

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

ONCOPROSIM: UMA FERRAMENTA *IN SILICO* PARA ANÁLISE DA EVOLUÇÃO TUMORAL

Dissertação de mestrado  
apresentada ao programa  
de Pós-Graduação em  
Biologia Celular e  
Molecular da Universidade  
Federal do Rio Grande do  
Sul como requisito parcial  
para obtenção do grau de  
Mestre

DARLAN CONTERNO MINUSSI  
ORIENTADOR: Dr GUIDO LENZ  
CO-ORIENTADOR: Dr MANUEL MENEZES DE OLIVEIRA NETO

PORTO ALEGRE, AGOSTO DE 2015

*“Study hard what interests you the most in the most undisciplined, irreverent and original manner possible”*

**Richard P. Feynman**

## **AGRADECIMENTOS**

Aos meus pais, pelo amor e apoio incondicional durante mais essa etapa. A coragem e determinação que sempre demonstram é meu maior exemplo. Obrigado pela educação e auxílio em todos os momentos, esta conquista pertence a vocês.

A Lia Mara Gomes Paim, é ela que compartilha o dia-a-dia comigo, que compartilha a paixão pela ciência, que compartilha os dias difíceis e os momentos felizes. E por ser o motivo de meu sorriso ao início de cada dia.

A toda a minha família, pelas risadas, churrascos e apoio.

Aos amigos de Boston, Florianópolis, Marau, Passo Fundo, Porto Alegre, Vienna. Tenho muita sorte em possuir tantos amigos queridos em tantos lugares. Obrigado por todos os momentos compartilhados. Vocês também são partes fundamentais dessa conquista

Ao governo do Brasil e as agências de fomento, que forneceram os recursos estruturais e financeiros para que este trabalho fosse realizado. Ao governo dos Estados Unidos, por disponibilizar de maneira gratuita bancos de dados e o PubMed, fundamentais para esse trabalho.

Aos amigos e professores da UFCSPA e da UFRGS pelos ensinamentos.

Ao pessoal do instituto de informática da UFRGS, especialmente ao Bernardo Henz pelo trabalho em conjunto e sem o qual este trabalho não teria sido possível e, ao meu co-orientador, Manuel Menezes de Oliveira Neto pelos ensinamentos e disponibilidade em aceitar esse desafio em conjunto conosco.

Ao meu orientador Guido Lenz, por todos os ensinamentos e paciência, por acreditar em minha capacidade e corrigir minhas (não tão infrequentes) crises de autocrítica. Sua paixão pela ciência e seu conhecimento são os dois aspectos que almejo reproduzir durante minha carreira científica. Também a sua esposa Márcia Rosângela Wink que muito me ajudou durante a faculdade e também sempre me apoiou.

A todos os amigos do laboratório de sinalização celular, incluindo aqueles que não mais trabalham ali, mas os quais eu já tive o prazer de conviver. Obrigado por tornarem esse um ambiente de trabalho incrível e do qual me orgulho em fazer parte. E, especialmente, por aguentar todo dia o meu mau humor matinal.

A Lia Mara Gomes Paim, Mardja Manssur Bueno e Silva e Pitia Flores Ledur pela leitura crítica do manuscrito e sugestões.

Aos grandes nomes da ciência sobre os ombros os quais me apoio e a todas as pessoas que apoiam o progresso da ciência.

## SUMÁRIO

1. LISTA DE FIGURAS.....	5
2. LISTA DE TABELAS .....	5
3. LISTA DE ABREVIATURAS.....	5
4. RESUMO.....	6
5. ABSTRACT .....	7
6. INTRODUÇÃO .....	8
7. EVOLUÇÃO TUMORAL .....	8
7.1 Iniciação tumoral. ....	9
7.2 Número de mutações por amostra e classificação da importância das mutações .....	10
7.3 Célula de origem .....	15
7.4 Dinâmica evolutiva de tumores .....	17
8. MODELOS DE EVOLUÇÃO TUMORAL .....	23
9. OBJETIVO GERAL.....	29
10. OBJETIVOS ESPECÍFICOS.....	29
11. ARTIGO CIENTÍFICO.....	30
12. PERSPECTIVAS .....	78
13. CONCLUSÕES.....	79
14. REFERÊNCIAS .....	80

## 1. LISTA DE FIGURAS

**Figura 1:** Frequência de mutações somáticas em diferentes tipos tumorais

**Figura 2:** Relação de possíveis células de origem em diferentes tumores

**Figura 3:** Progressão do câncer do cólon e reto

**Figura 4:** Ilustração de inter e intra-heterogeneidade tumoral

**Figura 5:** Dinâmicas de evolução tumorais

**Figura 6:** Tempo de sucessão entre linhagens clonais

## 2. LISTA DE TABELAS

**Tabela 1:** Número de mutações somáticas em diferentes tipos tumorais

## 3. LISTA DE ABREVIATURAS

COSMIC	Catalogue of Somatic Mutations in Cancer
TCGA	The Cancer Genome Atlas
TAF	Tumor Associated Factor
TSG	Tumor supressor gene
TTP	Tool for Tumor Progression

#### 4. RESUMO

O conhecimento da evolução tumoral ocorre, até então, através da informação que podemos extrair de biópsias. Essas amostras são de suma importância para o diagnóstico e tratamento clínico, entretanto, por se tratarem de “instantâneos” são limitadas a uma única etapa do desenvolvimento tumoral e não revelam toda a história da evolução tumoral. Nesse trabalho, propomos simular a evolução clonal de diferentes tipos tumorais, com foco nos mecanismos moleculares que levam à progressão tumoral. Para isso, geramos células com duas matrizes que representam a característica diploide do genoma humano e, a cada divisão celular, incluímos mutações que podem alterar as probabilidades de proliferação e morte celular basal. Através desse modelo e, com o auxílio de um gerador de números pseudoaleatórios, podemos simular diferentes padrões de evolução tumoral para aferir a incidência em razão dos diferentes espectros mutacionais presentes nas distintas evoluções clonais. Além disso, podemos reproduzir características essenciais do genoma de tumores como a interação de oncogenes com genes supressores tumorais, frequência de uma mutação de acordo com diferentes valores de *fitness* e, até mesmo, simulações de tratamento e reincidência tumoral. Assim, com o desenvolvimento do conhecimento acerca da biologia tumoral, acreditamos que nosso modelo poderá oferecer uma visão única da evolução de tumores aliados a velocidade que somente modelos *in silico* são capazes de oferecer.

## 5. ABSTRACT

The current knowledge regarding tumor evolution, until now, has been obtained from what we can gather through biopsy samples. In spite of the importance of these samples to diagnosis and treatment, they represent only a glimpse of the whole tumor evolutionary path, whereas the majority of the tumor development remains hidden. In this work, we attempt to simulate the clonal evolution of different types of tumors, focusing on the molecular mechanisms that lead to tumor progression. In order to do that, we generate cells with two arrays that represent the diploid characteristic of the human genome and, in case of cell division, mutations can be inserted in the genome that may alter the default probabilities of proliferation and death. With the help of a pseudorandom number generator, we can use our model to simulate different patterns of tumor evolution and investigate the effects of different mutational spectrum in the incidence of distinct tumors. Moreover, we can use our model to reproduce essential characteristics of the tumor genome such as: the synergy between oncogenes and tumor suppressor genes, changes in mutation frequency with distinct fitness values and even simulate tumor treatment. Therefore, with the increasing knowledge in tumor biology, we believe that our model can offer a unique perspective of tumor evolution, allied to the speed and reproducibility that only *in silico* models are capable of offering.

## 6. INTRODUÇÃO

Câncer é uma doença perversa, que, anualmente, transforma a vida de milhões de pessoas em uma batalha. Para o ano de 2015, no Brasil, estão previstos 576 mil novos casos da doença (Inca, 2014) e nos Estados Unidos, somente para o ano de 2014, foram registrados mais de 1,5 milhão de novos casos com 580 mil mortes previstas em decorrência da doença (Siegel *et al.*, 2014).

Trata-se de uma doença de difícil detecção em estágio inicial e de difícil tratamento em estágios mais avançados, sendo que grande parte dos fármacos utilizados para tratar a doença tem efeitos colaterais impactantes na vida dos pacientes. Ainda, mesmo quando tratada com sucesso em uma primeira instância, em grande parte dos casos existe recorrência, etapa em que o tumor retorna e deixa de responder aos fármacos utilizados no tratamento.

Até então percebida e tratada erroneamente como uma doença homogênea, a atenção destinada ao estudo da biologia básica do câncer passa por uma revolução. O centro das atenções voltou-se para a tentativa de compreender passo-a-passo o desenvolvimento da doença. Descobertas primordiais foram feitas mostrando que tumores de diferentes tecidos são altamente heterogêneos, bem como tumores do mesmo tecido também são altamente heterogêneos (Navin *et al.*, 2011; Hodis *et al.*, 2012; Nik-Zainal *et al.*, 2012; Brennan *et al.*, 2013; Patel *et al.*, 2014) e, não obstante, recentes evidências apontam para a enorme heterogeneidade intra-tumoral (Patel *et al.*, 2014). Somente a compreensão dos mecanismos responsáveis por tamanha diferença permitirá que se possa vencer o último álamo desta doença: a resistência tumoral.

## 7. EVOLUÇÃO TUMORAL

Apesar de ser um termo conhecido como sinônimo de avanço e responsável pelo desenvolvimento de estruturas como organismos multicelulares, quando se trata de câncer, evolução passa a ser uma das mais temidas palavras, sinônimo de progressão da doença (Bozic e Nowak, 2013). Por isso, para a compreensão da biologia básica do



câncer é necessário o estudo de sua progressão desde a célula de origem até os mecanismos de resistência.

### **7.1 Iniciação tumoral.**

Câncer é uma doença genética em que uma célula saudável sofre erros (epi)genéticos durante seu processo de replicação que podem resultar em ganhos e/ou perdas cromossômicas; mutações pontuais ou alterações epigenéticas que resultam em vantagens competitivas para a célula quando comparadas as outras células existentes em seu microambiente.

Mas qual seria a quantidade de erros que uma célula suporta até que a mesma adquira uma vantagem proliferativa em relação às demais células de seu microambiente? As primeiras tentativas de responder essa pergunta surgiram quando Armitage e Doll em 1954, baseados em dados epidemiológicos de mortalidade para 17 tipos de tumores, propuseram o aparecimento do câncer em humanos como resultado de sucessivas mudanças celulares durante o decorrer do tempo, teoria que ficou conhecida como a teoria dos multi-estágios (Armitage e Doll, 1954). Três anos depois, a teoria multi-estágios acabou contestada pelos próprios autores, que apresentaram a hipótese de que apenas dois *hits* seriam o suficiente para a iniciação tumoral (Armitage e Doll, 1957). A sucessão entre dois *hits* ou múltiplos estágios perdurou durante algum tempo e, nesse ínterim, outros trabalhos analisaram dados epidemiológicos de câncer gástrico que corroboravam a hipótese de múltiplos *hits* (Ashley, 1969).

Em 1971, Knudson apresentou um estudo que se tornou um marco no desenvolvimento da biologia tumoral. Ele analisou 48 casos de retinoblastoma, um tipo de câncer hereditário, e demonstrou que somente duas mutações eram suficientes para a iniciação do tumor, sendo a primeira mutação na linhagem germinativa e a segunda ao longo da vida do paciente. (Knudson, 1971).

Entretanto, com o avanço dos estudos, novos mecanismos de iniciação tumoral foram descobertos, complicando o panorama em relação ao número de *hits* necessários para que uma célula desenvolva um potencial para se tornar cancerígena. Alguns tumores podem surgir de um único evento, como o caso da leucemia mielóide crônica,

em que uma única translocação do cromossomo 9 com o 22 resulta no desenvolvimento da doença (Rowley, 1973). Além disso, outros mecanismos mais recentes foram descobertos, como *chromotripsis*, em que um único cromossomo sofre uma série de eventos catastróficos de quebra que podem ser tidos como um mecanismo de iniciação tumoral (Cai *et al.*, 2014).

Seria possível então adotar um número preciso de eventos necessários para a evolução tumoral? Sem conhecer a taxa de divisão mitótica das células que irão originar o tumor, as taxas exatas de mutação por mitose nesse tipo celular e os processos seletivos que ocorrem durante as etapas da evolução para um estado tumoral, parece ser extremamente difícil determinar uma quantidade precisa de eventos necessários para a origem tumoral (Knudson, 2001)

## **7.2 Número de mutações por amostra e classificação da importância das mutações**

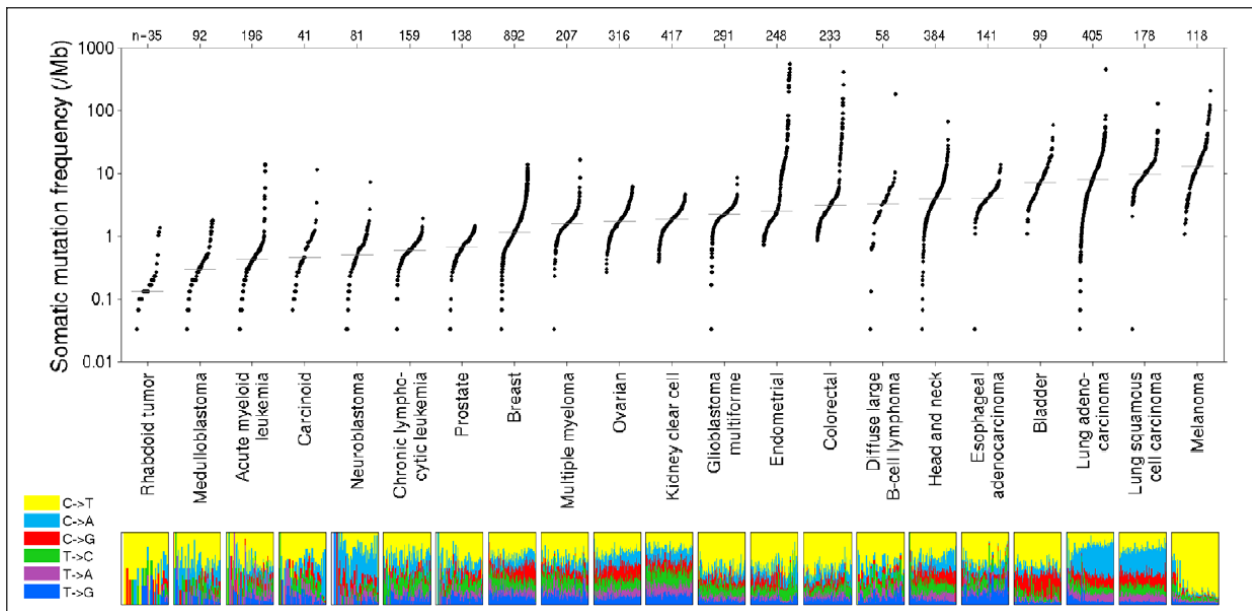
O desenvolvimento do sequenciamento de nova geração tem se mostrado uma ferramenta vital para responder a questão acima. A análise genômica dos diferentes tumores demonstrou uma grande discrepância entre os números de eventos previstos e a quantidade de mutações existentes nas amostras (Tabela 1).

**Tabela 1.** Número de mutações por tipo tumoral obtido por dados de sequenciamento de nova geração

<b>Tumor</b>	<b>n</b>	<b>Número de mutações somáticas anotadas</b>	<b>Média de mutações por tumor</b>	<b>Referência</b>
Ovário	316	19.356	61	(Cancer Genome Atlas Research, 2011)
Cólon e reto	188 (poucas mutações)	90.059	58	(Cancer Genome Atlas, 2012)
	36 (hipermutados)		728	
Pulmão	178	48.690	273	(Cancer Genome Atlas Research, 2012)
Melanoma	121	86.813	714	(Hodis <i>et al.</i> , 2012)
Glioblastoma	291	21.540	74	(Brennan <i>et al.</i> , 2013)

Em posse desses dados, Lawrence *et al.* em 2014 passaram a realizar interessantes trabalhos comparativos entre as diferenças amostras tumorais. Um destes

trabalhos demonstrou o número de mutações somáticas por megabase para 21 diferentes tipos tumorais na figura reproduzida abaixo (Figura 1) (Lawrence *et al.*, 2014).



**Figura 1:** Frequência de mutações somáticas por megabase entre 21 diferentes tipos tumorais e suas frequências de transições e transversões. Retirado de (Lawrence *et al.*, 2014).

De posse da informação a respeito da quantidade de mutações que de fato existem em uma amostra populacional de um tumor, passou a ser necessário definir se todas essas mutações exerciam ou não alguma função na evolução e progressão tumoral. Para isso, foi feita a diferenciação entre mutações “*drivers*” e “*passengers*”.

Mutações *drivers* tem implicação direta na causa dos tumores. Elas foram selecionadas em alguma etapa do crescimento tumoral, embora não necessariamente estarão presentes na etapa final, uma vez que podem servir como plataforma de modificação do microambiente tumoral (i. e. mutações que propiciam angiogênese), para outra mutação que estará melhor adaptada (Spencer *et al.*, 2006). Primariamente, então, elas conferem vantagens seletivas para as células que as sofreram no microambiente do qual pertencem.

Já mutações *passengers* são encontradas em células que sofrem uma mutação *driver* e que carregarão as mutações *passengers* durante a expansão clonal (Stratton *et al.*, 2009). Entretanto, alguns autores defendem que mutações *passengers* podem sim exercer influência durante a progressão tumoral (Mcfarland *et al.*, 2013), porém isso não será aqui explorado.

Logo, para determinar de maneira empírica o número necessário de mutações para o desenvolvimento de um tumor em um dado tecido, é fundamental a criação de um catálogo das mutações existentes em todos os tipos tumorais e, principalmente, definir quais dessas alterações são de fato responsáveis pela origem e progressão tumoral. Assim, uma quantidade significativa de recursos tem sido direcionada para o sequenciamento de tumores com projetos como o The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research, 2008) e o Catálogo de Mutações Somáticas (COSMIC).

O COSMIC trata-se de uma base de dados disponível na Web contendo mais de 2 milhões de mutações pontuais catalogadas manualmente e/ou de maneira semi-automática em mais de 1 milhão de amostras tumorais (Forbes *et al.*, 2015). Outro projeto que merece destaque é o International Cancer Genome Consortium que atua em larga escala em todo o mundo e pretende catalogar as mutações de mais de 50 tipos de tumores diferentes, além de disponibilizar os dados de maneira online (International Cancer Genome *et al.*, 2010).

A construção desses catálogos permitiu que grande parte dos trabalhos focassem seus esforços na identificação de mutações *drivers*, os quais serão brevemente descritos. A primeira abordagem utilizada é uma simples medição da frequência com que um gene aparece mutado nas diferentes amostras em relação as mutações existentes no *background*. Se essas mutações aparecem em alta frequência nos diferentes tumores, é razoável supor que elas são mutações essenciais para o desenvolvimento tumoral. Essa análise esbarra na dificuldade de identificar qual a verdadeira taxa de mutação do *background* devido a grande diferença de quantidade de mutações dos diferentes tumores e também devido a dificuldade em detectar mutações pouco frequentes (Pon e Marra, 2015).

Uma segunda abordagem consiste em avaliar o impacto causado pela mutação na função da proteína afetada. Dessa maneira, parte-se do pressuposto de que quanto maior o impacto causado à proteína, maior será a importância dessa mutação para os tumores. Algoritmos como CanPredict verificam se a mudança do aminoácido causado pela mutação tende a ser importante para o câncer ou não e identifica, assim, um gene candidato a *driver* (Kaminker *et al.*, 2007). Outras ferramentas como o MutationAssessor criam uma pontuação para o impacto funcional que a mutação representará na proteína baseado na conservação evolutiva da posição do nucleotídeo alterado (Reva *et al.*, 2011). Apesar de possuir vantagens como a possibilidade de realizar predições a partir de uma única amostra, a avaliação através da função também esbarra na possibilidade de que nem toda a mutação em região conservada pode ser fundamental para o desenvolvimento tumoral. Um estudo mais aprofundado sobre as vantagens e desvantagens para ambos os métodos pode ser encontrado em (Pon e Marra, 2015).

Não surpreendentemente, a integração das estratégias apresentadas parece ser a melhor maneira de trabalhar para a identificação dos genes verdadeiramente importantes para a progressão tumoral. Ferramentas de bioinformática têm sido os melhores aliados nessa tarefa. Um exemplo é o MutSig, uma ferramenta que combina a frequência com que as mutações aparecem em relação ao *background*, o impacto funcional delas em relação a conservação evolucionária da posição do nucleotídeo e o aparecimento dessa mutação no mesmo nucleotídeo (*hotspots*) em diversas amostras (Lawrence *et al.*, 2013). Essa ferramenta se mostrou capaz de identificar novos genes *drivers* em 21 diferentes tipos tumorais e evitar que genes sem importância para a evolução tumoral fossem detectados erroneamente (Lawrence *et al.*, 2014). Ainda assim, nesse mesmo trabalho os autores destacam a dificuldade de detecção de genes *drivers* que aparecem mutados em menos de 5% das amostras tumorais, e propõe como solução para esse problema o aumento do número de amostras analisadas. Assim, conseqüentemente, aumenta o poder estatístico desses achados, especialmente em casos de tumores hipermutados como melanomas; porém tal alternativa é restringida pelos altos custos envolvidos nesse processo.

### 7.3 Célula de origem

Conhecidos os mecanismos genéticos pelo qual o câncer tem início, passou a ser necessário descobrir a célula precursora, ou seja, a célula que adquiriu as mutações anteriormente apresentadas e que propiciaram as condições para que a célula obtenha as vantagens seletivas e escape do controle da homeostase em que se encontra.

Muito se conhece sobre os momentos finais dos tumores, mas traçar a sua origem é um tópico de extrema dificuldade, principalmente visto que quase todo o conhecimento que possuímos em relação a tumores humanos provém de biópsias, normalmente obtidas quando os tumores já são capazes de apresentar sinais clínicos nos pacientes e que, mesmo quando em tumores detectados precocemente, estão condicionados aos limites de detecção dos métodos de diagnóstico por imagem em que os tumores já possuem em torno de 6 mm de diâmetro (PET-CT) (Kennedy *et al.*, 2006) e mais de 100 milhões de células (3 milhões de células em 1 mm<sup>3</sup> (Spinney, 2006)). Portanto, é somente através de pesquisa *in vitro*, manipulações gênicas *in vivo* e regressões inferidas pelos dados obtidos de biópsias que podemos criar hipóteses sobre as células de origem.

Grande parte dos dados obtidos até então apontam para células com características de células-tronco como células de origem de tumores (Figura 2) (Barker *et al.*, 2009; Visvader, 2011; Som *et al.*, 2013), fato interessante visto que são células que inerentemente possuem capacidade proliferativa e de auto renovação. Além disso, recentemente foi demonstrado que o risco de desenvolver câncer em um determinado tecido está altamente correlacionado com a quantidade de divisões que as células-tronco desse tecido sofrem ao longo da vida de um indivíduo (Tomasetti e Vogelstein, 2015).

Em câncer de intestino, um modelo animal utilizando *Cre recombinase* aponta que células da cripta que expressam o receptor Lgr5, característico de células-tronco, e que tem o gene APC deletado (um importante supressor tumoral) iniciam um processo de proliferação que resulta em um microadenoma. Além disso, dentro de um intervalo de tempo de 3 a 5 semanas, um adenoma macroscópico passa a ser passível de detecção (Barker *et al.*, 2009). Para glioblastoma, um elegante modelo que utiliza o

método de análise de mosaico com duplo marcador resultando em dupla inativação de P53/Nf1 através de *Cre recombinase* em camundongos demonstrou que crescimento maligno aconteceu predominantemente em célula precursora de oligodendrócitos ou quando a mutação tumorigênica ocorreu em uma outra célula precursora que dará origem a uma célula precursora de oligodendrócitos, mas não houve formações de tumores quando a mutação se deu em células precursoras neurais (Liu *et al.*, 2011).

Outro tipo tumoral do qual já existem evidências para a célula de origem é o de próstata. O tecido normal divide-se em dois tipos celulares principais, células basais e luminais, ambas com capacidade de sustentar a sua própria linhagem. Em camundongos, demonstrou-se que os tumores de próstata são originados principalmente de células provenientes da linhagem luminal, através de marcações das duas linhagens e utilização de *Cre recombinase* condicionada a receptor de estrogênio. (Wang *et al.*, 2014).

Competentes revisões acerca das células de origem de tumores podem ser encontradas em (Visvader, 2011) e (Blanpain, 2013). A importância do conhecimento sobre a célula de origem para a simulação de tumores reside especialmente no fato de que é essa informação que será usada para determinar a população celular que pode dar início ao câncer.



Tumour type	Genetic model	Promoter-Cre construct	Lineage tracing	Cell of origin
<b>Mouse models</b>				
Brain: Glioblastoma	RAS, AKT activation (RCAS-TVA system: nestin, <i>Gfap</i> promoters)	NA	-	Neural progenitor cell <sup>40</sup>
	p16 <sup>INK4</sup> /p19 <sup>ARF</sup> , BMI1 inactivation; mutant EGFR	NA	-	Neural progenitor and astrocyte <sup>46,47</sup>
	p53, NF1 and/or PTEN inactivation	Nestin-CreERT2, Adeno-Cre	-	Multipotent progenitor <sup>41</sup>
	PDGFB activation (RCAS-TVA system)	NA	-	Oligodendrocyte progenitor <sup>45</sup>
	RAS, AKT activation; p53 inactivation	GFAP-Cre	-	Multipotent progenitor <sup>42</sup>
	Mutant p53 expression	GFAP-Cre	-	Neural progenitor or transit-amplifying cell <sup>45</sup>
	PTEN, p53 inactivation	GFAP-Cre	-	Multipotent progenitor <sup>44</sup>
Medulloblastoma	Patched inactivation	MATH1-Cre, GFAP-Cre	-	Multipotent progenitor and granule neuron progenitor <sup>49</sup>
	Smoothed activation	GFAP-Cre, MATH1-Cre, OLIG2-TVA-Cre	-	Multipotent progenitor and granule neuron progenitor <sup>48</sup>
	RB, p53, PTEN inactivation	GFAP-Cre	-	Multipotent progenitor <sup>44</sup>
	RB, p53 inactivation	Adeno-Cre	-	Neural progenitor cell <sup>50</sup>
	$\beta$ -catenin mutant, p53 inactivation	BLBP-Cre, ATOH1-Cre	-	Dorsal brainstem progenitor <sup>51</sup>
Ependymoma (supratentorial)	p16 <sup>INK4</sup> /p19 <sup>ARF</sup> inactivation; EPHB2 activation	NA	-	Embryonic cerebral stem/progenitor cell <sup>14</sup>
Intestine	APC inactivation	AhCre, LGR5-CreERT2	+	Stem cell <sup>26</sup>
	Mutant $\beta$ -catenin	CD133-CreERT2	+	Stem cell <sup>27</sup>
	Mutant $\beta$ -catenin	BMI1-CreER	+	Stem cell <sup>28</sup>
Lung	Kras activation	Adeno-Cre	-	Bronchioalveolar stem cell <sup>77</sup>
Mammary	NOTCH1 activation in cell subsets	NA	-	Luminal progenitor <sup>65</sup>
	BRCA1, p53 inactivation	BLG-Cre, K14-Cre	-	Luminal progenitor <sup>63</sup>
Pancreas	Kras activation, Inflammation	RIP-CreER	+	Endocrine cell <sup>69</sup>
Prostate	PTEN inactivation	NKX3.1-CreERT2	+	Luminal stem cell <sup>54</sup>
	ERG1, PI(3)K and/or AR expression	NA	-	Basal progenitor <sup>59</sup>
	PTEN inactivation	PB-Cre	-	Basal progenitor <sup>58</sup>
	PTEN inactivation	PSA-Cre	-	Luminal cell <sup>56,57</sup>
Skin/basal cell carcinoma	Smoothed activation	K14-CreER	+	Interfollicular epidermal progenitor <sup>72</sup>
Stomach	APC inactivation	LGR5-CreERT2	+	Stem cell <sup>59</sup>
<b>Human tissue</b>				
Breast (basal-like subtype) <sup>8</sup>	Preneoplastic BRCA1 <sup>-/-</sup> cell subsets	NA	-	Luminal progenitor <sup>57</sup>
Prostate	PI(3)K, ERG, AR into cell subsets	NA	-	Basal progenitor <sup>60</sup>

**Figura 2:** Possíveis células de origem e mecanismos de detecção para diferentes tipos de tumores. Retirado de (Visvader, 2011)

## 7.4 Dinâmica evolutiva de tumores

Uma vez estabelecidas as premissas da iniciação tumoral, deve-se passar aos conceitos da dinâmica evolutiva do câncer. Quais características levam ao contínuo desenvolvimento do tumor? Por que acontecem seleções clonais? Qual a proporção da intra-heterogeneidade tumoral existente nos tumores?

O acúmulo de mutações genéticas, sejam elas em pequena escala como as mutações somáticas ou grandes eventos como inserções, deleções, translocações ou

*chromotripsis* são as principais ferramentas para que a célula escape dos seus mecanismos de controle.

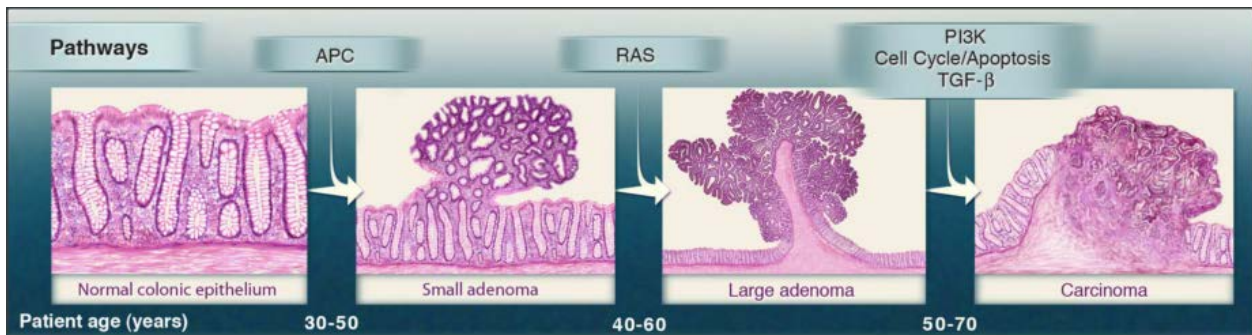
Em 2000, Hanahan e Weinberg publicaram um divisor de águas na literatura tumoral quando estabeleceram as 6 características fundamentais que praticamente todos os tipos tumorais compartilham (Hanahan e Weinberg, 2000). Este trabalho foi posteriormente atualizado em 2011 com a inclusão de mais duas características tumorais e duas características não essenciais mas sim ferramentas utilizadas pelos tumores para atingir de maneira mais eficiente essas condições (Hanahan e Weinberg, 2011). Essas ferramentas são: evasão da apoptose; autossuficiência de sinais proliferativos; insensibilidade a sinais anti-proliferativos; capacidade de sustentar angiogênese; potencial replicativo ilimitado; invasão de tecidos e metástases; desregulação do metabolismo celular; e evasão do sistema imune.

Todas essas condições são adquiridas através das diferentes alterações sofridas pelo (epi)genoma celular, sendo que algumas ocorrem em etapas mais iniciais do desenvolvimento tumoral, como a evasão da apoptose e potencial replicativo ilimitado, enquanto outras ocorrem mais tardiamente durante a evolução, como a invasão de tecidos e metástases.

Estudar como ocorre a evolução e a ordem dos eventos é um enorme desafio, especialmente quando se busca dados humanos. Enquanto para organismos unicelulares como leveduras é possível estudar quais mutações ocorrem em um determinado tempo através da análise de sucessivas gerações (Lang *et al.*, 2013), os tumores humanos são descobertos, tratados e analisados nas mais diversas etapas de estadiamento tumoral.

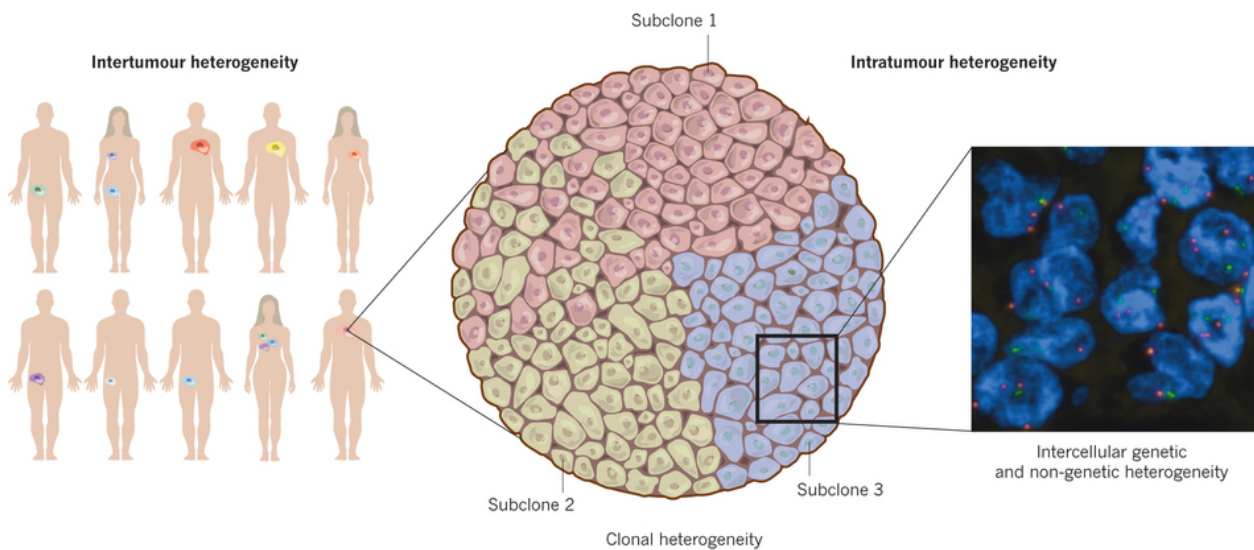
A melhor descrição das etapas de evolução tumoral em humanos dá-se pelo estudo do câncer do cólon e do reto (Figura 3). A doença tem início muitos anos antes da detecção quando uma célula do epitélio intestinal recebe uma mutação suficiente para tornar inativo o gene supressor tumoral APC (Kinzler e Vogelstein, 1996), uma proteína que controla a ativação de  $\beta$ -catenina (Yang *et al.*, 2006). Quando uma célula que já porta a inativação do APC sofre a ativação da uma poderosa via oncogênica através da mutação de KRAS/BRAF, inicia-se a formação de um pequeno adenoma (Phelps *et al.*, 2009). Sucessivamente, novas expansões clonais e mutações em vias

como TP53, PIK3CA e TGF- $\beta$  são os passos necessários para que o adenoma que até então era benigno passe a ser um tumor maligno com capacidade de invadir as diferentes camadas do epitélio intestinal (Jones *et al.*, 2008).



**Figura 3:** Progressão do câncer colorretal. Após a perda do supressor tumoral APC, um pequeno adenoma é formado; RAS mutado resulta em um grande adenoma e a perda de mais um supressor tumoral resulta no carcinoma propriamente dito. Retirado de (Vogelstein *et al.*, 2013)

Essa estruturação retoma o pensamento clássico da evolução de tumores em que as mutações permitem vantagens a células e, após sucessivas passagens, a célula mais apta irá dominar todo o compartimento (Nik-Zainal *et al.*, 2012). Entretanto, é importante lembrar que uma vez formado o carcinoma, ele não é de maneira alguma uma estrutura homogênea composta por células com mesmos padrões de mutações somáticas e expressão gênica. É então que entra a intra-heterogeneidade tumoral.



**Figura 4:** Exemplos de inter e intra-heterogeneidade tumoral. Tumores são diferentes entre distintos tipos tumorais e também apresentam alta heterogeneidade com a existência de diversos subclones dentro do mesmo tumor (Retirado de (Burrell *et al.*, 2013)).

Análises de diferentes tumores em espaços temporais diferentes ou a nível de células únicas mostram a grande heterogeneidade existente entre as células e comprovam a existência de subclones celulares que também fazem parte da população tumoral (Navin *et al.*, 2010; Navin *et al.*, 2011; Nik-Zainal *et al.*, 2012; Francis *et al.*, 2014; Patel *et al.*, 2014; Kim *et al.*, 2015) (Figura 4). O impacto dessas populações subclonais presentes nos tumores é de extrema relevância clínica, visto que essa heterogeneidade é considerada uma das grandes culpadas pela resistência aos tratamentos existentes (Diaz *et al.*, 2012), especialmente quando teorias prevêem que cada lesão metastática detectável por radiografias pode conter pelo menos 10 subclones resistentes (Bozic e Nowak, 2014).

O conhecimento correto das populações subclonais permite o desenvolvimento de terapias personalizadas em que os alvos são mutações *drivers* clonais, assim, será possível identificar mutações essenciais para o início do tumor mas que não são eficazes como alvos terapêuticos ou mutações subclonais que serão letais em caso de resistência tumoral e que devem ser prontamente exterminadas. (Hiley *et al.*, 2014). Paradoxalmente, amostras de pacientes com glioblastoma com alta quantidade de

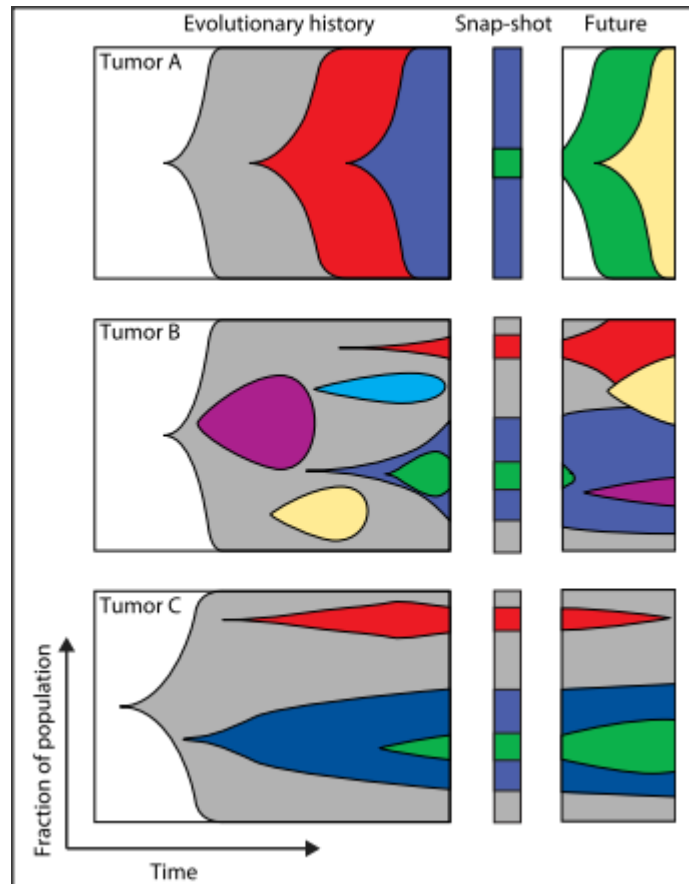
populações subclonais em seus tumores possuem uma sobrevida significativamente maior em relação a tumores majoritariamente clonais (Kim *et al.*, 2015).

Dados de sequenciamento de nova geração já podem ser utilizados para inferir o histórico da evolução clonal e presença de subclones em diferentes tipos tumorais aliados a ferramentas de bioinformática como CloneHD (Fischer *et al.*, 2014) ; PyClone (Roth *et al.*, 2014) e SciClone (Miller *et al.*, 2014), que permitem a identificação e quantificação de populações subclonais existentes em tumores através da análise das frequências de variantes alélicas existentes em dados obtidos por sequenciamento de nova geração.

Em 21 diferentes amostras de câncer de mama, grande parte das mutações foram encontradas somente em uma fração das células existentes na população e, para todos os casos, pelo menos um subclone dominante que compõe mais de 50% da população tumoral foi encontrado (Nik-Zainal *et al.*, 2012). Em glioblastoma, amostras de tumores primários e de recorrência tumoral dos mesmos pacientes foram sequenciadas e analisadas quanto a suas composições subclonais, esses dados foram utilizados para inferir padrões de evolução tumoral comparando as etapas de antes e depois da ocorrência de resistência. Duas possíveis hipóteses para a trajetória da progressão de tumores foram levantadas. A primeira diz que a barreira terapêutica eliminou todos os subclones existentes e a recorrência foi proveniente de uma célula ancestral mais próxima da célula de origem tumoral. Já a segunda hipótese estabelece que a terapia eliminou somente uma parte dos subclones existentes na população primária. É interessante notar que no mesmo trabalho algumas amostras não se enquadram em nenhuma das hipóteses aventadas (Kim *et al.*, 2015).

Assim, é possível conjecturar diferentes hipóteses para a dinâmica evolutiva de tumores. A figura 5 retirada de (Hiley *et al.*, 2014) sumariza esses eventos: (I) sucessivas evoluções clonais resultam em *sweeps* clonais das células mais aptas que dominam toda a população e permitem uma dinâmica evolutiva linear; (II) uma evolução ramificada em que existe uma heterogeneidade ocasionada pela ausência de uma população subclonal dominante. A figura ainda destaca um importante fator quando mostra que as biópsias do tumor B e do tumor C podem propiciar uma falsa

impressão de que os dois tumores são idênticos, embora suas progressões ocorram de maneira distinta.



**Figura 5:** Diferentes dinâmicas de evoluções tumorais. Tumores progridem de diferentes maneiras e pequenas amostras podem gerar a falsa impressão de que eles possuiriam a mesma evolução e a mesma quantidade de subclones. Entretanto, suas evoluções serão completamente distintas. Retirado de (Hiley *et al.*, 2014)

Extensas revisões recentes que detalham os diferentes padrões de evolução tumoral e os últimos conceitos acerca desse tópico foram publicadas e podem ser consultadas, tais como (Merlo *et al.*, 2006; Greaves e Maley, 2012; Aktipis *et al.*, 2013; Sidow e Spies, 2015).

## 8. MODELOS DE EVOLUÇÃO TUMORAL

A importância de modelos na biologia não pode ser negligenciada. Descobertas fundamentais foram influenciadas pela construção de modelos baseados em dados empíricos, tais como a estrutura da alfa-hélice protéica por Linus Pauling e a estrutura do DNA por Watson e Crick.

Na área do câncer, os modelos são um campo de estudo relativamente novo e com poucas publicações. Apesar disso, modelos de evolução tumoral já não podem ser considerados como novidades na literatura. Inicialmente, grande enfoque foi dado para modelos matemáticos e, mais recentemente, a revolução digital provocou uma forte tendência de utilização de modelos computacionais.

Como explicado anteriormente, os tumores originam-se de uma única célula através de mutações genéticas. Tais células multiplicam-se e passam a competir por recursos, estimulando a etapa vascular em que há a secreção de fatores para angiogênese. A limitação de recursos leva a morte celular por necrose em algumas regiões dos tumores e a progressão do tumor estimula a invasão de tecidos adjacentes e a consequente ocorrência de metástases. Com o intuito de simplificar e facilitar a compreensão de características por vezes demasiado complexas ou ainda não completamente estudadas, a literatura procura simular separadamente cada uma dessas etapas, dividindo os modelos em avasculares, vasculares e metastáticos (Roose *et al.*, 2007). Todas estas simulações têm como objetivo reproduzir o que é observado *in vivo* de maneira controlável, rápida e capaz de realizar predições condizentes com dados experimentais *in vitro* ou *in vivo*.

Grande parte dos modelos avasculares preocupava-se em simular o crescimento das células em relação a sua capacidade de receber nutrientes através de difusão (Ward e King, 1999). Utilizaram-se desses modelos os trabalhos de (Burton, 1966; Deakin, 1975; Adam e Maggelakis, 1989; Casciari *et al.*, 1992; Marusic *et al.*, 1994; Please *et al.*, 1999; Ward e King, 1999), dentre outros mais detalhados na excelente revisão do processo histórico de modelos avasculares em (Araujo e McElwain, 2004). Esses trabalhos buscavam explicar até que ponto um tumor poderia crescer e a partir de quando as células atingem uma etapa em que a escassez de recursos predomina e

o crescimento ocorre de maneira mais lenta. Para isso, eles utilizaram modelos de esferas em que, na fase inicial, todas as células recebem os nutrientes necessários para sustentar a proliferação celular. Com o crescimento da esfera, somente a parte externa recebe nutrientes suficientes para sustentar a proliferação e as células da região central desenvolvem uma região necrótica devido a ausência de recursos (Jiang *et al.*, 2005) resultando em um limite para o crescimento do tumor em que somente uma pequena coroa da esfera tumoral possui células vivas e é capaz de proliferar (Burton, 1966). Mais recentemente, novos trabalhos adaptaram o modelo de esferas para incluir uma rede de proteínas que controla a transição entre diferentes fases do ciclo celular e os gradientes de difusão de nutrientes e metabólitos do ambiente extra-celular (Jiang *et al.*, 2005) na tentativa de incorporar maior detalhamento ao já consagrado modelo.

A simulação de tumores vasculares procura detalhar uma segunda fase da evolução tumoral. Tumores a partir de 2 mm deixam de crescer e só retomam o crescimento quando possuem novas fontes vasculares para nutrição de suas células (Folkman e Klagsbrun, 1987), (Chaplain *et al.*, 2006).

O processo de angiogênese só acontece durante processos patológicos, com exceção do desenvolvimento embrionário e durante o processo de recuperação de ferimentos (Chaplain *et al.*, 2006). Sua importância especial em tumores ocorre por ser a via de nutrição da massa tumoral e, também, uma das fontes de disseminação das células para outras regiões durante o processo de metástase. Para a simulação e modelagem do processo angiogênico influenciado por tumores, procura-se entender como as células endoteliais são estimuladas a proliferar e a migrar para a região tumoral e quais são os estímulos fornecidos pelas células tumorais para que esse processo ocorra.

Stokes e Lauffenburger em 1991 mostraram um trabalho pioneiro de simulações de padrões aleatórios de migração de células endoteliais por estímulos quimiotáticos e a importância desses aspectos para a construção da microvasculatura (Stokes e Lauffenburger, 1991). Uma vez que células tumorais secretam os fatores quimiotáticos que exercem grande influência no recrutamento de células endoteliais à região tumoral muitos trabalhos criaram modelos para simular os coeficientes de difusão no microambiente e como isso afeta a migração de células endoteliais em virtude desses



fatores denominados *Tumor Angiogenesis Factors* (TAF) (Chaplain e Stuart, 1993). Não demorou muito para que os modelos de angiogênese em tumores captassem os aspectos essenciais e, passassem a desenvolver estratégias que poderiam resultar em atividades anti-angiogênicas (Orme e Chaplain, 1997).

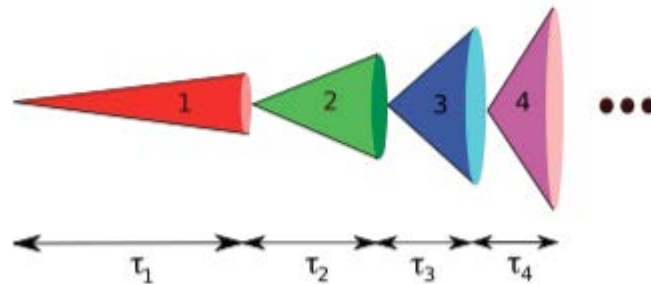
Para um estudo mais detalhado de modelos de angiogênese tumoral, excelentes revisões foram publicadas incluindo modelos matemáticos e computacionais e podem ser encontradas nos trabalhos de (Mantzaris *et al.*, 2004) e (Chaplain *et al.*, 2006).

De maior relevância para este trabalho são os modelos computacionais já produzidos, que apesar de não serem uma novidade na literatura, tratam-se de um ramo recente e em franca expansão. Modelos computacionais já foram utilizados para a verificação da emergência de populações subclonais em glioblastomas (Kansal *et al.*, 2000); crescimento de esferas multicelulares avasculares como as dos modelos matemáticos (Dormann e Deutsch, 2002) e inclusive em modelos de crescimento e difusão das células de gliomas no tecido cerebral e como essa tecnologia pode auxiliar na decisão da extensão da ressecção que deve ser aplicada quando o tratamento envolve cirurgia (Swanson *et al.*, 2002; Swanson *et al.*, 2003).

Beerenwinkel *et al* em 2007 desenvolveram um modelo computacional que utiliza do processo de Wright-Fisher com o principal objetivo de modelar a evolução do câncer de colon a partir de adenomas através do número de mutações e, principalmente, o tempo em que esse adenoma acumulou um número suficiente de mutações e como o *fitness* conferido por essas alterações moleculares afeta o tempo de progressão do processo evolutivo tumoral (Beerenwinkel *et al.*, 2007).

Outro trabalho fundamental que teve por objetivo demonstrar e caracterizar através de modelo computacional o tempo de sucessão entre diferentes ondas clonais caracterizadas por novas mutações *drivers* foi desenvolvido por Bozic *et al.* Nele, os autores aplicaram um modelo em que, a cada geração, as células da simulação poderiam proliferar ou morrer e, em caso de proliferação, receber uma nova mutação que acarretará em um acréscimo no *fitness* das células-filhas. Assim, através dessa simulação foi demonstrado que o tempo de sucessão entre diferentes ondas clonais é menor quanto maior for o número de mutações *drivers* que uma célula possuir, conforme demonstrado pelo esquema representado na figura 6. Não obstante, o

mesmo trabalho mostra que simulações também são capazes de prever a correlação entre o número de mutações *drivers* existentes e o número de mutações passageiras para amostras em tumores pancreáticos e glioblastomas quando os dados obtidos pelas simulações computacionais são comparados a dados clínicos existentes (Bozic *et al.*, 2010).



**Figura 6:** Tempo de sucessão entre diferentes linhagens clonais de acordo com o número de mutações que aumentam o *fitness* celular. O tempo de aparecimento de novas linhagens clonais é menor pelo aumento de mutações *driver* em novos clones. Retirado de (Bozic *et al.*, 2010)

Proveniente do mesmo grupo do trabalho supracitado é a ferramenta Tool for Tumor Progression (TTP): Trata-se de um *software* que, através de simulações computacionais estocásticas, é possível realizar previsões do comportamento de células sob diferentes valores de *fitness* celular; número de mutações ou tempo de divisão celular. Para isso, a cada geração, as células podem morrer ou proliferar e o programa permite a modulação de quatro posições (cada uma atuando como um possível gene que será ou não mutado) em que diferentes valores para cada posição atuam como multiplicadores de *fitness* celular. Interessantemente, também existe um mecanismo de densidade populacional para que seja possível simular modelos em que tumores estão restritos pelo espaço físico disponível para sua proliferação. Ao final, o programa disponibiliza gráficos em que é possível acompanhar a dinâmica evolutiva do tumor simulado através do acúmulo de células que possuem uma ou mais das quatro posições possíveis mutadas (Reiter *et al.*, 2013).

O trabalho mais semelhante ao nosso e que compartilha diversas características com o trabalho que produzimos e ao qual buscamos aprimorar foi desenvolvido por Abbott *et al* em 2006. Nele, os autores desenvolveram um *software* denominado CancerSim com o objetivo principal de simular os Hallmarks of Cancer em que, partindo de uma célula única em um grid de três dimensões, as células podem proliferar ou morrer de maneira estocástica e, a cada divisão, podem receber aleatoriamente mutações em uma das 6 características principais do câncer em que os autores se propuseram a simular. Várias características interessantes e fundamentais em relação a evolução tumoral foram implementadas nesse programa e merecem destaque, por exemplo, em cada quadrado do grid os autores simulam vasos sanguíneos que simulam gradientes nutricionais que suportam a proliferação celular. Além disso, uma interessante simulação de inibição por contato também ocorre em que as células estão limitadas em seu crescimento em razão do espaço físico que possuem, ou seja, uma célula com todos os seus espaços vizinhos preenchidos não pode proliferar (Abbott *et al.*, 2006).

Esse mesmo *software* foi utilizado em outro trabalho publicado por Spencer e colaboradores também em 2006 e destaca como modelos computacionais podem ser utilizados para inferir características fundamentais da tumorigênese e auxiliar pesquisadores em direcionar suas áreas de pesquisa a tópicos até então de pouco destaque na biologia tumoral. Nesse trabalho, através da utilização do CancerSim os autores puderam verificar como a instabilidade genética influencia na heterogeneidade tumoral; como a sequência de mutações diferem entre tumores que iniciam precocemente comparado a tumores tardios e a importância da angiogênese em estabelecer as condições ideais de microambiente para que células com outras mutações aproveitem a nova conjuntura para a sua proliferação (Spencer *et al.*, 2006).

Modernos modelos matemáticos e computacionais permitem prever a probabilidade de detecção de metástases em câncer de pâncreas assim como estimar a probabilidade de sobrevivência de um paciente em relação ao tamanho do câncer primário; o número de metástases encontradas e a taxa de crescimento das metástases. Mais do que isso, o mesmo modelo é translacional e auxilia na escolha de tratamento

adjuvante e cirurgia em razão da predição da sobrevida do paciente quanto ao seu estágio no momento do diagnóstico (Haeno *et al.*, 2012).

Assim, modelos computacionais buscam auxiliar na elucidação dos fatores fundamentais da tumorigênese e, futuramente, podem servir como plataformas rápidas de simulação evolutiva de tumores presentes em pacientes caracterizados através de suas distintas anormalidades moleculares.

## 9. OBJETIVO GERAL

Criar um modelo *in silico* de evolução tumoral com base em pequenos eventos somáticos e seus impactos no *fitness* celular.

## 10. OBJETIVOS ESPECÍFICOS

- Simular aspectos básicos da biologia celular como divisão e morte celular, genoma, mutações por divisão e senescência.
- Simular aspectos básicos da biologia tumoral como oncogenes e genes supressores tumorais, impacto de mutações e senescência induzida por oncogenes.
- Permitir a customização do programa para a simulação de diferentes tumores, com as mutações catalogadas pela literatura atual e seus impactos.
- Permitir a simulação de tratamento através da alteração das probabilidades de proliferação e morte para simular o impacto de um quimioterápico na evolução tumoral.

## **11. ARTIGO CIENTÍFICO**

Neste capítulo está apresentado o artigo científico resultante do trabalho desenvolvido durante o período do mestrado. Este artigo será submetido ao periódico “Plos Computational Biology”, por isso a formatação deste capítulo está de acordo com as regras desse periódico. A discussão será feita com base nas figuras aqui apresentadas.

## OncoProSim: A tool for *in silico* tumor evolution analysis

Darlan Conterno Minussi<sup>1¶</sup>; Bernardo Henz<sup>2¶</sup>; Eduardo Cremonese Filippi-Chiela<sup>3</sup>; Manuel Menezes Oliveira<sup>2</sup>; Guido Lenz<sup>1\*</sup>

<sup>1</sup> Biophysics department, Federal University of Rio Grande do Sul, RS, Brazil

<sup>2</sup> Institute of Informatics, Federal University of Rio Grande do Sul, RS, Brazil

<sup>3</sup> FAMED, Federal University of Rio Grande do Sul, RS, Brazil

\* Corresponding author:

E-mail: lenz@ufrgs.br

¶ These authors contributed equally to this work

## **Abstract**

The current knowledge regarding tumor evolution, until now, has been obtained from what we can gather through biopsy samples. In spite of the importance of these samples to diagnosis and treatment, they represent only a glimpse of the whole tumor evolutionary path, whereas the majority of the tumor development remains hidden. In this work, we developed a software to, simulate the clonal evolution of different types of tumors, focusing on the molecular mechanisms that lead to tumor progression. In order to do that, we generate cells with two arrays that represent the diploid characteristic of the human genome and, in case of cell division, mutations can be inserted in the genome that may alter the default probabilities of proliferation and death. With the help of a pseudorandom number generator, we can use our model to simulate different patterns of tumor evolution and investigate the effects of different mutational spectrum in the incidence of distinct tumors. Moreover, we can use our model to reproduce essential characteristics of the tumor genome such as: the synergy between oncogenes and tumor suppressor genes, changes in mutation frequency with distinct fitness values and even simulate tumor treatment. Therefore, with the increasing knowledge in tumor biology, we believe that our model can offer a unique perspective of tumor evolution, allied to the speed and reproducibility that only *in silico* models are capable of offering.

## **Author summary**

Cancer is a perverse disease that affects millions of people throughout the world. Unfortunately, current treatment is not always able to eliminate the totality of the transformed cells that compose the tumor mass and, therefore, some of these patients will present a more malignant version of the disease that might be unresponsive to the previous treatment regimen. Since the



path from the disease origin until its clinic diagnostic remains unknown, it is currently impossible to determine the many molecular transformational stages that the tumor has surpassed until its detection. Therefore, tumor biology could hugely benefit from tools that help uncover the hidden tumor evolutionary story. Here, we sought to develop a software to fill that gap, where tumors are recreated in silico by reproducing, stochastically, distinct possible pathways of the cell's molecular machinery mistakes, based on the current knowledge about significantly mutated genes in different types of cancers. Also, different weights for the impact of a given gene mutation in the cell fitness can be tested and, therefore, new pathways can be discovered based on gene mutational impact. Finally, we also propose a simple way of test treatment efficacy based on how the proliferation or death probability is affected.

## **Introduction**

Decades of cancer research have already provided some evidence of how tumors behave after diagnosis. For most cases, a prognosis regarding tumor progression can be made based on histological and genetic information of the tumor. However, key aspects of cancer biology remain unknown and it is still unclear how tumors progress from its origin until they become detectable.

Our current understanding of cancer is limited to the knowledge we can gather through biopsies, which are the equivalent of a photograph of the final stage of a tumor, and little is known about the disease behavior until that moment. A complex disease such as cancer cannot be completely understood unless it is analyzed dynamically. To overcome its last alamo - cancer resistance - it is imperative to fully understand the complete course of the disease. This is only obtained with comprehension of its stages of clonal evolution both for tumor establishment and recurrence.

Attempts to simulate tumor evolution are by no means novel. Since Armitage and Doll multi-stage theory (1) and Knudson two hit hypothesis (2), several mathematical models of avascular and vascular tumors simulated tumor growth, angiogenesis, and treatment (reviewed in (3-5)). Recently, a strong trend towards *in silico* models has emerged, especially since the dawn of the big data era, where data from microarray to next generation sequencing became available. Likewise, this tendency towards computational models is supported by the ever increasing knowledge of cancer molecular and cellular biology.

Proof that *in silico* models can be used to generate biological insight was given by softwares such as CancerSim, which simulates a vascular replica of tumor evolution. Concerned especially with the six mutational pathways described by the Hallmarks of Cancer, it was used to predict features that are commonly found in tumors, such as the importance of genetic instability on early onset cancers, tumor heterogeneity, and how some mutations serve as stage-setting to boost and sustain proliferation (6-9). Likewise, other computational models have been used to estimate the waiting time for the onset of cancer based on the number of driver mutations in a cell genome (10) or, demonstrate that the rate of clonal expansion increases with each new driver mutation, while also estimating the number of passenger mutations in a tumor as a function of driver mutations (11).

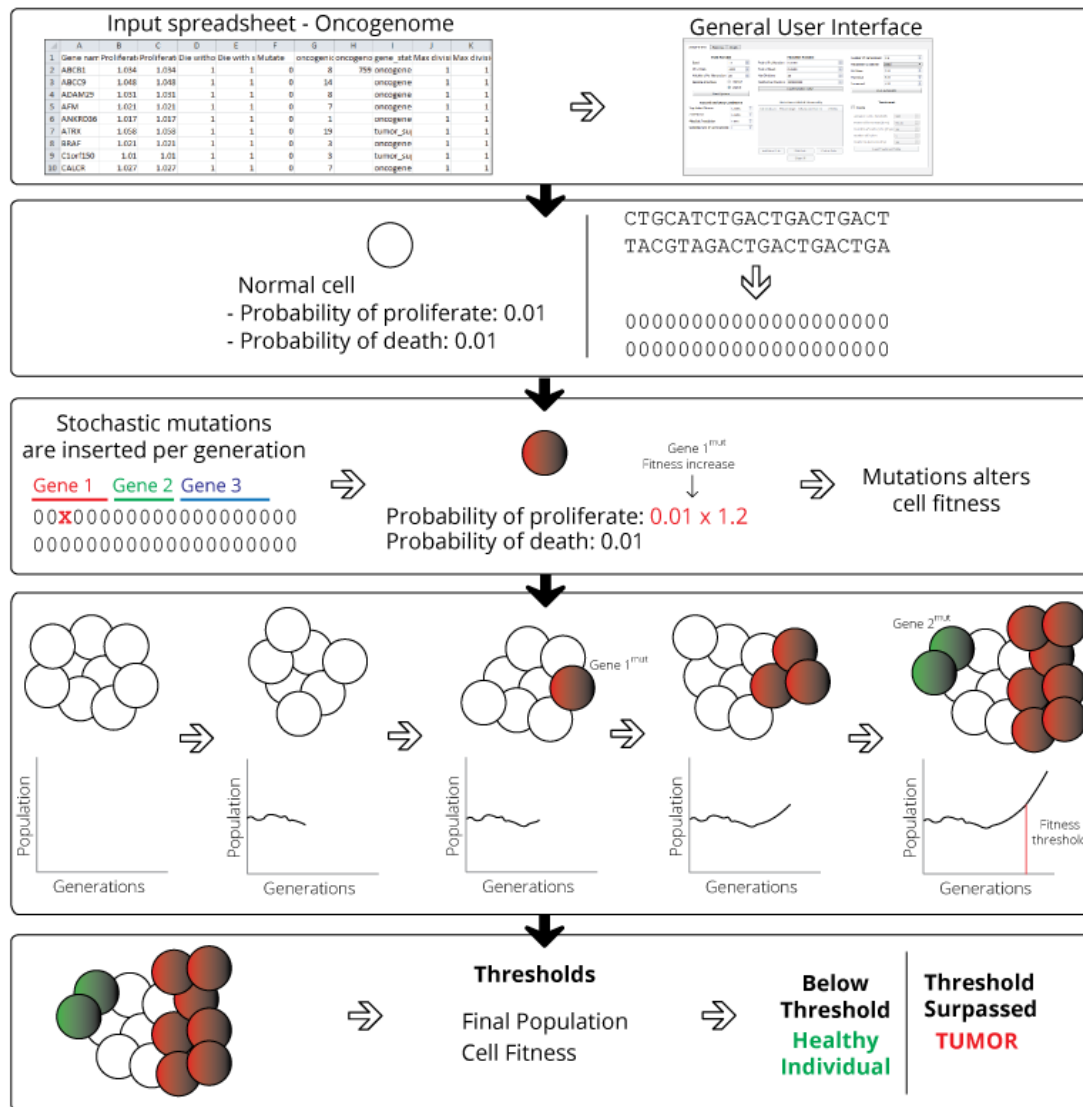
Here we simulate cancer evolution through a computational process. In order to construct a reliable model of tumor evolution, first and foremost, it is necessary to reproduce the most basic aspects of cell biology. However, such an intricate and yet to be fully understood system, poses serious computational challenges, where it is inevitable to resort to some kind of simplification during the process. Hereby, we endeavor not only to reproduce aspects of the basic cell biology but also to emulate one of its critical mistakes: the cancer initiation and evolution process. To the best of our knowledge, there is no model that permits to specifically model any tumor, chiefly allowing integrating the current knowledge about the mutational spectrum of different types of cancer with a unique weight for each genome alteration.

Therefore, we developed the Oncogenomic Progression Simulator (OncoProSim). With our model, it is possible to estimate different evolutionary trajectories for tumors, as well as estimate the impact of particular mutations. We developed a tumor model that can provide clues regarding tumor initiation, progression, and treatment, notably using cell fitness as the leading component in pursuance to dissect the essence of all tumors: its damaged cell molecular machinery.

## Results

### OncoProSim Overview

Every cell in our software is initiated with two arrays that represent the diploid characteristic of the human genome. Each element of such an array corresponds to a base. Groups of consecutive bases represent genes that can be classified as either an oncogene or as a tumor suppressor gene. The gene list, as well as each gene classification, oncogenomic size and multiplier weights are given by a customizable user input spreadsheet (table 1). At every generation, each cell can stochastically proliferate or die according to their default probabilities and, in case a cell does proliferate, a number of mutations are inserted into the genome arrays that can either affect a gene or a non-gene region. Importantly, if the mutation affects a gene, the probability of proliferation and death will be altered based on the modifier given by the input table provided by the user. A pseudorandom number generator, when initiated with different seeds, allows the test of several evolutionary histories. Thresholds for either fitness or cell population number level can be set to denote what is considered an *in silico* tumor (Figure 1).



**Figure 1: Overview of OncoProSim:** General scheme of OncoProSim including input, cell biology, mutation, tumor evolution, and outcome.

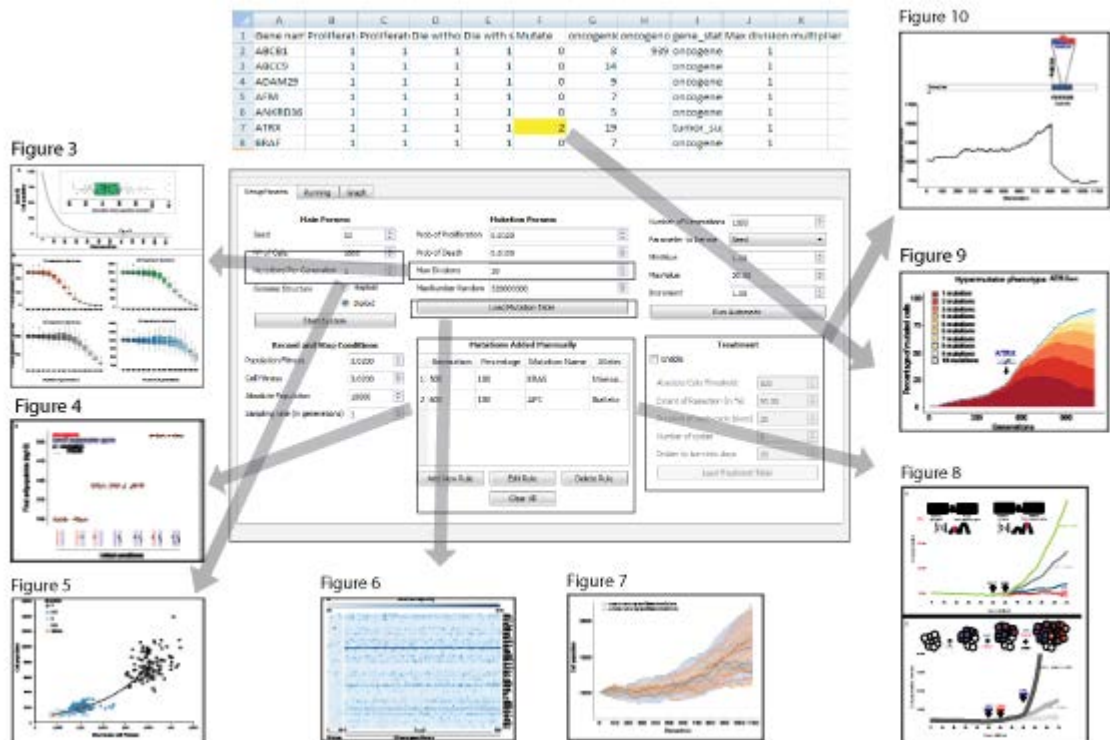
### OncoProSim Interface

We strived to deliver an experience where the user has absolute control of the input for each parameter. Every characteristic from the initial number of cells to the genome size can be customized, including a spreadsheet where specific gene that might be randomly mutated during

the course of the simulation are loaded and, the impact that the mutation (i.e. any kind of genetic alteration) of this gene has on the cell fitness. (table 1 and supplementary table 2).

In this manuscript, we will describe potential uses for the different parameters available to the user, therefore, each figure will refer to a unique aspect of OncoProSim and, mostly, will serve as a proof of concept that the respective tumor biology characteristic is adequately reproduced by OncoProSim (Figure 2)

Additionally, a different spreadsheet controls the therapy parameters, in which the time of therapy, the “concentration” of chemotherapy (i.e. the effect the therapy has on proliferation, death), and mutation of the cells (table 3 and supplementary table 3).

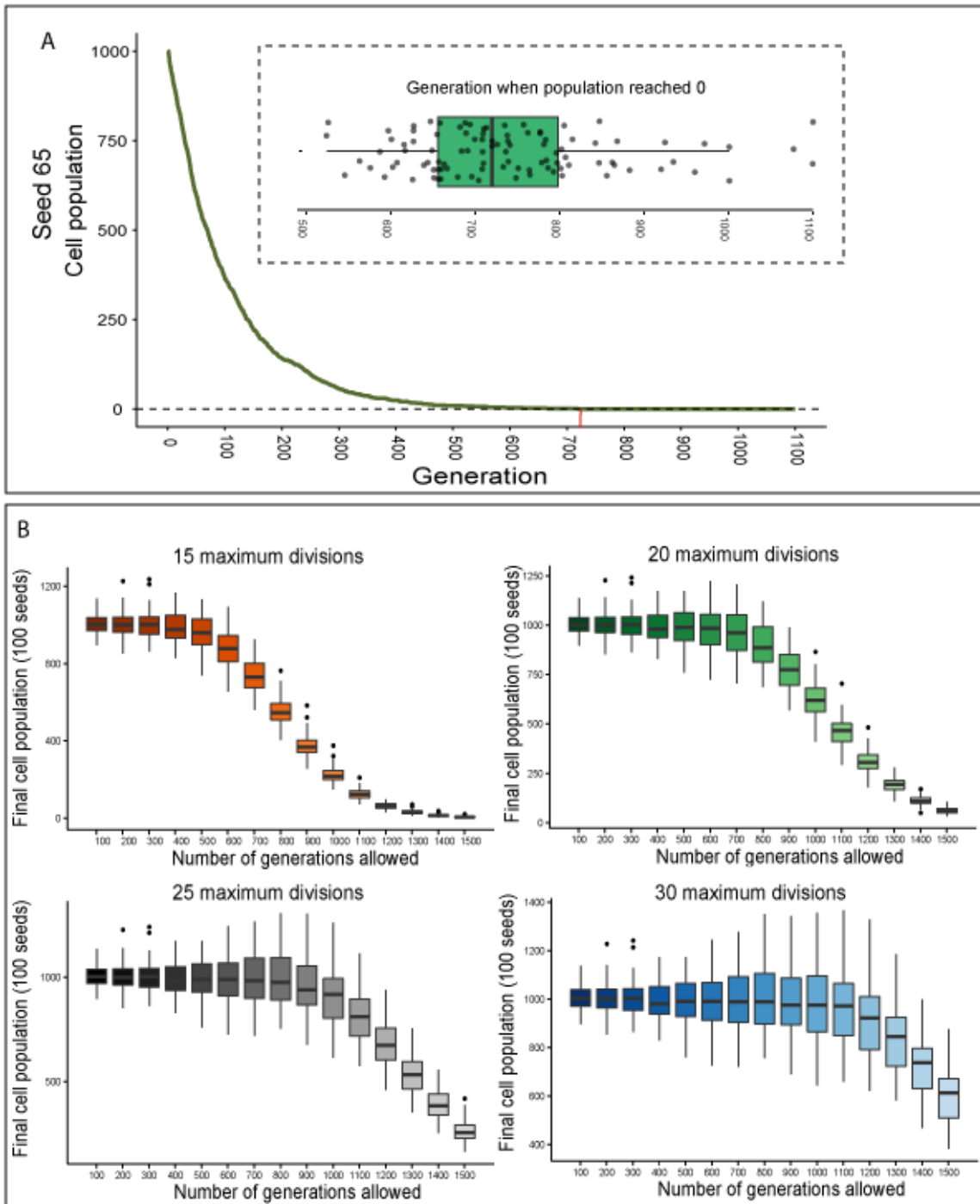


**Figure 2: General User Interface of OncoProSim:** Every input characteristic can be controlled by the user. Each figure is used as proof of concept of what can be achieved with OncoProSim.

### **Number of maximum cell divisions is used to simulate cell senescence and the length of each simulation**

Normal human cells generally enter a state of proliferative arrest called senescence due limit in cell divisions known as the Hayflick limit, controlled mostly by the telomere length (12-16). To simulate the Hayflick limit, we provide the maximum number of cell division for each cell as a parameter. All cells are initiated with the same number of maximum divisions given by user input and at each division, this number is reduced by 1. When it reaches zero, the cell is not allowed to proliferate anymore, though it can still die stochastically. If we force all cells to start with the maximum-cell-division parameter set to zero, each simulation will reach a null population number at a different generation due to random cell death (Figure 3A)

Therefore, the maximum number of cell divisions can be used to establish the length of each simulation. Higher values will lead to an increase in the number of generations before the simulation starts to significantly decrease the cell population. As expected, an increase in the maximum number of division increases the number of generations required for reducing the cell population (Figure 3). For this purpose all simulations were performed with the maximum number of divisions parameter set to 30 and until the 1100 generation. The large variation in the size of the final cell population obtained in 100 simulations is a key aspect that indicates the stochastic nature of the simulation process and the reason why each run can be considered as a different patient.



**Figure 3: Maximum divisions limits the number of times that a cell can proliferate.** Cells that reach zero maximum divisions cannot proliferate anymore but are still susceptible to die stochastically. A) Sample simulation with maximum divisions starting at 0. Each distinct seed will reach a null population in a different generation B) Maximum division number can limit the

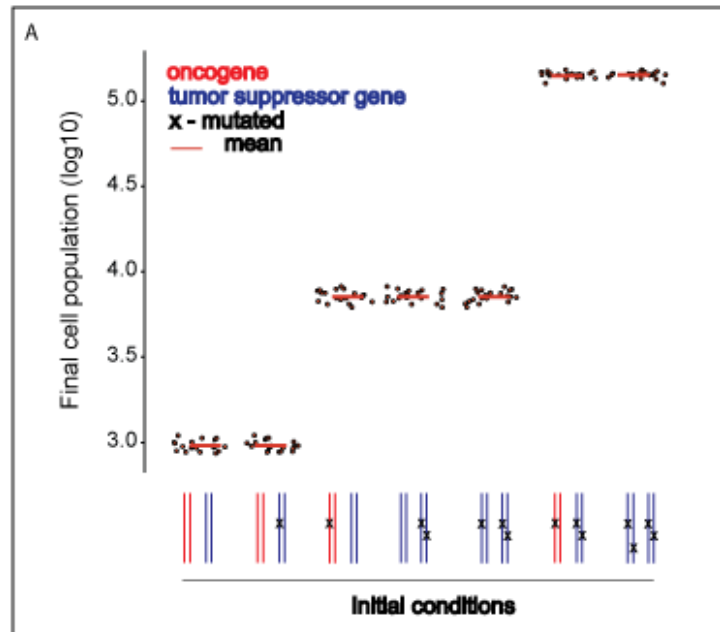
length of the simulation. Smaller values will result in a faster decay in the number of generations that the simulations runs through before reaching a null population.

### **OncoProSim distinguishes between oncogene or tumor suppressor genes**

Oncogenes are activated when a mutation occurs within a proto-oncogene; a widely known example is the RAS oncogene, where a missense mutation results in an amino acid substitution that turns the protein permanently on, sending downstream proliferating signals in many types of human cancer (17-20). On the other hand, tumor suppressor gene mutations result in loss-of-function of a given protein (21, 22). Mutation of one allele of a proto-oncogene is normally sufficient to initiate the oncogenic state of the affected protein, whereas tumor suppressor genes generally need mutations in both alleles to become inactivated (23). Therefore, for the multiplicative effect of an oncogene to be activated in OncoProSim, it suffices that only one of the gene alleles receives a mutation, whereas tumor suppressor genes are required to have both alleles mutated before any multiplicative effect is triggered.

Using as input only two equally sized genes with the same multiplicative values, we tested OncoProSim performance under different conditions for those genes: one being oncogene and the other a tumor suppressor gene, and a second case where both genes were tumor suppressor genes (Figure 4). All cells received initial mutations either in one allele or in both alleles simultaneously; no further mutations were allowed. In each condition, final cell population numbers for 20 different cases were analyzed.





**Figure 4: OncoProSim distinguishes between mono-allelic or bi-allelic mutations.** Oncogenes are activated with a mono-allelic mutation whereas tumor suppressor genes require mutations in both alleles to be inactivated.

A mono-allelic mutation of a tumor suppressor gene did not significantly increase the number of cells over non-mutated initial state. As expected, mono-allelic mutation of an oncogene or biallelic mutation of a tumor suppressor gene had the same effect on cell proliferation, producing 7-fold increase in cell number. The same effect was observed when two tumor suppressor genes are used, being the first mono-allelic and the second biallelic mutated, after all, the first tumor suppressor needs both alleles to be mutated to activate its multiplier effect. Lastly, when using a mono-allelic mutated oncogene and a biallelic mutated tumor suppressor gene, we can observe their synergistic effect, as both proliferation multiplicative effects were activated and, accordingly, cell population number is 147-fold higher than in non-mutated cases.

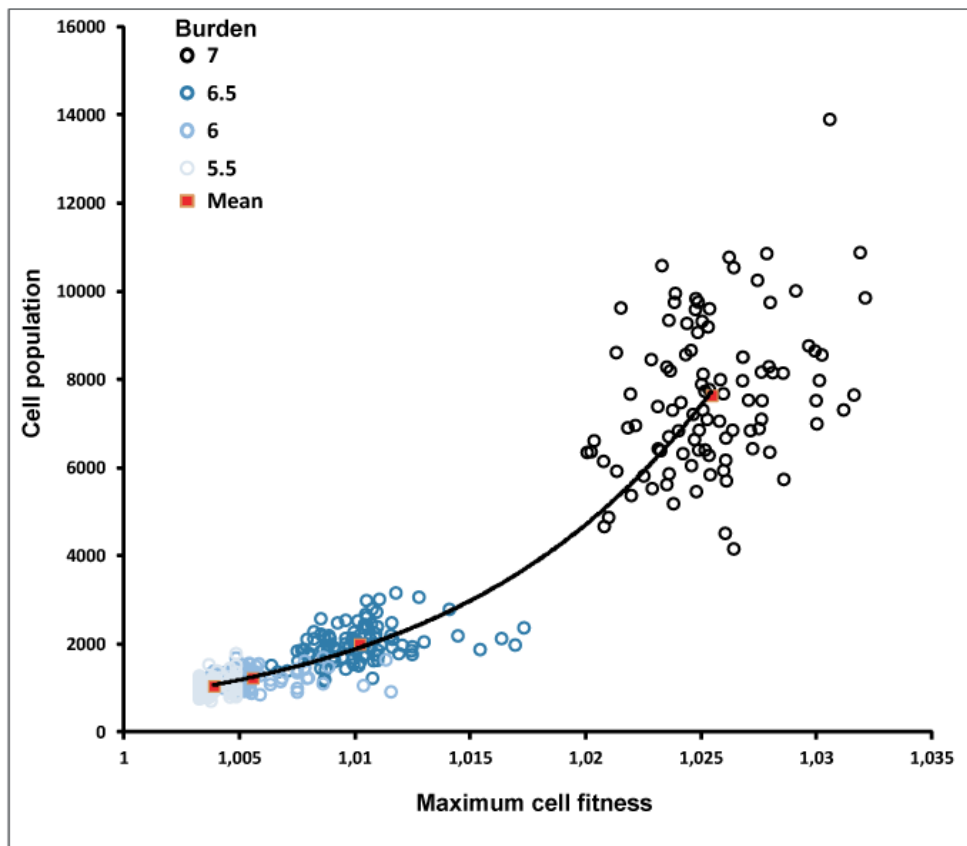
## **Higher mutation number per generation leads to increased cell fitness heterogeneity**

Tumors are highly heterogeneous entities. Heterogeneity can be seen among patients with the same type of tumor (24-26) and even among the primary tumor itself (27, 28). Here, we determine cell fitness as a measure of each cell adaptability to proliferate or die within the simulation, where cell fitness is given by the formula:

$$fitness = 1 + Probability\ of\ Proliferate - Probability\ of\ Death$$

At every generation the maximum cell fitness achieved in the simulation as well as the population average fitness are registered. Moreover, fitness thresholds can be defined and are marked in the simulation so the user can specifically verify the position where the threshold was triggered. We can use OncoProSim to observe the distribution of cell fitness as we increase the number of mutations per generation. In order to do so, we used glioblastoma as our disease model. A table with the significantly mutated genes detected in glioblastoma (25) served as input for the genome of each cell. Their respective gene mutation frequencies were used as proliferation multiplicative values in case of mutation. One hundred simulations were performed for each condition and all parameters were kept unaltered except for the number of mutations allowed per generation.

As the number of mutations per generation increases, so does the distribution of cell fitness among the simulations and the final cell population. Notably, these results might generate some insight into how hypermutated samples can have a high distribution in their fitness. It is also worth mentioning how such a small increase in cellular fitness can dramatically change the final cell population (Figure 5).



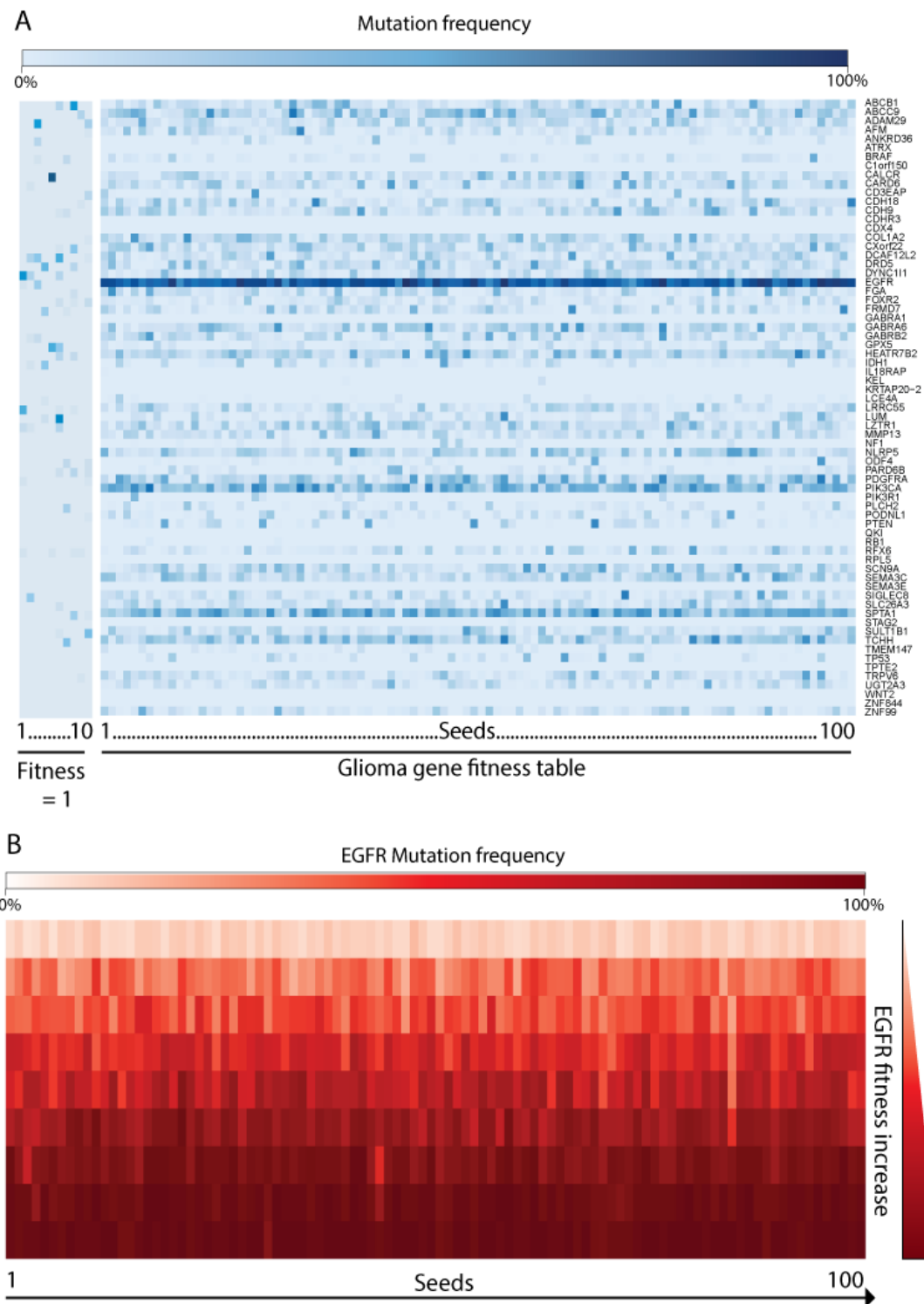
**Figure 5: Increase in mutation rate results in highly heterogeneous distribution of fitness.**

With low mutation rate, few cells are affected by mutations and have only a small fitness increase. With higher mutational rate, all simulations had cells mutated and present a high heterogeneous distribution of fitness. Also, increases in mutational rate result in increase in the final population. It is worth noting that small increases in cell fitness are enough to sustain a high surge in the number of cells.

### **Multiplicative values can be precisely controlled and lead to different outcomes**

OncoProSim can provide the frequency that each mutation appears in the population at the end of the simulation. The quantification of how much a mutation affects cell fitness remains an open question. In this work we used the rationale that gene mutation frequency (25, 29) is an

indicator of their impact on cell fitness. Although we are aware that fitness may change along the evolution of the tumor, and that their representation at the end of evolution is the endpoint, with limited temporal information about the importance of a given gene through oncogenesis. Unfortunately, this is just one piece of a still unknown equation about the impact of individual genetic alterations in cancer. OncoProSim provides full control of each mutation impact. Multiplicative values for each gene highly influence how that mutation will be selected throughout the simulation, importantly, when all mutations are set to have equal effects, we do not observe the dramatic changes on mutation frequency heterogeneity in different genes. Therefore, OncoProSim is capable of simulating selection of the cells that received the mutations with the highest impact (Figure 6A). In order to demonstrate how different multiplicative values impact the percentage of affected cells by the end of the simulation, we used EGFR, a commonly altered gene in glioblastoma (25, 30). All parameters were kept constant except for EGFR mutation probability of proliferation values. As fitness was increased, the proportion of cells that presented EGFR mutation in their genome at the end of 100 simulations also increased (Figure 6B). It is worth noting how different fitness values have a huge impact on whether a determined mutation will be more frequent. Again, it is interesting to note how each single run is representative of a very unique case, in conformation with human samples that are highly heterogenic.



**Figure 6: Mutational spectrum for different simulations.** Each row represents a gene and each column a different seed. Colors are darker according to the mutational frequency of that

gene at the end of the simulation. A) Glioblastoma significantly mutated genes are used as example of the variability observed in OncoProSim due to its stochastic simulations. Two major aspects influence mutation frequency of a given gene: Gene oncogenomic size and its multiplicative values. Due to the necessity of homozygous mutation for tumor suppressor genes, oncogenes frequency are higher than tumor suppressors. It is worth noting that a control spreadsheet does not show any specific selection on gene mutational frequency. B) Proof of concept that serial increases in the multiplicative proliferation values of a gene will increase its mutational prevalence at the end of different simulations.

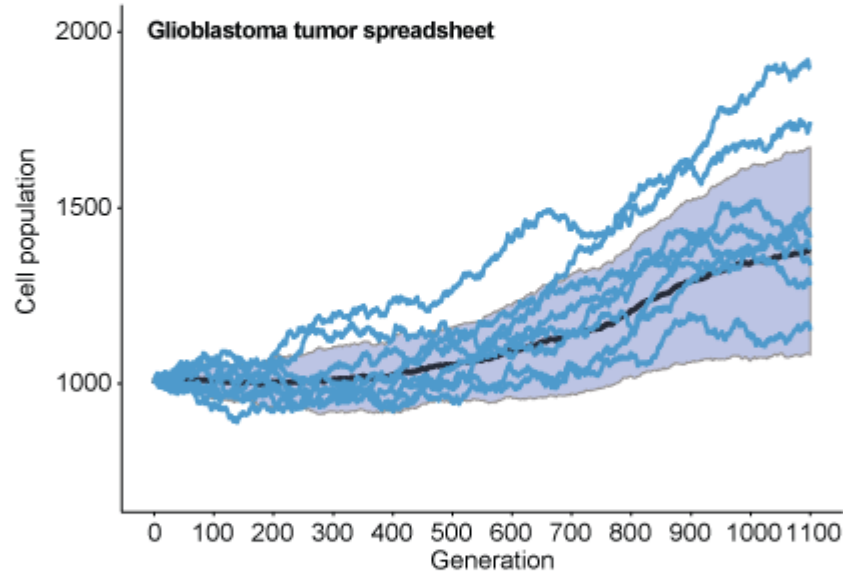
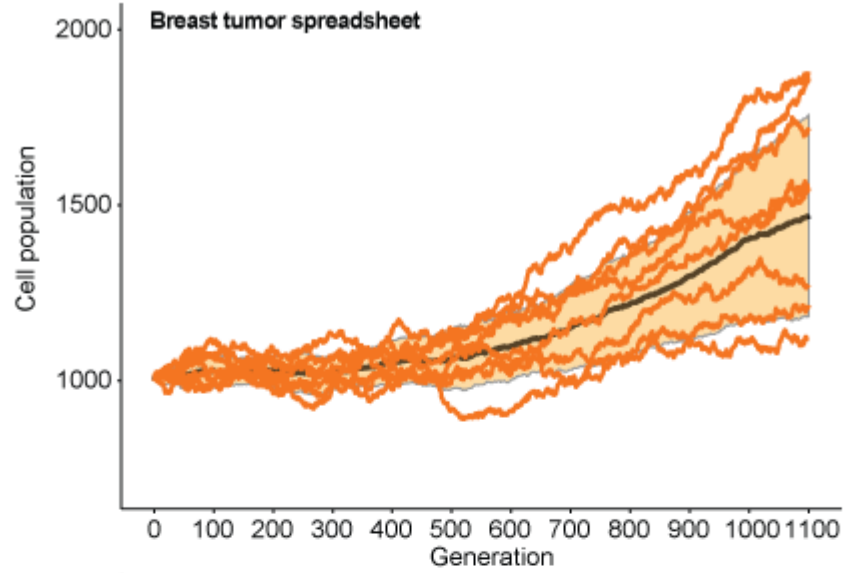
### **OncoProSim simulates different tumor evolutionary trajectories and distinct tumors**

One major aspect of OncoProSim is the inherently efficiency that only *in silico* platforms can provide. In OncoProSim it is possible to run more than a hundred different simulations within one hour on a typical personal computer, with every simulation distinct from each other. Such a powerful aspect can be used to efficiently simulate several types of cancer and its incidence in regards to the threshold chosen by the user. With the help of a pseudorandom number generation the user can use different seeds where each one of them will lead to a distinct evolutionary pathway during the course of the simulation (Figure 7A).

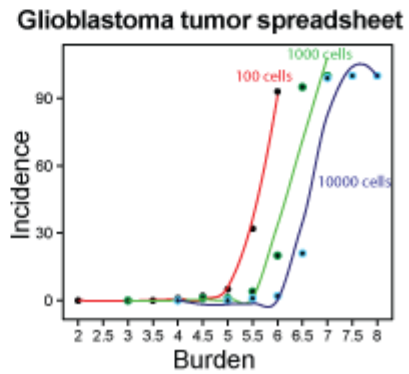
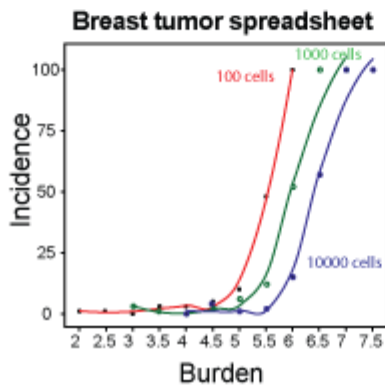
Allied to this fact, by changing the input, one can simulate a variety of different tumors. For demonstration purposes, we used glioblastoma and breast invasive carcinoma specific input tables (supplementary table 2) to observe the dose-dependent response on the incidence of those tumors, with cell population as the base threshold criteria. A simulation was considered tumoral whenever its cell population reached a number of more than 3-fold of the mean of a simulation with no mutations allowed.

We used different tumor burden, i.e. the product of the initial population by the mutations per division, to analyze the incidence of tumors per burden with different number of initial cells. As the initial population increases, so does the burden needed to produce tumoral simulations. Importantly, using the breast cancer input data, tumors generally appear sooner than when using the glioblastoma ones (Figure 7B).

A



B





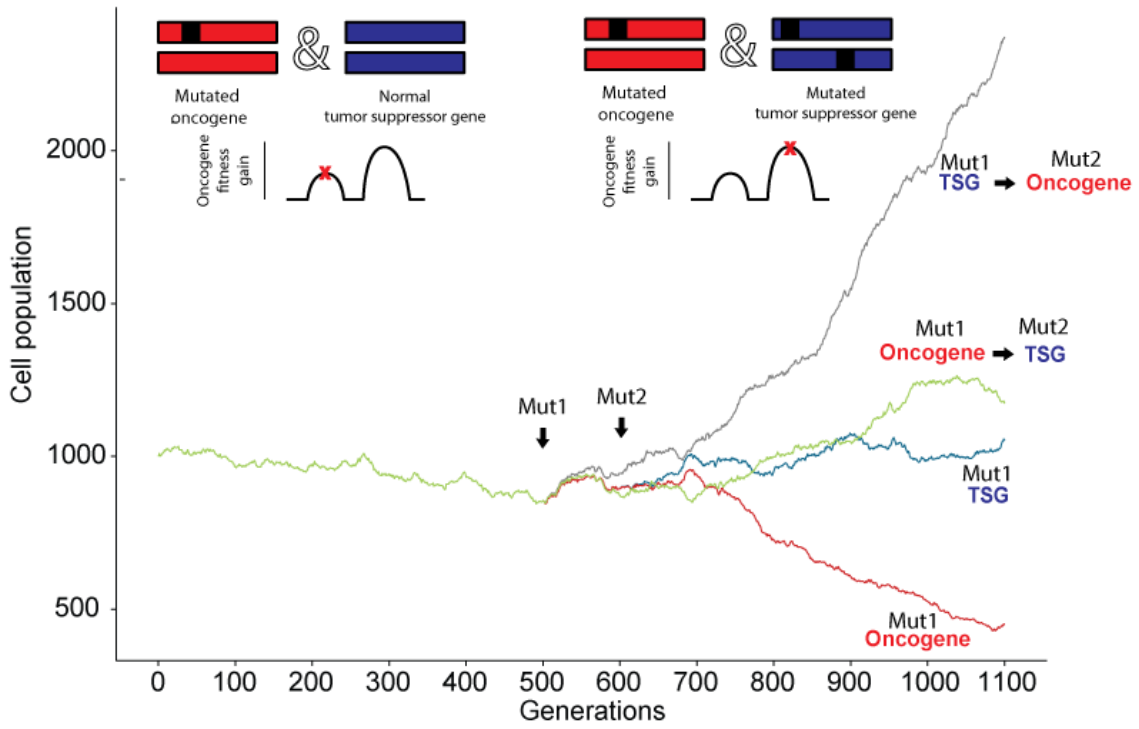
**Figure 7: Simulations and incidence for two different tumors.** Input data for breast and glioblastoma cancer were used. A) Simulations for 8 different seeds for each tumor. Orange background represents standard deviation for 20 breast simulation whereas blue background represents standard deviation for 20 glioblastoma simulation samples. B) Tumor incidence for different burdens. Tumor was established when cell population was 3 times higher the standard deviation of OncoProSim stochastic fluctuation.

### **Multiplicative values can be controlled in a step-wise manner**

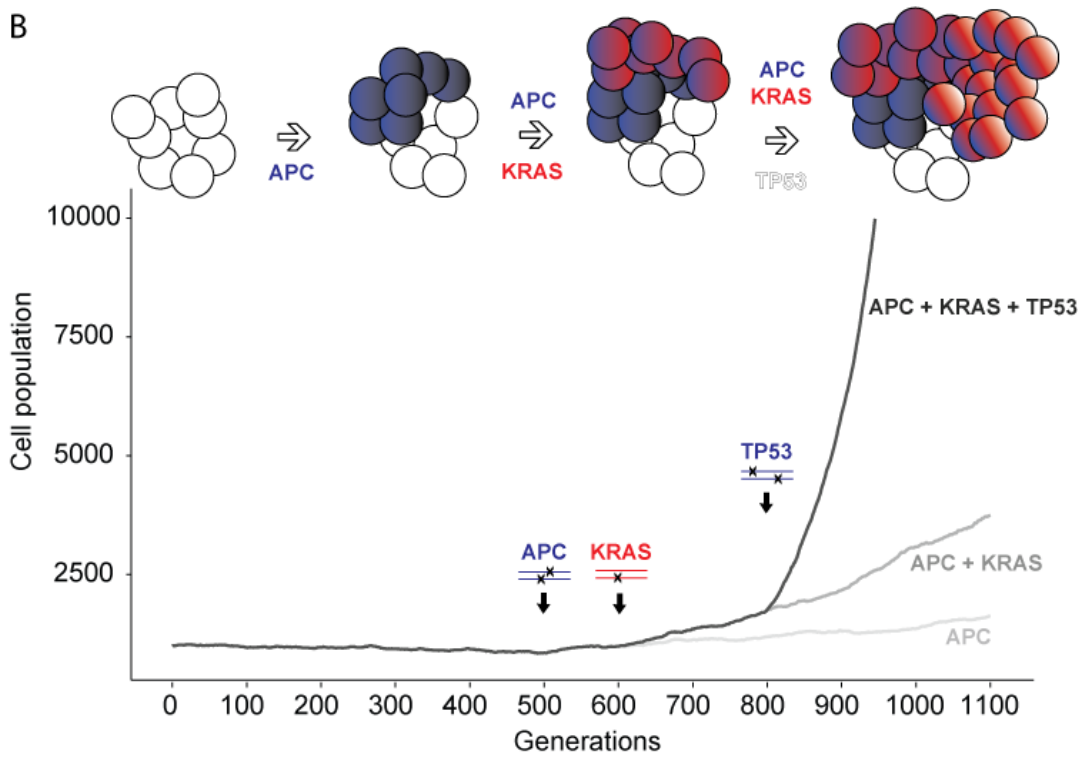
Few cancers have their step-by-step evolution described. Colorectal carcinoma is among one of the rare examples where at least one major pathway of tumor evolution from a normal cell to a full blown carcinoma is known. Lgr5+ cells in the intestinal crypt (31) are believed to be the cell of origin for this type of disease. Those cells have their APC gene (a WNT-pathway acting gene) inactive, which results in the release of the B-catenin protein. This first mutation is enough to result in abnormal cell proliferation consolidated in a small adenoma. If a second mutation is introduced. In the absence of APC, like a mutation in KRAS, higher proliferation ensues and the tissue now develops a large adenoma. If, by chance, a third mutation in either the TP53 tumor suppressor gene or in the PI3K pathway happens, the result is a full blown carcinoma (32-35).

OncoProSim allows the user to reconstruct this kind of data by applying the known cancer biology to its input parameters (Figure 8). In order to obtain a fine control over the influence of an oncogene mutation we introduced a step-wise activation of gene multiplicative values. For example, the full effect of a determined oncogene will only be activated if the same cell has also one tumor suppressor gene mutated. Otherwise, even though an oncogene mutation happens, the user has full control if that mutation will cause a small or a big effect in cell fitness and even if that mutation is controlled by an active tumor suppressor gene.

A



B



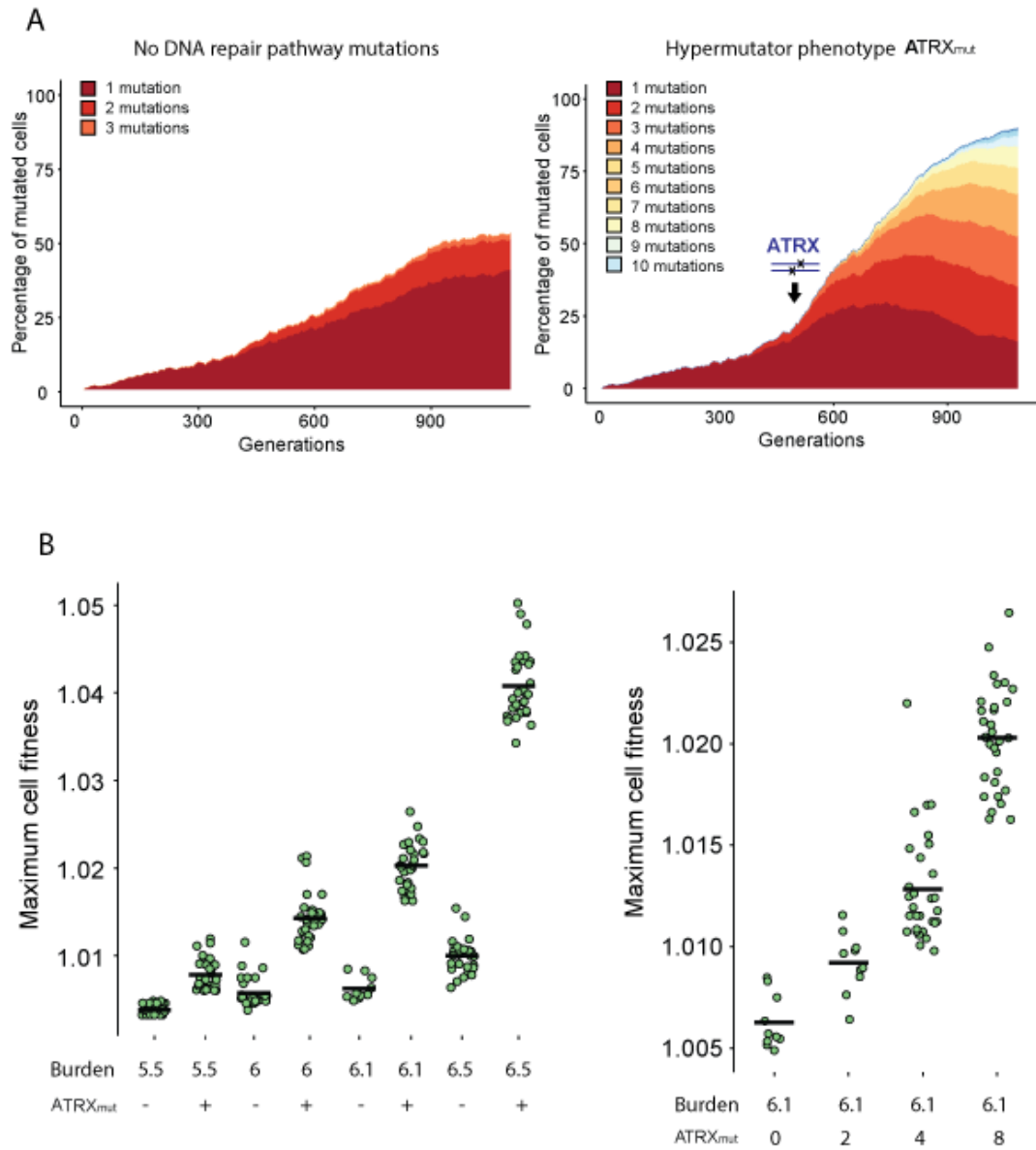
**Figure 8: Mutations can have synergistic effect and OncoProSim is able to simulate tumor evolution.** A) Oncogenes may not have their full multiplicative effect if there is not at least one tumor suppressor gene mutated. An oncogene mutation alone triggers oncogene induced cellular senescence. Only if a tumor suppressor gene is mutated before an oncogene, their combined effect triggers a substantial increase in the cell number by their synergistic effect. B) Colorectal carcinoma evolution is reproduced with OncoProSim. Multiplier values for each gene mutations were adjusted accordingly the most common described evolutionary pathway.

That concept is illustrated in Figure 8A, in the absence of a mutated tumor suppressor gene, KRAS has little effect over the cell probability of proliferation. However, if there is indeed a tumor suppressor gene mutated, the full multiplicative effect of KRAS is activated and the tumor will have a huge boost in its fitness besides the one given by the tumor suppressor gene mutation. Also, it is important to note that oncogene induced senescence can also be simulated when KRAS is mutated before at least one tumor suppressor gene is mutated.

### **Number of mutations per cell can be traced and mutated DNA repair pathway genes results in hypermutated simulations**

Lineage tracing is one of the most important aspects of tumor evolution. Being able to keep track of the mutations that a given cell has suffered throughout its history, and predict how it will behave in the near future, will be crucial to personalized cancer treatment. We sought to provide means for the user to track the different clonal expansions that appear during the tumor evolution. OncoProSim tracks each clonal expansion by recording the frequency of cells that carry  $n$  driver mutations. OncoProSim does not yet provide a thorough detailed explanation of the clonal waves that appear during the simulation with the specific gene for each wave. This will be implemented in future versions of the software.

OncoProSim also allows simulations of hypermutator phenotypes when a gene that is involved in the DNA repair pathway is mutated. Glioblastoma has been described as a type of tumor known to have some hypermutated samples (36). Genes such as ATRX have been described as important tumor suppressor gene whose inactivation results in activation of alternative telomere lengthening and extensive genome rearrangements (37). Here we used ATRX as an example of how a mutation in a caretaker gene can increase the number of mutations applied in each generation and how that can simulate tumors with hypermutator phenotypes, we have arbitrarily chose an 8-fold increase in mutations per division after ATRX mutation. However, the user has the freedom to set any number that he may desire (Figure 9A). It is also interesting to note how, as expected, the maximum cell fitness obtained is consistently higher in samples with hypermutation phenotype (Figure 9B).



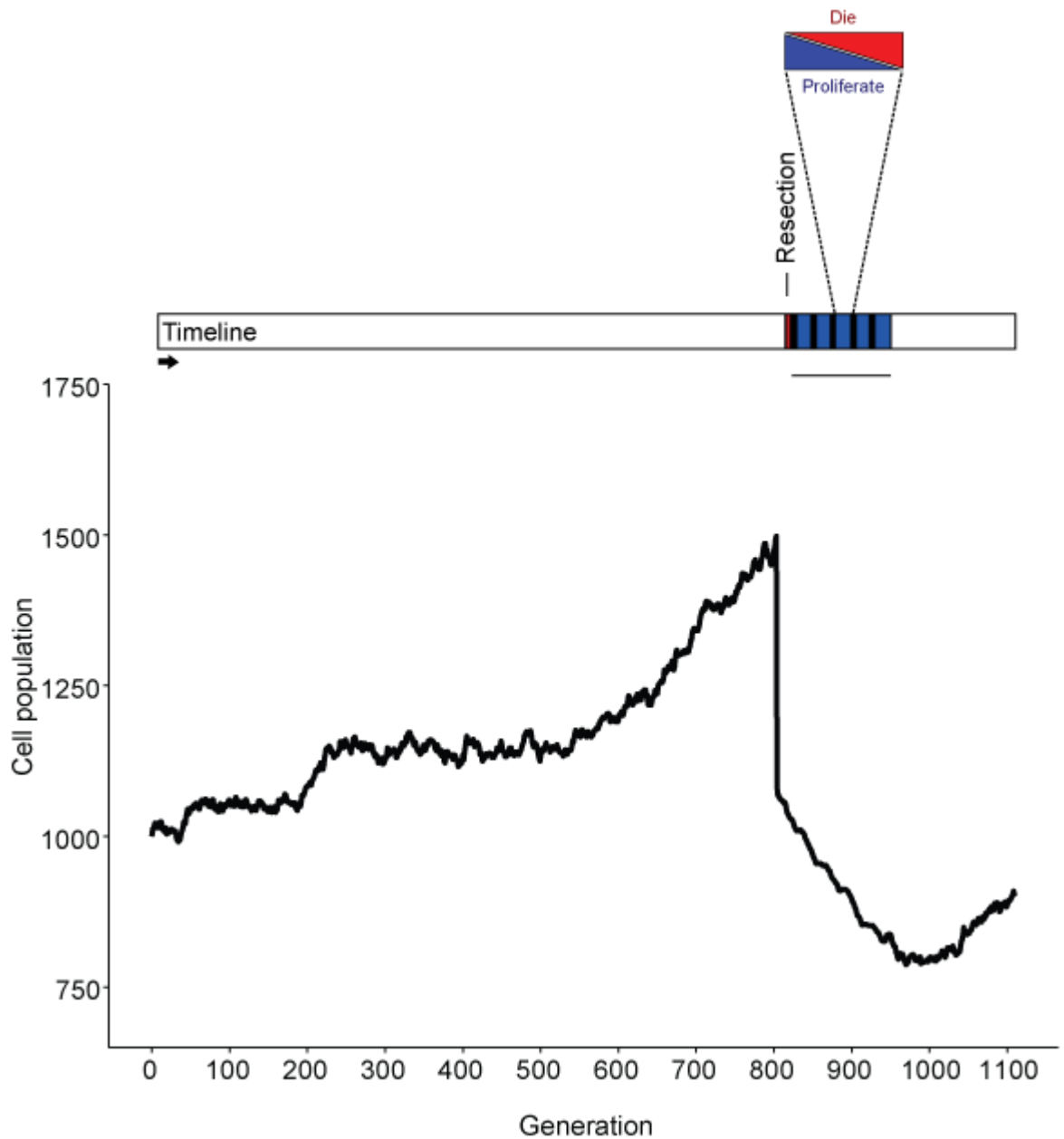
**Figure 9: Mutations in DNA repair genes boost the number of mutations that happens per cell division.** A)  $ATR_{mut}$  mutation increases the number of mutations per division by 8-fold. B) Distribution of fitness increases due to the surge in mutation per division caused by a mutation in a DNA repair pathway gene.

## Treatment

Lastly, it is important to offer a glimpse of how the modeled tumor can be treated. Cancer treatment depends on several factors including: tissue affected, stage, and size, among others. Usually a triad composed by surgery, radiotherapy, and chemotherapy is used. OncoProSim uses a simple model of treatment where those three treatment pillars are simulated.

After reaching a threshold (set by the user), the treatment begins by simulating resection, removing any cell that has at least one mutation. The extent of resection will be defined by the user and the mutated cells will be removed randomly. After that, a new table, specific for treatment, is used as input for the user to simulate chemotherapy, where probabilities of proliferation and death will be changed only during the duration of the treatment. This process will be repeated  $n$  times to represent the number of treatment cycles.

Also, it is known that chemotherapy might be involved in the evolution of tumor by increasing the rate that mutations are inserted into the cell DNA. Temozolomide, for example, may be an example of chemotherapy that is thought to be responsible for therapy driven evolution of glioma (38). Therefore, OncoProSim is able to increase the mutational rate during the course of treatment. It is important to note that the chemotherapy phase of treatment also affects non-mutated cells, as an indicative of therapy toxicity (Figure 10).



**Figure 10: Example of treatment simulation.** Treatment is triggered when the cell number reaches a value of 500 cells above the initial cell population and comprises basically three steps. First, surgery where only cells affected by mutations are removed (in this example 90% of the affected cells). Second, cells have their proliferation and death probabilities altered according to the values established by the therapy spreadsheet. Lastly, the second stage is repeated  $n$  times

to simulate several cycles of treatment. Non-mutated cells are also affected by the second stage.

## **Discussion**

As already noted in the discussion of several works prior to ours (8, 11, 39), computer modeling of tumor still has various limitations, and our work is no different. In order to mimic a complex disease, where so many signaling pathways are altered, simplifications are still needed. It is important to note that we do not aim to substitute any of those models but to implement the knowledge already given by them and help build a better model for tumor initiation.

We opted for an avascular model of tumor evolution, as we believe that angiogenesis is not a key factor during the initial stages of a tumor. During these initial steps of tumorigenesis, it is reasonable to conceive that all cells would have all necessary conditions to survive under normal homeostasis, though we do not underestimate the importance of vasculogenesis to cancer progression. Also, some models have already elegantly implemented with relevant conclusions as noted by Spencer *et al.*, where the angiogenesis characteristic of some cells can be exploited by other cells to boost their proliferation (9). Also, there are many other works that specifically aim to mathematically or computationally model vascular tumors and are reviewed in (40, 41).

OncoProSim strongest characteristic resides in its fully customizable interface, that allows users to simulate different types of tumors with a custom spreadsheet that specifies which genes might receive genomic events. This way, new information regarding tumorigenesis published in the literature can easily be incorporated to the simulation. The CancerSim software, developed by Abbot *et al.* and used by Spencer *et al.* remarkably used the six Hallmarks of Cancer as the key signaling pathways that could be altered in the simulation (8, 9). However, the hallmarks themselves were updated in 2011 to include two new features and two new emerging



hallmarks. Therefore, since OncoProSim resorts to user provided gene tables, new characteristics could be readily added to the software.

Another important aspect of OncoProSim is that multiplicative values for proliferation and death are used to establish the fitness of a cell. Other works have tried their hands on this concept. Bozic *et al.* have used a simple fixed fitness gain for each acquired driver mutation (11). Though their model nicely fits the evolution of some tumors, we find it unlikely that mutations of different genes would always return the same fitness advantage to the cell, especially when it is known that different mutations may have distinct impacts on the protein function (42). Foo *et al.* generated a distribution of mutational fitness where each mutated cell would receive a random fitness gain (39), This does not provide control on how key driver genes such as P53 might have a higher weight on fitness impact. However, they did introduce an interesting concept where the cell fitness is their primary threshold of tumor initiation. OncoProSim also provides the user the possibility of setting a fitness threshold for tumorigenesis, in addition to the possibility of using the cell population as a threshold. Interestingly, OncoProSim allows the user to combine both triggers and check what was the maximum fitness reached by at least one cell when the cell population trigger was reached.

We hope that OncoProSim will be used to create networks that will contribute to gain a proper knowledge of how much a specific gene mutation affects cell fitness. Though the establishment of how much of a fitness gain each driver mutation provides will remain a challenge. Particularly when it is known that different driver genes will affect the cell fitness in a different manner and that the same mutation might have a completely distinct effect in two separate individuals due to differences in their genetic background (43) and environment exposure.

Our model only grants that minor somatic events are simulated, even though several major somatic events are behind cancer initiation or progression. In chronic myelogenous leukemia, it is widely known that a translocation is present in virtually all cases (44) and is

proven to be a key initiation factor of this disease (45); In glioblastoma, copy number variations such as the gain of chromosome 7 and loss of chromosome 10 are believed to be the first genetic events for all non G-CIMP tumors (46) and even catastrophic chromosomal fragmentation events such as chromotripsis-like patterns have been identified in several types of cancer (47). Events like translocations or huge chromosomal number variations are still difficult to implement together with minor events in a computational model. One prominent feature of bioinformatics is that all those limitations can be easily converted into features in a more complex version of the software. Therefore, we strongly encourage the scientific community to use and improve upon our model.

To the best of our knowledge, our model is the first to simulate the diploid condition of the human genome, which results in vital information, especially in cancer. A tumor suppressor gene will normally only unleash its effects on cell fitness if both alleles are mutated (hereafter characterizing loss-of-function), whereas a proto-oncogene will affect cell fitness and become an oncogene with an event in just one of the gene alleles (hereafter characterizing gain-of-function). More importantly, in our model, an oncogene can be setup to have two different effect levels; a smaller or even negative effect in the absence of any tumor suppressor gene mutated beforehand in the same cell; and a second, stronger effect, if the cell already contains some tumor suppressor gene mutated. This situation resembles colon carcinoma evolution where the loss of APC and subsequent KRAS/BRAF mutation results in adenoma and after TP53/PTEN inactivation it results in a full blown carcinoma (48). Moreover, it also allows RAS mutations before any tumor suppressor inactivation to result in cell death, exactly like oncogene induced senescence (49), an important endogenous anti-tumor suppressor mechanism (50).

Though we understand that computer simulations are still in their infancy and are just a model, it is important to remember that even established methods, such as cell culture, barely resemble the reality, and the difference between cell culture cancer cells and primary cancer cells can be clearly seen in the literature (28, 51). Nevertheless, in the same way that

mathematical cancer modeling can act as an intermediate role (5), we strongly believe that computational modeling of cancer provide a reliable and faster way to produce insights regarding tumor evolution.

Another vital aspect of the model here presented is that we sought to permit the software to be as adaptable as possible so that users could simulate several different kinds of tumors. We recognize that this is highly ambitious, especially when our primary concern is to model a disease that looks to be one of its kind in each patient. However, since the user has control of almost all input features, from the genes that are going to be mutated until the chosen regimen of treatment, our model is certainly capable of simulating not only glioblastoma, our model of choice, but in fact different types of cancers.

Hereby, we endeavor to generate some insight into the molecular basis of tumor evolution and it is certainly feasible to imagine a translational application in the near future of our model. Though we do not claim that the use of OncoProSim will be able to predict tumor evolution history during these initial stages, it is certainly feasible to envisage that capacity in the future versions. With the likelihood of cheaper DNA sequencing machines flooding the clinical ambient, it will be viable for the oncologist to input data resulting from patient cancer biopsy and, using our model, be able to reconstruct parts of the evolutionary history of that tumor. Moreover, it is possible that with the sequencing of a patient tumor and the knowledge of its current mutations, OncoProSim could traceback as well as predict the next steps in that particular tumor evolution. Such a strategy would be especially helpful to decide the regimen of treatment in a personalized manner and might help the clinician in the choice of treatment regimen.

## **Methods**

### **The Code**

The source code was written in C++, with the help of the QT library for creating the interface. The paradigm used was the Object-oriented programming (OOP). In this Section we are going to explain the logic that runs under our program. We will start by defining the main objects of the system, and, after that, present the main routines of the program.

### **Main objects**

The program can be basically divided into four main types of objects: CellSystem, Cell, MutationTable, and Treatment.

CellSystem: this is the main object of the program, and the program only has one instance of this type. It stores all Cells of the system, the MutationTable, the manually-specified mutation rules, the Treatment, the random number-generator engine, and all intrinsic characteristics of the whole system (such as population history, current generation, number of affected cells, and others).

Cell: correspond to a cell inside the system. Each instance of this type will correspond to a single cell, storing information such as: life time, number of telomeres, probability of reproduction, probability of death, list of mutations, and structures representing the (first and second) strands.

MutationTable: holds information about the table loaded by the user. It stores information about all the mutations listed on the loaded file, along with the name of the mutations, number of basis, type of mutation, and modifiers for probabilities (when affecting a cell) for reproduction, death, and mutagen effects.

Treatment: is the object that holds information about the Treatment (loaded by the user), such as: threshold for treatment start, extent of the resection, duration of the treatment, number of cycles, and probabilities modifiers.

**Main Routines:** We now describe the main routines of the program, starting by 'processGeneration()', which is the one that the CellSystem runs for each generation:

This routine runs each time we want to advance in the generation counter. Note that the CellSystem calls each cell's routine 'tryReproduce()', and test if the treatment is running or if it should start on the current generation. Now we present the routine that runs for each cell:

Note how each cell will perform this routine taking its own probabilities (which may or not be modified by the treatment). Then, by asking to the CellSystem a random probability number (a number between [0, 1] generated by the pseudonumber-generator engine), the cell check if a new cell will be generated (whenever the cell has available telomeres). The same happens to check if the cell should die. Whenever a cell proliferates, a new cell is added to the system, by using the following procedure:

---

**Algorithm 1** CellSystem::processGeneration()

---

```
for each Cell c in ArrayOfCells[ ] do
  c.tryReproduce()
end for
if (Treatment.treatmentEnable()) then
  if (Treatment.duringTreatment()) then
    Treatment.incrementDayOfCycle()
  else
    Treatment.checkIfStart()
    if (Treatment.hasStarted()) then
      Treatment.performResection()
    end if
  end if
end if
populationHistory.push(numberOfCells)
affectedHistory.push(numberOfAffecteds)
generation++
```

---

---

**Algorithm 2** Cell::tryReproduce( )

---

```
lifeTime++
currentProlProb = Cell.getProlProby()
currentDeathProb = Cell.getDeathProb()
if (CellSystem.isDuringTreatment()) then
    currentProlProb *= Treatment.getProlModifier()
    currentDeathProb *= Treatment.getDeathModifier()
end if
if (CellSystem.getRandomProbNumber() < currentProlProby) then
    if (telomeresi > 0) then
        telomeresi -
        CellSystem.addNewCell(this)
    end if
end if
if (CellSystem.getRandomProbNumber() < currentDeathProb) then
    CellSystem.informNaturalDeath(this)
end if
```

---

---

**Algorithm 3** CellSystem::addNewCell(CellFather)

---

```
Require: CellFather
newCell.copy(CellFather)
prolModifier = 1
deathModifier = 1
if (Treatment.duringTreatment()) then
    prolModifier *= Treatment.getProlModifier()
    deathModifier *= Treatment.getDeathModifier()
end if
for i=0 to mutationPerReproduction do
    basisSelected = CellSystem.getRandomNumber()
    mutSelected = MutationTable.getMutation(basisSelected)
    if (mutSelected is valid) then
        mutType = MutationTable.getMutationType(mutSelected)
        strandSelected = CellSystem.getRandomNumber()%2
        if (strandSelected=1) then
            firstStrand.addMutation(mutSelected)
        else
            secondStrand.addMutation(mutSelected)
        end if
        if (MutationTable.appearedNewMutation(mutSelected,mutType,firstStrand,secondStrand))
        then
            if (newCell.hasTumorSuppressor()) then
                prolModifier *= mutationTable.getProlModifierAfterSup(mutSelected)
                deathModifier *= MutationTable.getDeathModifierAfterSup(mutSelected)
            else
                proliferationModifier *= MutationTable.getProlModifierBeforeSup(mutSelected)
                deathModifier *= MutationTable.getDeathModifierBeforeSup(mutSelected)
            end if
        end if
        newCell.updateProlRate(prolModifier)
        newCell.updateDeathRate(deathnModifier)
    end if
end for
CellSystem.registerCell(newCell)
numberOfCells++
```

---

These are the main routines of the program. Note how the pseudorandom numbers determine when the cell will reproduce or die, and also what basis and which strand will happen the mutation. In those pseudocode we have abstracted lot of fine details such as memory control, objects interaction and optimization modifications.

### **Databases used**

Significant mutated genes for glioblastoma and for breast cancer were taken from (25, 29, 52). Number of oncogenomic bases for each gene was obtained in the following manner: Data from TCGA level 2 somatic mutations for breast cancer, glioblastoma, and colon adenocarcinoma were downloaded from tcga data portal (<https://tcga-data.nci.nih.gov/tcga/>). Then, unique mutations for each gene were counted with the help of R software version 3.2.1 (53). The sum of all possible mutations was used as the oncogenome for each simulation.

### **Tumor Suppressor Genes and Oncogene classification**

Classification of genes by either tumor suppressor gene or oncogene for the examples was adapted from what was described in (11). Briefly, a ratio of inactivating mutations (frame shift deletions or insertions, nonsense, nonstop, and splice site) by activating mutations (in frame deletions or insertions, missense, and silent) was calculated. Values above 0.2 were considered tumor suppressor genes whereas values below 0.2 were considered oncogenes.

### **Results output**

We provide the user with a plethora of output results that can be used to evaluate the outcome of the performed simulations. Recorded parameters and their meaning are described in (Table 3), and several of those presented here will offer a glimpse of how they can be used to provide important resources to tumor biology.

**Table 1.** Input parameters accepted by OncoProSim



	<b>Parameter</b>	<b>Action</b>
Essential parameters	Seed	Controls the pseudorandom number generator
	Number of cells	Starting number of cells
	Mutations per division	Number of mutations that will be inserted into each cell at each division
	Probability of proliferation	Default probability of proliferation for all cells
	Probability of death	Default probability of proliferation for all cells
	Maximum divisions	Number of divisions that a cell can go through
	Genome size	Size of the two matrices that will correspond to the cell genome.
Automatic runs	Number of generations	Number of maximum generations allowed per seed
	Parameter to iterate	Parameter that will be automatically iterated
	Minimum value	Starting iterated parameter value
	Maximum value	Ending iterated parameter value

Threshold parameters	Population fitness	Records in which generation population has reached the specified fitness
	Cell fitness	Records in which generation a single cell has reached the specified fitness
	Absolute population	Maximum cell population number
	Sampling rate	Interval in which OncoProSim will record the desired conditions
Treatment	Absolute cell threshold	Number above the starting number of cells when treatment will be started
	Extent of resection (%)	Simulates surgery. Removes the desired percentage of only cells that contains mutations within genes
	Duration of each cycle (days)	Probability of proliferation and death are divided by the treatment cycle in days to transform the generations in days
	Number of cycles	Number of cycles that the treatment will be applied

---

**Table 2.** Input tables

	<b>Parameter</b>	<b>Action</b>
Mutation table	Proliferate without suppressor	Multiplicative proliferation value for a given gene when at least one tumor suppressor gene is mutated
	Proliferate with suppressor	Multiplicative proliferation value for a given gene when there are no tumor suppressor genes mutated
	Die without suppressor	Multiplicative death value for a given gene when at least one tumor suppressor gene is mutated
	Die without suppressor	Multiplicative death value for a given gene when there are no tumor suppressor genes mutated
	Mutate	Multiplies the number of mutations per generation in case the gene is mutated
	Oncogenic bases	Number of bases that can receive a mutation for a given gene
	Oncogenomic size	Sum of all the oncogenic bases introduced by the oncogenic bases column

	Gene status	Classification of the gene: either oncogene or tumor suppressor gene
	Maximum division multiplier without suppressor	Multiplicative value for the maximum divisions parameter when at least one tumor suppressor gene is mutated
	Maximum division multiplier with suppressor	Multiplicative value for the maximum divisions parameter when there are no tumor suppressor genes mutated
Therapy table	Day	Each row represents one day of treatment
	Treatment proliferation multiplier	Multiplicative value for proliferation parameter active only during the course of treatment
	Treatment death multiplier	Multiplicative value for death parameter active only during the course of treatment
	Treatment mutation multiplier	Multiplies the number of mutations per generation during the course of treatment

---

**Table 3.** Output parameters by OncoProSim

<b>Output</b>	<b>Data presented</b>
Automatic runs information	All the input parameters by seed; if any threshold was set returns the generation that the parameter was surpassed; returns maximum population and cell fitness obtained during the simulation; returns treatment generation start, end, and the number of cells removed during the resection
Mutations percentage	Returns the frequency of mutation by gene for each seed in absolute values
Intermediate results	Returns the absolute number of cells that present a given gene mutation by generation
Histogram results	Returns the number of mutations per cell per generation

## REFERENCES

1. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer*. 1954;8(1):1-12.
2. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*. 1971;68(4):820-3.
3. Byrne HM, Alarcon T, Owen MR, Webb SD, Maini PK. Modelling aspects of cancer dynamics: a review. *Philosophical transactions Series A, Mathematical, physical, and engineering sciences*. 2006;364(1843):1563-78.
4. Roose T, Chapman S, Maini P. *Mathematical Models of Avascular Tumor Growth*. *SIAM Review*. 2007;49(2):179-208.
5. Byrne HM. Dissecting cancer through mathematics: from the cell to the animal model. *Nature reviews Cancer*. 2010;10(3):221-30.
6. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.
7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
8. Abbott RG, Forrest S, Pienta KJ. Simulating the hallmarks of cancer. *Artificial life*. 2006;12(4):617-34.
9. Spencer SL, Gerety RA, Pienta KJ, Forrest S. Modeling somatic evolution in tumorigenesis. *PLoS computational biology*. 2006;2(8):e108.
10. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, et al. Genetic progression and the waiting time to cancer. *PLoS computational biology*. 2007;3(11):e225.
11. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(43):18545-50.

12. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. *Experimental cell research*. 1961;25:585-621.
13. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature*. 1990;345(6274):458-60.
14. Campisi J. The biology of replicative senescence. *European journal of cancer*. 1997;33(5):703-9.
15. Campisi J, d'Adda di Fagagna F. Cellular senescence: when bad things happen to good cells. *Nature reviews Molecular cell biology*. 2007;8(9):729-40.
16. Kuilman T, Michaloglou C, Mooi WJ, Peeper DS. The essence of senescence. *Genes & development*. 2010;24(22):2463-79.
17. Shih C, Weinberg RA. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell*. 1982;29(1):161-9.
18. Bos JL, Verlaan-de Vries M, Jansen AM, Veeneman GH, van Boom JH, van der Eb AJ. Three different mutations in codon 61 of the human N-ras gene detected by synthetic oligonucleotide hybridization. *Nucleic acids research*. 1984;12(23):9155-63.
19. Bos JL. ras oncogenes in human cancer: a review. *Cancer research*. 1989;49(17):4682-9.
20. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. *Nature reviews Cancer*. 2011;11(11):761-74.
21. Weinberg RA. Tumor suppressor genes. *Science*. 1991;254(5035):1138-46.
22. Macleod K. Tumor suppressor genes. *Current opinion in genetics & development*. 2000;10(1):81-93.
23. Knudson AG. Two genetic hits (more or less) to cancer. *Nature reviews Cancer*. 2001;1(2):157-62.
24. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994-1007.

25. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155(2):462-77.
26. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. *Cell*. 2012;150(2):251-63.
27. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90-4.
28. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396-401.
29. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330-7.
30. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-8.
31. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*. 2009;457(7229):608-11.
32. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell*. 1996;87(2):159-70.
33. Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. *Cell*. 2000;103(2):311-20.
34. Phelps RA, Chidester S, Dehghanizadeh S, Phelps J, Sandoval IT, Rai K, et al. A two-step model for colon adenoma initiation and progression caused by APC loss. *Cell*. 2009;137(4):623-34.
35. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546-58.
36. Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome research*. 2015;25(3):316-27.



37. Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, et al. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS genetics*. 2012;8(7):e1002772.
38. Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*. 2014;343(6167):189-93.
39. Foo J, Leder K, Michor F. Stochastic dynamics of cancer initiation. *Physical biology*. 2011;8(1):015002.
40. Mantzaris NV, Webb S, Othmer HG. Mathematical modeling of tumor-induced angiogenesis. *Journal of mathematical biology*. 2004;49(2):111-87.
41. Chaplain MA, McDougall SR, Anderson AR. Mathematical modeling of tumor-induced angiogenesis. *Annual review of biomedical engineering*. 2006;8:233-57.
42. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*. 2011;39(17):e118.
43. Sidow A, Spies N. Concepts in solid tumor evolution. *Trends in genetics : TIG*. 2015.
44. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;243(5405):290-3.
45. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science*. 1990;247(4944):824-30.
46. Ozawa T, Riester M, Cheng YK, Huse JT, Squatrito M, Helmy K, et al. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer cell*. 2014;26(2):288-300.

47. Cai H, Kumar N, Bagheri HC, von Mering C, Robinson MD, Baudis M. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *BMC genomics*. 2014;15:82.
48. Jones S, Chen WD, Parmigiani G, Diehl F, Beerewinkel N, Antal T, et al. Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(11):4283-8.
49. Mason DX, Jackson TJ, Lin AW. Molecular signature of oncogenic ras-induced senescence. *Oncogene*. 2004;23(57):9238-46.
50. Lenz G. Endogenous anticancer mechanisms (EACMs). *Frontiers in bioscience*. 2012;4:1017-30.
51. Li A, Walling J, Kotliarov Y, Center A, Steed ME, Ahn SJ, et al. Genomic changes and gene expression profiles reveal that established glioma cell lines are poorly representative of primary human gliomas. *Molecular cancer research : MCR*. 2008;6(1):21-30.
52. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
53. R Development Core Team. *R: A Language and Environment for Statistical Computing*. 2008.

## DISCUSSÃO

Uma doença com uma intrincada biologia como o câncer e, ainda, com uma grande parte de seu funcionamento desconhecida, traz uma infinidade de características complexas e de difícil estudo. Por isso, alguns desses aspectos somente agora começam a ser esclarecidos pela ciência, especialmente com projetos como o TCGA, o qual produziu informações até então desconhecidas quanto ao número e a frequência de mutações em distintos tipos tumorais (Parsons *et al.*, 2008; Pleasance *et al.*, 2010; Cancer Genome Atlas, 2012; Cancer Genome Atlas Research, 2012; Brennan *et al.*, 2013; Lawrence *et al.*, 2014)

Ao simular a evolução tumoral, partimos do conhecimento obtido por uma etapa intermediária ou avançada do estadiamento tumoral, obtida através de uma amostra de biópsia de um paciente e, retroativamente, tentamos montar a sequência de eventos (epi)genéticos que construiu a história evolutiva da doença. Assim, tentamos realizar uma regressão dos eventos prévios a uma dada etapa tumoral, e aumentar o conhecimento atual, utilizando como aliado a compreensão dos mecanismos tumorais já obtidos pela biologia experimental.

Diversos projetos já tentaram descrever, matematicamente e computacionalmente, a história evolutiva de tumores (Spencer *et al.*, 2004; Abbott *et al.*, 2006; Spencer *et al.*, 2006; Beerenwinkel *et al.*, 2007; Gerlee e Anderson, 2008; Bozic *et al.*, 2010; Reiter *et al.*, 2013). Entretanto, assim como em nosso modelo, todos os projetos, admitidamente, recorrem a simplificações e priorizam diferentes características encontradas em tumores, visto que um modelo 100% fidedigno a realidade ainda se apresenta como um enorme desafio.

Spencer e colaboradores utilizaram um software desenvolvido por Abbott e colaboradores que prioriza a simulação de um modelo vascular dos chamados *Hallmarks of Cancer*. Trata-se de um elegante modelo, que demonstra o poder dos programas em produzir resultados com ganho de conhecimento biológico, tais como a importância de mutações que estimulam a angiogênese em estabelecer o ambiente necessário para que outros clones proliferem, porém, ele peca em assumir que somente seis mutações de vias de sinalização centrais são suficientes para simular as

diferentes características tumorais existentes (Hanahan e Weinberg, 2000; Abbott *et al.*, 2006; Spencer *et al.*, 2006). Para aumentar a verossimilhança com tumores reais buscamos programar OncoProSim em torno de tabelas personalizadas pelo usuário, assim, qualquer via de sinalização pode ser atingida, contanto que os genes necessários para essa condição estejam presentes na tabela. Os próprios *Hallmarks of Cancer* foram atualizados em 2011 para inclusão de mais duas características fundamentais e duas características emergentes (Hanahan e Weinberg, 2011), a utilização de uma tabela facilmente customizável contorna essas situações, visto que torna possível facilmente implementar novas descobertas descritas pela literatura.

Bozic e colaboradores demonstraram que modelos computacionais vão ao encontro de dados experimentais quando, em seu modelo teórico, a comparação do número de mutações *drivers* com o número de mutações *passengers* existentes em glioblastoma e câncer de pâncreas, foi similar a encontrada em dados reais. Ainda, o modelo teórico foi capaz de demonstrar como novos clones aparecem em um intervalo de tempo menor quando comparados a geração anterior, sendo proporcional ao aumento do número de mutações *drivers* na linhagem clonal (Bozic *et al.*, 2010). Em nosso modelo, a reprodução de fenótipo hipermutador confirma os achados de que o aparecimento de novos clones ocorrem em um intervalo de tempo menor conforme o número de mutações existentes aumenta.

No mesmo trabalho, Bozic e colaboradores fixaram o ganho de fitness dado por uma nova mutação em aproximadamente 0.4% de acordo com as deduções de seu modelo a partir de dados de glioblastoma e câncer pancreático (Bozic *et al.*, 2010). Entretanto, acreditamos ser vital que diferentes mutações possuam valores distintos que afetem o *fitness* celular. Conforme várias ferramentas demonstram, diferentes mutações podem apresentar distintos impactos na função das proteínas de acordo com o códon que foi afetado (Reva *et al.*, 2011). Foo e colaboradores também buscaram resolver a questão do *fitness* celular utilizando um diferente esquema, em que toda vez que uma célula recebe uma mutação, um valor aleatório de *fitness* é retirado de uma distribuição probabilística. Apesar de ser um modelo interessante, novamente ele carece de um controle maior sobre o quanto as mutações impactam a proliferação ou morte celular, como o proporcionado por OncoProSim. Todavia, o mesmo trabalho

propõe um interessante conceito, em que a iniciação tumoral acontece quando uma célula atinge um determinado valor de *fitness* celular (Foo *et al.*, 2011), esse mesmo parâmetro pode ser adotado em OncoProSim visto que o momento em que a célula atinge um determinado limite de *fitness* é registrado e pode ser consultado pelo usuário.

Entretanto, também utilizamos de simplificações que limitam a verossimilhança do modelo com a complexidade existente na biologia tumoral. Primeiramente, optamos por um modelo avascular de simulação tumoral. Apesar de ser uma característica vital para a progressão de tumores sólidos, nosso objetivo principal consiste em demonstrar as etapas iniciais de desenvolvimento da doença, em que a célula precursora tumoral dispõe de todos os recursos necessários à sua proliferação. Além disso, uma grande quantidade de modelos, principalmente matemáticos, preocupam-se exclusivamente em entender como a angiogênese e a difusão de nutrientes atua no ambiente tumoral (Chaplain e Stuart, 1993; Orme e Chaplain, 1997).

Para aumentar a eficiência das simulações, o número inicial de células ainda está aquém dos estimados. Por exemplo, para glioblastoma, em que evidências apontam para células precursoras de oligodendrócitos como as células de origem tumoral (Liu *et al.*, 2011), estimadas em ter uma prevalência de 3 a 4% do total de células do sistema nervoso central de um total de aproximadamente 84,6 bilhões de células não neuronais (Dawson *et al.*, 2003; Azevedo *et al.*, 2009), deveríamos modular 3 a 4 bilhões de células. Trabalhar com esse número de células, mesmo com as máquinas modernas, consome muito tempo e recursos computacionais, reduzindo a eficiência com que as simulações podem ser realizadas, e tornando inviável para pesquisadores sem acesso a computadores de última geração.

Ainda, nosso modelo simula apenas pequenos eventos mutacionais. Alguns tumores já foram descritos como originados por mutações genômicas de grande porte, por exemplo, a translocação do cromossomo 9 e 22 presente na grande maioria dos casos de leucemia mielóide crônica (Rowley, 1973) ou o aumento do número de cópias de cromossomos inteiros como descrito para glioblastoma (Ozawa *et al.*, 2014). Um modelo que combina translocações e ampliações ou deleções de grande porte ainda apresenta dificuldades técnicas de difícil reprodução quando comparadas a um modelo de pequenos eventos somáticos.

Apesar dessas limitações, estamos confiantes de que nosso modelo representa um avanço nos modelos de evolução tumoral. Com ele, é possível customizar o tipo tumoral modelado e quais genes serão possivelmente afetados pelas mutações, ainda, cada mutação confere um peso único para o *fitness* celular e, mutações em genes supressores tumorais e oncogenes são distintas e podem atuar de maneira sinérgica. Para preencher as lacunas existentes em modelos apresentados anteriormente, OncoProSim foca em simular uma das raízes dos tumores, os eventos mutacionais que ocorrem ao decorrer do tempo no genoma celular, com uma interface extremamente customizável e intuitiva que permite ao usuário simular diferentes tipos tumorais e, assim, viabilizar o estudo da interferência de diferentes aspectos da biologia tumoral na evolução da doença. Ainda, pretendemos disponibilizar um modelo completo em que até mesmo a terapia e a reincidência tumoral possam ser estudadas.

É importante ressaltar que, aqui, não almejamos substituir os modelos já existentes e acima descritos, mas sim agregar ao conhecimento já produzido e, é preciso lembrar que técnicas experimentais como cultura de células também sofrem com limitações e são, relativamente, distante da complexidade apresentada pela biologia tumoral em pacientes (Li *et al.*, 2008; Patel *et al.*, 2014). Por último, cabe destacar a grande vantagem da biologia tumoral *in silico*, na qual toda e qualquer limitação pode ser contornada com eficiência e velocidade uma vez que novas idéias podem ser rapidamente implementadas e adaptadas ao modelo.

## **12. PERSPECTIVAS**

Modelos *in silico* permitem, com rapidez e eficiência, a implementação de novas características e funções. Várias das características aqui mencionadas como limitações podem, e devem, ser implementadas em versões futuras.

Em curto prazo, pretendemos aprimorar o módulo de tratamento, com a personalização da terapia em que se busca atacar a característica específica do tumor de cada paciente, é importante que nosso modelo permita também essa customização de acordo com as mutações que aconteceram na evolução tumoral.

Outros passos promissores é a integração do modelo com programas de predição de impacto específico de cada mutação.

Ainda, por ser uma ferramenta *open source*, a utilização do OncoProSim pela comunidade acadêmica poderá contribuir imensamente com o desenvolvimento e melhora do modelo através do uso e sugestões críticas.

Por último, em longo prazo, esperamos chegar a um patamar em que o programa, uma vez que em posse da sequência genômica de uma biópsia, possa auxiliar em traçar a história evolutiva daquele tumor e, assim, auxiliar o oncologista na escolha de fármacos específicos para o paciente de acordo com a probabilidade de mutações e clones determinada por OncoProSim.

### **13. CONCLUSÕES**

Aqui, apresentamos um modelo altamente customizável capaz de mimetizar as mutações presentes no genoma de diferentes tipos tumorais e, também, determinar o peso com que cada mutação influencia no *fitness* da célula que a recebeu. Assim, é possível simular diferentes histórias evolutivas tumorais que, futuramente, podem auxiliar o oncologista na escolha dos diferentes tratamentos para tumores.

OncoProSim é uma ferramenta simples e intuitiva para simulação da evolução tumoral e com grande potencial adaptativo para diferentes tipos tumorais.

## 14. REFERÊNCIAS

ABBOTT, R. G.; FORREST, S.; PIENTA, K. J. Simulating the hallmarks of cancer. **Artif Life**, v. 12, n. 4, p. 617-34, Fall 2006.

ADAM, J. A.; MAGGELAKIS, S. A. Mathematical models of tumor growth. IV. Effects of a necrotic core. **Math Biosci**, v. 97, n. 1, p. 121-36, Nov 1989.

AKTIPIIS, C. A. et al. Life history trade-offs in cancer evolution. **Nat Rev Cancer**, v. 13, n. 12, p. 883-92, Dec 2013.

ARAUJO, R. P.; MCELWAIN, D. L. A history of the study of solid tumour growth: the contribution of mathematical modelling. **Bull Math Biol**, v. 66, n. 5, p. 1039-91, Sep 2004.

ARMITAGE, P.; DOLL, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. **Br J Cancer**, v. 8, n. 1, p. 1-12, Mar 1954.

ARMITAGE, P.; DOLL, R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. **Br J Cancer**, v. 11, n. 2, p. 161-9, Jun 1957.

ASHLEY, D. J. The two "hit" and multiple "hit" theories of carcinogenesis. **Br J Cancer**, v. 23, n. 2, p. 313-28, Jun 1969.

AZEVEDO, F. A. et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. **J Comp Neurol**, v. 513, n. 5, p. 532-41, Apr 10 2009.

BARKER, N. et al. Crypt stem cells as the cells-of-origin of intestinal cancer. **Nature**, v. 457, n. 7229, p. 608-11, Jan 29 2009.

BEERENWINKEL, N. et al. Genetic progression and the waiting time to cancer. **PLoS Comput Biol**, v. 3, n. 11, p. e225, Nov 2007.

BLANPAIN, C. Tracing the cellular origin of cancer. **Nat Cell Biol**, v. 15, n. 2, p. 126-34, Feb 2013.

BOZIC, I. et al. Accumulation of driver and passenger mutations during tumor progression. **Proc Natl Acad Sci U S A**, v. 107, n. 43, p. 18545-50, Oct 26 2010.

BOZIC, I.; NOWAK, M. A. Cancer. Unwanted evolution. **Science**, v. 342, n. 6161, p. 938-9, Nov 22 2013.

BOZIC, I.; NOWAK, M. A. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. **Proc Natl Acad Sci U S A**, v. 111, n. 45, p. 15964-8, Nov 11 2014.

BRENNAN, C. W. et al. The somatic genomic landscape of glioblastoma. **Cell**, v. 155, n. 2, p. 462-77, Oct 10 2013.



BURRELL, R. A. et al. The causes and consequences of genetic heterogeneity in cancer evolution. **Nature**, v. 501, n. 7467, p. 338-45, Sep 19 2013.

BURTON, A. C. Rate of growth of solid tumours as a problem of diffusion. **Growth**, v. 30, n. 2, p. 157-76, Jun 1966.

CAI, H. et al. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. **BMC Genomics**, v. 15, p. 82, 2014.

CANCER GENOME ATLAS, N. Comprehensive molecular characterization of human colon and rectal cancer. **Nature**, v. 487, n. 7407, p. 330-7, Jul 19 2012.

CANCER GENOME ATLAS RESEARCH, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. **Nature**, v. 455, n. 7216, p. 1061-8, Oct 23 2008.

CANCER GENOME ATLAS RESEARCH, N. Integrated genomic analyses of ovarian carcinoma. **Nature**, v. 474, n. 7353, p. 609-15, Jun 30 2011.

CANCER GENOME ATLAS RESEARCH, N. Comprehensive genomic characterization of squamous cell lung cancers. **Nature**, v. 489, n. 7417, p. 519-25, Sep 27 2012.

CASCIARI, J. J.; SOTIRCHOS, S. V.; SUTHERLAND, R. M. Mathematical modelling of microenvironment and growth in EMT6/Ro multicellular tumour spheroids. **Cell Prolif**, v. 25, n. 1, p. 1-22, Jan 1992.

CHAPLAIN, M. A.; MCDUGALL, S. R.; ANDERSON, A. R. Mathematical modeling of tumor-induced angiogenesis. **Annu Rev Biomed Eng**, v. 8, p. 233-57, 2006.

CHAPLAIN, M. A.; STUART, A. M. A model mechanism for the chemotactic response of endothelial cells to tumour angiogenesis factor. **IMA J Math Appl Med Biol**, v. 10, n. 3, p. 149-68, 1993.

DAWSON, M. R. et al. NG2-expressing glial progenitor cells: an abundant and widespread population of cycling cells in the adult rat CNS. **Mol Cell Neurosci**, v. 24, n. 2, p. 476-88, Oct 2003.

DEAKIN, A. S. Model for the growth of a solid in vitro tumor. **Growth**, v. 39, n. 1, p. 159-65, Mar 1975.

DIAZ, L. A., JR. et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. **Nature**, v. 486, n. 7404, p. 537-40, Jun 28 2012.

DORMANN, S.; DEUTSCH, A. Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. **In Silico Biol**, v. 2, n. 3, p. 393-406, 2002.

FISCHER, A. et al. High-definition reconstruction of clonal composition in cancer. **Cell Rep**, v. 7, n. 5, p. 1740-52, Jun 12 2014.

FOLKMAN, J.; KLAGSBRUN, M. Angiogenic factors. **Science**, v. 235, n. 4787, p. 442-7, Jan 23 1987.

FOO, J.; LEDER, K.; MICHOR, F. Stochastic dynamics of cancer initiation. **Phys Biol**, v. 8, n. 1, p. 015002, Feb 2011.

FORBES, S. A. et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. **Nucleic Acids Res**, v. 43, n. Database issue, p. D805-11, Jan 2015.

FRANCIS, J. M. et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. **Cancer Discov**, v. 4, n. 8, p. 956-71, Aug 2014.

GERLEE, P.; ANDERSON, A. R. A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype. **J Theor Biol**, v. 250, n. 4, p. 705-22, Feb 21 2008.

GREAVES, M.; MALEY, C. C. Clonal evolution in cancer. **Nature**, v. 481, n. 7381, p. 306-13, Jan 19 2012.

HAENO, H. et al. Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. **Cell**, v. 148, n. 1-2, p. 362-75, Jan 20 2012.

HANAHAN, D.; WEINBERG, R. A. The hallmarks of cancer. **Cell**, v. 100, n. 1, p. 57-70, Jan 7 2000.

HANAHAN, D.; WEINBERG, R. A. Hallmarks of cancer: the next generation. **Cell**, v. 144, n. 5, p. 646-74, Mar 4 2011.

HILEY, C. et al. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. **Genome Biol**, v. 15, n. 8, p. 453, 2014.

HODIS, E. et al. A landscape of driver mutations in melanoma. **Cell**, v. 150, n. 2, p. 251-63, Jul 20 2012.

.

INCA. Estimativa 2014.

INTERNATIONAL CANCER GENOME, C. et al. International network of cancer genome projects. **Nature**, v. 464, n. 7291, p. 993-8, Apr 15 2010.

JIANG, Y. et al. A multiscale model for avascular tumor growth. **Biophys J**, v. 89, n. 6, p. 3884-94, Dec 2005.

JONES, S. et al. Comparative lesion sequencing provides insights into tumor evolution. **Proc Natl Acad Sci U S A**, v. 105, n. 11, p. 4283-8, Mar 18 2008.

KAMINKER, J. S. et al. CanPredict: a computational tool for predicting cancer-associated missense mutations. **Nucleic Acids Res**, v. 35, n. Web Server issue, p. W595-8, Jul 2007.

KANSAL, A. R. et al. Emergence of a subpopulation in a computational model of tumor growth. **J Theor Biol**, v. 207, n. 3, p. 431-41, Dec 7 2000.

KENNEDY, J. A. et al. Super-resolution in PET imaging. **IEEE Trans Med Imaging**, v. 25, n. 2, p. 137-47, Feb 2006.

KIM, H. et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. **Genome Res**, v. 25, n. 3, p. 316-27, Mar 2015.

KINZLER, K. W.; VOGELSTEIN, B. Lessons from hereditary colorectal cancer. **Cell**, v. 87, n. 2, p. 159-70, Oct 18 1996.

KNUDSON, A. G. Two genetic hits (more or less) to cancer. **Nat Rev Cancer**, v. 1, n. 2, p. 157-62, Nov 2001. ISSN 1474-175X (Print)  
1474-175X (Linking).

KNUDSON, A. G., JR. Mutation and cancer: statistical study of retinoblastoma. **Proc Natl Acad Sci U S A**, v. 68, n. 4, p. 820-3, Apr 1971.

LANG, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. **Nature**, v. 500, n. 7464, p. 571-4, Aug 29 2013.

LAWRENCE, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. **Nature**, v. 505, n. 7484, p. 495-501, Jan 23 2014.

LAWRENCE, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. **Nature**, v. 499, n. 7457, p. 214-8, Jul 11 2013.

LI, A. et al. Genomic changes and gene expression profiles reveal that established glioma cell lines are poorly representative of primary human gliomas. **Mol Cancer Res**, v. 6, n. 1, p. 21-30, Jan 2008.

LIU, C. et al. Mosaic analysis with double markers reveals tumor cell of origin in glioma. **Cell**, v. 146, n. 2, p. 209-21, Jul 22 2011.

MANTZARIS, N. V.; WEBB, S.; OTHMER, H. G. Mathematical modeling of tumor-induced angiogenesis. **J Math Biol**, v. 49, n. 2, p. 111-87, Aug 2004.

MARUSIC, M. et al. Analysis of growth of multicellular tumour spheroids by mathematical models. **Cell Prolif**, v. 27, n. 2, p. 73-94, Feb 1994.

MCFARLAND, C. D. et al. Impact of deleterious passenger mutations on cancer progression. **Proc Natl Acad Sci U S A**, v. 110, n. 8, p. 2910-5, Feb 19 2013.

MERLO, L. M. F. et al. Cancer as an evolutionary and ecological process. **Nat Rev Cancer**, v. 6, n. 12, p. 924-935, 12/print 2006.

MILLER, C. A. et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. **PLoS Comput Biol**, v. 10, n. 8, p. e1003665, Aug 2014.

- NAVIN, N. et al. Tumour evolution inferred by single-cell sequencing. **Nature**, v. 472, n. 7341, p. 90-4, Apr 7 2011.
- NAVIN, N. et al. Inferring tumor progression from genomic heterogeneity. **Genome Res**, v. 20, n. 1, p. 68-80, Jan 2010.
- NIK-ZAINAL, S. et al. The life history of 21 breast cancers. **Cell**, v. 149, n. 5, p. 994-1007, May 25 2012.
- ORME, M. E.; CHAPLAIN, M. A. Two-dimensional models of tumour angiogenesis and anti-angiogenesis strategies. **IMA J Math Appl Med Biol**, v. 14, n. 3, p. 189-205, Sep 1997.
- OZAWA, T. et al. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. **Cancer Cell**, v. 26, n. 2, p. 288-300, Aug 11 2014.
- PARSONS, D. W. et al. An integrated genomic analysis of human glioblastoma multiforme. **Science**, v. 321, n. 5897, p. 1807-12, Sep 26 2008.
- PATEL, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. **Science**, v. 344, n. 6190, p. 1396-401, Jun 20 2014.
- PHELPS, R. A. et al. A two-step model for colon adenoma initiation and progression caused by APC loss. **Cell**, v. 137, n. 4, p. 623-34, May 15 2009.
- PLEASANCE, E. D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. **Nature**, v. 463, n. 7278, p. 191-6, Jan 14 2010.
- PLEASE, C. P.; PETTET, G. J.; MCELWAIN, D. L. S. Avascular tumour dynamics and necrosis. **Mathematical Models and Methods in Applied Sciences**, v. 09, n. 04, p. 569-579, 1999.
- PON, J. R.; MARRA, M. A. Driver and passenger mutations in cancer. **Annu Rev Pathol**, v. 10, p. 25-50, Jan 24 2015.
- REITER, J. et al. TTP: Tool for Tumor Progression. In: SHARYGINA, N. e VEITH, H. (Ed.). **Computer Aided Verification**: Springer Berlin Heidelberg, v.8044, 2013. cap. 6, p.101-106. (Lecture Notes in Computer Science).
- REVA, B.; ANTIPIN, Y.; SANDER, C. Predicting the functional impact of protein mutations: application to cancer genomics. **Nucleic Acids Res**, v. 39, n. 17, p. e118, Sep 1 2011.
- ROOSE, T.; CHAPMAN, S.; MAINI, P. Mathematical Models of Avascular Tumor Growth. **SIAM Review**, v. 49, n. 2, p. 179-208, 2007/01/01 2007.
- ROTH, A. et al. PyClone: statistical inference of clonal population structure in cancer. **Nat Methods**, v. 11, n. 4, p. 396-8, Apr 2014.

ROWLEY, J. D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. **Nature**, v. 243, n. 5405, p. 290-3, Jun 1 1973.

SIDOW, A.; SPIES, N. Concepts in solid tumor evolution. **Trends Genet**, Feb 27 2015.

SIEGEL, R. et al. Cancer statistics, 2014. **CA Cancer J Clin**, v. 64, n. 1, p. 9-29, Jan-Feb 2014.

SOM, A.; WEN, S.; TU, S. M. Stem cell origin of testicular seminoma. **Clin Genitourin Cancer**, v. 11, n. 4, p. 489-94, Dec 2013.

SPENCER, S. L. et al. An ordinary differential equation model for the multistep transformation to cancer. **J Theor Biol**, v. 231, n. 4, p. 515-24, Dec 21 2004.

SPENCER, S. L. et al. Modeling somatic evolution in tumorigenesis. **PLoS Comput Biol**, v. 2, n. 8, p. e108, Aug 18 2006.

SPINNEY, L. Cancer: caught in time. **Nature**, v. 442, n. 7104, p. 736-8, Aug 17 2006.

STOKES, C. L.; LAUFFENBURGER, D. A. Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. **J Theor Biol**, v. 152, n. 3, p. 377-403, Oct 7 1991.

STRATTON, M. R.; CAMPBELL, P. J.; FUTREAL, P. A. The cancer genome. **Nature**, v. 458, n. 7239, p. 719-24, Apr 9 2009. ISSN 1476-4687

SWANSON, K. R.; ALVORD, E. C., JR.; MURRAY, J. D. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. **Br J Cancer**, v. 86, n. 1, p. 14-8, Jan 7 2002.

SWANSON, K. R.; ALVORD JR, E. C.; MURRAY, J. D. Virtual resection of gliomas: Effect of extent of resection on recurrence. **Mathematical and Computer Modelling**, v. 37, n. 11, p. 1177-1190, 6// 2003.

TOMASETTI, C.; VOGELSTEIN, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. **Science**, v. 347, n. 6217, p. 78-81, Jan 2 2015. ISSN 1095-9203.

VISVADER, J. E. Cells of origin in cancer. **Nature**, v. 469, n. 7330, p. 314-22, Jan 20 2011.

VOGELSTEIN, B. et al. Cancer genome landscapes. **Science**, v. 339, n. 6127, p. 1546-58, Mar 29 2013.

WANG, Z. A. et al. Luminal cells are favored as the cell of origin for prostate cancer. **Cell Rep**, v. 8, n. 5, p. 1339-46.

WARD, J. P.; KING, J. R. Mathematical modelling of avascular-tumour growth. II: Modelling growth saturation. **IMA J Math Appl Med Biol**, v. 16, n. 2, p. 171-211, Jun 1999.

YANG, J. et al. Adenomatous polyposis coli (APC) differentially regulates beta-catenin phosphorylation and ubiquitination in colon cancer cells. **J Biol Chem**, v. 281, n. 26, p. 17751-7, Jun 30 2006.