97
Bexiga	7.200	7,26	2.110	9,79	2.470	2,39	830	3,21
Esôfago	7.950	8,04	1.460	6,75	2.860	2,76	610	2,27
Ovário	-	-	-	-	6.150	5,95	2.170	8,92
Linfoma de Hodgkin	1.460	1,46	450	1,74	1.010	0,93	400	1,33
Linfoma não Hodgkin	5.210	5,27	1.550	7,15	5.030	4,88	1.670	7,02
Glândula Tireoide	1.090	1,08	350	1,27	5.870	5,70	1.800	7,46
Sistema Nervoso Central	5.440	5,50	1.290	5,86	4.830	4,68	1.250	5,20
Leucemias	5.540	5,63	1.370	6,38	4.530	4,38	1.180	4,88
Corpo do Útero	-	-	-	-	6.950	6,74	2.530	10,47
Pele Melanoma	3.000	3,03	840	3,86	2.670	2,59	740	2,96
Outras Localizações	51.850	52,38	11.890	55,45	47.840	46,36	11.820	49,33
Subtotal	214.350	216,48	52.750	245,63	205.960	199,57	61.710	257,55
Pele não Melanoma	80.850	81,66	17.370	80,90	94.910	91,98	21.910	91,65
Todas as Neoplasias	295.200	298,13	70.120	326,51	300.870	291,54	83.620	348,99

*Números arredondados para múltiplos de 10.

Tabela 2 - Distribuição dos dez tipos de câncer mais comuns estimados para 2016 por sexo (INCA, 2016).

Localização Primária	Casos	%	Homens	Mulheres	Localização Primária	Casos	%
Próstata	61.200	28,6%			Mama feminina	57.960	28,1%
Traqueia, Brônquio e Pulmão	17.330	8,1%			Cólon e Reto	17.620	8,6%
Cólon e Reto	16.660	7,8%			Colo do útero	16.340	7,9%
Estômago	12.920	6,0%			Traqueia, Brônquio e Pulmão	10.890	5,3%
Cavidade Oral	11.140	5,2%			Estômago	7.600	3,7%
Esôfago	7.950	3,7%			Corpo do útero	6.950	3,4%
Bexiga	7.200	3,4%			Ovário	6.150	3,0%
Laringe	6.360	3,0%			Glândula Tireoide	5.870	2,9%
Leucemias	5.540	2,6%			Linfoma não Hodgkin	5.030	2,4%
Sistema Nervoso Central	5.440	2,5%			Sistema Nervoso Central	4.830	2,3%

*Números arredondados para múltiplos de 10.

3.2 Câncer como um processo evolutivo

Algumas linhas de pesquisa evidenciam que o desenvolvimento de tumores é um processo composto por várias etapas e que essas etapas refletem mudanças genéticas que conduzem a progressiva transformação de células normais para tumorais. Estas transformações são causadas por mutações que ocorrem de forma estocástica. Estas mutações podem ser classificadas de acordo com a sua influência no estabelecimento e desenvolvimento do tumor. Mutação do tipo *driver* garantem às células uma maior adaptabilidade e sobrevivência no organismo, e a cada nova mutação *driver* podemos

observar um aumento na vantagem adaptativa do tumor. Mutações do tipo *passenger* não presentam nenhum efeito (RENAN, 1993).

Além disso, podemos observar um comportamento onde a cada nova mutação do tipo *driver*, o tempo necessário para a ocorrência de uma nova mutação é menor (Figura 1). Portanto, é fato a interdependência do número de mutações do tipo *driver* e o crescimento tumoral.

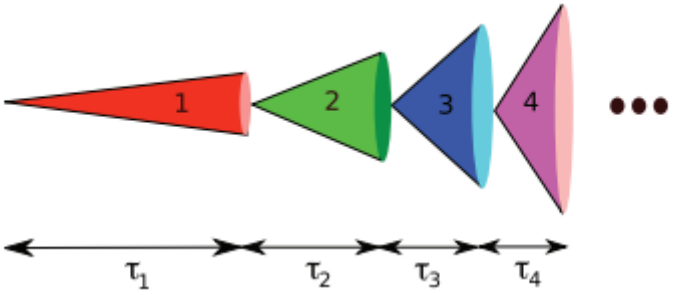

Figura 1 - Tempo necessário para uma nova mutação do tipo *driver* (BOZIC et al., 2010).

Nesse contexto, pesquisadores relacionaram uma lista de capacidades que uma célula mutada pode adquirir para sua adaptação (HAHN et al., 1999). Elas são sustentadas pelas novas mutações que ocorrem a partir da primeira mutação e que conferem ao tumor uma alta vantagem adaptativa (Figura 2).

Figura 2 - Capacidades adquiridas do câncer (adaptado de HANAHAN et al., 2000).

3.2.1 Autossuficiência em sinais de crescimento

Células normais requerem sinais de crescimento antes de mudarem de um estado de quiescência - que é um estado onde a célula se encontra metabolicamente ativa porém não prolifera, para um estado ativo de proliferação. Esses sinais são transmitidos para a célula por receptores transmembrana que ligam distintas classes de sinalizadores: fatores de crescimento, componentes da matriz extracelular e moléculas de adesão/interação célula-célula. Não existe na natureza nenhum tipo de célula normal que consiga se proliferar sem a presença de fatores de crescimento. Muitos oncogenes, genes relacionados ao surgimento de tumores, agem imitando um comportamento exacerbado de sinalização de crescimento (HANAHAN; WEINBERG, 2000).

3.2.2 Insensibilidade à sinais anti-crescimento

Em um tecido normal, múltiplos sinais anti-proliferativos operam para manter a quiescência celular e homeostase. Entretanto, células cancerígenas são capazes de evadir esses sinais, de modo a continuar seu crescimento. As células agem alterando diretamente seu ciclo de replicação, passando despercebidas por estes sinais (HANAHAN; WEINBERG, 2000).

3.2.3 Evasão de apoptose

A habilidade de celulares tumorais se expandirem está ligada não apenas com a habilidade de se proliferar, como também com a habilidade de impedir a morte celular programada – a apoptose. A apoptose ocorre quando o ambiente ou fatores internos enviam sinais para que as células morram, uma vez que não conseguem executar suas funções propriamente. Sendo assim, um tumor em expansão deve perpassar esta barreira, de modo a tornar esses sinais de morte programada não funcionais (HANAHAN; WEINBERG, 2000).

3.2.4 Potencial de replicação ilimitado

As capacidades explicadas anteriormente poderiam constituir sozinhas as vantagens necessárias para o crescimento completo de um tumor, no entanto recentes estudos apresentaram que este crescimento desordenado e evasão da comunicação entre células que gera a apoptose, em si, não é suficiente para garantir o crescimento tumoral. Um exemplo é o câncer de mama, onde muitos e talvez todos os tipos de células mamárias carregam um programa autônomo que limita seu crescimento. Este programa opera independente das conexões intercelulares, sendo assim, constitui-se mais uma

barreira a ser transposta no sentido de que o tumor se torne macroscópico e letal (HANAHAN; WEINBERG, 2000).

3.2.5 Manutenção da angiogênese

Uma grande barreira no crescimento tumoral se encontra no fato de que em certo ponto de seu crescimento ele assume um tamanho muito grande. A corrente sanguínea, e consequentemente os nutrientes, não conseguem alcançar o interior do tumor e células podem entrar em processo de necrose. O tumor, para permanecer vivo, deve criar canais sanguíneos que cheguem até o seu interior e levem alimento e oxigênio para as células. A partir do momento em que essa barreira é quebrada, o tumor irá trabalhar como um organismo autossuficiente (HANAHAN; WEINBERG, 2013).

3.2.6 Invasão de tecidos & Metástase

Finalmente, após sucessivos incrementos em sua vantagem adaptativa o tumor poderá invadir outros tecidos e se alastrar pelo corpo. Este processo é conhecido como metástase, onde as células cancerígenas, agora com alto valor adaptativo, podem se desprender do tecido original, migrar via circulação sanguínea ou linfática e estabelecer novos focos tumorais em outras partes do corpo (HANAHAN; WEINBERG, 2013).

3.3 Genoma do Câncer

Como em todas as células do nosso corpo, as células cancerígenas são descendentes diretas de divisões mitóticas e, sendo assim, carregam uma cópia do seu código genético. Entretanto, este DNA que as células cancerígenas carregam possui uma série de diferenças do seu progenitor, devido a mutações e outras alterações no genoma da célula.

Estas alterações incluem substituição de bases, inserções ou deleção de pequenos ou grandes segmentos do DNA, rearranjos de segmentos quebrados em outras partes do genoma, etc. Estas alterações podem ser resultado de vários fatores como predisposição genética advinda dos progenitores ou provocadas por fatores externos, como é o caso de queimaduras solares que induzem o câncer de pele.

Um ponto que se deve ter em mente é o de que todos os cânceres são resultado de mutações, porém nem toda mutação está envolvida no desenvolvimento do câncer. Para isso, é necessário introduzir o conceito de mutações do tipo *driver* e mutações do tipo *passenger*.

Uma mutação *driver* está intimamente relacionada ao processo de oncogênese, ela proporciona ao câncer um aumento em sua vantagem adaptativa, conferindo-lhe um aumento no seu crescimento. Já as mutações *passenger* não conferem ao câncer aumento no crescimento, nem, portanto, possuem influência no seu desenvolvimento. Elas são mutações decorrentes desse processo de tumorigênese e têm, comumente, seu número aumentado de acordo com o aumento das mutações do tipo *driver*.

3.4 Modelos de Dinâmica Evolutiva do Câncer

Modelos de dinâmica evolutiva ajudam no entendimento de como tumores se desenvolvem e podem desempenhar um papel essencial no estabelecimento do prognóstico como na previsão de resultados de intervenções médicas (WACLAW et al, 2015). Um exemplo disso é na terapia focada em alvos, na qual a medicação é liberada no local alvo (tumor), onde poderá ter maior efetividade. Bozic e colaboradores desenvolveram um modelo de crescimento tumoral densidade-dependente, e levando em consideração o tamanho do tumor, foi possível identificar que tanto o crescimento tumoral quanto a distribuição de células-filhas são limitadas pela geometria e fatores metabólicos. Nesse sentido alguns modelos matemáticos têm sido uma ferramenta essencial para informar que combinação de terapias podem minimizar a resistência do câncer (BOZIC et al, 2013).

Em 2013, Bozic e colaboradores apresentaram um modelo matemático capaz de descrever a dinâmica evolutiva de lesões em resposta ao tratamento. Foram estudados vinte pacientes com melanoma onde se reproduziram as características e submeteram-se ao tratamento em diversas modalidades (sequencial, simultâneo, combinações diferentes de drogas, etc.). Identificou-se que muitas vezes o melhor tratamento combinado e simultâneo com duas drogas é bastante eficaz no controle da doença, e que em caso de necessidade maior, tumor mais agressivo, uma boa opção seria a aplicação de até três drogas.

Na mesma linha, diversos modelos matemáticos foram desenvolvidos e implementados computacionalmente com o intuito de auxiliar o entendimento e testar hipóteses acerca da biologia do câncer. A Teoria Multiestágio tem sua base na probabilidade do desenvolvimento do câncer advir da idade, porém este método não observa a cinética da criação do tumor e sua progressão, considerando apenas que

existem etapas que devem ser alcançadas durante o crescimento tumoral (ARMITAGE, 1954).

Posteriormente, observou-se que a metodologia mais interessante na modelagem de dinâmicas evolutivas estava nos modelos estocásticos, onde o comportamento é muito próximo do biológico das mutações, ou seja, aleatório, não observando características determinísticas sem base biológica. Modelos estocásticos foram desenvolvidos tanto para populações mistas quanto para populações estruturadas, onde é possível implementar diferentes parâmetros (NOWAK, 2006).

Ainda existem modelos determinísticos, onde é possível, através de sistemas dinâmicos, descrever o comportamento médio de sistemas evolutivos, sendo estes mais fáceis de implementar, mas podem negligenciar o caráter aleatório que existe na biologia (BEERENWINKEL et al, 2014).

Dentre os estudos feitos, uma proposição mais adequada é que a evolução tumoral está intimamente ligada com o número de mutações *driver* que o tumor possui, pois conferem ao tumor a vantagem adaptativa capaz de transpor as barreiras descritas anteriormente. Dessa forma, o seu número representa uma grande vantagem adaptativa para o câncer. Assim, um interessante exemplo de modelo matemático baseado no caráter estocástico da biologia e que ainda traduz uma relação das mutações *driver* com as mutações *passenger* é o caso do modelo proposto por Bozic em 2010. Nele, foram estabelecidos parâmetros da biologia do câncer e feitas simulações de acordo com os mesmos. Ao final, os dados simulados foram comparados com os dados reais, câncer de pâncreas e gliobastoma, em relação a proporção de células com mutações *driver* e *passenger*, para classificação das *drivers* foi utilizado o filtro CHASM (Cancer-specific High-throughput Annotation of Somatic Mutations) (CARTER, 2009). Foi constatado que ambos gráficos possuem um comportamento semelhante com uma curva logarítmica, e ainda que em decorrência do aumento das mutações *driver* houve também um aumento das mutações *passenger* (BOZIC et al., 2010).

3.5 Modelo estocástico de Galton-Watson

O processo de Galton-Watson é um processo de ramificação estocástico que foi utilizado inicialmente para determinar por quanto tempo o sobrenome de uma família iria persistir através de gerações. Por sua característica aleatória este método é de interesse e aderente à pesquisa, uma vez que este método combina elementos de caráter

estocástico e possível estabelecer parâmetros de crescimento populacionais e separá-los em resultados possíveis.

Equação 2 - Equação base do método de Galton-Watson.

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)} \qquad \{\xi_j^{(n)} : n, j \in \mathbb{N}\}$$

Nesta equação temos X_n como sendo o número de descendentes (ao longo da linhagem masculina) da n-ésima geração e $\varepsilon_j^{(n)}$ o número de filhos do sexo masculino da j-ésima linhagem desses descendentes.

Na tentativa de estimar o tempo de vida de um sobrenome familiar foi calculado a probabilidade de cada progenitor ter nenhum filho (assexuado), um filho ou dois filhos. Portanto, a partir da metodologia de ramificação, foi estimado o resultado possível de gerações às quais aquele sobrenome poderia sobreviver.

Com este método é possível estabelecer as probabilidades relacionadas a determinados comportamentos (continuidade, duplicação, morte) como exemplificado na Figura 3. Mais tarde, na metodologia desse trabalho, iremos identificar as semelhanças dessa situação com os cenários prováveis de uma divisão celular.

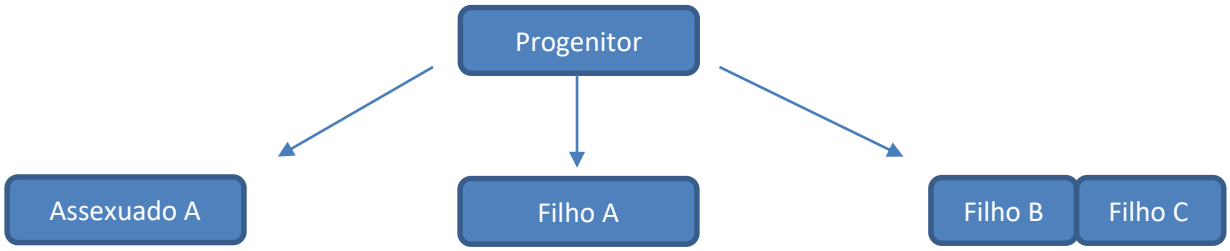

Figura 3 - Comportamento básico de um processo de ramificação.

4 METODOLOGIA

O desenvolvimento do projeto foi organizado em três macro etapas. Estas e suas subdivisões estão relacionadas no fluxograma apresentado na Figura 4.

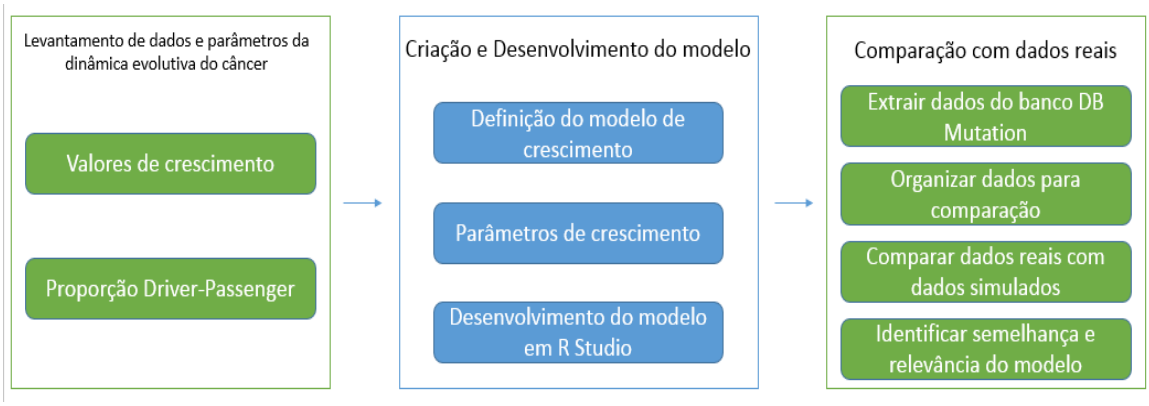

Figura 4 - Fluxograma das etapas do projeto.

4.1 Levantamento de dados e parâmetros da dinâmica evolutiva

Nesse estudo foi necessário adquirir informações relativas à dinâmica evolutiva do câncer e seus parâmetros. Para tanto, foi buscado na literatura valores representativos do crescimento tumoral e da relação existente entre mutações *driver* e *passenger*.

4.1.1 Valores de crescimento

Para o modelo em questão existem alguns parâmetros que possuem influência na modelagem, entre eles estão: taxa de mutação pontual, válido para qualquer tipo de mutação, taxa de mutação *driver* - u, vantagem adaptativa - s, tempo de dobramento (vida útil de uma célula até a próxima divisão, nosso passo) - T, número de mutações *driver* – k, número de mutações *passenger* – n e taxa de aquisição de mutações neutras – v.

Dentre estes parâmetros existem algumas considerações a se fazer. A taxa de mutação pontual não possui aplicabilidade no nosso modelo. Uma vez que estamos querendo observar a influência que mutações do tipo *driver* tem no crescimento tumoral, podendo, assim, o uso dessa taxa implicaria na observância desnecessária de valores como as mutações *passenger*.

Como dito no parágrafo anterior, o interesse desse estudo está na observação da consequência que as mutações *driver* geram. Sendo assim, parâmetros como taxa de mutação *driver*, vantagem adaptativa e o número de mutações *driver* serão fator preponderante na construção do modelo. Em adição, temos fatores que não podem faltar como tempo de dobramento e a taxa de aquisição de mutações neutras, para, em aspecto futuro, realizar a contagem das mutações *passenger* que aconteceram ao longo da simulação.

4.1.2 Proporção *Driver-Passenger*

Embora o modelo seja estocástico, a literatura, de acordo com nosso estudo, ainda não encontrou uma modelagem definitiva para determinar o número de mutações *passenger*. Sendo necessário que esta parte do nosso modelo seja determinada através de um cálculo aproximado que será feito ao final de cada simulação.

Para isso, utilizaremos a seguinte equação:

Equação 2 – Determinação do número de mutações *passenger*.

$$\boldsymbol{n(t) = \frac{vt}{T}}$$

Nesta equação temos que *n* é o número total de mutações *passenger*, *v*, taxa de aquisição de mutações neutras e *T*, tempo de dobramento. *T* e *v* são valores constantes e dependentes de qual tipo de câncer se quer simular, já *t* está condicionado ao número de passos dados enquanto o tumor existe.

4.2 Implementação e simulação

O modelo foi gerado em software R Studio© (*2016 The R Studio*) a partir dos parâmetros disponíveis na literatura e um pacote auxiliar de computação paralela, uma vez que foram feitas cerca de cem mil simulações com parâmetros para câncer de mama. As simulações foram executadas em um do Instituto Metrópole Digital em parceria com o BioMe (Núcleo Interdisciplinar de Bioinformática) da Universidade Federal do Rio Grande do Norte. Além disso, foi necessário adicionar fatores limitantes ao código como tempo de execução total e número máximo compreensível de mutações *driver*.

4.2.1 Modelo – dinâmica evolutiva

A dinâmica do modelo evolutivo de câncer, semelhante ao modelo de Galton-Watson com os sobrenomes, se baseia em três possíveis cenários: geração de duas células filhas iguais à progenitora; geração uma célula filha igual à progenitora e outra com uma nova mutação *driver*; e morte. Isso pode ser observado na Figura 5, a seguir:

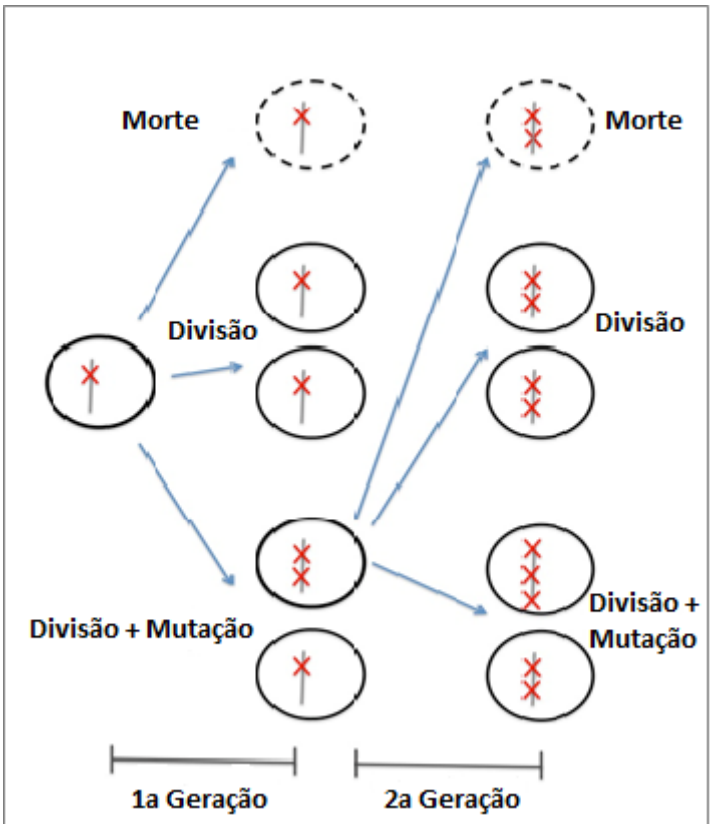

Figura 5 - Dinâmica de evolução tumoral e os possíveis resultados para cada célula, em cada iteração (adaptado de ROGOES, 2010).

Para o modelo em questão as probabilidades de cada evento ocorrer são dependentes da Equação 3, que representa a probabilidade de estagnação de uma célula, não há alteração na quantidade total do tumor, com *k* mutações *driver*, sendo *s* a vantagem adaptativa inicial.

Equação 3 – Cálculo da probabilidade de estagnação.

$$\boldsymbol{d_k = \frac{1}{2}(1 - s)^k}$$

Esta probabilidade, caso s seja igual a 0,04%, será de 49,8%. O que representa que a população de células tem uma chance menor de morrer ou estagnar, mantendo assim o tamanho populacional anterior. E, sendo assim, possui a chance de 50,2% de se dividir em duas novas células mutadas.

4.2.2 Parâmetros

Os valores dos parâmetros foram obtidos na literatura e estão relacionados na Tabela 3. O valor para a taxa de aquisição de mutações *driver* foi calculado com base na multiplicação da taxa de aquisição de mutações pontuais em tumores, $\sim 5\ x\ 10^{-10}$, pelo número de posições no genoma que podem se tornar mutações driver, $\sim 3{,}4\ x\ 10^{4}$. Adicionalmente, foi estabelecido um limite temporal para a simulação, e o tempo escolhido foi de 20 anos. Este valor também foi observado na literatura e representa um intervalo confiável para análise comparativa.

Em razão da dificuldade de determinar o número de mutações *passenger*, foi determinado que o código irá contar os passos dados até a perpetuação da última linhagem. Com isso, utilizaremos a Equação 2 para estimar o valor de mutações *passenger* a partir de *t*, número de passos para determinado tempo de dobramento. Por fim, estipulamos a população inicial como sendo de apenas 1 célula com mutação *driver*.

Tabela 3 – Parâmetros e respectivos valores utilizados no modelo.

Parâmetro	Câncer de mama	Referência
Taxa de mutação *driver*	u = 3,4 x 10^{-5}	(BOZIC, 2010)
Vantagem adaptativa	s = 0,4% (0,004)	(BOZIC, 2010)
Tempo de dobramento	T = 2 dias	(DAI, 2015)
Taxa de aquisição de mutações neutras	v = 0,016	(BOZIC, 2010)
População inicial	1 célula	Escolhido
Tempo total de simulação	20 anos	Escolhido

4.3 Bioinformática

Para avaliar a qualidade do modelo, comparamos os resultados obtidos pela simulação com os resultados obtidos da análise dos genomas disponíveis no banco de dados DB Mutation.

4.3.1 Banco de dados DB Mutation

Os dados para análise do número de mutações *drivers* e *passengers* de indivíduos foram obtidos do banco de dados DB MUTATION, desenvolvido pelo BioMe (Núcleo Interdisciplinar de Bioinformática) na Universidade Federal do Rio Grande do Norte (BioInfo-IMD, 2016). O DB Mutation compreende informações dos bancos de dados COSMIC (Catalog of Somatic Mutations In Cancer) e TCGA (The Cancer Genome Atlas). Estes bancos possuem amostras de sequenciamento dos mais diversos tipos de câncer, sendo o conjunto deles a maior fonte de dados sobre genoma disponível atualmente (THE SCIENTIST, 2012). Destes, utilizamos 30 amostras de cada banco, referentes ao câncer de mama.

O DB Mutation acrescenta algumas informações ao usar um conjunto de técnicas para analisar se as mutações encontradas em cada amostra possuem fator de influência no crescimento tumoral. As mutações são avaliadas através de nove parâmetros, os chamados preditores, que são contabilizados para formar o indicador ***NDamage***. Caso a mutação tenha passado em pelo menos cinco deles, declara-se a mutação como sendo *driver*. Este estudo tomou como base esses critérios de classificação para determinar as proporções *driver*/*passenger*.

4.3.2 Análise das mutações genéticas através de preditores

Os parâmetros mencionados anteriormente constituem técnicas já existentes na literatura. Eles provêm do dbNSFP, que nada mais é do que um banco de dados que carrega consigo mecanismos de predição (LIU, 2013). Estes mecanismos observam em geral características como homogeneidade na sequência da amostra, informações do RNA e proteínas, substituição de nucleotídeos, etc. Muitos repetem técnicas uns dos outros, porém acredita-se que um critério de proporção de 5/9 efetivamente apresenta indícios suficientes para classificar uma mutação como *driver* e caso não, como *passenger*. A Tabela 4 descreve brevemente cada um destes preditores.

Tabela 4 – Preditores utilizados no DB Mutation e suas definições (LIU, 2003).

Mecanismo de predição	Fator de avaliação
SIFT	Possui alcance de 0 a 1. Analisa a probabilidade de efeito danoso de SNP, quanto menor a pontuação maior a probabilidade de dano.
PolyPhen-2-HDIV	Ferramenta automática para predição de possíveis impactos de substituição de aminoácidos na estrutura e função da proteína humana. Esta predição é baseada no número de características compreendidas na sequência, informações filogênicas e estruturais caracterizam a substituição.
PolyPhen-2-HVAR	Ferramenta automática para predição de possíveis impactos de substituição de aminoácidos na estrutura e função da proteína humana. Esta predição é baseada no número de características compreendidas na sequência, informações filogênicas e estruturais caracterizam a substituição.
Mutation Taster	Emprega um classificador de Bayes para eventualmente predizer o potencial de alteração de uma doença. O classificador de Bayes é alimentado com o resultado de todos os testes e características das alterações e calcula as probabilidades de alterações para uma alteração ser tanto uma doença por meio de mutação quanto um polimorfismo sem efeito.
Fathmm	Servidor capaz de predizer as consequências funcionais de ambas variantes de código e não-variantes de código. Retorna predições capazes de distinguir entre promoção do câncer/mutações driver e outros polimorfismos germinativos.
LTR	É um eficiente programa para encontrar LTR *retranspsons* completes em sequências gênicas. Possui alcance de 0 a 1. Prediz se determinada mutação é deletéria (D), neutra (N) ou indeterminada (U), porém essas definições não se baseiam unicamente no valor resultado encontrado.

MetaSVM	Máquina de vetor suporte (SVM) determina um resultado de predição que incorpora dez outros preditores (*SIFT, PolyPhen-2 HDIV, Polyphen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LTR, SiPhy, PhyloP*) e ainda a máxima frequência observada em 1000 populações gênicas. Quanto maior o valor obtido mais provável de ser danoso. Alcance de -2 a 3 no dbNSFP.
MetaLR	Mecanismo de predição baseado em regressão logística, que incorpora dez outros preditores (*SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP*) e ainda a máxima frequência observada em 1000 populações gênicas. Quanto maior o resultado mais provável d ser danoso. Alcance de 0 a 1.
Mutation Assessor	Servidor capaz de predizer o impacto da substituição de aminoácidos em proteínas, como mutações descobertas no câncer ou *missense* polimorfismos. O impacto funcional é acessado baseado na conservação evolucionária do aminoácido afetado em proteínas homólogas.

Em vista desses preditores é possível determinar com confiança quais das mutações disponíveis no banco de dados são *driver* e quais são *passenger*. E, portanto, comparar com efetividades os dados reais triados com os dados simulados adquiridos.

4.4 Análise Estatística – Modelo de regressão

Utilizamos o coeficiente de determinação, conhecido como R^2, pois é uma informação que determina se o modelo consegue explicar os valores observados. Trata-se de uma medida baseada em regressão linear e com isso consegue determinar se os valores do nosso *x*, número de *drivers*, são parâmetro condicional para estimar o valor do nosso *y*, número de *passengers*. Quanto maior o R^2, mais explicativo é o modelo, mais ele se ajusta à amostra. Seu valor é regido pela Equação 4, abaixo indicada.

Equação 4 – Valor do R^2.

$$\boldsymbol{R^2 = 1 - \frac{SQ_{res}}{SQ_{tot}}}$$

Nesta equação temos que SQ_{tot} é a soma total dos quadrados, ou seja, a soma dos quadrados das diferenças entre a média e cada valor observado, SQ_{res} é a soma dos quadrados dos resíduos, que calcula a parte não explicada do modelo, ou seja, valores que não possuem explicação no modelo. O valor do R^2 varia entre 0 e 1 e indica em porcentagem o quanto o modelo consegue explicar os valores observados.

5 RESULTADOS E DISCUSSÃO

Esta seção compreende os resultados obtidos no estudo. Além disso, também serão apresentados alguns pontos para discussão, que estão correlacionados com os resultados.

5.1 Modelo Estocástico e Simulações

Devido a importância das mutações no desenvolvimento tumoral, um dos objetivos foi analisar se os tempos entre mutações decrescem a partir do aumento do número delas. Com isso, empregamos o modelo em ramificação pois este representa melhor o processo biológico de aquisição de mutações, tornando as simulações numéricas factíveis. No modelo, a cada passo, dois dias, todas as células irão se (i) dividir sem adquirir nova mutação *driver*, (ii) dividir adquirindo uma nova mutação *driver*, ou (iii) morrer. A possibilidade da célula entrar em quiescência ou se diferenciar é modelada dentro do grupo 'morte'. Nós assumimos, também, que mutações *driver* reduzem a possibilidade de morte celular.

Foram utilizados os parâmetros listados na tabela 3, e tempo de dobramento adequado ao câncer de mama (2 dias). Ao todo foram realizadas cem mil simulações.

Dessas, foram desconsideradas as que obtiveram apenas a primeira mutação *driver* (98.951 simulações), ou seja 98,95% de todas as simulações. Muitas dessas simulações não passaram para o segundo passo da simulação, morrendo logo em seguida. Isso se deve ao fato de que inicialmente a probabilidade de morte celular é maior, uma vez que a vantagem adaptativa da célula com mutação não é alta. Ela, por sua vez, não conseguirá perdurar no organismo a não ser que alcance uma vantagem adaptativa relevante para sua sobrevivência. Além disso, como mesmo Bozic afirma, nem todas as mutações simuladas são bem-sucedidas porque flutuações estocásticas podem levar a extinção de uma linhagem.

Na figura 6 é apresentado o número de simulações obtidas por quantidade de mutações *drivers* (linhagens) ao total.

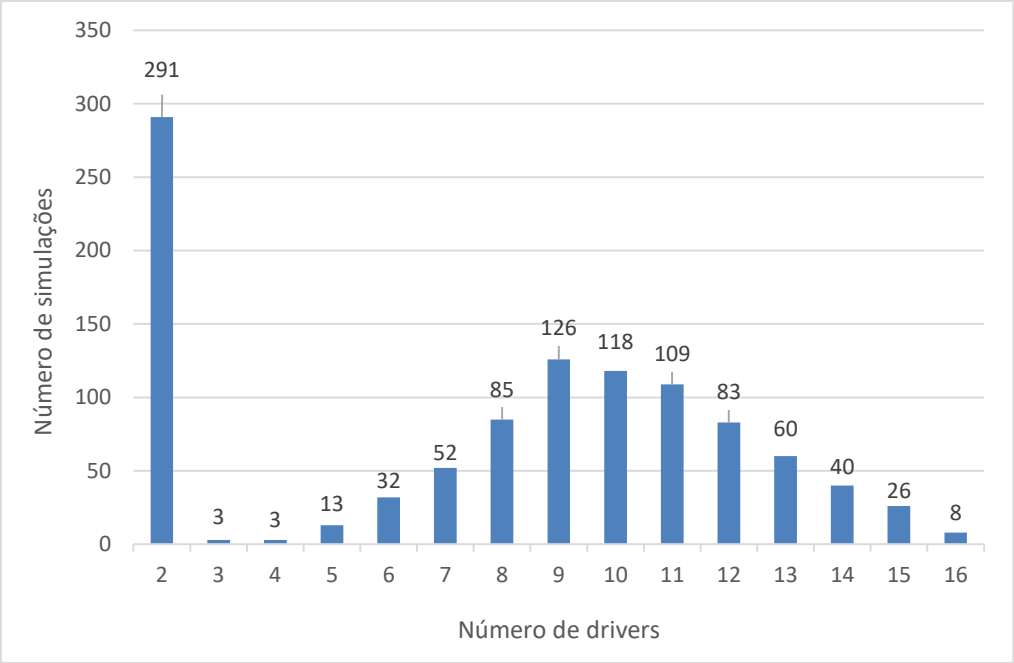

Figura 6 – Quantidade de simulações obtidas com o número máximo de *drivers*.

É interessante notar a distribuição do número destas simulações. As que obtiveram apenas duas mutações *drivers* foram muito mais frequentes do que todas as outras; mas as simulações com três e quatro *drivers* ficaram bem abaixo do esperado. Isto pode ser explicado pelo valor do parâmetro *s*, que representa a vantagem adaptativa. Provavelmente, após o tumor após adquirir certo número (três ou quatro) de mutações *driver*, já obteve vantagem adaptativa suficiente para perdurar no organismo por mais tempo, garantindo assim sua sobrevivência. Dessa forma, encontramos poucas simulações que terminaram com três ou quatro mutações, pois este número de mutações ainda não foi suficiente para contrabalançar a chance de extinção das linhagens.

Outro ponto importante se encontra no também alto número de mutações *driver* entre seis e quatorze mutações, novamente, isso se explica devido ao alto fator adaptativo que as células cancerígenas já obtiveram. Uma vez que após adquirir mutações *drivers* suficientes é possível encontrar um aumento na taxa de clonagem de células, acelerando, assim, sua expansão e ficando o organismo impossibilitado de inviabilizar o crescimento a partir de seus mecanismos próprios.

Escolhemos algumas imagens representativas, mostradas na Figura 7, do crescimento populacional a partir do número de mutações. Nesta figura apresentamos oito exemplos das simulações, onde cada cor representa uma linhagem de mutação *driver* e seu comportamento de crescimento populacional ao longo do tempo. Em 7a,

por exemplo, temos uma linhagem que iniciou e cresceu rapidamente (curva em preto), mas que após um certo número de iterações não conseguiu perdurar e se extinguiu pouco depois da morte da linhagem de duas mutações *driver* (pontos em vermelho). Nas imagens seguintes temos que após adquirir o número de mutações limitante, cerca de três ou quatro mutações, o tumor cresceu com velocidade ao longo do tempo.

Uma observação importante foi encontrada nas simulações representadas nas figuras 7d, 7e, 7f, 7g, 7h, onde após certo tempo de simulação as linhagens seguintes superaram o número de células da primeira linhagem. Na figura 7b a segunda mutação apenas ocorreu cerca de 10 anos após a inicial, mas para a terceira mutação ocorrer foi necessário apenas três anos. Na figura 8h temos que as novas mutações chegam a ser anuais, após 15 anos de simulação. Isso confirma o exposto na Figura 2, onde se diz que a cada nova mutação o tempo até a próxima diminui e consequentemente a taxa de divisão celular aumenta, ocorrendo, assim, a superposição do número de células das linhagens seguintes em relação a inicial.

Para a simulação com um total de 16 mutações *driver* (Figura 8h), o número de células total do tumor foi de $6{,}85\ x\ 10^{17}$ células, número que não apresenta respaldo biológico, uma vez que o número total de células no corpo gira em torno de 10^{12} e 10^{16} (BIANCONI, 2013). Este comportamento na simulação é devido ao tempo e ação das mutações, e poderia corresponder a um tumor altamente agressivo, mas na prática a agressividade da ação tumoral já teria provocado a morte do paciente.

Os resultados apresentados foram baseados em parâmetros de crescimento encontrados na literatura, especificamente os valores das taxas de aquisição de mutações *driver* e o tempo de dobramento específico do câncer de mama. Sem isso, não haveria respaldo biológico para compararmos a eficácia do modelo em comparação ao utilizado por Bozic com gliobastoma e câncer de pâncreas (BOZIC et al., 2010). Utilizando esta abordagem baseada em premissas biologicamente relevantes podemos evoluir de um modelo de matemática aplicada para um modelo matemático com perspectivas de aplicações biológicas.

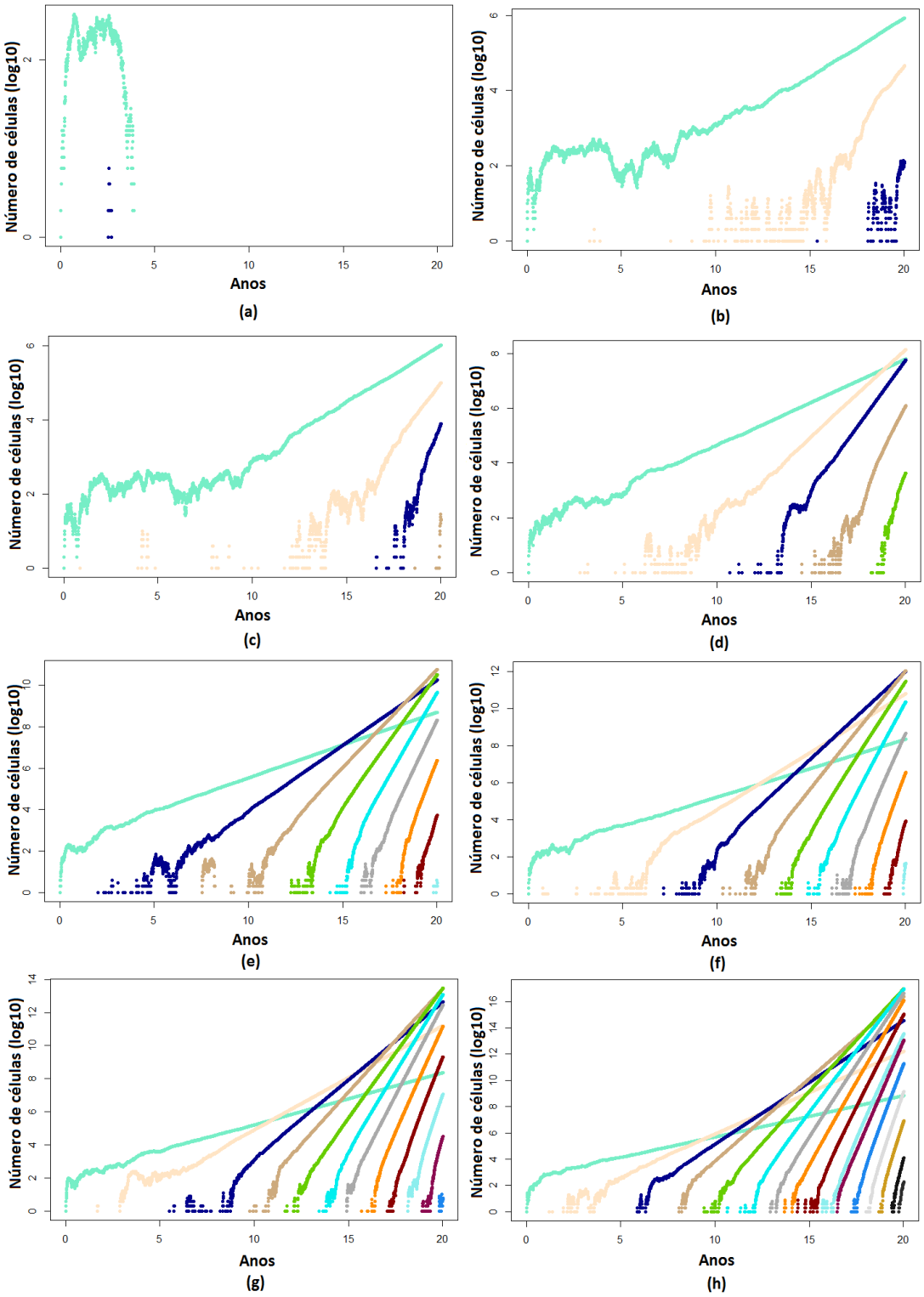

Figura 7 – Representação das simulações – número de células por tempo de simulação (idade do tumor). Cada gráfico representa uma realização do mesmo modelo estocástico, utilizando os mesmos parâmetros e cada cor representa uma linhagem com "n" mutações *drivers*. 7a, dois *drivers*; 7b, três *drivers*; 7c, quatro *drivers*; 7d, cinco *drivers*; 7e, nove *drivers*; 7f, dez *drivers*; 7g, doze *drivers*; 7h, dezesseis *drivers*.

Para determinar no número de *passengers* das simulações foi utilizada a Equação 2, onde *t* é calculado como o número de passos (iterações) do início da simulação até o início da última linhagem. Através do gráfico da Figura 8, é possível observar uma dependência do crescimento do número de *passengers* em relação ao número de *drivers*, e após quatro mutações *driver*, os números de *passengers* estabilizaram em 28 a 29 mutações.

É provável que o método utilizado não se aproxima o bastante do comportamento biológico, porém observa-se que há um crescimento inicial do número de mutações *passengers* em relação às *drivers*, e que após certo tempo o número de mutações entra em estagnação. Com isso, há indício para se acreditar que há realmente relação entre o número de *passengers* e sua dependência das mutações *drivers*. A explicação está no fato de que as mutações *passenger* nada mais são do que o produto da multiplicação defeituosa das células e, sendo assim, não desempenham nenhum papel na dinâmica do crescimento tumoral (mutações neutras). E este acúmulo provavelmente ocorre porque as mutações drivers afetam algumas vias metabólicas (capacidades adquiridas do câncer) que levam a uma maior proliferação celular, seja por uma aumento na frequência do ciclo celular, seja pela diminuição do tempo do ciclo.

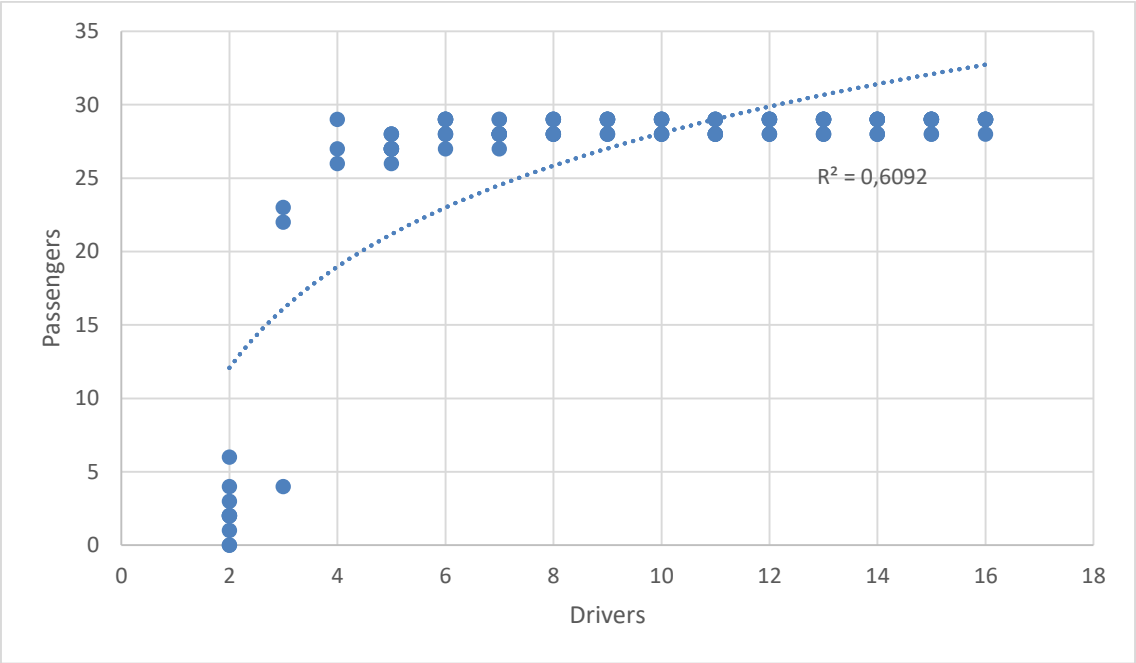

Figura 8 – Número de *passengers* obtidos em simulações com número máximo de *drivers*.

Comparando os nossos resultados com os obtidos por Bozic e colaboradores em uma questão importante a se comentar é de que, entre as premissas consideradas está a de que não houve muita divisão celular antes da iniciação do tumor, dessa forma podemos considerar o início de tudo a partir da primeira mutação *driver*. Informações mais completas acerca do câncer de mama são difíceis de encontrar, como por exemplo a existência ou não de acúmulos de mutações *passengers* antes da aquisição da primeira *driver*. Adicionalmente, o tempo de dobramento para células epiteliais mamárias varia bastante, de acordo com a literatura, entre 24 e 120h (DAI, 2015), que para o estudo realizado influencia e muito nos resultados.

No entanto, como Bozic afirma, é esperado que a média da vantagem adaptativa de uma mutação *driver* será similar, do ponto de vista de todos os tipos de tumores, uma vez que os caminhos pelos quais essas mutações agem se sobrepõem em um grau considerável. Esta linha de raciocínio pode estar correta observando a Figura 8 onde o comportamento logarítmico se mantém para o caso do câncer de mama. Onde o tempo acumulado para adquirir k mutações aumenta logaritmicamente com k.

5.2 Bioinformática e Análise de Dados Genômicos

Para avaliar a relevância do modelo, comparamos os resultados da simulação com informações obtidas da análise bioinformática (dados DB Mutation) em relação a proporção de mutações *passenger* e *driver*. As Figuras 9 e 10 apresentam as proporções de *drivers* e *passengers* para câncer de mama encontradas no DB Mutation, seguindo os critérios apresentados na metodologia. Devido a certas discrepâncias nos dados de cada banco, cerca de 80% das amostras provenientes do TCGA não passaram pelos filtros classificadores do DB Mutation, e por isso os dados obtidos de cada banco foram analisados separadamente. Além disso, desconsideramos amostras que possuíam muitas mutações (amostras com mais de 30 mutações) com o propósito de haver mais complacência na comparação com os dados gerados a partir das simulações.

Como podemos notar, as distribuições de *passenger*/*driver* apresentam um comportamento semelhante ao do gráfico de relação *passenger*/*driver* das simulações, onde temos em quantidade amostras com poucas mutações *driver* e, após um intervalo de poucas mutações, temos novamente uma concentração de amostras em números mais altos para mutações *driver*.

Baseando-se no coeficiente de determinação R^2 da curva, é possível observar que os dados provenientes da base de dados COSMIC possuem mais relevância que os da base TCGA. O R^2 varia entre 0 e 1, e indica, em porcentagem, o quanto o modelo consegue explicar os valores observados. Isto se remete ao problema que encontramos ao selecionar as amostras, onde houve discrepâncias entras as bases de dados, com o agravante que os dados do TCGA em sua maioria não foram filtrados pelos classificadores do DB Mutation. Portanto, os dados analisados podem não fornecer uma informação completa, uma vez que não foi possível pelo DB Mutation classificar com eficiência as informações disponíveis nas amostras. Superar essa deficiência tanto no DB Mutation quanto no TCGA nos ajudará futuramente para fins de comparação de dados simulados com reais.

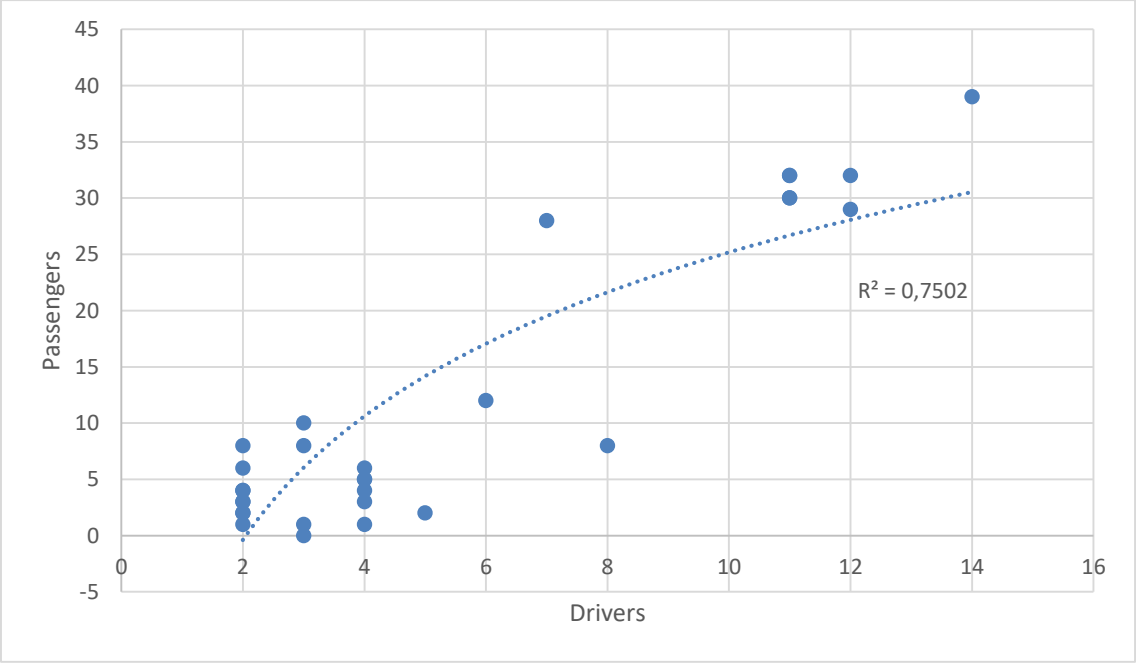

Figura 9 – Proporção de *drivers* e *passengers* para a base dados COSMIC.

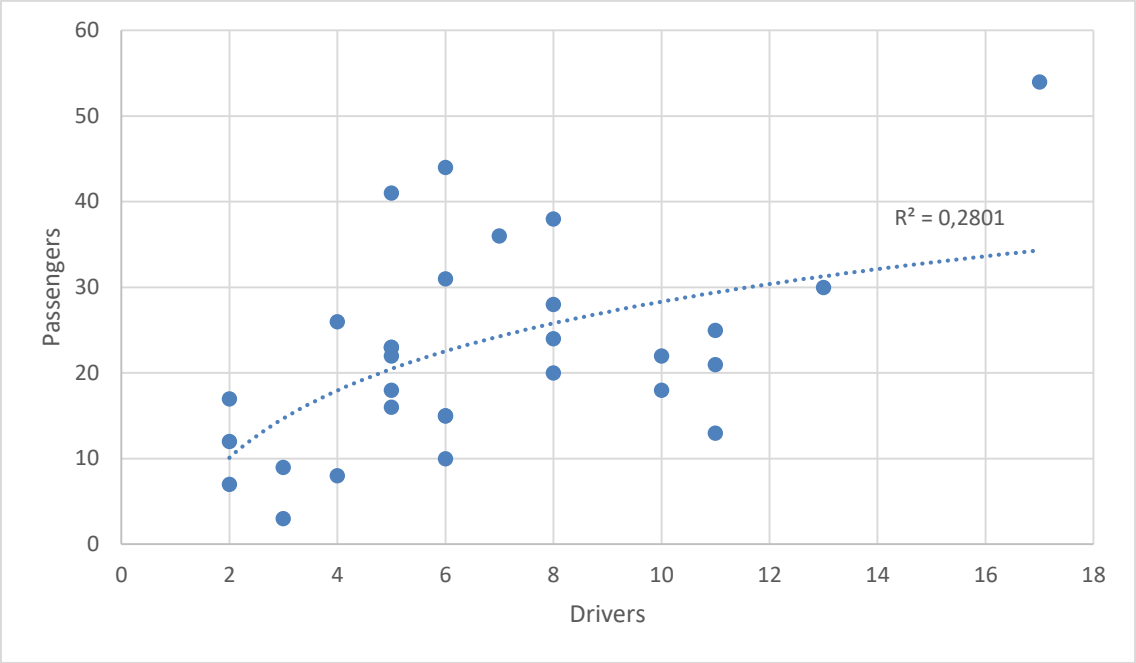

Figura 10 – Proporção de *drivers* e *passengers* para a base dados TCGA.

Outra comparação a se fazer com o trabalho realizado por Bozic e colaboradores está no fluxograma que ela utilizou para classificar mutações e comparar os dados reais (glioblastoma multiforme e pâncreas) com os dados simulados. Os autores utilizaram uma definição simplista para classificar mutações de uma amostra como *drivers*, considerando para tanto que caso a mutação fosse *missense* ela seria invariavelmente uma mutação *driver*. Esta consideração pode não ser muito aplicável do ponto de vista biológico, sendo necessário uma definição mais formal e que não deixe espaço para falsos negativos e positivos.

No entanto, as predições que ela fez e seus resultados estão de acordo também com os resultados obtidos durante esse estudo. Como em todos os modelos existentes, é necessário fazer premissas limitantes, como as que foram usadas neste modelo, porém essas premissas não deixam de fornecer as conclusões chave do estudo, que podem auxiliar um melhor entendimento dos processos biológicos.

Através disso, vê-se que o contexto real do modelo e seus resultados suportam informações o suficiente para acreditar que tumores e seu crescimento são dependentes das mutações *driver*. Em adição a isso temos que os números da relação de mutações *driver* com mutações *passenger* também é dependente e se apresenta com comportamento logarítmico.

6 CONCLUSÃO

Diante do que foi exposto podemos entender que o estudo alcançou seu objetivo de apresentar um modelo de dinâmica evolutiva e analisar, comparativamente, os resultados da simulação aos dados referentes ao câncer de mama.

Nossos resultados mostraram que após adquirir mutações *driver* o tumor acelera seu processo de aquisição de novas mutações. O tempo decorrido para a aquisição da segunda mutação foi de dez anos e para a terceira a custar dois. Além disso, a aquisição de mutações *passenger* é dependente do crescimento tumoral e consequentemente do número de mutações *driver* que este possui.

Observamos ainda a relação logarítmica entre os tipos de mutações, *driver* e *passenger*, onde após certo tempo de simulação o número de mutações *passenger* entra em um estado de platô, decorrente do número máximo de passos possíveis para o tempo de simulação definido. Em adição, comprovamos que o fator de vantagem adaptativa, *s*, realmente caracteriza um parâmetro seguro para futuras simulações.

Entendemos que para o futuro da pesquisa de dinâmica evolutiva de tumores devemos estabelecer parâmetros cada vez mais próximos da biologia, uma vez que muitos são fatores limitantes da dinâmica.

Futuramente poderemos testar tratamentos para o câncer a partir de simulações para então testar *in vivo*. Entretanto, entendemos que ainda há um longo caminho a ser trilhado no sentido de modelar a dinâmica de crescimento tumoral e os fatores que corroboram para seu desenvolvimento.

7 REFERÊNCIAS

1. ARMITAGE, Peter; DOLL, Richard. The age distribution of cancer and a multi-stage theory of carcinogenesis. **British journal of cancer**, v. 8, n. 1, p. 1, 1954.
2. BALMAIN, Allan et al. How many mutations are required for tumorigenesis? Implications from human cancer data. **Molecular carcinogenesis**, v. 7, n. 3, p. 139-146, 1993.
3. BIOME – NÚCLEO MULTIDISCIPLINAR DE BIOINFORMÁTICA. **DB Mutation**, 2016. Disponível em: http://www.bioinformatics-brazil.org/dbmutation/. Acesso em: 26 de nov. 2016.
4. BISHOP, J. Michael; WEINBERG, Robert Allan (Ed.). **Scientific American molecular oncology**. Scientific American, 1996.
5. BOZIC, Ivana et al. Accumulation of driver and passenger mutations during tumor progression. **Proceedings of the National Academy of Sciences**, v. 107, n. 43, p. 18545-18550, 2010.
6. BOZIC, Ivana; ALLEN, Benjamin; NOWAK, Martin A. Dynamics of targeted cancer therapy. **Trends in molecular medicine**, v. 18, n. 6, p. 311-316, 2012.
7. CARTER, Hannah et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. **Cancer research**, v. 69, n. 16, p. 6660-6667, 2009.
8. DAI, Xiaofeng et al. Breast cancer intrinsic subtype classification, clinical use and future trends. **American journal of cancer research**, v. 5, n. 10, p. 2929, 2015.
9. DURRETT, Rick et al. Evolutionary dynamics of tumor progression with random fitness values. **Theoretical population biology**, v. 78, n. 1, p. 54-66, 2010.
10. DURRETT, Rick et al. Intratumor heterogeneity in evolutionary models of tumor progression. **Genetics**, v. 188, n. 2, p. 461-477, 2011.
11. FOULDS, L. The experimental study of tumor progression: a review. **Cancer research**, v. 14, n. 5, p. 327-339, 1954.

12. HAHN, William C. et al. Creation of human tumour cells with defined genetic elements. **Nature**, v. 400, n. 6743, p. 464-468, 1999.

13. HANAHAN, Douglas; FOLKMAN, Judah. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. **cell**, v. 86, n. 3, p. 353-364, 1996.

14. HANAHAN, Douglas; WEINBERG, Robert A. The hallmarks of cancer. **cell**, v. 100, n. 1, p. 57-70, 2000.

15. HANAHAN, Douglas; WEINBERG, Robert A. Hallmarks of cancer: the next generation. **cell**, v. 144, n. 5, p. 646-674, 2011.

16. INSTITUTO NACIONAL DO CÂNCER JOSÉ ALENCAR GOMES DA SILVA**. Estimativa 2016 – Incidência do Câncer no Brasil**, 2016. Disponível em: < http://www.inca.gov.br/estimativa/2016/index.asp?ID=2 />. Acesso em: 20 de nov. 2016.

17. LIU, Xiaoming; JIAN, Xueqiu; BOERWINKLE, Eric. dbNSFP v2. 0: a database of human non-synonymous SNVs and their functional predictions and annotations. **Human mutation**, v. 34, n. 9, p. E2393-E2402, 2013.

18. NOWAK, Martin A. **Evolutionary dynamics**. Harvard University Press, 2006.

19. R CORE TEAM. R: A language and enviroment for statistical computing. **R Foundation for Statistical Computing**, Viena, Austria. Disponível em: <http://www.R-project.org/>. Acesso em: 26 de nov. 2016.

20. REGOES, Roland R. Population genetics meets cancer genomics. **Proceedings of the National Academy of Sciences**, v. 107, n. 43, p. 18241-18242, 2010.

21. STEWART, B. et al. World cancer report 2014. **World**, 2016.

22. STRATTON, Michael R.; CAMPBELL, Peter J.; FUTREAL, P. Andrew. The cancer genome. **Nature**, v. 458, n. 7239, p. 719-724, 2009.

23. THE SCIENTIST. **Combing the Cancer Genome**, 2012. Disponível em: http://www.the-scientist.com/?articles.view/articleNo/31747/title/Combing-the-Cancer-Genome/. Acesso em: 05 de dez. 2016.

24. WACLAW, Bartlomiej et al. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. **Nature**, v. 525, n. 7568, p. 261-264, 2015.